United States
Environmental Protection
Agency

Office of Research and
Development
Cincinnati, Ohio 45268

EPA/625/R-95/004
September 1995

EPA

Seminar Publication

National Conference on Environmental Problem-Solving with Geographic Information Systems

Cincinnati, Ohio
September 21-23, 1994

EPA/625/R-95/004
September 1995

Seminar Publication

National Conference on Environmental Problem-Solving with Geographic Information Systems

September 21-23, 1994
Cincinnati, Ohio

U.S. Environmental Protection Agency
Office of Research and Development
National Risk Management Research Laboratory
Center for Environmental Research Information
Cincinnati, Ohio

Notice

This document has been subjected to EPA's peer and administrative review and has been approved for publication as an EPA document. The views expressed are those of the authors and do not necessarily reflect those of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

When an NTIS number is cited in a reference, that document is available from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
703-487-4650

Please note that the Soil Conservation Service of the U.S. Department of Agriculture is now the Natural Resources Conservation Service; "Soil Conservation Service" is used throughout this document because this was the name in use at the time of the conference.

Contents

Page

Contents (continued)

Page

Contents (continued)

Contents (continued)

Page

Acknowledgments

The success of the conference and this document is due largely to the talents and efforts of many individuals. Gratitude goes to each person involved.

Authors

A special thanks goes to the many authors of the papers presented in this document. Their efforts in preparing papers made this document possible and led to the overall success of the conference.

Peer Reviewers

The following individuals peer reviewed this document:

- Jim Goodrich, U.S. Environmental Protection Agency, National Risk Management Research Laboratory, Cincinnati, Ohio.
- Nancy Phillips, Hollis, New Hampshire.
- John Shafer, Earth Sciences & Resources Institute, University of South Carolina, Columbia, South Carolina.

Editorial Review and Document Production

Heidi Schultz, Eastern Research Group, Inc., Lexington, Massachusetts, directed the editorial review and production of this document.

Technical Direction and Coordination

Daniel Murray and Susan Schock, U.S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Center for Environmental Research Information, Cincinnati, Ohio, coordinated the preparation of this document and provided technical direction throughout its development.

Special Thanks

Special thanks goes to the following individuals for their support:

- Tom Davenport, U.S. Environmental Protection Agency, Region 5, Chicago, Illinois.
- Mike Forrester, Urban and Regional Information Systems Association, Washington, DC.
- Bill French, American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland.
- Carol Griffin, Idaho Falls, Idaho.
- Lyn Kirschner, Conservation Technology Information Center, West Lafayette, Indiana.
- Kevin Klug, Association of American Geographers, Washington, DC.

- John Lyon, Ohio State University/American Society of Photogrammetry and Remote Sensing, Columbus, Ohio.
- Mark Monmonier, Department of Geography, Syracuse University, Syracuse, New York.
- Sam Noe, Joint Center for GIS & Spatial Analysis, University of Cincinnati, Cincinnati, Ohio.
- Nancy Phillips, Hollis, New Hampshire.
- Javier Ruis, U.S. Department of Agriculture, Soil Conservation Service, Fort Worth, Texas.
- Don Schregardus, Ohio Environmental Protection Agency, Columbus, Ohio.

Introduction

The National Conference on Environmental Problem-Solving with Geographic Information Systems was held in Cincinnati, Ohio, September 21 to 23, 1994. The conference was a forum for over 450 environmental professionals to exchange information and approaches on how to use geographic information systems (GIS) to define, assess, and solve environmental problems.

Cross-media pollutant transport and watershed-based decision-making have made the process of solving environmental problems more complex. The application of GIS to environmental problem-solving has greatly increased our ability to manipulate and analyze relational and spatial data, providing environmental decision-makers with a powerful tool for analyzing multimedia environmental data over increasingly broader areas (e.g., watersheds, states, regions). While the approach to using GIS varies from application to application, a common, technically sound framework for applying GIS to environmental problems should be developed and implemented. This conference was an initial step in defining this framework by examining the following areas:

- Problem identification and definition.
- Data requirements (e.g., coverage, scale), availability, documentation, reliability, and acquisition.
- Approaches considered and selected for problem-solving.
- Unique challenges and pitfalls encountered.
- Interpretation of results, including level of confidence achieved based on data quality and approach taken.

Presenters were requested to address one or more of these areas in papers and posters that focused on applications of GIS to specific environmental problems.

This document presents peer-reviewed papers from the conference. The papers have been organized by general topic area as follows:

- GIS Concepts
- Ground-Water Applications
- Watershed Applications
- Wetlands Applications
- Water Quality Applications
- Environmental Management Applications
- Other GIS Applications

The purpose of this document is to share the information presented at the conference with individuals who were unable to attend. This document will be useful to individuals who are currently applying GIS to environmental situations or considering GIS for application in environmental problem-solving. These individuals include environmental regulatory personnel at the federal, state and local level; university professors, researchers, and students; private sector personnel, including industry representatives and environmental consultants; and other interested persons. The goal of sharing this information with a broader audience is to help users apply GIS to environmental problem-solving with a greater awareness of the power and limitations of this very useful tool.

GIS Concepts

GIS Uncertainty and Policy: Where Do We Draw the 25-Inch Line?

James E. Mitchell
Institute for Environmental Studies, Louisiana State University, Baton Rouge, Louisiana

Abstract

The growing availability of improved hardware and software for geographic information systems (GIS) has outstripped most users' ability to identify and represent uncertainty in the available data. In practice, the proliferation and compounding of errors and uncertainty increase as information becomes more easily handled and combined from different sources.

Various stages of GIS database development and analysis generate different forms and amounts of error and uncertainty. In most cases, inherent uncertainty within source data is simply ignored and its nature eventually lost through subsequent processing. Both the location of features and their attributes can include error and uncertainty. By the time decision-makers receive mapped information, it is typically represented as correctly located and attributed.

The use of weather and climate information provided by the National Climatic Data Center (NCDC) is a common example of this scenario. Weather station locations provided by NCDC are reported to the nearest truncated degree-minute. A minute is one-sixtieth of a degree of arc. In the center of the continental United States, 1 minute of latitude averages approximately 6,000 feet and 1 minute of longitude averages approximately 4,800 feet. Thus, the station location is only known to lie within a box of approximately 1 square mile. Map representations of these data should reflect this uncertainty.

Under the Municipal Solid Waste Landfill (MSWLF) Criteria, the U.S. Environmental Protection Agency has dictated that the 25-inch precipitation contour line be used as a regulatory boundary for the level of protection required at municipal landfill sites. The way in which these lines are created and interpreted has important policy implications. Indeed, the cost and practicality of a given location must take this into account. If the 25-inch precipitation figure is critical, characterizing its uncertainty is also important.

In this work, uncertainty is considered a property of the data (1). A Monte Carlo procedure is used to represent the stochastic character of contour lines generated from point data with known locational uncertainty. The 30-year normal precipitation data for Kansas are used as an example. The results of this study are compared with the 25-inch contour used for regulatory purposes in Kansas. This study demonstrates that the method of interpolation greatly influences the resulting contours. In addition, locational uncertainty changes the results unpredictably using four different contouring methods. Finally, the differences have potentially significant policy implications. The nature and origin of these factors are discussed.

Problem Statement

The increasing power of geographic information systems (GIS) and the availability of digital data have enabled users and decision-makers to perform complex spatial analyses for a great variety of environmental applications (2). The rapid expansion of GIS has resulted in a parallel growing concern about the quality of data (3).

An understanding of error and uncertainty is critical for proper use of spatial information. For the purposes of this discussion, error is defined as a deviation between the GIS representation of a feature and its true value (4). For a location, this might arise from rounding or truncating digits. Attribute error can involve misclassification of a feature or some other form of incorrectly accounting for its nature. Error is a measurable value quantifying these differences.

Furthermore, uncertainty shall refer to a characteristic for which the exact location and/or quantity cannot be calculated (5) or an attribute whose value represents a distribution or some other ensemble (composite) measure. Locational uncertainty often arises when inappropriate measurement systems are used. An example of this is the use of a Public Land Survey designation (often referred to as a legal location) to specify a point location. This system is designed to represent a tract of land (an area). It is not accurate for locating points (1).

The uncertainty associated with an attribute is an important characteristic of that feature. It quantifies the precision of a stochastic quantity; that is, one that is not accurately represented by a single value. Annual precipitation is represented by a single number, typically the 30-year mean annual precipitation (30-year normal). This number varies each year, however, and that uncertainty can be quantified by the variance or other statistical measures. In this sense, uncertainty is a known or calculable value that can be used in spatial analyses.

Unreliable GIS data and products may lead to adverse environmental and legal consequences. The National Center for Geographic Information Systems and Analysis (NCGISA) chose data quality as the first initiative on its GIS research agenda (6). Many efforts have been made prior to this, and since, to understand and manage error and uncertainty in GIS applications.

GIS analyses are inherently subject to propagation of error and uncertainty (4, 7). No data set can represent every spatial reality of a geographically dispersed phenomenon. Monmonier (8) points out that as long as the three-dimensional earth's surface is transformed to a two-dimensional plane, error and uncertainty of various forms will be produced. Goodchild and Min-Hua (9) point out two issues that are important when dealing with error and uncertainty:

- Minimization of error in the creation of GIS products.
- Measurement and presentation of error and uncertainty in a useful fashion.

GIS technology introduces error and uncertainty through two major sources: (1) inherent error and (2) operational error. Inherent error is the error present in source data. It is generated when the data are collected. Operational error is generated during data entry and manipulation (7, 10-13). Examples include locational shifts due to projection or combining information from different source scales.

Most error and uncertainty contained in GIS data cannot be eliminated. Instead, they are actually created, accentuated, and propagated through GIS manipulation procedures (14-16). Most operational errors are difficult to estimate.

The selection by the U.S. Environmental Protection Agency (EPA) of a 25-inch per year local precipitation limit as one of the criteria to determine whether small municipal solid waste landfills (MSWLF) are subject to the provisions of Subtitle D provides an excellent example of how uncertainty and errors enter into a GIS analysis and its subsequent products. It demonstrates all of the major forms and purveyors of error and uncertainty:

- Spatial (locational) error
- Statistical (sampling) uncertainty
- Temporal (time domain) error
- Error proliferation (processing error)
- Analytical (choice of methodology) error
- Cartographic representation error

Many of these are avoidable; some are known and understood, yet they remain largely ignored by users of GIS technology. This work presents each of these factors, discusses their origins, and shows how GIS could have been used to better serve the policy and regulatory processes. The Kansas example demonstrates that ignoring the factors influencing error and uncertainty can result in incorrect conclusions and inappropriate policy decisions.

Data Requirements and Sources

To perform an analysis of precipitation, data are typically obtained from the National Climatic Data Center (NCDC), located in Asheville, North Carolina. This is the national repository for such data. These data are also available through state or regional climate centers. The Kansas Weather Library at Kansas State University provided data for this study. The 1990 "normal precipitation" data (17) and locations were obtained and generated into an ARC/INFO point coverage. Figure 1 displays the locations of the precipitation stations used in this study.

Normal precipitation is defined as the average annual precipitation for a three-decade (30 years) period at each station for which reliable data are available. To avoid "edge effects" (processing anomalies due to a lack of data along edges of an area), all stations in Kansas and some from neighboring states were used. A total of 380 stations compose this data set. In addition, precipitation contours from the "Availability of Ground Water in Kansas Map" (18) were digitized from a [paper] source map. The Geohydrology Section of the Kansas Geological Survey provided base map coverages of cartographic features.

All data represent the best available information from the source institutions noted above. Those organizations use the data in their analytical and cartographic research and production operations.

Methodology

To examine the influence of locational uncertainty on the representation of three-dimensional, natural phenomena, a Monte Carlo approach was adopted (1). Using this technique, random realizations of point locations are generated for each rain gauge, in each of 50 separate simulations. From this, 50 possible representations of the unknown locations of each gauge are used to create

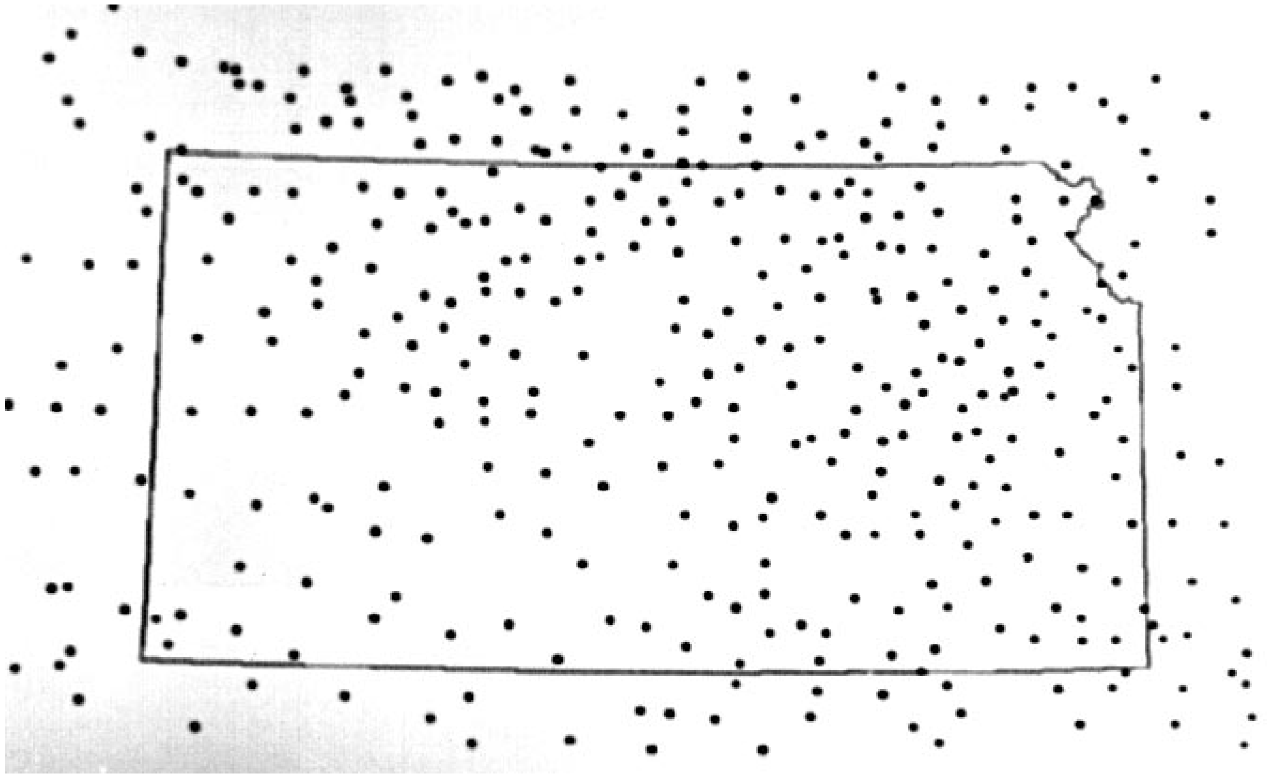


Figure 1. Location of rain gauges used in this study.

50 different sets of contours. All Monte Carlo calculations and data generation were performed using Statistical Analysis System (SAS) (19, 20).

These 50 simulations were sequentially processed using the four different contouring methods available within ARC/INFO. This provided a means to examine analytical error propagation. The first of the four methods is kriging (21, 22). This is referred to in the paper as the UK method, for its use of linear universal kriging (23). The other three are manipulations of the triangular-irregular network (TIN) contouring algorithm available in ARC/INFO. These differ by the number of interpolation points used along the edges of the elements in the TIN data structure (24). The first used the default 1, the second used 5, and the third used 10 (the largest value available). These are labeled D1, D5, and D10, respectively.

GIS operations used in this work include overlay analysis, areal calculation, and arc intersection. ARC/INFO was used for all GIS and cartographic production in this work.

Identifying the Sources of Uncertainty

Spatial Error

Data obtained from NCDC is provided with the knowledge that weather station locations are reported using truncated degrees and minutes of longitude and latitude. NCDC cannot provide any better locational accuracy at this time. Because each location is reported with error, this clearly has the potential to affect any contours or other three-dimensional features interpolated from the data. The magnitude and nature of this influence are unknown and unpredictable (1).

In addition to the poorly defined station locations, examination of the data revealed other anomalies. The locations in the publication reporting normal precipitation (17) were not identical to those identified by the Kansas State climatologist and NCDC. Some of the discrepancies were quite large. These anomalies were brought to the attention of all parties involved. No resolution was provided to this investigator's satisfaction, however.

The contours digitized from the "Availability of Ground Water in Kansas Map" are stated to originate from the 1960 normals (18). No documentation exists, however, concerning the way the lines were derived or the number of rain gauges used. Presumably, they were contoured by hand.

Statistical Uncertainty

This is a sampling consideration based on the size of the data, the nature of the process being sampled, and its variability. Unfortunately, precipitation is a particularly "patchy" phenomenon. That is, rain falls in a discontinuous fashion, and adjacent gauges can depict very different patterns. This is confounded by the fact that most contouring algorithms and other approaches to represent three-dimensional surfaces assume a relatively smooth (locally) and continuous process.

Areal processes are almost always sampled as point information. Most contouring algorithms require a regular grid from which to interpolate surface features. In practice, rain gauges, as well as other environmental sampling programs, are irregularly distributed. Placement often depends on factors other than grid sampling (e.g., convenience, access to communications, finances, Congressional districts). This creates a "nonexperimental sampling" design (25). Nonexperimental sampling can contribute to uncertainty (26, 27).

Temporal Error

The normals are recalculated each decade and can change drastically in local areas. These changes arise for various reasons. First, some stations enter and drop from the database. Stations are deleted due to changes in location or extended periods of data collection problems. On occasion, new stations are added. Thus, the size and areal coverage of the data set changes with time.

In addition, weather patterns change with time. Extended periods of drought or excess rain or snow alter measured precipitation. In turn, three-dimensional representations change unevenly.

Error Proliferation

Once an error enters into the database and is included in GIS operations, spatial analysis, or spatial interpolation, its effect passes into the next stage of processing. In the 1990 normal precipitation data for Kansas, two stations are reported in Garden City. Despite the fact that the two are only a few miles apart, their annual total precipitation differs by 2 inches! In consultation with the state climatologist (Mary Knapp, Kansas State University), one was eliminated from the analysis. This process was repeated for an additional six stations where reported values appeared to be anomalous compared with nearby stations or the previous normal precipitation (1951 to 1980).

Errors can also proliferate through the normal handling of data. With geographic data, this often occurs while converting data from raster to vector and vector to raster forms (28). Some GIS operations are best accomplished in one form or another. As a result, transformations are often "hidden" from the user. Commonly, features "move" slightly after each step in an analysis.

Analytical Error

Different techniques have been developed for performing spatial interpolation, and an abundance of software is available for this purpose. All these methods have strengths and weaknesses. Each is based on a specific set of assumptions about the form and nature of the data. Some are more robust (less sensitive to data anomalies) than others. Most importantly, some provide additional information useful in data analysis.

Unfortunately, users often "take the defaults" when using sophisticated techniques and ignore the assumptions behind the method. Parameters can be varied and their effect evaluated, as in a sensitivity analysis (29). Often, the best approach is to try several methods and evaluate their joint performance (30, 31).

Another difficulty is the need to assign values to areas. By definition, polygons in a GIS are considered to be homogeneous. In reality, they bound areas that are a gradation from one characteristic to another. On the other hand, contours are commonly used to depict surface gradients but are useless (within a GIS) for analytical or modeling purposes. Ultimately, data sampling is accomplished as a point process (except, perhaps, in remote sensing), while many forms of data analysis and processing require areal information.

Cartographic Representation Error

Communicating the uncertainty of map features is not a trivial endeavor. Maps can be produced in two basic forms: as a raster (e.g., orthophotoquads or satellite images) or a composition of vectors (e.g., contour maps). The printing process, however, often reduces all of this to a raster representation at a very fine pixel size. Each method poses its own problems in depicting uncertainty.

Rasters can be used effectively in conjunction with color information theory to produce a continuum of shading within a thematic map layer (32). The choice of colors, however, can influence the interpretation of the data, and no universal scheme exists for depicting thematic variability. For example, blue shades often represent water or cold, while yellow and/or red often represent temperature or heat.

Vectors present a different suite of problems. Contouring is the primary technique for using vectors to depict areal variation. By definition, however, contour lines represent an exact isoline or single value along its length. Uncertainty cannot be represented in a line. Rather, a composite of lines can be displayed that represents a set of possible interpretations of the data. This is not a practical solution for mapping, however, as it can create a jumble of intersecting lines that makes interpretation difficult and is not an aesthetic means of presentation.

Challenges Encountered in This Study

This work attempts to discover and account for sources of error and uncertainty in GIS analysis. Given this information, the challenges are to find the best way to incorporate it into the analysis and to represent it in a useful manner. Another challenge is finding ways to use

GIS uncertainty to support policy and management decisions. Addressing these manifold problems starts with identifying the sources of error and uncertainty, the way they enter the analysis, and the manner in which they are propagated through the use of GIS.

This study includes a number of known sources of uncertainty. In practice, this is not always the case. Users of GIS data and technology should always assume that the sources of uncertainty discussed in this paper are present and attempt to determine their nature. Uncertainty should be considered a property of the data and appropriately represented (1). This is the approach taken in this work.

After examining these factors, a Monte Carlo simulation was deemed the most appropriate approach to capture the nature of the locational uncertainty. Four different methods of contouring were used to examine analytical uncertainty (uncertainty due to the choice of a contouring algorithm). In addition, the contribution of statistical (sampling) uncertainty could have been addressed through incorporating information about the standard error of the point precipitation measurements (normals) used as the base data. Time limitations precluded examining this dimension of the question. Comparing the contours resulting from the 1960 and 1990 precipitation normals demonstrates the effect of temporal variation.

The greatest challenge is communicating the uncertainty in a manner useful to decision-makers. This paper presents a series of maps, figures, and tables aimed at addressing this problem. Some of the maps (see Figures 2 through 5) show the uncertainty resulting from each of the contouring methods. Figure 6 depicts the union (overlay) of the four approaches and displays their correspondence. The pie chart in Figure 7 is a nonspatial representation of this correspondence and the relative area represented within the different combinations of overlapping regions of uncertainty. Tables 1 through 3 further compare these quantities. Figure 8 is the map that the Kansas Department of Health and Environment (KDHE) chose to define the regulatory boundary (the 25-inch contour). The contours resulting from this study can be seen in Figure 9. Finally, the map in Figure 10 is a cartographic comparison of the differences between the contours used by KDHE (based on the 1960 normals) and those generated by the currently available data (the 1990 normals).

There is no single best approach for meeting these challenges, and there may never be one. The real challenge to address is how to educate technical GIS professionals and the users of their work to look for uncertainty and consider its influence on their decision-making process.

Results

The zones of uncertainty defined by the results from the four contouring methods used in this study are displayed in Figures 2 through 5. For each method, these zones represent the areal extent of the overlain contour lines produced in the 50 simulations. Each region is bounded by the furthest west or east contour generated along any length of the region. Table 1 shows the relative area falling within each of these zones as they traverse the state of Kansas. Clear differences exist between the total areas of uncertainty. It is their placement and relative location, however, that have policy and management implications. GIS is required to examine these questions.

Table 1. Comparison of Absolute and Relative Area of Uncertainty Arising From Four Methods of Determining the 25-Inch Precipitation Contour

Method[a]	Total Area (square miles)	Difference Between This and UK Method (square miles)	Percentage of UK Method	Percentage of Combined Area[b] of Uncertainty
UK	289.33	—	—	19.65
D1	494.38	205.05	170.87	33.58
D5	602.13	312.80	208.11	40.90
D10	631.11	341.78	218.13	42.87

[a] UK = universal kriging with linear drift, D1 = TIN interpolation with 1 subdivision, D5 = TIN interpolation with 5 subdivisions, D10 = TIN interpolation with 10 subdivisions (23, 24).
[b] A union (overlay) of all four sets of regions of uncertainty creates a combined area of 1,472.07 square miles. This includes the zones of uncertainty for each method of contouring and areas not included within any of the four regions of uncertainty (gaps between them).

Figures 2 through 5 clearly show differences in both the extent of uncertainty in the 25-inch contour line and its positional interpretation. Each method has a slightly different bend or twist. Islands (isolated regions where the 25-inch line appears as a closed loop) are manifested differently depending on the interpolation scheme. It is interesting to note the relative correspondence between the general shape of the D1 and UK methods. In the south-central border region, D1 and UK represent the local uncertainty as a bulge, while D5 and D10 depict it as an island of lower precipitation.

Areal correspondence and difference are depicted in Figure 6 and Table 2. Figure 7 is a pie chart visualizing the information in Table 2. These results are somewhat surprising in that areas where none of the four methods located the 25-inch line (“None Present”) represent the second largest composite area. Because of the large number of “sliver polygons,” the graphic representation of the overlay is somewhat difficult to interpret. Table 2 clarifies these interrelationships by breaking down the various categories. The average area per polygon value

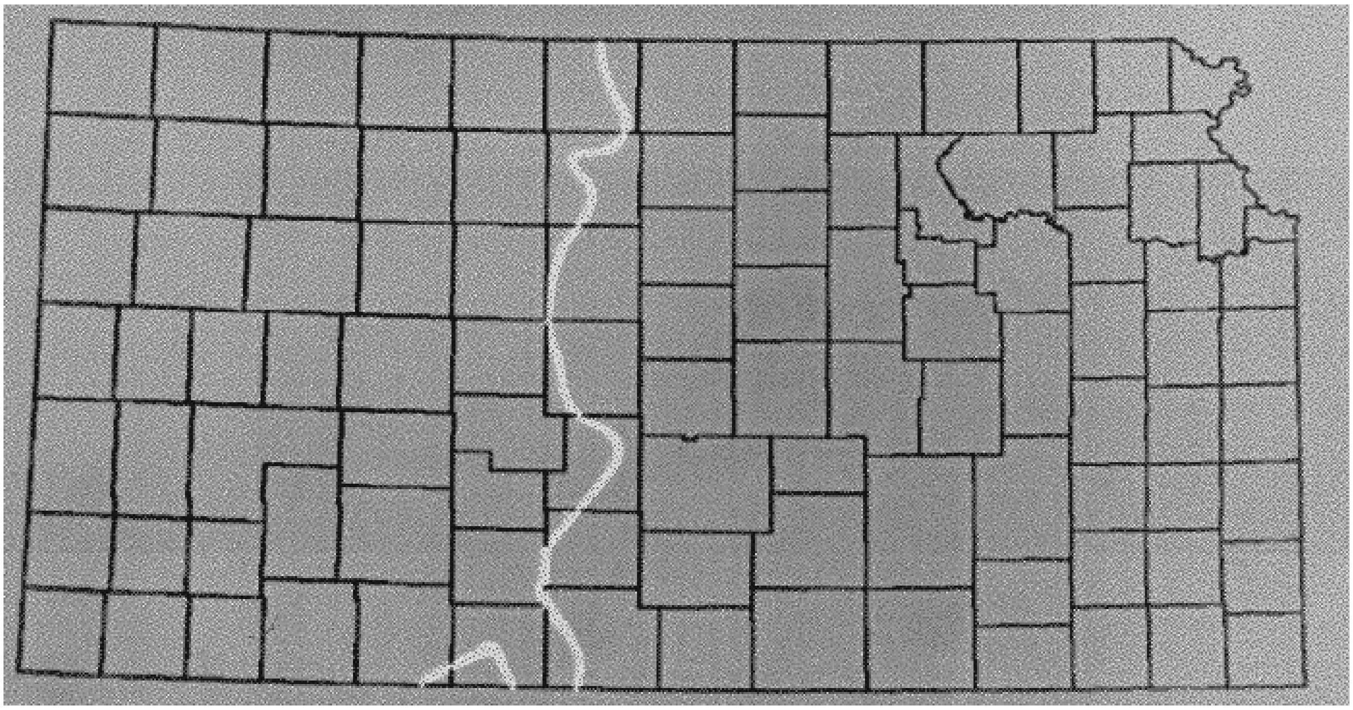


Figure 2. Regions of uncertainty produced by the UK method of contouring.

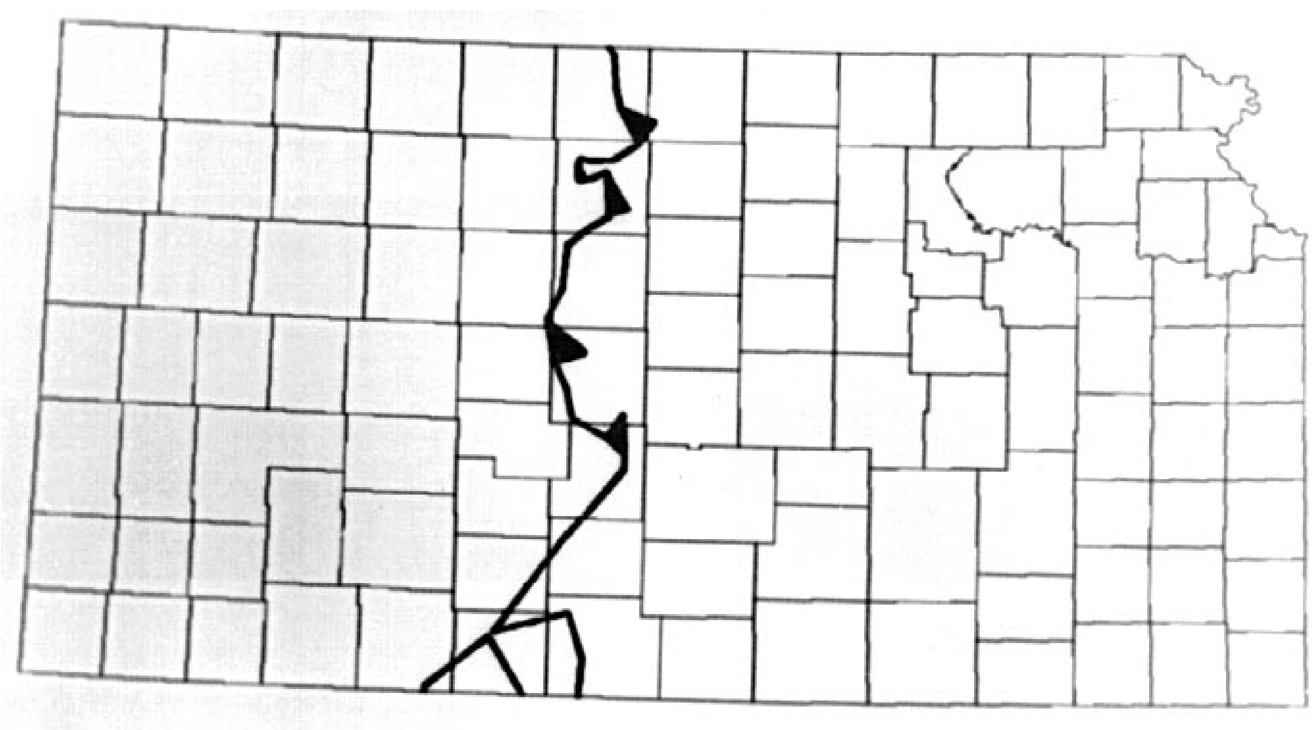


Figure 3. Regions of uncertainty produced by the D1 method of contouring.

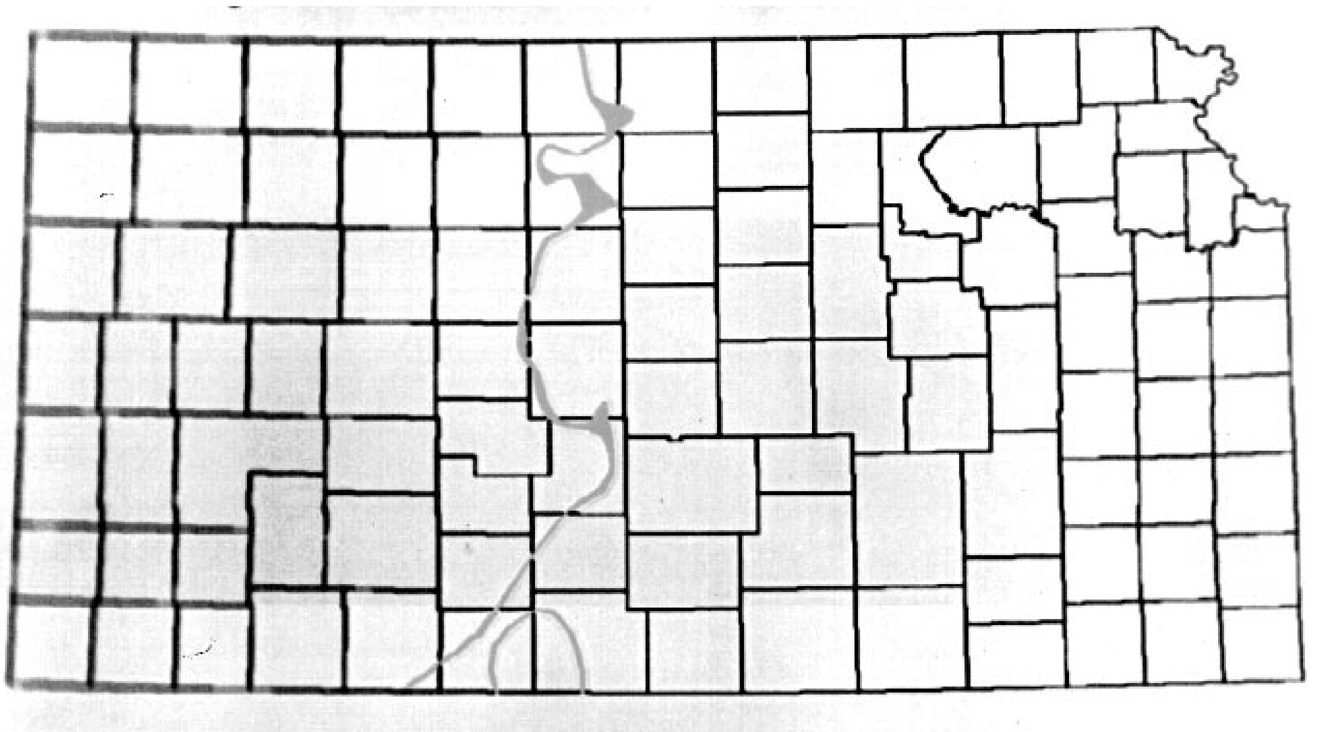


Figure 4. Regions of uncertainty produced by the D5 method of contouring.

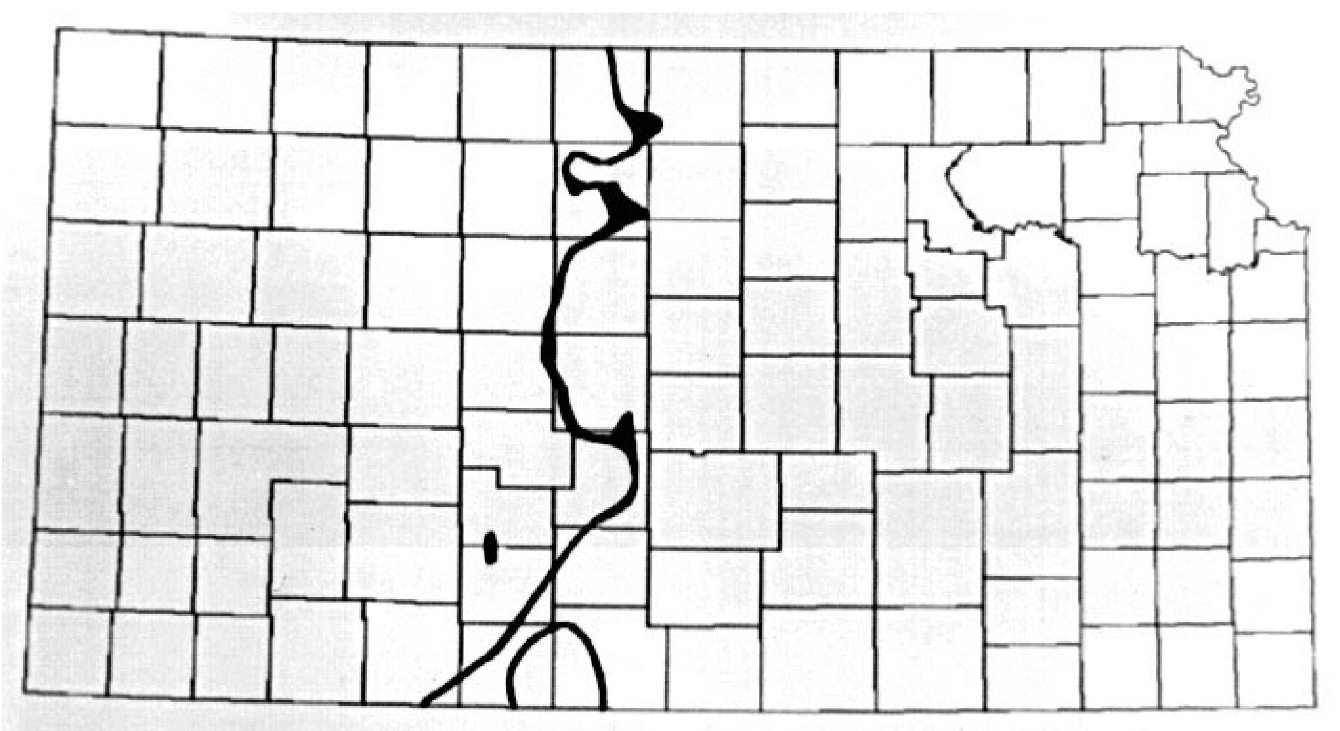


Figure 5. Regions of uncertainty produced by the D10 method of contouring.

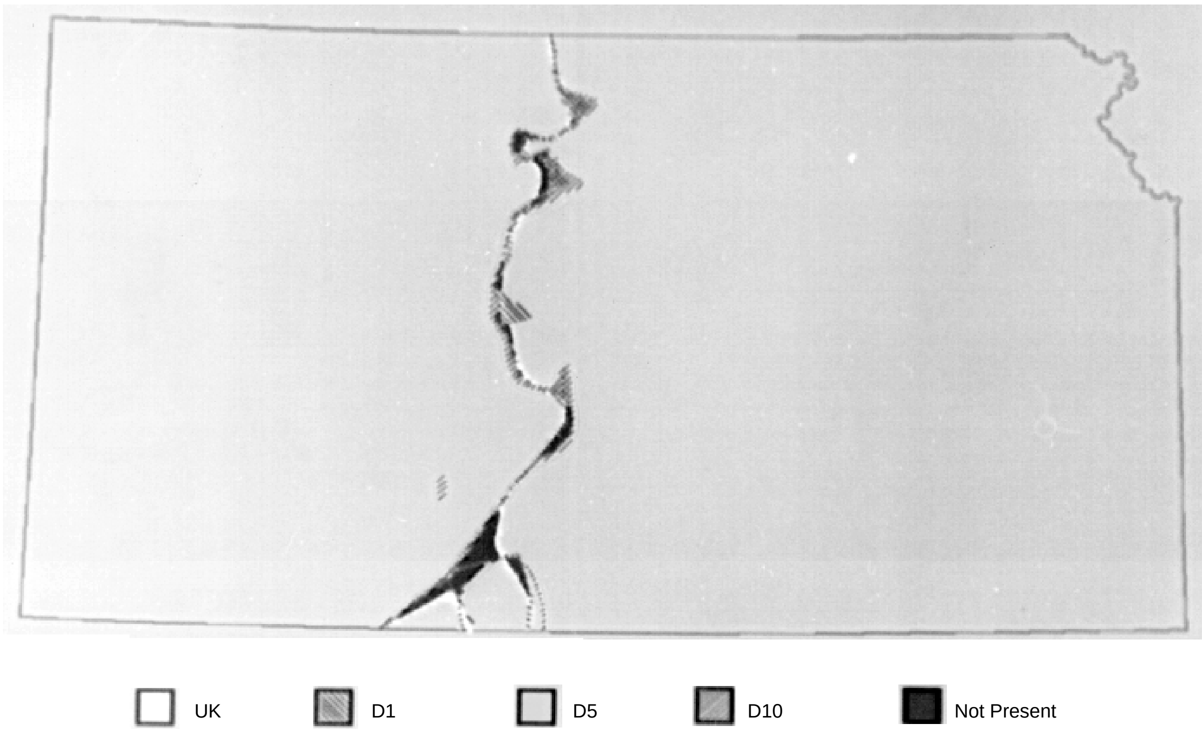


Figure 6. **Union of the regions of uncertainty from all four methods of contouring. The numerous "sliver" polygons make this a difficult presentation to interpret at this scale. The black areas appear prominently, however. These represent areas where no method placed contour lines.**

adds to the interpretation of the relative areas by incorporating the number of polygons in each category. An inspection of this column makes those categories with a multitude of very small polygons stand out. It also displays a number of large jumps in magnitude. As this number increases, the significance of the correspondence increases.

Table 2 highlights the correspondence between the D5 and D10 approaches and the D1 and UK methods of interpolation. This relationship is interesting because the algorithms used by the UK and D1 methods both are forms of linear interpolation. The D5 and D10 algorithms are designed to provide more "smoothing" and appear to create increasingly more "bull's eyes." Only D5 and D10 generate these features. Of the 50 simulations, a particular bull's eye appears west of the 25-inch contour (see Figures 3 and 4) four times using D5 and 39 times using D10. The size and location of these anomalies also vary with the input data. Polygons containing contours from all four methods rank ninth in total area and seventh in average area (out of 17 categories). This supports the conclusion that the four chosen methods have a relatively low spatial correspondence.

Table 3 breaks down the area of uncertainty by county. Although the zones of uncertainty appear to be relatively small when displayed on a statewide basis, they have important impacts in local areas. In particular, combining this information with soils, topography, ground water, and other information can clearly indicate whether a specific location is suitable for a landfill. Often, information developed at one scale is used in another. In this case, statewide information is being used for a site-specific application.

Figure 8 is a copy of the map the KDHE used to delineate the 25-inch precipitation contour. The results from

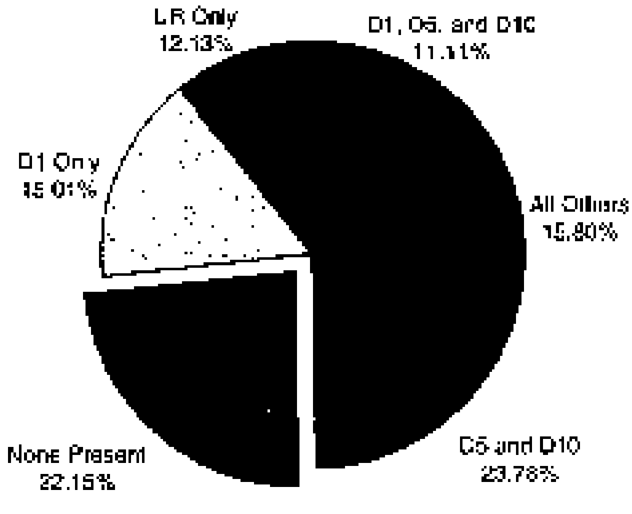


Figure 7. **Breakdown of the total area in each category resulting from the union (overlay) of the regions of uncertainty from all four contouring methods. The "None Present" category represents a surprisingly large proportion among the 17 possible combinations.**

Table 2. Comparison of Absolute and Relative Area of Uncertainty Arising From Four Methods of Determining the 25-Inch Precipitation Contour

Methods of Contouring Found Within Area	N	Total Area (square miles)	Percentage of Total Combined Area	Average Area per Polygon[a] (square miles)
D5 and D10	28	350.13	23.78	12.50
None present	28	326.06	22.15	11.64
D1 only	37	221.02	15.01	5.97
UK only	27	178.60	12.13	6.61
D1, D5, and D10	24	163.61	11.11	6.82
D10 only	69	70.07	4.76	1.02
UK and D1	19	65.82	4.47	3.46
D5 only	73	34.68	2.36	0.48
Common to all	10	23.45	1.59	2.35
UK, D5, and D10	14	15.47	1.05	1.11
D1 and D5	38	10.68	0.73	0.28
D1 and D10	38	6.50	0.44	0.17
UK and D5	19	2.21	0.15	0.12
UK, D1, and D5	12	1.89	0.13	0.15
UK, D1, and D10	12	1.41	0.10	0.12
UK and D10	13	0.47	0.03	0.04
Combined total	461	1472.07	100.00	

[a]Average Area per Polygon = (Total Area) / N. This a useful measure to compare the relative size of each polygon in each classification.

Table 3. Area of Uncertainty for Each County Arising From Four Methods of Determining the 25-Inch Precipitation Contour

County	UK Area (square miles)	D1 Area (square miles)	D5 Area (square miles)	D10 Area (square miles)
Barber	36.38	24.29	28.05	28.31
Barton	28.47	96.84	56.16	59.56
Clark	12.64	11.49	12.97	13.02
Comanche	44.30	52.94	51.43	52.23
Edwards	—[a]	—	—	9.44
Ellis	—	6.14	9.27	8.90
Jewell	—	6.65	23.05	24.26
Kiowa	15.15	14.32	17.81	31.12
Osborne	43.87	99.82	168.2	172.6
Pawnee	—	—	0.11	0.30
Pratt	18.84	11.30	11.56	11.43
Rush	0.04	8.51	31.32	29.51
Russell	27.17	22.71	28.30	28.00
Smith	27.59	68.47	67.38	66.34
Stafford	34.88	70.90	96.49	96.05
Total	289.3	494.4	602.1	631.1

[a]A dash indicates that no contours appeared in that county for the method specified.

the UK method were selected as the best available representation of normal precipitation across Kansas (see Figure 9). The figure displays the unclipped contour lines generated from the data. This is done to point out the importance of “edge effect.” Note the incoherent behavior of the contour lines at their termini. If a smaller window of data points were used, interpolation problems would have lain across the region of interest. When present, these features require more handling and time for analysis. They often introduce additional error and uncertainty.

The policy implications of this example are demonstrated in Figure 10. Here, the map shows the combination of the “official” KDHE map and the data interpreted in this study. The pattern of noncorrespondence is noteworthy. The lightest areas are regions that currently experience higher annual precipitation than forecast by the 1960 normals (from the KDHE map). Black areas are expected to have lower precipitation under current climatic conditions. Therefore, large areas of Kansas that should be under regulation according to the MSWLF regulations are not.

In summary, the figures and tables clearly show that locational uncertainty of data measured as points is propagated into contour lines. The nature and magnitude of that uncertainty varies with location and method of interpolation and shows no regular (predictable) pattern. Perhaps most importantly, uncertainty that appears small at one scale can be relatively more significant at another. In addition, seemingly small geographic feature and uncertainty can be an important factor in decision-making.

Discussion and Conclusions

GIS is an established and accepted technology, especially in applications related to natural resource and environmental management. Despite the widespread proliferation of GIS into these areas, the available data are not always appropriate for the intended application. Furthermore, adequate documentation is not always available to determine whether the data are adequate for a given use. The development of metadata standards will play an important role in addressing this problem. Errors and uncertainty will always be present in GIS data. Recognizing their presence, incorporating them into the analysis, and representing them in GIS products will remain a constant challenge.

This study demonstrates the influence that various sources of GIS uncertainty can leverage on the results of an analysis. The example of the 25-inch precipitation

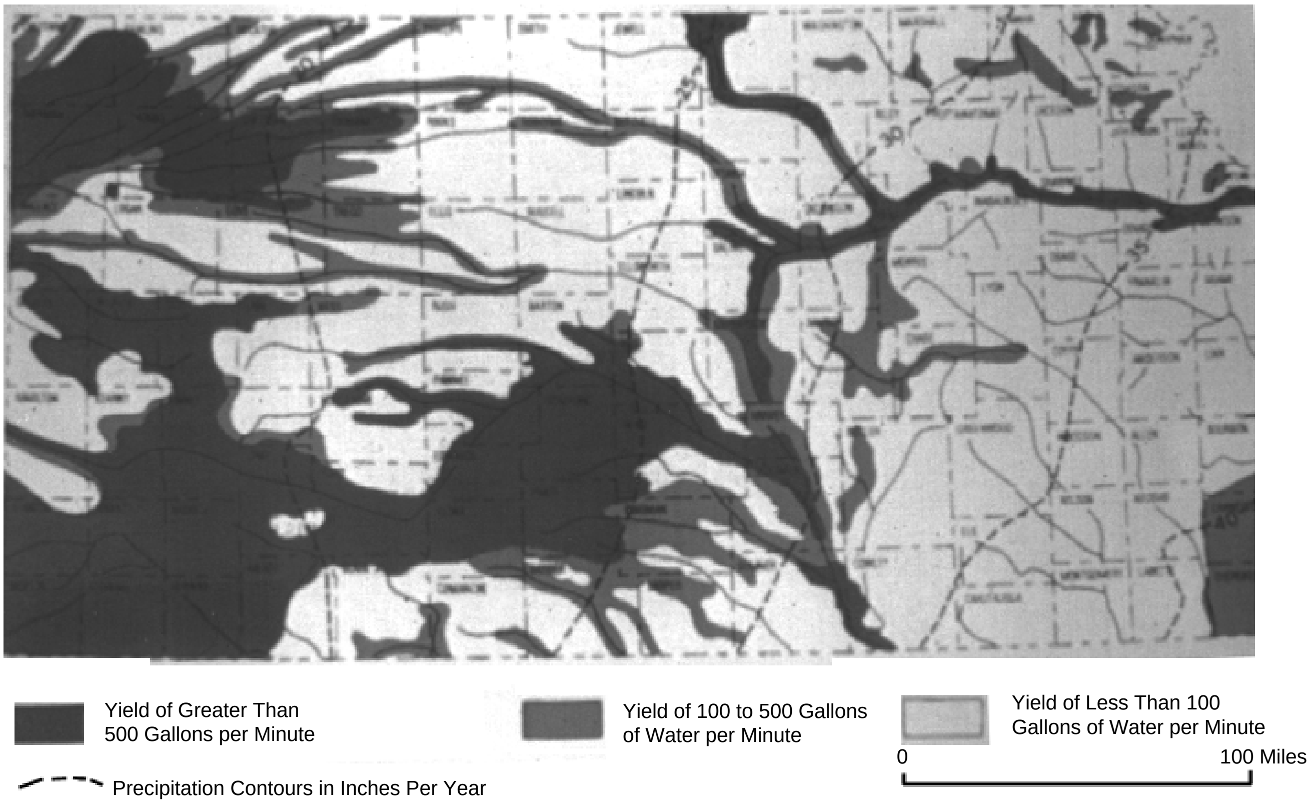


Figure 8. The map that the KDHE selected as the definitive source for the location of the 25-inch precipitation contour (18).

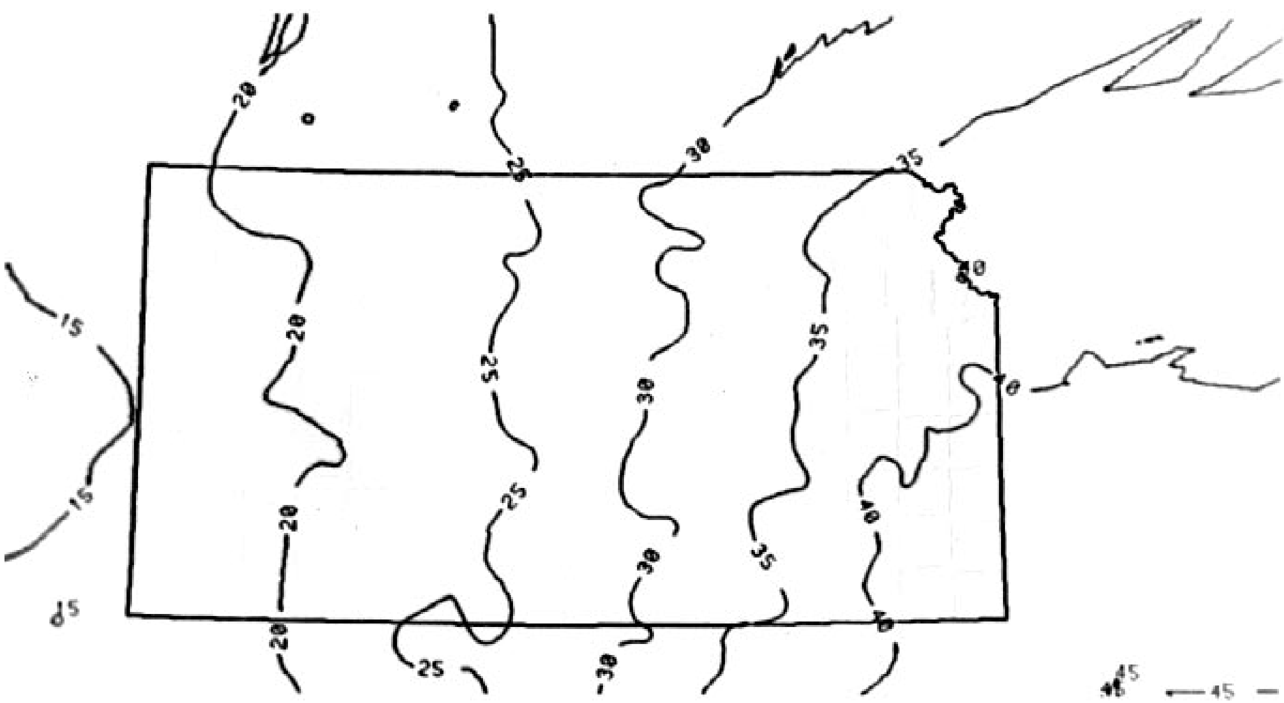


Figure 9. Map of the precipitation contours resulting from applying the UK method with linear drift to the 1990 normals (17). Here, the contour interval is 5 inches. Note the incoherent behavior of the contours around the margins of the map. This is referred to as "edge effect."

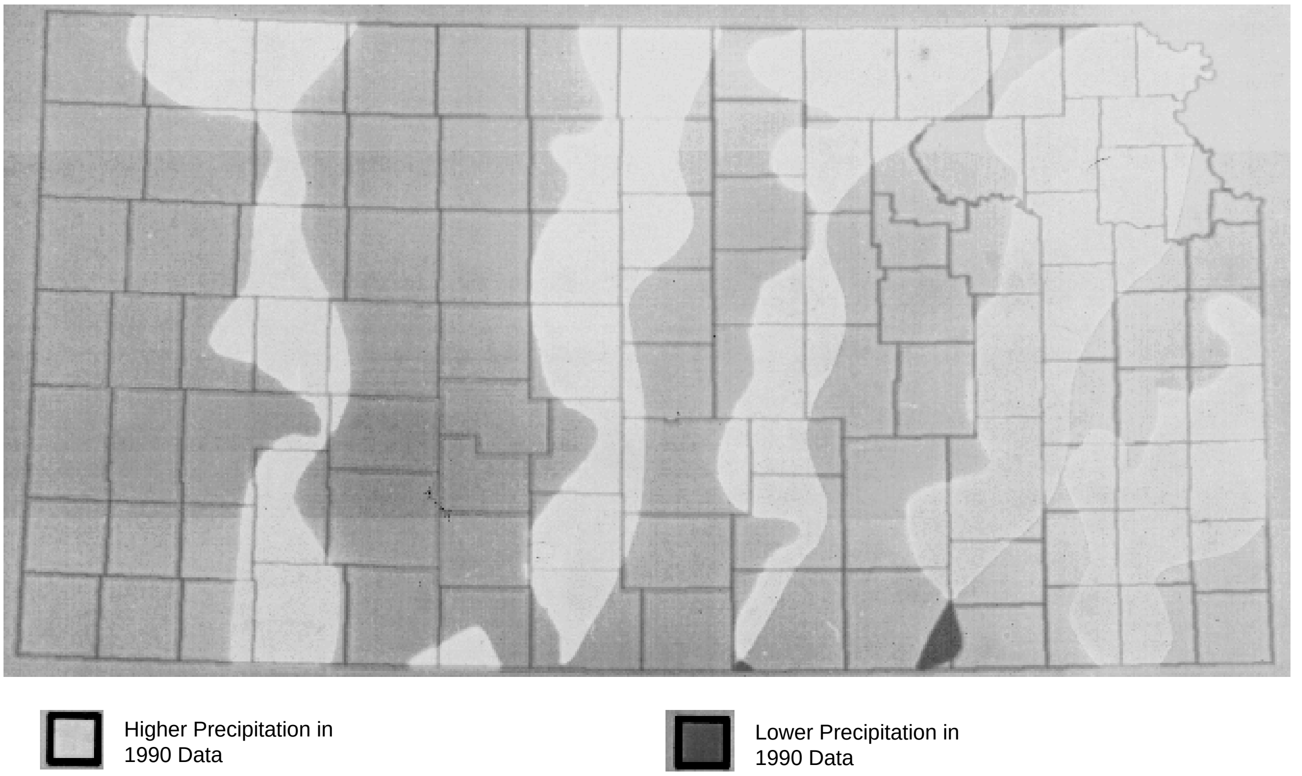


Figure 10. This map represents the union (overlay) of the information in Figures 8 and 9. The lighter regions represent areas exhibiting higher precipitation in the 1990 normals (1961 to 1990) than was apparent in the 1960 normals (1931 to 1960). The darker areas show the opposite relationship.

line in Kansas is a clear example of how the use of inappropriate data can have far-reaching effects on policy and management. The regulatory agency, KDHE, chose the wrong map upon which to base its regulatory authority. As a result, numerous potential sites for small municipal solid waste landfills will be considered that are in violation of the letter and intent of the law.

Ultimately, the responsibility for proper use of GIS technology lies in the hands of practitioners. Technical staff performing GIS analysis must be knowledgeable about sources of error and uncertainty and ensure that users of their work are aware of their influence on GIS output.

The problems demonstrated in the Kansas example could have been avoided simply by investigating the appropriateness of the data. Instead, a convenient source was chosen without seeking any other sources of “better” information. Indeed, familiarity with the nature of the data (30-year normals) should have led the policy analyst to select the most current data and not data that are 30 years out of date! An understanding that contour lines represent a generalization of the point precipitation measurements should also have led to the conclusion that locations near the boundary line ought to be monitored for compliance. Both the temporal and spatial characteristics of climate can change, as exemplified by the difference in the 1960 and 1990 normals. The “Dust Bowl” periods of the 1930s and 1950s significantly influenced the 1960 normals (33). As a result, they are not appropriate for this application.

Although a powerful tool, GIS does not hold all the answers. The technical community and policy-makers must work together to ensure its proper use. In reality, no 25-inch precipitation line floats over Kansas. It is merely the interpretation of scientists and policy-makers who select its location. The only way to arrive at a reasonable answer is to gather the best available information and allow all parties to scrutinize it. GIS can be a wonderful tool to do this.

Acknowledgments

The author wishes to express his gratitude to the Kansas Geological Survey for its computing support before, during, and after his move to Louisiana. Also, the author wishes to thank Mr. M. Schouest, who made the move both possible and tolerable. Mr. C. Johnson of Johnson Controls in Baton Rouge, Louisiana, produced the excellent presentation graphics. Finally, this work could not have been performed without the Internet.

References

1. Mitchell, J.E. 1993. The influence of (x,y) uncertainty on prediction error and contour lines from a three-dimensional surface. Water Resources Bull. 29(5):863-870.

2. Northeast Waste Management Officials Association. 1992. Catalogue and directory of New England states and regional GIS organizations and activities, and an assessment of their future GIS needs. EPA/G-D001456/90-2.

3. Campbell, W.G., and D.C. Mortenson. 1989. Ensuring the quality of geographic information system data: A practical application of quality control. EPA/600/J-89/306.

4. Chrisman, N.R. 1991. The error component in spatial data. In: Maguire, Goodchild, and Rhine, eds. Principles of geographic information systems, Vol. 1. pp. 164-174.

5. Isaaks, E.H., and R.M. Srivastava. 1989. An introduction to applied geostatistics. New York, NY: Oxford University Press.

6. National Center for Geographic Information and Analysis. 1989. The research plan of the National Center for Geographic Information and Analysis. Int. J. Geo. Info. Sys. 3(2):117-136.

7. Vitek, J.D., S.J. Walsh, and M.S. Gregory. 1984. Accuracy in geographic information systems: An assessment of inherent and operational errors. Proceedings of the PECORA IX Symposium. pp. 296-302.

8. Monmonier, M.S. 1977. Maps, distortion, and meaning. Association of American Geographers Resource Paper Number 75-4.

9. Goodchild, M.F., and W. Min-Hua. 1988. Modeling error in raster-based spatial data. Proceedings of the Third International Symposium on Spatial Data Handling, Sydney, Australia (August 17-19). pp. 97-106.

10. Walsh, S.J. 1989. Inherent and operational error within a GIS: Evaluation from a user perspective. In: Goodchild, M.F., ed. Accuracy of spatial databases. p. 290.

11. Burrough, P.A. 1986. Principles of geographic information systems for land resources assessment. Oxford, England: Clarendon Press.

12. Marble, D.F., and D.J. Peuquet. 1983. Geographic information systems and remote sensing: Manual of remote sensing, 2nd ed. Falls Church, VA: American Society of Photogrammetry and Remote Sensing. pp. 923-958.

13. Mead, D.A. 1982. Assessing data quality in geographic information systems. In: Johannsen, C.J., and J.L. Sanders, eds. Remote sensing for resource management. Ankeny, OH: Soil Conservation Society of America. pp. 51-59.

14. Heuvelink, G.B.M., and P.A. Burrough. 1989. Propagation of errors in spatial modeling with GIS. Int. J. Geo. Info. Sys. 3(4):303-322.

15. Dunn, R., A.R. Harrison, and J.C. White. 1990. Positional accuracy and measurement error in digital databases of land use: An empirical study. Int. J. Geo. Info. Sys. 4(4):385-398.

16. Walsh, S.J., D.R. Lightfoot, and D.R. Butler. 1987. Recognition and assessment of error in geographic information systems. Photogrammetric Eng. and Remote Sensing 53(10):1,423-1,430.

17. Owenby, J.R., and D.S. Ezell. 1992. Climatography of the United States: Monthly station normals of temperature, precipitation, and heating and cooling; degree days 1961-1990, Kansas. No. 81. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Climatic Data Center, Asheville, NC (January).

18. Bayne, C.K. 1975. General availability of ground water and normal annual precipitation in Kansas (1:500,000 map). Kansas Geological Survey Map Series M4-A.

19. SAS Institute, Inc. SAS/GRAPH Software: Reference, Version 6, Vol. 1, 1st ed. Cary, NC: Statistical Analysis System Institute, Inc. 1990.

20. SAS Institute, Inc. SAS/GRAPH Software: Reference, Version 6, Vol. 2, 1st ed. Cary, NC: Statistical Analysis System Institute, Inc. 1990.

21. Olea, R.A. 1992. Understanding allays intimidation. Geobyte 7(5):12-17.

22. Bastin, G., B. Lorent, C. Duqué, and M. Gevers. 1984. Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. Water Resour. Res. 20(4):463-470.

23. ESRI. 1991. Cell-based modeling with GRID. Redlands, CA: Environmental Systems Research Institute, Inc.

24. ESRI. 1991. Surface modeling with TIN, surface analysis, and display. Redlands, CA: Environmental Systems Research Institute, Inc.

25. Davis, J.C. 1982. Statistics and data analysis in geology, 2nd ed. New York, NY: John Wiley and Sons.

26. Mitchell, J.E. 1991. A regionalization approach based on empirical Bayes techniques with applications to hydrologic forecasting, Ph.D. dissertation. Duke University, Durham, NC.

27. Mitchell, J.E., H. Sun, and S. Chakrabarti. 1992. Predicting the form of a surface from random, nonexperimental samples using neural networks. Proceedings of the First Midwest Electro-Technology Conference, Iowa State University, Ames, IA (April 10-11).

28. Congalton, R.G., and D.M. Schallert. 1992. Exploring the effects of vector to raster and raster to vector conversion. EPA/600/R-92/166.

29. Mitchell, J.E., and H. Sun. 1992. Generalization in higher order neural networks when predicting polynomial surfaces, implications for the free-form surface problem. In: Intelligent engineering systems through neural networks, Vol. 2. Proceedings of the Artificial Neural Networks in Engineering (ANNIE '92) Conference, St. Louis, MO (November 15-18).

30. Bates, J.M., and C.W.J. Granger. 1969. The combination of forecasts. Operational Res. Quarterly 20:451-468.

31. Granger, C.J., and R. Ramanathan. 1984. Improved methods of combining forecasts. J. Forecasting 3:197-204.

32. Carr, D.B., A.R. Olsen, and D. White. 1994. Hexagon mosaic maps for display of univariate and bivariate geographical data. EPA/68-C8/0006.

33. Rule, G.K. 1960. Dust bowl. In: The world book encyclopedia, Vol. 4. pp. 315-316.

Data Quality Issues Affecting GIS Use for Environmental Problem-Solving

Carol B. Griffin
Henry's Fork Foundation, Island Park, Idaho

"Abandon hope, all ye who enter here." Dante's quote might well be the advice that experienced geographic information system (GIS) users give to nonusers about to confront data quality issues associated with GIS use. Indeed, after reading this paper, some decision-makers might abandon attempts to use a GIS because of the error associated with it. Others may want to spend an inordinate amount of time and money trying to eliminate all error associated with GIS use. Neither option is prudent.

Data quality is important because it affects how reliable GIS-generated information is in the decision-making process. Too often, the availability of inexpensive digital data overshadows data quality concerns; people frequently use digital data because they are available, not because they have the necessary accuracy.

A GIS can help decision-makers use spatial information more fully than manual methods allow, but sometimes data quality issues cause concern about using GIS-generated outputs. Making environmental decisions without adequate consideration to data quality may lead to an erroneous decision, erode public confidence, or cause an agency to incur liability. This paper attempts to encourage decision-makers to become more aware of data quality issues, including the sources and magnitude of error.

GIS error research has necessarily progressed in a linear fashion, beginning with identifying and classifying sources of error. This paper discusses both inherent (source) error and the error that GIS operations introduce (operational error) during data input, storage, analysis/manipulation, and output (1). Strategies for coping with error and research into error reduction techniques have only recently received attention. Unfortunately, the answers to error management questions such as, "How will the error affect decision-making?" are not clear. The end of this paper covers several error management suggestions and anticipated software improvements designed to reduce errors, however.

Data Quality Concepts and Their Importance

Data quality is a major issue for GIS-generated maps, much more so than it is for paper maps. In part, this is because a GIS can perform operations on spatial data that would be nearly impossible without a GIS because of scale, complexity, and generalization issues (2). Cartographers adjust for these problems when they manually manipulate and instantly combine paper maps by adhering to long-standing cartographic principles, but GIS personnel may not be fully trained in these principles. A GIS enables an analyst, whether trained in cartographic principles or not, to combine or manipulate data in appropriate or in inappropriate, illogical, and erroneous ways. Lack of training coupled with the speed of spatial data manipulation can have serious consequences for an agency whose personnel produce and use GIS-generated maps.

Limited scientific understanding, limited ability to measure data, sampling error, inherent variability, and inadequacy of mathematical representations all contribute to uncertainties associated with spatial data. Uncertainty about spatial data consists of two parts: ignorance and variability. Ignorance means that variables have a "true" value, but it is unknown to us, whereas variability means one value cannot represent the variables.

Data quality defies a simple definition. For this paper, data quality can roughly mean how "good" the data are for a given purpose. People usually think of data quality in terms of error, but the term is broader and encompasses the six components outlined in the next section. Error can mean the difference between the observed values and the "true" value. The "true" value of a variable is usually unknown and unknowable, but for this paper's purposes, "true" could be the known value or the value one would obtain from field measurements (the discussion of data collection tries to dispel the notion that there is one "true" value for many variables, such as soil type in a given area or water temperature in a lake). Imperfect equipment or observers and environmental effects cause spatial error. According to Thapa and Bossler (3), errors fall into three categories:

- Gross errors and blunders (people or equipment).
- Systematic errors (which introduce bias).
- Random errors (due to imperfect instruments and observers).

In addition, another view divides spatial error into two different components: accuracy and precision. Accuracy means how close a value is to the "true" value or a known standard (absence of bias). Precision can have two definitions: it can be a measure of dispersion (standard deviation) of observations about a mean, or it can refer to the number of decimal digits used to represent a value (4). In the first definition of precision, a measurement of 6 feet plus or minus 1 foot is more precise than one of 6 feet plus or minus 3 feet. In the second definition, a value of 6.1794 feet is more precise than one of 6.1 feet. Figure 1 provides a graphic explanation of the difference between error, accuracy, and precision.

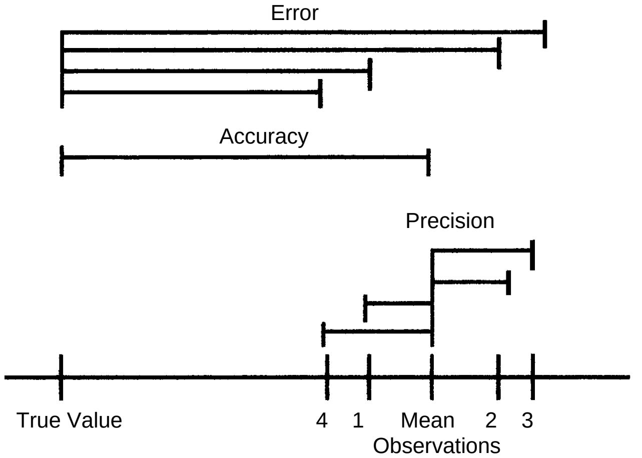


Figure 1. Relationship between error, accuracy, and precision.

Data are not accurate or inaccurate. Instead, data accuracy exists on a continuum, ranging from low to high accuracy. Although people strive for accurate (error-free) data, obtaining 100-percent accurate data is impractical. The list below provides some of the reasons why total accuracy is not obtainable (5):

- Objects to be measured are often vaguely defined.
- Some phenomena are variable in nature.
- Classification schemes are imprecise.
- Measurements are inherently imprecise.
- Gross errors of a nonstatistical nature can occur during measurement.
- Attributes encoded on an ordinal scale (high, medium, low) are approximate.
- Data represent a past state of reality.

Users of geographic data should strive for data that are only as accurate as they need. A variety of factors, of course, can determine need:

- Intended use of the data
- Budget constraints
- Time constraints
- Data storage considerations
- Potential liability

The main barrier to highly accurate data is lack of funds. Nale (6) suggests that rather than abandoning a GIS project because funds are not sufficient to achieve the desired accuracy, an agency should collect data at the desired accuracy from smaller areas, such as areas being developed or redeveloped. Over time, data collection at the desired accuracy can expand to include areas that lacked data due to budgetary constraints. Smith and Honeycutt (7) outline the use of a value of information approach in determining the need for more data (or more accurate data) based on the expected costs and benefits associated with data collection. If the benefits of increased data accuracy are greater than the expected costs, additional funds should be allocated to obtain more accurate data.

The intended use of data affects the type of data, as well as the data quality needed. Beard (8) divides GIS applications into six types (see Table 1). The specific type of data quality one needs (e.g., positional accuracy, attribute accuracy) also varies with the intended application. Analysts with inventory applications such as agricultural production are less concerned about positional accuracy than with an accurate assessment of anticipated crop yields (attribute accuracy). Decision-makers must

Table 1. Types of GIS Applications (8)

Application	Example
Siting	Finding optimal location (fire station, waste site)
Logistic	Movement or distribution through space (emergency response, military movement)
Routing	Optimal movement through a known network (mail, school bus)
Navigation	Way finding; may or may not involve a known network (ground, sea, air)
Inventory	Count and location of objects for a given time (census, tax rolls)
Monitoring/ Analysis	Examining processes over space and time (ecological, zoological, geological, epidemiological studies)

decide which data quality component is the most important for their use because optimizing all six components can be very expensive (9). An obvious conflict arises when local and state governments must meet multiple application needs simultaneously and thus feel forced to try to optimize several data quality components.

The nature of the decision may also help decision-makers determine the data quality they need. Beard (8) lists several of these factors (see Table 2). A political, high-risk decision requires higher quality data than a nonpolitical, low-risk decision because more public attention focuses on the former decision.

Table 2. Factors That May Affect the Data Quality Needed for Decision-Making (8)

Lower Data Quality Possibly Needed	Higher Data Quality Possibly Needed
Routine	Nonroutine
Nonpolitical	Political
Minimal risk	High risk
Noncontroversial	Controversial
Indefinite	Immediate
Local implication	Global implication

Components of Data Quality

The National Committee for Digital Cartographic Data Standards (9) identifies six components of digital cartographic data quality. This section discusses each of these components:

- Lineage
- Positional accuracy
- Attribute accuracy
- Logical consistency
- Completeness
- Temporal accuracy

Most components of data quality apply to both source and operational error.

Lineage

Because uses and users of data change, those at the national level have noted a recent push to include documentation when disseminating spatial data. Data lineage, also known as metadata or a data dictionary, is data about data. Metadata consists of information about the source data such as:

- Date of collection
- Short definition
- Data type, field length, and format
- Control points used
- Collection method, field notes, and maps
- Data processing steps
- Assessment of the reliability of source data
- Data quality reports

Access to this information can help GIS personnel determine if the data are appropriate for their use, thereby minimizing risks associated with using the wrong data or using data inappropriately. According to Chrisman (10), the only ethical and probably best legal strategy for those who produce spatial data is to reveal more information about the data (metadata) so that users can make informed decisions. Eagan and Ventura's article (11) contains a sample of a generic environmental data lineage report. The U.S. Environmental Protection Agency's (EPA's) new locational data policy requires contractors to estimate data accuracy and provide information about the lineage of the data (12).

Positional Accuracy

Anyone who has used a map has probably come across features that are not located where the map says they should be located and has experienced low positional accuracy. (Undoubtedly, they have also detected features that were not on the map, but that is a different issue.) Positional accuracy, frequently referred to as horizontal error, is how close a location on a map is to its "true" ground position. Features may be located inaccurately on maps for many reasons, including (13):

- Poor field work.
- Distortion of the original paper map (temperature, humidity).
- Poor conversion from raster to vector or vector to raster data.
- Data layers are collected at different times.
- Natural variability in data (tides, vegetation, soil).
- Human-induced changes (altering reservoir water levels).
- Movement of features (due to scale of the map and printing constraints) so they can be easily discerned by the map reader.
- Combining maps with different scales.
- Combining maps with different projection and coordinate systems.
- Different national horizontal datum in source materials.
- Different minimum mapping units.

Positional accuracy has two components: bias and precision. Bias reflects the average positional error of the sample points and indicates a systematic discrepancy (e.g., all locations are 7 feet east of where they should be). Estimating precision entails calculating the standard deviation of the dispersion of the positional errors. Usually, root mean square error (RMSE) is reported as the measure of positional accuracy, but it does not distinguish bias from precision (14). RMSE is frequently monitored during digitizing to minimize the introduction of additional positional error into the GIS.

To determine positional accuracy, one must compare the location of spatial data with an independent source of higher accuracy. Federal agencies that collect data and produce maps adhere to National Map Accuracy Standards (NMAS) for positional accuracy. Maps such as United States Geological Survey (USGS) topographic maps that conform to NMAS carry an explicit statement on them. Other groups also have developed standards for large-scale mapping (15).

NMAS for positional accuracy require that not more than 10 percent of well-defined points can be in error by more than one-thirtieth of an inch for maps at a scale of 1:20,000 or larger. For smaller scale maps, not more than 10 percent of well-defined points can be in error by more than one-fiftieth of an inch (16). Thus, less than 10 percent of the well-defined locations on a USGS 1:24,000 map can stand more than 40 feet from their “true” location; the other 90 percent of the well-defined points must stand less than 40 feet from their “true” location. Table 3 shows the acceptable positional accuracy for commonly used maps. Note that as scale decreases from 1:1,200 to 1:100,000, positional accuracy decreases.

Several important issues relate to NMAS. First, not all maps adhere to NMAS, which means their positional accuracy may be lower than NMAS or may be unknown. Second, NMAS do not indicate the location of points in error. Third, 10 percent of the well-defined points can have a positional error greater than the standards allow, but neither the location nor the magnitude of these errors are known. Fourth, NMAS apply to well-defined points; therefore, areas that are not well defined may have even lower positional accuracy. The implication of these errors in location is that users should use caution in making decisions that require high positional accuracy. Positional accuracy issues are particularly troublesome for GIS operations on small-scale maps or when combining large-scale maps (1:1,200) with small-scale maps (1:100,000).

Table 3. NMAS Horizontal (Positional) Accuracy

Scale	1 Inch = x Feet	Horizontal Accuracy +/- Feet
1:1,200	100	3.33
1:2,400	200	6.67
1:4,800	400	13.33
1:12,000	1,000	33.33
1:24,000	2,000	40.00
1:63,360	5,280	105.60
1:100,000	8,333	166.67

Recently, global positioning systems (GPS), which the U.S. military developed, have helped to obtain more accurate feature locations. GPS is not without error, however. The list below notes some of the possible sources of error associated with GPS use, some of which can be controlled while others cannot (17):

- Errors in orbital information.
- Errors in the satellite clocks.
- Errors in the receiver clocks.
- Ionospheric or tropospheric refraction.
- Deliberate degrading of the satellite signal.
- Obstructions that block the signal.
- Reflection of the GPS signal off buildings, water, or metal.
- Human error.

The importance of positional accuracy depends on the intended use of the data. In an urban area, a positional error of 1 foot on a tax map may be unacceptable because 1 foot may be worth millions of dollars. In a rural area, however, tax boundaries mapped within 10 feet of their surveyed location may be accurate enough (6). Somers (18) reports that positional accuracy of 10 to 20 feet may be sufficient for environmental analysis. She says the cost of increasing accuracy to 5 feet could increase the cost of data collection by a factor of 10. The decision-maker must determine the needed positional accuracy.

Attribute Accuracy

Attribute accuracy refers to how well the description of a characteristic of spatial data matches what actually exists on the ground. For some spatial data, the location does not change over time, but the value of the attribute does (e.g., the location of a census tract does not change, but the population within a census tract changes). Attribute accuracy is reported differently for continuous data (i.e., elevation, which has an infinite number of values) or discrete data (i.e., gender, which has a finite number of values).

NMAS exist for elevation contour lines on topographic maps. NMAS for vertical accuracy state that not more than 10 percent of the points tested shall be in error by more than one-half of the contour interval (16). A well-defined point on a USGS topographic map with a 10-foot contour interval could vary by 10 feet because the actual elevation could be 5 feet higher or lower than the map indicates. The implications of these errors are similar to the ones for positional accuracy. In addition, errors in elevation are important because small changes in elevation may significantly affect some GIS analysis operations such as the determination of aspect, slope, viewshed, and watershed boundaries.

NMAS do not exist for discrete variables such as land use derived from satellite imagery. Instead, a classification matrix reports attribute accuracy. Field checking or checking a portion of the classified image against a map of higher accuracy determines the accuracy of the land use classification. The result of the comparison is a table from which to calculate overall, producer's, and user's accuracy. Table 4 is an example of a classification accuracy matrix.

Table 4. Example of a Classification Accuracy Matrix (19)

Classified Data (Satellite Image) Number of Cells	Reference Data ("Ground Truth") Number of Cells: Forest	Water	Urban	Total
Forest	28	14	15	57
Water	1	15	5	21
Urban	1	1	20	22
Total	30	30	40	100

Overall Accuracy (sum of the main diagonal)

$$\frac{63}{100} = 63\%$$

Producer's Accuracy (column total)		User's Accuracy (row total)	
Forest	$= \frac{28}{30} = 93\%$	Forest	$= \frac{28}{57} = 49\%$
Water	$= \frac{15}{30} = 50\%$	Water	$= \frac{15}{21} = 71\%$
Urban	$= \frac{20}{40} = 50\%$	Urban	$= \frac{20}{22} = 91\%$

Overall accuracy is the percentage of correctly classified cells calculated as the sum of the main diagonal (19). Producer's accuracy is the total number of correct pixels in a category divided by the total number of pixels of that category as derived from the reference data (column total). It corresponds to how well the person classifying the image (the "producer") can correctly classify or map an area on the earth. In this example, the producer most accurately classified forested land (93 percent). User's accuracy describes the probability that a sample from the classified area actually represents that category on the ground. The map "user" is concerned about the map's reliability. In this example, the most accurately classified land use from the user's perspective is urban (91 percent).

The significance of overall, producer's, and user's accuracy depends on the intended use of the data. As an example, Chrisman (20) says that the error in distinguishing wetland from pasture may not matter to someone estimating open space, but the difference is critical if the person is estimating the amount of wildlife habitat available. Story and Congalton (19) provide an example of how to interpret a classification matrix. A forester looks at the classification matrix and sees that forest classification is 93 percent accurate (producer's accuracy); therefore, the analyst did not identify only 7 percent of the forest on the ground. Once the forester field checks the supposed forested area, she finds that only 49 percent (28 cells) of the sites mapped as forest are actually forest; the rest are water (14 cells) or urban (15 cells) areas.

A report of overall, producer's, and user's accuracy can help decision-makers determine the appropriateness of the classified image for their use by identifying potential errors in classification. This can help direct field work, which can improve the classification of the image and perhaps subsequent images. Because GIS analysis frequently uses land use, decision-makers need to know that significant variability can result when several analysts classify the same image. Bell and Pucherelli (21) found that consistency in classification can improve by having one person classify the entire image. McGwire (22) even found significant differences between analysts in unsupervised classification of Landsat imagery. Computers primarily perform unsupervised classification, which implies that different analysts would classify the same image in the same way.

Logical Consistency

Logical consistency focuses on flaws in the logical relationships among data elements. For example, a vector GIS should label all polygons with only one label per polygon, and all polygons should be closed. Logical inconsistency can also occur by collecting data layers at different times or from different scale maps with different positional accuracies. For example, the edge of a lake on the hydrology data layer should coincide with the edge of land in the land use data layer. If data on the lake were collected during a wet year rather than a dry year, the lake's volume would be higher than normal, affecting its location on the map. If land use data for the same area were collected during a dry year, the boundary of the lake on the two layers would not be the same.

Logical inconsistencies usually do not appear until the two maps are overlaid and the boundaries do not coincide (see Figure 2). The user must determine the "correct" location of the feature that appears misaligned on one or more data layers. The inconsistency between the

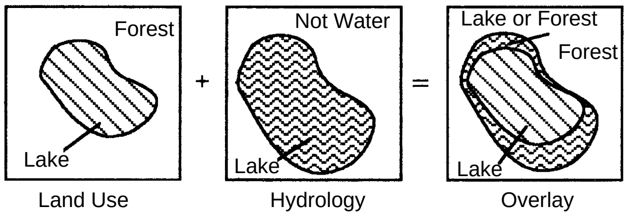


Figure 2. Logical inconsistency in lake and forest location.

location of the two layers resolves through a process called conflation. All maps are adjusted so that the feature on each data layer lines up with the same feature on the base map.

Completeness

Completeness focuses on the adequacy of data collection procedures. Robinson and Frank (5) discuss two kinds of uncertainty associated with collecting spatial data that can lead to error. One type of uncertainty is the inability to measure or predict an inherently exact characteristic or event with certainty. Examples of this are blunders in data collection or measurement error, neither of which can be accurately predicted. The other kind of uncertainty is associated with concepts that are inherently ambiguous. Crisp data sets, such as property boundaries, have little ambiguity; the only issue related to error is the positional accuracy in measuring the boundary. Because land use data are not crisp data sets, the challenge is to accurately represent an inherently inexact concept.

Although we know spatial data are variable, our classification systems generally ignore the second type of uncertainty. Analysts map data as though all variables had exact boundaries and all polygons consisted of homogeneous data. Burrough (4) reports that spatial variation of natural phenomena is "not just a local noise function or inaccuracy that can be removed by collecting more data or by increasing the precision of measurement, but is often a fundamental aspect of nature that occurs at all scales. . . ."

Mapping spatial data is a function of how humans aggregate and disaggregate data either in space, categories, quantities, or time; spatial data seldom exist in nature the way maps depict them (23). Data and relationships between data are sensitive to the scale and the zoning system in which the data are reported (24, 25). The modifiable area unit problem occurs because an analyst can recombine a given set of units or zones into the same total number of units producing very different results (see Figure 3).

The scale problem occurs because an analyst can combine a set of small units into a smaller number of larger units, which can change the inferences that can be

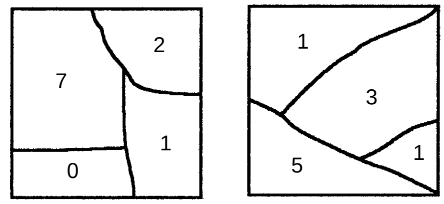


Figure 3. Modifiable area unit. (Number of units is constant; location of units changes.)

made from the data. In Figure 4, the area containing the highest values changes from the southwest corner in the first picture to the northern half in the second picture. For example, water quality data are scale-dependent because they vary based on the size and location of the collection area (e.g., adjacent to a point source discharge, a stream segment, the entire river, or the lake the river discharges into).

Kennedy (25) reports on a similar problem known as the small number problem. This problem occurs when calculations use a percentage, ratio, or rate for a geographic area for which the population of interest (denominator) is sparse or the numerator is a rare event (1 case of cancer per 1 million people). GIS-generated maps may highlight a statistically insignificant change in rare events. Small, random fluctuations in the numerator may cause large fluctuations in the resulting percentage, ratio, or rate. If policy-makers use these maps, priorities for public health policy may change because of the erroneous belief that an area is experiencing more unwanted rare events.

Data can be collected using a tag- or count-based system, which affects their usefulness. The tag approach categorizes items based on the dominant or average attribute and is ideal for planners who want only one value for each area. For example, each polygon in a county soil survey is tagged with one soil type. According to soil taxonomy rules, however, only about 35 percent of a delimited area on a soil survey must match its classification, and up to 10 percent may be a radically different soil (26). Although the text in the soil survey sets limits on data accuracy by listing major impurities found with each soil type, the GIS seldom carries that information because analysts only digitize soil boundaries and label data with the dominant attribute. This

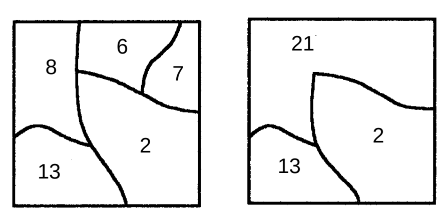


Figure 4. Scale problem (number of units changes).

leads to the depiction of apparently homogeneous soil units although the text specifies that the data are not homogeneous (27).

Some soil or land cover phenomena, even though present in small quantities and thus not mapped, may have great significance for hydrologic models, which makes the tag approach to data collection troublesome. Data collected using the count system allow the analyst to tabulate the frequency of occurrence or areal extent of a particular phenomenon. Environmental modelers prefer count data but are usually forced to use tag data, which can introduce error into their models (26). The new digital soils databases, STATSGO and SSURGO, are collected and depicted using a count format, which will help experienced analysts use the data more fully. Figure 5 shows the difference between tag and count methods of data collection.

Data are seldom complete because analysts use classification rules to indicate how homogeneous an area must be before it is classified a particular way (e.g., more than 50 percent, more than 75 percent). Another decision an analyst must make is where to draw the boundary between two different areas; it is seldom clear where a forest leaves off and a rural development begins. Analysts must also decide how or if to show inclusions (e.g., a forested area in the middle of agricultural land uses).

Temporal Accuracy

Collecting data at different times introduces error because the variable may have changed since data collection. The effect of time, reported as the date of the source material, depends on the intended use of the data. Some natural resource data have daily, weekly, seasonal, or annual cycles that are important to consider. For example, obtaining land use data from remotely sensed imagery in November for North Dakota produces a very different land use map than data analysts obtain during the July growing season.

In addition, demographic and land use information changes quickly in a rapidly urbanizing area. Data collected at several times can produce logical inconsistency between data layers, forcing the analyst to adjust the location of features to coincide with the base map. Another problem with collecting data at different times is that data may be collected using different standards, which may not be apparent to the user (4).

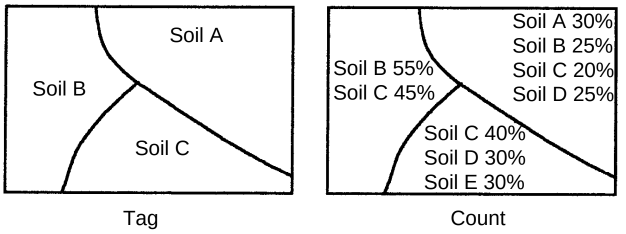


Figure 5. Tag and count methods of data collection.

Source Errors in a GIS

Source (or inherent) error derives from errors in data collection. The amount of error present in collected data is a function of the assumptions, methods, and procedures used to create the source map (28). Primary data refers to data collected from field sampling or remote sensing. Causes of the errors associated with this data are (3, 4, 8, 14, 29):

- Environmental conditions (e.g., temperature, humidity).
- Sampling system (e.g., incomplete or biased data collection).
- Time constraints.
- Map projection.
- Map construction techniques.
- Map design specifications.
- Symbolization of data.
- Natural variability.
- Imprecision due to vagueness (e.g., classifying a forest).
- Measurement error from unreliable, inaccurate, or biased observers.
- Measurement error from unreliable, inaccurate, or biased equipment.
- Lab errors (e.g., reproducibility between lab procedures and between labs).

The process of converting primary data to secondary data (usually a map) introduces additional error. Many of the data layers that a GIS analyst acquires are secondary data. Some of the errors associated with mapmaking are (3):

- Error in plotting control points.
- Compilation error.
- Error introduced in drawing.
- Error due to map generalization.
- Error in map reproduction.
- Error in color registration.
- Deformation of the material (temperature, humidity).
- Error introduced due to using a uniform scale.
- Uncertainty in the definition of a feature (boundary between two land uses).
- Error due to feature exaggeration.
- Error in digitization or scanning.

Converting paper maps to digital data for entry into a GIS (tertiary data) introduces still more error (the errors generated from converting paper maps into a digital format are discussed in the section on input error), in part because the purpose for which the data was collected differs from the intended use of the data.

Many types of error are associated with data collection:

- Data for the entire area may be incomplete.
- Data may be collected and mapped at inappropriate scales.
- Data may not be relevant for the intended application.
- Data may not be accessible because use is restricted.
- Resolution of the data may not be sufficient.
- Density of observations may not be sufficient.

The following discussion explains these types of errors.

Data for the Entire Area May Be Incomplete

An incomplete data record may be due to mechanical problems that interrupt recording devices, cloud cover or other types of interference, or financial constraints. Possible solutions to this problem include collecting additional data for the incomplete area, using information from a similar area, generalizing existing large-scale maps to match the less detailed data needed, or converting existing small-scale maps to large-scale maps to obtain data at the desired scale. Collecting additional data may not be a feasible solution because of time or money constraints. Extrapolating data from the surrogate area to the desired area can cause problems because the areas are not identical and the scale, accuracy, or resolution of the surrogate area data may be inappropriate for the intended use. The section on analysis/manipulation of data within a GIS covers the effect of generalization on data quality as well as the effect of converting small-scale maps to large-scale maps.

Data May Be Collected and Mapped at a Scale That Is Inappropriate for the Application

A variety of guidelines suggest the appropriate map scale to use for various applications (see Table 5). Also, some maps and digital databases suggest the type of application for which they are appropriate (e.g., the STATSGO digital soil database is suitable for state and regional planning, whereas SSURGO is suitable for local level planning). Tosta (30) cites an example of combining wetland data with parcel boundaries to determine ownership of the land containing a wetland. If wetland mapping was done to plus or minus 100 feet positional accuracy and parcels are 40 feet wide, then the scale of the wetland map is inappropriate for determining if a wetland is located on a specific parcel.

Table 5. Relationship Between Map Scale and Map Use (6)

Map Scale	Map Use
1:600 or larger	Engineering design
1:720 to 1:1,200	Engineering planning
1:2,400 to 1:4,800	General planning
1:6,000 and smaller	Regional planning

Identifying the optimal scale of the necessary data is crucial because at some point, the cost of collection and storage exceeds the benefits of increasing the map scale. Lewis Carroll (1893) summed up the quest for data mapped at an ever larger scale and the problems associated with large-scale maps:

> "What do you consider the *largest* map that would be really useful?"
>
> "About six inches to the mile."
>
> "Only *six inches!"* exclaimed Mein Herr. "We very soon got to six *yards* to the mile. Then we tried a *hundred* yards to the mile. And then came the grandest idea of all! We actually made a map of the country, on the scale of *a mile to the mile!"*
>
> "Have you used it much?" I enquired.
>
> "It has never been spread out, yet," said Mein Herr. "The farmers objected: they said it would cover the whole country, and shut out the sunlight! So now we use the country itself, as its own map, and I assure you it does nearly as well."

Data Collected May Not Be Relevant for the Intended Application

Frequently, using surrogate data is quicker or cheaper than collecting needed data (e.g., Landsat imagery rather than data field collection used to determine land use) (4). The accuracy and classification scheme used in collecting the data depends on the intended use of the data, which may not coincide with the analyst's purpose. For instance, soil maps were developed to aid farmers in determining what crops they should plant and for estimating crop yield. Soil maps, however, see wide use for very different purposes (e.g., hydrologic and other environmental models). In addition, STORET data, collected at points, are typically extrapolated to represent water quality in an entire stream stretch.

Data May Not Be Accessible Because Use Is Restricted

An example of restricted data is Census data on individual households. An agency may not want to release data that reveal the location of endangered species. Another example is that people may not even want the information mapped. For example, some cavers do not want to reveal the location of caves to the U.S. Forest Service, which is charged under the federal Cave Resources Protection Act with protecting caves, because they think the best way to protect the caves is to not map them (31). The National Park Service is putting the location of petroglyphs in the Petroglyph National Monument into a GIS. Making their location known to the public, however, is troublesome because this may, in fact, encourage their vandalism (32). Other problems in obtaining data include difficulty in acquisition even if access is not restricted, expensive collection or input, or unsuitable format (4, 14).

Resolution of the Available Data May Not Be Sufficient

Spatial resolution is the minimum distance needed between two objects for the equipment to record the objects as two entities; that is, resolution is the smallest unit a map represents. To obtain an approximation of a map's resolution, divide the denominator of the map scale by 2,000 to get resolution in meters; for instance, a 1:24,000-scale map has a resolution of approximately 12 meters (33).

Resolution relates to accuracy in that different map scales conform to different accuracy standards. Two air photos shot from the same camera at the same distance above the ground have the same scale. If one photo has finer grain film, however, smaller details are evident on it, and this photo produces a map with higher resolution (34). According to Csillag (33), analysts cannot simultaneously optimize attribute accuracy and spatial resolution. As spatial resolution increases, attribute complexity increases (35). Also, the finer the spatial resolution, the greater the probability that random error significantly affects a data value.

Resolution of the data is not necessarily the same as the size of a raster cell in a database. Statistical sampling theory suggests using a raster cell size that is half the length (one-fourth of the area) of the smallest feature an analyst wishes to record. Raster data have a fixed spatial resolution that depends on the size of the cell employed, but a GIS analyst can divide or aggregate cells to achieve a different cell size. Frequently, an analyst transforms data collected at one level of resolution to a higher level of resolution than existed in the original source material. According to Everett and Simonett (23), "Geographic analysis, however, can be no better than that of the smallest bit of data which the system is capable of detecting." Vector data are limited by the resolution of input/output devices, limits on data storage, and the accuracy of the digitized location for individual points (36). The spatial resolution of the database and the processes that operate on it should be reduced to a level consistent with the data's accuracy.

The spatial resolution needed depends on the intended use of the data, cost, and data storage considerations. As resolution increases, so does the cost of collection and storage. Resolution sufficient to detect an object means that an analyst can reveal the presence of something. Identification, the ability to identify the object or feature, requires three times the spatial resolution of detection. Analysis, a finer level of identification, requires 10 to 100 times the resolution that identification needs (23). Increasing resolution increases the amount of data for storage, with storage requirements increasing by the square of the resolution of the data. For example, if the resolution of the data needs to change from 10-meter to 1-meter pixels, file size increases by 10^2 or 100 times (14).

Density of Observations May Be Insufficient

The density of observations serves as a general indicator of data reliability (4). Users need to know if sampling was done at the optimum density to resolve the pattern. Burrough determined that boulder clay in The Netherlands could be resolved by sampling at 20-meter intervals or less, whereas coversand showed little variation in sampling from 20- to 200-meter intervals.

Some strategies for reducing data collection errors are to:

- Adhere to professional standards
- Allocate enough time and money
- Use a rigorous sampling design
- Standardize data collection procedures
- Document data collection procedures
- Calibrate data collection instruments
- Use more accurate instruments
- Perform blunder checks to detect gross errors

Documenting data collection procedures and distributing them along with data allows potential users to determine if the data are suitable for their purposes. By not documenting procedures, errors in the source material are essentially "lost" by inputting the data to a GIS, and the errors become largely undetectable in subsequent GIS procedures. The result is that agencies that make decisions based on the GIS-generated map assume the source data are accurate, only to discover later that the map contains substantial errors in part due to errors in the source material.

Operational Errors in a GIS

Data input, storage, analysis/manipulation, and output can introduce operational errors. Digital maps, unlike paper maps, can accumulate new operational errors through GIS operations (8). Even if the input data were totally error-free, which the last section demonstrated is not the case, GIS operations can produce positional and attribute errors. The GIS operation itself determines to a large extent the types of errors that result.

Input Errors

The process of inputting spatial and attribute data can introduce error. The major sources of input error are manual entry of attribute features and scanning or digitizing spatial features. Manual entry errors include incomplete entry of attribute data, entering the wrong attribute data, or entering the right attribute data at the wrong location. Digitizing errors originate from equipment, personnel, or the source material (see Table 6).

Digitizing errors, such as under- and overshoot of lines and polygons that are not closed, can introduce error (see Figure 6). GIS software can "snap" lines together that really do not connect. Depending on the tolerance

Table 6. Types of Digitizing Errors (4, 14, 37)

Personnel Errors
Changes in the origin
Incorrect registration of the map on the digitizing table
Creation of over- and undershoots
Creation of polygons that are not closed
Incomplete spatial data when data are not entered
Duplication of spatial data when lines are digitized twice
Line-following error (inability to trace map lines perfectly with the cursor)
Line-sampling error (selection of points used to represent the map)
Physiological error (involuntary muscle spasms)
Equipment Errors
Digitizing table (center has higher positional accuracy than the edges)
Resolution of the digitizer
Differential accuracy depending on cursor orientation
Errors in Source Material
Distortion because source maps have not been scale-corrected
Distortion due to changes in temperature and humidity
Necessity of digitizing sharp boundary lines when they are gradual transitions
Width of map boundaries (0.4 mm) digitized with a 0.02-mm accuracy digitizer

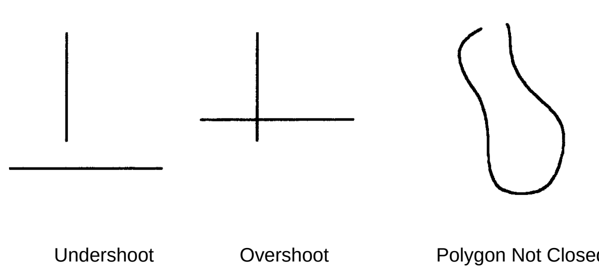


Figure 6. Common digitizing errors.

selected, this can result in the movement of both lines, which can decrease the accuracy of the resultant map.

Despite the long list of personnel errors associated with digitizing, a good operator probably contributes the least error in the entire digitizing process (38). Giovachino discusses methods that can help determine equipment accuracy, including checking the repeatability, stability, and effect of cursor rotation. Digitizing accuracy varies based on the width, complexity, and density of the feature being digitized but typically varies from 0.01 to 0.003 (3).

One problem with digitized data is that the data can imply a false sense of precision. Boundaries on paper maps are frequently 0.4 mm wide but are digitized with 0.02-mm accuracy. The result is that the lines are stored with 0.02-mm accuracy, implying a level of precision that far exceeds the original data.

Minimizing digitizing errors is important because the errors can affect subsequent GIS analysis. Campbell and Mortenson (39) provide a list of procedures they used to reduce errors associated with digitizing and labeling:

- Use log sheets to ensure consistency and accountability, and to provide documentation.
- Check for completeness in digitizing all lines and polygons.
- Check for complete and accurate polygon labeling.
- Set an acceptable RMSE term for digitizing (usually 0.003).
- Always overshoot rather than undershoot when digitizing.
- Overlay a plot of the digitized data with the source map to check lines and polygons. If light passes between the digitized line segment and source map, redigitize it.
- Check digitized work immediately to provide feedback to the digitizer operator and to help identify and correct systematic errors.
- Limit digitizing to less than 4 hours a day.
- Involve people in doing GIS-related jobs other than digitizing to decrease turnover and increase the level of experience.

Storage Errors

Data storage in a GIS usually involves two main types of errors. First, many GIS systems have insufficient numerical precision, which can introduce error due to rounding. Integers are stored as 16 or 32 bits, which have four significant figures. Real numbers are stored as floating point numbers either in single precision (32 bit, 7 significant figures) or double precision (64 bit, 15 or 16 significant figures). If the data in a GIS range from fractions of a meter to full UTM coordinates, typical 32-bit GIS systems cannot store all the numbers. Using double precision (64 bits) reduces this problem but increases storage requirements.

Second, GIS processing and storage usually ignore significant digits (data precision). As a result, the precision of GIS processing frequently exceeds the accuracy of the data (40). When a GIS converts a temperature recorded and entered as 70 degrees Fahrenheit (nearest degree) to centigrade, the GIS stores the temperature as 21.111 degrees rather than 21 degrees, which the significant figures in the original temperature measurement would dictate. Using the accuracy of the data, not the precision of floating point arithmetic, partially resolves this but requires the user to make a special effort because the GIS does not automatically track significant figures.

Analysis/Manipulation Errors

GIS analysis/manipulation functions, designed to transform or combine data sets, also can introduce errors. These errors originate from the measurement scale used or during data conversion (vector to raster and raster to vector), map overlay, generalization, converting small-scale to large-scale maps, slope, viewshed, and other analysis functions. One of the biggest problems associated with GIS use is that data in digital form are subject to different uses than data in paper form because the user has access to multiple data layers.

Measurement Scale

Four measurement scales can depict spatial data: nominal, ordinal, interval, or ratio scales. A name or letter describes nominal data (e.g., land use type, hydrologic soil group C). Performing mathematical operations such as addition and subtraction on nominal data is meaningless. Ordinal or ranked data have an order to them such as low, medium, and high. Interval data have a known distance between the intervals such as 0, 1 to 5, 6 to 9, more than 9. Ratio data are similar to interval data except ratio data have a meaningful zero (e.g., temperature on the Kelvin scale).

Often during GIS operations, analysts convert interval or ratio data into nominal data (e.g., low slope is 0 to 3 percent, medium slope is 4 to 10 percent), resulting in a loss of information. Analysts should preserve the original slope values in the GIS in case the user later wants to modify the classification scheme. Robinson and Frank (5) describe the tradeoff between information content and the meaning that can be derived from it, which partly helps explain why interval data are frequently converted to nominal data. The authors identify a continuum progressing from nominal data at one end that is highly subjective, has low information content, and high meaning (low slope means something to the average user) to ratio data that has low subjectivity, high information content, and low meaning (a slope of 7 percent may not mean much to the average user).

Data Conversion

Errors can occur in converting a vector map to a raster map or a raster map to a vector map. For instance, remotely sensed data are collected using a raster-based system. Using a vector GIS, however, requires conversion from raster to vector data. The size of the error depends on the conversion algorithm, complexity of features, and grid cell size and orientation (13).

A line on a vector map converted to a raster map has lower accuracy in the raster representation because vector data structures store data more accurately than raster ones. When polygons in a vector GIS are converted to a raster GIS, the coding rule usually used assigns the value that covers the largest area within the cell of a categorical map to the entire cell (see Figure 7). For example, when placing a grid over a vector map with an urban land polygon adjacent to an agricultural polygon, the cell placement can include part of both polygons. If the resultant cell comprises 51 percent urban and 49 percent agricultural land, the cell is assigned 100 percent urban. Converting a numerical map between raster and vector systems requires spatial interpolation procedures. GIS software packages use different interpolation methods that can produce a different output even when using the same input data.

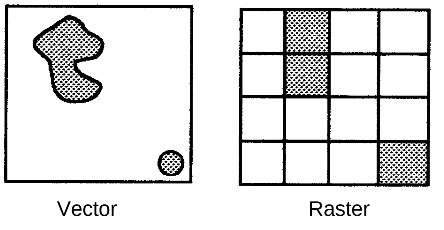


Figure 7. Polygon conversion from vector to raster data.

Map Overlay

Map overlay, used extensively in planning and natural resource management, is the combining of two or more data layers to create new information. In a vector GIS, slivers or spurious polygons can result from overlaying two data layers to produce a new map (slivers cannot be formed in a raster-based GIS). When combining the data layers, lines do not coincide, resulting in the creation of a new polygon or sliver that did not exist on either layer (see Figure 8). Unfortunately, as accuracy in digitizing increases, so does the number of slivers (41). Positional error in the boundaries can occur because of

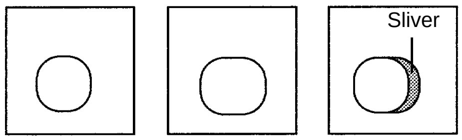


Figure 8. Sliver example.

mistakes in measuring or converting the data to digital form, incremental expansion or recession of a real world boundary over time, or the fact that certain boundaries are difficult to determine and thus are generalized differently (42).

The number of map layers, accuracy of each map layer, and the coincidence of errors at the same position from several map layers all determine the accuracy of the map overlay procedures (43). Using probability theory, Newcomer and Szajgin determined that the highest accuracy to expect from a map overlay is equal to the accuracy of the least accurate map layer. The lowest accuracy in map overlay occurs when errors in each map occur at unique points.

In the quest for more accurate results, GIS modelers have increased the complexity of their models and therefore have increased the number of data layers needed. Guptill (44) states, "Conventional wisdom would say that as you add more data to the solution of a problem, the likelihood of getting an accurate solution increases. However, if each additional data layer degrades the quality of the combined data set, and hence the accuracy of the solution, then additional data sets may be counterproductive."

Generalization

Monmonier (45) provides an extensive discussion of geometric and content generalization procedures used in map-making. Table 7 lists common types of generalization. Generalizing data on a map helps to focus the user's attention on one or two types of information and to filter out irrelevant details. Generalizing is performed by reducing the scale of the data; a 1:24,000-scale map can be generalized to a 1:100,000-scale map so that all data layers have the same scale. With generalizing, areas on a large-scale map become point or line features on a small-scale map (35). Obtaining some measurements from small-scale maps, however, requires caution. For example, a map may depict a 40-foot wide road as a single line one-fiftieth of an inch wide. On a 1:100,000 map, one-fiftieth of an inch translates into a 160-foot wide road—four times the actual width of the road.

Several studies have pointed to errors that can result from generalization. Wehde (46) compared soil maps generated from 0.017-acre grid cells and 11 progressively increasing grid cell sizes. He found that as grid cell size increased, map accuracy decreased. More recently, Stoms (47) found that generalizing a habitat map from 1 to 25, 100, and 500 hectares decreased the number of habitat types and the number of species predicted.

Table 7. Generalization Operations (45)

Geometric Generalization
Generalizing a Line
Simplification
Displacement
Smoothing
Enhancement
Selection
Generalizing a Point
Selection
Displacement
Graphic association
Abbreviation
Aggregation
Area conversion
Generalizing an Area
Selection
Simplification
Displacement
Smoothing
Enhancement
Aggregation
Dissolution
Segmentation
Point conversion
Line conversion
Content Generalization
Selection
Classification

Transforming Small-Scale Maps to Large-Scale Maps

Converting small-scale maps (1:250,000) to large-scale maps (1:24,000) is advisable only if the analyst fully appreciates the effect of this procedure on map quality. Data mapped at a small scale are subject to different accuracy standards than data mapped at a large scale. Connin (48) reports, "Problems with accuracy arise when positions are reported to decimal parts of a foot or meter, but the method of data capture may cause the positional error to be as much as hundreds of feet or meters." Yet when converting the data from small- to large-scale, the data appear to have the accuracy of the large-scale map. Theoretically, data should not be transformed and used at a scale larger than the scale of the document from which the data are derived (3).

Slope and Viewshed

GIS software packages use a variety of algorithms to calculate slope and viewsheds and can produce very different results. Algorithms are an unambiguous set of rules or a finite sequence of operations used to carry out a procedure. Smith, Prisley, and Weih (49) used six different GIS algorithms to determine slope on 5,905 acres of land in order to calculate the amount of land deemed unsuitable for timber harvesting. They found that unsuitable land varied from 175 to 1,735 acres, indicating that different algorithms produce very different results. Felleman and Griffin (50) found that GIS packages with different algorithms generate alternate viewsheds (the area that can be seen from a point).

Output Errors

A variety of errors are associated with data output:

- Output devices create error.
- Paper shrinks and swells.
- Line implies certainty that may not exist because boundaries are gradual.
- A cell or polygon implies homogeneity.
- Scale can be modified to imply higher accuracy than exists in the source data.
- Precision can be modified to imply higher precision than exists in the source data.
- Depiction of symbols and colors may not follow convention.

An important problem associated with GIS-generated maps is that users make informal assessments about data quality, partially based on how they perceive the quality of the output. A hand-drawn map connotes a lower level of accuracy than a five-color, GIS-produced map complete with scale and agency logo. Another problem with output is that distinguishing highly accurate data from less accurate data is impossible on a GIS-generated map. Users want the output from a GIS to look like maps they usually see, perpetuating the notion that lines mark exact boundaries and that polygons or cells are homogeneous. Maps that federal mapping agencies produce frequently follow NMAS, but GIS-generated maps seldom adhere to published map accuracy standards. An agency could require that GIS map products meet NMAS, which would establish and maintain data standards from data collection to output.

A pen stroke of one-fiftieth of an inch on an output device translates to an error of 40 feet on the ground for a 1:24,000-scale map (6). Small changes in paper maps due to changes in temperature and humidity can represent several feet on the ground. As previously noted, analysts can modify the scale of GIS maps to whatever they desire. The basic rule of informational integrity is that the implied precision of data output should not exceed the precision (spatial, temporal, or mathematical) of the least precise input variable (26).

GIS-generated maps probably do not differ significantly from paper maps in their implication that lines and polygons on the map represent certainty and homogeneity. GIS-generated maps, however, may not depict standard symbols, sizes, shapes, colors, and orientation. For example, paper geological maps use dashed lines to show inferred, rather than actual, field collected data, but geological maps in a GIS may not follow the same convention (27). Cartographers conventionally use blue lines to indicate water, but a GIS map-maker can show water as red rather than blue.

Even more troublesome are the color schemes that some analysts use in depicting model output. Analysts often give little thought to assigning the colors to model results depicted as ordinal rankings. For example, areas of high erosion might be blue, medium erosion might be red, and low erosion might be green. This selection of colors ignores the intuitive meaning that people assign to colors. It has been suggested that the color ordering used in stop lights might provide a better option. In that case, areas of high erosion would be red, medium erosion would be yellow, and low erosion would be green.

Error Reduction Techniques

Although GIS users and researchers develop error reduction strategies, ultimately users must rely on GIS software developers to implement new error reduction techniques in GIS packages. Error reduction techniques range from simple software warnings to prohibiting a user from performing selected GIS procedures. Dutton (51) predicts that future GIS programs will automate data manipulation (i.e., size, format, and placement of feature labels on maps) in keeping with standard cartographic principles. Dutton (51) and Beard (8) also predict that future GIS packages will enforce metadata-based constraints such as operations that are illegal or illogical (e.g., determining the average value of nominal data such as land use), or are inadvisable (e.g., overlaying maps with widely different scales).

Another change Dutton anticipates is that software vendors will include information in manuals that explains how executing a specific command may affect the database. Graphic techniques to depict error are being developed for nonexpert users while experts tend to use spatial statistics. Felleman (52) and Berry (53) present an interesting graphic portrayal of an error map that may indicate the future of error maps. Additional research must determine what effect errors will have on decision-making.

Error Management

Ultimately, the decision-maker must determine what to do with the information in this paper. A decision-maker has a variety of possible courses of action, ranging from prudent steps that attempt to minimize error and the effect it has on decisions, to other less useful options. Possible actions are to:

- Abandon use of a GIS.
- Ignore the error associated with GIS use.
- Attempt to collect “error-free” data.
- Determine if the data are accurate enough for the intended purpose.
- Develop and use data quality procedures.
- Obtain and use an error report with GIS-generated output.
- Ask that GIS-generated maps show potential errors.
- Continually educate users about the appropriate use of spatial data.

First, the decision-maker could abandon any attempt to use a GIS because of the errors associated with its use. At times, this may be the appropriate strategy, but this approach ignores the potential benefits associated with GIS use.

Second, the decision-maker could ignore the error associated with GIS use and continue to use the GIS for decision-making. This type of “head in the sand” approach is not advisable because of the potential liability associated with making decisions based on inaccurate data.

Third, the decision-maker could engage in an expensive and time-consuming effort to collect highly accurate error in hopes that error becomes a nonissue. Depending on the intended use of the data, the cost of collecting more accurate data may exceed the benefit.

Fourth, the decision-maker could assess whether the information available is accurate enough for the intended purpose. If data quality is too low, the decision-maker may opt to collect new data at the desired quality. If collecting additional data is not possible, the decision-maker can explore what types of decisions are possible given the attainable data quality. For instance, Hunter and Goodchild (54) found that the data they were using were suitable only for initial screening rather than for regulatory and land-purchasing decisions.

Fifth, procedures to ensure high quality data could be developed and used in the data collection, input, and manipulation stages of building a GIS database.

Sixth, the decision-maker could require a quantitative or at least a qualitative report on the sources, magnitude, and effects of errors. The absence of an error report does not mean the map is error-free (36). Dutton (51) predicts that in the near future users of geographic data will demand error reports, confidence limits, and sensitivity analyses with GIS-generated output.

Seventh, the decision-maker could ask for GIS-generated maps that adequately portray the error in the final map. For example, areas where the uncertainty is high could appear in red on maps. Another option is to place a buffer around lines to indicate the relative positional accuracy of a line or to show transition zones. Finally, an analyst can present the output in ways other than a dichotomous yes or no; instead, the analyst may use yes, maybe, or no depictions or even more gradations.

Finally, Beard (8) introduced the concept of directing efforts toward educating users about use error. She defines use error as the misinterpretation of maps or misapplication of maps to tasks for which they are not appropriate. “We can’t assume that GIS will automatically be less susceptible to misuse than traditional maps, and it may, in fact, exacerbate the problem by expanding access to mapped information.” Beard argues that money directed to reducing source and operational error, while important, may not matter if use error is large.

Conclusions

GIS is a powerful tool for analyzing spatial data. Everyone who uses GIS-generated output, however, must be aware of source errors and operational errors introduced during data input, storage, analysis/manipulation, and output. Increased awareness of the sources and magnitude of error can help decision-makers determine if data are appropriate for their use. Decision-makers cannot leave data quality concerns to GIS analysts because efforts to improve data quality are not without cost, and the decision-makers typically control funding.

Decision-makers must not get caught up in the glamour of the spatial analyses and outputs that a GIS can produce. These attributes may lead decision-makers to ignore issues associated with uncertainty, error, accuracy, and precision. Inexpensive digital data can make analysts and decision-makers ignore data quality. If subsequent management decisions are made based on poor quality data, the resultant decisions may turn out wrong. This would give decision-makers a jaded view of the usefulness of GIS. An adequate understanding of data quality issues can help decision-makers ask the right questions of analysts and avoid making decisions that are inappropriate given the data quality.

References

1. Walsh, S.J., D.R. Lightfoot, and D.R. Butler. 1987. Recognition and assessment of error in geographic information systems. Photogrammetric Eng. and Remote Sensing 53(10):1,423-1,430.

2. Openshaw, S. 1989. Learning to live with errors in spatial databases. In: Goodchild, M., and S. Gopal, eds. The accuracy of spatial databases. Bristol, PA: Taylor and Francis. pp. 263-276.

3. Thapa, K., and J. Bossler. 1992. Accuracy of spatial data used in geographic information systems. Photogrammetric Eng. and Remote Sensing 58(6):835-841.

4. Burrough, P.A. 1986. Principles of geographical information systems for land resources assessment. New York, NY: Oxford University Press.

5. Robinson, V.B., and A.U. Frank. 1985. About different kinds of uncertainty in collections of spatial data. In: Proceedings of Digital Representations of Spatial Knowledge, Washington, DC, Auto-Carto 7. Falls Church, VA: ASPRS and ACSM. pp. 440-449.

6. Nale, D.K. 1992. Do traditional map standards conflict with a GIS landbase? GIS World 5(7):50-53.

7. Smith, G.R., and D.M. Honeycutt. 1987. Geographic data uncertainty, decision-making, and the value of information. In: Proceedings of GIS '87, San Francisco, CA (October 26-30). Falls Church, VA: ASPRS and ACSM. pp. 300-312.

8. Beard, K. 1989. Dimensions of use and value of geographic information. In: Onsrud, H., H.W. Calkins, and N.J. Obermeyer, eds. In: Proceedings of use and value of geographic information: Initiative four specialist meeting, Tenants Harbor, ME (August). Vol. 89-7. Santa Barbara, CA: NCGIA.

9. National Committee for Digital Cartographic Data Standards. 1988. The proposed standards for digital cartographic data. Amer. Cartographer 15(1):11-140.

10. Chrisman, N. 1994. A vision of digital libraries for geographic information, or how I stopped trying to find the on-ramps to the information superhighway. Geo. Info. Sys. 4(4):21-24.

11. Eagan, P.D., and S.J. Ventura. 1993. Enhancing value of environmental data: Data lineage reporting. J. Environ. Eng. 119(1):5-16.

12. U.S. EPA. 1990. Information resources management policy manual. Washington, DC.

13. Lunetta, R.S., R.C. Congalton, L.K. Fenstermaker, J.R. Jenson, K.C. McGwire, and L.R. Tinney. 1991. Remote sensing and geographic information system data integration: Error sources and research issues. Photogrammetric Eng. and Remote Sensing 57(6):677-688.

14. Aronoff, S. 1989. Geographic information systems: A management perspective. Ottawa, Ontario: WDL Publications.

15. ASPRS. 1990. ASPRS accuracy standards for large-scale maps. Photogrammetric Eng. and Remote Sensing 56(7):1,068-1,070.

16. Thompson, M.M. 1987. Maps for America: Cartographic products of the U.S. Geological Survey and others, 3rd ed. Washington, DC: U.S. Geological Survey.

17. Ferguson, J. 1993. What is it we obtain from GPS? Professional Surveyor 13(4):3-4.

18. Somers, R. 1994. Is a GIS department the answer? Geo. Info. Sys. 4(6):29-32.

19. Story, M., and R.G. Congalton. 1986. Accuracy assessment: A user's perspective. Photogrammetric Eng. and Remote Sensing 52(3):397-399.

20. Chrisman, N.R. 1987. The accuracy of map overlays: A reassessment. Landscape and Urban Planning 14:427-439.

21. Bell, A.D., and M.J. Pucherelli. 1992. Audubon Lake island erosion study using aerial photography and geographic information systems. Report No. R-92-12. Denver, CO: U.S. Department of the Interior, Bureau of Reclamation.

22. McGwire, K.C. 1992. Analyst variability in labeling of unsupervised classifications. Photogrammetric Eng. and Remote Sensing 58(12):1,673-1,677.

23. Everett, J., and D.S. Simonett. 1976. Principles, concepts, and philosophical problems in remote sensing. In: Lintz, J., and D.S. Simonett, eds. Remote sensing of the environment. Reading, MA: Addison-Wesley.

24. Fortheringham, A.S. 1989. Scale-independent spatial analysis. In: Goodchild, M., and S. Gopal, eds. The accuracy of spatial databases. Bristol, PA: Taylor and Francis. pp. 221-228.

25. Kennedy, S. 1989. The small number problem and the accuracy of spatial databases. In: Goodchild, M., and S. Gopal, eds. The accuracy of spatial databases. Bristol, PA: Taylor and Francis. pp. 187-196.

26. Brown, D.A., and P.J. Gersmehl. 1987. Maintaining relational accuracy of geocoded data in environmental modeling. In: Proceedings of GIS '87, San Francisco, CA (October 26-30). Falls Church, VA: ASPRS and ACSM. pp. 266-275.

27. Fisher, P.F. 1987. The nature of soil data in GIS—error or uncertainty. In: Proceedings of the International Geographic Information Systems (IGIS) Symposium, Arlington, VA (November 15-18). Vol. III, pp. 307-318. Arlington, VA: Association of American Geographers.

28. Walsh, S.J. 1989. User considerations in landscape characterization. In: Goodchild, M., and S. Gopal, eds. The accuracy of spatial databases. Bristol, PA: Taylor and Francis. pp. 35-43.

29. Stoms, D. 1987. Reasoning with uncertainty in intelligent geographic information systems. In: Proceedings of GIS '87, San Francisco, CA (October 26-30). Falls Church, VA: ASPRS and ACSM. pp. 693-700.

30. Tosta. 1993. The data wars: Part II. Geo. Info. Sys. 3(2):22-26.

31. Anonymous. 1994. Between a rock and a hard place. Backpacker 22(8):16-18.

32. Fletcher, M., and D. Sanchez. 1994. Etched in stone: Recovering Native American rock art. GPS World 5(10):20-29.

33. Csillag, F. 1991. Resolution revisited. In: 1991 ACSM-ASPRS Annual Convention, Baltimore, MD (March 25-28). Auto-Carto 10, Volume 6. Bethesda, MD: ACSM and ASPRS. pp. 15-28.

34. Star, J.L., and J.E. Estes. 1990. Geographic information systems: An introduction. Englewood Cliffs, NJ: Prentice Hall.

35. Goodchild, M. 1993. The state of GIS for environmental problem-solving. In: Goodchild, M.R., B.O. Parks, and L.T. Steyaert, eds. Environmental modeling with GIS. New York, NY: Oxford University Press. pp. 8-15.

36. Veregin, H. 1987. Error modeling in geographic information systems: A review. Unpublished manuscript.

37. Keefer, B.J., J.L. Smith, and T.G. Gregoire. 1991. Modeling and evaluating the effects of stream mode digitizing errors on map variables. Photogrammetric Eng. and Remote Sensing 57(7):957-963.

38. Giovachino, D. 1993. How to determine the accuracy of your graphic digitizer. Geo. Info. Sys. 3(3):50-53.

39. Campbell, W.G., and D.C. Mortenson. 1989. Ensuring the quality of geographic information system data: A practical application of quality control. Photogrammetric Eng. and Remote Sensing 55(11):1,613-1,618.

40. Goodchild, M., and S. Gopal. 1989. Preface. In: Goodchild, M., and S. Gopal, eds. The accuracy of spatial databases. Bristol, PA: Taylor and Francis.

41. Goodchild, M. 1993. Data models and data quality: Problems and prospects. In: Goodchild, M.F., B.O. Parks, and L.T. Steyaert, eds. Environmental modeling with GIS. New York, NY: Oxford University Press. pp. 94-103.

42. Lester, M., and N. Chrisman. 1991. Not all slivers are skinny: A comparison of two methods for detecting positional error in categorical maps. In: Proceedings of GIS/LIS '91, Atlanta, GA (October 28-November 1). Bethesda, MD: ASPRS. pp. 648-658.

43. Newcomer, J.A., and J. Szajgin. 1984. Accumulation of thematic map errors in digital overlay analysis. Amer. Cartographer 11(1):58-62.

44. Guptill, S.C. 1989. Inclusion of accuracy data in a feature-based, object-oriented data model. In: Goodchild, M., and S. Gopal, eds. The accuracy of spatial databases. Bristol, PA: Taylor and Francis. pp. 91-98.

45. Monmonier, M. 1991. How to lie with maps. Chicago, IL: University of Chicago Press.

46. Wehde, M. 1982. Grid cell size in relation to errors in maps and inventories produced by computerized map processing. Photogrammetric Eng. and Remote Sensing 48(8):1,289-1,298.

47. Stoms, D.M. 1992. Effects of habitat map generalization in biodiversity assessment. Photogrammetric Eng. and Remote Sensing 58(11):1,587-1,591.

48. Connin, R. 1994. The trouble with data. Professional Surveyor 14(1):24-25.

49. Smith, J.L., S.P. Prisley, and R.C. Weih. 1991. Considering the effect of spatial data variability on the outcomes of forest management decisions. In: Proceedings of GIS/LIS '91, Atlanta, GA (October 28-November 1). Bethesda, MD: ASPRS. pp. 286-292.

50. Felleman, J., and C.B. Griffin. 1990. The role of error in GIS-based viewshed determination: A problem analysis. Report No. IEPP-90-2. Syracuse, NY: Institute of Environmental Policy and Planning, State University of New York, College of Environmental Science and Forestry.

51. Dutton, G. 1991. Improving spatial analysis in GIS environments. In: 1991 ACSM-ASPRS Annual Convention, Baltimore, MD (March 25-28). Auto-Carto 10, Volume 6. Bethesda, MD: ACSM and ASPRS. pp. 168-185.

52. Felleman, J. 1992. Letter to the editor. Geo. Info. Sys. 2(8):12-13.

53. Berry, J.K. 1994. The this, that, there rule. GIS World 7(7):22.

54. Hunter, G.J., and M.F. Goodchild. 1993. Managing uncertainty in spatial databases: Putting theory into practice. In: Proceedings of the annual meeting of the Urban and Regional Information Systems Association, Atlanta, GA (July 25-29). Washington, DC: Urban and Regional Information Systems Association. pp. 1-15.

You Can't Do That With These Data! Or: Uses and Abuses of Tap Water Monitoring Analyses

Michael R. Schock
Drinking Water Research Division, U.S. Environmental Protection Agency, Cincinnati, Ohio

Jonathan A. Clement
Black & Veatch, Cambridge, Massachusetts

Introduction

Linkage between human health and drinking water quality has been an area of interest in the United States for many years. Over the past approximately 10 years, drinking water monitoring requirements have expanded rapidly under the Safe Drinking Water Act (SDWA). Growing public and governmental interest in this environmental area makes the aggregation and consolidation of data on the occurrence and distribution of many organic and inorganic contaminants and background constituents of drinking water an important process. These data can then be made available for systemization and visualization to regulators, municipalities, water utilities, public interest groups, health researchers, consulting engineers, and water treatment scientists.

Given a sufficient number of data points and a convenient computerized database/mapping platform, a wide variety of maps can be generated to use in research and decision-making processes. The validity of doing so, however, rests inseparably upon the basis of the sampling plan and protocols, as well as the precision and accuracy of the analytical methods used for the constituents of interest. The well-known problem of matching the proper scale of the source data to that employed in the maps for interpretation is a critical problem with drinking water sampling, where many unappreciated small-scale variations render many, if not most, attempts to make generalizations inaccurate or meaningless.

This paper introduces and describes many concepts related to what generates or controls the concentrations of metals and other constituents in drinking water, ways in which the sampling protocol affects apparent levels of constituents, and the magnitude of temporal and spatial variability present in both municipal and private water supplies. Illustrations from water quality studies show in practical terms how generalizations must be kept to a minimum and how the data input into a geographic information system (GIS) for interpretation and evaluation must be carefully analyzed and screened to determine the appropriateness for various well-intended purposes. The discussion and examples show how many apparently significant trends and assessments of exposures or occurrences turn out to be merely artifacts of critical (yet subtle) inconsistencies or errors in the planning and execution of the sample collection process, or inconsistencies caused by the fact that regulatory (and not research) requirements govern the origin of the data.

The concepts this paper covers are equally valid in many other disciplines using or contemplating the use of GIS for interpretation of all kinds of “field” data.

Why Maps Are Useful for Drinking Water Studies

Maps and GIS databases could have wide applicability to drinking water studies. For example, they could provide the basis for investigating the occurrence of regulatory contaminants or related constituents, either to estimate the costs of compliance with a regulation or to estimate human health effects. Mapping could be useful to utilities and consultants investigating process changes for a utility or determining the effectiveness of some existing treatment such as corrosion control or chlorination. Use of GIS could also assist in assessing the feasibility and impact of system expansion. Another promising application would be GIS assistance in developing and implementing wellhead protection plans. Many other areas of application may be possible now, or will be discovered in the future, as GIS technology and regulatory requirements continue to develop.

Sampling Protocols for Data Usable in GIS

Several SDWA regulations have resulted or will result in the collection of geographically diverse drinking water quality data that may interest mappers. The Lead and Copper Rule, the Surface Water Treatment Rule, the proposed Information Collection Rule, and the Disinfection/Disinfection Byproduct Rule are but four examples. Many states have their own variations on federal drinking water regulations, so their data collection requirements may differ somewhat. Considerable data may also be collected for specific research studies of either academic or purely practical nature.

Chemical Factors in Constituent Behavior

For the purposes of this discussion, chemical constituents in drinking water may be classified as being generally reactive or nonreactive. Reactive constituents may change concentrations or chemical form for a variety of reasons, such as:

- A result of interaction with the background composition of the drinking water.
- By precipitation or dissolution reactions with pipe material used for the distribution system.
- By chemical reactions with disinfectants added at water treatment plants.
- By slow chemical reactions started at water treatment plants.

Nonreactive constituents may play an important role by providing a chemical background that indirectly influences the speed or extent of other chemical reactions and transformations. Table 1 gives a summary of many common constituents of drinking water and identifies whether they function essentially as reactive or nonreactive constituents.

Table 1. General Reactivity Trends for Common Drinking Water Constituents

Constituent	General Reactivity Tendency
pH	Highly reactive
Dissolved oxygen	Reactive
Calcium	Nonreactive (reactive when cementitious pipe linings are present)
Magnesium	Nonreactive
Total carbonate	Nonreactive
Total alkalinity	Reactive, particularly with pH changes
Chlorine residual	Reactive
Temperature	Either
Iron	Reactive
Copper	Reactive
Lead	Reactive
Zinc	Reactive
Silica	Nonreactive
Sulfate	Nonreactive
Orthophosphate	Reactive
Polyphosphate	Reactive
Total phosphate	Reactive
Nitrate	Nonreactive
Chloride	Nonreactive
Fluoride	Nonreactive
Trihalomethanes	Reactive
Haloacetic acids	Reactive

Reactive Constituents

Clearly, chemical species or compounds that can change in concentration or transform into other species or compounds during distribution make mapping on very large scales difficult to justify. Reactive constituents may also change concentration in the same place over time, such as water standing overnight in a home, school, or building, which is discussed in a later section. Some examples of reactions during water distribution follow:

- During lime softening processes at some central water treatment plants, a supersaturated state is used for the compound calcium carbonate to remove calcium (and sometimes magnesium) ions from the water. This condition is sometimes maintained into the distribution system as well to assist in maintaining chemical conditions useful for corrosion control of lead and copper. Thus, calcium levels, pH, and carbonate concentrations (and consequently, alkalinity) drop as water passes away from the plant (1, 2).
- The metals in pipe materials, such as iron, copper, zinc (in galvanized pipe), and lead, are oxidized by oxygen, free chlorine, chloramines, ozone, and other disinfectants, which renders them into a form that water can transport, unless other chemical conditions are such that a highly insoluble scale deposits on the pipe, immobilizing the metal (1, 3).
- Prolonged contact with chlorine disinfectant species converts a fraction of natural organic matter present in many distributed waters into regulated “disinfection byproduct” compounds, such as trihalomethanes, chloroform, and haloacetic acids (4, 5).
- Following the addition of chlorine or after increasing pH to enable some corrosion control for copper and lead, iron present in well waters in dissolved ferrous (Fe^{2+}) form oxidizes into Fe^{3+} form, which is much less soluble. Obnoxious “red water” results, as ferric oxyhydroxide precipitate forms and clouds the water.
- Polyphosphate chemicals added to “sequester” iron or manganese in well waters break down into simpler

polyphosphate forms of shorter chain lengths, plus orthophosphate. The orthophosphate frequently becomes present at high enough concentration to aid in controlling lead or copper (1, 6-8).

- Water passes through newly installed cement mortar-lined pipes, or aggressive water passes through older asbestos-cement pipes. Because of the particular chemical nature of the water, calcium carbonate and calcium hydroxide in the cement dissolve, raising the pH and hardness of the water (1).
- Free chlorine is added to disinfect water and is such a strong oxidant that it is unstable in water at normal concentrations. Additionally, it reacts with miles of unlined cast iron pipe, accelerating the decomposition of hypochlorous acid or hypochlorite ion to chloride. Consequently, the overall redox potential of the water supply and the effectiveness of disinfection decrease.
- A concentration of 1 milligram per liter (as PO_4) phosphoric acid is added to a distributed water at pH 7.5 to control lead corrosion. The orthophosphate reacts with exposed iron in the distribution main, however, and the residual concentration of orthophosphate decreases throughout distribution passage to the point where the level is no longer adequate to create the lead orthophosphate passivating film needed (1, 6, 8, 9).

Unless a constituent is known to be nonreactive, maps may be falsely generated under the premise that the concentration of a constituent is essentially a constant over some geographic area. Following the changes in concentration or chemical form of reactive constituents would also seem to be a useful application of GIS technology. One major restriction applies to the viability of that approach, however. Presuming that the analytical techniques used can adequately quantify the concentration and concentration changes observed, the scale of the variability or concentration change relative to the scale of the mapping perspective becomes critical to accurate mapping. A later section of this paper considers this critical factor in more specific detail.

Nonreactive Constituents

Almost no inorganic constituents in natural or drinking water are purely chemically inert. Under some conditions, and at some concentrations, significant reactions can occur. Some constituents that are actually reactive may act as if they are nonreactive constituents, however, because they are present in high enough concentrations relative to the extent of chemical reactions taking place that no discernible change in their concentration results. An obvious example is the dissolved inorganic carbonate (DIC = H_2CO_3* + HCO_3- + $CO_3$2-) concentration (1, 10). Complexation and formation of passivating basic carbonate solid films of lead and copper by carbonate and bicarbonate ion dominate the corrosion control chemistry of copper(II) and lead(II) (1, 11). The concentration of DIC in water on either a molar or weight basis, however, is normally a factor of 500 to 10,000 higher than the lead or copper concentrations. Hence, changes in the DIC content from these reactions normally are analytically undetectable.

Another example is fluoride ion, which is often used as a distribution system water flow “tracer” because of its relative inertness. Actually, fluoride ion can form strong complexes with aluminum left in water following coagulation treatment with alum. The solubility of fluoride-containing solids with other major drinking water components (such as calcium and sodium) is very high, however, and fluoride reacts only weakly with metallic plumbing materials in the distribution system. Therefore, total fluoride concentrations tend to remain constant.

Relatively accurate maps of the occurrence and distribution of nonreactive constituents can be made, but their usefulness depends on the scale of the mapping relative to their occurrence and the particular question under investigation. All of this supports the need to ensure that the question asked can be answered correctly at the map scale.

Scale of Drinking Water Constituent Sources

More than 59,000 public water suppliers exist in the United States (12). Of these, approximately 660 are considered large water systems, which serve over 50,000 in population. These municipal systems use source water supplies that can be ground-water wells, “surface” waters (i.e., rivers, reservoirs, lakes), or a combination of both. Some water suppliers perform minimal water treatment of their own and purchase water from another water system or systems to satisfy their needs.

Surface Water Sources

Many water utilities use a single water treatment plant to treat surface waters, which could satisfy the entire water demand of the community all year. In many cases, however, utilities combine several surface water sources and use a different treatment plant to treat each water source. The water plants usually discharge into the distribution system at different points, and system hydraulics dictate the areas of the system in which waters mix. This is important because the water quality characteristics, which often differ among treatment plants, influence the corrosivity of the waters to various plumbing materials in the distribution system. Different water constituents also may affect the disinfection effectiveness of the treatment and the formation of unwanted disinfection byproducts.

For surface water systems, the chemical composition of the water depends on the upstream or watershed geochemistry, the seasonal nature of the water body used as the source, and the characteristics that the treatment imparts, such as coagulation with ferric sulfate or alum (aluminum sulfate), lime softening, filtration, pH adjustment, corrosion control treatment, chlorination, etc.

The scale of the source water chemical data, therefore, is large, driven by the geology, soil nature, land use, and climate. The chemical nature of the treated water, however, may differ significantly from that of its source.

Ground-Water Supplies

Many water utilities use multiple ground-water wells. A water supply of medium to large size usually uses multiple wells, instead of or in addition to the surface water supplies. Wells number from only two or three to more than 100 for very large water systems. Wells normally operate in different patterns, and only rarely do all wells operate at the same time. The yield of the wells and their water quality dictates the combination and number of wells used at a particular moment. The wells may or may not be from the same aquifer, and even if they are, local inhomogeneities frequently exist in water composition (especially with iron and manganese) that limit the usefulness of certain wells without substantial treatment.

Historically, utilities have treated some (but not necessarily all) wells with a chemical such as a polyphosphate or sodium silicate to sequester the iron and manganese from wells. Some utilities install physical removal processes such as ion-exchange softeners, reverse-osmosis plants, aeration systems for iron removal, air stripping towers for volatile organic compound or radon removal, or "greensand" filters for the removal of iron and manganese. These facilities sometimes exist at only certain well sites or at some point where water from multiple wells is combined.

The scale of chemical controls on ground-water supplies, therefore, becomes only hundreds of feet. Contaminants of raw waters, such as arsenic, nitrate, or chromium, are geologically and geochemically controlled. Therefore, their occurrence is geographically variable on even a small scale, and the variability exists vertically in the subsurface as well as horizontally. A municipality may use wells of different depths into different aquifers, or even approximately the same depth spread out over hundreds of feet to many miles in the same aquifer or a variety of geologic units.

The variability of individual ground-water wells over time (such as seasonally) is usually less apparent than with surface water sources, but the fact that many wells are frequently used in different combinations and for different lengths of time (hours to days, usually) makes characterizing "influent" water quality complicated. The same observation applies to water systems that allow different amounts of water to bypass treatment processes (e.g., ion-exchange, reverse osmosis) depending on the levels of targeted undesirable contaminants (e.g., nitrate, sulfate, arsenic).

These characteristics of the nature of chemical composition, use, and treatment of ground-water supplies clearly show that generalizations over areas such as states or geographic regions (e.g., New England, Upper Midwest) are at least very gross and uncertain and at worst, entirely misleading when decisions are to be made about risk and health assessments, or estimates of the necessity for certain treatments or economic impacts of different potential drinking water regulations.

Combination Systems

Some municipalities combine the use of surface water supplies and ground-water wells. Therefore, general water chemical characteristics vary throughout the system in a regular manner in response to the location and use of different sources, as well as relative amounts of water that the different sources produce and deliver.

Distribution System Mains

The next lower level of scale is the distribution system network of pipes and storage. Common materials used for distribution system piping include cast iron, ductile iron, cement mortar-lined iron, iron with organic coatings, asbestos-cement (A-C), and various forms of plastic. Pipe diameters range from about 4 inches to many feet, depending on size of the water utility and community, size of the neighborhood fed by the line, and distance of travel for the water. Here, because of the large volume of water involved relative to the pipe diameter, the major chemical interactions involve such constituents as hardness (calcium and magnesium) ions, pH, iron, bicarbonate and carbonate ions, and chlorine residual species, and possibly microbiological parameters such as total plate counts, heterotrophic plate counts, and assimilable organic carbon. Disinfection byproducts (DBPs) may change in concentration and type because of the time involved in the water traveling through the piping from the treatment plant. Trace metal contamination, such as lead and copper, is usually negligible from this source, unless it is present when distributed from the wells or water treatment plants.

Depending on prevalent economics and construction practices during periods of water system growth, the materials will not be either randomly or uniformly distributed geographically within system boundaries. Water flow often varies greatly within the distribution system, and water lines sometimes terminate in dead-end areas with minimal flow rates. Water quality often differs substantially in these dead ends from that in the fully flowing parts of the distribution system.

Household Service Lines

Service lines represent the connection between the house or building and the distribution main. Sometimes, the service lines are joined to the mains by a flexible, approximately 2- to 3-foot long pipe called a "gooseneck" or "pig-tail." Historically, this connector was often made of lead. Recently, copper has been the most widely used material, with plastic gaining in acceptance. Service lines for homes are usually 0.75 to 1 inch in diameter, with service lines for many commercial buildings or multifamily dwellings ranging in size from 1.5 to 3 inches in diameter. Service line for homes and buildings have usually been made of lead, brass, copper, galvanized steel, or plastic. The material used depends on the age of the water connection and the construction practices of the area involved. A recent report estimated that approximately 6.4 million lead connections ("goosenecks") still exist in the United States, and about 3.3 million lead service lines still exist (13). In many communities, old lead service lines remain a major source of lead in drinking water.

Like distribution system materials, service line materials may vary greatly within a distribution system by space and time. For instance, in large eastern cities, very old neighborhoods may have many (or even mostly) lead service lines. New neighborhoods likely have copper or plastic service lines. Galvanized steel or copper pipes may have been installed between the era when lead was used and modern times. With the exception of Chicago, where lead service lines were occasionally installed into the 1980s, the use of lead for service lines generally stopped in the late 1940s or early 1950s. An example of nonuniform distribution of service line materials is shown by Figure 1, a map indicating Cincinnati sampling sites for Lead and Copper Rule (14-17) monitoring. Erratic clustering of different service line materials is evident.

Rehabilitation of old houses or replacement of failed piping results in a mixture of new and old material in areas where houses are predominantly old. Following completion of the construction, maps of service line material would show many clusters representing prevalent plumbing codes and economics.

Interior Plumbing

Interior plumbing of buildings and houses reflects even more variability than service lines. This is the dominant contributor to lead and copper levels at most sites covered under the Lead and Copper Rule (14-17). Interior plumbing consists of piping, plus a large number of valves, connectors, fixtures, and perhaps soldered joints and a water meter. Any or all of these components are replaced at varying intervals as a result of failures or remodeling. Therefore, even generalizations within a small neighborhood are risky, unless the neighborhood is very new and uniformly constructed. When attempting to survey the composition of plumbing materials that might be the source of drinking water contamination, merely asking for the age of the house or building is insufficient. Questions must be asked to obtain the necessary precise information on the age and type of plumbing materials and components in the building.

Typical interior plumbing materials include lead, galvanized steel, copper, and different plastics for pipes. Some brass and black steel have been used for short times in some areas. Faucets are almost always made with either brass or plastic internal parts, which differ in composition from the exteriors, which are usually plated with chrome or other metal. Interior faucet volumes typically range from about 30 milliliters to 120 milliliters, depending upon design. Valves and meters are also frequently made of brass or bronze, which are copper/zinc alloys usually containing 2 percent to 6 percent lead. Until recently, solders used to join copper drinking water pipe sections were usually a tin and lead combination, containing 40 percent to 60 percent lead. Occasionally, connector lines to fixtures include copper, stainless steel, aluminum, or flexible plastic sections.

Private Water Systems

The many possible designs of domestic water systems originating from wells or cisterns are too numerous to illustrate. Figure 2 gives an example of one such system layout. Private systems share many features with domestic systems supplied by water utilities, however. Interior plumbing shares most of the same configurations and materials. For private water systems, additional plumbing that could cause contamination or water chemistry changes includes well casing material, submersible pump casing and fittings, pressure tank feed and control plumbing, and nonsubmersible pump interior materials. Therefore, problems with determining the frequency and distribution of levels of potential contaminants include those present for domestic situations in general, plus those complications arising from cycling of the pumps, pressure tank system, or both.

Water Samples Representing Distances

One of the most important fundamentals of understanding drinking water sampling is that volumes of water (e.g., 1-liter samples, 250-milliliter samples) represent the linear distance of plumbing material in contact with the water sampled. Because of water mixing and flow during use or sampling, they are also integrated samples of that volume. This understanding is at the heart of designing accurate water sampling programs and making viable interpretations of existing monitoring data that may be contained in (or mappable by) a GIS.

Table 2 summarizes some interesting and important relationships between pipes of different inside diameters

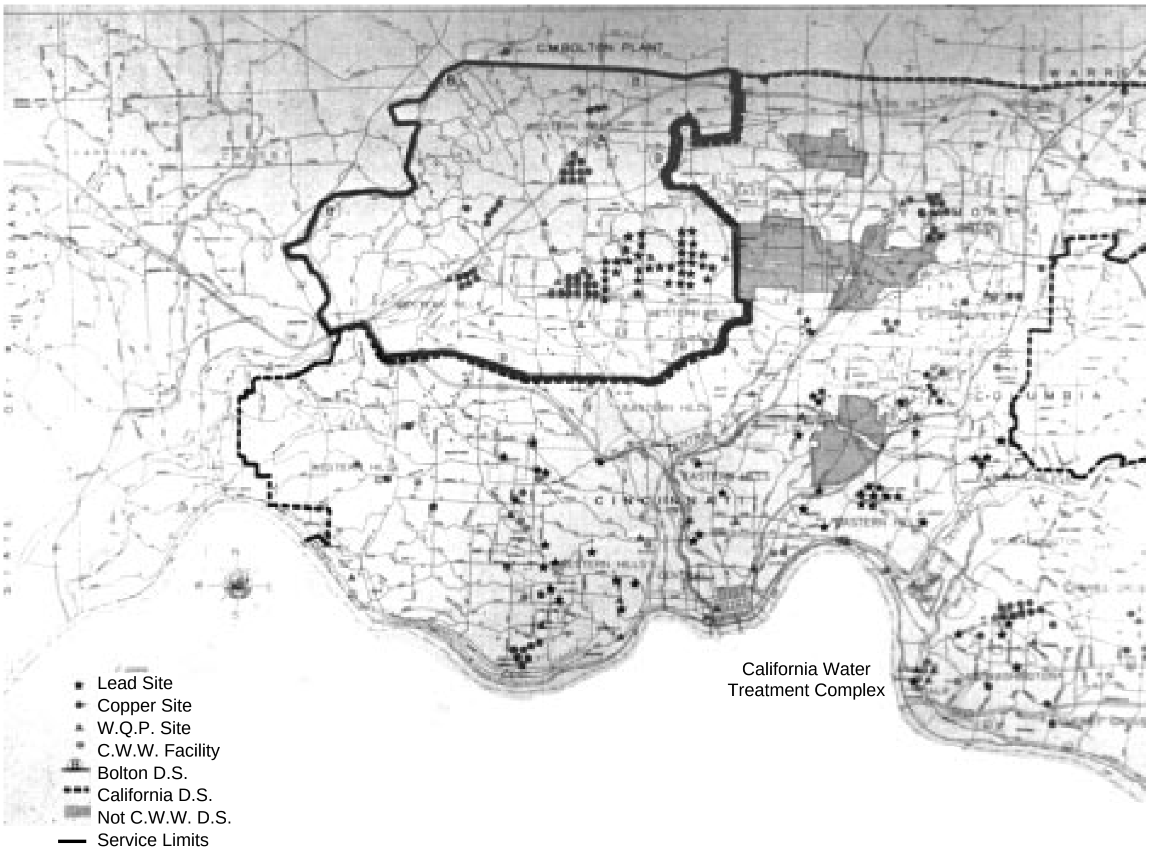


Figure 1. Cincinnati Water Works Lead and Copper Rule compliance monitoring, July to December 1992.

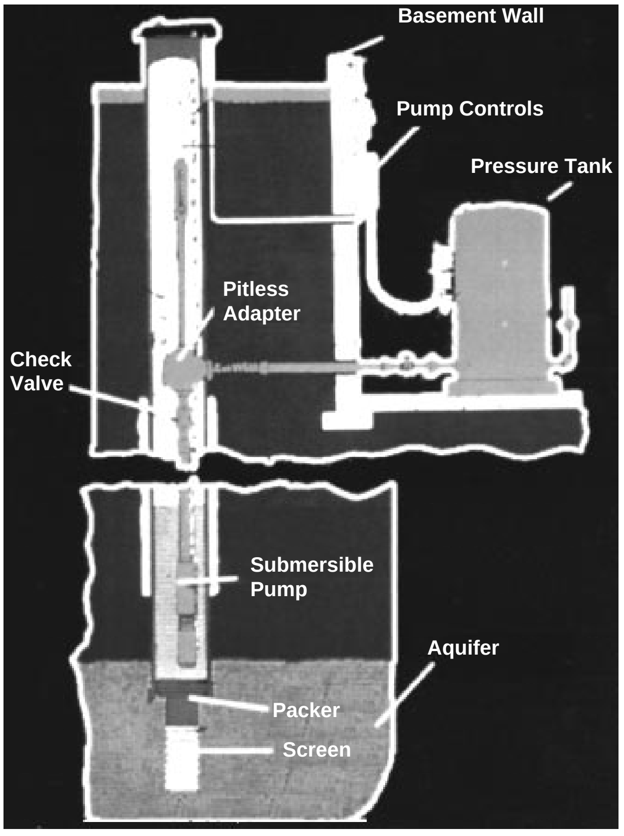


Figure 2. Distribution system.

(IDs) and the volumes of water they contain per unit of length (7). Much domestic interior plumbing has an ID of approximately 0.5 inches, depending upon the material.

Figure 3 shows schematically what parts of a plumbing system would likely be represented by samples of different volumes taken after water was allowed to stand in the pipe for many hours. Faucets, bubblers, and other terminating fixtures vary widely in volume. Kitchen-type fixtures usually contain from 60 to 120 milliliters of water. Bathroom-type fixtures may contain only about 30 to 60 milliliters of water. Bubblers, such as those frequently found on school or office drinking fountains, are smaller still. As can be seen schematically in Figure 3a, a small volume such as 125 milliliters captures the faucet and a short distance of pipe immediately leading to it. In many plumbing systems, this volume catches water in contact with numerous soldered joints. On the other hand, if a single 1-liter first-draw sample is taken, the water in the bottle represents a much longer distance back into the plumbing system. In a situation where the source of lead in drinking water is a new brass faucet, or soldered joints of lead-tin solder, this larger volume usually gives a lower lead concentration than the smaller volume because more water in the sample is not in intimate contact with materials containing lead.

Other sampling schemes logically follow. For instance, if examining copper pipe corrosion, discarding the first

Table 2. Interrelationships Among Pipe Length, ID, and Internal Volume for Selected Common Plumbing Materials and Pipe Sizes

Material	Identification/ Type	True ID	True OD	Length for 1,000 Milliliters (Feet)	Milliliters per Foot
Copper tubing	0.5-inch, type L, annealed	0.545	0.625	22	46
Copper tubing	0.5-inch, type L, drawn	0.545	0.625	22	46
Copper pipe	0.5-inch, schedule 40	0.622	0.840	17	60
Galvanized steel pipe	0.5-inch, schedule 40	0.616	0.840	17	59
Lead pipe or tube	0.5-inch ID, 0.25-inch wall	0.50	1.00	26	39
Lead pipe or tube	0.75-inch ID, 0.25-inch wall	0.75	1.25	11.5	87
PVC or CPVC pipe	0.5-inch, schedule 80	0.546	0.840	22	46

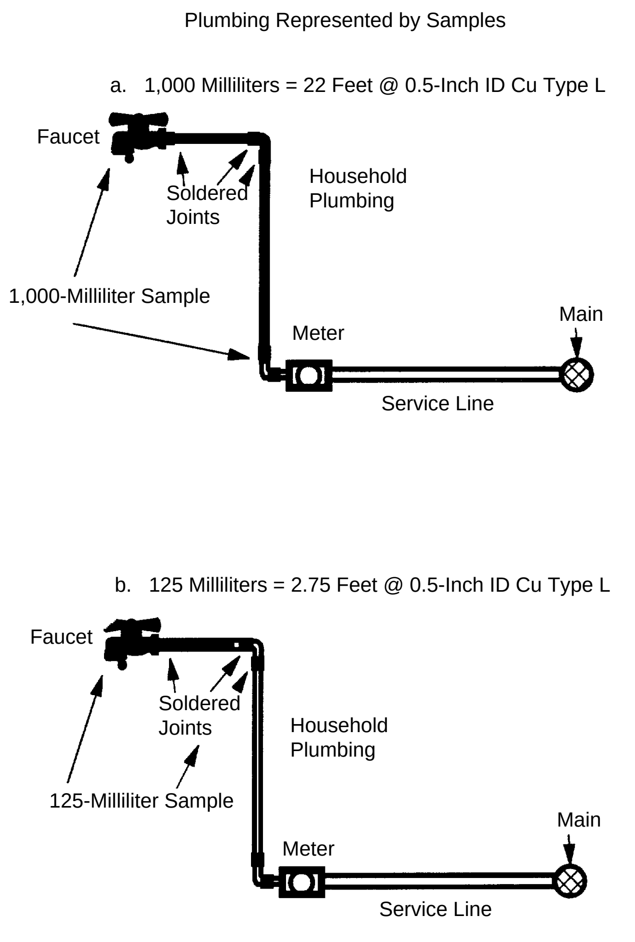


Figure 3. Schematic diagram of plumbing materials represented by sample volumes of a) 1 liter and b) 125 milliliters.

125 or 250 milliliters of water is likely to give more accurate information because it minimizes the effects of the faucet material as well as piping that connects the faucet to the interior line. Often, this connecting piping is not copper. If examining the corrosivity of the water to lead service lines, wasting a volume of water corresponding to the distance from the outlet to the service line better estimates the effect, although not without uncertainty (18). Many other sampling schemes are possible and useful, but users must be aware that the sampling protocol may have as much or more influence on the observed metal concentration than water quality or other variables. Hence, incorporation of monitoring data into a GIS database must be done only when the source represents equivalent samples.

Because of turbulent mixing during flow, local high concentrations of lead (or other contaminant) may become broadened and diluted by the time the water to be sampled reaches the sample collection bottle (18). In many cases, therefore, numerous small-volume sequential samples can be taken and used to profile a plumbing system to locate brass valves, connectors, soldered joints, etc. Figure 4 illustrates sequential sampling results for one room of a building. Peaks in the distribution of samples physically correspond to the location of a chrome-plated brass faucet and to a later concentration of fresh Sn:Pb soldered joints. Unfortunately, even small-volume sequential tap water samples must pass over other potentially contaminating or altering surfaces on the way through the sampling tap.

Kuch and Wagner have shown how water can dissolve large amounts of lead simply by traveling through long distances in lead pipes with small IDs (9, 19). Although this study specifically examined lead, the principle applies to other metallic piping materials. This phenomenon is inseparable from the aspect of time, which is the next subject.

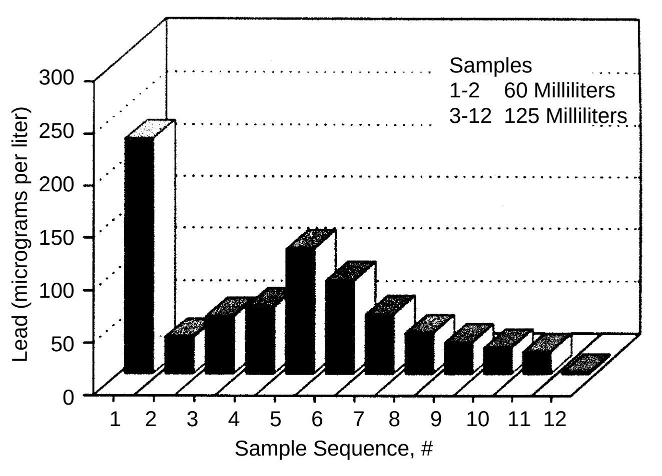


Figure 4. Sequential sampling results from a room on the ground floor of a building.

Effect of Time

While some chemical reactions are instantaneous, many dissolution and precipitation reaction steps that are important in controlling metal levels in water take many hours to many days to reach equilibrium. In fact, passivation films and scales on pipes that inhibit corrosion and reduce leaching of trace metals may take months to decades to develop substantially. Some other chemical transformations, such as creation of trihalomethanes from chlorination of natural organic matter during disinfection, or processes such as inactivation of pathogens, may occur over hours (20, 21).

Many steps in an overall chemical reaction process could be rate-limiting. Figure 5 shows how lead levels increase in 0.5-inch ID pipe given two different assumptions. The top curve shows how lead increases and levels off after about 8 to 12 hours (9, 19). This curve is closely applicable to any metal, as long as the limiting factor on the rate of metal migration into the water is the radial diffusion of the soluble metal species away from the pipe surface. The second curve shows, schematically, the effect of a diffusion barrier film (e.g., calcium carbonate, adsorbed iron hydroxide mixed with organic matter, aluminosilicate mineral deposits) or inhibition of metal oxidation rates on lead migration into water after different amounts of time.

In some water systems, significant chemical changes can occur while the water is standing that can drastically affect the oxidation or solubility of the plumbing material. For example, dissolved oxygen and free chlorine react quickly in new copper pipe or brass. If the water stands

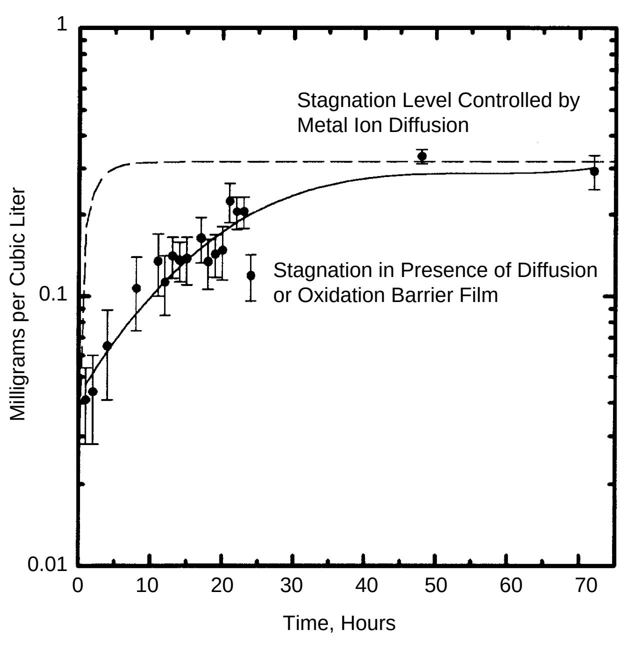


Figure 5. Comparison of lead concentrations that would be observed after water stands different amounts of time given different controlling chemistry factors.

sufficiently long, their concentrations may become negligible, which would significantly alter the redox conditions governing metal solubility. In the absence of oxygen or chlorine species, the dominant form of copper in water and on plumbing material then becomes copper(I) instead of copper(II), resulting in different solubility characteristics after consumption of the oxidant than at initiation of the standing period (11).

Seemingly identical water samples collected from the same taps in houses, schools, or other buildings yield different metal concentrations, depending on the time the water was in contact with the faucet, solder, or piping material. Similarly, samples taken for disinfection byproducts after different chlorine contact times may produce different concentrations and different speciation (e.g., trihalomethanes, haloacetic acids). This factor causes considerable confusion in many investigations of contamination of school or building drinking water taps and water coolers and complicates estimating human exposure for health-effects studies.

Interconnectedness of Distance and Time

In innumerable situations, the effects of distance and time are impossible to separate. Some generalizations and examples follow.

Dead ends and slow rate areas produce long residence times for the distributed water. This results in long contact times with pipe materials, so reactive constituents can change considerably in concentration. The process is totally interactive, in that concentration changes of reactive constituents are in response to contact with the pipe materials, and in turn, the materials respond to the water composition. Water may take hours to days to reach a particular home or building and may traverse many miles of distribution system piping of the same or differing composition. Water thoroughly run through a household faucet for 5 or 10 minutes to purge the lines is “fresh” from the resident’s perspective, but may be “old” from the distribution system perspective.

The profile of the water line shown in Figure 4 was made on the basis of filling small volume (60 or 125 milliliters) sample bottles, one after another, without wasting any water. If the objective were to capture only the highest risk of lead contamination from a lead service line after some hours of stagnation, then the sampling process would be different. Instead of collecting all water between the tap and the service line, the water can be run until either a target volume is wasted (representing the linear plumbing distance to the service line) or can be run at a given rate until, after the appropriate length of time passes, the sample bottle can “intercept” the slug of water residing in the service line. Beware that differences may exist in the peak concentration of the contaminant and the “width” of the slug of elevated contaminant level, depending upon the rate of water flow before and during sampling (18, 22, 23).

Other Sources of Variability in Water Samples

Variability in water samples can stem from many sources aside from those discussed in this paper (18). The nature of the errors and their likely magnitude may vary with each episode of sampling and analysis and is far beyond the present scope of discussion. A brief listing to consider, however, when drawing conclusions from "field" data includes:

- Analytical imprecision or bias.
- Flow rate of the water during sampling.
- Temperature.
- Particulate erosion from plumbing materials during sampling.
- Effect of container material.
- Effect of air contact or other handling effects during sample collection and shipment.

A Case Study of Easy Misinterpretation

Interpretations of water quality problems based on aggregate monitoring data can be very misleading unless analysis is performed at the appropriate scale. The situation of one utility described below provides a good example of how using a GIS approach could have helped solve the problem but also highlights how carefully data would need to be matched and consolidated only at the proper scale if GIS were to be employed for evaluating some kinds of water quality problems.

The utility at Hopkinton, Massachusetts, found very high lead and copper levels exceeding the regulatory action levels under the Lead and Copper Rule (14-17). The 90th percentile copper levels even exceeded 6 milligrams per liter, compared with an action level requirement of only 1.3 milligrams per liter. Some sites with lead service lines are present in the system. Figure 6 shows a schematic representation of the distribution system for this utility. Five wells feed the system, and four (1, 2, 4, and 5) are used regularly.

The background hardness, alkalinity, and carbonate concentrations are fairly similar for all wells. The pH of the ground water from the wells is usually slightly above 6. Chlorine solution is dosed for disinfection. High iron levels are present in wells 1, 2, and 3, and high manganese is also present in well 3. Generally, high dosages of a polyphosphate chemical were added to wells 1 and 2 to respond to consumer complaints about the "red water" that results from iron oxidation and precipitation. Water from different wells mixes in the distribution system, but the water tends to partition into two zones as

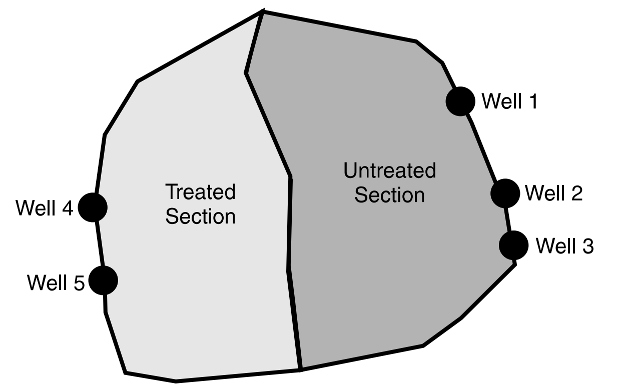


Figure 6. Schematic representation of the utility's distribution system.

Figure 6 indicates. The lead and copper levels tended to be distinctly lower in the section where the polyphosphate was dosed, marked as the "treated" part of the system. The ongoing research study has employed approximately 22 monitoring sites.

From the information presented thus far, the system clearly cannot be characterized by a discrete value for lead or copper contamination, as well as the chemical background of water throughout the system. Hence, putting data at the "whole system" scale into a statewide or countrywide data system would be tempting, but it could be very misleading in solving the treatment problem. Having accurate spatially distributed data for background water qualities, monitoring site characteristics, and metal levels at the subsystem scale, such as that which could be integrated into GIS, would have been extremely convenient, however. Yet, even more information at a smaller scale is necessary to understand and solve the whole treatment problem.

The utility initially observed that because the lowest lead and copper levels coincided with the area of the system fed by the polyphosphate chemical, that chemical likely caused the corrosion inhibition. Median lead levels, for example, were between about 200 and 300 milligrams per liter in the "untreated" section, compared with about 10 to 15 milligrams per liter in the "treated" section. Median copper levels were approximately 4 to 5 milligrams per liter in the untreated section, but only about 0.3 to 0.5 milligrams per liter in the treated section. The utility and the researchers wondered whether the polyphosphate chemical should also be added to the other wells. This is a matter of significant concern because some studies indicate polyphosphate chemicals can enhance lead corrosion (1, 6) and the subject has rarely been studied under statistically valid controlled conditions.

Additional site-by-site investigation, however, first revealed that the sites with lead service lines all lay in the untreated area of the distribution system. Because the research sampling program included two successive 1-liter samples, the additional contamination from the service lines was confirmed by higher lead levels in the second 1-liter sample than in the first in many cases. Therefore, physical reasons, in addition to chemical ones, explained the discrepancy in the lead levels. Further, when considering only the treated system sites, the lead levels were still high enough to be of concern.

Focusing on the copper sites resulted in the collection of more important and interesting small-scale information. Figure 7 shows the difference between average copper levels in the two sections of the system. Almost all sites in both parts of the system had copper interior plumbing with 50:50 or 60:40 Sn:Pb soldered joints and faucets with brass-containing internal materials. Though the chemical added for iron control was ostensibly a polyphosphate chemical, it also contained an initially present fraction of orthophosphate and also tended to partially break down to orthophosphate in the presence of iron and calcium (as most polyphosphates do). Figure 8 shows the orthophosphate concentrations in the two different parts of the system. While the levels of orthophosphate present in the treated section would be far too low to significantly inhibit lead leaching at the background pH (1, 6, 7, 9, 22), the orthophosphate plausibly may significantly inhibit copper dissolution, in concordance with recent research projections (11).

Having determined through detailed small-scale sampling and analysis that the chemistry affecting metal levels in the system is generally consistent with modern knowledge, a new treatment plan is being implemented to control copper and lead levels through pH adjustment in conjunction with iron control through a compatible sodium silicate/oxidation treatment. Incorporation of system and monitoring site physical characteristic data, plus monitoring results, into GIS could have saved considerable investigatory effort. The importance of this case history, however, is that the data must be of the appropriate scale and highly documented to be useful in problem-solving. Failure to use data meeting these re-

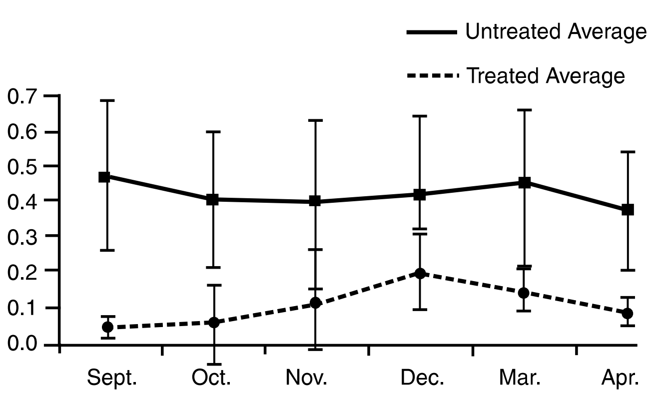


Figure 7. Average copper levels in treated and untreated sections of the system.

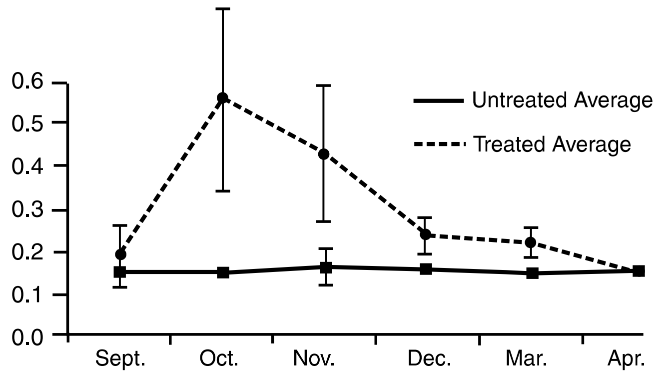


Figure 8. Average orthophosphate levels in treated and untreated sections of the system.

quirements, as well as overgeneralization to a large mapping scale, can lead to ineffective if not damaging water treatment choices that could adversely affect public health.

Conclusions

The examples and discussion above lead to several general conclusions about the use of GIS with drinking water monitoring data:

- Temporal and spatial variability stems from many causes, down to a very small scale.
- Sampling protocols must be keyed to the precise questions under investigation.
- Regulatory sampling, whose results are generally readily available, is usually inappropriate to assess human exposure to trace metals or other parameters of interest (such as DBPs).
- Generalizations on a large scale are often impossible because of the geology and water chemistry variations.

Additionally, some considerations apply to the types of mapping that could be employed by GIS. For example, a mapping technique such as contouring may be especially inappropriate for use with drinking water data. Major problems could result from:

- Discrete, small-scale (such as within an individual house) variability in distributions of certain contaminants, such as lead and copper.
- Physical constraints of the distribution system network.
- The small number of monitoring sites in relation to the size of the distribution network.
- Different chemical or hydraulic zones in the distribution system.

Employing GIS could be very useful in solving a variety of drinking water problems. Users must be extremely conscious of the nature of the source information, however, to avoid abusive extrapolations and generalizations.

Acknowledgments

Some of the examples cited were investigated as part of a U.S. Environmental Protection Agency cooperative agreement (Darren A. Lytle, Project Officer) with the New England Water Works Association. Jack DeMarco of the Cincinnati Water Works kindly provided Figure 1 on distribution of sampling sites for lead and copper monitoring.

References

1. Schock, M.R. 1990. Internal corrosion and deposition control. In: Water quality and treatment: A handbook of community water supplies, 4th ed. New York, NY: McGraw Hill.
2. Benefield, L.D., and J.S. Morgan. 1990. Chemical precipitation. In: Water quality and treatment: A handbook of community water supplies, 4th ed. New York, NY: McGraw Hill.
3. Glaze, W.H. 1990. Chemical oxidation. In: Water quality and treatment: A handbook of community water supplies, 4th ed. New York, NY: McGraw Hill.
4. Cotruvo, J.A., and C.D. Vogt. 1990. Rationale for water quality standards and goals. In: Water quality and treatment: A handbook of community water supplies, 4th ed. New York, NY: McGraw Hill.
5. Tate, C.H., and K.F. Arnold. 1990. Health and aesthetic aspects of water quality. In: Water quality and treatment: A handbook of community water supplies, 4th ed. New York, NY: McGraw Hill.
6. Schock, M.R. 1989. Understanding corrosion control strategies for lead. JAWWA 81:7:88.
7. AWWARF. 1990. Lead control strategies. Denver, CO: American Water Works Association Research Foundation and American Water Works Association.
8. AWWARF. 1994. Corrosion control in water distribution systems, 2nd ed. Denver, CO: American Water Works Association Research Foundation/Engler Bunte Institute.
9. Schock, M.R., and I. Wagner. 1985. The corrosion and solubility of lead in drinking water. In: Internal corrosion of water distribution systems. Denver, CO: American Water Works Association Research Foundation/DVGW Forschungsstelle.
10. Pankow, J.F. 1991. Aquatic chemistry concepts. Chelsea, MI: Lewis Publishers, Inc.
11. U.S. EPA. 1995. Effect of pH, DIC, orthophosphate and sulfate on drinking water cuprosolvency. Cincinnati, OH: Office of Research and Development. In press.
12. AWWA. 1993. Initial monitoring experiences of large water utilities under U.S. EPA's Lead and Copper Rule. Denver, CO: Water Industry Technical Action Fund/American Water Works Association.
13. AWWA. 1990. Lead service line replacement: Benefit to cost analysis. Denver, CO: Water Industry Technical Action Fund/American Water Works Association.
14. AWWA. 1991. Lead and copper: A working explanation of the Lead and Copper Rule. Denver, CO: American Water Works Association.
15. Federal Register. 1991. Lead and copper: Final rule correction. Fed. Reg. 56:135:32,112 (July 15).
16. Federal Register. 1991. Lead and copper: Final rule. Fed. Reg. 56:110:26,460 (June 7).
17. Federal Register. 1992. Lead and copper: Final rule correction. Fed. Reg. 57:125:28,785 (June 29).
18. Schock, M.R. 1990. Causes of temporal variability of lead in domestic plumbing systems. Environ. Monit. Assess. 15:59.
19. Kuch, A., and I. Wagner. 1983. Mass transfer model to describe lead concentrations in drinking water. Water Res. 17:10:1,303.
20. Ireland, J.C. 1993. Alternatives to chlorine for disinfection of drinking water. In: Strategies and technologies for meeting SDWA requirements. Lancaster, PA: Technomic.
21. Miltner, R.J. 1993. Pilot scale treatment for control of disinfection byproducts. In: Strategies and technologies for meeting SDWA requirements. Lancaster, PA: Technomic.
22. Sheiham, I., and P.J. Jackson. 1981. Scientific basis for control of lead in drinking water by water treatment. J. Inst. Water Engrs. and Scientists 35:6:491.
23. Heumann, D.W. 1989. Los Angeles Department of Water and Power: Solid lead gooseneck slug dispersion in consumer plumbing system. Proceedings of the AWWA Water Quality Technology Conference, Philadelphia, PA.

Ground-Water Applications

Using GIS/GPS in the Design and Operation of Minnesota's Ground Water Monitoring and Assessment Program

Tom Clark, Yuan-Ming Hsu, Jennifer Schlotthauer, and Don Jakes
Minnesota Pollution Control Agency, St. Paul, Minnesota

Georgianna Myers
Water Management Consultants, Denver, Colorado

Abstract

Minnesota's Ground Water Monitoring and Assessment Program (GWMAP) is administered by the Minnesota Pollution Control Agency (MPCA) to evaluate baseline ground-water quality conditions regionally and statewide. The program uses a systematic sampling design to maintain uniform geographic distribution of randomly selected monitoring stations (wells) for ground-water sampling and data analysis. In 1993, geographic information system (GIS) and global positioning system (GPS) technologies were integrated into GWMAP, automating the selection of wells and the field determination of well locations.

GWMAP consists of three components: the statewide baseline network, regional monitoring cooperatives, and a trends analysis component. In the statewide baseline network, Minnesota is divided into over 700 121-square-mile grid cells, each with a centralized, 9-square-mile sampling region. Within each target area, single-aquifer, cased and grouted wells are sampled for about 125 metals, organic compounds, and major cations and anions. We are currently finishing the second year of a 5-year program to establish the statewide grid. When complete, the statewide baseline component will consist of about 1,600 wells representing Minnesota's 14 major aquifers.

In 1993, approximately 4,000 well construction records were selected for geologic and hydrologic review, using a GIS overlay, from a database of 200,000 water well records maintained in the state's County Well Index (CWI). Using GPS, 364 wells were sampled and field located. The semiautomatic well selection process uses existing electronic coverage of public land survey (PLS) data maintained in CWI in conjunction with the digitized systematic sampling grid. GIS has greatly reduced the time needed for selecting sampling stations. With the combination of GIS and GPS, program costs have decreased, allowing more resources to be applied toward sampling, while efficiency and quality of data have improved.

Introduction

Quantitative assessment of ground-water quality conditions requires a highly organized data collection program that includes statistical evaluation of monitoring results (1, 2). States have difficulty providing the staff and financial resources necessary to generate statewide quantitative ground-water information. With the use of geographic information system (GIS) and global positioning system (GPS) technologies, however, states have the potential to improve the quality of environmental monitoring programs and to reduce the amount of staff time needed to collect and evaluate data, thus decreasing costs. The degree to which states realize these potential benefits depends largely on how effectively the technology can be incorporated into the design of the monitoring program. This paper describes how GIS and GPS technologies are being integrated into the design and operation of Minnesota's Ground Water Monitoring and Assessment Program (GWMAP) to improve overall effectiveness.

The Minnesota Pollution Control Agency (MPCA) has sampled and analyzed ambient ground-water quality in the state's 14 principal aquifers since 1978. In 1990, the MPCA began a redesign of its ground-water monitoring program to better assess water quality conditions statewide (3). Three program components resulted from the redesign: a statewide baseline network for complete geographic coverage, a trends analysis component for intensive studies of how ground-water quality in specific areas changes with time, and a regional monitoring cooperative link to governmental units such as counties to meet specific local ground-water assessment needs. This paper describes the design and operation of the statewide baseline network.

The design of the statewide network is geographically and statistically based to automate well selection and data interpretation. In 1993, the MPCA began integrating GIS and GPS technologies into this part of the program. The implementation of GIS and GPS surpassed our expectations by reducing staff time required to select wells and evaluate analytical results (see Table 1). In addition, through the elimination of previously uncontrollable variables, the use of GIS and GPS has increased the accuracy of GWMAP data.

Monitoring Program Description

Since 1992, GWMAP has selected 150 to 250 existing water supplies yearly for ground-water sampling and analysis of about 125 parameters, including major cations and anions, metals, and volatile organic compounds. Well selection is a fundamental element of GWMAP that, if efficiently performed, supports the program objectives by upholding the quality of the monitoring data and minimizing the operating costs.

A key to the interpretation of monitoring data is the technique used to select wells for sampling (2, 4, 5). Minnesota has over 200,000 active water wells with approximately 10,000 new installations annually. For each well selected for GWMAP monitoring, a hydrologist must individually review many well construction records. An automated prescreening mechanism to facilitate well selection can result in considerable time (and therefore cost) savings. GWMAP chose GIS as the best tool for this task. GIS enables the program to combine a systematic sampling technique with hydrogeologic criteria to ensure an efficient and consistent selection process. As Table 1 shows, GIS allowed us to more than triple our geographic coverage and wells initially selected, while dramatically reducing the records that must be individually reviewed. We realized a time savings of 2 months compared with the time required before GIS implementation.

In general, systematic sampling techniques use a randomly generated uniform grid to determine sampling locations in space and/or time (5). Systematic sampling was initially implemented in GWMAP in 1991 using a manually generated spatial grid defined by the public land survey (PLS) (3). Although the PLS is not 100 percent geographically uniform, it was selected for the grid to expedite well selection from existing digital databases in which wells are organized by PLS location.

Systematic Sample Site Selection Using GIS

Systematic sample site selection is a three-step process. First, a database search of Minnesota's County Well Index (CWI) (6), containing nearly 200,000 driller's records, is conducted to include all available water wells in the region of interest. Second, the candidate pool is reduced to those wells located within regularly spaced grid cells. Third, further wells are eliminated from the candidate pool by applying geologic and well construction criteria mandated in the GWMAP design (7).

Generating the Sampling Grid

The statewide sampling grid was generated from a randomly selected origin (8). This grid consists of approximately 700 square cells, 11 miles on a side (see Figure 1). The centroid of each cell is consecutively numbered and was extracted to produce the origin of the sampling zone.

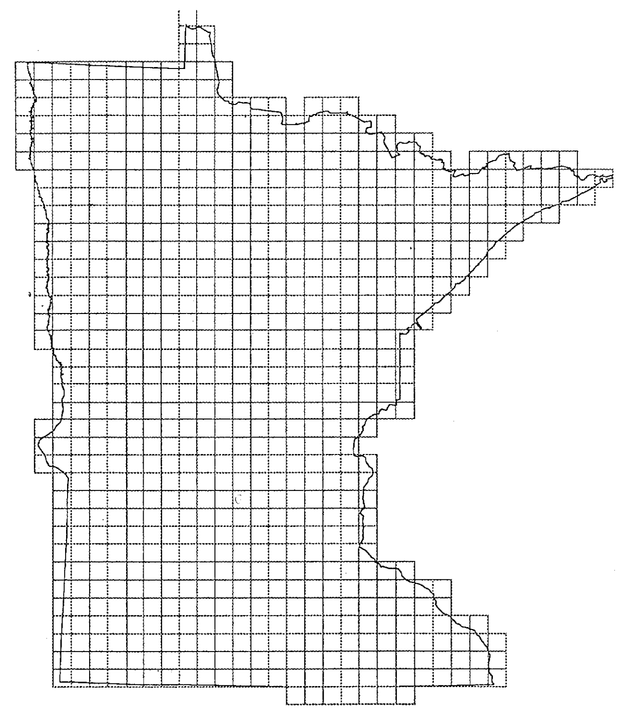


Figure 1. Statewide baseline network sampling grid.

Table 1. Well Selection in 1992 and 1993

Year	Area Covered	PLS Sections Selected	Well Logs Selected	Well Logs Reviewed	Wells Sampled	Time Spent
1992	9 counties	500	3,000	3,000	158	6 months
1993	26 counties	1,659	11,000	834	206	4 months

Establishing the Sampling Zone

Each sampling zone consists of a 3- by 3-mile box from which potential sampling sites are selected. It is generated by computing the coordinates of the four corners of the box using the grid cell's centroid as the origin. To link the sampling zone and grid cell, both are identified with the same numerical code.

These sampling "target" zones, a series of regularly spaced, 9-square-mile boxes, are then made into a GIS coverage and overlaid on top of the PLS coverage to extract those sections that are associated with each of the sampling zones. Ideally, each sampling zone should cover exactly nine PLS sections (3). Due to irregularities in the PLS system, however, portions of 16 to 20 sections usually fall within the sampling zone of each cell (see Figure 2).

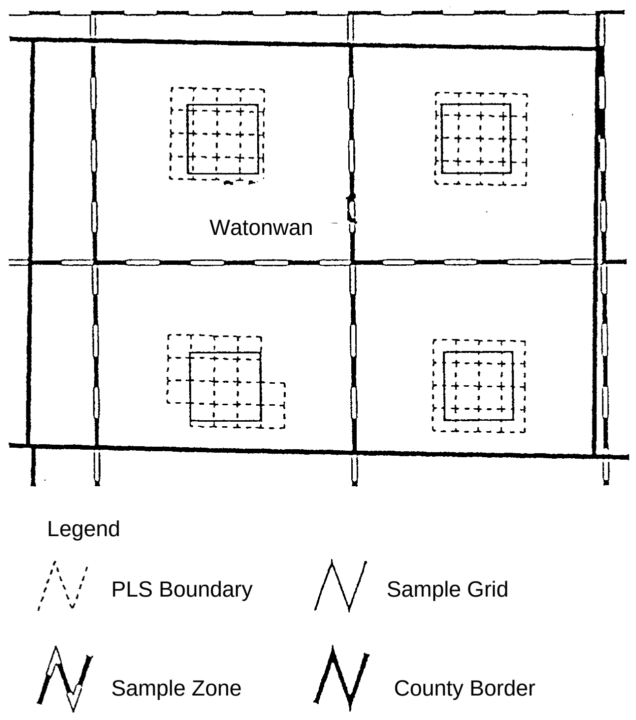


Figure 2. PLS and the sampling grid, Watonwan County.

Selection of PLS Sections

The PLS coverage was derived from the Minnesota Land Management Information Center (LMIC) "GISMO" file. It was originally created in 1979 by digitizing every section corner in Minnesota from the U.S. Geological Survey (USGS) 7.5-minute quadrangle map series.

The PLS section information is necessary in the well selection process because the original well construction logs, maintained by the Minnesota Geological Survey (MGS), are organized by PLS. Although most of the well selection process can be automated, manual file searches for well records are still necessary and require the PLS information.

Well Selection

After identifying the PLS sections within the sampling grid, the statewide well database is imported as a point coverage and overlaid with the selected PLS section coverage. Thus, all wells that fall within the 16 to 20 sections are selected as potential candidates. The accuracy of the well locations in CWI varies; most of the point locations are approximated to four quarters (2.5 acres). The CWI does not contain all well construction information, however, requiring that copies of driller's logs be made for GWMAP files.

The final well selection is done after applying the 9-square-mile sampling zone over the potential pool of candidates. For wells that fall within the zone, the well construction records are pulled from MGS files, copied, and submitted for hydrologist review. Depending on the target cell location, the number of candidate wells requiring review may range from a few to more than 100. For newly installed water wells whose records have not yet been digitized by LMIC, the PLS locations of the wells are manually plotted onto a map to confirm whether they fall into a sampling grid cell. Typically, from 5 percent to as many as 20 percent of selected wells that meet the location criteria are sampled. This accounts for the hydrogeologic and well construction criteria and the cooperation of well owners participating in the program. Currently, interest in ground-water protection programs runs high in rural Minnesota, with an acceptance rate of up to 80 percent.

The implementation of GIS in well selection helped GWMAP excel in two major areas. First, the development of the statewide GIS grid eliminated previously uncontrolled variables by removing the PLS spatial inconsistencies from the systematic grid. Second, the GIS reduced the manual workload with the automation of two important steps in the well selection process: the generation of PLS section information to facilitate the database search, and the identification of wells that meet the geographic location criteria. The success of GWMAP relies largely on the ability to use existing GIS coverages. In using coverages created by other entities, this program identified the need for a uniform standard for data conversion and transfer.

Application of Global Positioning Systems in Ground-Water Sampling

In 1991, the U.S. Environmental Protection Agency (EPA) established a policy that all new data collected after 1992 should meet an accuracy goal of 25 meters or better (9). The purpose of EPA's Locational Data Policy (LDP) is to establish principles for collecting and documenting consistently formatted locational data to facilitate cross-programmatic, multimedia analyses. Accurate geographic information is important to the spatial analysis of well sampling results. Any uncertainty in sample location can compromise hydrogeologic analysis (10). GPS is an easy, cost-effective solution.

Global Positioning System Field Application

Beginning in October 1992, GWMAP employed GPS in the field to assist in locating sample sites. Applying GPS in the field has proven to be quite easy. The program uses a multichannel C/A code receiver with internal data logging capability. Typically, the receiver is placed directly on top of the wellhead and logs 100 to 150 GPS readings into the receiver's internal memory in approximately 5 minutes.

The GPS is also used for navigation in the field to locate sampling sites. Because sampling sites are predetermined, their locations can be extracted from a topographic map. The approximate coordinates can then be loaded into a GPS receiver. In most cases, the receiver successfully led the field team within visual range of the sampling site.

Because of the inherent selective availability (SA) of the GPS, raw field data must go through a differential correction process to achieve the goal of 25-meter accuracy (9, 11).

Data Management and Processing

Once the GPS receiver is brought back from the field, data are downloaded to a personal computer (I486 processor at a speed of 50 MHz) and differentially corrected (11). The average or mean of the 100 or more readings collected onsite is calculated and reported as the site location.

The MPCA does not operate a GPS base station for the purpose of differential correction. The base station data are obtained through a computer network (Internet) from the Minnesota Department of Health (MDH) base station located in Minneapolis.

To facilitate future data integration and document data accuracy for secondary application, GWMAP proposed quality assurance codes for GPS data collected by the MPCA. The value of the accuracy proposed is a nominal value rather than an absolute number (see Table 2). Each of the seven processing methods is assigned a separate code.

In the field experience of GWMAP, a nominal accuracy of 2 to 5 meters has been consistently achieved after the postdifferential correction and averaging have been applied to the data. This technology is suitable for any program that is designed to conduct either large-area or intensive monitoring activities. It helps to cut costs by increasing efficiency and accuracy of the data. The data collected by GWMAP can be used not only in a regional study but could be used directly in a site-specific investigation as well.

Table 2. Proposed Nominal Accuracy Reference Table

Type of GPS Receiver Used	Processing Method Used To Correct Data	Nominal Accuracy (meters)
Navigational quality C/A code receiver	Postdifferential corrected	2-5
	Real-time differential corrected (RTCM)	2-5
	Autonomous mode (no correction)	15-100
Navigational quality with carrier aid receiver	Postdifferential corrected	< 1
	Real-time differential corrected (RTCM)	< 1
	Autonomous mode (no correction)	15-100
Survey quality receiver (dual or single frequency)	Postdifferential corrected	< 0.1

GWMAP also found that GPS can be used most efficiently by separating the two roles of field operator and data manager. The field operators receive only the brief instructions necessary to operate a GPS receiver before going into the field. The data manager handles the data processing details. The field operators can then concentrate their efforts on obtaining ground-water samples and conducting the hydrogeologic investigation.

Conclusions

GIS and GPS technologies made it possible for the MPCA to implement the statewide GWMAP project by optimizing the available funding and staff time. GIS minimized staff time spent on identifying sampling areas, manipulating the sampling grid, and selecting monitoring sites. In addition, GIS enabled GWMAP to integrate a variety of databases and maps of different scales.

Using GPS to locate sampling sites enabled GWMAP to efficiently obtain accurate geographic locational data with relative ease. This eliminated the degree of uncertainty that previously might have compromised the statistical evaluation of the hydrogeologic data.

GWMAP's success in integrating existing digital data to automate the well selection process clearly demonstrated the importance of the ability to share information with others and the great need for a broadly applied standard for data conversion and transfer.

Acknowledgments

The authors wish to thank Renee Johnson of the Minnesota Department of Natural Resources for her work to convert the PLS data layer to GIS coverage. Susanne Maeder of LMIC supplied the statewide CWI coverage, and Susan Schreifels of MPCA conducted research on the LDP and made valuable suggestions on implementing GPS.

References

1. Nelson, J.D., and R.C. Ward. 1981. Statistical considerations and sampling techniques for ground-water quality monitoring. Ground Water 19(6):617-625.
2. Ward, R.C. 1989. Water quality monitoring—A systems approach to design. Presented at the International Symposium on the Design of Water Quality Information Systems, Colorado State University, Ft. Collins, CO.
3. Myers, G., S. Magdalene, D. Jakes, and E. Porcher. 1992. The redesign of the ambient ground water monitoring program. St. Paul, MN: Minnesota Pollution Control Agency.
4. Olea, R.A. 1984. Systematic sampling of spatial functions. Series on Spatial Analysis, No. 7. Kansas Geological Survey, University of Kansas, Lawrence, KS.
5. Gilbert, R.O. 1987. Statistical methods for environmental pollution monitoring. New York, NY: Van Nostrand Reinhold.
6. Wahl, T.E., and R.G. Tipping. 1991. Ground-water data management—The county well index. Report to the Legislative Commission on Minnesota Resources. Minnesota Geological Survey, University of Minnesota, St. Paul, MN.
7. Clark, T., Y. Hsu, J. Schlotthauer, and D. Jakes. 1994. Ground-water monitoring and assessment program—Annual report. St. Paul, MN: Minnesota Pollution Control Agency.
8. ESRI. 1993. ARC/INFO version 6.1, ARCPLOT command references. Redlands, CA: Environmental Systems Research Institute, Inc.
9. U.S. EPA. 1992. Global positioning systems technology and its application in environmental programs. GIS Technical Memorandum 3. EPA/600/R-92/036. Washington, DC.
10. Mitchell, J.E. 1993. A characterization of the influence of (x,y) uncertainty on predicting the form of three-dimensional surfaces. Proceedings of the AWRA Spring Symposium on Geographic Information Systems and Water Resources, Mobile, AL. pp. 559-567.
11. Trimble Navigation, Ltd. 1993. GPS Pathfinder System, general reference. Sunnyvale, CA: Trimble Navigation, Ltd.

Use of GIS in Modeling Ground-Water Flow in the Memphis, Tennessee, Area

James Outlaw and Michael Clay Brown
University of Memphis, Memphis, Tennessee

Abstract

Memphis, Tennessee relies solely on ground water for its municipal and industrial water supply. Memphis Light, Gas, and Water (MLGW) Division owns and operates over 160 water wells in 10 production fields throughout Shelby County. MLGW produces an average of approximately 200 million gallons per day, excluding much of the industrial demand. The city obtains its water from a thick, prolific aquifer known as the Memphis Sand, which was thought to be separated from a surficial aquifer by a thick confining layer. In recent years, evidence of leakage from the surficial aquifer to the Memphis Sand has been found.

The University of Memphis Ground Water Institute (GWI) is developing a hydrogeologic database of the Memphis area to study the aquifer. The database serves as the basis for several ground-water flow models that have been created as well as part of the wellhead protection programs currently being developed for Memphis and other municipalities in Shelby County. A geologic database was developed and is constantly being updated from borehole geophysical logs made in the area. Well locations are being field verified using a global positioning system (GPS).

Use of the database has allowed the development of a three-dimensional model of the Memphis area subsurface. The database also contains locations of and information on both private and public production and monitoring wells, Superfund sites, underground storage tanks, city and county zoning, land use, and other pertinent information. Procedures for linking the database to ground-water flow and solute transport models have been developed. The data visualization capabilities and the ability to link information to geographic features make geographic information systems (GIS) an ideal medium for solving ground-water problems.

An example of GIS use in ground-water flow modeling is the study of the Justin J. Davis Wellfield. The water quality parameters of alkalinity, hardness, sulfate, and barium have significantly increased over the past 10 years at this facility. To understand why these changes are occurring, MLGW, the GWI, and the U.S. Geological Survey (USGS) participated in a joint investigation of the wellfield.

In the spring of 1992, a series of 12 monitoring wells was drilled into the surficial aquifer near the production wells. Geophysical logging and split-spoon sampling revealed an absence of the confining layer, referred to as a window, at one of the monitoring wells. All other wells penetrated various thicknesses of clay. This window in the confining layer suggests that the water quality changes could be due to leakage from the surficial aquifer to the Memphis Sand.

The GIS database was used to construct a flow model of the Davis area. Also, using the surface modeling capabilities of GIS, the extent of the confining layer window was estimated and used to calculate leakage between the two aquifers. The results of these analyses also indicate that further subsurface exploration is needed to more accurately define the extent of the confining layer window.

Introduction

Memphis, Tennessee, relies solely on ground water for its municipal and industrial water supply. The Memphis Light, Gas, and Water (MLGW) Division owns and operates over 160 water wells in 10 production fields throughout Shelby County, as shown in Figure 1. MLGW produces an average of approximately 200 million gallons per day, excluding much of the industrial demand.

The city obtains its water from a thick, prolific aquifer known as the Memphis Sand, which was thought to be separated from the surficial aquifer by a thick confining layer. In recent years, evidence of leakage from the surficial aquifer to the Memphis Sand has been found. The University of Memphis Ground Water Institute (GWI) is developing a hydrogeologic database of the Memphis area to study the aquifer. Several ground-water

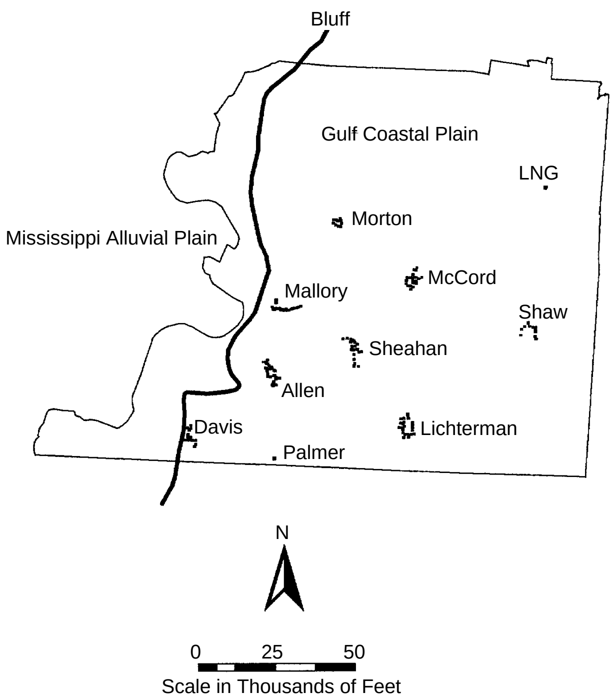


Figure 1. Physiographic description and MLGW Wellfields in Shelby County.

flow models have been developed using the database. Also, the database is an integral part of wellhead protection programs being developed for Memphis and other municipalities in Shelby County. A geologic database was developed and is constantly being updated from borehole geophysical logs made in the area. Well locations are being field verified using a global positioning system (GPS).

Use of the database has allowed the development of a three-dimensional model of the subsurface of the Memphis area. The database also contains locations of and information on both private and public production and monitoring wells; Superfund sites; underground storage tanks; city and county zoning; land use; and other pertinent information. Water quality measurements for every MLGW production well have been obtained, and a history of water quality for the Memphis Sand is being developed. Procedures have been developed for linking the database to ground-water flow and solute transport models (1). The data visualization capabilities and the ability to link information to geographic features make geographic information systems (GIS) an ideal medium for solving ground-water problems.

GIS Database

The GWI has developed and is continuing to update a hydrogeologic database for the Memphis area. ARC/INFO (marketed by Environmental Systems Research Institute, Redlands, California) is the GIS program that the GWI is using. The program runs on a network of 10 SUN SPARC stations. The capabilities of ARC/INFO and the computational speed of the SPARC stations allow very sophisticated ground-water analyses to be performed and have allowed the development of an extensive electronic database.

The basic unit of data storage in ARC/INFO is a coverage. A coverage is a digital representation of a single type of geographic feature (e.g., points may represent wells, lines may represent streets or equipotential lines, and polygons may represent political boundaries or zoning classifications). Information may be associated with an individual geographic feature in a feature attribute table. This information may then be queried and used in analyses. ARC/INFO also has its own macro language that allows the customization and automation of many ARC/INFO procedures.

A relatively new feature of ARC/INFO is address matching. This procedure compares a file containing the street address of a particular feature with an address coverage. This coverage is basically a library of addresses that are linked to a geographic coordinate. As the addresses from the input file are compared with the address coverage, the matching points are written to a second coverage. Any addresses in the input file that do not match an address in the address coverage are written to a "rejects" file. These can be matched by hand on a one-by-one basis.

Address matching serves as an alternative to digitizing, as long as a good address coverage for a specific area exists. The GWI has used this capability extensively and has developed a coverage of underground storage tank (UST) locations inside Shelby County. A database of private and monitoring wells is also being developed and updated using address matching. The raw information was obtained in an ASCII format from the appropriate regulating agencies (i.e., the Tennessee Department of Environment and Conservation Division of Underground Storage Tanks and the Memphis/Shelby County Health Department). The ASCII information was imported into ARC/INFO and address matched. The creation of a suitable address coverage and completion of the address matching of the UST file has taken almost 2 years. The private well coverage is currently being updated from historical information provided by regulating agencies and local well drilling companies.

An important part of the database is the geologic information obtained from geophysical logs in the area. Gamma logs, resistivity logs, and spontaneous potential (SP) logs are three major types of electric geophysical logs. Gamma logs measure naturally occurring radiation emitted from soil in the borehole. Clays and shales emit gamma rays. A high gamma count indicates the presence of clay or shale, and a low gamma count implies that little or no clay is present. Sand layers that contain fresh water are located using resistivity logs. Maximum values of resistivity indicate the possibility of a sand layer. Clays and sands that contain salt water may exhibit similar resistivities. SP logs are used to differentiate between the two (2).

Corroborating data such as formation logs, geologic studies, and available material samples should be consulted when reading and interpreting geophysical logs. The accuracy and reliability of an application based on well logs is completely dependent on a realistic interpretation of the geophysical data. A sample interpretation of a set of geophysical logs is shown in Figure 2. The results of interpretations like this are entered into the point attribute file of a well coverage.

The Triangulated Irregular Network (TIN) module of ARC/INFO is used to create a three-dimensional surface from information stored in a coverage. TIN creates a surface from a set of nonoverlapping triangles defined by a set of irregularly or regularly spaced points. In this study, the points defining the triangular TIN surfaces are the locations of the wells in the model area. TIN uses various interpolation routines to estimate surface values. Once the surfaces have been developed, two-dimensional profiles can be made that show the relative thicknesses of the various soil strata, as shown in Figure 3. These profiles aid in selecting boundary conditions and defining layers in ground-water flow models (3, 4).

In addition to the creation of profiles, a process that extracts surface values for use in a ground-water flow model has been developed. The GWI uses the United States Geological Survey (USGS) flow model, MODFLOW (5). Being a cell-based model, MODFLOW requires a value for each hydrogeologic parameter for each cell in the model grid. A series of FORTRAN programs and arc macro language (AML) programs were coupled to extract the required hydrogeologic data from surface models. For example, piezometric surface values are required to set initial conditions for the model. A coverage of the piezometric surface of the Memphis Sand was created, converted to a TIN surface, and the required values for each cell in the model were extracted using the procedure described above.

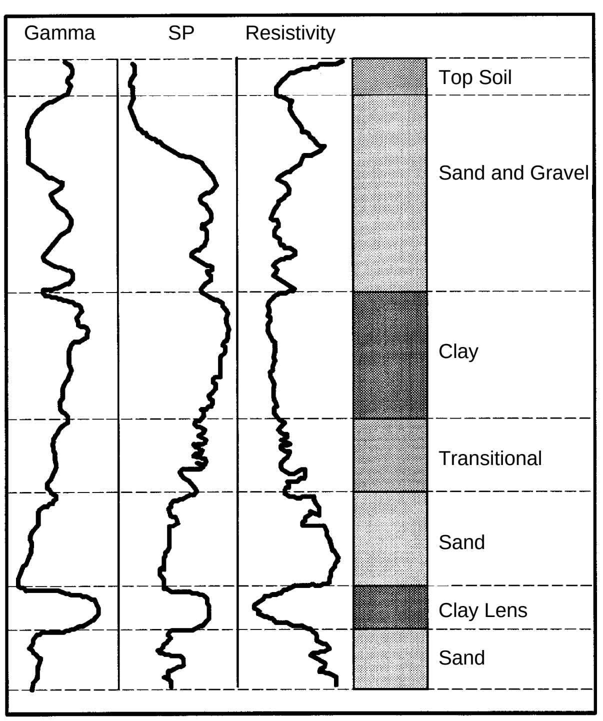


Figure 2. Example interpretation of geophysical logs.

The results and hydrogeologic data from the calibrated model can be read back into the database and converted into coverages. This allows piezometric contours to be developed and displayed with other information in the database to aid in decision-making. Also, capture zones for the wells can be brought into the database and compared with surface features like industries, landfills, Superfund sites, UST locations, or other sites that may have an impact. This has proved especially helpful in developing wellhead protection programs where a complete contaminant source inventory must be performed for the capture zone of each well and within a fixed radius around the well. The procedure used to develop model data from the GIS database is summarized in Figure 4.

McCord Wellfield Wellhead Protection Program

MLGW and the GWI performed a demonstration project funded by the U.S. Environmental Protection Agency (EPA) for the C.M. McCord Wellfield Wellhead Protection Program (6). This wellfield was selected because of multijurisdictional problems that will be encountered during plan implementation. The City of Memphis owns all the wells, but many of them, and all future well lots, are located within the city limits of Bartlett. A wellhead protection plan will have to involve the cooperation of both municipalities. The existing wellfield is shown in Figure 5.

Tennessee wellhead protection regulations require the delineation of two zones of protection for a city the size of Memphis: a 750-foot radius around the wellhead and a 10-year capture zone for the well. The 10-year capture zone (called the Zone 2 area) was delineated using two flow models and information obtained from the GWI database. Results were imported into the GIS database and compared with existing information. The Zone 1 area was delineated by buffering each well point in the coverage by the appropriate radius. Each well location was verified using a Trimble GPS unit and is accurate to within 2 meters. A contaminant source inventory was performed using the coverages developed by address matching. The primary potential sources of contamination in this area are USTs. A windshield survey located other potential sources, such as dry cleaners. These locations were entered into the database also by using

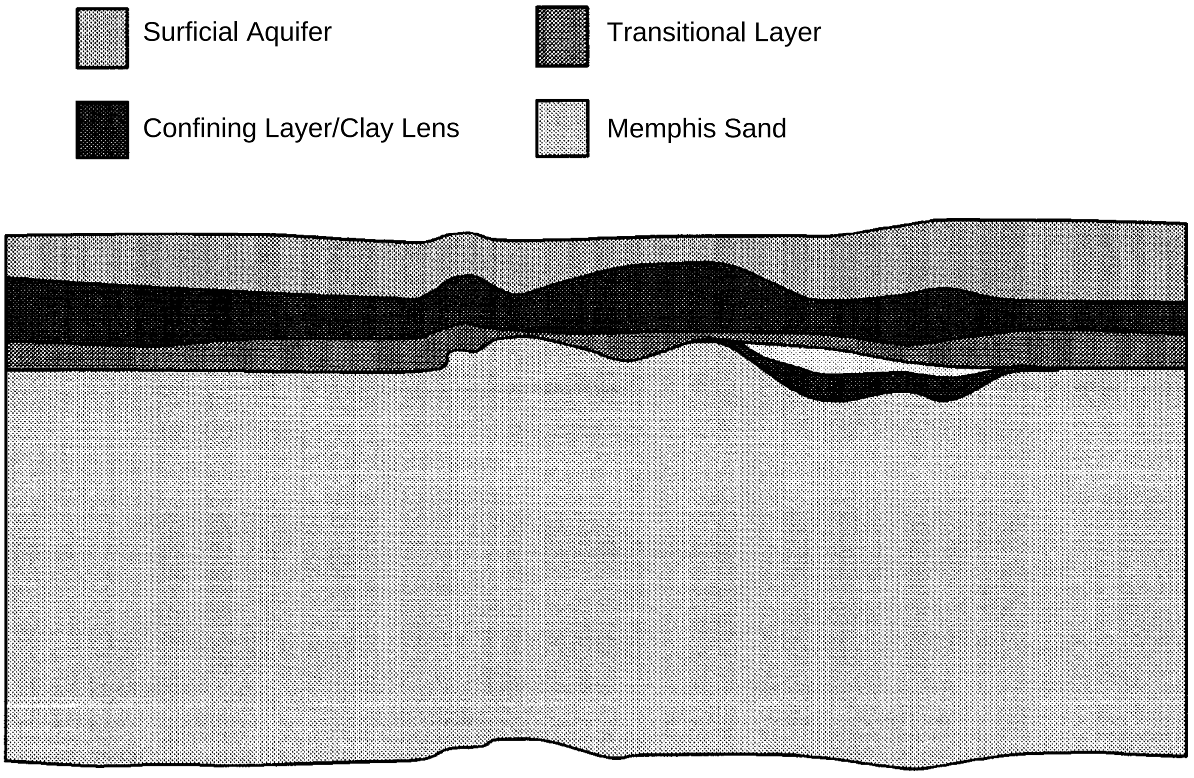


Figure 3. Two-dimensional profile constructed from surface models.

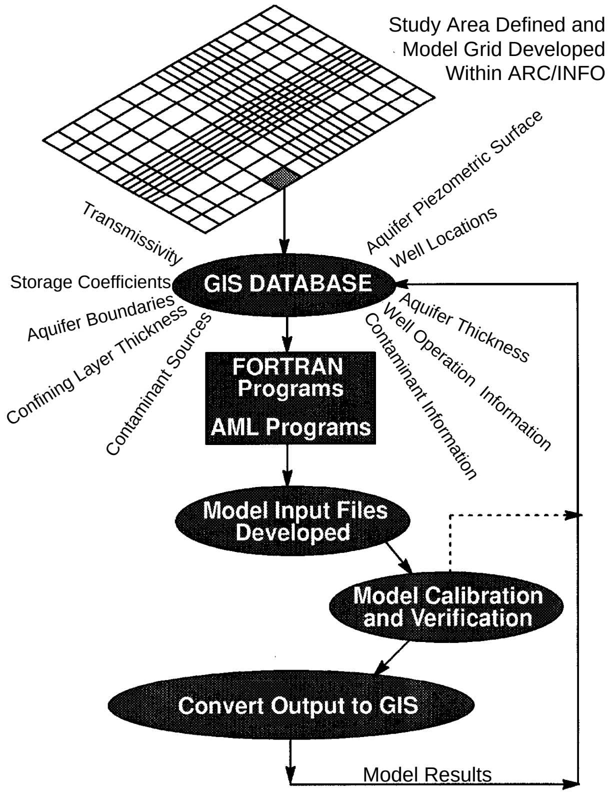


Figure 4. Procedure for integrating GIS and flow model.

the address-matching capabilities of ARC/INFO. The Zone 2 areas for present wells and future wells, along with the potential sources of contamination, are shown in Figure 6.

Davis Wellfield Study

An example of GIS use in ground-water flow modeling is the study of the Justin J. Davis Wellfield. The Davis Wellfield is one of 10 producing fields operated by MLGW. It is located in the southwestern corner of Shelby County and consists of 14 wells, as shown in Figure 1. Production at the Davis Wellfield began in 1971, and an estimated 13 million gallons per day are currently withdrawn from the Memphis Sand aquifer. Since 1972, MLGW has collected water quality data from the wells at the Davis Wellfield, including values for alkalinity, hardness, chloride, sulfate, iron, and barium. Water quality parameters of alkalinity, hardness, sulfate, and barium have significantly increased in the past 10 years.[1] A possible explanation for the change in water quality is water leakage from the upper aquifer through the confining unit to the Memphis Sand aquifer.[1] The water chemistry from the two aquifers is noticeably different. The surficial aquifers generally have a higher total dissolved solids concentration, hardness, and alkalinity than water from the Memphis Sand.[1]

[1] Webb, J. 1992. Memphis Light, Gas, and Water Division. Personal interview.

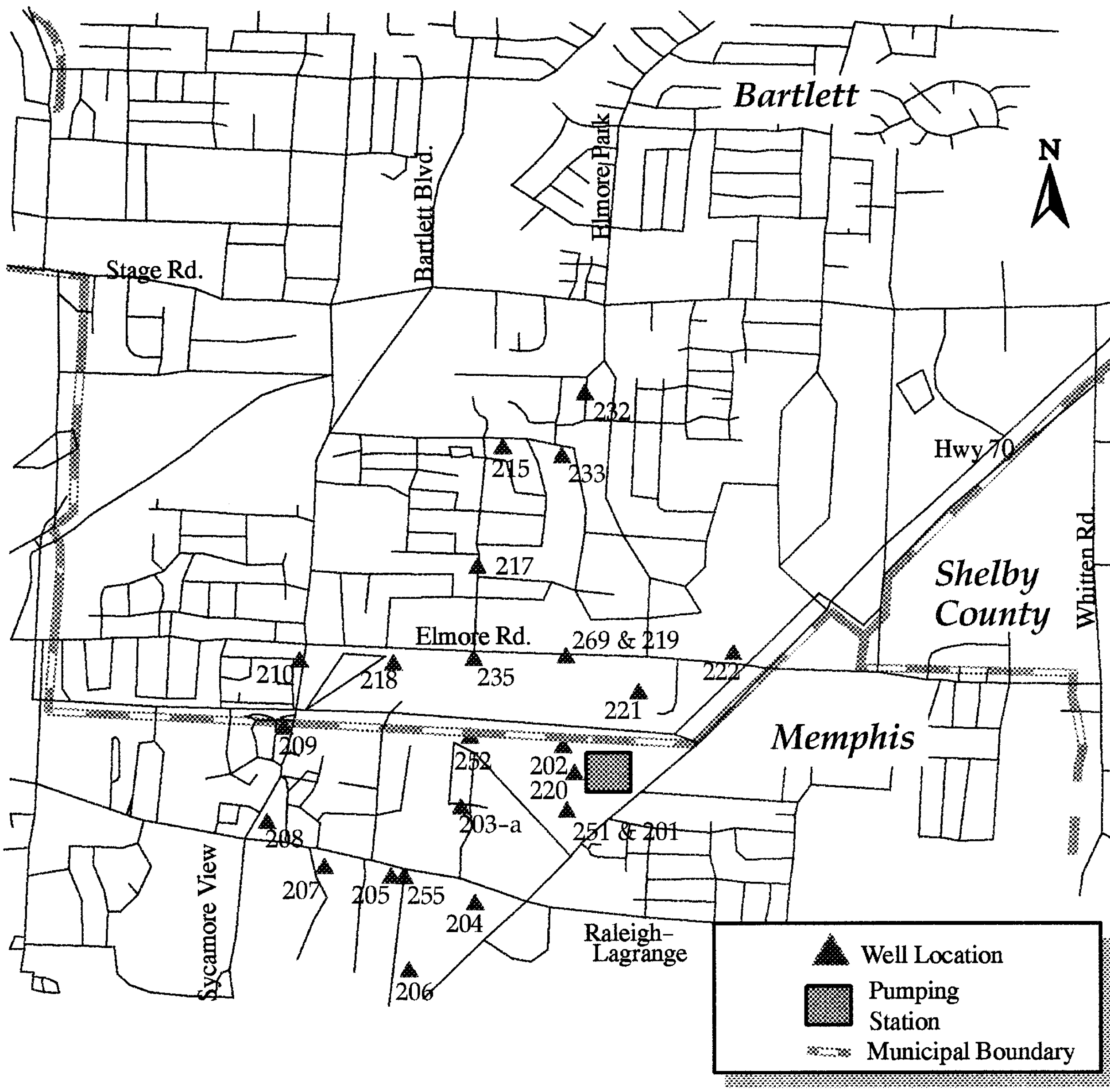


Figure 5. Existing McCord Wellfield.

MLGW, the GWI, and the USGS participated in a joint investigation of the wellfield to determine why the water quality changes are occurring. In the spring of 1992, 12 monitoring wells were drilled into the surficial aquifer near the production wells, as shown in Figure 7.

Geophysical logging and split-spoon sampling revealed an absence of the confining layer at one monitoring well, GWI-3. All other wells penetrated various thicknesses of clay. This “window” in the confining layer suggests that the water quality changes could be due to leakage from the surficial aquifer to the Memphis Sand. The logs for these monitoring wells were combined with an existing geophysical log coverage. The extent of the confining layer window was estimated using GIS surface modeling capabilities, as shown in Figure 8.

Two-dimensional profiles were created to further show the extent of the confining layer window. The locations of the profiles in relation to various surface features are shown in Figure 9. Profiles 1 and 2 were taken across the river bluff, and Profiles 2 and 3 were taken across the window. The profiles are shown in Figure 10.

Many important features of this area’s geology can be inferred by looking at the profiles. A connection of the alluvial and fluvial aquifers is shown in Profile 2. Elsewhere along the bluff, the connection of the two aquifers is less prominent, as shown in Profile 1. The connection of the two aquifers in Profile 2 may be the cause of a peculiar mounding effect in the water table of the alluvial aquifer in that area. The thinning of the top soil in Profiles 1 and 3 may indicate a local recharge area for the

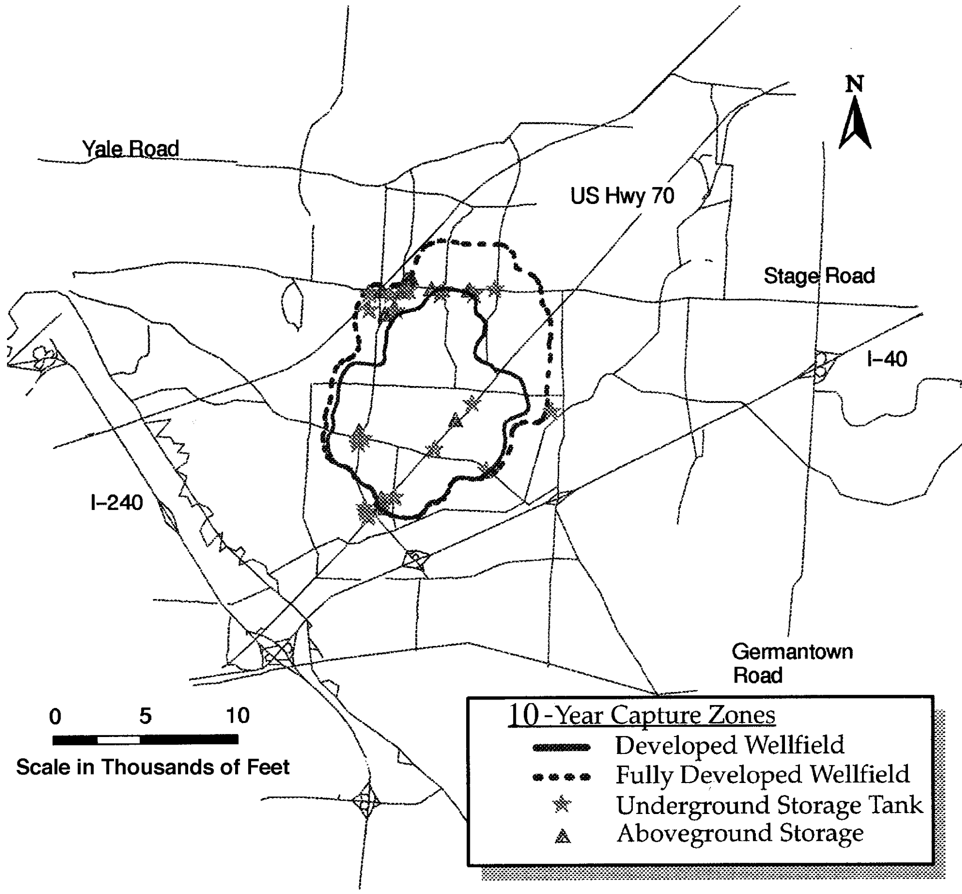


Figure 6. Underground storage tanks and aboveground storage locations.

alluvial aquifer. Profiles 2 and 3 show the confining layer window that suggests a connection between the surficial aquifer and the Memphis Sand.

Following the convention of the USGS, a "window" is defined as any area where the aggregate clay thickness is less than 10 feet (8). A surface of the thickness of the confining layer was generated from the geophysical log coverage. The surface model was converted to a contour line coverage on a 5-foot interval. Using the ARC-EDIT module of ARC/INFO, the contour line coverage was converted to a polygon coverage. The area bounded by the 10-foot contour of the surface model was calculated to be 840,000 square feet (about 19 acres). The area was calculated by adding the areas between the 10- and 5-foot contours and the area within the 5-foot contour.

A flow model of the area was developed based on the hydrogeologic data contained in the database. A steady state model was calibrated to hydraulic conditions recorded during fall 1992 by the USGS and the GWI; the root mean square (RMS) error for this model was 1.76 feet. A second steady state model was developed to simulate conditions recorded during spring 1993 (a period of high water levels in area lakes). The RMS error for this simulation was 5.19 feet. This higher error may indicate that the high water levels in surface water bodies are not realistic for steady state boundary conditions. Realistically, monitoring wells that are relatively far from a surface water body are affected more by average water levels over time rather than relatively short periods of highs and lows.

Using average values of h_1, h_2 (head in upper and lower aquifers), *l* (vertical flow distance), and *VCONT* (a parameter used in MODFLOW to allow for vertical conductance), the estimated flow rate through the window for fall 1992 may be computed as:

$$k = \text{VCONT } 1 = 1.76e^{-3} \times 199.2 = 0.351 \,\frac{\text{ft}}{\text{day}}$$

$$i = \frac{h_1 - h_2}{l} = \frac{186.5 - 156.1}{199.2} = 0.153$$

$$Q = kAi = 0.351 \times 840{,}000 \times 0.153 = 45{,}111 \,\frac{\text{ft}^3}{\text{day}}$$

$$45{,}111 \,\frac{\text{ft}^3}{\text{day}} \times 7.48 \,\frac{\text{gallons}}{\text{ft}^3} = 337{,}430 \,\frac{\text{gallons}}{\text{day}} = 0.34 \text{ MGD}$$

The flow rate calculated from average spring 1993 (a period of high water levels in the surficial aquifers) model results was computed as:

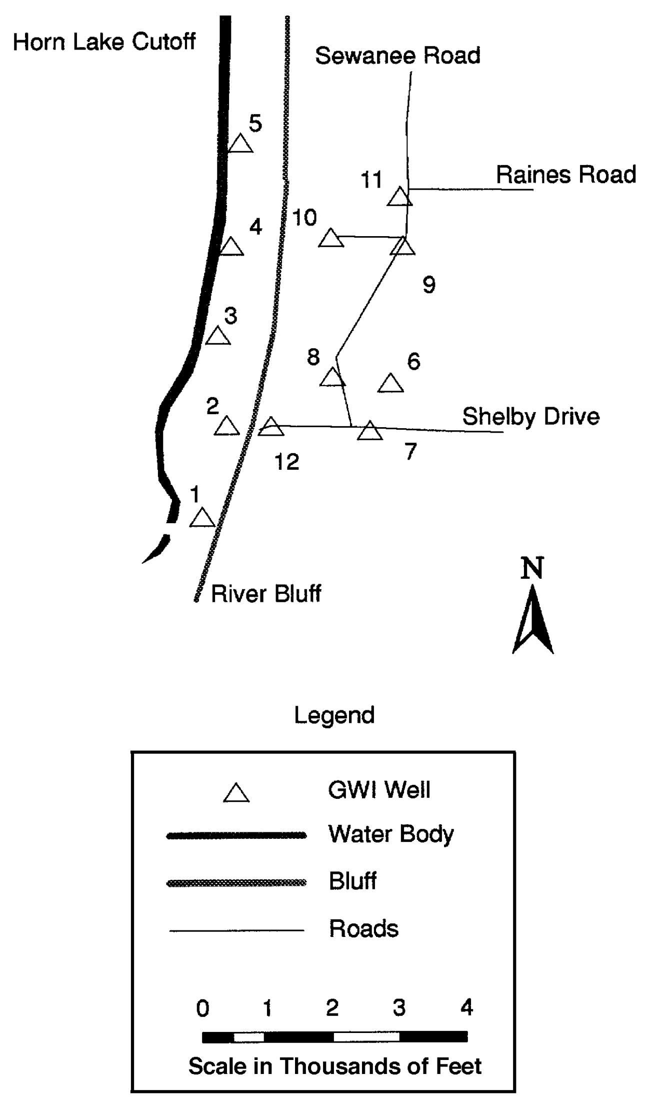


Figure 7. Location of GWI monitor wells.

$$k = VCONT\ l = 1.76e^{-3} \times 199.2 = 0.351\ \frac{ft}{day}$$

$$i = \frac{h_1 - h_2}{l} = \frac{196.4 - 156.8}{199.2} = 0.199$$

$$Q = kAi = 0.351 \times 840{,}000 \times 0.199 = 58{,}673\ \frac{ft^3}{day}$$

$$58{,}673\ \frac{ft^3}{day} \times 7.48\ \frac{gallons}{ft^3} = 438{,}874\ \frac{gallons}{day} = 0.44\ \text{MGD}$$

Using the GIS-delineated window, an estimated 0.34 to 0.44 million gallons per day flow from the alluvial aquifer to the Memphis Sand aquifer. This variation in the flow rate was due to seasonal variations of water level in the alluvial aquifer.

Since the wellfield pumps approximately 13 million gallons per day, the effect of the window on the entire Davis Wellfield may not be significant. The window lies within the 30-year capture zone of two wells in the field, however.

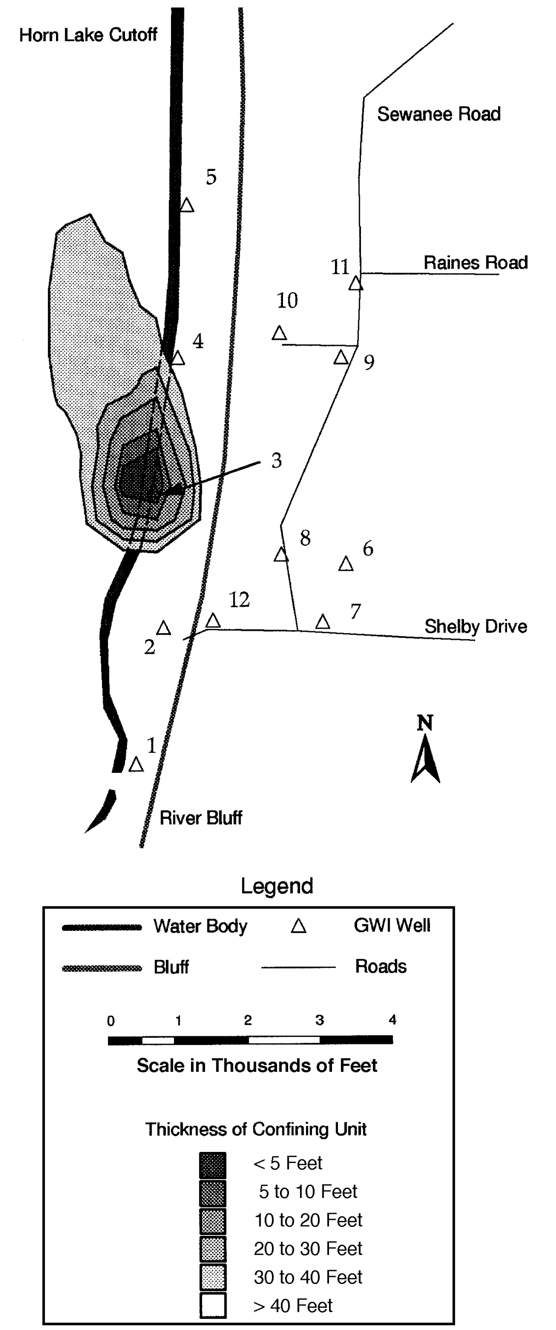


Figure 8. Location and extent of window in confining unit.

The flow rate through the window is approximately 20 percent of the total production of these two wells. Additionally, since the wells would probably not operate simultaneously, the flow rate through the window may account for approximately 40 percent of the flow at either well.

Particle tracking in the Memphis Sand was developed using MODPATH (9) from MODFLOW results. Particles were placed at model well screens and tracked backward for 30 years. The output from MODPATH was read into the GIS database for comparison with other data, as shown in Figure 11. The hole in the confining unit lies

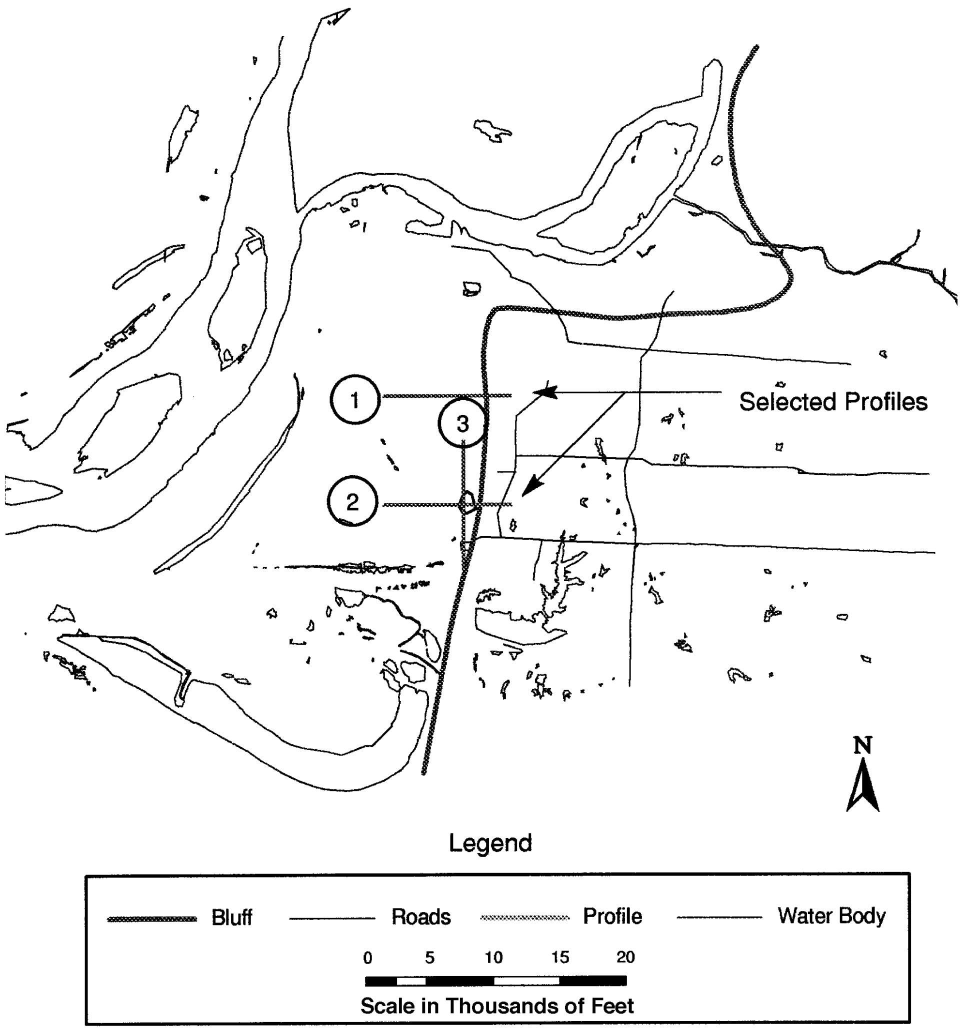


Figure 9. Location of selected profiles.

within the 30-year capture zone of MLGW wells 418 and 419. Historically, 419 was the first well that experienced water quality changes. The water quality is becoming similar to water found in the alluvial aquifer.

A change in the water quality in well 418 is not as immediately noticeable as in well 419. This inconsistency in data may indicate that the window does not extend northward from GWI-3, as the TIN model predicted. To determine which capture zone (418 or 419) encompasses GWI-3, particles were tracked backward for 40 years, as shown in Figure 12. GWI-3 lies on the edge of the capture zone for 419. The flow lines from 418 and 419 move toward the northwest and southwest in the Memphis Sand, pass up through the window, and emerge in the upper aquifer.

Conclusions

The explanation of the Davis Wellfield investigation addressed some limitations of the database. The utility that this hydrogeologic database provides greatly outweighs the disadvantages, however. Without the ability to map and define hydrogeologic features, this project may not have been completed in the allotted time frame or may not have been completed in the same level of detail. GIS greatly enhances the development and evaluation of ground-water flow models.

Specific conclusions that can be drawn from the analysis performed in this project are:

- The delineation of a window in the confining layer using a GIS database is possible.
- Based on the GIS-generated window, an estimated 0.34 to 0.44 million gallons per day flow from the upper aquifer to the Memphis Sand, which may account for as much as 40 percent of the flow at either well 418 or 419.
- The drilling of more monitoring wells north, south, east, and west of GWI-3 may provide for a more accurate delineation of the window.

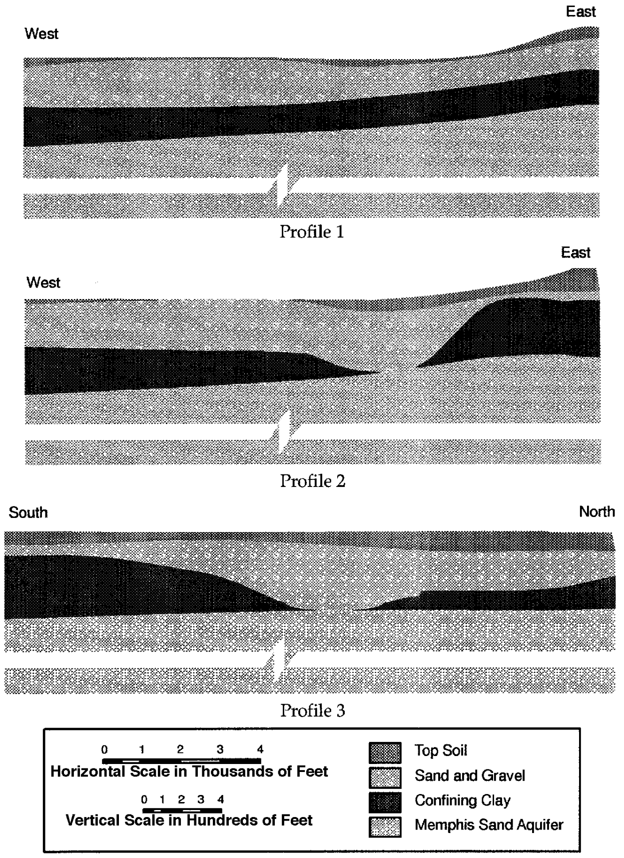


Figure 10. Selected subsurface cross sections in the Davis Wellfield.

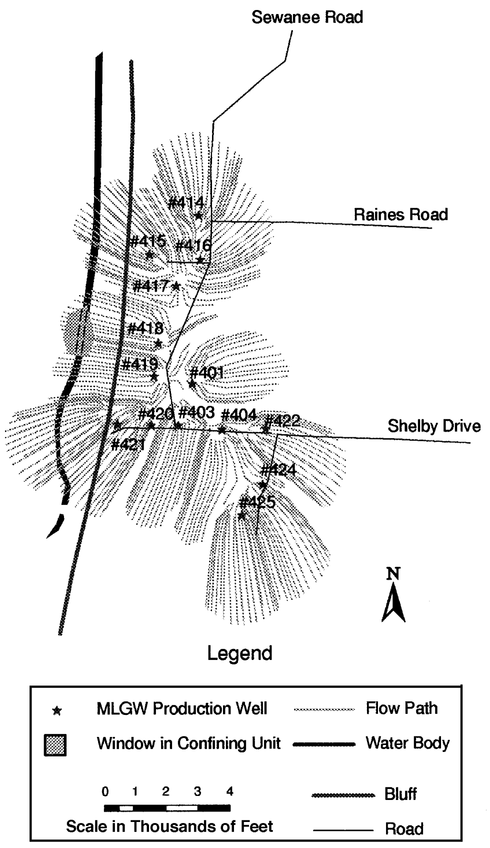


Figure 11. Backward tracking for 30 years.

General conclusions that may be drawn from this discussion are:

- GIS provides a convenient method of viewing flow model input and output.
- A flow model may be developed and evaluated in a relatively short time using GIS.
- GIS provides a convenient means of compiling and managing the information required to develop a wellhead protection program.

Some GIS disadvantages that have been noted are:

- The time required to develop a database and learn to apply the GIS program in a particular situation may be prohibitive.
- GIS-generated results from a limited database may be misleading and should be corroborated with other analysis methods.

Acknowledgments

The authors would like to acknowledge the efforts of the professors and students of the GWI, both present and past, for their contributions to the database and this

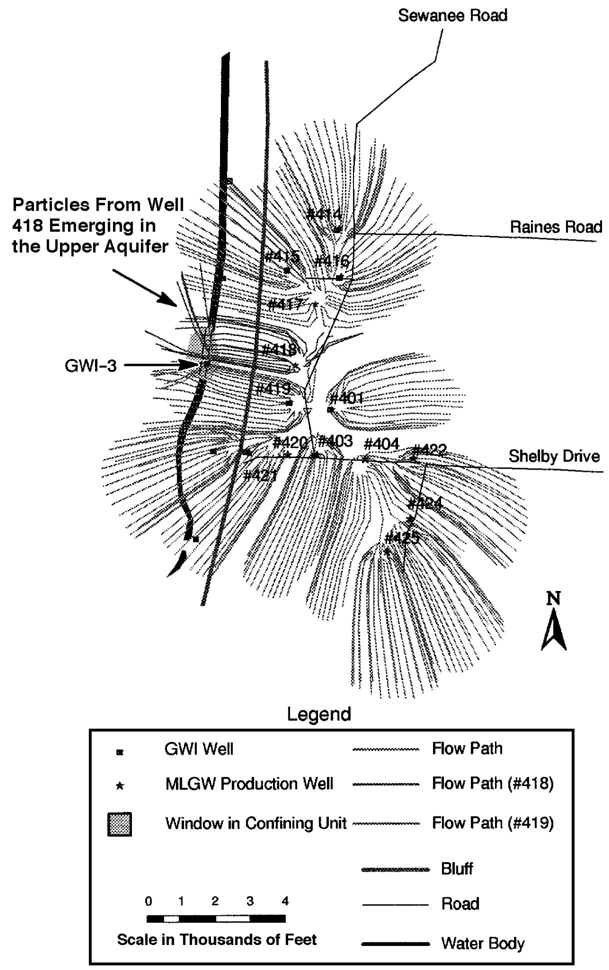


Figure 12. Backward tracking for 40 years.

project, especially: Dr. John W. Smith, Director, and Dr. Charles V. Camp for their patience and guidance in interpreting the model results and hydrogeologic conditions of the area; and David W. Kenley, Brian A. Waldron, and Robert B. Braun for their help in creating the well log database.

References

1. Camp, C.V., J.E. Outlaw, and M.C. Brown. 1994. GIS-based ground-water modeling. Microcomputers in Civil Eng. 9:281-293.

2. Driscoll, F.G. 1986. Groundwater and wells. St. Paul, MN: Johnson Division.

3. Camp, C.V., and M.C. Brown. 1993. A GIS procedure for developing a three-dimensional subsurface profile. J. Computing in Civil Eng. (July).

4. Camp, C.V., and J.E. Outlaw. 1993. Constructing subsurface profiles using GIS. Adv. in Eng. Software. 18:211-218.

5. McDonald, M., and A. Harbaugh. 1988. A modular three-dimensional finite-difference ground-water flow model. Open File Report 83-875. U.S. Geological Survey.

6. Palazolo, P.J., J.W. Smith, J.L. Anderson, and C.V. Camp. 1994. C.M. McCord wellhead protection demonstration grant. Final report. Herff College of Engineering Ground Water Institute, University of Memphis, Memphis, TN.

7. Richardson, G. 1989. A study of potential sources of leakage into the Memphis sand aquifer beneath the Davis well field in Memphis, Tennessee. M.S. thesis. Memphis State University, Memphis, TN.

8. Parks, W.S. 1990. Hydrology and preliminary assessment of the potential for contamination of the Memphis aquifer in the Memphis area, Tennessee. Water Resources Investigation Report 90-4092. U.S. Geological Survey.

9. Pollock, D. 1989. Documentation of computer programs to compute and display pathlines using results from the U.S. Geological Survey modular three-dimensional finite difference ground-water flow model. Open File Report 89-381. U.S. Geological Survey.

MODRISI: A PC Approach to GIS and Ground-Water Modeling

Randall R. Ross
Robert S. Kerr Environmental Research Laboratory, U.S. Environmental Protection Agency, Ada, Oklahoma

Milovan S. Beljin
University of Cincinnati, Cincinnati, Ohio

Abstract

It is widely accepted that ground-water contamination problems cannot be adequately defined or addressed until the governing physical, chemical, and biological processes affecting the transport and fate of contaminants are adequately characterized. Recent research has led to a better understanding of these complex processes and their effect on the movement of contaminants in the subsurface. The compilation and application of such information has yet to be accomplished at many hazardous waste sites, however. Too often, copious quantities of data are collected, only to be stored, ignored, or misplaced, rather than used for problem-solving. Geographic information systems (GIS) are computer-based tools that are relatively new to many environmental professionals. GIS allows the manipulation, analysis, interpretation, and visualization of spatially related data (e.g., hydraulic head, ground-water velocity, and contaminant concentration). GIS is more than a cartographic utility program, however. The analytical capabilities of GIS allow users to display, overlay, merge, and identify spatial data, thereby providing the basis for effective environmental decision-making.

IDRISI is a widely used PC-based raster GIS system that provides numerous analytical capabilities that are directly applicable to hydrogeologic studies. Raster systems are particularly well suited for analysis of continuous data such as elevation (e.g., water table, land and bedrock surfaces), precipitation, recharge, or contaminant concentrations and may be readily integrated with finite-difference ground-water models. Because the formats for IDRISI and ground-water model input data sets are different, a need exists for a program to integrate these two types of robust tools.

MODRISI is a collection of utility programs that allows easy manipulation and transfer of data files between IDRISI and ground-water models (e.g., MODFLOW, ASM, MOC). In addition, MODRISI integrates other widely used commercial and private domain software packages, such as SURFER, Geopack, GeoEas, AutoCAD, CorelDraw, and various spreadsheet programs. Two-dimensional arrays of models' input data sets can easily be created from IDRISI image files. AutoCAD vector files obtained by digitizing model boundaries, well locations, rivers and streams, or U.S. Geological Survey digital elevation model (DEM) files can also be translated into model input file formats. MODRISI can process model output files and prepare GIS image files that can be displayed and manipulated within IDRISI. Thus, MODRISI is more than a pre- and postprocessor for ground-water models; it is a complete GIS/ground-water modeling interface that is accessible to most ground-water hydrologists.

Introduction

Hydrogeologists collect and analyze large volumes of data during a ground-water modeling process. These data are stored and presented in many different forms such as maps, graphs, tables, computer databases, or spreadsheets. To most hydrogeologists, geographic information systems (GIS) are relatively new tools. They have been developed and applied in other natural and social science fields for over two decades, however, and can also be used in the ground-water modeling process.

GIS represents a new, powerful set of tools that can significantly improve the usefulness of results obtained during the ground-water modeling process. Bridging the disciplines of ground-water modeling, computer graphics, cartography, and data management, GIS represents a computer-based set of tools to display and analyze spatial data (e.g., water level elevations, ground-water quality data, modeling results, ground-water pollution potential). Efficient use of increasingly large volumes of data can be achieved only with powerful systems capable of acquiring information from a variety of sources, scales, and resolutions.

GIS can be defined as a computer-assisted system for the efficient acquisition, storage, retrieval, analysis, and representation of spatial data. Most GIS platforms consist of numerous subsystems that perform the listed tasks. The subsystems have the ability to query spatially related information and incorporate statistical analyses and modeling of relations and their temporal changes within the database. More than just a mapping system, GIS allows the user to analyze spatially related data and visualize results in either paper map form or graphically on screen. The data to be analyzed are a collection of spatial information represented by points, lines, and polygons and their associated attributes (characteristics of the features such as elevation or concentration). The cartographic tools of GIS allow the analyst to display, overlay, measure, merge, and identify the data to support a particular analysis. By allowing spatial data analysis and display, GIS provides the means necessary for effective environmental decision-making and implementation of environmental management plans.

GIS uses two basic map representation techniques: vector and raster. Vector representations describe features with a number of connected points. Raster representations subdivide a study area into a mesh of grid cells, each cell containing either a quantitative attribute value or feature identifier. Raster systems are well suited for analysis of continuous data (e.g., water level elevations, infiltration and recharge rates). This makes raster-based systems ideal for integration with ground-water models that use regularly spaced nodes. The objective of this paper is to illustrate such an integration of MODFLOW (1), a widely used U.S. Geological Survey (USGS) finite-difference ground-water flow model, and IDRISI (2), a raster-based GIS.

Previous Studies

To enhance understanding of a hydrogeologic system, and also to develop a credible ground-water model of the system, hydrogeologic features such as lithological logs, recharge and withdraw rates, estimates of spatial distribution of hydraulic conductivity, or specific storage can be plotted using GIS capabilities of data retrieval and overlay options to interactively define an area of interest (3-6). Two previous studies that combined GIS and ground-water modeling are briefly described.

Torak and McFadden (7) used GIS to facilitate finite-element modeling of ground-water flow. Complex aquifer geometry and irregularly distributed aquifer-system characteristics that influence ground-water flow affect the design of the finite-element mesh. GIS systems represent the complex arrangement of nodes and elements and the distribution of aquifer properties to provide input to the flow model. Point-data coverages of pertinent aquifer characteristics are rated from a relational database and are displayed using GIS.

Contoured surfaces based on point-data coverages produce triangulated irregular networks (TINs) that are superposed on the finite-element mesh to delineate zones of elements having similar aquifer properties. Zone boundaries are identified using the contoured TIN surface and by manually determining where boundaries align with the element sides. The allocation of well pumping rates to nodes in the finite-element mesh is performed efficiently with GIS for model input. Well pumping rates are accumulated by element from the combined coverages of the pumping data and the mesh, and element data are distributed to the node points for input. GIS is also used to prepare data for model input and to assess the adequacy of the data prior to simulation.

Three-dimensional perspectives showing TIN coverages of aquifer-property data are used to analyze and interpret complexities within the flow system before zonation. Additionally, GIS is used to display computed hydraulic heads over the finite-element mesh to produce contour maps of the simulated potentiometric surface. Because the node points in the finite-element mesh are not arranged in an orthogonal fashion, such as a finite-difference grid, a map display of the computed values of hydraulic head at the nodes is prepared for efficient and accurate interpretation of simulation results.

Harris et al. (8) conducted the Remedial Investigation/Feasibility Study (RI/FS) of the San Gabriel basin. Vast amounts of hydrogeologic data have been gathered, and a comprehensive systematized GIS database has been developed. The identified hydrologic boundaries, recharge basins, stream locations, well locations, and contaminant distributions are some of the features considered in developing a base map. The GIS-generated base map has allowed development of a finite-element grid for the basin. For each finite element, the initial estimates of the hydraulic conductivity, specific yield, recharge rates, and other input parameters were provided.

Using simple interfacing programs, the retrieval GIS nodal and elemental data were converted to required formats for the input files of the Couple Fluid, Energy, and Solute Transport (CFEST) code. Simulated ground-water levels were compared with the GIS-generated potentiometric surfaces. In areas of wide variations between simulated and observed data, the zonal distribution of controlling parameters was reevaluated, analyzed, and updated. Data processing, development of input files for computerized analysis of ground-water flow, and analysis of simulation results with different alternative conceptualizations is time consuming and tedious. Efficient use of GIS and CFEST not only eased the burden of conducting multiple simulations but reduced the probability of errors as well as the amount of time and effort required for each simulation.

IDRISI

IDRISI is a grid-based geographic information and image processing system developed by the Graduate School of Geography at Clark University and supported by the United Nations Institute for Training and Research (UNITAR) and the United Nations Environment Programme Global Resource Information Database (UNEP/GRID) (9). IDRISI is a collection of over 100 program modules that are linked through a menu system. These programs are organized into several groups:

- The core modules provide data entry and database management capabilities.
- The geographic analysis modules provide tools for database analysis.
- The statistical analysis modules allow statistical characterization of images.
- The peripheral modules provide a series of utilities.

IDRISI and other raster-based systems divide data sets into map layers; each layer contains data for a single attribute. For the example of a ground-water model, these layers could correspond to the MODFLOW two-dimensional arrays (e.g., initial water levels, transmissivity distribution, IBOUND arrays, computed hydraulic heads). IDRISI provides many analytical tools that are useful in hydrogeologic studies.

Three of the most important categories of these tools are database query, map algebra, and context operator. A semihypothetical case described below illustrates the use of these analytical tools. IDRISI provides an extensive set of tools for image processing, geographic and statistical analysis, spatial decision support, time series analysis, data display, and import/export and conversion. In addition, as a set of independent program modules linked to a broad set of simple data structures, the system is designed such that researchers may readily integrate into the system their own modules, written in any programming language.

IDRISI uses three types of data files: image, vector, and attribute. Image files contain rasterized information relating to a spatial variable. Vector files contain the coordinates of points, lines, and polygonal features. An attribute file lists the identifiers of features and the associated attribute values. Values files can be extracted from the existing image files, or image files can be created from existing values files. The values files can be combined and stored in a dBASE format. Each image, vector, or attribute file has a corresponding documentation file that contains information about the data file (e.g., title, number of rows and columns).

MODRISI: MODFLOW/IDRISI Interface

MODRISI is a set of utility programs that allows the transfer of data files between MODFLOW, IDRISI, Golden Software SURFER, GeoEas, and other software. Preparation of two-dimensional arrays for the MODFLOW input files is generally tedious and time consuming. The arrays can be created easily from the IDRISI image files, however. Thus, MODRISI serves as a preprocessor for MODFLOW. For example, when the values of a variable are available only for irregularly spaced points, interpolation routines in SURFER or GeoEas may be used to estimate the values of the variable on regularly spaced grid-nodes. MODRISI translates SURFER or GeoEas files into IDRISI image files for manipulation, analysis, and display. IDRISI recognizes either latitude and longitude geodetic coordinates or arbitrary Cartesian plane coordinates. IDRISI assigns the lower left grid-block of a raster image as a zero-row, zero-column block.

Vector files, such as model boundaries, well locations, and rivers, may be created within IDRISI and translated into a MODFLOW input file format. For example, the location of a river may be digitized on screen in IDRISI. The vector-to-raster function may be invoked, assigning all blocks through which the river passes as river nodes. Similarly, the positions of wells may be digitized and translated into the row-column positions and saved as a MODFLOW input file for the well package. Once the MODFLOW input files are prepared, MODFLOW simulations may be initiated. The MODFLOW hydraulic head output files may be read by MODRISI and modified to create IDRISI image files. Again, the image files may be displayed and evaluated within IDRISI. Thus, MODRISI is used as a postprocessor for MODFLOW.

Case Study

The utility of MODRISI was demonstrated at a hazardous waste site. Previous investigations provided site-specific information, including water level and bedrock and land surface elevations, which was then analyzed using GeoEas, a public domain geostatistical software program. These data were kriged to produce a grid of regularly spaced data. These data were imported into MODRISI and converted to IDRISI image files. IDRISI was used to visualize the surfaces that GeoEas generated.

Several prominent features are obvious upon inspection of the kriged bedrock topography (see Figure 1). A bedrock ridge trending northwest to southeast is flanked by a minor trough to the east and a major trough to the southwest. The outline of the site is visible on all figures. The kriged water level elevation map (see Figure 2) illustrates the general hydraulic gradient to the west. Land surface elevations (see Figure 3) range from greater than 200 feet in the northeastern portion of the site to a low of 166 feet on the western boundary.

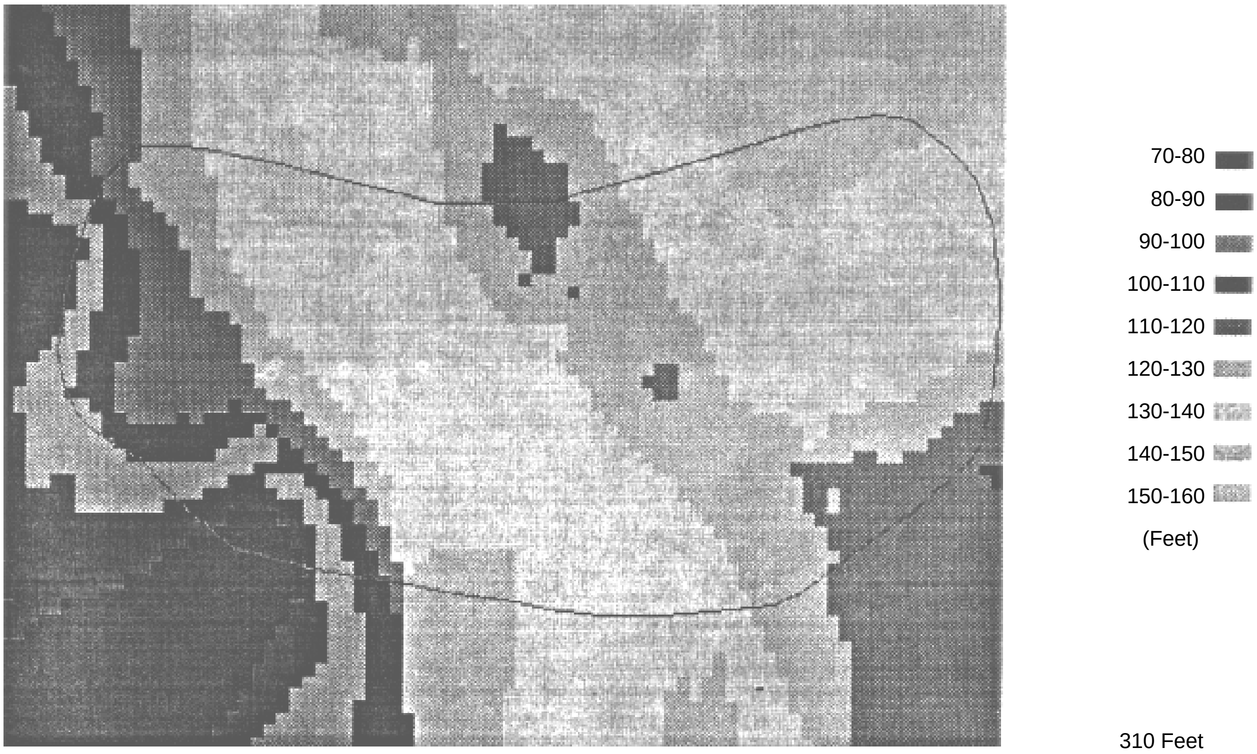


Figure 1. Bedrock elevation contour map derived from kriged data and transformed by MODRISI into IDRISI image file format.

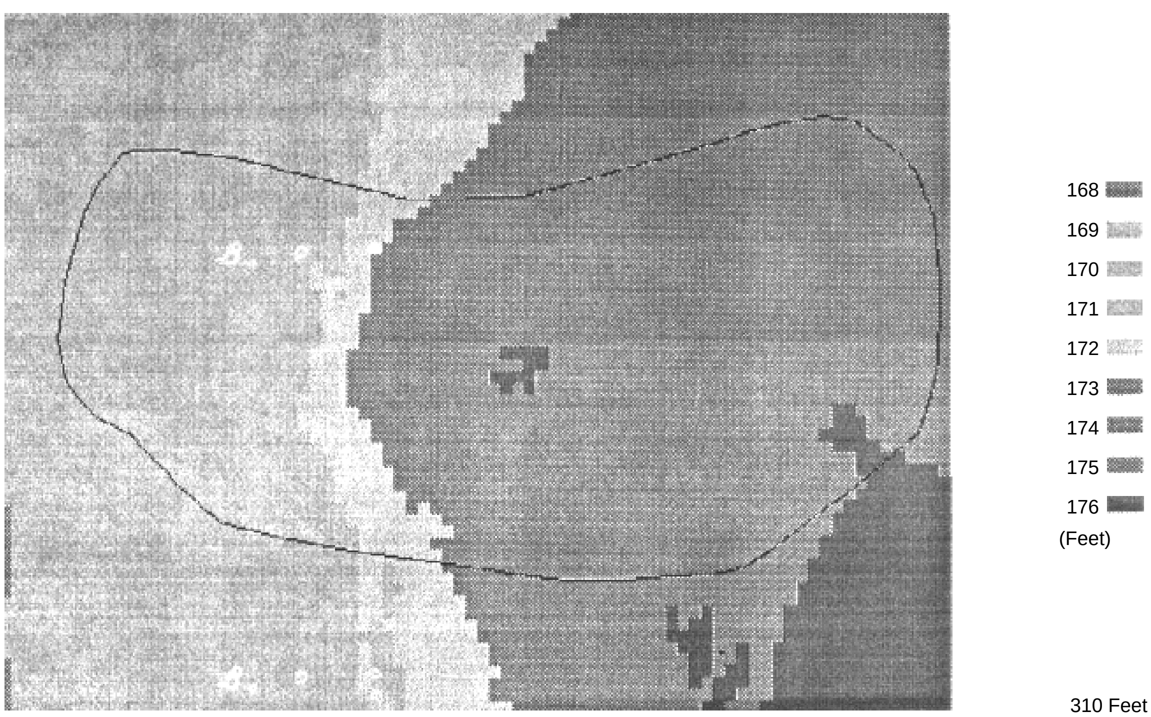


Figure 2. Water level contour map derived from kriged water level data and transformed by MODRISI into IDRISI image file format.

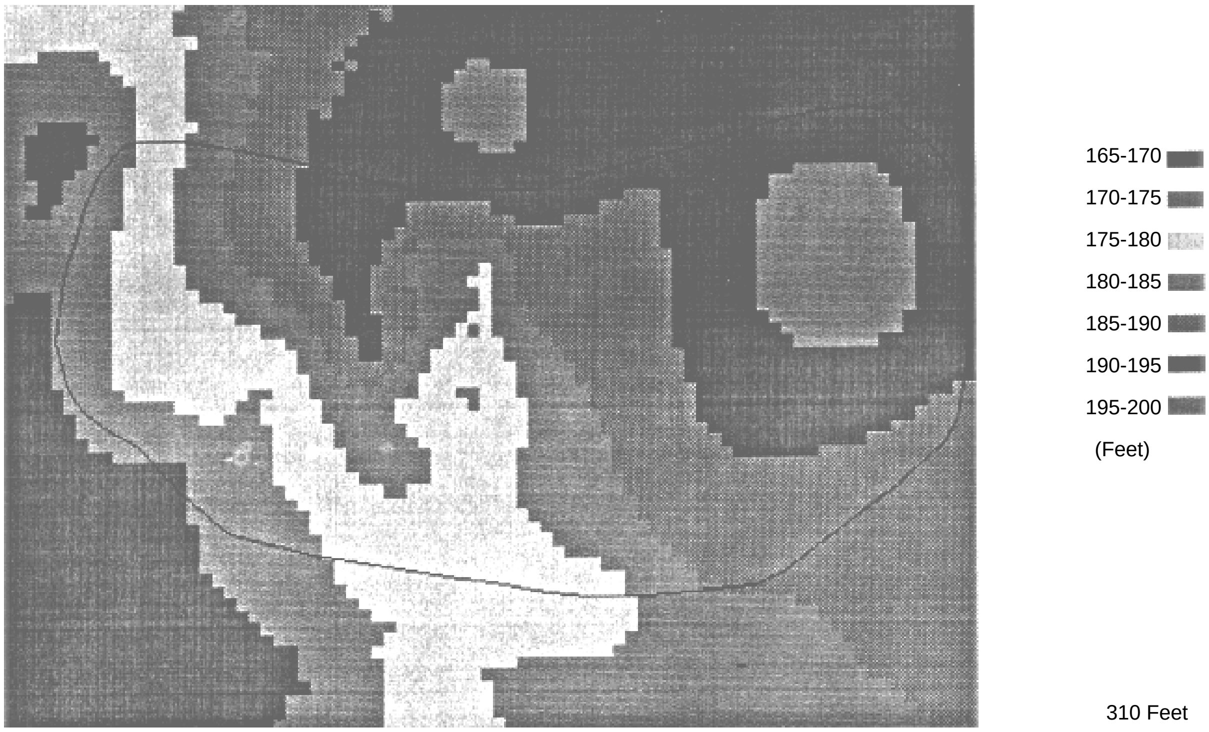


Figure 3. Land elevation contour map generated from kriged data and transformed by MODRISI into IDRISI image file format.

Additionally, the value (e.g., elevation) and (x,y) coordinates may be queried for any point, line, or area of an image file.

The analytical capabilities of IDRISI are illustrated in the following example. The saturated thickness was determined by subtracting the bedrock surface from the water level surface. The results illustrate the spatial variability of saturated thickness of the overburden aquifer (see Figure 4). The results correspond favorably with the general bedrock topography, as would be expected given the relatively low hydraulic gradient across the site. This OVERLAY (subtract) function may also readily be used to evaluate the adequacy of model calibration. For example, predicted values may be compared with observed values graphically, allowing the modeler to quickly visualize and identify areas of the model domain requiring additional consideration and manipulation.

Conclusions

One of the most tedious tasks in a ground-water modeling process is preparing input data and postprocessing model results. GIS allows rapid incorporation and evaluation of new site characterization information. The proposed combination of IDRISI, a raster-based GIS system, and MODRISI, a set of utility programs, could significantly reduce the amount of time necessary for entering data in required array formats. The visualization capabilities of IDRISI in conjunction with MODRISI and MODFLOW allow project managers to better understand the three-dimensional nature of subsurface environmental problems.

Acknowledgments

The authors wish to thank Chet Janowski, U.S. Environmental Protection Agency Region I, for his support on this project.

References

1. McDonald, M., and A. Harbough. 1988. A modular three-dimensional finite-difference ground-water flow model: USGS techniques of water resources, Book 6, Chapter A1.
2. Eastman, J.R. 1992. IDRISI: Technical reference. Clark University, Worcester, MA (March).
3. Van Metre, P.C. 1990. Structure and application of an interface program between a geographic-information system and a ground-water flow model. USGS Open File Report 90-165. Denver, CO: U.S. Geological Survey.
4. Hinaman, K.C. 1993. Use of a geographic information system to assemble input-data sets for a finite-difference model of ground-water flow. In: Proceedings of the AWRA Symposium on Geographic Information Systems and Water Resources, Mobile, AL, (March 14-17). pp. 405-412.
5. Rifai, H.S., L.A. Hendricks, K. Kilborn, and P.B. Bedient. 1993. A geographic information system (GIS) user interface for delineating wellhead protection areas. Ground Water 31(3):480-488.
6. Orzol, L.L., and T.S. McGrath. 1992. Modifications of the USGS modular three-dimensional finite-difference, ground-water flow model. U.S. Geologic Survey Open File Report 92-50.

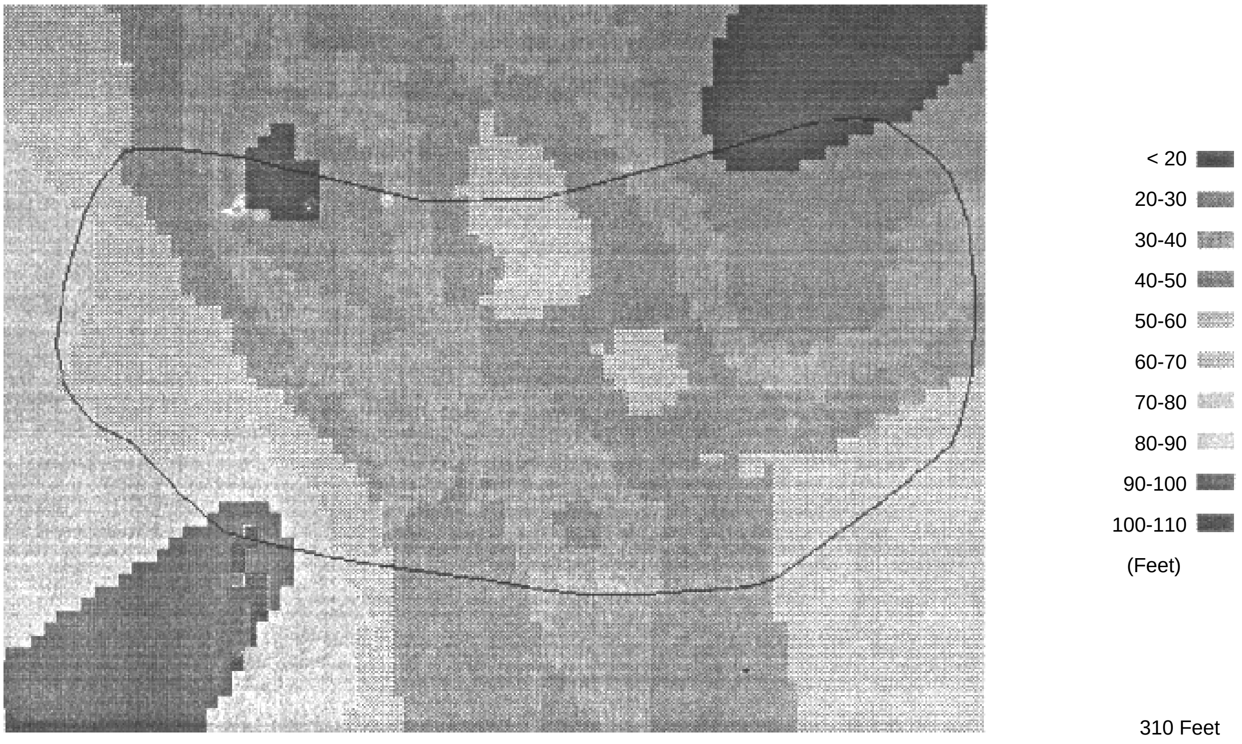


Figure 4. Saturated thickness contour map developed by subtracting bedrock image file from water level image file within IDRISI.

7. Torak, L.J., and K.W. McFadden. 1988. Application of a geographic information system to finite-element modeling of ground-water flow. In: Proceedings of the GIS Symposium Integrating Technology and Geoscience Applications, Denver, CO (September 26-30).

8. Harris, J., S. Gupta, G. Woodside, and N. Ziemba. 1993. Integrated use of GIS and three-dimensional, finite-element model: San Gabriel basin ground-water flow analyses. In: Goodchild, M., B. Parks, and L. Steyaert, eds. Environmental modeling with GIS. New York, NY: Oxford University Press.

9. Eastman, J.R., P.A. Kyem, J. Tolendano, and W. Jin. 1993. GIS and decision making. Explorations in geographic information systems technology, Vol. 4. United Nations Institute for Training and Research, Geneva, Switzerland.

GIS in Statewide Ground-Water Vulnerability Evaluation to Pollution Potential

Navulur Kumar and Bernard A. Engel
Department of Agricultural Engineering, Purdue University, West Lafayette, Indiana

Abstract

The ground-water vulnerability of Indiana to pollution potential was evaluated using a geographic information systems (GIS) environment. The Geographic Resources Analysis Support System (GRASS) and the GRID submodule of ARC/INFO were used to conduct the analysis and to identify and display the areas sensitive to ground-water pollution potential. The state soils geographic (STATSGO) database was employed to retrieve statewide soils information required for the analysis. The information from the STATSGO database was used in two models, DRASTIC (acronym representing the following hydrogeologic settings: **D**epth to water table, aquifer **R**echarge, **A**quifer media, **S**oil media, **T**opography, **I**mpact of vadose zone, and hydraulic **C**onductivity of the aquifer) and SEEPAGE (System for Early Evaluation of Pollution Potential of Agriculture Ground-Water Environments). These models employ a numerical ranking system and consider various hydrogeologic settings that affect the ground-water quality of a region. Ground-water vulnerability maps were prepared for the state of Indiana based on DRASTIC and SEEPAGE results. Continuing work is planned to determine the accuracy of the results by comparing the existing well-water quality data. The DRASTIC Index and SEEPAGE Index number (SIN) maps show great potential as screening tools for policy decision-making in ground-water management.

Introduction

Ground-water contamination due to fertilizer and pesticide use in agricultural management systems is of wide concern. In 1989, reports of ground-water contamination in New York wells led the U.S. Environmental Protection Agency (EPA) to conduct a nationwide survey on well contamination in the United States. These wells were tested for presence of nitrate, pesticides, and pesticide breakdown products (1). Statistically, the wells selected represent more than 94,600 wells in approximately 38,300 community water systems. Over 52 percent of the community water systems and 57 percent of the rural domestic wells tested contained nitrates (2).

Indiana has abundant ground-water systems providing drinking water for 60 percent of its population. A study on well-water quality detected pesticides in 4 percent of wells tested in Indiana. Also, 10 percent of private wells and 2 percent of noncommunity wells contained excessive nitrate levels (3).

Statewide maps showing the areas vulnerable to ground-water contamination have many potential uses such as implementation of ground-water management strategies to prevent degradation of ground-water quality and monitoring of ground-water systems. These maps will be helpful in evaluating the existing and potential policies for ground-water protection. Ground-water models such as SEEPAGE (System for Early Evaluation of Pollution Potential of Agriculture Ground-Water Environments) and DRASTIC (acronym representing the following hydrogeologic settings: **D**epth to water table, aquifer **R**echarge, **A**quifer media, **S**oil media, **T**opography, **I**mpact of vadose zone, and hydraulic **C**onductivity of the aquifer) can be applied on a regional scale to develop such maps.

The data layers required for these models are commonly available data such as pH and organic matter content. For most states, the statewide ground-water vulnerability maps generated using DRASTIC were produced from 1:2,000,000-scale data (4). EPA (2) found that these maps did not correlate well with the water quality analysis performed for the national survey of pesticides in drinking water wells. States need more detailed and accurate maps to implement ground-water management programs. The state soils geographic (STATSGO) database at the 1:250,000-scale might be useful for studies at a larger scale.

The geographic information systems (GIS) environment is widely applied for diverse applications in resources management and other areas. It offers the facilities to store, manipulate, and analyze data in different formats and at different scales. The DRASTIC and SEEPAGE models can be integrated within the GIS environment to produce the final ground-water vulnerability maps.

Objectives

The purpose of the study was to prepare maps showing areas in Indiana vulnerable to ground-water pollution. This goal was accomplished by considering hydrogeologic factors in each region that affect the mobility and leaching of the contaminant reaching the aquifer. The prime objectives of this research were to:

- Evaluate Indiana's ground-water vulnerability to pollution potential using the DRASTIC and SEEPAGE models:
  - Integrate and evaluate the models in a GIS environment (Geographic Resources Analysis Support System [GRASS] ARC/INFO).
  - Develop a graphic user interface (GUI) in ARC/INFO to conduct the analyses.
- Compare the pollution potential map from the DRASTIC model with the map developed using the SEEPAGE Index number (SIN).
- Validate the accuracy of the present approach by comparing the vulnerability maps with the existing well-water quality data sampled across the state.

DRASTIC

DRASTIC is a ground-water quality model for evaluating the pollution potential of large areas using the hydrogeologic settings of the region (4-6). EPA developed this model in the 1980s. DRASTIC includes different hydrogeologic settings that influence a region's pollution potential. A hydrogeologic setting is a mappable unit with common hydrogeologic characteristics. This model employs a numerical ranking system that assigns relative weights to parameters that help evaluate relative ground-water vulnerability to contamination.

The hydrogeologic settings that make up the acronym DRASTIC are:

- *[D] Depth to water table:* Compared with deep water tables, shallow water tables pose a greater chance for the contaminant to reach the ground-water surface.
- *[R] Recharge (net):* Net recharge is the amount of water per unit area of soil that percolates to the aquifer. This is the principal vehicle that transports the contaminant to the ground water. Higher recharges increase the chances of the contaminant being transported to the ground-water table.
- *[A] Aquifer media:* The material of the aquifer determines the mobility of the contaminant traveling through it. An increase in travel time of the pollutant through the aquifer increases contaminant attenuation.
- *[S] Soil media:* Soil media is the uppermost portion of the unsaturated zone/vadose zone characterized by significant biologic activity. This, in addition to the aquifer media, determines the amount of water percolating to the ground-water surface. Soils with clays and silts have larger water holding capacity and thus increase the travel time of the contaminant through the root zone.
- *[T] Topography (slope):* The higher the slope, the lower the pollution potential due to higher runoff and erosion rates, which include pollutants that infiltrate the soil.
- *[I] Impact of vadose zone:* The unsaturated zone above the water table is referred to as the vadose zone. The texture of the vadose zone determines the travel time of the contaminant. Authors of this model suggest using the layer that most restricts water flow.
- *[C] Conductivity (hydraulic):* Hydraulic conductivity of the soil media determines the amount of water percolating through the aquifer to the ground water. For highly permeable soils, the travel time of the pollutant is decreased within the aquifer.

The major assumptions outlined in DRASTIC are:

- The contaminant is introduced at the surface.
- The contaminant reaches ground water by precipitation.
- The contaminant has the mobility of water.
- The area of the study site is more than 100 acres.

DRASTIC evaluates pollution potential based on the seven hydrogeologic settings listed above. Each factor is assigned a weight based on its relative significance in affecting the pollution potential. Each factor is also assigned a rating for different ranges of the values. Typical ratings range from 1 to 10, and weights range from 1 to 5. The DRASTIC Index, a measure of pollution potential, is computed by summation of the products of rating and weights of each factor as follows:

DRASTIC Index =
$D_rD_w + R_rR_w + A_rA_w + S_rS_w + T_rT_w + I_rI_w + C_rC_w$

where:

D_r = Ratings for the depth to water table
D_w = Weights for the depth to water table
R_r = Ratings for different ranges of aquifer recharge
R_w = Weights for the aquifer recharge
A_r = Ratings for the aquifer media
A_w = Weights for the aquifer media
S_r = Ratings for soil media
S_w = Weights for soil media
T_r = Ratings for topography (slope)
T_w = Weights for topography
I_r = Ratings for the vadose zone

Iw = Weights for the vadose zone
Cr = Ratings for different rates of hydraulic conductivity
Cw = Weights for hydraulic conductivity

DRASTIC assigns two different weights depending upon the type of contaminant. Pesticides are given different weights than general contaminants. In assigning the weights, DRASTIC considers the different properties of pesticides as they travel through the vadose zone and root zone of the soil media.

The higher the DRASTIC Index, the greater the relative pollution potential. The DRASTIC Index is divided into four categories: low, moderate, high, and very high. The sites with high and very high categories are more vulnerable to contaminations and hence should be reviewed by the site specialist. These weights are relative, however. Low pollution potential does not necessarily indicate that a site is free from ground-water contamination. It indicates only that the site is less susceptible to contamination than sites with high or very high DRASTIC ratings.

SEEPAGE

The SEEPAGE model is a combination of three models adapted to meet the Soil Conservation Service's (SCS's) need to assist field personnel (7, 8). SEEPAGE considers hydrogeologic settings and physical properties of the soil that affect ground-water vulnerability to pollution potential. SEEPAGE is also a numerical ranking model that considers contamination from both concentrated and dispersed sources.

The SEEPAGE model considers the following parameters:

- Soil slope
- Depth to water table
- Vadose zone material
- Aquifer material
- Soil depth
- Attenuation potential

The attenuation potential further considers the following factors:

- Texture of surface soil
- Texture of subsoil
- Surface layer pH
- Organic matter content of the surface
- Soil drainage class
- Soil permeability (least permeable layer)

Each factor is assigned a numerical weight ranging from 1 to 50 based on its relative significance, with the parameter that has the most significant effect on water quality assigned a weight of 50 and the least significant assigned a weight of 1. The weights are different for concentrated or site-specific sources, and dispersed or nonspecific sources.

Similar to DRASTIC, each factor can be divided into ranges and ratings, varying from 1 to 50. The ratings of the aquifer media and vadose zone are subjective and can be changed for a particular region. Once the scores of the six factors are obtained, they are summed to obtain the SIN. These values represent pollution potential, where a high SIN implies relatively more vulnerability of the ground-water system to contamination. The SIN values are arranged into four categories of pollution potential: low, moderate, high, and very high. A high or very high SIN category indicates that the site has significant constraints for ground-water quality management (7).

GIS

GIS has been widely used for natural resources management and planning, primarily during the past decade. A GIS can be combined with a ground-water quality model to identify and rank the areas vulnerable to pollution potential for different scenarios and land use practices. Many GIS software packages are available. GRASS is a raster-based public domain software developed by the U.S. Army Construction Engineers Research Laboratory (9). This software can assign different weights to, or reclass, the data layers and combine map layers, and is suitable for implementing the DRASTIC and SEEPAGE models. ARC/INFO is a GIS software developed by Environmental Systems Research Institute (ESRI) in Redlands, California. The GRID submodule of ARC/INFO facilitates the handling of raster data. Also, the capability to develop a menu-based GUI helps users easily implement the models. The GRID submodule also can reclass and manipulate the map layers suitable for conducting the analyses.

Methodology

Developing the Data Layers in GRASS and ARC/INFO

The STATSGO database from SCS comes at a scale of 1:250,000 and is distributed in different data formats. This study used the STATSGO database in the ARC/INFO format. The database is organized into map units that have up to 21 components. These map components have information assigned to layers of soil horizons. Each layer is attributed various soil properties such as pH or organic matter content (10). Each property is assigned a high and a low value for a map unit. The STATSGO map for Indiana is available in the vector format. This map was exported to GRASS as a vector coverage (11) and was converted into a raster coverage within the GRASS GIS environment. This was used as the base map for the DRASTIC and SEEPAGE analyses. The hydrogeologic parameters required for the models were identified from the corresponding INFO data tables and were exported into an ASCII file. Code was developed to generate a GRASS reclass file assigning the weighted values of the parameters to the corresponding map units in the base map. The STATSGO base map imported into GRASS was reclassed for each hydrogeologic setting (e.g., topography, pH) to create the data layers required for DRASTIC and SEEPAGE analyses.

The map layers of the hydrogeologic parameters in GRASS were then exported to ARC/INFO as raster coverages. A GRASS command was developed that allows the output ASCII file from GRASS to be imported into ARC/INFO directly without further modifications to the header in the ASCII file.

Developing a Graphic User Interface in ARC/INFO

The dynamic form-menu option (12) was used to develop a GUI for both DRASTIC and SEEPAGE analyses (see Figures 1 and 2). Because ratings for some parameters are subjective, the GUI provided an option to change the weights assigned to hydrogeologic settings. The coverages must already be assigned ratings before using the interface, however. The interface also allows users to reclassify the final vulnerability maps qualitatively (13) into four categories (low, moderate, high, and very high) after viewing the range of DRASTIC Index or SIN values.

Conducting the Analyses

The data layers were developed separately for the high and low values of the hydrogeologic settings. Once all the data layers were compiled, the corresponding ratings and weights were assigned and the analyses were conducted using the GUI. The data layers aquifer recharge, aquifer media, and vadose zone media were not available, so the analyses were conducted without these base maps. The SEEPAGE analysis was performed for concentrated/point sources of pollution. The final vulnerability indexes from the analyses were classified into four categories (low, medium, high, and very high) (see Table 1) to generate the final statewide vulnerability maps.

Table 1. Pollution Potential Categories Using SEEPAGE and DRASTIC Indexes

	Range of DRASTIC/SEEPAGE Index			
Analysis	**Low**	**Moderate**	**High**	**Very High**
SEEPAGE	1-24	25-48	49-70	> 70
DRASTIC	30-70	71-100	101-110	> 110

Validating the Accuracy of the Vulnerability Maps

The ground-water vulnerability maps produced by DRASTIC analysis were compared with those generated using the SEEPAGE model. The final statewide ground-water vulnerability maps from either approach were compared with the well-water quality data sampled from over 2,500 wells (see Figure 3), and the number of wells falling into each vulnerability category was tabulated.

Results and Discussion

Statewide analysis of ground-water vulnerability to pollution potential was conducted using the DRASTIC and SEEPAGE analyses at a scale of 1:250,000 in the raster format. The analyses were conducted for both the high and low values of the hydrogeologic settings, and the final vulnerability maps were prepared for the state of Indiana (see Figures 4 and 5). The vulnerability maps from both approaches were compared in the GRASS environment.

In both analyses, the low values of hydrogeologic settings resulted in more areas being classified as high and very high categories, compared with the high values of hydrogeologic settings in a map unit. The DRASTIC analysis placed more areas in the very high vulnerability category, compared with the SEEPAGE analysis, which categorized the same areas as high vulnerability. The nitrate-nitrogen concentrations observed in the wells were compared with the final vulnerability maps, and the number of wells falling into each of the four vulnerability categories (low, moderate, high, and very high) were summarized (see Tables 2 and 3).

Approximately 80 percent of the wells with concentrations less than 5 parts per million are classified under the moderate vulnerability category in SEEPAGE analysis. Overall, the results from the analyses did not correlate satisfactorily with the observed well-water quality data. Unavailability of the data layers aquifer media, aquifer recharge, and vadose zone media may have caused these results. The well-water quality data of nitrate-nitrogen contaminations was considered only for testing map accuracy, whereas the analyses do not account for the type of contaminant, its severity, and its volume in the generation of vulnerability maps. Other limitations of these approaches, including that the factors influencing aquifer contamination (e.g., direction of water flow, land use, population at risk, point sources of pollution) are not considered in the ground-water vulnerability evaluation, might also have led to the observed results.

The DRASTIC and SEEPAGE analyses can be improved by incorporating data layers such as land use and nitrate loadings in computing the DRASTIC and SEEPAGE Indexes. The STATSGO database can be used for

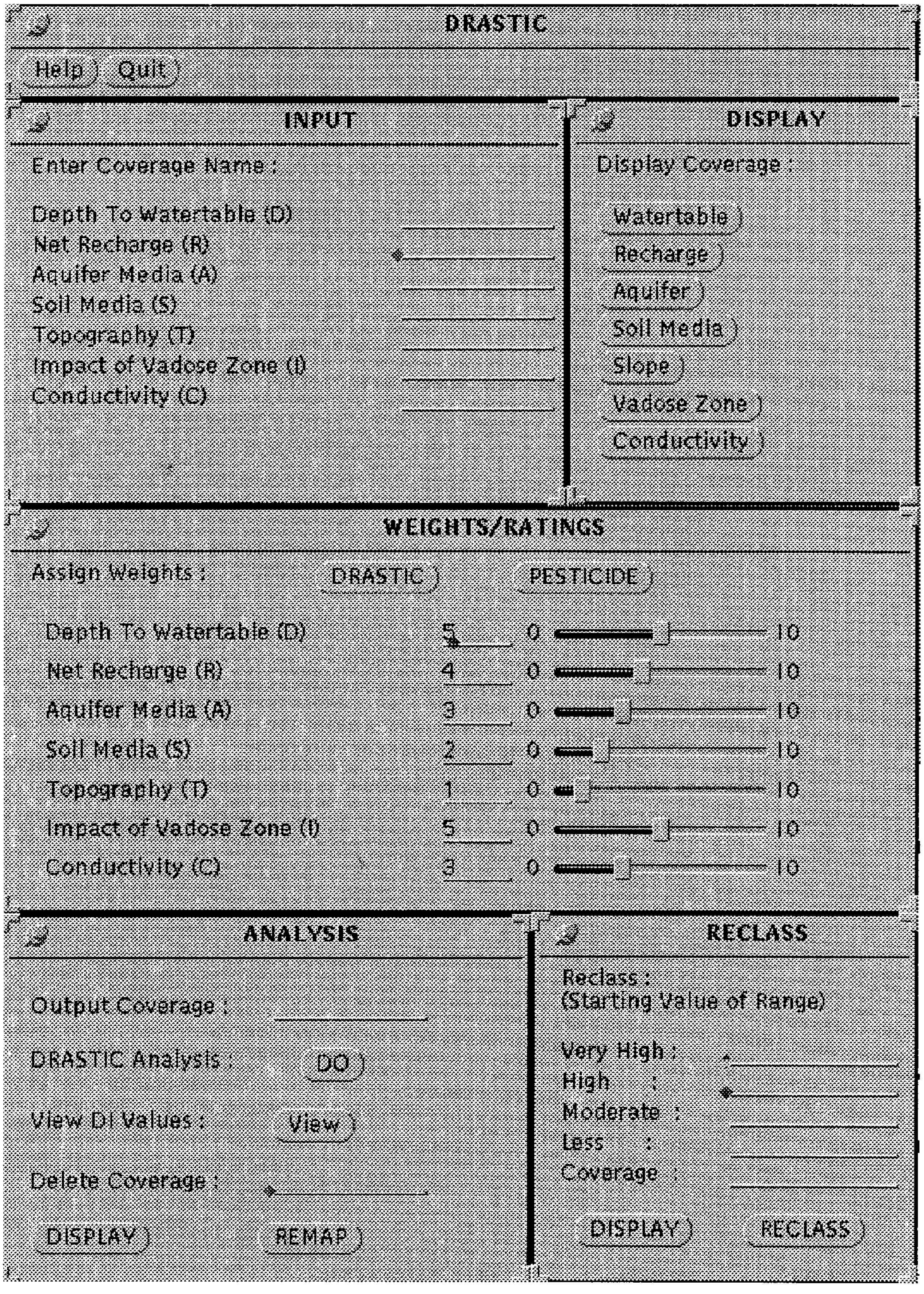


Figure 1. GUI for DRASTIC analysis.

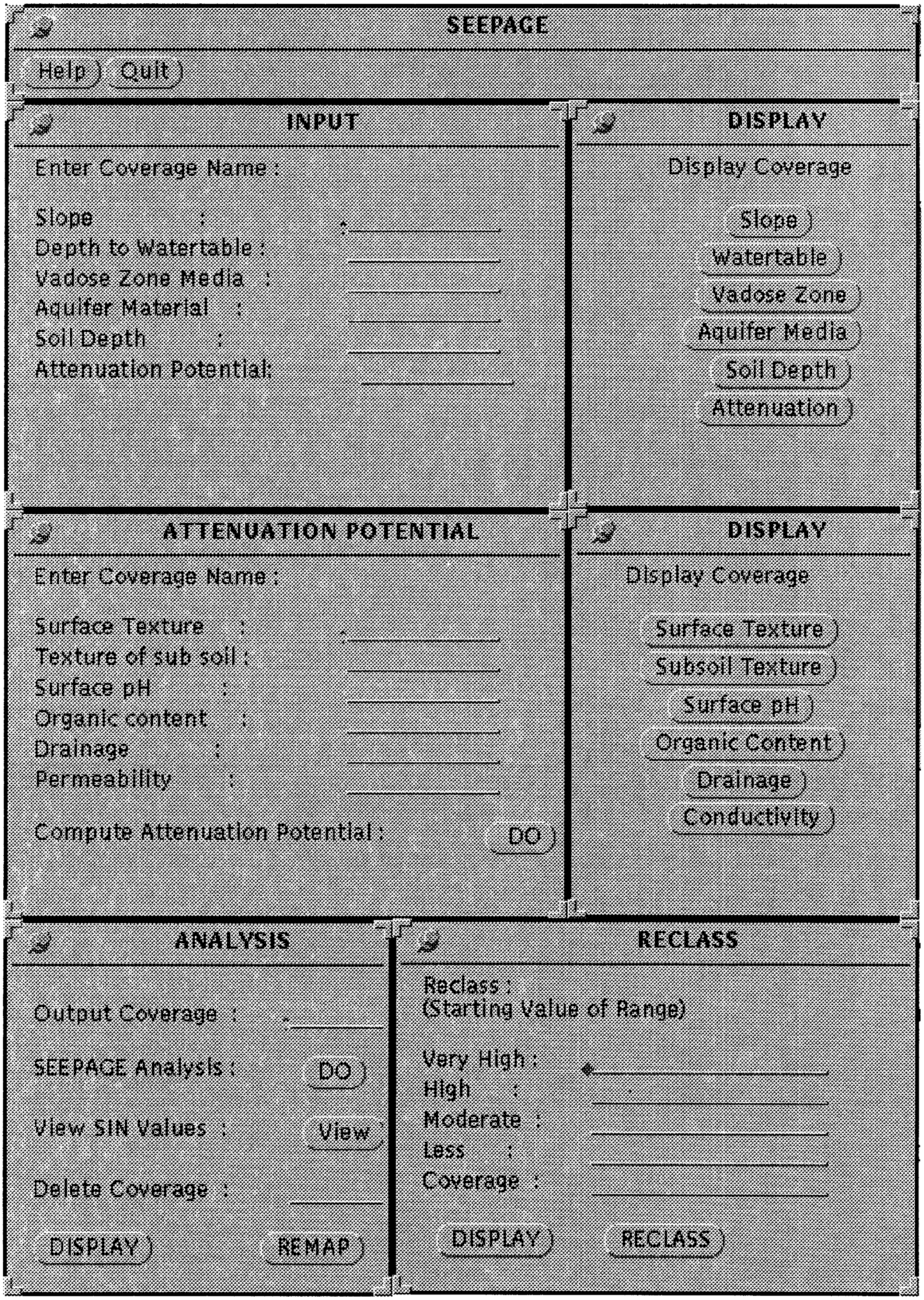


Figure 2. GUI for SEEPAGE analysis.

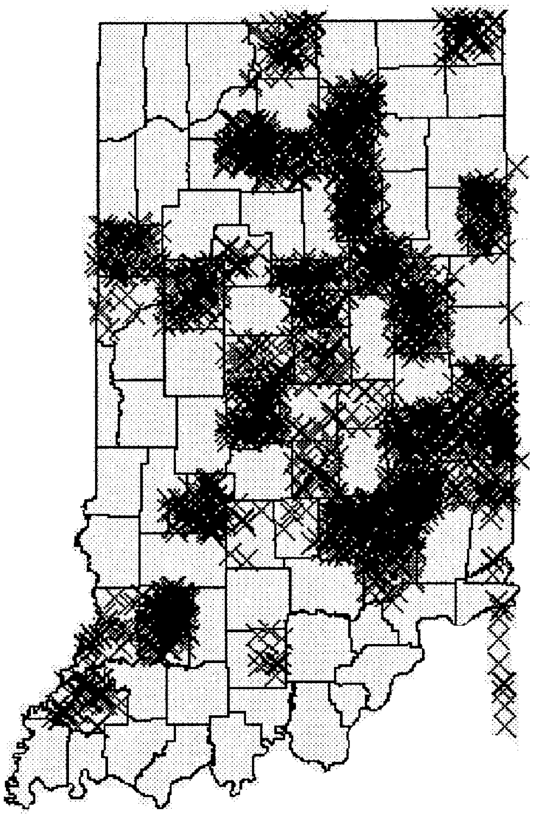


Figure 3. Sampling sites (wells) for water quality data.

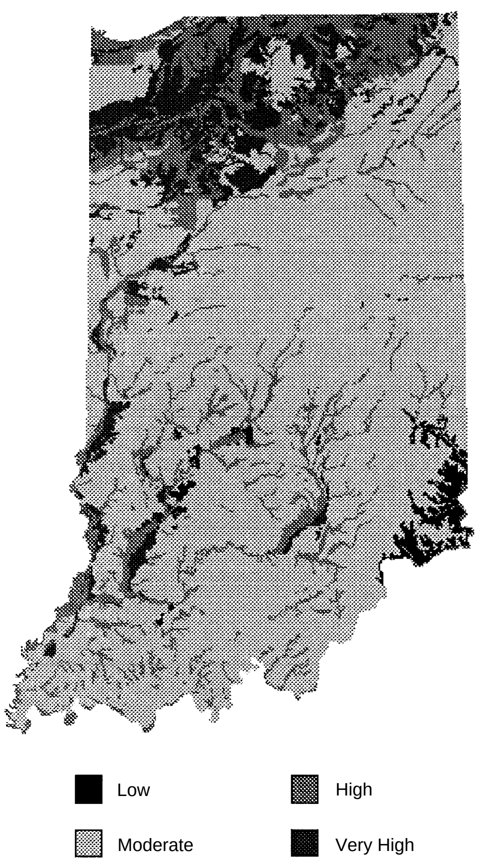


Figure 4. Ground-water vulnerability map using DRASTIC analysis.

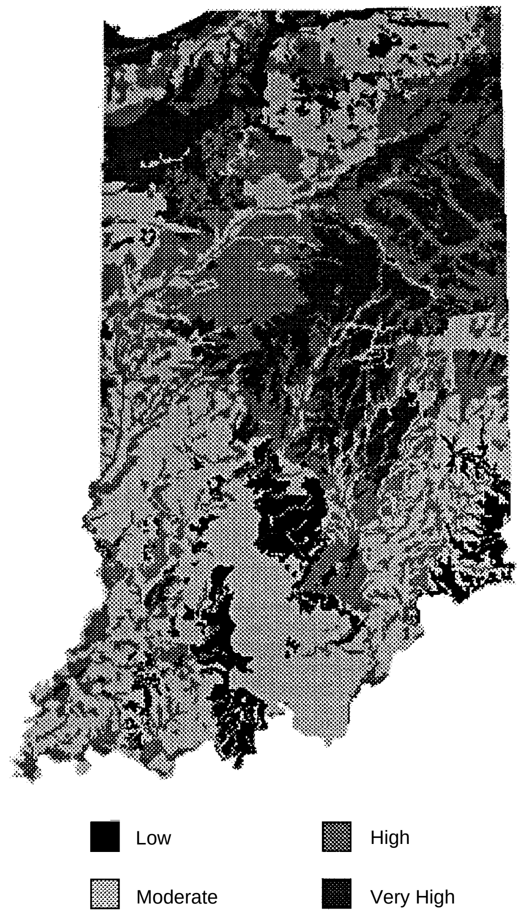


Figure 5. Ground-water vulnerability map using SEEPAGE analysis.

Table 2. Comparison of SEEPAGE Results With Observed Nitrate-Nitrogen Concentrations in Wells

	Nitrate-Nitrogen Levels		
Category	**0-5**	**5-10**	**> 10**
Low	7	2	8
Moderate	1,322	76	384
High	249	11	1
Very high	194	2	64

Table 3. Comparison of DRASTIC Results With Observed Nitrate-Nitrogen Concentrations in Wells

	Nitrate-Nitrogen Levels		
Category	**0-5**	**5-10**	**> 10**
Low	11	2	9
Moderate	541	41	290
High	720	39	213
Very high	500	9	65

developing most of the data layers required for the analyses. GIS is a useful tool for integrating ground-water quality models and facilitates testing the models for different scenarios. The GUI helps users easily conduct analyses and facilitates changing the weights for subjective hydrogeologic settings. DRASTIC and SEEPAGE approaches show great potential as screening tools for policy decision-making in ground-water management.

Summary

Ground-water pollution from agricultural management systems is of wide concern. Few models address ground-water vulnerability on a regional scale. The DRASTIC and SEEPAGE models are numerical ranking models that consider various hydrogeologic settings affecting the contamination of a region. The data required for these models are commonly available data, and the STATSGO database at 1:250,000-scale was used in this study. These models were integrated in the GIS environment of GRASS and ARC/INFO in the raster format. A menu-based GUI was developed in ARC/INFO for conducting the analyses. The vulnerability maps generated from DRASTIC and SEEPAGE analyses were compared. The statewide vulnerability maps also were compared with the well-water quality data to validate the accuracy of the models.

References

1. U.S. EPA. 1990. National pesticide survey: Phase I report. Washington, DC.
2. U.S. EPA. 1992. Another look: National survey of pesticides in drinking water wells, Phase II report. EPA/579/09-91/020. Washington, DC.
3. Department of Environmental Management, Groundwater Section. 1989. Indiana ground-water protection: A guide book (June).
4. U.S. EPA. 1987. DRASTIC: A standardized system for evaluating ground-water pollution potential using hydrogeologic settings. EPA/600/2-87/035. Washington, DC.
5. U.S. EPA. 1985. DRASTIC: A standardized system for evaluating ground-water pollution potential using hydrogeologic settings. EPA/600/2-85/0108. U.S. EPA, Robert S. Kerr Environmental Research Laboratory, Ada, OK.
6. Deichert, L.A., and J.M. Hamlet. 1992. Nonpoint ground-water pollution potential in Pennsylvania. ASAE Paper No. 922531. American Society of Agricultural Engineers International Winter Meeting, Nashville, TN.
7. Richert, S.E., S.E. Young, and C. Johnson. 1992. SEEPAGE: A GIS model for ground-water pollution potential. ASAE Paper No. 922592. American Society of Agricultural Engineers International Winter Meeting, Nashville, TN.
8. Engel, B.A., and D.D. Jones. 1992. Technique for developing ground-water vulnerability to nitrate maps for large areas: Research proposal 12-1992.
9. U.S. CERL. 1990. GRASS: Geographical Resources Analysis Supporting System user's manual. Champaign, IL: U.S. Army Construction Engineers Research Laboratory.
10. SCS. 1992. State Soils Geographic Database (STATSGO) user's guide. Soil Conservation Service Publication No. 1492.
11. ESRI. 1992. ARC/INFO command references: Arc command references. Redlands, CA: Environmental Systems Research Institute.
12. ESRI. 1992. ARC/INFO user's guide: AML user's guide. Redlands, CA: Environmental Systems Research Institute.
13. ESRI. 1992. ARC/INFO user's guide: Cell-based modeling with GRID. Redlands, CA: Environmental Systems Research Institute.

Verification of Contaminant Flow Estimation With GIS and Aerial Photography

Thomas M. Williams
Clemson University, Georgetown, South Carolina

Abstract

Estimation of contaminant movement in ground water requires interpolation of data from sampling wells that represent a very small sample of aquifer volume. Spatial statistics and kriging provide the best unbiased estimator of interpolated concentrations. Hurricane Hugo provided an opportunity to compare these estimators with actual forest mortality caused by saltwater inundation associated with the tidal surge. During the 9- to 15- month period after the hurricane, salt from the tidal surge moved within the shallow water table aquifer, causing widespread tree mortality on Hobcaw Forest in eastern Georgetown County, South Carolina. A small watershed (12 acres) was instrumented with 24 multilevel sampling wells. Piezometric potential and samples for salt concentration were collected for 12 months (months 18 to 30 after the tidal surge). These data produced three-dimensional estimations of flow directions and two-dimensional maps of chloride concentration. These maps led to the identification of important heterogeneities in the water table aquifer. Apparently, the infiltrated salt water moved to the bottom of the aquifer (15 feet) and emerged, killing the forest, where aquifer heterogeneity resulted in upward movements of ground water.

Georgetown County implemented a geographic information system (GIS) for tax mapping in 1988 and prepared 1:400-scale orthophotographs of the entire county with true ground accuracy of less than 5 feet. Color infrared aerial photographs were taken from a Cessna 150 platform annually after the hurricane. ERDAS GIS software and the accurate photo base allowed removal of scale irregularities and distortion that resulted from using a small aircraft. Scanned images, using a 10-square-foot pixel, were compared with kriged chloride concentration maps, also using a 10-foot cell size. Grid cells with estimated chloride concentration of more than 500 milligrams per liter also exhibited low reflectance in the infrared-enhanced color band, indicating tree mortality. Here, a small number of sampling wells accurately predicted ground-water movement of a contaminant (NaCl), and GIS and remote sensing verified this movement.

Introduction

Estimating contaminant flow in ground water is difficult because we cannot “see” the aquifer. We know that aquifers comprise sediments that vary from place to place, that changes in hydraulic conductivity determine the rate of water movement, and that the spatial variability of the aquifer sediment determines the hydraulic conductivity. Our inability to accurately represent spatial variability of the aquifer limits our ability to predict ground-water flow and, thereby, contaminant transport.

A large variety of prediction models are available (1-3), and stochastic methods of estimating spatial heterogeneity have been developed (4, 5) and tested (6, 7). On well-characterized field sites, these techniques can produce predictions of tracer movements that accurately predict experimental plumes in terms of mass behavior. Even at these research sites, the spatial distribution of hydraulic conductivity is not known well enough to predict behavior at any particular point.

Ground-water measurements generally derive from wells that are single points. To understand movement of an entire plume, these single point samples must be extended to represent areas. The geostatistical approach allows quantitative estimation of the spatial variation of point estimates (8). Kriging is a technique that uses spatial covariance to estimate values at points where no measurement exists (9). It produces the best linear unbiased estimator of nonmeasured points (8).

Following Hurricane Hugo, these techniques were used to study saltwater movement in the water table aquifer in forested stands of eastern Georgetown County, South Carolina. Clemson University received a grant from the U.S. Forest Service to examine forest mortality and regeneration success within the forest zone covered by salt water during the tidal surge. In this study, we used a small sample of ground water to estimate the direction and concentration of salt moving in the aquifer. Geographic information systems (GIS) proved to be a useful tool to verify conclusions based on the small sample size. Onsite sampling, aerial photography, vector and raster GIS, and spatial statistics were combined into one analysis system. The system estimated and verified directions of salt movement within the aquifer. GIS and remote sensing of forest mortality produced an independent indicator of salt movement that could be compared with the geostatistical technique.

Problem Statement

The main goal of the research project was to evaluate problems for forest regeneration in areas covered by salt water during the hurricane. In many of these areas, the mature trees died during the summer following the hurricane. These areas have very low elevation, little relief, and abundant rainfall, causing the water table to remain near the soil surface. The hypothesis was that salt movement within the aquifer killed the mature trees and could limit regeneration success. We divided the problem into three tasks: to determine if salt concentrations in the aquifer were high in areas where mature trees died, to determine pathways of salt movement within the aquifer that could explain high salt concentrations, and to predict regeneration success from the pattern of salt movement.

GIS contributed both to testing the initial hypothesis and to extending predictions to areas not initially studied. GIS has been used primarily to store and display spatial data in a way that preserves and presents the spatial relationships as well as the data. For this project, we collected two dissimilar data types. To determine salt movement, we measured salt concentrations and piezometric pressures in a series of wells. GIS had to represent the well data and the domain of the kriging procedures in a coordinate system compatible with the mortality data. We determined forest mortality data from infrared-enhanced color aerial photography. GIS also had to allow separation of the infrared signature of the photography, transform the signature into data that were comparable with the well data, and ensure that the coordinate systems of the well data and the mortality data represented the same true ground positions.

To use the GIS ability for this project, we needed to choose several GIS parameters. In this case, we estimated ground-water chloride concentration using kriging, which produced data on a grid comparable with mortality interpreted from photographs. A raster GIS representation could be compared with individual grid chloride values. Each grid cell was 10 square feet so that each cell would be within a single tree crown.

Methods

Study Location

The study was located on 12 acres of a small watershed located on the eastern side of Hobcaw Forest, an experimental forest managed by Clemson University, Department of Forest Resources. Hobcaw Forest is located on the end of a peninsula between the Winyah Bay and the Atlantic Ocean in eastern Georgetown County, South Carolina. The study watershed is located immediately west of the salt marsh and barrier island separating the forest from the Atlantic Ocean and is in Pleistocene-aged beach sediment. Watershed divides were created by former low dune lines, and the stream is within a small depression between these former dune lines. Divides are from 7 to 8 feet above sea level and the stream from 4 to 6 feet above sea level.

The study watershed is 50 miles northeast of Charleston, South Carolina, where the eye of Hurricane Hugo struck the U.S. coastline. Along this portion of the South Carolina coast, the tidal surge was approximately 10 feet above mean sea level (10), covering the entire watershed. After the hurricane, shallow auger holes contained water with sodium concentrations of 4,000 milligrams per liter (11). The hurricane winds did little damage to the watershed forest, but 25 percent of the large oaks were windthrown (12). Beginning in the spring of 1990, however, many hardwoods and pines began dying. By the winter of 1990 and 1991, a large portion of the forest on the watershed had died. Tree mortality did not correspond with high salinity measured by the initial auger-hole method, suggesting movement in the water table aquifer.

Well Installation

The water table aquifer is about 20 feet thick, consisting of fine sand similar to the present beach, with thin beds of shells 10 feet beneath the stream. The bottom of the aquifer is a bed of clay up to 3 feet thick over a leaky artesian aquifer composed of shell and sand. Local rainfall recharges the water table aquifer. Recharge for the lower aquifer is provided by leakage from the water table aquifer beneath the center of the peninsula, about 2 miles west of the watershed, where land elevations are 15 to 25 feet above sea level. Piezometric potential in the lower aquifer is generally a few inches above the water table aquifer, making it only weakly artesian (13).

We installed 24 multilevel ground-water samplers (14) in the water table aquifer. Five samplers were located in regeneration measurement plots (15) placed within the stream. Two samplers, one at each edge of the hardwood wetland, formed a line perpendicular to the stream at the regeneration plot. Two more samplers were located along these lines near the watershed divides on each side of the watershed. The 24 samplers formed five transects across the stream (see Figure 1). Piezometric potential and ground-water chloride concentrations were measured from these samplers from March 31, 1991, through April 1, 1991. Williams (16) provides a complete description of samplers, sampling procedures, and laboratory analysis.

GIS Implementation and Measures

A GIS system for Hobcaw Forest management was developed in 1987 using Environmental Systems Research Institute's PC ARC/INFO software (17). The initial system consisted of forest stand boundaries digitized onto 1:100,000 digital line graphs (DLGs) purchased from the U.S. Geological Survey. These relatively crude maps were combined with stand records and used for management decisions that did not require exact locations of stand boundaries. Later, management of the endangered red-cockaded woodpecker's habitat required that mapped stand lines be closer to true ground locations than the original DLG data scale allowed. A program of ground surveys and aerial photography was conducted in the late 1980s to locate stand boundaries more accurately (18).

In 1988, Georgetown County began a program to convert county tax mapping to computer-based systems. The first step was to acquire survey grade orthophotography. Copies of 1:400-scale orthophotographs, with guaranteed ground accuracy of plus or minus 5 feet, became available in 1990. Roads and stand boundaries were digitized from these photographs into the PC ARC/INFO database. The new, accurate map was combined with stand records of previous coverages to create stand record coverages on a map that was true to the ground within 5 feet, plus or minus 0.3 percent.

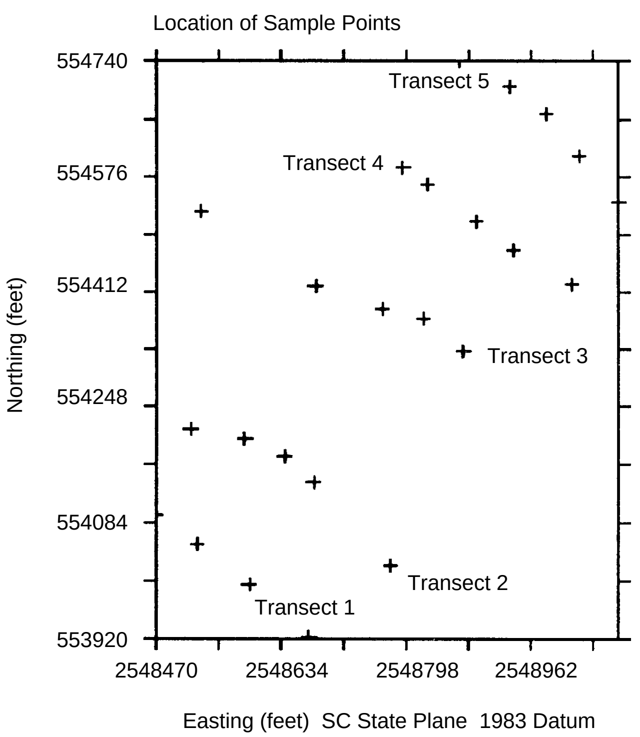


Figure 1. Position of sampling wells in rectangle defined for estimation of salinity movements.

In 1991, the GIS programs ERDAS VGA and LIVE LINK were obtained. ERDAS VGA programs allow image processing and raster GIS to be done on personal computers with VGA and some Super VGA monitor adapters. The LIVE LINK program allows display of both ARC/INFO and ERDAS images on the same monitor screen. Orthophotographs were scanned using 5-square-foot pixels and rectified with less than 1-pixel mean error, giving accurate ground locations of plus or minus 8 feet.

Ground surveys from a nearby benchmark provided accurate locations of the sampling wells. PC-TRAVERSE performed coordinate geometry from the survey notes, which was then plotted on the ARC/INFO forest stand database. Accuracy was checked by plotting the recognizable points on the survey with the scanned orthophotograph using the LIVE LINK software.

In February 1991, the Hobcaw Forest was photographed with infrared-enhanced color film. In this film, the red layer is sensitive to near infrared radiation that is strongly reflected by chlorophyll. Red colors in resulting prints indicate living vegetation. The color photography was not corrected for scale variation (from small fluctuation in aircraft altitude) or for distortion (caused by slight variations in the aircraft attitude). One photograph (1:1,320 scale) covering the study watershed was scanned into the ERDAS program. This image was rectified to the 1988 orthophotograph image using control points visible on both. A 10-square-foot pixel was used to sample individual tree crowns. The mean error of rectification was 1.5 pixels for a ground location, plus or minus 15 feet.

Ground-water chloride values at any one point varied over three orders of magnitude in all three dimensions and over two orders of magnitude with time. Annual averages of piezometric potential and chloride concentration, however, yielded interpretable results that were also statistically significant (16). Averaged values could then be combined with the surveyed sampler locations in the ARC/INFO system. The GIS also calculated the corner coordinates for a rectangle that would include all the sampler locations.

Geostatistical calculations were performed using the GS+ software. The data input to this program is an ASCII file of sample point locations in x,y coordinates and data values. The program allows calculation of semivariograms with various combinations of active and maximum lag distance and fitting of various model types to best fit the semivariogram (19). An active lag of 65 feet and a Gausian model produced the best fit:

$$\tau(h) = 0.001 + 1.337\ (1 - \exp\ [-h^2/20736]),\ r^2 = 0.616$$

where $\tau(h)$ is the semivariance at lag distance h.

This best fit model was then used in a block kriging (9) procedure. The procedure used eight nearest neighbors and calculated average values for 10- by 10-square-foot blocks. Values were calculated within the rectangle defined by the corner coordinates from the ARC/INFO procedure described above.

Finally, the rectangle defined around the sampler position formed the region of comparison between the rate of mortality, as sampled by infrared reflection, and estimated average chloride concentrations. The first comparison involved mapping reflection in the red band of the aerial photograph as a gray-scale map and comparing it with the contoured map of chloride concentration. The rectangle coordinates were used in the ERDAS software to create a subset of the scanned aerial photograph that included only the red band in the 5,145 pixels defined by the rectangle surrounding the samplers. In this subset, the infrared reflection was scaled as a gray-scale value between 0 and 255 for each 10- by 10-square-foot block defined in the concentration map. In addition to mapping, a regression of chloride concentration to gray-scale value was performed using the individual blocks.

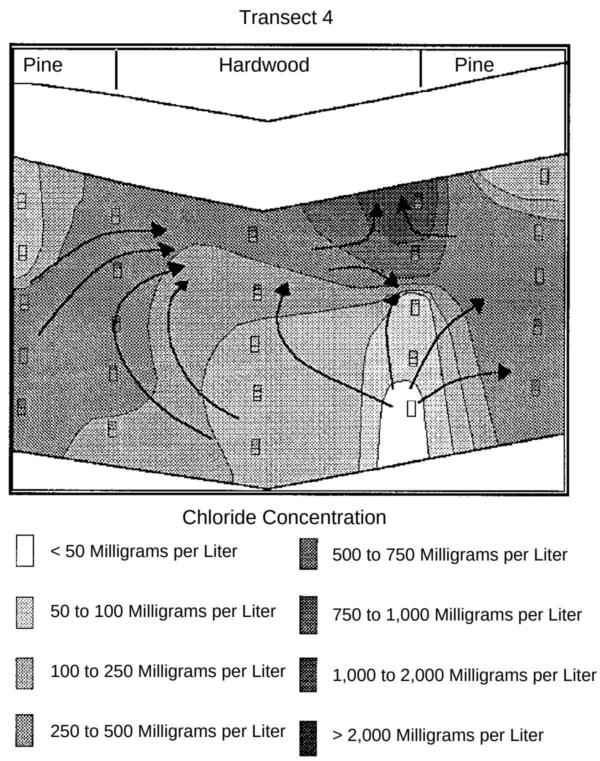


Figure 2. Cross section of aquifer at transect 4; gray scale represents chloride concentration, boxes at sampler positions represent 95-percent confidence limits of average chloride concentration in same gray scale, and arrows are perpendicular to contours of piezometric potential and represent two-dimensional vectors of streamlines.

Results

Ground Water

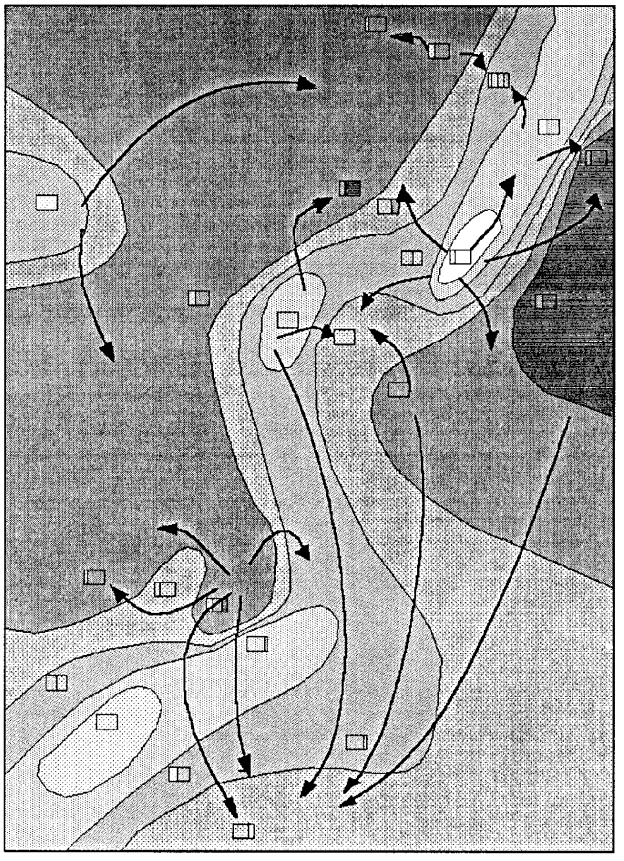
Ground-water chloride data reflected a consistent explanation of salt movement. Initial auger-hole data collected within a month of the hurricane indicated most salt was near the surface of the pine ridges, where, presumably, salt water had filled the aquifer to the soil surface during the hurricane. Data collected 30 months after the hurricane indicated the bulk of the salt had moved to the bottom of the aquifer under the pine ridges. Figures 2 and 3 represent the most significant results interpreted from the piezometric potential and chloride concentration measurements.

Figure 2 represents a cross section of the chloride concentrations and directions of ground-water flow in transect 4, the second most northern transect. The common information represented in this cross section was the west to east movement of ground water, representing the regional flow toward the forest edge. Also, there is an area of upwelling just east of the stream at the bottom of the aquifer, representing a leaky spot in the underlying clay layer. Upwelling causes the west to east streamlines to rise toward the surface along the western edge of the wetland. Chloride concentrations indicate large reservoirs of salt beneath each of the pine ridges and small pockets of fresher water near the surface, probably the result of rain infiltration during the 30 months since the hurricane. The water upwelling from the artesian aquifer was consistently fresh. Flow passing through the concentrated zone beneath the western ridge was pushed to the surface beneath the stream, where evaporation caused chloride concentrations to average more than 1,000 milligrams per liter.

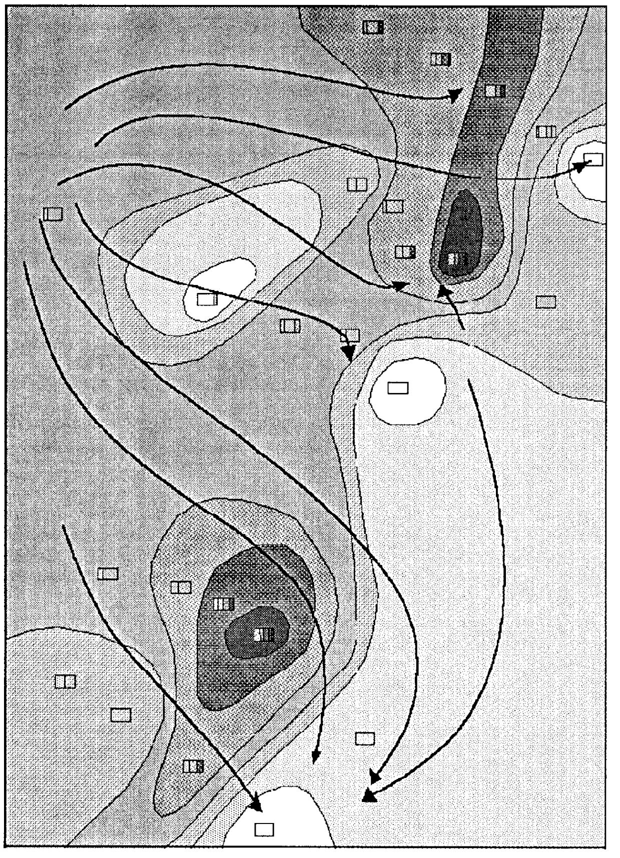
Figure 3 represents two plan views of the site at depths of 4 and 12 feet below the surface. The 12-foot plan view indicated that the east to west flow in transect 4 is only the east vector of a southeast flow. Other areas of upwelling exist beneath the stream. Chloride concentrations are highest beneath both ridges and lowest in the stream center. At the 4-foot depth, the east and southeast flows are also obvious. Also, as deeper flows from the western ridge are turned to the surface, high concentrations of chloride are present near the surface. High concentrations within the wetland result from water being carried to the surface due to upwelling within the wetland.

GIS Evaluations

The ground-water interpretations show a consistent explanation of salt movement. These interpretations are



Plan View 12-Foot Depth



Plan View 4-Foot Depth

Figure 3. Plan views of aquifer at depths of 4 and 12 feet below the surface with same chloride scale and two-dimensional plan vectors of streamlines.

based on only 24 sample points. Interpolation was linear using the nearest neighbor. The samples removed from the aquifer represent only 0.000014 percent of the aquifer volume. Interpretation of a three-dimensional flow regimen from such small sampling does not produce great confidence in the validity of the interpretation.

Salt at the 4-foot depth is most likely to interact with tree roots, and concentrations at this depth were used for kriging. Kriged results, mapped in the same manner as the aerial photography, show general agreement of high mortality and predicted chloride concentrations over 500 milligrams per liter (see Figure 4). A regression of chloride concentration (chloride) to gray-scale value (G) for the individual points yielded a significant negative correlation. The regression line G = 169 - 0.12 (chloride) explained only 27 percent of the variation in gray-scale value, however.

Conclusions

GIS was successfully used to verify interpretations of ground-water flow and salt movements in a shallow water table aquifer. A variety of computer software combined to create a system of analysis that allowed integration of field sampling, aerial photography, vector and raster GIS, and spatial statistics. Using this system, we compared chloride movement, measured by subsurface samplers, with remotely sensed tree mortality caused by soil salinity. The overall pattern of mortality was predicted by a 500-milligram-per-liter chloride contour estimated by kriging averaged concentrations. Estimation of mortality on a single tree basis was less successful, with a regression of chloride to infrared reflection explaining only 27 percent of the variation in reflection. The regression did not fit values of high reflection well but did predict reflection values of 100 or below (regions of high mortality) for concentrations above 500 milligrams per liter.

The most important factor in the success of this project was the availability of large-scale orthophotography. Georgetown County's investment in accurate mapping allowed creation of a map base that made scale correction of less costly aerial photography possible. Without assurance that pixels on the aerial photograph corresponded to the same locations as the subsurface samplers, correlations would have been meaningless.

Another factor that contributed greatly to the research was the fact that most of the computer software exported or imported data from simple ASCII files. The standard (x coordinate, y coordinate, data value) format in ASCII allowed files to be manipulated with spreadsheets

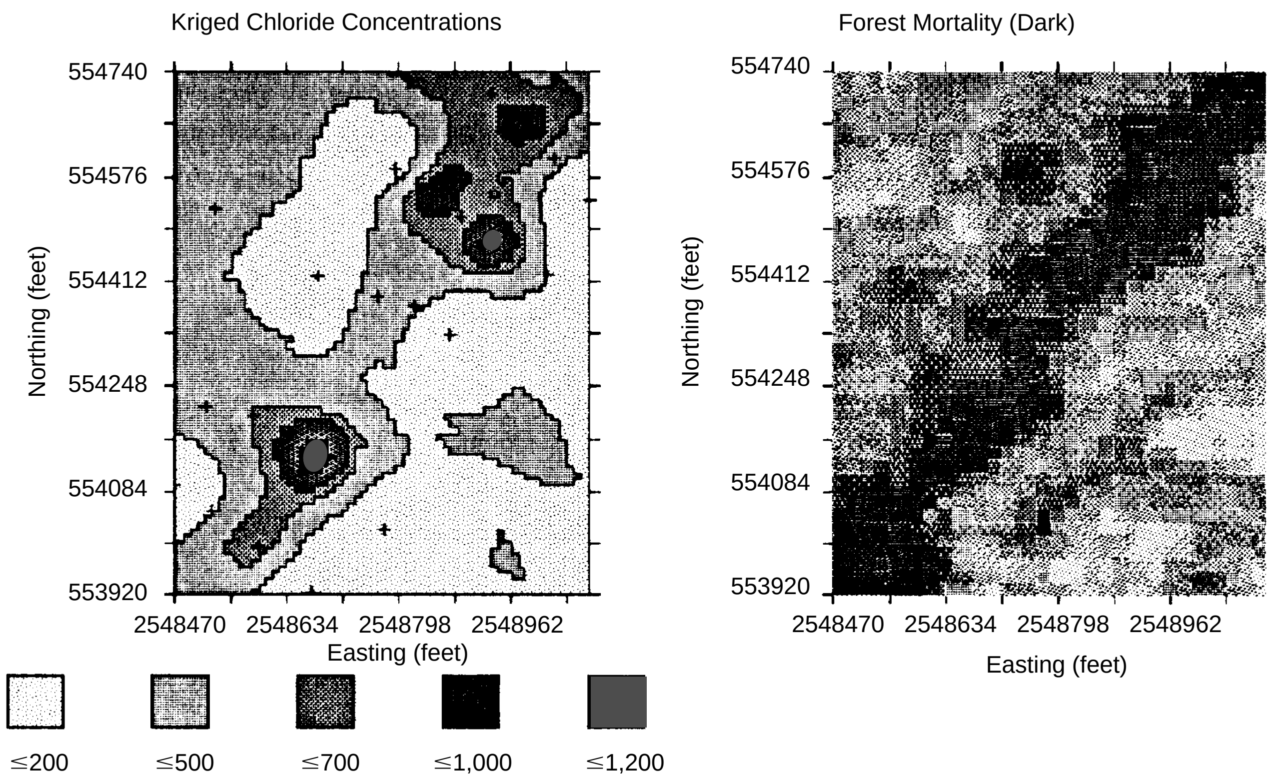


Figure 4. Plan view of chloride concentrations from krig analysis and gray scale of infrared reflection from aerial photograph. Lighter tones represent greater infrared reflection and less forest mortality.

or word processors. Creation of headers, positioning of columns, or changing order of rows or columns could be done for import into the next program. Although more difficult than point and click file transfers of the modern software, the simple standard format creates freedom to use the software in ways not anticipated by the software developers.

Finally, a clear problem statement aided in selecting the most applicable GIS techniques. GIS software allows several methods of data representation. In this example, we chose a raster representation with a cell the size of a tree crown. Criteria for choosing these parameters included physical dimensions of the phenomenon of interest, dimensions of GIS accuracy, and a desire for automated determination of values for individual comparisons. A careful review of the problem to be solved, data available, and capabilities of the GIS software are all necessary ingredients for a useful problem statement.

References

1. Van Genutchen, M.T., and W.J. Alves. 1982. Analytical solutions of the one-dimensional convective-dispersive solute transport equation. U.S. Department of Agriculture Technical Bulletin 1661.

2. Wang, H.F., and M.P. Anderson. 1982. Introduction to groundwater modeling: Finite difference and finite element methods. San Francisco, CA: W.H. Freeman.

3. Bear, J., and A. Verruijt. 1987. Modeling ground-water flow and pollution. Dordrecht, The Netherlands: D. Reidel Publishing Co.

4. Gelhar, L.W. 1993. Stochastic subsurface hydrology. Englewood Cliffs, NJ: Prentice Hall.

5. Dagan, G. 1987. Theory of solute transport by ground water. Ann. Rev. Fluid Mech. 19:83-215.

6. Garabedian, S.P., D.R. LeBlanc, L.W. Gelhar, and M.A. Celia. 1991. Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts, 2. Analysis of spatial moments for a nonreactive tracer. Water Resour. Res. 19(1):1,387-1,397.

7. Rehfield, K.R., J.M. Boggs, and L.W. Gelhar. 1992. Field study of dispersion in a heterogeneous aquifer, 3. Geostatistical analysis of hydraulic conductivity. Water Resour. Res. 28(12):3,309-3,324.

8. de Marsily, G. 1986. Quantitative hydrogeology. San Diego, CA: Academic Press.

9. Krige, D.G. 1966. Two-dimensional weighted moving average trend surfaces for ore-evaluation. J. South African Instit. Mining and Metallurgy 66:13-38.

10. Coch, N.K., and M.P. Wolfe. 1991. Effect of Hurricane Hugo storm surge in coastal South Carolina. J. Coastal Res. SI 8:201-208.

11. Gardner, L.R., W.K Michner, E.R. Blood, T.M. Williams, D.J. Lipscomb, and W.H. Jefferson. 1991. Ecological impact of Hurricane Hugo—salinization of a coastal forest. J. Coastal Res. SI 8:301-318.

12. Gresham, C.A., T.M. Williams, and D.J. Lipscomb. 1991. Hurricane Hugo wind damage to southeastern U.S. coastal tree species. Biotropica 23:420-426.

13. Nittrouer, P.L. 1988. Ground-water flow patterns in a forested beach ridge-scale system adjacent to a South Carolina salt marsh. Master's thesis, Department of Geological Sciences, University of South Carolina.

14. Williams, T.M., and J.C. McCarthy. 1991. Field-scale tests of colloid-facilitated transport. In: National Research and Development Conference on the Control of Hazardous Materials, Anaheim, California. Greenbelt, MD: Hazardous Materials Control Research Institute.

15. Conner, W.H. 1993. Artificial regeneration of bald cypress in three South Carolina forested wetland areas after Hurricane Hugo. In: Brissette, J.C., ed. Proceedings of the Seventh Southern Silvicultural Research Conference. U.S. Department of Agriculture Forest Service, General Technical Report SO-93. Southern Forest Experiment Station, Mobile, AL.

16. Williams, T.M. 1993. Salt water movement within the water table aquifer following Hurricane Hugo. In: Brissette, J.C., ed. Proceedings of the Seventh Southern Silvicultural Research Conference. U.S. Department of Agriculture Forest Service, General Technical Report SO-93. Southern Forest Experiment Station, Mobile, AL.

17. Lipscomb D.J., and T.M. Williams. 1988. A low-cost, map-based data retrieval system for forest managers. In: Proceedings of the American Congress of Surveying and Mapping/American Society for Photogrammetry and Remote Sensing, Fall Convention. Falls Church, VA: American Congress of Surveying and Mapping.

18. Lipscomb, D.J., and T.M. Williams. 1990. Developing a GIS for forest management in the 1990s. In: Resource technology 90: Second international symposium on advanced technology in natural resource management. Bethesda, MD: American Society of Photogrammetry and Remote Sensing.

19. Rossi, R.E., D.J. Mulla, A.G. Journel, and E.H. Franz. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecol. Monographs 62:277-314.

Geology of Will and Southern Cook Counties, Illinois

Edward Caldwell Smith
Illinois State Geological Survey, Champaign, Illinois

Introduction

The Silurian dolomite aquifer is the primary source of ground water in northeastern Illinois. It is overlain by glacially derived sands and gravels or tills. The sands and gravels within the glacial drift hydrologically interact with the fractured and creviced dolomite bedrock.

The purpose of this study was to define the extent of major glacial drift aquifers and their relationship to the shallow bedrock aquifer surface. The study succeeded in identifying two principal sand and gravel aquifers: an “upper” drift aquifer within the glacial tills and a “basal” drift aquifer overlying the bedrock. Bedrock topography, drift thickness, thickness of the Silurian dolomite, and thickness of major sand and gravel units were mapped to help define the geologic and hydrologic system and the interaction of the upper bedrock aquifer and the drift aquifers.

The data collected to create the various maps came from well records, engineering borings, oil and gas tests, and structure tests on file at the Illinois State Geological Survey (ISGS). Reviewing published reports, manuscripts, and unpublished reports on open file at the ISGS provided an overall perspective of the geology of the study area. Previously, no detailed studies of the hydrogeology of the entire area had been conducted. Incorporating water well and other data into a computer database greatly facilitated map construction. Preliminary maps were developed using Interactive Surface Modeling (ISM) software and a geographic information system (GIS).

Past regional geologic studies of the northeastern Illinois area that have encompassed this study area include Thwaites (1), Bretz (2), Bergstrom et al. (3), Bretz (4), Suter et al. (5), Hughes et al. (6), and Willman (7). Bogner (8) and Larsen (9) included interpretive maps of the surficial geology of the area as a part of planning studies for northeastern Illinois.

Map Construction

Creating the database used in the construction of the maps for this project entailed inputting information from well driller’s logs into a PC-based computerized spreadsheet (Quattro Pro). Well logs were primarily from water wells and engineering borings. Data items input into the spreadsheet included:

- Well identification (ID) number
- Owner name
- Location of well
- Thickness of drift
- Depth to top and bottom of the bedrock
- Depth to top and bottom of each sand unit

The ground surface elevation of each well was interpolated from United States Geological Survey (USGS) 7.5-minute quadrangles. Elevations of the top of bedrock and top and bottom of sand bodies were calculated based on the interpolated elevations. Locations were verified wherever possible using plat books by matching either landowner names or the address location from the well log. After compilation, the data were converted to ASCII text and transferred into an ARC/INFO (Versions 5.0.1 and 6.0) database on a SUN SPARC workstation. ARC/INFO is a product of Environmental Systems Research Institute, Inc., of Redlands, California.

Of the more than 10,000 records reviewed for this project, over 5,100 were input into the database. Subsequently, numerous data quality checks ensured that duplicate well ID numbers were corrected, locations were corrected, thicknesses were checked so that the sand thickness data reported did not exceed drift thickness, and elevations were checked so that elevation of a sand body was not below the bedrock surface. After running the data quality checks and removing questionable data from the database, approximately 5,000 records remained.

ISM, a contouring package from Dynamic Graphics, Inc., of Alameda, California, helped to create two dimensional grid representations of:

- Surface topography
- Drift thickness

- Bedrock topography
- Bedrock isopach
- Intermediate sand body isopach
- Basal sand isopach

ISM also allowed for the creation of contoured output of the grids. Grids are regularly spaced rectangular arrays of data points (nodes) that allow for efficient mathematical calculations and contouring. ISM uses a minimum tension gridding technique, allowing for the curvature (change in slope) of the surface to be spread throughout the surface rather than being concentrated at the input data points. The ISM program uses a biharmonic inverse cubic spline function (algorithm) to assign data values to grid nodes. This function assumes that for any grid node assignment, input data points farther away from the node being evaluated have less influence on that node's value than nearer data points. To determine each grid node value, ISM calculates an average value from the surrounding scattered input data (up to 15 input data points) and finds the standard deviation. ISM continues to refine the values of the grid nodes until the standard deviation is minimized (10).

Several grid spacings were reviewed to determine which would best represent the density of the data. The grids that ISM uses, as described above, determine the fineness to which the data control the resultant contours. Experimentation was necessary to determine a grid spacing that adequately represented the data. Too fine a grid spacing can exaggerate or overly weight individual points, causing the resultant contours to be overly jagged. With too large of a grid spacing, the contours can become overgeneralized and become much less data dependent because the calculated grids are overaveraged.

The two-dimensional grid of the land surface topography was based on surface topography lines and spot elevations digitized from USGS 7.5-minute quadrangles. The linework for each quadrangle was converted to ASCII files of data points. The ASCII files contained x and y coordinates and the elevation value of each data point. After inputting the ASCII files into ISM, a two-dimensional grid for each quadrangle was created. ISM also generated contour lines from each grid. Comparing plots of the generated lines with USGS 7.5-minute topographic maps allowed for the correction of errors and ensured that the grid elevation values were within 10 feet of the elevations shown on the USGS maps. An ISM two-dimensional grid of the entire area's surface topography was created by combining the grids. After creating a contoured surface of the grid, an ARC/INFO coverage of the output was produced. ARC/INFO was used to edit the coverage and produce the final map.

The two-dimensional grid of the bedrock surface topography was based on data from water well and engineering boring logs, ISGS field observations of outcrop locations, and previous ISGS mapping (9). An ASCII file of x and y coordinates and the elevation of each bedrock top was input into ISM. Subtracting the bedrock topography grid from the land surface grid produced a grid of the drift thickness. A contoured output of the grid was produced, and an ARC/INFO coverage of the output was created. Again, ARC/INFO was used to edit the coverage and produce the final map.

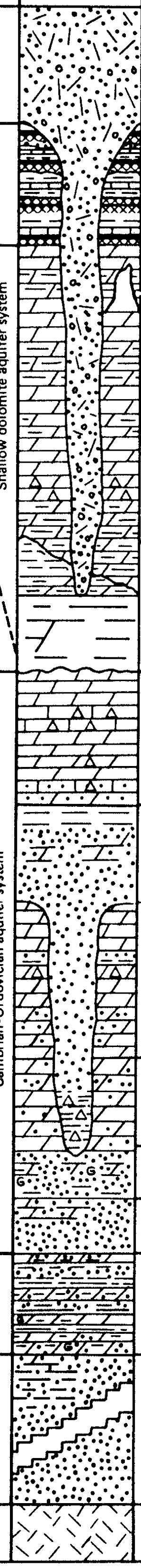
Creating the isopach maps entailed subtracting the top and bottom elevations of each unit to calculate the thickness of each unit. ASCII files of the x and y coordinates and the thickness values for each data point were input into ISM. ISM then created two-dimensional grids of each isopach. Contoured output of each grid was produced, which allowed for the creation of ARC/INFO coverages of the output. ARC/INFO was used to edit the coverages and produce the final maps.

Bedrock Geology of the Study Area

All the sedimentary bedrock units are of the Paleozoic Era. The Paleozoic bedrock comprises sequences of sandstones, dolomites, limestones, and shales. The stratigraphic column of Figure 1 illustrates the vertical succession of the bedrock. Major tectonic activity of the area includes the formation of the Kankakee Arch in Ordovician time (11) and faulting along the Sandwich Fault Zone. Faulting along the Sandwich Fault Zone (see Figure 2) may have occurred coincidentally with the formation of the Lasalle Anticlinorum in early Pennsylvanian time (12). No further faulting has been noted since deposition of glacial sediments. Bedrock units gently dip to the east (7). The majority of the area lies on the Niagara cuesta, a south and west facing scarp that comprises the resistant Silurian strata that have an eastward dip of roughly 15 feet per mile (13). The Silurian strata are absent west of the Kankakee River as well as in an area west of the Des Plaines River in west-central Will County (see Figure 2). This study relates to the hydrogeology of the Silurian strata and the drift materials, and details only the uppermost bedrock units. The report, however, does briefly summarize units below the Maquoketa Group using information from Hughes et al. (6) and Visockey et al. (14).

Precambrian Bedrock

Granites or granitic rock compose the Precambrian basement of northern Illinois. Few details about the nature of the basement rocks are known because few wells have completely penetrated the sedimentary bedrock of the region. The elevation of the top of the Precambrian basement probably stands at 4,000 feet below mean sea level in the study area.



SYSTEM	SERIES	GROUP OR FORMATION	AQUIFER	LOG	THICKNESS (FT)	DESCRIPTION
QUATERNARY	PLEISTOCENE		Sands and Gravels		0 – 250	Unconsolidated glacial deposits-pebbly clay (till), silt, sand and gravel Alluvial silts and sands along streams
PENNSYLVANIAN	DES MOINESIAN	Spoon and Carbondale			0 – 110	Shale, sandstone, clay, limestone, and coal
SILURIAN	NIAGARAN	Racine	Silurian (Shallow dolomite aquifer system)	reef	0 – 350	Dolomite, very pure to argillaceous, silty, cherty; reefs in upper part
		Sugar Run				Dolomite, slightly argillaceous and silty
		Joliet				Dolomite, very pure to shaly and shale, dolomitic; white, light gray, green, pink, maroon
	ALEXANDRIAN	Kankakee			0 – 100	Dolomite, pure top 1' - 2', thin green shale partings, base glauconitic
		Elwood				Dolomite, slightly argillaceous, abundant layered white chert
		Wilhelmi				Dolomite, gray, argillaceous and becomes dolomitic shale at base
ORDOVICIAN	CINCINNATIAN	Maquoketa			80 – 250	Shale, red to maroon, oolites Shale, silty, dolomitic, greenish gray, weak (Upper unit) Dolomite and limestone, white, light gray, interbedded shale (Middle unit) Shale, dolomitic, brown, gray (Lower unit)
	CHAMPLAINIAN	Galena	Galena-Platteville (Cambrian-Ordovician aquifer system)		310 – 380	Dolomite, and/or limestone, cherty (Lower part) Dolomite, shale partings, speckled
		Platteville				Dolomite and/or limestone, cherty, sandy at base
		Glenwood	Glenwood-St. Peter		125 – 600	Sandstone, fine and coarse-grained; little dolomite; shale at top
		St. Peter				Sandstone, fine to medium-grained; locally cherty red shale at base
	CANADIAN	Shakopee	Prairie du Chien		0 – 410	Dolomite, sandy, cherty (oolitic); sandstone
		New Richmond				Sandstone interbedded with dolomite
		Oneota				Dolomite, white to pink, coarse-grained cherty (oolitic)
		Gunter				Sandstone, medium-grained, slightly dolomitic
CAMBRIAN	CROIXAN	Eminence	Eminence Potosi		0 – 280	Dolomite, light colored, sandy, thin sandstones
		Potosi				Dolomite, fine-grained gray to brown, drusy quartz
		Franconia	Franconia		110 – 160	Dolomite, sandstone and shale, glauconitic, green to red, micaceous
		Ironton	Ironton-Galesville		135 – 235	Sandstone, fine to coarse-grained, well sorted; upper part dolomitic
		Galesville				
		Eau Claire			390 – 570	Shale and siltstone, dolomitic, glauconitic; sandstone, dolomitic, glauconitic
		Elmhurst Member	Elmhurst-Mt. Simon aquifer system		2200	Sandstone, coarse-grained, white, red in lower half; lenses of shale and siltstone, red, micaceous
		Mt. Simon				
PRE-CAMBRIAN						Granitic rocks

Figure 1. Generalized stratigraphic column of rock units and aquifers in northern Illinois (prepared by M.L. Sargent, ISGS).

Cambrian

The Elmhurst-Mt. Simon Sandstone comprises the oldest sedimentary units in Illinois and consists of medium-grained sandstones. It has a total thickness of approximately 2,500 feet. The upper part of this unit has acted as an aquifer in the Chicago region in the past; ground-water mining of the aquifer (a nonreplenished lowering of the static water level), however, has led to a discontinuation of its use for that purpose. The Eau Claire Formation, the Basal Sandstone Confining Unit (14), consists of dolomitic shale and siltstone with thin beds of sandstone. It has a thickness of 300 feet to 400 feet and separates the Elmhurst-Mt. Simon aquifer from the Ironton-Galesville Sandstones. The Ironton-Galesville Sandstones have a thickness of 150 feet to 250 feet and serve as a source of ground water in northern Illinois (6). The Galesville Sandstone is fine-grained, while the Ironton Sandstone is coarser grained and contains more dolomite. The Knox Megagroup, the Middle Confining Unit (14), comprises all the bedrock units between the Ironton-Galesville Sandstones and the Ancell Group. It includes the:

- Cambrian Franconia Formation
- Potosi Dolomite
- Eminence Formation
- Jordan Sandstone
- Ordovician Prairie du Chien Group

The Knox Megagroup is primarily dolomitic in composition, though it contains thin sandstones. Its thickness ranges from 400 feet in the northern portion of the study area to about 700 feet in the southernmost tip of Will County. The sandstones tend to be somewhat discontinuous and, where present, offer a localized source of ground water. The group as a whole acts as a confining unit between the Ironton Sandstone and the Ancell Group.

Ordovician

The Ancell Group, which contains the St. Peter Sandstone and Glenwood Sandstone, has a thickness of roughly 200 feet throughout the study area except in north-central Will County where it is over 400 feet. The thickness of the Ancell Group varies considerably in northern Illinois because it rests on an erosion surface. The Ancell Group is the shallowest aquifer present in this area below the Silurian dolomite aquifer. The elevations of the top of the Ancell Group range from just over sea level in the northwest corner of Will County to 500 feet below mean sea level in the southwestern corner. The Galena and Platteville Groups provide a sequence of carbonate rocks that are primarily dolomitic in composition. The Platteville Group conformably overlies the Ancell Group. The two units have a combined thickness of 350 feet throughout this part of the state. The Galena and Platteville Groups, combined with the overlying Maquoketa Shale Group, act as an aquitard between the Ancell aquifer and the Silurian dolomite aquifer.

Maquoketa Shale Group

The study area has three subaerially exposed bedrock units. The oldest of these that this report details are Ordovician-aged strata comprising the Cincinnatian Series Maquoketa Shale Group. The thickness of the Maquoketa Group ranges from 260 feet in eastern Will County to 120 feet in the northwestern corner of Will County and is unconformably overlain by Silurian strata (15). The Maquoketa Group comprises four formations:

- Scales Shale
- Fort Atkinson Limestone
- Brainard Shale
- Neda Formation

The Scales Shale forms the lowermost unit and consists of gray to brown dolomitic shale. Thin layers with phosphatic nodules and pyritic fossils occur near the top and base of the unit. The Scales Shale may attain a thickness of up to 120 feet in this region (15). The Fort Atkinson Limestone, a coarse-grained crinoidal limestone to fine-grained dolomite, may range up to 60 feet thick (15). The Brainard Shale comprises greenish gray dolomitic shale and has a thickness of generally less than 100 feet (16). The Neda Formation, the youngest formation in the Maquoketa Group, is relatively thin with a thickness of usually less than 10 feet. In some places, it may attain a maximum thickness of 15 feet. The Neda is exposed along the Kankakee River, and the Silurian-aged Kankakee Formation typically overlies it. The Neda Formation consists mostly of red and green shale with interbedded goethite and hematite oolite beds (7, 16).

Silurian System

Silurian-aged rocks consist almost solely of dolomites and dolomitic limestones. The Silurian is divided into the Alexandrian and Niagaran Series. The Alexandrian Series is about 25 feet thick and is represented by the Kankakee, Elwood, and Wilhelmi Formations. These formations are a fine- to medium-grained, white, gray to pinkish gray dolomite. The Kankakee Formation is exposed along the Kankakee River in southern Will County (17).

The Niagaran Series comprises much of the bedrock surface of this area and includes three formations. The Joliet Formation has a lower member of dolomite with interbedded red and green shale, and two upper members with an increasing purity of dolomite toward the top of the formation (7). The Sugar Run Formation, formerly termed the Waukesha Formation (17), is an argillaceous, fine-grained, medium- to thick-bedded, brownish gray dolomite (7). The Racine Formation is the thickest unit in the Niagaran Series, attaining a thickness of as much as 300 feet (17). The Racine Formation contains large reefs that are as high as 100 feet and consist of vugular gray dolomite. The inter-reef rock consists of dense, cherty gray dolomite. The Racine Formation is exposed in the bluffs along the Des Plaines River from Joliet to Blue Island, Illinois (17).

Figure 2 is an isopach of the Silurian dolomite indicating the thickness of the unit in the study area and the boundary of the Silurian rocks. The Silurian dolomite aquifer has a maximum thickness of just over 500 feet in the southeast corner of Will County and becomes thicker to the east and south. It rapidly increases in thickness from its margin along the western border of Will County, where it has eroded. The contact between the Silurian dolomite and the underlying Maquoketa Shale Group has relatively little relief. Thus, the major differences in thickness of the unit result from erosion of the bedrock surface. Joints and fracture patterns within the upper bedrock have a dominantly northwest-southeast and northeast-southwest orientation (18).

Pennsylvanian System

Pennsylvanian-aged bedrock is found in the southeastern portion of Will County west of the Kankakee River with an outcropping at the confluence of the Des Plaines and Kankakee Rivers. The lowermost unit, the Spoon Formation, is very thin and consists of clay beds with scattered occurrences of coal formed in channel-like depressions (19). The Spoon Formation overlies the Maquoketa Shale Group. The overlying Carbondale Formation may attain a thickness of over 100 feet in the southwestern corner of Will County. The Carbondale Formation consists of shale with thin limestone beds. The lowermost unit, the Colchester (Number 2) Coal Member, outcrops in this area and attains a thickness of up to 3 feet. It has been extensively mined along the Will-Grundy-Kankakee County border where large areas of strip-mined land are evident. Most of the available coal has been mined out, and numerous gob piles exist

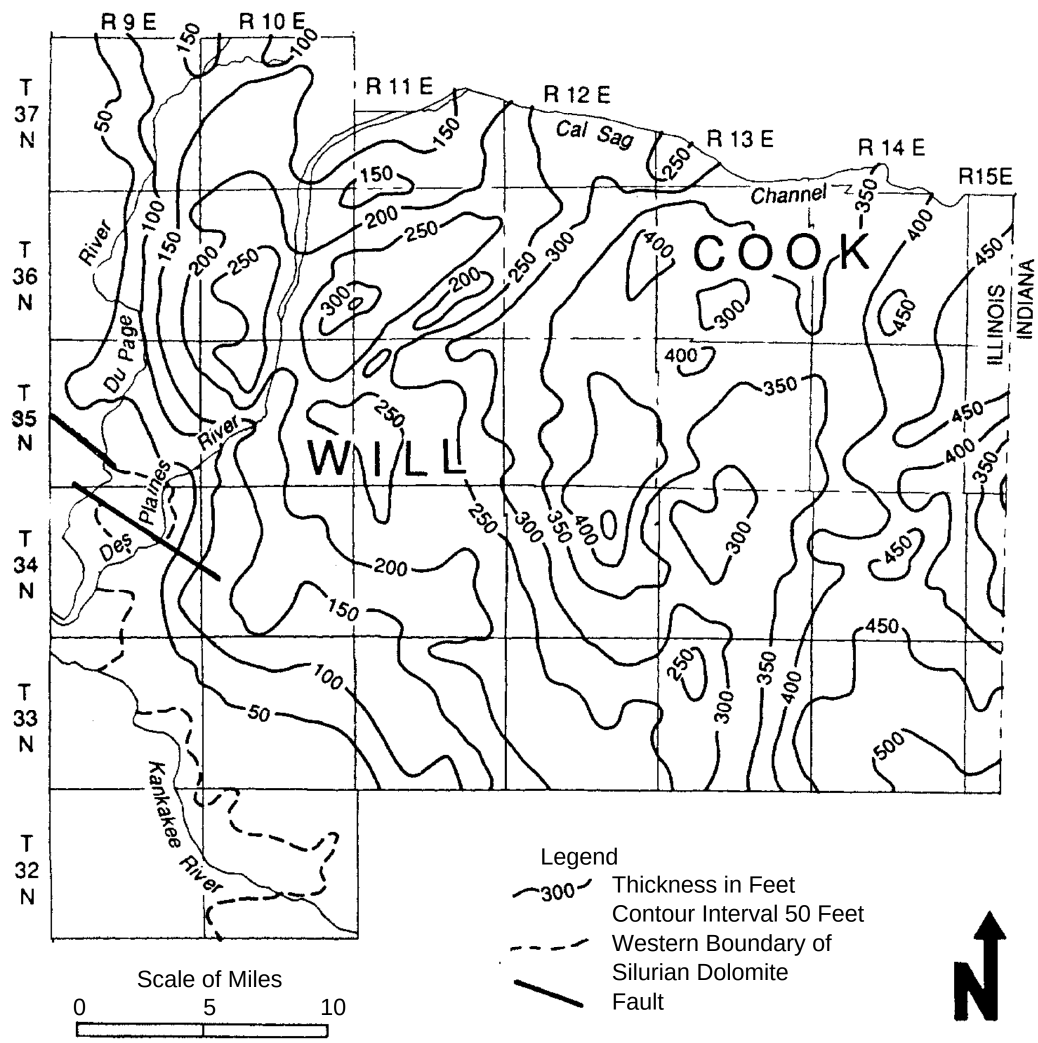


Figure 2. Thickness of the Silurian dolomite.

in the area of Braidwood. The Francis Creek Shale Member, which overlies the Number 2 coal, constitutes the remainder of Pennsylvanian units in the study area. The Francis Creek Shale is gray with numerous flattened concretions that contain the Mazon Creek flora of Pennsylvanian-aged fossils (19). Weathering of the mine slag materials may have exposed fossiliferous concretions in the gob piles (7).

Bedrock Topography

The highest bedrock elevations are in east-central Will County where the bedrock rises to over 700 feet above mean sea level (see Figure 3). Bedrock uplands occur as a broken curved ridge from the southeast to the northwest with bedrock elevations consistently rising over 650 feet above mean sea level. The bedrock surface slopes from the bedrock upland high westward to the Des Plaines River. It also has a regional downward slope to the south into Kankakee County, Illinois, to the northeast into the Lake Michigan basin, and to the east into Indiana. West of the Des Plaines River, the bedrock surface rises to over 650 feet above mean sea level in northeastern Kendall County. Elsewhere, the surface has relatively low relief.

The dominant features of the bedrock surface are the river valleys. The Des Plaines River valley is better expressed than the Kankakee River valley. This is true, in part, because it is older and acted as a drainageway for glacial meltwater where it may have become entrenched in the present valley. The Kankakee River valley may be less expressed partly because of the amount of scouring that occurred over a large area during the Kankakee flood event such that the river is not entrenched in most places. Also, smoothing of the study's contour maps has generalized some of the detail.

The buried Hadley Bedrock valley, described initially by Horberg and Emery (20), probably existed prior to glaciation and concurrently with the preglacial Des Plaines River. The valley may have acted as a drainageway for glacial meltwaters until the time that glacial debris buried it. Glacial scouring was originally believed to have formed the valley, but evidence presented by McConnel indicated a fluvial origin of the valley (21). Also, the base of the Hadley valley does not overhang or lie much below the Des Plaines valley but rather joins it at a smooth juncture.

The bedrock surface contains a number of sinkholes or closed depressions that are expressions of karst

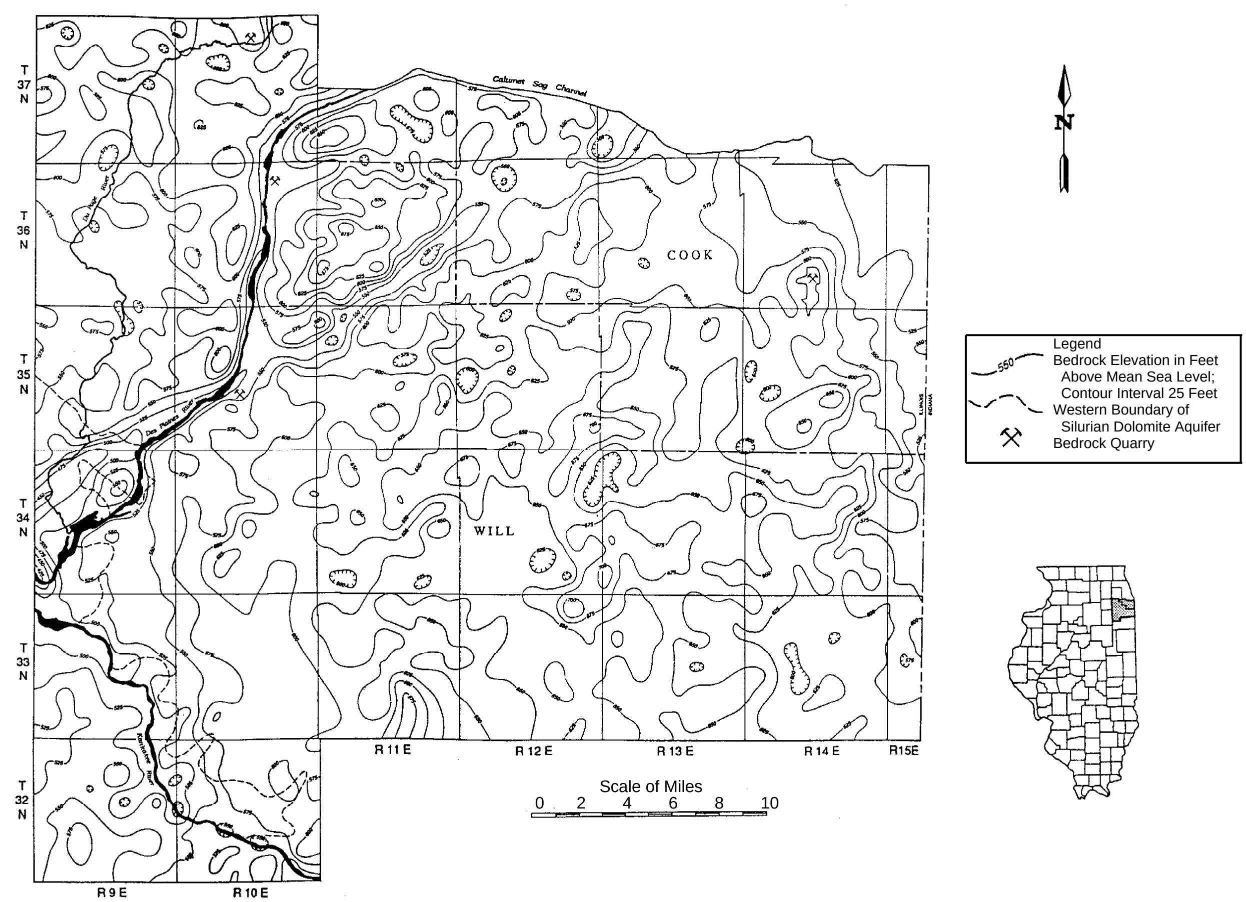


Figure 3. Topography of the bedrock in the study area.

development that formed prior to continental glaciation. Karst, a terrain developed on limestone or dolomite by solution or dissolving of the rock, is characterized by closed depressions and cavity development along joints and fractures. Fischer (22) first noted karst features in the Joliet area where early Pennsylvanian sediments of shale and clay filled cavities in the upper bedrock. Buschbach and Heim (23) indicated closed depressions in the Silurian dolomite surface in their bedrock topography map for the Chicago region. They speculated these depressions were expressions of karst development. McConnel (21) demonstrated the existence of sinkholes in the area of the buried Hadley Bedrock valley northeast of Joliet by using seismic refraction survey data.

Glacial Geology

The sediments overlying the bedrock comprise tills, sands and gravels, lacustrine deposits from glacial lakes, and surficial eolian deposits of loess and sand. The unconsolidated deposits are over 150 feet thick along the crest of the Valparaiso Morainic System. Figure 4, adapted from Willman (24), indicates the principal moraines. In the area where the Hadley Bedrock valley is present, the deposits attain a thickness of over 175 feet. Bedrock is mainly exposed along the Des Plaines River valley and its tributaries. It is also exposed in isolated areas in southeastern Cook County. The drift thickness map (see Figure 5) indicates the distribution of the earth materials overlying the bedrock and the locations of bedrock outcrops. The bedrock outcrop information for this map was derived from Piskin (25) and Berg and Kempton (26).

Erosion of the glacial sediments was a major factor in controlling the drift thickness of the area. Succeeding glaciers scraped off previously deposited sediments, but glacial meltwaters, which came from the east and north along the river channels, caused much of the erosion. Both the Kankakee River and Des Plaines River acted as meltwater channels as the glaciers melted. The Du Page River acted as a minor drainageway and was most active during large-scale flooding events. The thickness of the drift varies in the area also because of the topographic control that the bedrock on the overlying sediments exercises. The crest of the Valparaiso moraines coincides with the topographic high in the bedrock

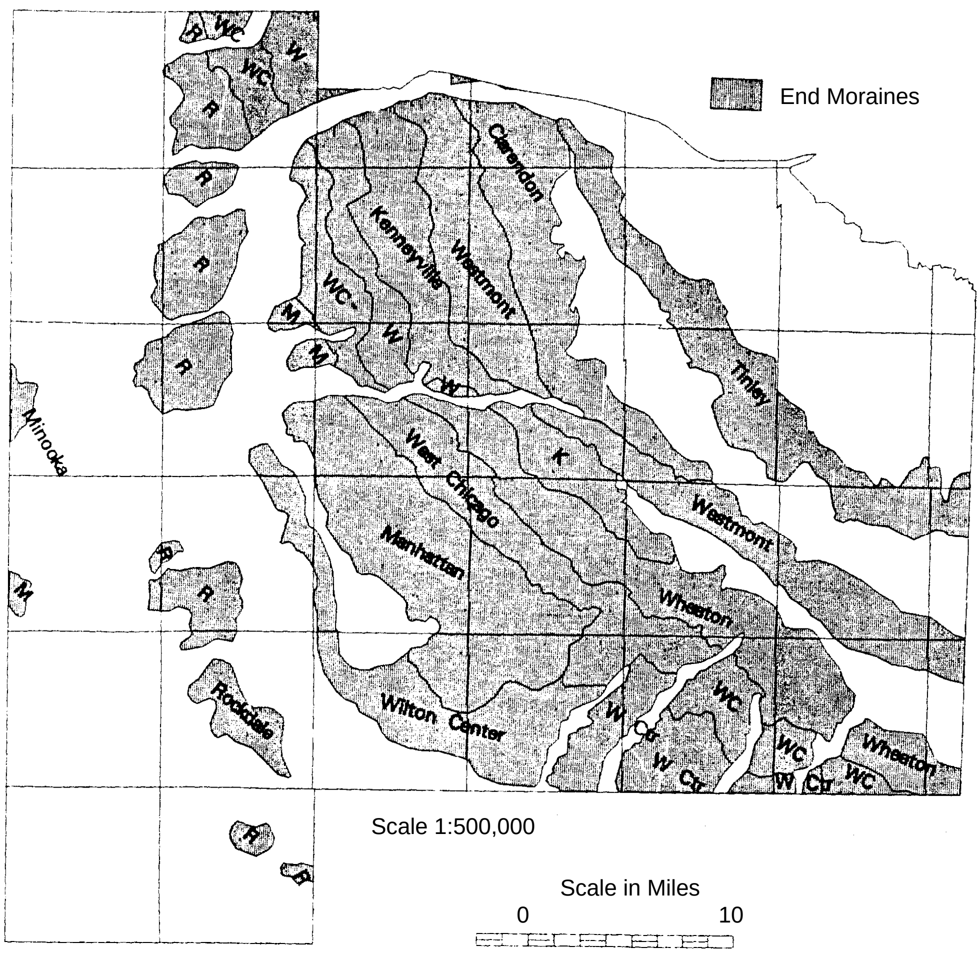


Figure 4. End moraines (late Wisconsin) in Will and southern Cook Counties, Illinois.

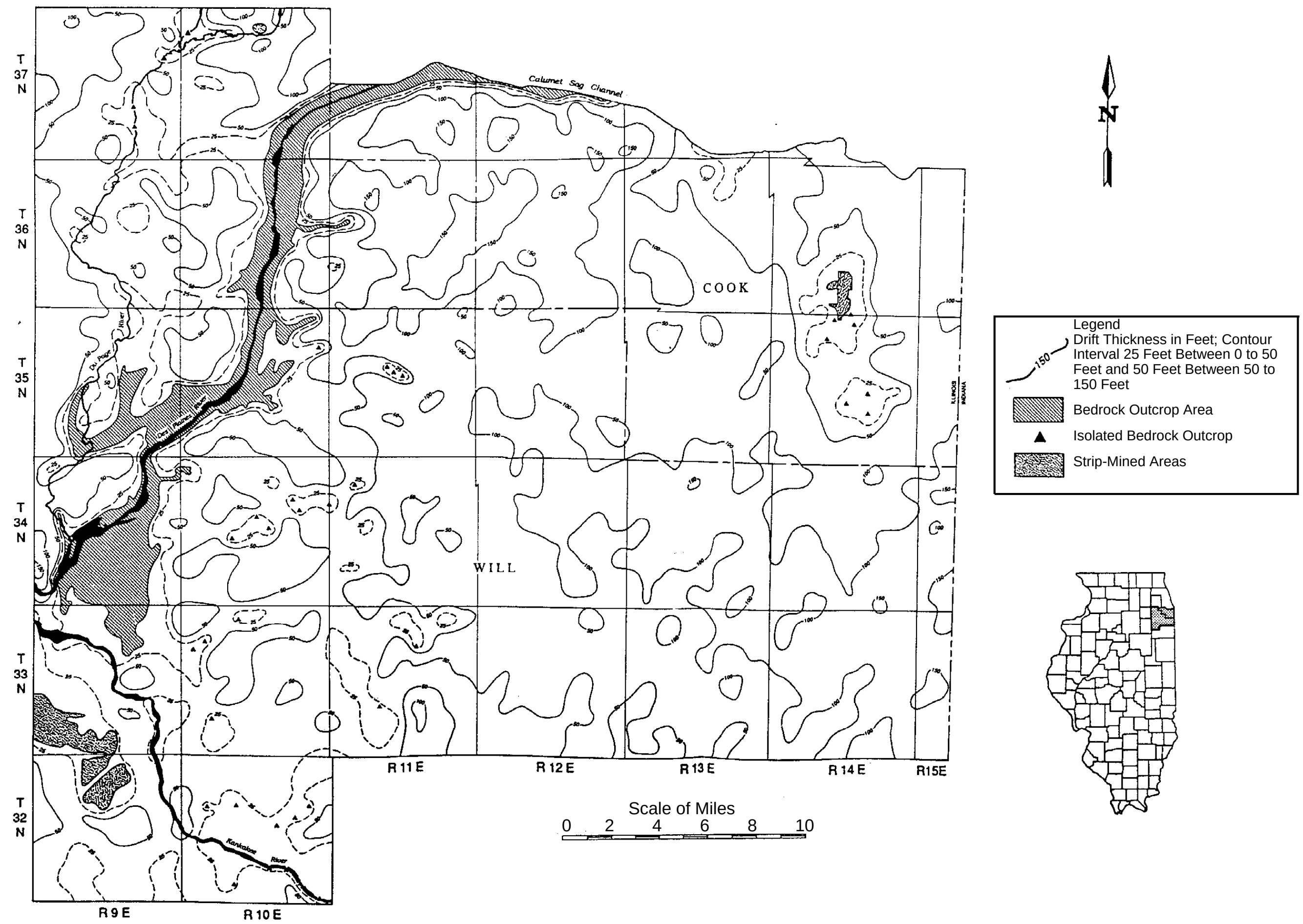


Figure 5. Thickness of the glacial drift in the study area and bedrock outcrop information (25, 26).

surface. The cross sections in Figure 6 also show this. The bedrock high may have caused late Woodfordian glaciers to stall repeatedly in the same area, causing moraines to build atop one another sequentially (27).

Descriptions by well drillers note few variations in the character of the unconsolidated sediments; therefore, we did not attempt to correlate these deposits. The drift materials present in the study area are late Wisconsinan or younger. Though this region experienced glaciating repeatedly prior to the Wisconsinan glaciation, no Illinoisan or pre-Illinoisan deposits have been identified (28). The drift units divide into three main units (29):

- The Lemont Drift
- The Yorkville Drift
- The Wadsworth Drift

The three drift units are all part of the Wedron Formation of Wisconsinan age. The Lemont Drift has a dolomitic character because the source material for the diamicton was glacially eroded Silurian dolomite. The Lemont Drift is the oldest of the three units and is found only underlying the Wadsworth Drift. The Yorkville Drift is the only drift unit present west of the Valparaiso Morainic System boundary within the study area. It overlies the bedrock surface wherever the basal sand unit is not present. The Wadsworth Drift comprises silty and clayey diamictons and is the youngest of the drifts (29). It overlies the Lemont Drift and the upper sand unit. In the cross sections (see Figure 6), where the upper sand unit is present, it roughly indicates the boundary between the Lemont and Wadsworth Drifts. The gradation between the different drift units at the Valparaiso System boundary is not well defined. The Wadsworth Drift appears to grade into the Yorkville Drift because they are very similar in composition near the boundary (9).

The large Kankakee flood left extensive deposits of sand and gravel and lacustrine sediments along the Kankakee River and Des Plaines River. The flood occurred as glacial meltwaters built up behind a constriction at the Marseilles Morainic System to the west (30). Large glacial lakes, which developed during the flood, subsequently emptied into the Illinois River valley after a breach in the moraines developed. The force of the flood waters eroded the glacial deposits along the river valleys, flattened the surface of the drift, and, in places,

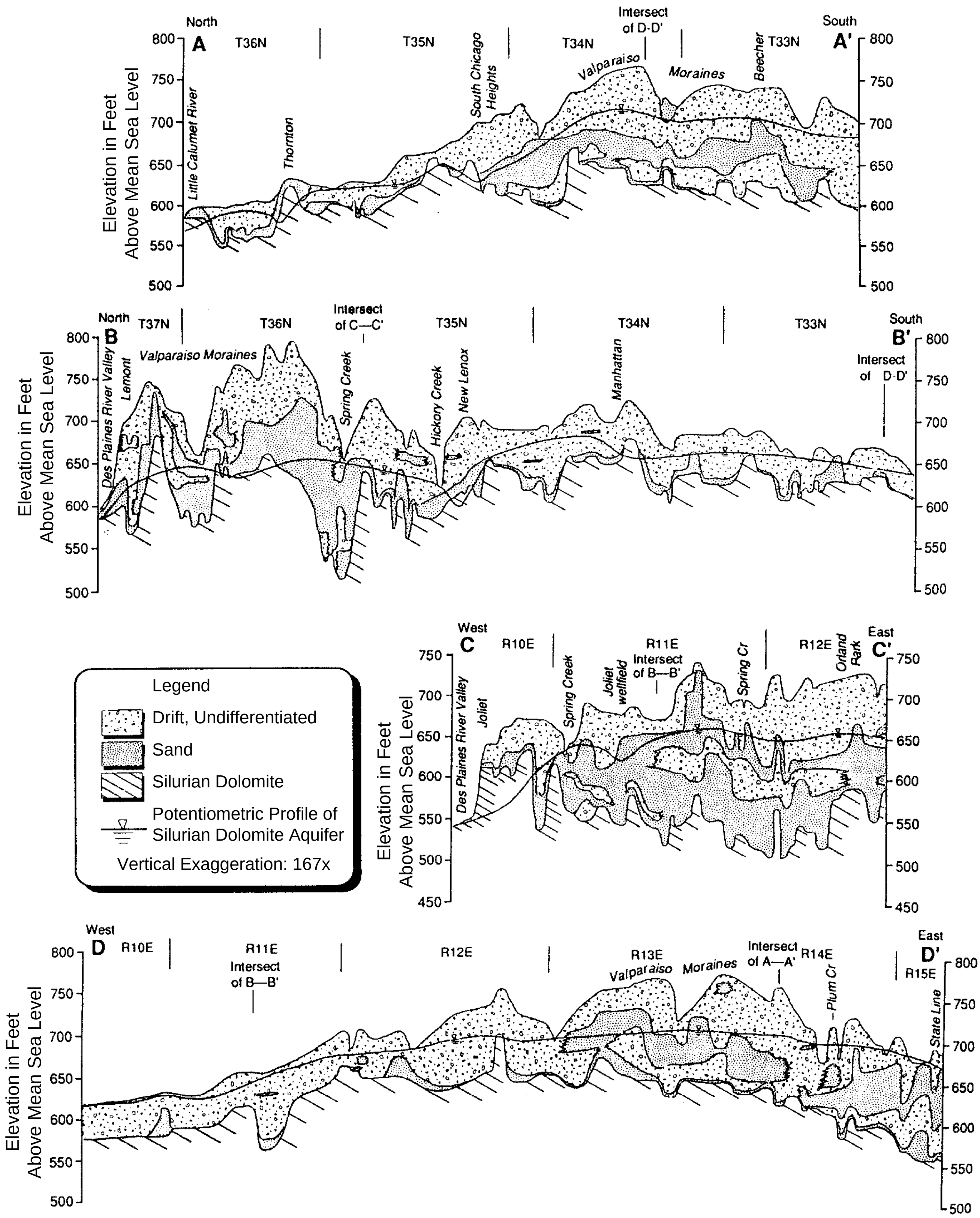


Figure 6. Geologic cross sections of the glacial drift and potentiometric profile of the Silurian dolomite aquifer.

exposed the underlying bedrock. The flood event formed thin, dispersed lake plain deposits of silt, clay, and sand in southwestern Will County. Some lacustrine deposits lie between morainic ridges in southern Cook County where small glacial lakes developed as the Valparaiso Moraines were being deposited (8).

Figure 7 shows the locations of some of the surficial materials. Sands and gravels were also deposited along tributary creeks and in abandoned channels that once connected the Du Page River and Des Plaines River north of their present juncture. Wind has reworked the surficial sand deposits forming low dunes along the Kankakee River in southern Will County. Masters (31) classified the sand and gravel deposits of the area by their origin, indicating that most of the deposits present in the valley of the Des Plaines River formed as well-sorted valley train deposits. In the Kankakee River valley, the sands and gravels were primarily deposited as riverine sediments during the Kankakee flood event.

Sand and Gravel Isopachs

The sand and gravel isopach maps (see Figures 8 and 9) indicate the variations in thickness of the upper and basal sand and gravel units. The most extensive deposits of both exist throughout the area overlain by the Valparaiso Morainic System. This may be associated with bedrock control on the formation of the moraines and associated deposits referred to earlier.

The thickest deposits lie in the buried Hadley Bedrock valley where thicknesses of both units can exceed over 100 feet. The upper sand unit may be found in the glacial drift within a wide range of elevations. For mapping purposes, we defined the upper sand unit as a sand unit greater than 1 foot thick that occurs between two fine-grained layers. The basal sand unit includes all coarse-grained materials that overlie the bedrock surface. Most of the basal sands present west of the Des Plaines River were formed as valley train deposits along the river

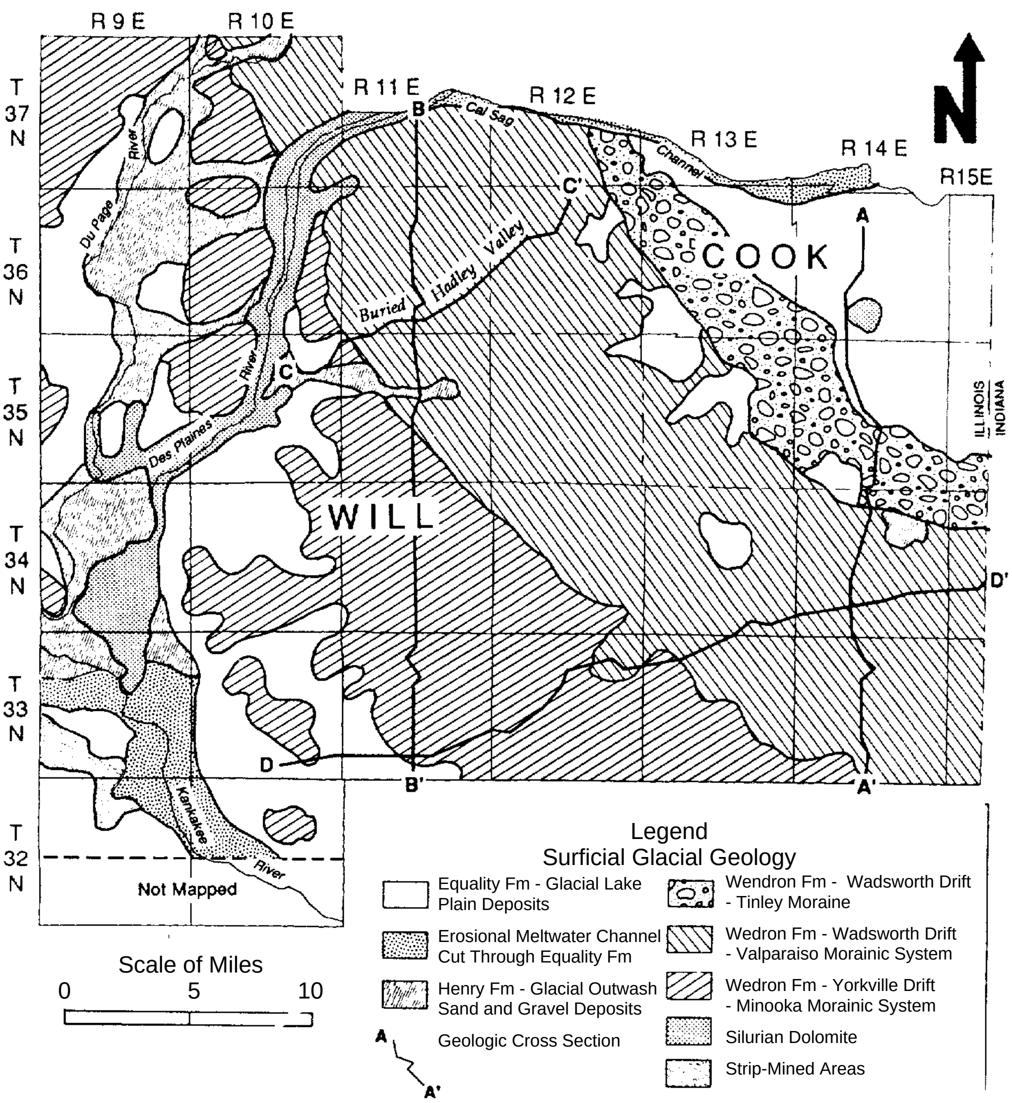


Figure 7. Surficial glacial geology (23).

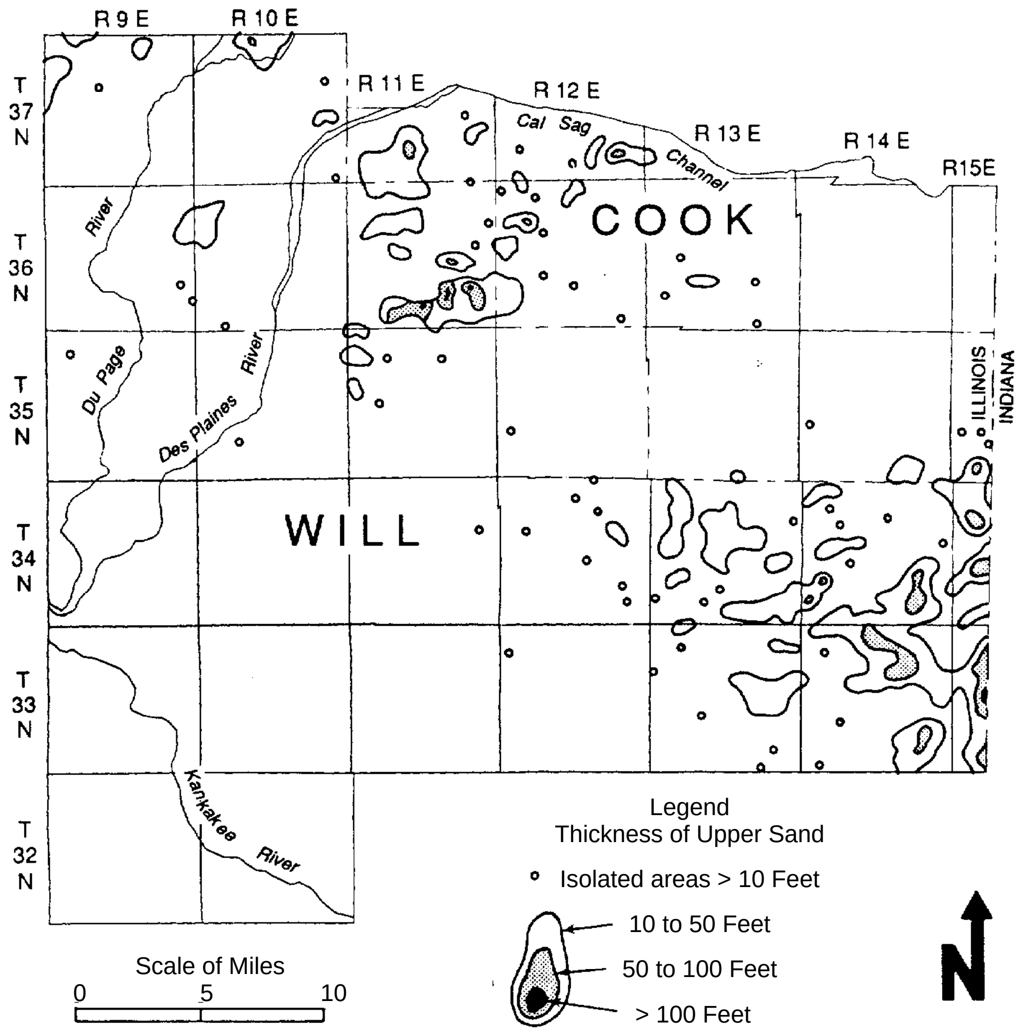


Figure 8. Thickness of the upper sand unit.

channels as the glaciers melted back. The origin of the extensive deposits underlying the Valparaiso Moraines is not clear. They may have been formed during early Wisconsinan glacial events as outwash plain deposits or they may have been deposited subglacially. The cross sections (see Figure 6) can reveal the variability and complexity of the sand and gravel layers as they occur within the drift. The sand and gravel deposits very seldomly act as aquifers in this region because almost all wells are completed in the Silurian dolomite aquifer. Clearly, Figures 8 and 9 indicate that some ground-water resource potential may exist within these deposits.

References

1. Thwaites, F.T. 1927. Stratigraphy and geologic structure of northern Illinois. Illinois State Geological Survey Report of Investigations 13.
2. Bretz, J.H. 1939. Geology of the Chicago region, Part I: General. Illinois State Geological Survey Bulletin 65.
3. Bergstrom, R.E., J.W. Foster, L.F. Selkregg, and W.A. Pryor. 1955. Ground-water possibilities in northeastern Illinois. Illinois State Geological Survey Circular 198.
4. Bretz, J.H. 1955. Geology of the Chicago region, Part II: The Pleistocene. Illinois State Geological Survey Bulletin 65.
5. Suter, M., R.E. Bergstrom, H.F. Smith, G.H. Emerich, W.C. Walton, and T.E. Larson. 1959. Preliminary report on the ground-water resources of the Chicago region, Illinois. Illinois State Geological Survey and Illinois State Water Survey Cooperative Ground-Water Report 1.
6. Hughes, G.M., P. Kraatz, and R.A. Landon. 1966. Bedrock aquifers of northeastern Illinois. Illinois State Geological Survey Circular 406.
7. Willman, H.B. 1971. Summary of the geology of the Chicago area. Illinois State Geological Survey Circular 460.
8. Bogner, J.E. 1976. Geology for planning in northeastern Illinois, Vol. V: Cook County. Illinois State Geological Survey Open-File Report.
9. Larsen, J.I. 1976. Geology for planning in northeastern Illinois, Part VI: Will County. Illinois State Geological Survey Open-File Report.

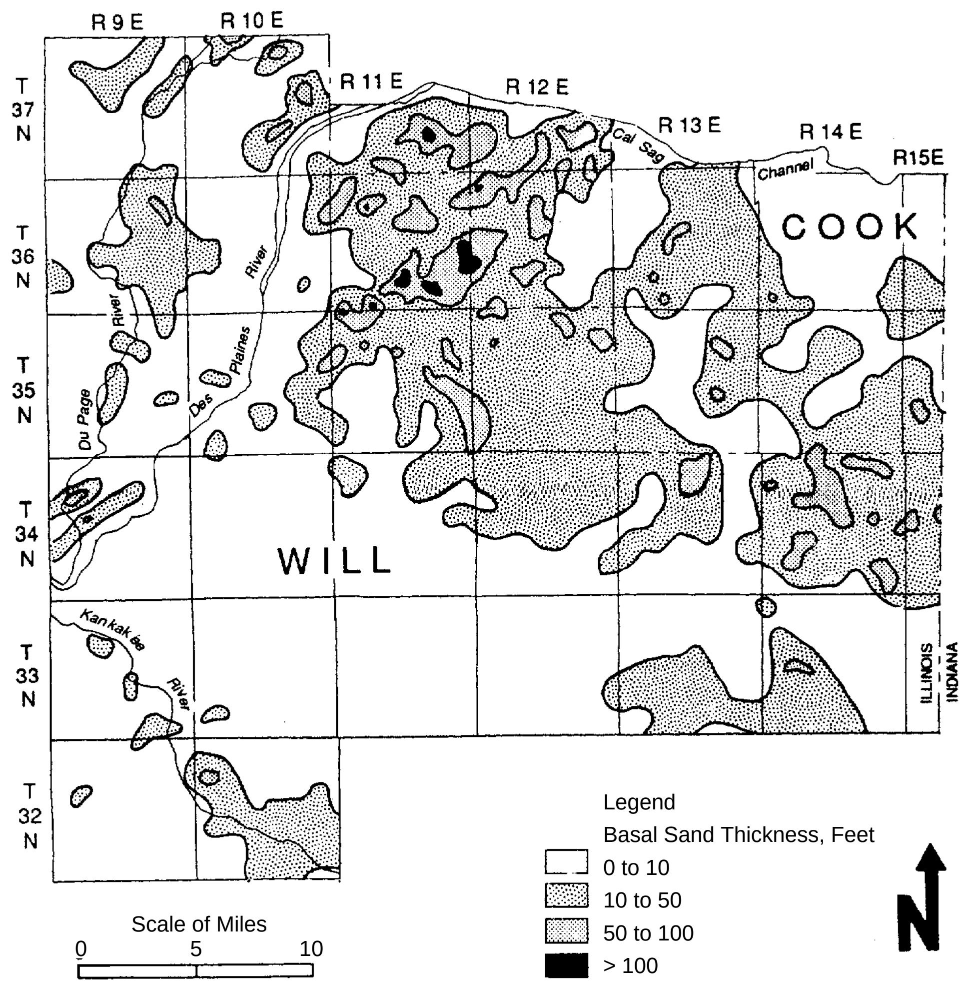


Figure 9. Thickness of the basal sand unit.

10. Dynamic Graphics, Inc. 1991. Interactive surface modeling users guide. Alameda, CA.

11. Ekblaw, G.E. 1938. Kankakee Arch in Illinois. Illinois State Geological Survey Circular 40.

12. Kolata, D.R., T.C. Buschbach, and J.D. Treworgy. 1978. The Sandwich Fault Zone of northern Illinois. Illinois State Geological Survey Circular 500.

13. Horberg, L. 1950. Bedrock topography in Illinois. Illinois State Geological Survey Bulletin 73.

14. Visocky, A.P., M.G. Sherrill, and K. Cartwright. 1985. Geology, hydrology, and water quality of the Cambrian and Ordovician systems in northern Illinois. Illinois Department of Energy and Natural Resources Cooperative Groundwater Report 10.

15. Kolata, D.R., and A.M. Graese. 1983. Lithostatigraphy and depositional environments of the Maquoketa group (Ordovician) in northern Illinois. Illinois State Geological Survey Circular 528.

16. Willman, H.B., et al. 1975. Handbook of Illinois stratigraphy. Illinois State Geological Survey Bulletin 95.

17. Willman, H.B. 1973. Rock stratigraphy of the Silurian system in northeastern and northwestern Illinois. Illinois State Geological Survey Circular 479.

18. Foote, G.R. 1982. Fracture analysis in northeastern Illinois and northern Indiana. Master's thesis, University of Illinois.

19. Smith, W.H. 1968. Strippable coal reserves of Illinois, Part 6: La Salle, Livingston, Grundy, Kankakee, Will, Putnam, and parts of Bureau and Marshall Counties. Illinois State Geological Survey Circular 419.

20. Horberg, L., and K.O. Emery. 1943. Buried bedrock valleys east of Joliet and their relation to water supply. Illinois State Geological Survey Circular 95.

21. McConnel, D.R. 1972. Bedrock topography and paleogeomorphology northeast of Joliet, Illinois. Master's thesis, University of Illinois.

22. Fischer, D.J. 1925. Geology and mineral resources of the Joliet quadrangle. Illinois State Geological Survey Bulletin 51.

23. Buschbach, T.C., and G.E. Heim. 1972. Preliminary geologic investigations of rock tunnel sites for flood and pollution control in the greater Chicago area. Illinois State Geological Survey Environmental Geology Notes 52.

24. Wilman, H.B., and J.C. Frye. 1970. Pleistocene stratigraphy of Illinois. Illinois State Geological Survey Bulletin 94.

25. Piskin, K. 1975. Thickness of glacial drift in Illinois. Illinois State Geological Survey 1:500,000 scale map.

26. Berg, R.C., and J.P. Kempton. 1988. Stack-unit mapping of geologic materials in Illinois to a depth of 15 meters. Illinois State Geological Survey Circular 542.

27. Hansel, A.K., and W.H. Johnson. 1987. Ice marginal sedimentation in a late Wisconsinan end moraine complex, northeastern Illinois, USA. In: van der Meer, J.J.M., ed. Tills and glaciotectonics. Proceedings of an INQUA Symposium on genesis and lithology of glacial deposits.

28. Horberg, L. 1953. Pleistocene deposits below the Wisconsinan drift in northeastern Illinois. Illinois State Geological Survey Report of Investigations 165.

29. Johnson, W.H., and A.K. Hansel. 1985. The Lemont section. In: Johnson, W.H., A.K. Hansel, B.J. Socha, L.R. Follmer, and J.M. Masters, eds. Depositional environments and correlation problems of the Wedron Formation (Wisconsinan) in northeastern Illinois. Illinois State Geological Survey Guidebook 16.

30. Gross, D.L., and R.C. Berg. 1981. Geology of the Kankakee River system in Kankakee County, Illinois. Illinois State Geological Survey Environmental Geology Notes 92.

31. Masters, J.M. 1978. Sand and gravel resources in northeastern Illinois. Illinois State Geological Survey Circular 503.

Watershed Applications

The Watershed Assessment Project: Tools for Regional Problem Area Identification

Christine Adamus
St. Johns River Water Management District, Palatka, Florida

The St. Johns River Water Management District of Florida recently completed a major water resources planning effort. As part of this planning effort, the St. Johns River Water Management District created a geographic information systems (GIS) project called the Watershed Assessment, which included a nonpoint source pollution load model. This paper introduces the planning project and the Watershed Assessment, and describes how the results of the model are being used to guide water management activities in northeast Florida.

Background

The St. Johns River Water Management District (District), one of five water management districts in Florida, covers 12,600 square miles (see Figure 1). The St. Johns River starts at the southern end of the District and flows north; it enters the Atlantic Ocean east of the city of Jacksonville. The cities of Orlando, Daytona Beach, and Jacksonville are partially or entirely within the District boundaries. Ad valorem taxes provide primary funding for the District.

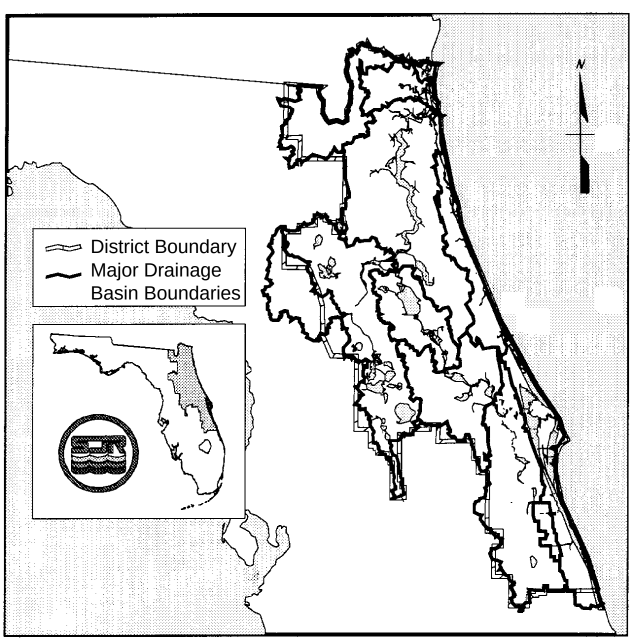


Figure 1. St. Johns River Water Management District, Florida.

The District boundaries are somewhat irregularly shaped because Florida water management districts are organized on hydrologic, not political, boundaries, which greatly improves the District's ability to manage the resources. On the north, the District shares the St. Mary's River with the state of Georgia and on the south, shares the Indian River Lagoon with another water management district. Most of the water bodies the District manages, however, have drainage basins that are entirely contained within the District's boundaries.

Water management districts in Florida have amassed extensive GIS libraries, which they share with local and statewide agencies. These libraries include basic data layers such as detailed land use, soils, and drainage basins. Districts also coordinate data collection and management to ensure data compatibility.

District Water Management Plan

All activities and programs of the water management districts are related to one or more of the following responsibilities: water supply, flood protection, water quality management, and natural systems management.

Each water management district recently completed a district water management plan (Plan). The main purpose of these Plans is to provide long-range guidance for the resolution of water management issues. The Florida Department of Environmental Protection will use these five Plans as the basis for a state water management plan. Each water management district used the same format, which comprised the following components:

- Resource assessment: What are the problems and issues related to each of the four responsibilities listed above?

- Options evaluation: What options are available for addressing the problems?
- Water management policies: What existing District policies influence the decisions that must be made?
- Implementation strategy: What is the best plan for addressing the problems?

The Watershed Assessment Project

The District created the Watershed Assessment project as part of is resource assessment. This GIS project examines the entire District to identify problems related to flood protection, ecosystems protection, and surface water quality.

The flood protection component is the only part of the Watershed Assessment that is not complete. It will involve simple overlays of floodplain boundaries with existing and future land use. Floodplain boundaries are defined as Federal Emergency Management Agency (FEMA) flood insurance rate map 100-year flood hazard areas. In many areas, these designations are not very accurate, yet we decided to proceed with their use because they are the best available information for many parts of the District. In areas where little hydrologic information is available and where the District has not conducted any related studies, the FEMA data are a helpful starting point. This echoes a theme of the Watershed Assessment project: the assessment is primarily intended to fill in gaps where we have not performed previous resource assessments, not to supplant existing information.

The ecosystems protection component of the Watershed Assessment is based heavily on a project identifying priority habitat in Florida, conducted by the Florida Game and Fresh Water Fish Commission (1). It is similar to gap analyses that the U.S. Fish and Wildlife Service currently is conducting in many parts of the country. For the Watershed Assessment, we modified the data somewhat and examined ways to protect the habitat in cooperation with local agencies.

The surface water quality component of the Watershed Assessment has two main parts. The first uses water quality data from stations that have been spatially referenced so that we can map them and combine the information with other information, such as the second part of the water quality component. This second part is a nonpoint source pollution load model, which is discussed in more detail below.

The Pollution Load Screening Model

The nonpoint source pollution load model is the Pollution Load Screening Model (PLSM), a commonly used screening tool in Florida. It is an empirical model that estimates annual loads to surface waters from stormwater runoff. Our goal in designing this model was to identify pollution load "problem areas" for examination in the Plan.

In these types of models, annual pollutant loads are a function of runoff volume and mean pollutant concentrations commonly found in runoff. Runoff volume varies with soil and land use, while pollutant concentrations vary with land use. For the PLSM, pollutant concentrations were derived from studies conducted solely in Florida. A report describing the model in detail is available (2).

Usually, this kind of model combines GIS with a spreadsheet: the GIS supplies important spatial information that is input into a spreadsheet where the actual calculations are made. The PLSM is different, however, because we programmed it entirely within GIS. The District's GIS software is ARC/INFO, and the model employs an ARC/INFO module called GRID, which uses cell-based processing and has analytical capabilities (3). All the model calculations are done in the GIS software, resulting in a more flexible model with useful display capabilities.

Model input consists of grids, or data layers, with a relatively small cell size (less than 1/2 acre). We chose this cell size based on the minimum mapping unit of the most detailed input data layer (land use) and the need to retain the major road features. The model has four input grids: land use, soils, rainfall, and watershed boundaries. For any given cell, the model first calculates potential annual runoff based on the land use, soil, and rainfall in that cell. It then calculates annual loads by applying land-use-dependent pollutant concentrations to the runoff.

For this model:

- Land use is from 1:24,000-scale aerial photography flown in 1988 and 1989. The model incorporates 13 land use categories.
- Soils are the Soil Conservation Service (SCS) SSURGO database, which corresponds to the county soil surveys. The PLSM uses the hydrologic group designation of each soil type.
- Rainfall was taken from a network of long-term rainfall stations located throughout the District.
- Watersheds were delineated by the United States Geological Survey (USGS) on 1:24,000-scale, 7.5-minute maps and digitized.

Model output consists of a runoff grid and six pollutant load grids. We calculated loads for total phosphorus, total nitrogen, suspended solids, biochemical oxygen demand, lead, and zinc. We chose these pollutants because reliable data were available and because they characterize a broad range of nonpoint pollution-generating land uses, from urban to agricultural. The model calculates runoff and loads for any point in space, allowing the user to see the spatial distribution of loads. An example of a total phosphorus load grid for one subbasin in the Jacksonville, Florida vicinity is shown in Figure 2.

The grids themselves provide a detailed view of model output. Model results can also be summarized by watershed, using the watershed boundary grid, and the information can be examined from a basinwide perspective.

We have applied PLSM results in other useful ways at the District. For example, District staff felt that previous sediment sampling sites were not appropriately located, so the District water quality network manager used model results to locate new sampling sites, focusing on problem areas as well as areas where we expect to see little or no nonpoint impact.

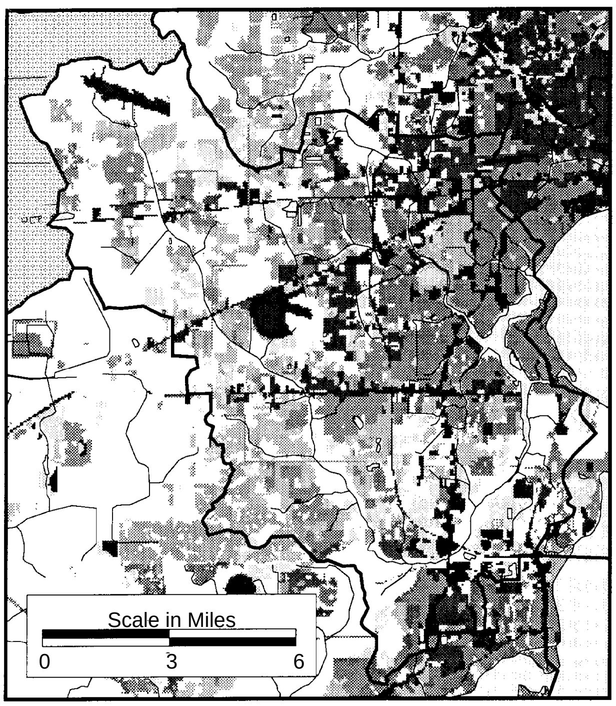


Figure 2. Distribution of total phosphorus loads, Ortega River subbasin (darker areas represent higher loads).

Application of Model Results in the Plan

Because the goal of the model was to identify potential stormwater runoff problem areas, we needed to simplify, or categorize, the model results for use in the Plan. We calculated the per acre watershed load for each pollutant and defined "potential stormwater runoff problem areas" as those individual watersheds with the highest loads for all pollutants. Problem areas for one major basin in the District, the lower St. Johns River basin, are depicted in Figure 3.

We also ran the model with future land use data obtained from county comprehensive plans. Because the

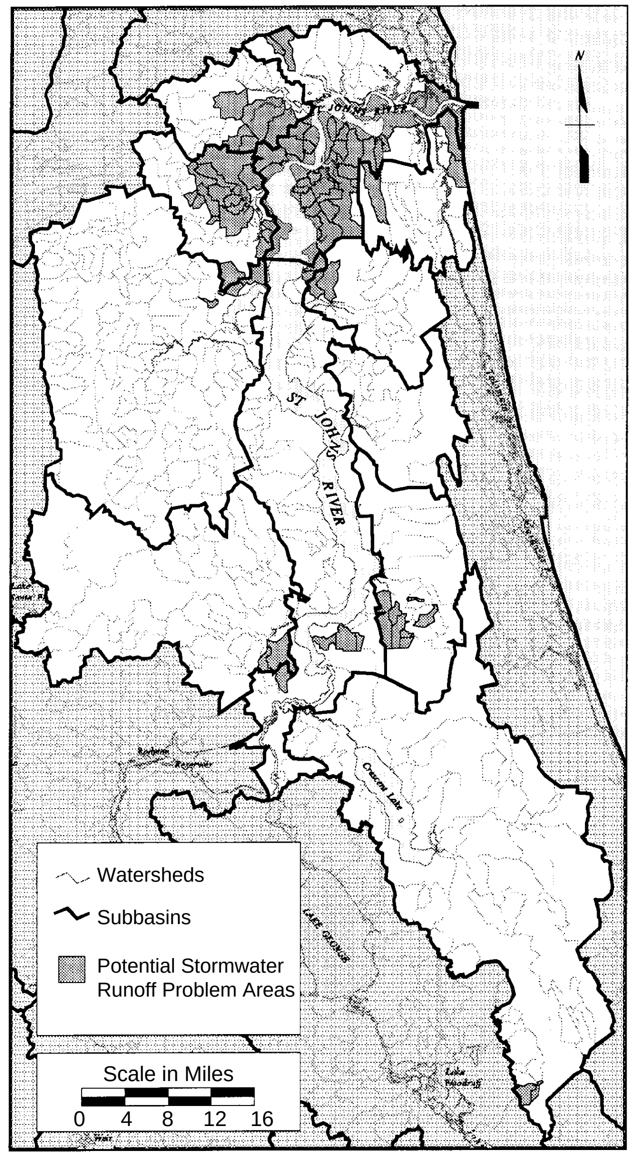


Figure 3. Potential stormwater runoff problem areas, lower St. Johns River basin.

county maps are guides to future development, and not predictions of actual development, we exercised caution when using the results. Problem areas were defined as those watersheds with projected loads greater than or equal to existing problem areas. Also, District planners combined model results with information about individual counties' regulations and policies to evaluate where problems are most likely to occur.

Prior to compiling the Plan, the District conducted workshops in each county in the District, in which problem areas identified by the PLSM were discussed with local agency staff, officials, and the public. We provided large, hard copy maps depicting stormwater runoff problem areas combined with results of a separate water quality analysis on county-based maps. These maps proved to be powerful tools for initiating discussions and gathering feedback.

In the Plan, stormwater runoff problem areas were reported for each of the 10 major drainage basins in the District. The information was also repackaged in a county-based format to create a quick reference for local agencies. District planners recommended strategies for addressing problems; these strategies vary as appropriate for each county. Examples include the need to assess compliance with existing stormwater permits, encourage stormwater reuse during the stormwater and consumptive use permitting processes, coordinate with municipalities that are implementing stormwater management plans, encourage and assist significantly affected municipalities to create stormwater utilities, and improve monitoring in problem areas that do not have sufficient water quality data.

In conclusion, the Watershed Assessment GIS project has proved to be useful not only to the St. Johns River Water Management District, but also to local governments. Large projects such as this could not be completed in a reasonable time without the use of GIS. Also, for ARC/INFO users who have been restricted to vector processing, the cell-based processing available in GRID is a powerful modeling tool.

References

1. Cox, J., R. Kautz, M. MacLaughlin, and T. Gilbert. 1994. Closing the gaps in Florida's wildlife habitat conservation system. Florida Game and Fresh Water Fish Commission, Office of Environmental Services, Tallahassee, FL.
2. Adamus, C.L., and M.J. Bergman. 1993. Development of a nonpoint source pollution load screening model. Technical Memorandum No. 1. Department of Surface Water Programs, St. Johns River Water Management District, Palatka, FL.
3. ESRI. 1992. Cell-based modeling with GRID. Redlands, CA: Environmental Systems Research Institute, Inc.

Watershed Stressors and Environmental Monitoring and Assessment Program Estuarine Indicators for South Shore Rhode Island

John F. Paul and George E. Morrison
Environmental Research Laboratory, U.S. Environmental Protection Agency, Narragansett, Rhode Island

Abstract

The U.S. Environmental Protection Agency has initiated the Environmental Monitoring and Assessment Program (EMAP), a nationwide ecological research, monitoring, and assessment program whose goal is to report on the condition of the nation's ecological resources. During the summers of 1990 through 1993, data were collected from approximately 450 sampling locations in estuarine waters of the Virginian Biogeographic Province (mouth of the Chesapeake Bay to Cape Cod). During this period, sampling stations were located in the coastal ponds and coastal area of south shore Rhode Island.

One objective of EMAP is to explore associations between indicators of estuarine condition and stressors in the watersheds of the sampled systems. Extensive watershed information for south shore Rhode Island is available in geographic information system (GIS) format. Watershed stressors along south shore Rhode Island were compared with EMAP indicators of estuarine conditions using GIS analysis tools. The indicator values for coastal EMAP stations (those offshore from coastal ponds) were associated with all of the aggregated south shore watershed stressors. The coastal pond indicator values were associated with stressors in the individual coastal pond watersheds. For the total south shore watershed, the major land use categories are residential and forest/brush land, followed by agriculture. Closer to the coast, residential land use is more prevalent, while further from the coast, forests/brush lands dominate. All coastal EMAP stations, with one exception, exhibited unimpacted benthic conditions, indicating no widespread problems. For the individual watersheds, the major land use categories are residential and forest/brush land. The population density (persons per square mile) shows an increasing trend from west to east. Impacted benthic conditions were observed at EMAP sampling sites in two coastal ponds. These two impacted benthic sites appear to be organically enriched.

Introduction

Since its inception in 1970, the U.S. Environmental Protection Agency (EPA) has had the responsibility for regulating, on a national scale, the use of individual and complex mixtures of pollutants entering our air, land, and water. The Agency's focus during this period centered primarily on environmental problems attributable to the use of individual toxic chemicals. Regulatory policy, while continuing to control new and historical sources of individual chemicals (i.e., "end of the pipe") and remediate existing pollution problems, will have to address the cumulative impacts from multiple stresses over large spatial and temporal scales.

In this decade, the focus of environmental problems, or "scale of concern," has shifted from point-source and local scales to regional and global scales. Concurrently, the focus has shifted from chemical to nonchemical stressors. The threat posed by nonchemical stresses (e.g., land use, habitat alteration and fragmentation, species loss and introduction) presents a substantial risk to the integrity of both specific populations and ecosystems, and entire watersheds and landscapes.

The shift in the scale of concern for environmental problems presents a unique challenge for environmental decision-making. Traditionally, environmental information has been collected over local spatial and short temporal scales, focused on addressing specific problems, limited in the number of parameters measured, and collected with a variety of sampling designs that were neither systematic nor probabilistic. It is not surprising, then, that several scientific reviews concluded that the information needed to assess, protect, and manage marine and estuarine resources was either insufficient or unavailable and recommended a national network of regional monitoring programs (1, 2). Two key recommendations resulted from these reviews: (1) the need for a national monitoring program designed to determine the status and trends of ecological resources, and (2) the need for an assessment framework for synthesizing and interpreting the information being produced in a timely manner and in a form that the public can understand and decision-makers can use. EPA's response to these recommendations was to institute a long-term monitoring program, the Environmental Monitoring and Assessment Program (EMAP), and to adopt a risk-based strategy for decision-making.

EMAP is a nationwide ecological research, monitoring, and assessment program whose goal is to report on the condition of the nation's ecological resources. During the summers of 1990 through 1993, data were collected from approximately 450 sampling locations in estuarine waters of the Virginian Biogeographic Province (mouth of the Chesapeake Bay to Cape Cod) (3-5). During this period, some of the sampling stations were located in the coastal ponds and coastal area of south shore Rhode Island. One objective of EMAP is to explore associations between indicators of estuarine condition and stressors in the watersheds of the sampled systems. Extensive watershed information for south shore Rhode Island is available in geographic information systems (GIS) format.

The intent of this paper is to compare watershed stressors with EMAP indicators of estuarine condition along south shore Rhode Island using GIS analysis tools. The indicator values for coastal EMAP stations (those offshore from coastal ponds) are associated with all of the aggregated south shore watershed stressors. The coastal ponds indicator values are associated with stressors in the individual coastal pond watersheds. The project reported on in this paper served as a pilot for integrating watershed information with wide-scale ecological data collected to assess condition of estuarine waters.

Ecological Risk Assessment Context

Robert Huggett, EPA's Assistant Administrator for Research and Development, is using the risk assessment-risk management paradigm as a framework to reorganize the EPA research laboratories (6). Huggett is also reorienting the research that EPA conducts to be risk based (both human and ecological). The major thrust of the research to be conducted in the EPA laboratories will be directed toward reducing the uncertainties in the risk assessment process. In this way, the risk assessment context provides the "why" for the research conducted.

Ecological risk assessment is defined as a process for evaluating the likelihood that adverse ecological effects have occurred, are occurring, or will occur as a result of exposure to one or more stressors (7). The value of the risk assessment framework lies in its utility as a process for ordering and analyzing exposure and effects information, and in its flexibility for describing past, present, and future risks.

One way of depicting the ecological risk assessment process is shown in Figure 1 (8). The key points are that the process is continuous; the process can be oriented in either direction, dependent upon the form of the question or issue being addressed; and monitoring is at the hub, providing information to all activities. The end result of the effort is to provide better information for making environmental management decisions.

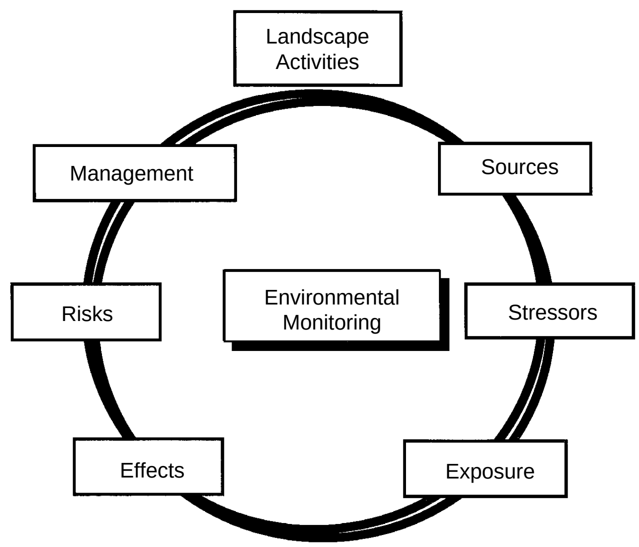


Figure 1. Ecological risk assessment framework (8).

Overview of EMAP and Estuarine Results

EMAP has been described as an approach to ecological research, monitoring, and assessment (9). It is not the only approach but is an approach that is driven by its goal to monitor and assess the condition of the nation's ecological resources. The objectives of the program to address this goal are to:

- Estimate the current status, trends, and changes in selected indicators of the condition of the nation's ecological resources on a regional basis with known confidence.
- Estimate the geographic coverage and extent of the nation's ecological resources with known confidence.
- Seek associations among selected indicators of natural and anthropogenic stress and indicators of ecological condition.
- Provide annual statistical summaries and periodic assessments of the nation's ecological resources.

The approach used by the program to meet its objectives and address its goal includes:

- Use of a large, regional scope that encompasses the entire county but provides information on the scale that federal and regional environmental managers require.
- Emphasis on ecological indicators to provide the information to assess condition (i.e., collect information on the ecological systems themselves to determine their condition or "health").
- A probability-based sampling design to produce statistically unbiased estimates on condition and to provide uncertainty bounds for these estimates.
- A vision of the program as long-term, continuing into the next century, which is consistent with the large, regional spatial scale being addressed.
- Development through partnerships with other agencies that have natural resource stewardship responsibility.

The estuarine component of EMAP was initiated in 1990, with monitoring in the estuarine waters of the Virginian Biogeographic Province (mouth of Chesapeake Bay northward to Cape Cod) (10). Figure 2 depicts the biogeographic provinces of estuarine resources of the country. These provinces have been delineated based upon major climatic zones and the prevailing offshore currents (11). This is comparable with the ecoregion approach used to describe the distribution of terrestrial ecosystems (12). The biogeographic province is the comparable approach for coastal ecosystems. Monitoring in the Virginian Province continued through 1993; monitoring was conducted in the Louisianian Province from 1991 to 1994; monitoring was initiated in the Carolinian Province in 1994; and monitoring will be initiated in the West Indian Province in 1995.

A suite of measurements was collected at each of the EMAP-Estuaries sampling sites that were selected with a probability-based sampling design (13, 14). As indicated above, the measurements emphasized ecological conditions indicators, which included biotic indicators such as benthic and fish abundance, biomass, diversity, and composition, and also included abiotic indicators such as dissolved oxygen, sediment contaminant concentration, and sediment toxicity (15).

In the Virginian Province, approximately 450 probability-based sampling sites were visited during the summer periods in 1990 through 1993 using consistent indicators and collection and analysis procedures. An example of the results is shown in Figure 3, which presents the condition of benthic resources (16). The benthic condition is reported using a benthic index, which is an aggregate of individual benthic measurements that were combined using discriminant analysis to differentiate impacted from unimpacted sites (3, 17). The figure presents results for values of the benthic index that were determined to be impacted. The bar chart is the standard EMAP format for results: province-scale results with 95-percent confidence intervals about estimates. The large, small, and tidal categories refer to the strata used in the probability-based sampling design: large systems are the broad expanses of water such as in Chesapeake Bay, Delaware Bay, and Long Island Sound; small

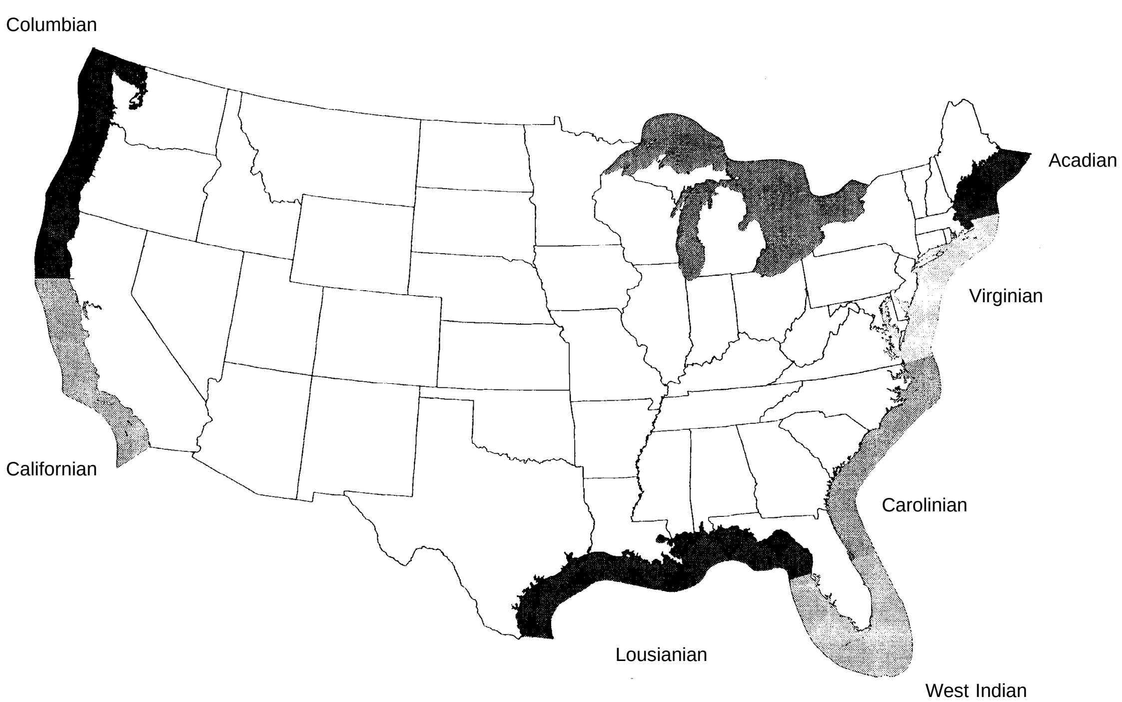


Figure 2. Biogeographic provinces used by EMAP-Estuaries (13).

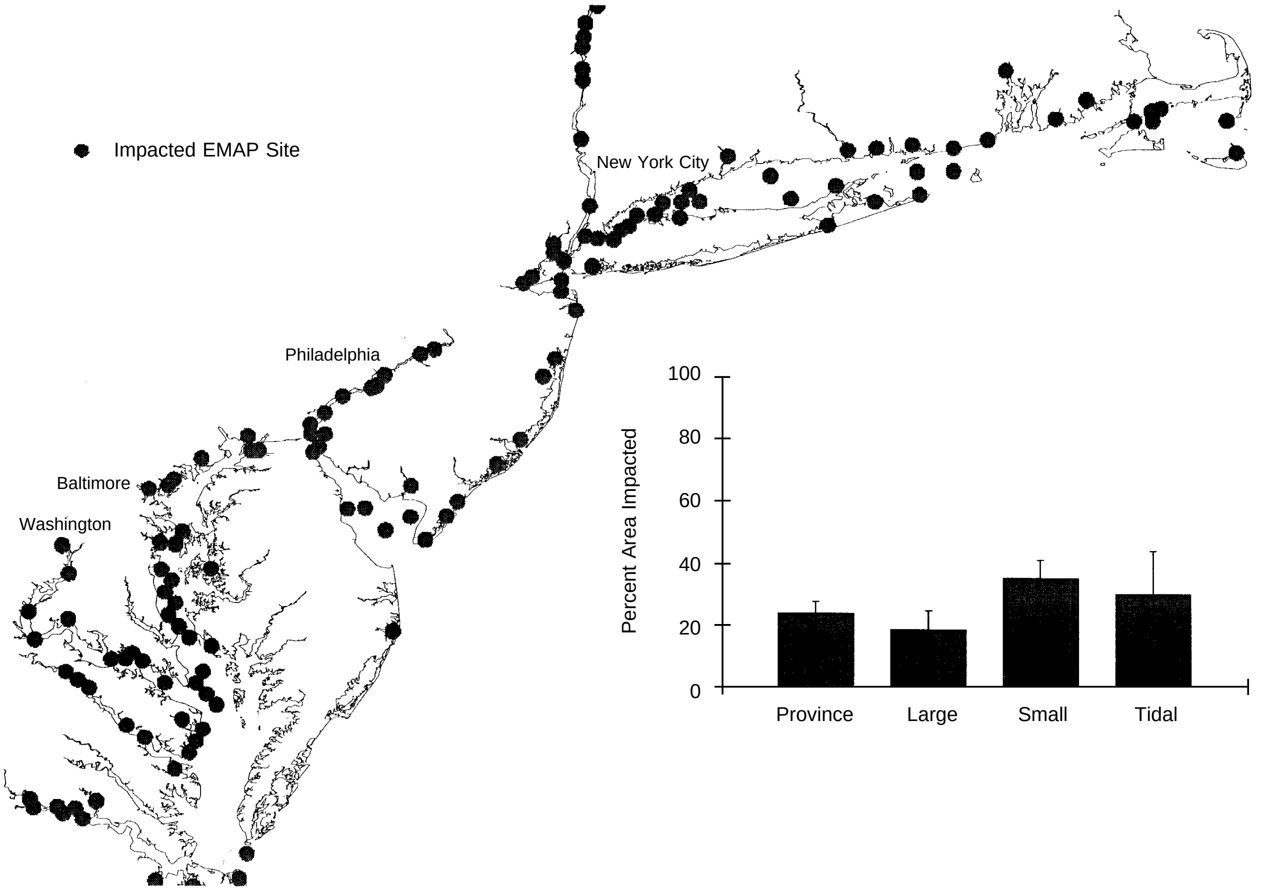


Figure 3. Condition of benthic communities in Virginian Province.

systems include the bays and harbors along the edges of the major systems and embayments along the coast; and large tidal rivers include the Potomac, James, Rappahannock, Delaware, and Hudson Rivers.

The results indicate that 24 percent ± 4 percent of the estuarine waters of the Virginian Province have impacted benthic communities. The small and tidal river systems have proportionately more impacted area than the large systems.

All the EMAP data are geographically referenced; therefore, the data can be spatially displayed to explore patterns. The spatial display of the impacted benthic community information is a simple spatial analysis of the EMAP data. This analysis shows that the impacted benthic resources are distributed across the entire province, with more impacted sites in the vicinity of the major metropolitan areas.

In addition to analyzing the EMAP results at the regional scale, analyses have been conducted at the watershed scale (see Figure 4). The probability-based sampling design permits the data to be aggregated (poststratified) in ways other than the way the original design was stratified. The only restriction to the aggregation is the number of available sample sites for the aggregation; a small number of sites leads to large uncertainties in the results. Figure 4 shows the aggregation for four major watersheds: Chesapeake Bay, Delaware Bay, Hudson-Raritan system, and Long Island Sound. This watershed scale is close to the practical scale at which environmental management decisions are implemented. The data need to be analyzed at smaller scales, however, to focus on environmental management of the smaller watersheds (e.g., contaminated sediments). This leads into the need to conduct the pilot project addressing watershed information.

South Shore Rhode Island Pilot Project

EMAP's third objective relates to exploring associations between indicators of estuarine condition and watershed stressors. Note that the word "watershed" was added. One way to address environmental management remediation strategies is to look at the watershed activities that could possibly be modified or changed to improve estuarine conditions.

Watershed stressors for the estuarine environment include land-based sources of pollution, such as point sources of pollution, and land use activities (i.e., how the land is actually used, including landscape patterns).

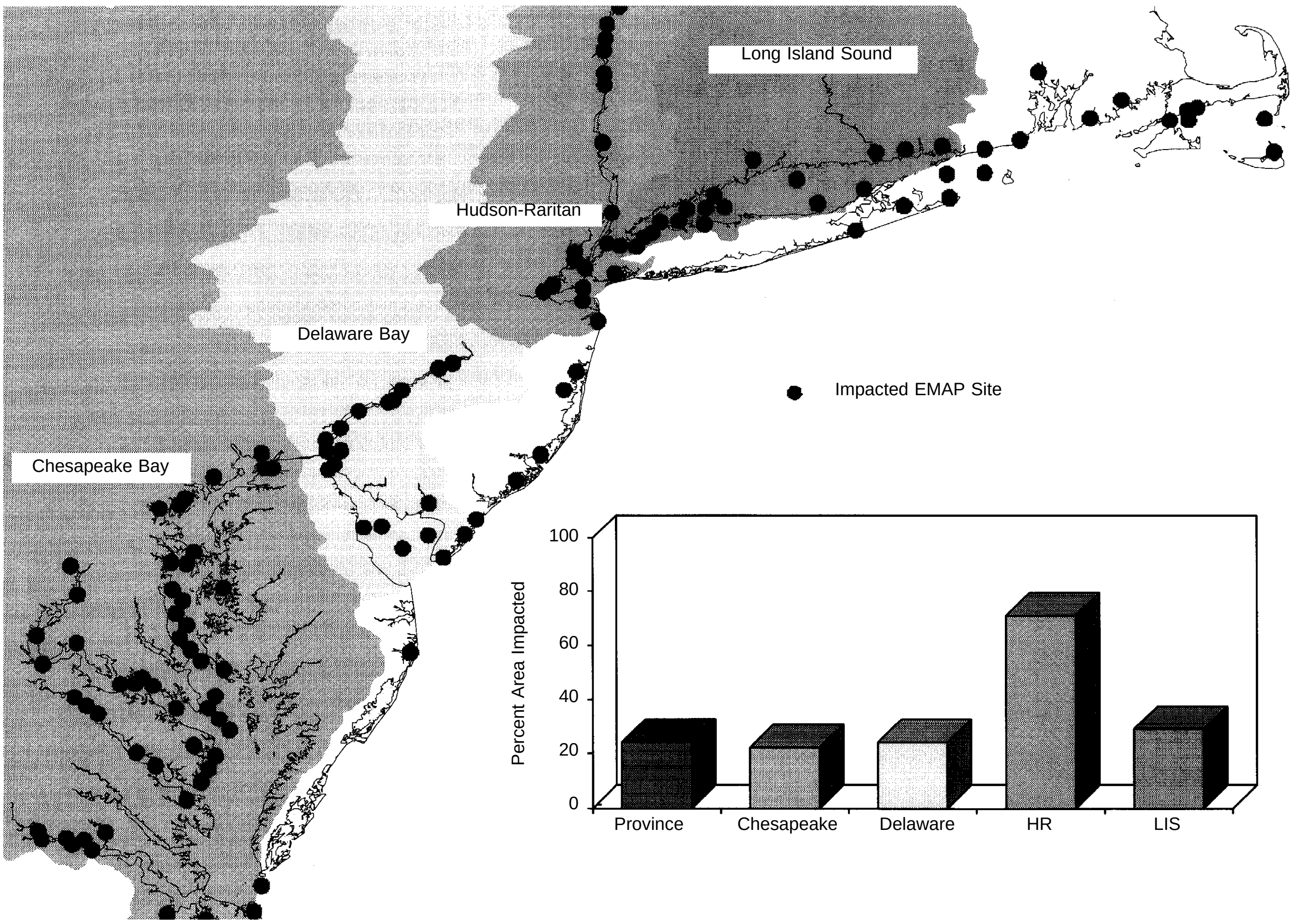


Figure 4. Condition of benthic communities in major watersheds in Virginian Province (Chesapeake Bay, Delaware Bay, Hudson-Raritan system, and Long Island Sound).

Which of these stressors is more important for a particular situation depends on the types of estuarine impact (localized or systemwide) and the management question that is being addressed.

The specific objective of the south shore Rhode Island pilot project was to compare watershed stressors with EMAP indicators of estuarine condition using GIS analysis tools. This project was not intended to be a definitive study by itself of south shore Rhode Island but to explore the process necessary to undertake the comparisons, to investigate the feasibility of pulling the necessary information together, and to identify potential problems before undertaking this comparison on a much larger scale.

The south shore Rhode Island study area is depicted in Figure 5. This coastal area drains into the coastal waters of Block Island Sound. The project was intentionally restricted to a limited geographic area to avoid being overwhelmed with the tremendous volumes of data that could have been encountered.

All data sources used in this project were available electronically. Digitized U.S. Geological Survey quad maps were available from the Rhode Island Geographic Information System (RIGIS) at the University of Rhode Island. The National Pollutant Discharge Elimination System (NPDES) was available for major dischargers from the National Oceanic and Atmospheric Administration's National Coastal Pollution Discharge Inventory (18). The 1990 census was also available from RIGIS. The EMAP 1990 through 1993 estuarine data were available from the EMAP-Estuaries Information System at the EPA Environmental Research Laboratory in Narragansett, Rhode Island. The RIGIS data were already available as ARC/INFO coverages. The NPDES and EMAP data had to be converted to ARC/INFO point coverages.

Two approaches were used to conduct spatial analyses. Buffer zones at 1, 3, 5, 10, and 20 kilometers from the south coast of Rhode Island were created and used to clip the south shore area coverages (e.g., land use, population, point sources). The watershed boundaries of three south shore coastal ponds were manually delineated and used to clip the south shore area coverages. The ponds were Quonochontaug, Ninigret, and Point Judith (west to east).

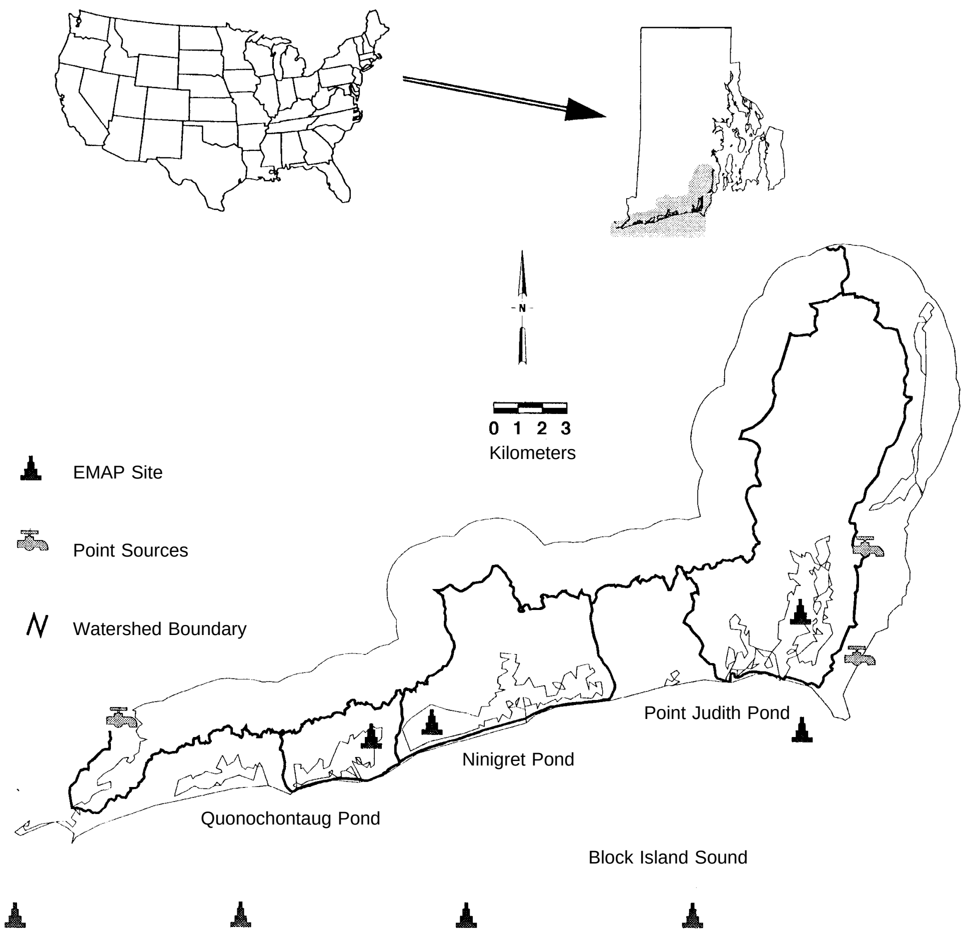


Figure 5. South shore Rhode Island study area.

South Shore Rhode Island Pilot Project Results

The results for land use by distance from the coast are presented in Figure 6. These results give the broad-scale coastal perspective. For the total south coast watershed, the major land use categories are residential and forest/brush land, followed by agriculture. Closer to the coast, residential land use is more prevalent, while farther from the coast, forests/brush lands dominate. Population (see Figure 7) increases with distance from the coast, but population density does not appear to be a function of distance from the coast. Only one out of five coastal EMAP stations exhibited impacted benthic conditions, indicating no widespread benthic problems in the coastal waters. The one station that was classified as impacted was dominated by an extremely high number of individuals of one species.

A smaller scale view can be gained by looking at the results for the individual watersheds. This view also provides an east-west perspective compared with the south-north perspective with the distance from the coast. Again, for the individual watersheds, the major land use categories are residential and forest/brush land (see Figure 8). The population increases from west to east, and population density shows an increasing trend from west to east (see Figure 9). Impacted benthic conditions were observed at the EMAP stations in Quonochontaug and Point Judith Ponds. These stations exhibited organic enrichment (total organic carbon in the sediments exceeding 2 percent), possibly from historically improperly treated sewage. No benthic data were available for the Ninigret Pond station; however, dissolved oxygen was observed to be low at this station. No major NPDES point sources are located in the coastal pond watersheds, although two are located on the eastern edge of the Point Judith Pond watershed boundary.

Discussion

A pilot project was conducted for south shore Rhode Island to compare watershed stressors with EMAP indicators of estuarine condition. The results indicate that such a comparison can be accomplished, with the watershed information providing a qualitative link to the estuarine conditions observed. One potential problem is the need to delineate the watershed boundaries for all

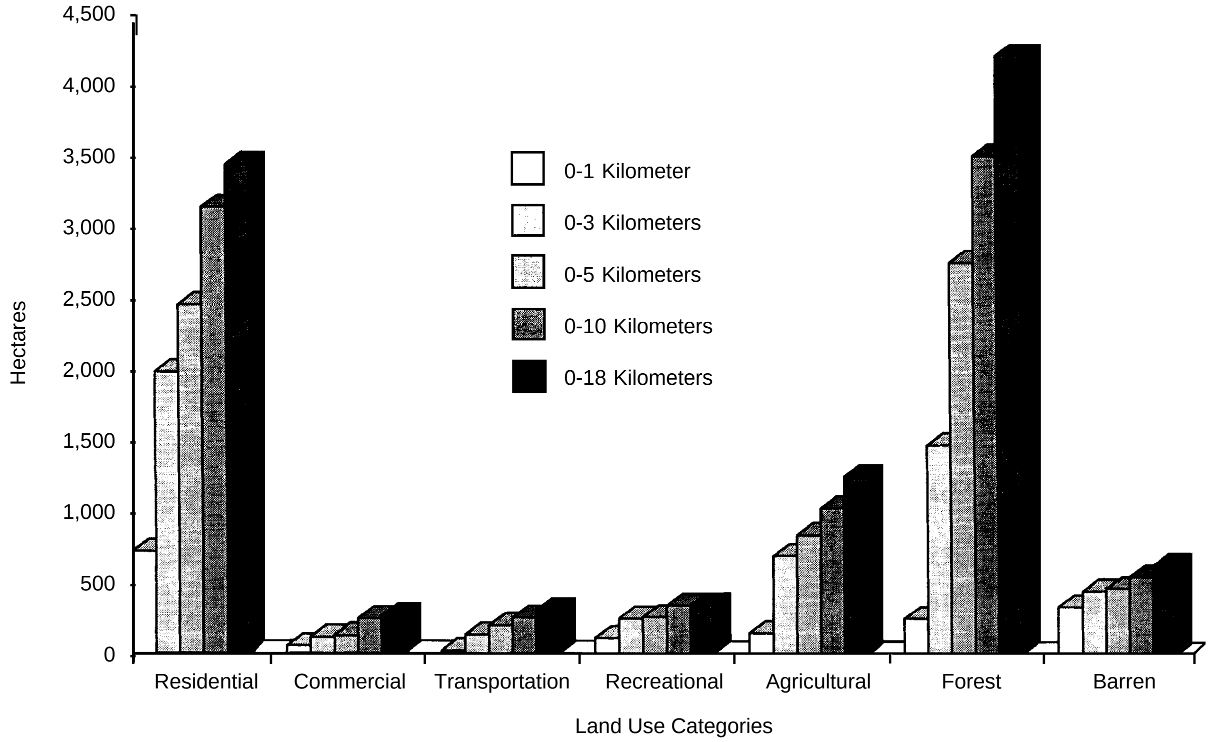


Figure 6. South shore Rhode Island land use by distance from south coast.

watersheds for which EMAP data are available. ARC/INFO provides tools for doing this, but practical application indicated that difficulties are encountered when the topographic relief is relatively flat, and on-screen corrections needed to be applied (19).

A restriction that needs to be understood before applying the procedures used in this project to a much wider geographic area is that the data sets for the watershed stressors need to be available over the wider geographic area. Further, these data sets need to be temporally consistent and constructed with consistent methods and land use classification schemes.

This project was conducted with only a small number of actual EMAP sampling sites, particularly for the individual watersheds. Because of this restriction, no statistical analyses were conducted with the EMAP data for comparison with the watershed information. Only qualitative comparisons were attempted. The next step is to increase the number of individual watersheds so that a rigorous statistical analysis can be conducted. This is being conducted by Comeleo et al. for comparing watershed stressors for subestuary watersheds in the Chesapeake Bay with estuarine sediment contamination (19). The steps after this will be to (1) apply the techniques to the entire EMAP Virginian Province data

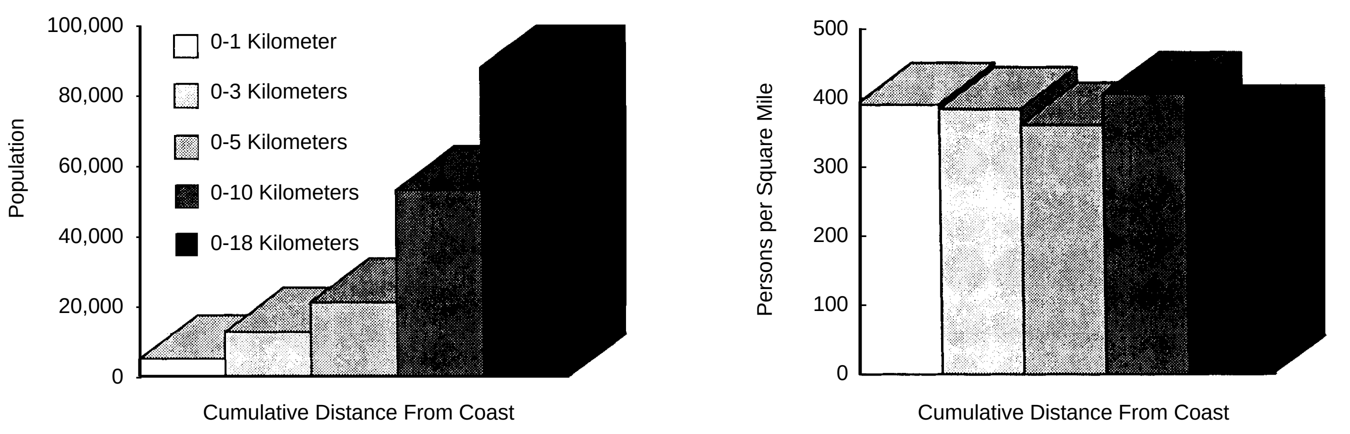


Figure 7. South shore Rhode Island population by distance from south coast.

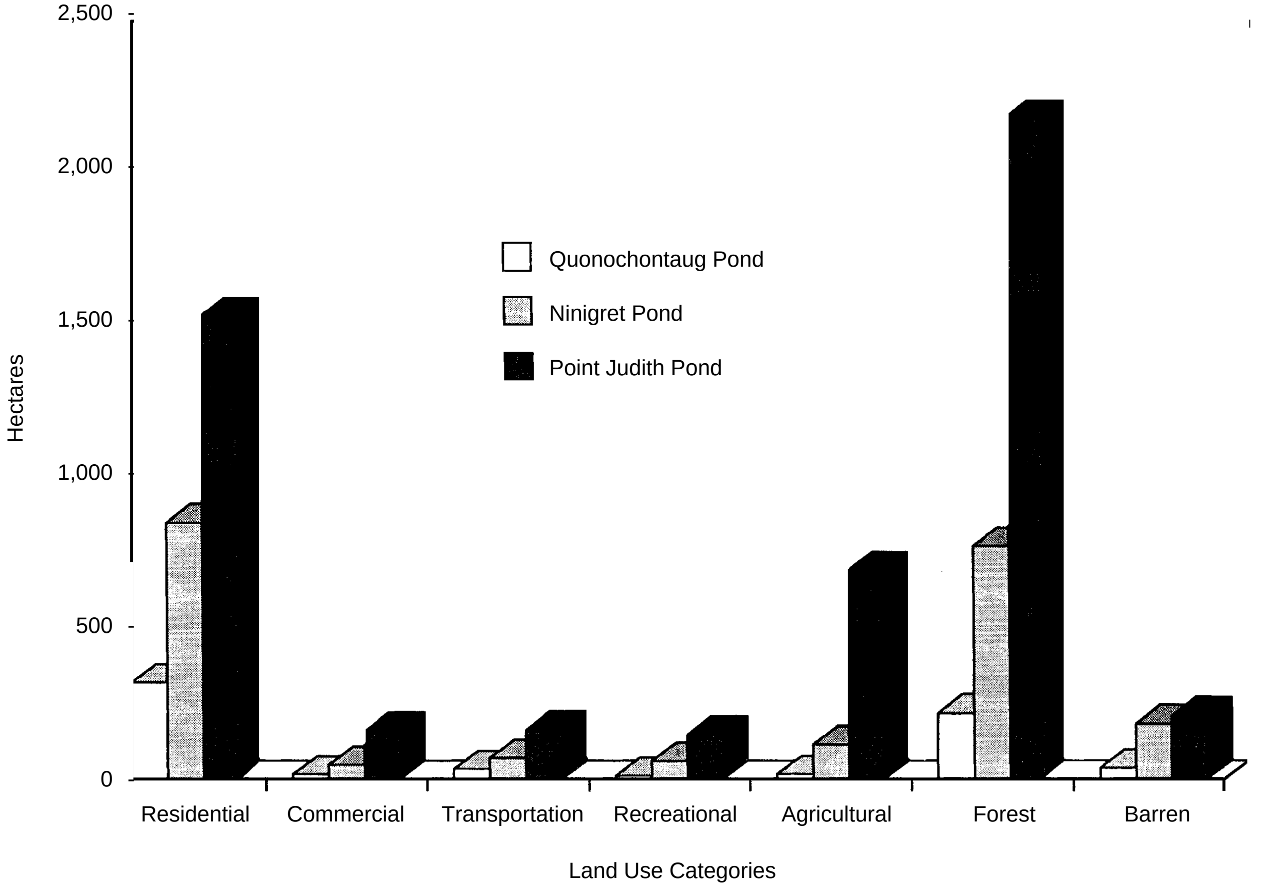


Figure 8. South shore Rhode Island land use by individual watershed.

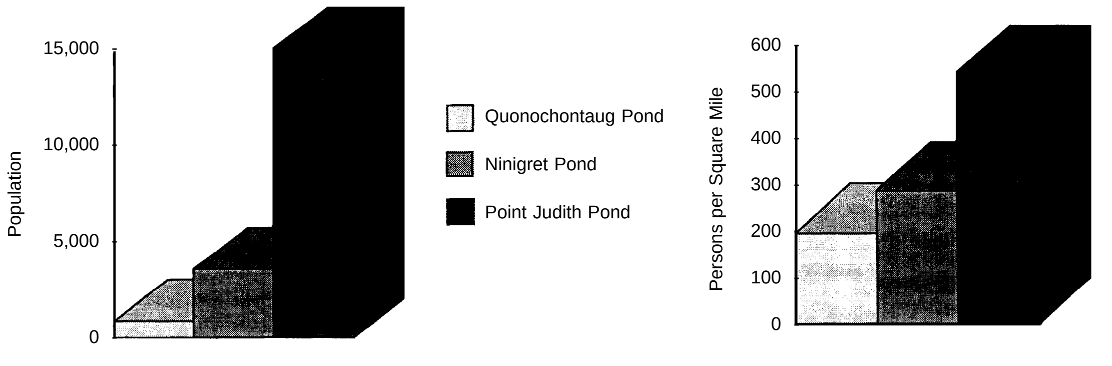


Figure 9. South shore Rhode Island population by individual watershed.

set, and (2) relate the watershed stressor information to estuarine benthic condition.

Acknowledgments

We thank S. Hale, J. Pesch, J. Copeland, and D. Keith for their review of this paper. This is contribution No. 1648 of the Environmental Research Laboratory-Narragansett.

References

1. SAB. 1988. Future risk: Research strategies for the 1990s. Science Advisory Board Research Strategies Committee, U.S. EPA, Washington, DC.
2. NRC. 1990. Managing troubled waters: The role of marine environmental monitoring. Washington, DC: National Academy Press.
3. U.S. EPA. 1993. Virginian Province demonstration project report, EMAP-Estuaries (1990). EPA/620/R-93/006. Washington, DC.

4. U.S. EPA. 1994. Statistical summary: EMAP-Estuaries Virginian Province (1991). EPA/620/R-94/005. Narragansett, RI.

5. U.S. EPA. 1994. Statistical summary: EMAP-Estuaries Virginian Province (1992). EPA/620/R-94/019. Narragansett, RI.

6. Newman, A. 1995. A new direction for EPA's Office of Research and Development. Environ. Sci. and Technol. 29(3):126A-129A.

7. U.S. EPA. 1992. Framework for ecological risk assessment. EPA/630/R-92/001. Washington, DC.

8. Gentile, J.H., J.F. Paul, K.J. Scott, and R.A. Linthurst. 1995. The role of regional monitoring in ecological risk assessment: An EMAP example for watersheds. Manuscript in review.

9. Messer, J.J., R. Linthurst, and W. Overton. 1991. An EPA program for monitoring ecological status and trends. Environ. Monitor. and Assessment 17(1):67-78.

10. Paul, J.F., K.J. Scott, A.F. Holland, S.B. Weisberg, J.K. Summers, and A. Robertson. 1992. The estuarine component of the U.S. EPA's Environmental Monitoring and Assessment Program. Chemistry and Ecology 7:93-116.

11. Bailey, R.G. 1983. Delineation of ecosystem regions. Environ. Mgmt. 7:365-373.

12. Omernik, J.M. 1987. Ecoregions of the conterminous United States. Annals of the Assoc. of Amer. Geographers 77:118-125.

13. U.S. EPA. 1990. Near coastal program plan for 1990: Estuaries. EPA/600/4-90/033. Narragansett, RI.

14. U.S. EPA. 1991. Design report of the Environmental Monitoring and Assessment Program. EPA/600/3-91/053. Chervils, OR.

15. U.S. EPA. 1990. Environmental Monitoring and Assessment Program: Ecological indicators. EPA/600/3-90/060. Washington, DC.

16. U.S. EPA. EMAP-Virginian Province four-year assessment (1990-1993). EPA report in review.

17. Engle, V.D., J.K. Summers, and G.R. Gaston. 1994. A benthic index of environmental condition of Gulf of Mexico estuaries. Estuaries 17(2):372-384.

18. NOAA. 1993. Point-source methods document. NOAA technical report. Silver Spring, MD: Strategic Environmental Assessments Division, National Ocean Service, National Oceanic and Atmospheric Administration.

19. Comeleo, R.L., J.F. Paul, P.V. August, J. Copeland, C. Baker, S. Hale, and R.W. Latimer. 1995. Relationships between landscape stressors and sediment contamination in Chesapeake Bay estuaries. Manuscript in review.

GIS Watershed Applications in the Analysis of Nonpoint Source Pollution

Thomas H. Cahill, Wesley R. Horner, and Joel S. McGuire
Cahill Associates, West Chester, Pennsylvania

Introduction

Geographic information systems (GIS) have been used to evaluate the impact of nonpoint source (NPS) pollution in a variety of watersheds and drainage systems over the past 20 years (1-6). During that period, our understanding of the sources and hydrologic transport mechanisms of NPS pollutants, both in particulate and soluble forms, has greatly increased (7-9). Our ability to create and manipulate land resource data, however, has advanced at a far more dramatic rate. Whereas 20 years ago, both computer system capabilities and peripheral hardware limited the process of encoding, storing, and displaying spatial data, today we can encode land resource data, analyze it, and produce stunning visual displays at a relatively low cost.

The question is: what has this experience told us regarding the yet unresolved problem of water quality degradation from NPS pollution in our streams, lakes, and coastal waters (10, 11)?

The purpose of this paper is to report on several recent studies of this nature that created a GIS as a tool to analyze NPS pollution. This paper will not cover all aspects of these studies; detailed reports on each project are available from the authors or respective clients. The objects of these studies were:

- A medium-sized lake draining a fairly small watershed
- A riverine system with multiple use impoundments
- A 100-mile stretch of Atlantic coastal estuary

These water bodies all have one common ingredient: NPS pollution significantly affects them. While the primary focus of these studies was to understand the dynamics of surface water quality, and specifically the NPS component, the further objective was to document the causal link between identified water resource problems and the watershed-wide management actions needed for their remediation. Thus, GIS serves not only as a mechanism for analysis of NPS pollution sources but also as the tool by which to evaluate alternative methods that would reduce or prevent this pollution.

Study Concepts

These three studies illustrate different approaches to both aspects of this problem. In the 93-square-mile Upper Perkiomen Creek watershed (UPW) study, the objective was to develop a management program that would reduce nutrient load in a system of reservoirs at the base of the watershed. An essential element in the analysis underlying GIS design (ARC/CAD) was to be able to differentiate and evaluate pollution sources in the watershed, while providing the technical basis for an innovative and far-reaching management program on all levels of government; that is, GIS was used not only to analyze the problem but to help formulate the solution.

In the more focused Neshaminy Creek study, Cahill Associates (CA) designed a detailed pixel/raster format for GIS to support detailed hydrologic modeling (12) and NPS loading analysis. This study, carried out under Pennsylvania's Act 167 stormwater management program, was under a legal requirement to translate technical findings into subdivision regulations that all 30 watershed municipalities would adopt. This mandate required much more geographically specific rigor in the GIS approach and in the management recommendations the law stipulated.

These two projects (see Figure 1), when taken together, illustrate the critical relationship between understanding the appropriate level of detail in GIS system design, GIS development with modeling and other analytical requirements, and ultimately, the proposed management actions for watershed-wide implementation.

In the New Jersey Atlantic Coastal Drainage (ACD) study, the objective was to document more completely the magnitude and sources of NPS pollutants, especially nutrients, entering New Jersey estuarine coastal waters. The GIS design placed special attention on the role of urban or developed land uses situated along the coastal fringe, particularly the maintained or landscaped portions of developed sites. Most previous studies have largely ignored this factor. Instead, they have focused water quality analysis typically on NPS loadings as a

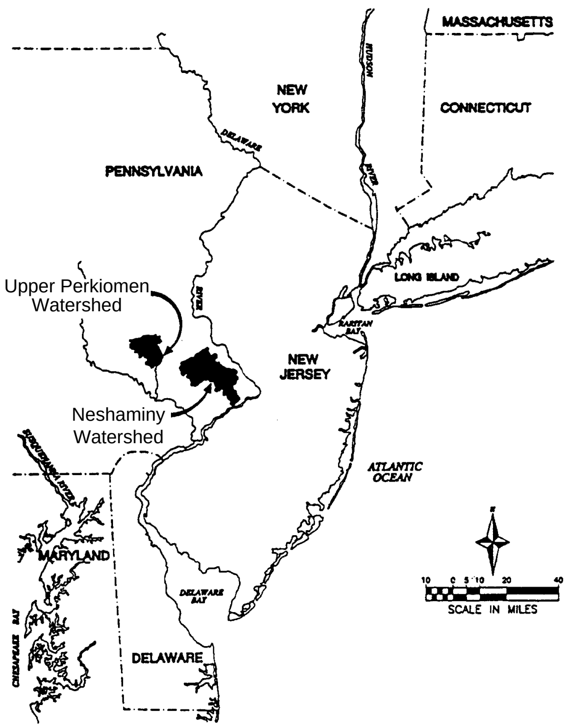


Figure 1. Regional location of Upper Perkiomen and Neshaminy basins.

function of impervious area coverage, with the assumption that loadings increase as imperviousness increases.

On the contrary, the CA thesis states that certain pollutant loadings, such as nutrients, maximize in areas with relatively moderate densities (1/2- to 1-acre lots) and percentage impervious cover but with large maintained lawnscapes. Because sandy soils allow soluble NPS pollutants to pass as interflow to points of surface discharge with surprising ease, they exacerbate the problem of nutrient applications in typical coastal drainage areas. GIS application in this case enabled estimation of the nutrient loading to coastal waters. Existing fertilized lawn areas were calculated to be a significant source of nutrient pollution, with loadings from new land development posited as an even more serious problem for New Jersey's coastal waters. GIS was then applied to evaluate the suitability of various best management practices (BMPs), based on the physical and chemical properties of the soil mantle and the existing and anticipated land use.

The Upper Perkiomen Creek Watershed Study

Background

The UPW in southeastern Pennsylvania is a tributary of the Schuylkill River in the Delaware River basin (see Figure 2). Serious eutrophication problems occurring in the system of reservoirs lying at the base of this relatively rural watershed prompted the study. The study

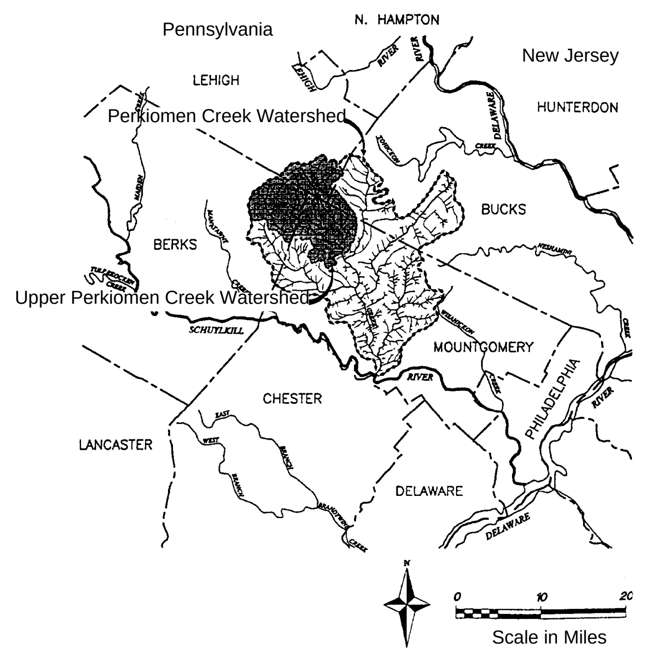


Figure 2. The Perkiomen Creek watershed in the Delaware River basin.

effort evolved from concerns on the part of the Delaware Riverkeeper, a private nonprofit environmental organization dedicated to promoting the environmental well-being of the Delaware River watershed. The Upper Perkiomen Creek has experienced various water quality problems, especially the eutrophication of Green Lane Reservoir, a large raw water supply storage reservoir (see Figure 3). Green Lane's highly eutrophic condition has been a constant since shortly after initial construction over 35 years ago, but the relative importance of NPS inputs has dramatically increased. Whereas 10 years ago point source input was the major source of phosphorus, elimination of some point sources and advanced waste treatment for others has greatly reduced that component of pollutant loading, while NPS sources have remained constant or increased. Current analysis indicates that NPS pollution constitutes over 80 percent of the annual load of phosphorus (see Figure 4) into the Green Lane Reservoir and is well in excess of the desired loading to restore water quality (see Figure 5).

Nonpoint Source Analysis

Calculating the NPS load was an essential ingredient in the study and relied on developing accurate measurement of NPS transport during stormwater runoff periods. Certain pollutants, specifically those associated with sediment and particulate transport such as phosphorus, have produced a "chemograph" that parallels but does not exactly follow the traditional form of the hydrograph (see Figure 6). The pollutant mass transport associated with this runoff flux frequently constitutes the major fraction of NPS discharge in a given watershed (8, 13). In

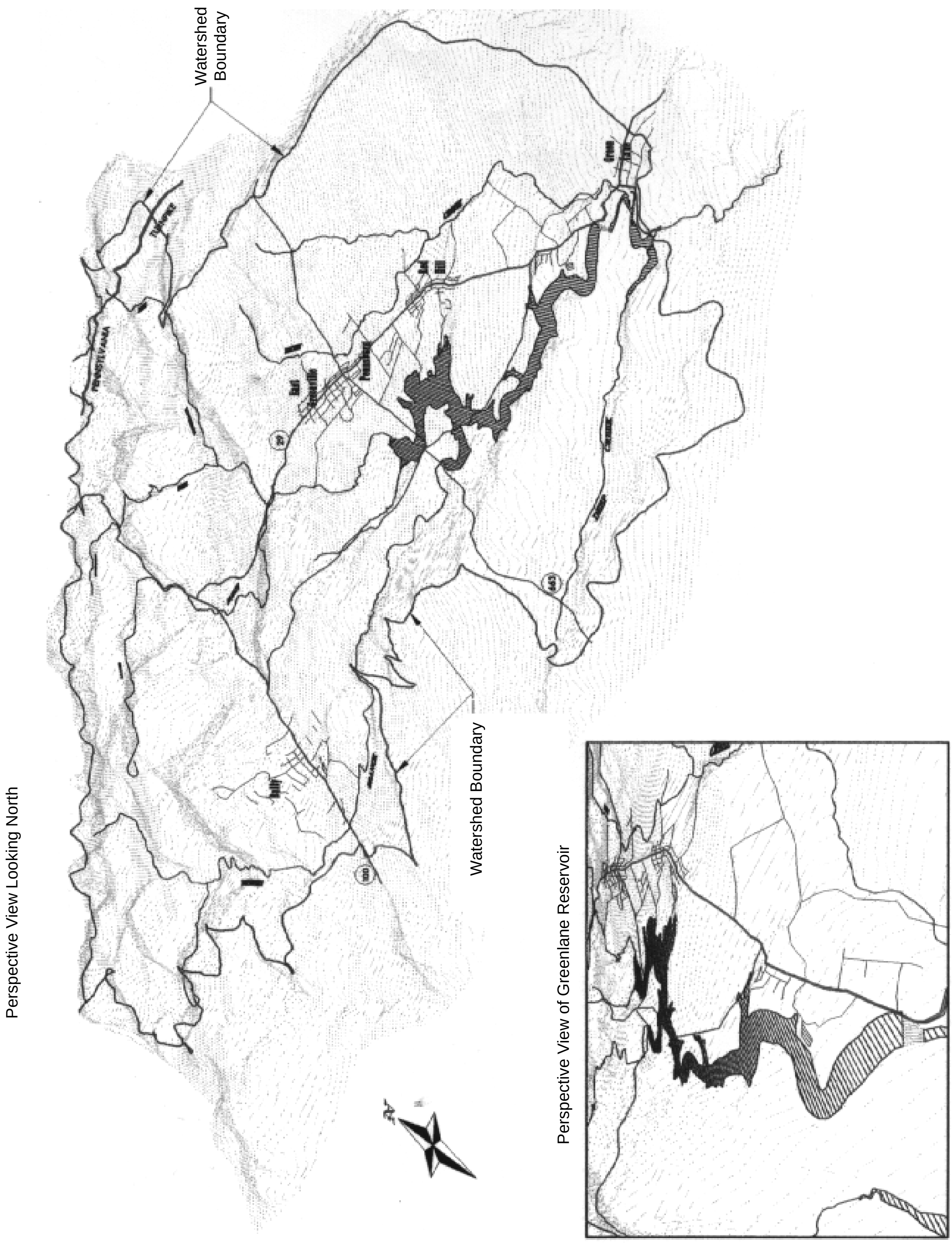


Figure 3. Green Lane Reservoir in the Upper Perkiomen watershed, 814 acres, 4.3 BG.

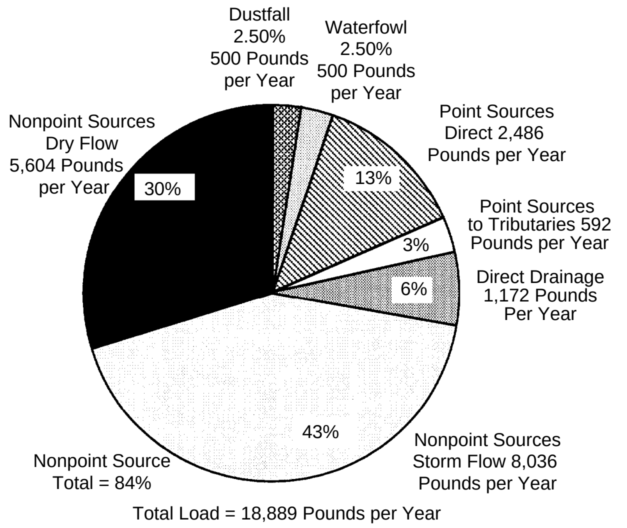


Figure 4. Sources of total phosphorus mass transport into the Green Lane Reservoir from the Upper Perkiomen watershed (71 square miles) in an average flow year—in pounds per year.

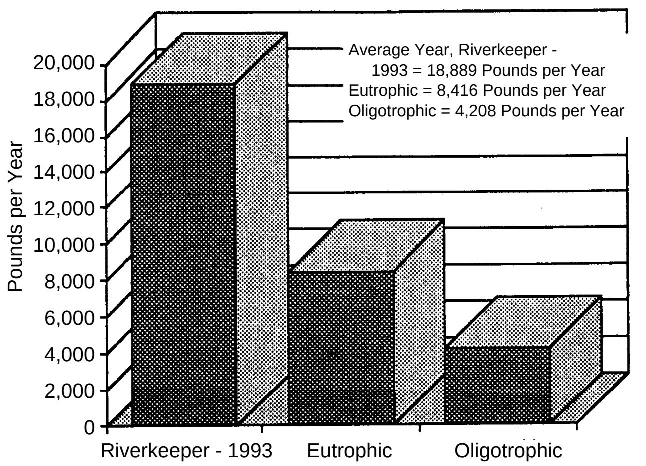


Figure 5. Reduction in annual phosphorus load required to achieve improved trophic level.

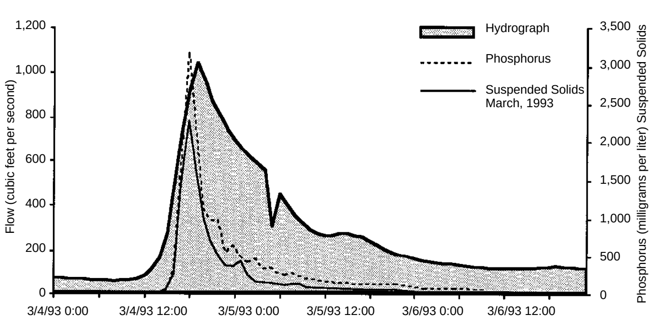


Figure 6. Storm hydrograph in the Upper Perkiomen watershed illustrating the dramatic increase in total phosphorus and suspended sediment during runoff.

the UPW study, operating continuous sampling stations at two key gage locations above the reservoir allowed the measurement of stormwater chemistry of this type and produced estimates of wet weather transport of phosphorus and sediment. Surprisingly, the NPS transport during dry weather, calculated by subtracting the point sources, was also significant and is attributed to livestock discharges and septage drainage.

But the wet weather proportion of NPS pollution still dominates lake water quality. Many have said that water quality in a given watershed is a function of land use, but that statement is as unsatisfying as saying that runoff is a function of rainfall. Experience has taught us that neither process is quite that simplistic, nor does either follow a direct linear relationship of cause and effect. The causal mechanisms that generate a certain mass load of pollutant in a drainage basin certainly result from how much mass of that pollutant is applied to the landscape within the drainage, which in turn is scoured from the landscape during periods of surface saturation, transported in, and diluted by runoff. The end result is a concentration of pollutant in the stormwater that might be several orders of magnitude greater than during dry weather flow, the hydrologic period traditionally used to measure and define water quality.

Developing NPS analysis or algorithms for stormwater quality modeling requires replicating the specific hydrograph and its associated chemograph, as well as defining the mechanisms by which pollutants are scoured from the land surface, transported in runoff, and pass through the river system. Total phosphorus (TP), for example, is transported with the colloidal soil particles (see Figure 7), so sediment transport and deposition constitute a key mechanism.

Adding to these complications is the question of whether to model single or multiple events. Is the chemodynamic process one in which the transport takes place over a series of storm events, so that each storm moves the pollutant mass a given distance in the drainage and then allows it to settle in the channel only to resuspend it with the next peak of flow? Or does the total mass transport occur in one single dynamic, from corn field or suburban lawn to lake, estuary, or other sink, that is hours or days downstream in the drainage? The issue of how stormwater transport of pollutants takes place is of paramount importance in current planning and regulatory implementation (11) because many of our current BMPs are relatively ineffective in removing NPS pollutants. This understanding is critical even as we attempt to intervene in the pollutant generation process by changing the way we cultivate the land, fertilize our landscapes, or for that matter, how we alter the land surface during growth.

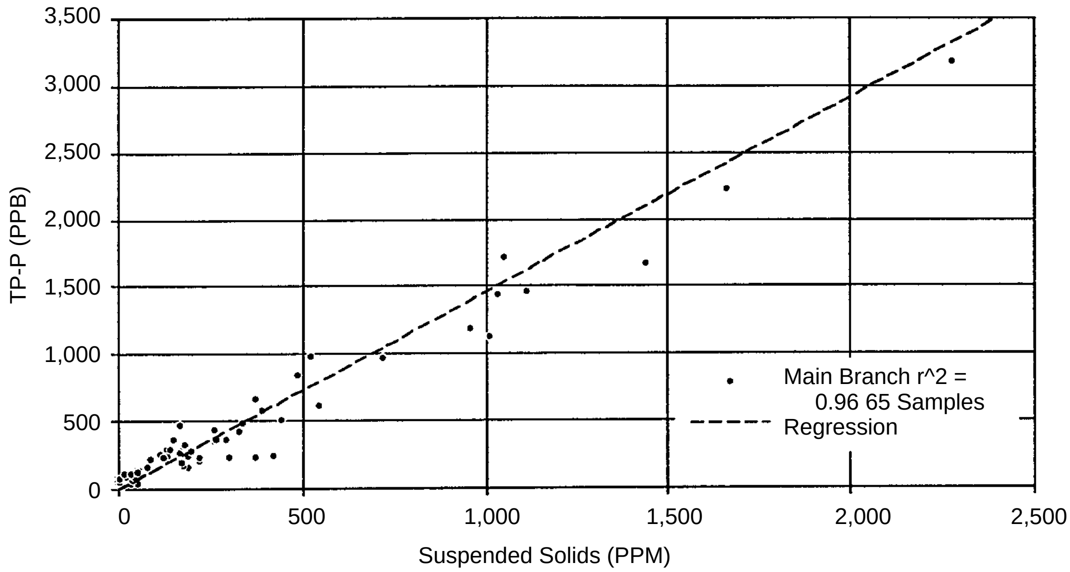


Figure 7. The relationship between total phosphorus and suspended sediment concentrations during runoff is strong but varies with different watersheds.

GIS Evaluation

The GIS data files on land use/land cover that were created for the UPW show that the bulk of the area is still quite undeveloped and rural (see Figure 8), with the steeply sloped and igneous rock areas in the headwaters in forest cover (38 percent) and the valleys in mixed agriculture (44 percent). The urban/suburban land composes the remaining 18 percent and largely consists of several older, historic boroughs linked together in a lineal pattern with widely scattered, low-density residential areas. Much of the existing housing is turn-of-the-century at quite high densities, mixed with a variety of commercial and other uses. This pattern contrasts sharply with typical large-lot suburban subdivisions. In fact, these watershed boroughs resemble the “village” concepts that innovative planning theorists advocate in a variety of important ways.

The watershed (see Figure 9) is blessed, or cursed, depending upon one’s perspective, with a multiplicity of local governments including four different counties and 18 different municipalities. This arrangement poses special challenges for management program implementation. Population projections indicate that additional development will occur at moderate rates throughout the watershed, reflecting recent trends.

Farming, both crop cultivation and dairying, is a major existing land use in the watershed, although agriculture is not especially robust and appears to be declining. This lack of agricultural vibrancy becomes a major factor in determining how to impose additional management measures on agricultural pollution sources. GIS tabulation of agricultural land totals some 19,000 acres above the reservoir, which can be compared with the estimated TP and suspended solids (SS) mass transport reaching the lake. Considering only the agricultural land to be the source of this NPS input (not quite true) suggests an average annual yield of 180#/acre/year-SS and 0.22#/acre/year-TP.

This sediment/phosphorus yield is more than sufficient to maintain a eutrophic condition in the reservoir system. The problem with this yield, however, is that it is two orders of magnitude less than commonly accepted methodologies of soil erosion, such as the universal soil loss equation (14), would suggest might come from such a watershed. Analysis of the cultivation practices taking place on farmland in the watershed estimates soil erosion to be approximately 5 to 10 tons per acre or more per year, far more than is observed passing out of the basin into the reservoir. The phosphorus applications on both cultivated and maintained residential landscapes also appear much greater than the mass transport actually measured in the flowing streams, which represent perhaps 7 percent or less of the annual land application.

The implication for NPS analysis is that the standard shopping list of either agricultural or urban BMPs might only reduce the mass transport by a relatively small fraction, even if successfully applied throughout the drainage. As Figure 7 illustrates, most of the phosphorus transport occurs on the colloidal fraction of sediment particles, which tend to remain in suspension as stormwaters pass through conventional detention structures, terraces, or grassed swales.

To consider more radical measures, GIS was used to determine possible alternatives, such as creating a stream buffer system (see Figure 10) with various setback distances from the perennial stream network, and to evaluate how great an impact this might have on agricultural land use and urban development. Land

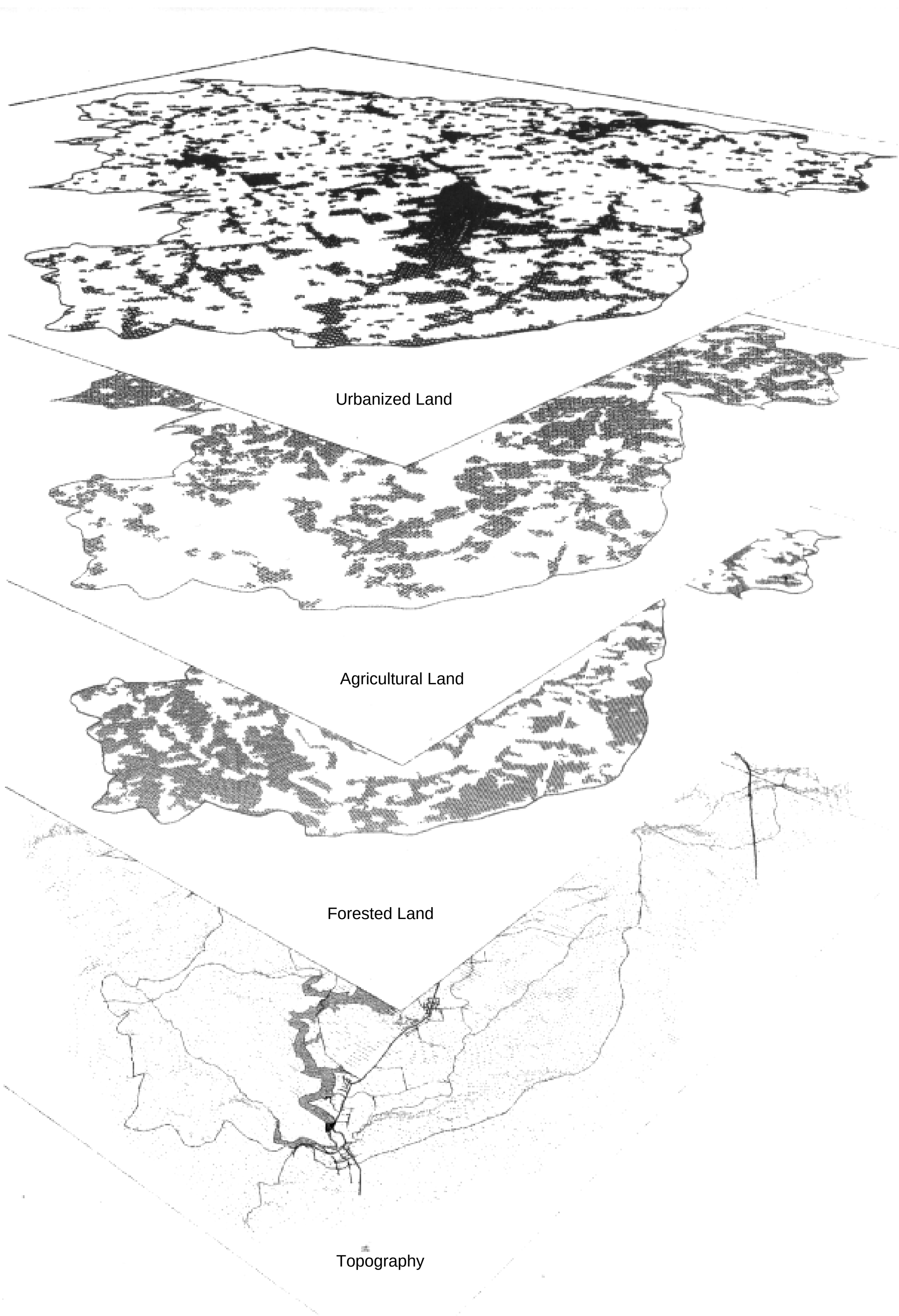


Figure 8. GIS data files showing land use/land cover characteristics for the Upper Perkomien watershed.

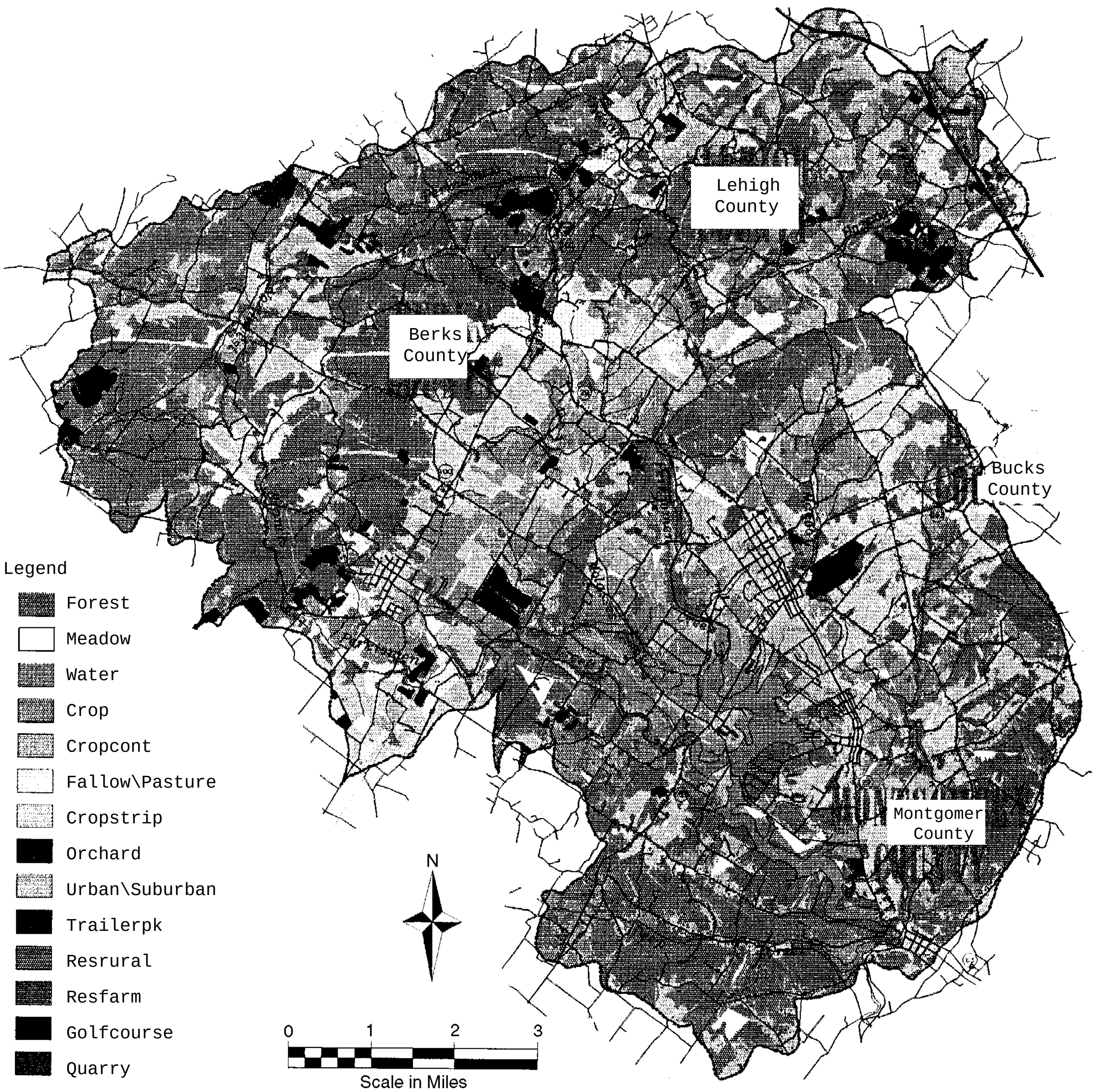


Figure 9. Existing land use/land cover GIS file for the Upper Perkiomen watershed. The 95-square-mile basin includes portions of four counties and 18 municipalities.

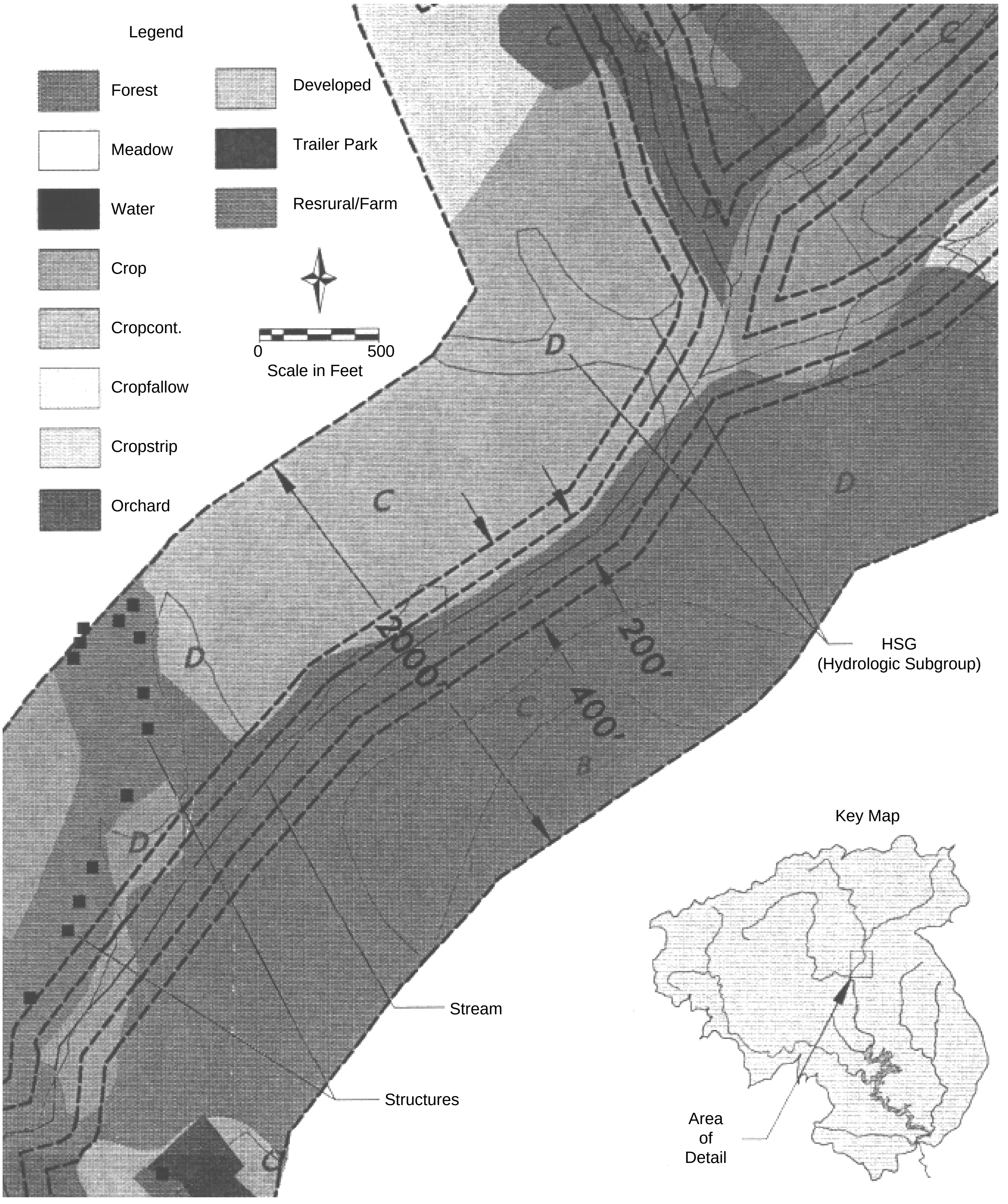


Figure 10. GIS analysis of stream corridors allows evaluation of riparian buffer systems, potential agricultural land loss, potential septic system discharges, and related NPS reduction with selected best management practices.

use at varying distances (100 feet, 200 feet, and 1,000 feet) from streams was tabulated, including all land area in the "active" agriculture categories. This GIS documentation allowed estimation of the significant NPS reduction in loadings that a riparian corridor management program could achieve.

In the same way, GIS analysis helped estimate pollutant loadings from malfunctioning onsite septic systems. Counts of structures in nonpublicly sewered areas within varying distances from the stream system were developed using GIS data files. The nearly 300 potential systems within a 200-foot radius of those streams draining into the Green Lane Reservoir identified in this manner, with pollutant generation factors applied, became the basis of a dry weather pollutant estimation. Although this approach was dependent on a variety of assumptions, alternative approaches of evaluating the problem, such as field visits to actual onsite systems throughout the watershed, would not have been feasible.

For urban and suburban development, the management focus was to estimate NPS loadings from future land development. GIS was used to demonstrate NPS pollutant load implications of future growth envisioned in the watershed's keystone municipality, Upper Hanover Township. Here, an increase of 15,000 residents would convert 1,772 acres into residential, commercial, and industrial uses. Nonpoint pollutant loadings generated by this new land development constituted significant increases in phosphorus, suspended solids, metals, oil/grease, and other pollutants, and would reverse any improvements in Green Lane Reservoir water quality that recent wastewater treatment plant upgrades achieved.

From a water quality perspective, future alternative land use configurations that concentrate development and minimize ultimate disturbance of the land surface yielded would substantially reduce NPS pollutant loadings into the reservoirs. This entire process of testing land use implications of different management approaches for their water quality impacts indicated that pollutant loads could be minimized far more cost effectively through management actions, both structural and nonstructural, which varied from the areawide to the site-specific.

Neshaminy Creek Watershed Stormwater Management Study

Background

The Neshaminy Creek watershed, including 237 square miles of mixed urban and rural land uses, lies primarily in Bucks County, Pennsylvania, and flows directly into the Delaware River (see Figure 1). The 1978 Pennsylvania 167 Stormwater Management Act, which required that counties prepare stormwater management plans for all 353 designated watersheds in the state, mandated the Neshaminy study. This act further stipulated that municipalities then needed to implement the watershed plans through adopting the necessary municipal ordinances and regulations. In fact, the Neshaminy study had three water resource management objectives:

- Prevent worsened flooding downstream caused by increased volumes of runoff from land development.
- Increase ground-water recharge.
- Reduce NPS pollutant loadings from new development.

In the initial study design, water quality and NPS issues were secondary to flooding concerns. When Pennsylvania's stormwater management program was conceived, the state focused on preventing watershed-wide flooding. Clearly, detention basins have become the primary mode of managing peak rates of stormwater discharge site-by-site in most communities. Because detention basins only control peak rates of runoff and allow significantly increased total volumes of water discharged from sites, however, the increased stormwater volumes can theoretically combine and create worsened flooding downstream. Consequently, most Act 167 planning has focused on elaborate hydrologic modeling designed to assess the seriousness of potential cumulative flooding in watersheds under study.

In the case of the Neshaminy, however, the record suggested that although localized flooding could be an issue, an existing network of eight multipurpose flood control structures constructed during the 1960s served to prevent significant flooding. Water quality certainly was a serious stormwater concern, however, especially in the areas flowing into the reservoirs where recreational use had become intense. Several of the existing impoundments were multipurpose, their permanent pools providing critical recreational functions for a burgeoning Bucks County population. At the same time, the proliferation of development in the watershed, with its increased point and nonpoint sources, had degraded streams and seriously affected the reservoirs. While the total stream system in the watershed was of concern, the future of the reservoirs came to be particularly important in developing the total stormwater management program for the Neshaminy watershed.

The Neshaminy lies at the heart of Bucks County, Pennsylvania's primary population and employment growth county (see Figure 11). Although the Neshaminy watershed has already experienced heavy development, especially in the lower or southern portions, farmsteads and large areas of undeveloped land still exist, especially in headwater areas. Agriculture has been a major land use in the past, but farms rapidly are converting to urban uses as the wave of urbanization moves outward from Philadelphia and from the Princeton/Trenton metropolitan areas. Growth projections indicate continu-

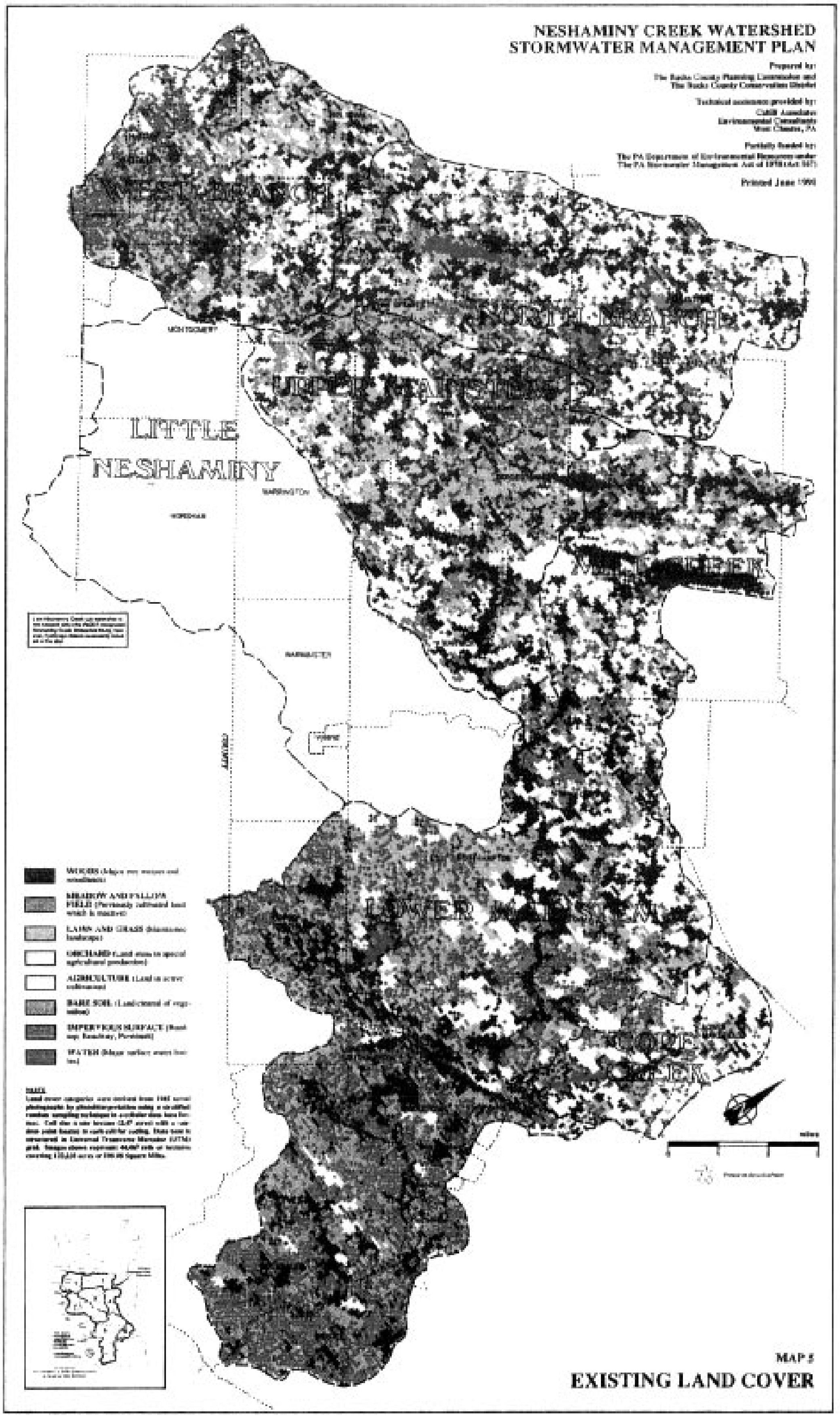


Figure 11. Land use/land cover in the Neshaminy basin of Bucks County, Pennsylvania. The watershed covers 237 square miles in southeast Pennsylvania.

ation of this rapid growth and a continuing change in existing land use/land cover, together with projected development within the required 10-year planning horizon.

Physiographically, the watershed spans both the Piedmont and Atlantic coastal plain provinces, with rolling topography and relatively steep slopes underlain by Triassic formation rock, including the Lockatong, Brunswick, and Stockton formations. This bedrock ranges from being a poor aquifer (Lockatong) to an excellent aquifer (Stockton) where the many rock fractures allow for considerable ground-water yields. Soils are quite variable, ranging from good loam (hydrologic soil group B) to clays and other types with poor drainage characteristics (e.g., high water table, shallow depth to bedrock). A large proportion of the soils in the watershed are categorized as hydrologic soil group C, which is marginal for many stormwater management infiltration techniques (see Figure 12) and produces a relatively large proportion of direct runoff. With an annual rainfall of 45 inches, base flow accounts for about 12 inches and direct runoff accounts for 10 inches.

The system of eight stormwater control structures, which were built over the past three decades under the federal PL 566 program, have altered the hydrology of the watershed (15). In addition, in heavily developed portions of the watershed, impervious surfaces combined with numerous detention basins prevent the bulk of the precipitation from being recharged, and the volume of total runoff proportionally increases. An elaborate system of municipal and nonmunicipal wastewater treatment plants also adds to this alteration of the hydrologic cycle. These plants discharge wastewater effluent that, in some cases, constitutes the bulk of the stream flow during dry periods. While the impact of NPS was evident throughout the drainage, it was of special interest in the impoundment network, especially those impoundments that were conceived as multipurpose in function and constitute major recreational resources in the watershed.

GIS Design

Act 167 requirements and the needs of the hydrologic and other modeling used in planning both heavily influenced the GIS developed for the Neshaminy. Spatial data files, including existing land use, future land use, and soil series aggregated by hydrologic soil groups, were created by digitizing at a 1-hectare (2.5-acre) cell resolution. The encoding process that helped design the GIS used a stratified random point sampling technique that similar studies had developed and applied (1, 3). The encoding process used a metric grid of 5-kilometer sections, subdivided into 2,500 1-hectare cells (100 meters on a side), aligned with the Universal Transverse Mercator (UTM) Grid System. This grid appears in blue on U.S. Geological Survey (USGS) topographic maps. These maps served as the framework of reference for all data compilation. Within each 100-meter cell, a randomly located point was chosen (see Figure 13) at which the specific factor was encoded as representative for the cell, using a digitizer tablet. This approach allowed extraction of the data from the respective source documents with some rectification necessary for many types of source maps and photographs.

The combination of soil series and cover in each cell helped to calculate the curve number and unit runoff per cell. The 45,000-cell data file was then used to calculate total runoff for a range of events in each of 100 subbasins that averaged 1.95 square miles each. The resultant hydrographs, used in combination with a separate linear data file in GIS describing the hydrographic network of stream geometry, routed and calibrated the hydrologic model (TR-20). NPS mass transport loadings were estimated on an annual basis by cell, again using the land use/land cover data file, and total loads summed by groups of subbasins above critical locations. This issue was particularly important with respect to the drainage areas above the impoundments, where NPS pollutants were of greatest concern.

The soil properties data file was especially useful in evaluating certain management objectives, such as the opportunity for recharging ground-water aquifers. The spatial variation in relative effectiveness of infiltration BMPs was considered for both quantity and quality mitigation because the best methods for NPS reduction usually include recharge where possible. The soil series corresponding with new growth areas were classified regarding their suitability for these BMPs, which are most efficient on well-drained or moderately well-drained soil. Thus, the alternative impacts of future growth could be considered in terms of potential generation (or management) of NPS loads. A BMP selection methodology (see Figure 14), which was developed for the 30 municipalities within the watershed, focused on new land development applications and considered both water quantity and quality management objectives. BMP selection is a function of several factors, including:

- The need for further peak rate reduction.
- The recharge sensitivity of the project site (defined as a function of headwaters stream location, areawide reliance on ground water for water supply, or presence of effluent limited streams).
- The need for priority NPS pollution controls (location within reservoir drainage).

Development of two “performance” levels of BMP selection techniques gave municipalities some degree of flexibility in developing their new stormwater management programs. This system required only the minimally acceptable techniques but recommended the more fully effective ones, hoping that municipalities would strive to incorporate

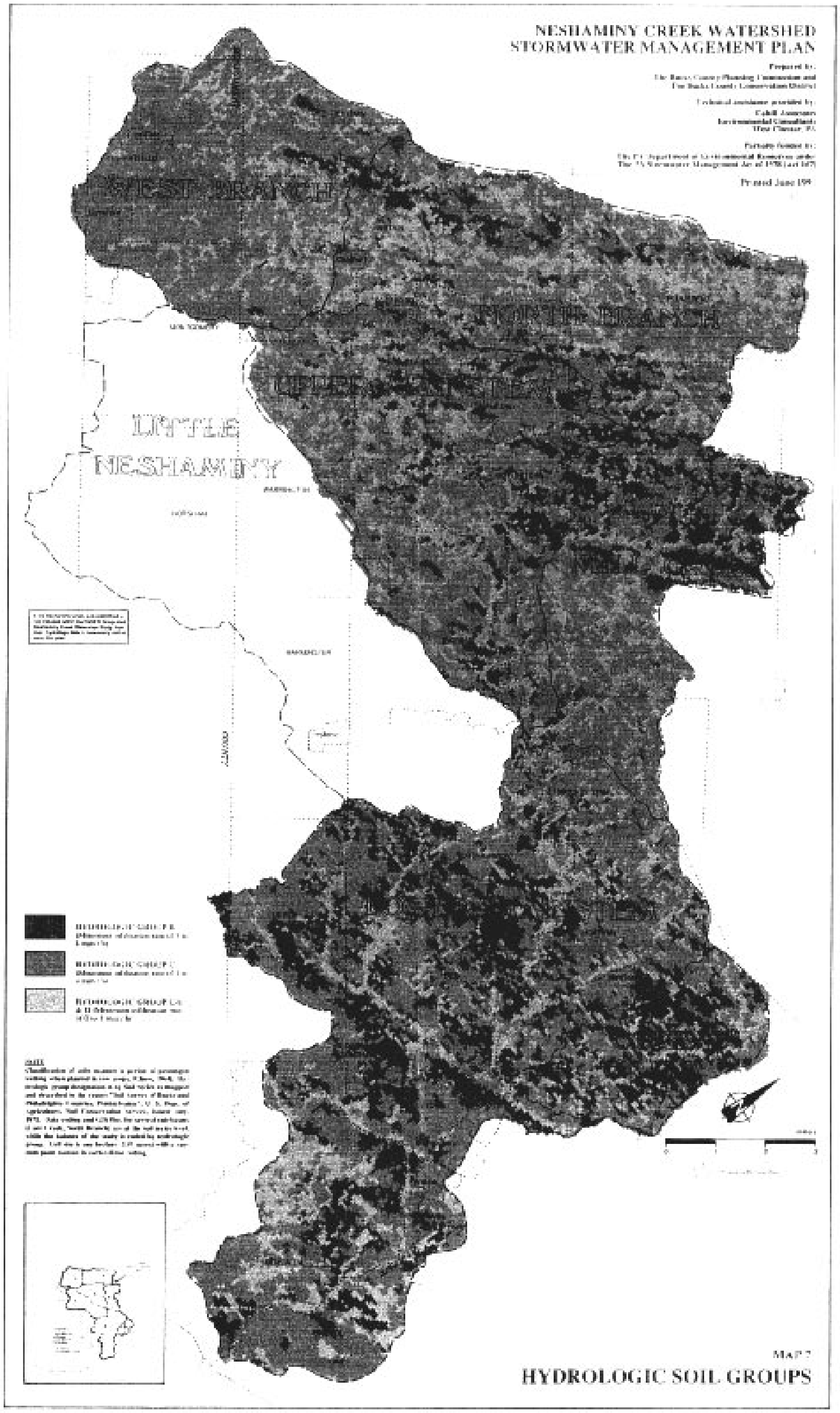


Figure 12. GIS file of hydrologic soil groups in the Neshaminy basin. The 31 soil series are digitized in 45,000 pixels of 1-hectare size.

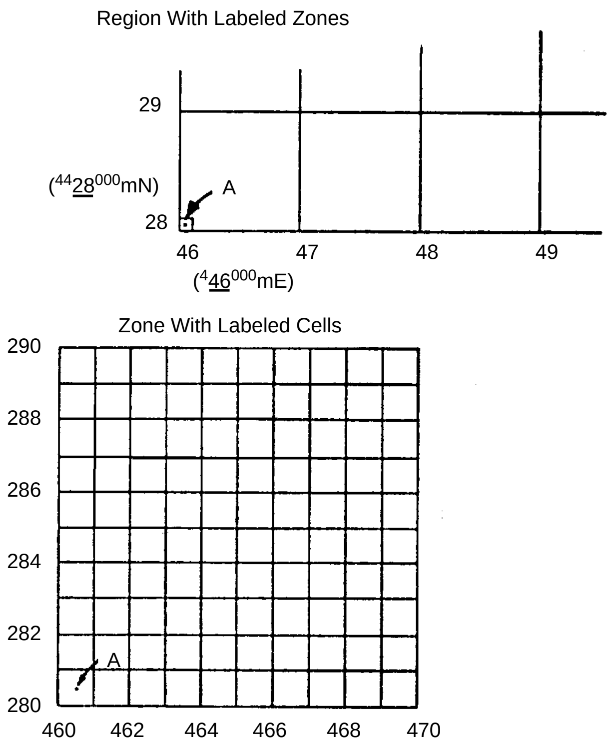


Figure 13. Raster/pixel design of GIS for Neshaminy modeling study. Each pixel is 1 hectare (2.47 acres).

recommended management measures wherever possible. The BMP selection methodology also was sensitive to type of land use or proposed development, assigning typical single-family residential subdivisions different BMPs than, for example, multifamily and other nonresidential proposals (including commercial and industrial proposals). The selection process also determined size of site to be a factor, differentiating between sites of 5 acres or more because of the varying degrees of cost and effectiveness of different BMP approaches. The methodology, if properly and fully implemented, should achieve the necessary stormwater-related objectives—both quantity and quality—that the analysis had deemed necessary (16).

GIS was especially important in its ability to test how reasonable the BMP selection methodology was. Such tests included the ability to evaluate, for each municipality, the following factors:

- The nature and extent of the projected development.
- The size of development/size of site assumptions.
- Other vital BMP feasibility factors such as soils and their appropriateness for different BMP techniques.

GIS also enabled analysis of the water quantity and quality impacts of projected growth on a baseline basis, assuming continuation of existing stormwater management practices. Water quality loadings to individual reservoirs and to the stream system could be readily demonstrated. Because overenrichment of the reservoirs was so crucial, researchers could estimate phosphorus and nitrogen loadings from projected development assuming existing stormwater practices, even on a municipality by municipality basis.

New Jersey Atlantic Coastal Drainage Study

Background

The third study considered a much larger coastal watershed in New Jersey (see Figure 15). The New Jersey

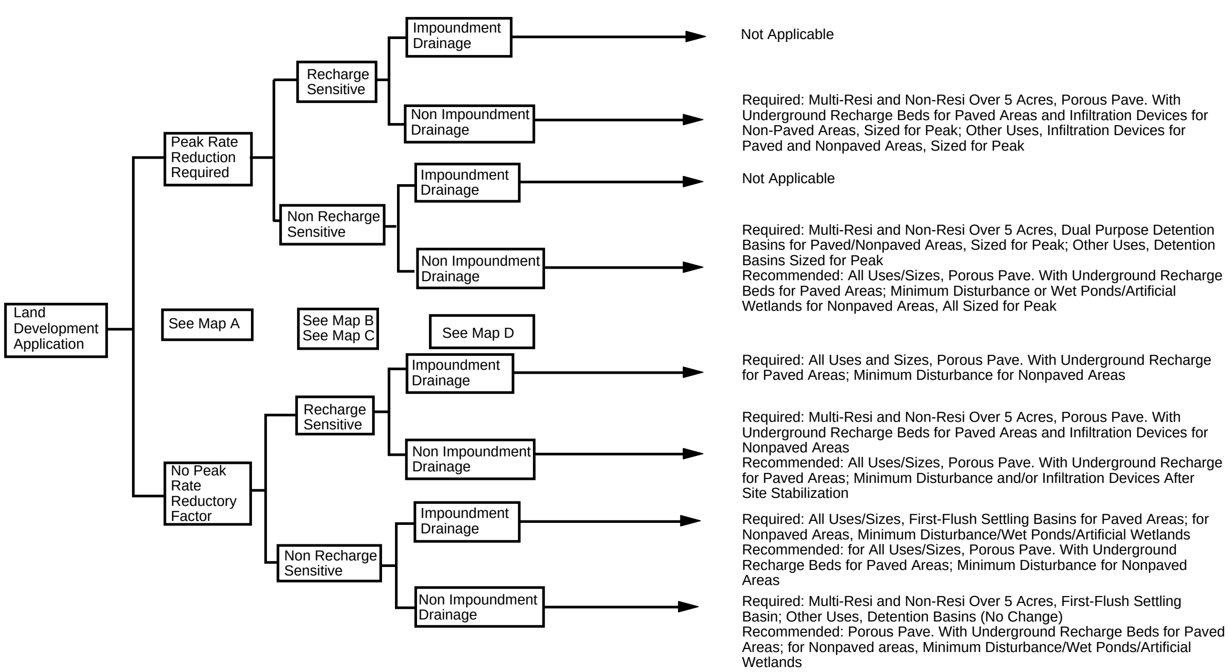


Figure 14. BMP selection methodology used with the GIS database in the Neshaminy basin modeling study.

Atlantic Coastal Drainage (ACD) includes an area of 2,086 square miles, with barrier islands (50 square miles), wetlands/bays/estuaries (285 square miles), and a unique scrubby pitch pine-cedar forest, known as the Pine Barrens, largely covering the 1,750 square miles of mainland interior (see Figure 16). This flat coastal plain comprises a series of unconsolidated sedimentary deposits of sand, marl, and clay, which increase in thickness toward the coastline. Over the past 16,000 years, as the ocean level has risen, the water's edge has progressed inland to its present position. Ocean currents and upland erosion and deposition have created a long, narrow series of barrier islands that absorb the energy of ocean storms and buffer the estuary habitats from the scour of waves and currents. Between the mainland and barrier islands are embayments and estuaries of different sizes and configurations. Inland erosion and marine sediments have gradually filled many of these areas, creating extensive wetlands (17).

In this ACD region, new land development and population growth have caused significant degradation of water quality from an increase in both point source and NPS pollution. Although the array of pollutants is ominously broad, increased nitrogen and phosphorus loadings have resulted in enrichment of back bays, estuaries, and nearshore waters, contributing to algal blooms, declining finfish and shellfish populations, diminished recreational

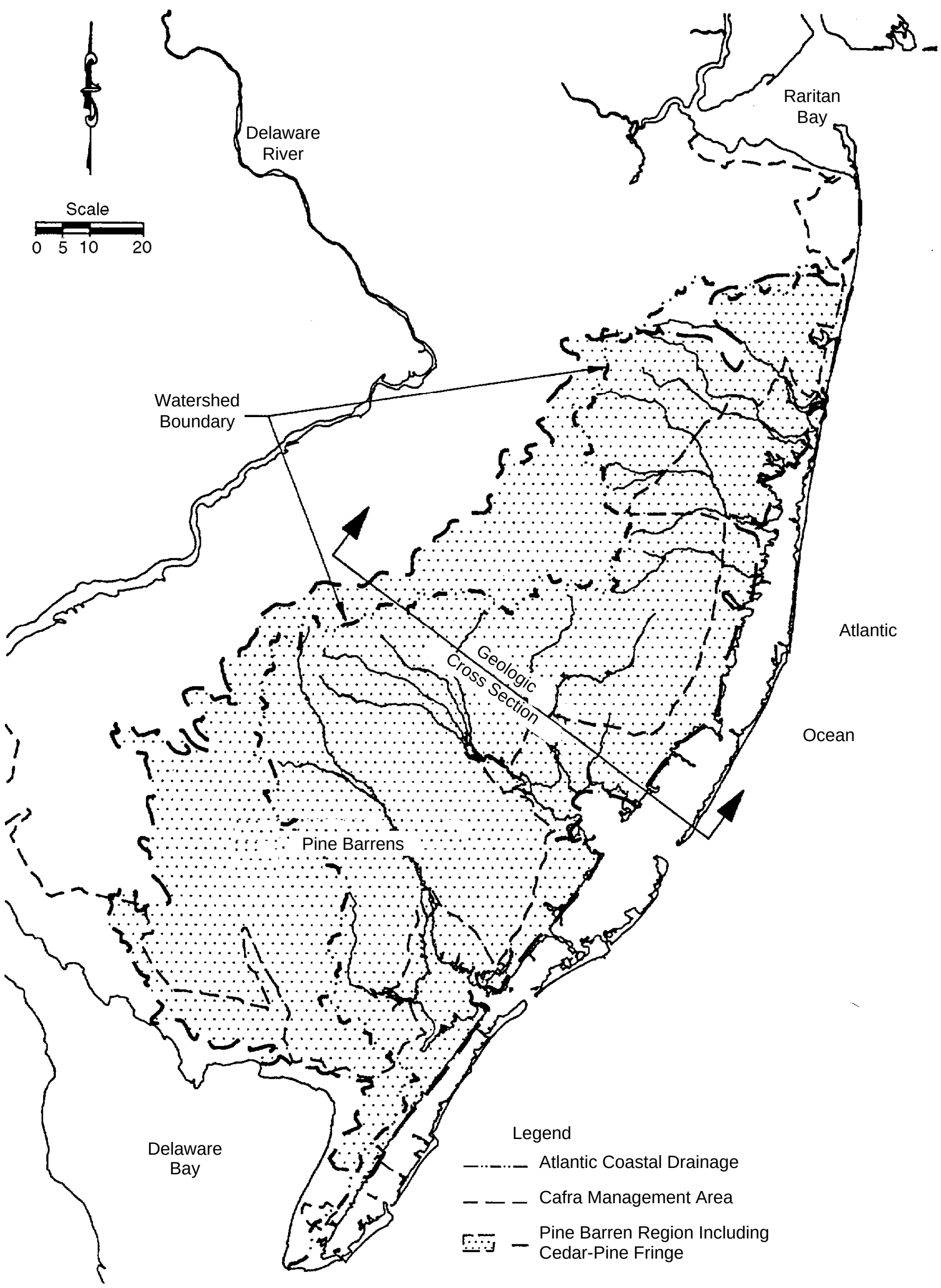


Figure 15. The ACD of New Jersey includes approximately 2,000 square miles of land area from the Manasquan River to Cape May.

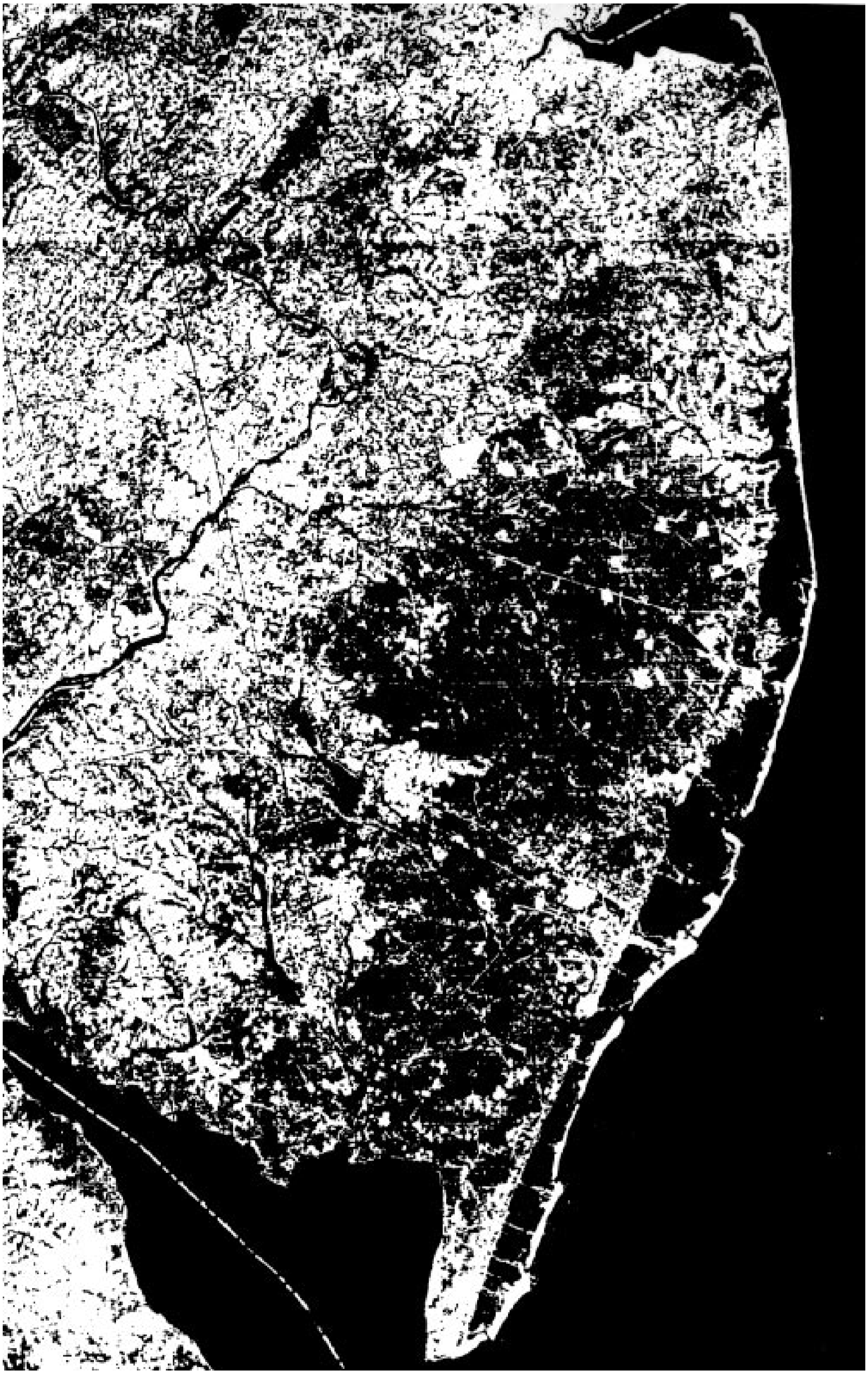


Figure 16. Aerial photograph of New Jersey illustrating the Barrier Islands and estuary system situated along the Atlantic coast.

opportunities, and a variety of other problems (18). A major source of these nutrients is point source sewage treatment plants (STPs), but the effluent outfalls of almost all these STPs discharge into nearshore ocean waters beyond the barrier islands. Thus, NPS pollutants almost totally dominate the water quality in the estuaries and back bays (19, 20).

These NPS pollutants, which rain scours from the land surface and flushes into coastal waters with each rainfall, comprise a largely unmeasured and unmanaged flux of contaminants. Prior research on coastal water quality has given considerable attention to NPS pollution generated from paved or impervious surfaces, particularly roadways and parking lots where hydrocarbons, metals, suspended solids, biologic oxygen demand (BOD), and other pollutants have been measured.

Although these NPS pollutants are certainly of concern in New Jersey's coastal waters, the enrichment issue has led to a focus on NPS pollution produced when creating large areas of pervious and heavily maintained landscape, such as lawns and other landscaped areas, in the sandy soil context of the coastal area. Typically, significant quantities of fertilizer and other chemicals, which are applied on these new pervious surfaces, are naturally low in nutrients. Although a modest portion of the applied fertilizer runs off directly into surface waters, larger quantities of soluble pollutants, such as nitrates and herbicides, quickly percolate down through the sandy soil, then move rapidly as interflow to the estuary system.

In this coastal drainage of unconsolidated sediments, the hydrologic cycle differs from inland watersheds. Of the 45-inch average annual rainfall, only a small fraction (2.5 inches per year) becomes direct runoff, with the balance rapidly infiltrating into the sand strata (21). Most of the infiltration that reaches the ground water (20 inches per year) discharges to surface streams (17 inches per year) within a few hours following rainfall, producing a lagging and attenuated hydrograph. This rapid infiltration, combined with the sand texture of the soil, has a major bearing on the water quality implications of new land development. Thus, urbanization of coastal regions has dramatically altered hydrologic response, with every square foot of new impervious surface converting what had been approximately 41.5 inches of infiltration into direct runoff to bays and estuaries, with a turbid soup of NPS pollutants.

Even in areas that have maintained infiltration, the coastal soils do not remove NPS pollutants as efficiently as other areas of New Jersey that overlie consolidated formations with heavier clay soils. These soils provide a much more thorough removal of NPS pollutants through physical, chemical, and biologic processes, as rainfall percolates through the soil mantle.

With development of coastal areas, increased impervious areas and changing flow pathways (inlets and storm sewers) convey nonpoint pollutants introduced by development (from both pervious and impervious surfaces) directly to the coastal waters. In addition, freshwater recharge to the underlying aquifer decreases with the increase in impervious surfaces, with resulting increases in saltwater intrusion into the sand aquifers and contamination of ground-water supply wells along the coast. Further compounding the loss of the stormwater for ground-water recharge are increased ground-water withdrawals necessary for new water supply. In sum, urban growth within the ACD, with its 1.13 million permanent residents (and still growing) and an additional 1.5 million summer tourists, has dramatically altered the natural drainage system (and landscape) in a way that significantly increases the discharge of NPS pollutants (22).

GIS Approach

New Jersey's Department of Environmental Protection already had developed a computerized GIS system (ARC/INFO) for environmental analysis and resource planning, so this study aimed to use existing GIS work and to refine this GIS system. Although data files for municipal boundaries, watershed areas, and a variety of other factors already existed, land use/land cover data had not been developed and constituted a major work task. The subsequent land use/land cover file included the entire 2,000 square miles of the ACD, but this focused on the urbanized area (212 square miles) that occupied about 11 percent of the coastal fringe. The end product was a polygon file that described about 2,500 polygons of urban/suburban land, each averaging about 0.1 square miles (see Figure 17).

Using aerial photographs combined with USGS base maps and extensive field reconnaissance, each polygon was classified by:

- Land use type.
- Percentage of impervious cover and maintained areas.
- Degree of maintenance (fertilization) being provided to these maintained areas.

Although classifying land use type and extent of impervious cover/maintained areas was a relatively straightforward evaluation process (rated within one of 11 categories by percentage, 0 to 5 percent, and so forth), the third variable, degree of maintenance, required special treatment and data development procedures. Degree of maintenance was translated into high, medium, and low categories, with high maintenance exemplified by golf courses or other intensively maintained areas. Medium maintenance assumed chemical application rates comparable with those recommended by Rutgers University state agronomists. Finally, low maintenance was typified by a wooded or otherwise naturally vegetated

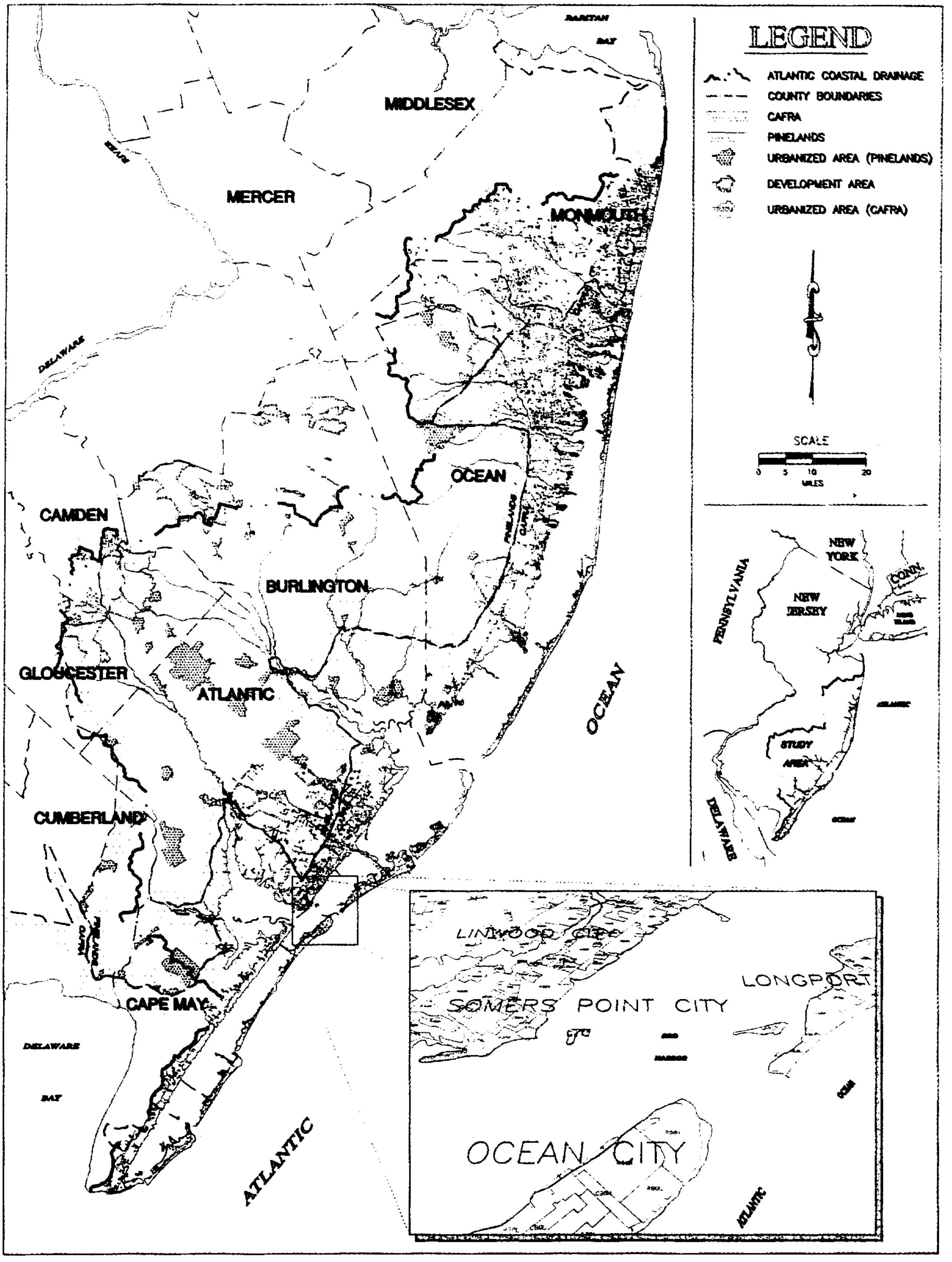


Figure 17. Urban land use polygons digitized for the New Jersey coastal drainage. The 2,500 polygons shown cover approximately 212 square miles (11 percent) of the ACD area of 2,000 square miles.

lot and assumed little or no regular chemical application. Research staff executed considerable field reconnaissance to objectify this judgment-based rating technique (see Figure 18).

Nonpoint Source Analysis

Because the drainage is almost entirely estuarine, the hydrologic aspects in this study were almost irrelevant except as a tool to describe the pollutant transport process. Such coastal drainage systems do not allow the measurement of hydrographs and chemographs (see Figure 6), except on inland riverine segments or selected infrastructure points of discharge (storm sewer outfalls). Thus, the NPS analysis focused on the pollutant production and transport process, especially the nutrients applied to the maintained landscapes, which are a major part of coastal urbanization.

This study required a great deal of effort to produce an index table relating urban cover characteristics (percentage impervious, amount of chemical application) to NPS production potential. For each of the 2,500 urban land polygons GIS described, estimates of the NPS loading for a number of pollutants were generated. Potential loadings were then aggregated by subwatershed. Total NPS loadings could then be compared with point source loadings for the entire coastal drainage. Interestingly, the NPS loading dominated water quality in the estuarine drainage while the point sources, discharged by ocean outfalls to nearshore waters beyond the barrier islands, were the major source of nutrient pollution in this portion of the coastal environment (see Figure 19).

Given the estimates of NPS pollution, the major question involves how to control or reduce these loads. The suitability of selected BMPs for the reduction/prevention of pollutant generation was then evaluated and spatially identified within the drainage (see Figure 20). This figure evaluated the use of constructed wetland systems as a structural measure. That is, GIS allowed state regulators to identify not only what works best in terms of water quality protection measures, but also where these methods could work successfully. This analysis was driven by a detailed evaluation of the combinations of natural conditions GIS identified within the study area. For example, certain BMPs can be applied on soils that have a certain set of characteristics (permeability, depth to seasonal high water table) and that are presently in a given land cover and planned for urbanization.

GIS also aided in evaluating alternative BMP techniques, including reduction in nutrient applications and land management BMPs such as elimination of artificial landscapes, again using a combination of natural features and land use patterns (see Figure 21). The result of this analysis considered the relative proximity of urban land uses to the coastal waters as significantly increasing the potential for NPS transport. State regulatory programs establish minimum setback criteria for development in sensitive areas, and these criteria may be modified to consider pollutant production potential based on GIS delineation of pollutant production.

New Jersey has been striving to develop NPS management programs for coastal areas to reduce existing sources of pollution as well as prevent the creation of

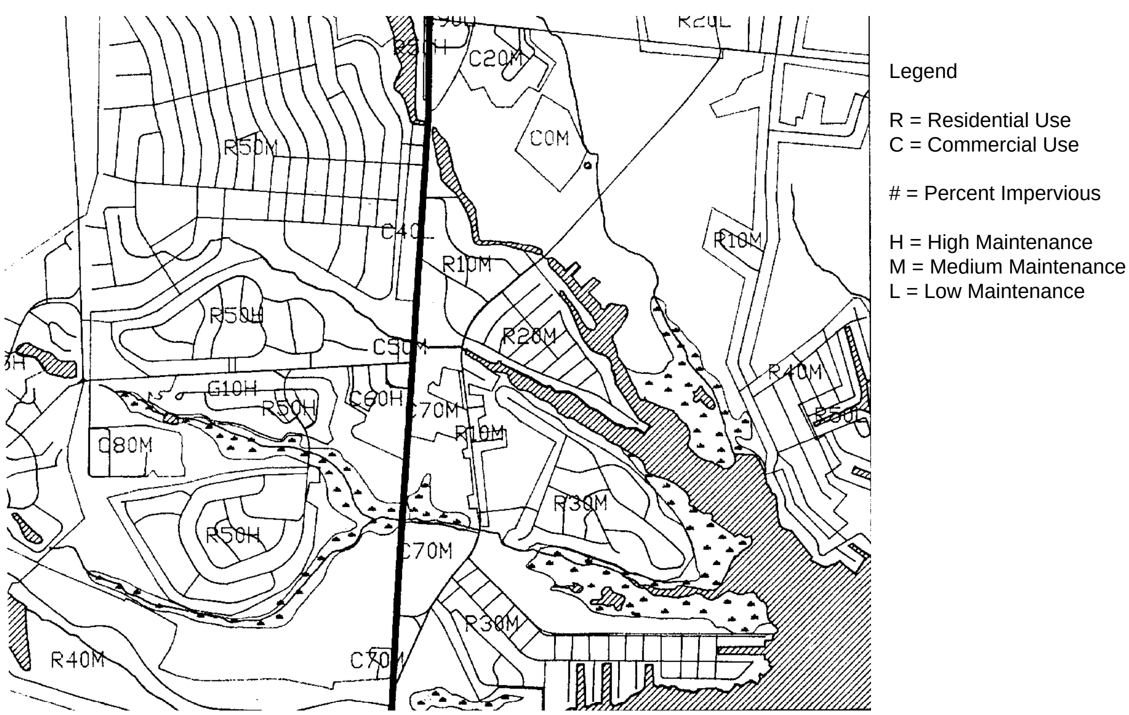


Figure 18. Classification of urban polygons by land use, percentage impervious cover, and degree of land fertilization.

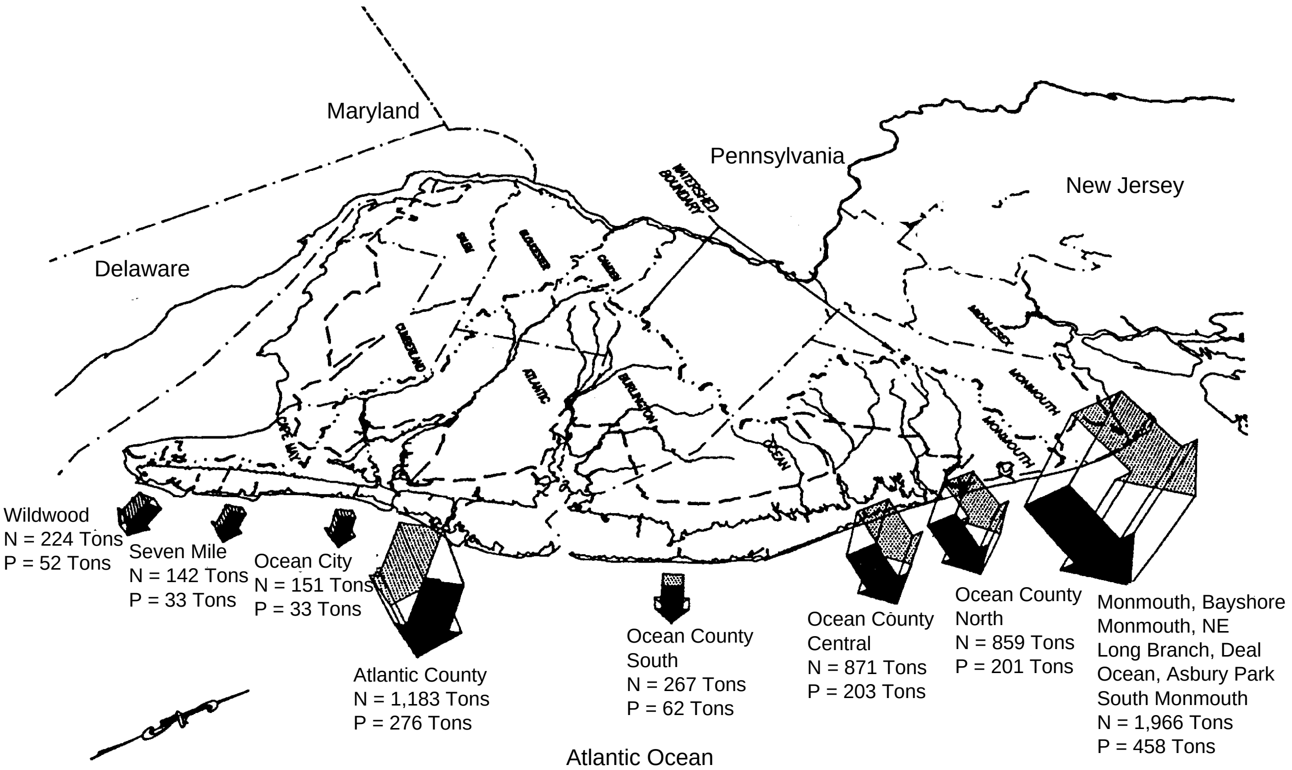


Figure 19. Point and NPS discharges to the ACD. Data shown are in tons of TP and NO_3-N per year.

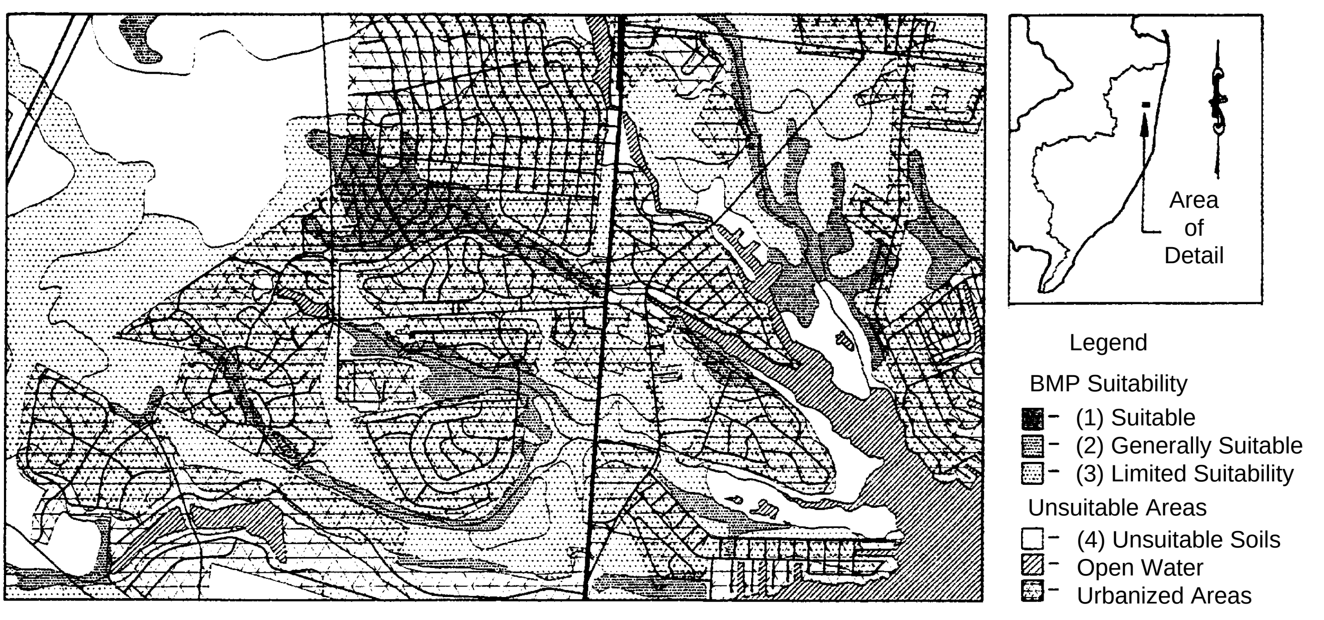


Figure 20. BMP analysis using GIS. Files consider soil suitability, current vegetative cover, and BMP criteria on vacant and developed land parcels.

pollution. As most regulatory agencies have discovered, NPS management programs can be difficult to implement, especially when confronting issues of land use management. To substantiate the need for new management programs amidst these controversies, the ability to document causal linkages (i.e., to generate data and statistics that make the case for NPS pollutant generators and resultant water quality degradation) is very important. The need for documentation of various types is especially great given the less than perfect data record of water quality in coastal and other waters. All of these factors come together to make the value of a GIS system for water quality management very real.

Conclusion

This GIS-driven analysis indicates that NPS pollutants, especially the nutrients phosphorus and nitrogen, generated from fertilized fields or maintained landscapes surrounding new residential, commercial, and other types of development in drainage systems, contribute significantly to water quality degradation. In effect, the particulate-associated phosphorus and the soluble nitrates serve as surrogates for the full spectrum of NPS pollutants that each rainfall washes from the land. A comprehensive water quality management program must include structural measures to remove pollutants this runoff conveys, as well as management of the contributing landscape to reduce (and perhaps eliminate) the application of these chemicals within the drainage.

In planning new development, management actions should occur on a variety of levels or tiers. On an areawide basis, growth should proceed (with guidance and management) in a manner that would reduce total pollutant discharges; therefore, the total amount of maintained area being created should be as concentrated as possible. On the more site-specific level, measures and construction techniques that reduce the quantity of pollutants generated are essential. Required development guidelines must include, but not be limited to:

- Prevention of excessive site disturbance and ongoing site maintenance (described as a policy of minimum disturbance and minimum maintenance).
- Use of special materials for reduction of stormwater runoff (porous pavement and ground-water recharge).
- Use of stormwater treatment systems (water quality detention basins, artificial wetlands).

In sum, the regulatory framework must contain both “how to build” guidelines, as well as “where not to build” guidelines. GIS can be a powerful tool in both of these processes.

While inland lakes serve as nutrient traps for these NPS pollutants, perhaps the greatest potential impact is the gradual process of excessively enriching our coastal waters. As population continues to migrate to coastal areas, the importance of protecting this fragile ecosystem

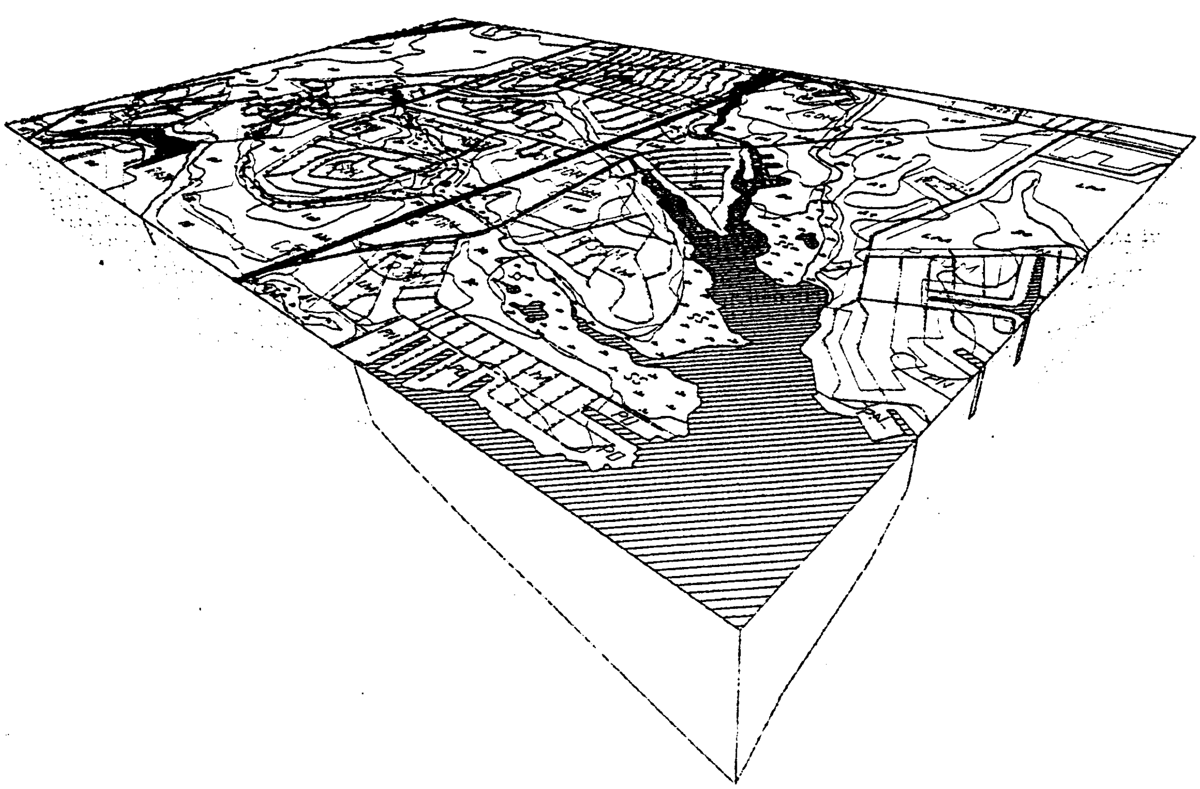


Figure 21. For certain regulatory criteria, the proximity of land uses to the water’s edge was a consideration in BMP selection.

increases. The pollution that new land development generates, including the discharge of point source wastes, should not be allowed to enter coastal waters; it should not be allowed to destroy the natural balance that exists between land and water. The concept of stormwater management takes on an entirely different meaning when viewed as one of the basic mechanisms of this NPS pollution transport. For centuries, engineering of the shoreline has intensively focused on protecting human developments from the ravages of ocean storms. Now, however, the converse seems to be emerging: ocean waters need protection from the impacts of human development.

References

1. Bliss, N., T.H. Cahill, E.B. MacDougall, and C.A. Staub. 1975. Land resource measurement for water quality analysis. Chadds Ford, PA: Tri-County Conservancy of the Brandywine.
2. Cahill, T.H., R.W. Pierson, Jr., and B.R. Cohen. 1978. The evaluation of best management practices for the reduction of diffuse pollutants in an agricultural watershed. In: Lohr, R.C., ed. Best management practices for agriculture and silviculture. Ann Arbor, MI: Ann Arbor Science.
3. Cahill, T.H., R.W. Pierson, Jr., and B.R. Cohen. 1979. Nonpoint source model calibration in the Honey Creek watershed. #R-805421-01. Athens, GA: U.S. EPA Environmental Research Laboratory.
4. Cahill, T.H. 1980. The analysis of relationships between land use and water quality in the Lake Erie basin. Burlington, Ontario: International Association of Great Lakes Research.
5. Cahill, T.H., J. McGuire, and C. Smith. 1993. Hydrologic and water quality modeling with geographic information systems. Proceedings of the Symposium on Geographic Information Systems and Water Resources, AWRA, Mobile, AL (March).
6. U.S. EPA. 1991. Remote sensing and GIS applications to nonpoint source planning. Proceedings of the U.S. EPA Workshop for Region 5 and Northeast Illinois Planning Commission, Chicago, IL (October 1990).
7. Baker, D.B., and J.W. Kramer. 1973. Phosphorus sources and transport in an agricultural basin of Lake Erie. Proceedings of the 16th Conference, Great Lakes Research, Ann Arbor, MI (September).
8. Cahill, T.H., and T.R. Hammer. 1976. Phosphate transport in river basins. Proceedings of the International Joint Committee on Fluvial Transport Workshop, Kitchener, Ontario (October).
9. Cahill Associates. 1989. Stormwater management in the New Jersey coastal zone. Trenton, NJ: State of New Jersey, Department of Environmental Protection, Division of Coastal Resources.
10. Delaware Riverkeeper. 1993. Upper Perkiomen Creek watershed water quality management plan. Lambertville, NJ: Delaware Riverkeeper/Watershed Association of the Delaware River.
11. U.S. EPA. 1993. Guidance specifying management measures for sources of nonpoint pollution in coastal waters. EPA/840/B-92/002. Washington, DC.
12. Soil Conservation Commission. 1982. TR-20, project formulation—Hydrology. Technical Release No. 20. PB83-223768. Landham, MD: Soil Conservation Service.
13. Cahill, T.H., M.C. Adams, and W.R. Horner. 1990. The use of porous paving for groundwater recharge in stormwater management systems. Presented at the 1988 Floodplain/Stormwater Management Symposium, State College, PA (October).
14. Soil Conservation Commission. 1974. Universal soil loss equation. Technical notes, Conservation Agronomy No. 32. Portland, OR: West Technical Service Center.
15. Cahill, T.H., M. Adams, S. Remalie, and C. Smith. 1988. The hydrology of flood flow in the Neshaminy Creek basin, Pennsylvania. Jamison, PA: The Neshaminy Water Resources Authority (May).
16. BCPC. 1992. Neshaminy Creek watershed stormwater management plan, Vol. 1: Policy document, and Vol. II: Plan implementation. Doylestown, PA: Bucks County Planning Commission (January).
17. Clark, J. 1977. Coastal ecosystems: Ecological considerations for management of the coastal zone. Washington, DC: The Conservation Foundation.
18. New Jersey Department of Environmental Protection. 1988. The state of the ocean: A report by the blue ribbon panel in ocean incidents. Trenton, NJ.
19. Cahill, T.H., M. Adams, C.L. Smith, and J.S. McGuire. 1991. GIS analysis of nonpoint source pollution in the New Jersey coastal zone. With Whitney, S., and S. Halsey, New Jersey Department of Environmental Protection, Division of Coastal Resources. Presented at the National Conference on Integrated Water Information Management, Atlantic City, NJ (August).
20. Cahill, T.H., M. Adams, C.L. Smith, and J.S. McGuire. 1991. Living on the edge: Environmental quality in the coastal zone. With Whitney, S., and S. Halsey, New Jersey Department of Environmental Protection, Division of Coastal Resources. Presented at the International Conference on Integrated Stormwater Management, Singapore (July).
21. Martin, M. 1989. Ground-water flow in the New Jersey coastal plain. Open File Report 87-528. West Trenton, NJ: U.S. Geological Survey.
22. National Oceanic and Atmospheric Administration. 1989. Selected characteristics in coastal states, 1980-2000. Rockville, MD: U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Strategic Assessment Branch, Ocean Assessment Division (October).

Using GIS To Identify Linkages Between Landscapes and Stream Ecosystems

Carl Richards, Lucinda Johnson, and George Host
University of Minnesota, Duluth, Minnesota

Introduction

Factors that operate on a variety of temporal and spatial scales influence the structural and functional components of stream ecosystems (1). Quantifying the effects of factors that operate across multiple scales has challenged aquatic scientists over the last several decades. Recently, scientists have recognized that they cannot successfully protect or restore ecosystem integrity without taking into account all appropriate scales; therefore, they are focusing on understanding interactions between terrestrial and aquatic components of entire watersheds (2). Although awareness of the importance of watershed and landscape-scale influences on streams is growing, the tools to examine these influences are still in their infancy.

Most watershed and landscape studies to date have focused on the role watershed-scale parameters play on water chemistry (3-5). These studies usually examined nutrient and sediment inputs from various watershed land covers. Methods for evaluating the patterns in the terrestrial segment of the watershed were awkward and laborious, involving use of planimeters or cutting and weighing maps. More recent watershed studies have attempted to integrate both longitudinal and lateral influences of the terrestrial ecosystems on water quality in streams and wetlands (6-8). This approach takes advantage of newly available tools (geographic information systems and multivariate statistics) for quantifying landscape structure.

Relatively few studies have examined how watershed features influence biological communities. Most studies examining stream biota have concentrated on single land use types (9, 10) or on the relationship between watershed land use and stream physical habitat (11, 12). Typically, study designs have not addressed questions concerning variability of stream communities over relatively large geographic scales.

Our own work centers on identifying linkages between landscape features (watershed scale) and stream reach environments (physical habitat, chemistry), and relating these parameters to major patterns of community variation. In this manner, landscape and reach environment interactions probably control the influence that specific landscape components have on biological communities (13). This is the general premise when using biological communities to assess watershed status. To represent stream biota, we examine benthic macroinvertebrates, which have been used extensively for biomonitoring numerous environmental stresses (14). Macroinvertebrates are sensitive to watershed conditions and exhibit sufficient stability in assemblage structure over time to make them useful as long-term monitors of stream health (15).

This paper presents an overview of our attempts to identify the relative strengths of landscape variables on macroinvertebrate communities. We classify landscape variables into two general categories. The first category, geology and landscape structure (GEOS), considers variables that are fixed on the landscape and are largely uncontrollable by management activities. The second category, land use (LU), includes variables that have anthropogenic origins and may be influenced by land management activities. By understanding the relative strengths these two sets of variables possess in determining community structure, we hope to identify specific species groups that can act as land use and land form indicators. We also hope to identify ways to predict the outcome of specific large-scale land management activities (e.g., silviculture, agriculture) or other large-scale environmental changes (e.g., global warming) on stream ecosystems.

Study Area

This study was conducted in the Saginaw River basin, a 22,562-square-kilometer watershed in east-central Michigan (see Figure 1) that flows into Lake Huron. The Saginaw River watershed was chosen for this study because its component drainages range from heavily affected agricultural to relatively pristine areas.

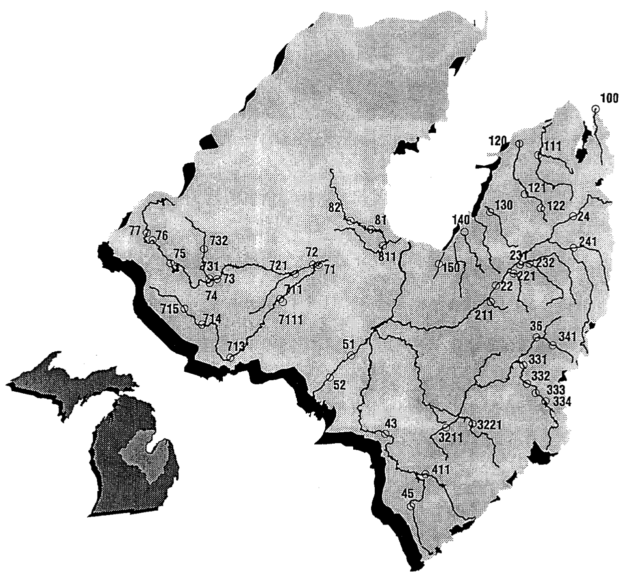


Figure 1. The Saginaw basin study area.

Dominating the soils in the lake plain are medium- and fine-textured loams to clays, with sand found in the outwash plains and channels. Artificial drainage and tile systems extensively drain the clay regions. Glacial features such as ground moraines and outwash plains are common. The western sector is characterized by rolling plains with coarse-textured ground moraines. This region contains a high percentage of the forested land, while agricultural land use dominates the eastern sector.

The Saginaw basin covers 16,317 square kilometers, including four major subbasins: the Tittabawassee (6,734 square kilometers), Shiawassee (3,626 square kilometers), Flint (3,108 square kilometers), and Cass (2,331 square kilometers) Rivers. The Tittabawassee subbasin further divides into three principal watercourses—the Chippewa, Pine, and Tittabawassee Rivers. Watersheds adjacent to Lake Huron (Kawkawlin and East basins) are characterized by low topographic relief and elevations averaging 203 and 206 meters, respectively. The Flint and Chippewa/Pine basins average about 278 meters in elevation. These drainages also exhibit the greatest variation in topography.

Study Design

The analysis covers 45 stream sites within the study area. These sites reflect a gradient of land use and physiographic conditions in the Saginaw River drainage. Researchers obtained biological, chemical, and physical samples at one 200-meter stream segment at each site. In addition, a geographic information system (GIS) database was compiled reflecting a number of landscape parameters for the watershed of each stream segment.

Sampling Methods

Macroinvertebrates

At each sampling site, we deployed Hester-Dendy artificial substrate samplers (16) for macroinvertebrate community characterizations twice, in early summer and during base flow conditions in late summer and fall of 1991 or 1992. We allowed samplers to colonize for 6 to 8 weeks. In the laboratory, macroinvertebrates were counted and identified to genus whenever possible. A series of derived variables from the original species abundance tables was used to describe community characteristics. We chose metrics based on their relative utility for examining macroinvertebrate communities, as suggested by Barbour et al. (17) and Karr and Kearns (18). Because macroinvertebrate assemblages are relatively stable through time (15) and preliminary analysis indicated no significant differences between sampling years at stations for which we had 2 years of data (unpublished data), we combined macroinvertebrate data into one database.

Chemistry

We assessed nutrients and other chemical properties related to water quality at each stream site during several periods in the summer and fall of 1991 or 1992. Stream flow during fall sampling was typically less than median flow rates and was considered to represent base flow levels. We used the maximum values of samples taken in June and July to represent summer conditions and the maximum values from September and October to represent fall base flow conditions. The nutrients measured were ammonium (NH_3), nitrate-nitrogen, total nitrogen (TN), orthophosphate (PO_4), and total phosphorus (TP). In addition, we assessed alkalinity (ALK), conductivity, total dissolved solids, and total suspended solids (TSS). Standard methods were used for all measurements (19).

Physical Habitat

We assessed physical habitat during base flow conditions at each stream site in a stream reach that is at least 8 to 12 times the width of the stream segment. A suite of quantitative habitat structure measurements and observations was made at each site. We derived values for six general habitat attributes:

- Substrate characteristics
- Instream cover
- Channel morphology
- Riparian and bank conditions

- Riffle/Run quality
- Pool quality

Landscape Descriptors

Land use patterns, surficial geology, hydrography, and elevation databases helped to quantify landscape characteristics in the study area (see Table 1). Land use patterns were derived from existing digital data at the Michigan Department of Natural Resources (Michigan Resource Information System [MIRIS] database) (see Table 2). We based classification of land use/cover categories on a modified version of the Anderson (20) scheme, which was constructed specifically for natural resource applications. The result was the following nine land use/cover categories:

- Urban
- Row crop/agriculture
- Other agriculture
- Herbaceous range land
- Shrubby range land
- Nonforested wetlands
- Forested wetlands
- Mixed hardwood forests
- Deciduous forests

In this region, nonrow-crop agriculture is largely represented by pasture, and range lands are predominantly abandoned fields (old fields).

The U.S. Department of Agriculture (USDA) STATSGO soils database enabled the compilation of soil data. The database consists of U.S. Soil Conservation Service (SCS) soil surveys and includes information on dominant texture and drainage in large landscape units. We aggregated soils into simplified categories based on glacial landform.

We delineated watershed boundaries for each sampling station on United States Geological Survey (USGS) topographic maps and digitized them using ARC/INFO (Environmental Systems Research Institute [ESRI], Redlands, California). We identified stream order for each stream segment and coded it as an attribute of the stream reach file. All databases were transformed into a common digital format as necessary and projected into a common coordinate system. We stored data in vector format and analyzed them in ARC/INFO.

Table 2 lists the landscape variables we derived for each watershed. Land use/cover values were reported and analyzed as a percentage of the total watershed area. Patch heterogeneity measured landscape fragmentation and was reported as the number of patches per hectare. We derived slope from elevation data using ARC/INFO. The standard deviation of elevation was used as a surrogate measure of topographic variability.

Statistical Analysis

Using redundancy analysis (RDA), a canonical extension of principal component analysis (PCA), we detected relationships among the individual multivariate data sets. RDA is a form of direct gradient analysis that describes variation in a multivariate data set (e.g., habitat variables or macroinvertebrate metrics) based upon environmental data (21). In RDA, the station scores from a PCA are regressed on a specified set of environmental variables with each iteration, and the fitted values of the regression become new station scores (22). Thus, environmental or predictor variables constrain PCA. Two important outputs from this method are the interset correlations of environmental variables with the RDA axes, which indicate the environmental variables that have the strongest influence in the ordination, and the fraction of

Table 1. Spatial Data Used for Landscape Characterization

Data Layer	Source	Scale	Format Received
Hydrology	U.S. EPA stream reach	1:100,000	ARC/INFO
Elevation	USGS-DEM	1:250,000	Digital elevation model
Land use/cover	USGS	1:100,000	Digital line graph
Land use/cover	MIDNR	1:24,000	Intergraph
Watershed boundary	USGS topographic maps	1:24,000	Manual delineation
Station locality	USGS topographic map	1:24,000	Manual digitizing
Soils	USDA SCS STATSGO	1:250,000	ARC/INFO
Major basin	USGS topographic	1:24,000	Manual digitizing
Quartenary geology	University of Michigan	1:250,000	Manual digitizing

Key: USGS = United States Geological Survey
DEM = Digital elevation model
MIDNR = Michigan Department of Natural Resources
USDA SCS = United States Department of Agriculture Soil Conservation Service

Table 2. Landscape Variables Measured or Derived for Each Watershed

Variable	Units
Land use/cover	Proportional area
Slope	Degrees
Standard deviation elevation	Meters
Patch heterogeneity	Patches/hectare
Soils	Proportional area
Total area	Hectares

total variance of each predicted variable that is explained by the RDA axes (22).

Performing Monte Carlo permutation tests determined the statistical validity of the association between predictor and predicted variables. Tests were conducted by random permutation of the site numbers in the predictor variables. We randomly linked the predictor data to the predicted data. We conducted 99 simulations to approximate a normal distribution with which to compare our data with random combinations.

We first determined which of the reach variables had strong influences on macroinvertebrate distributions by conducting separate RDAs with physical habitat and chemistry variables as environmental descriptors. We then examined the ability of the landscape data to predict the variation in the important reach variables.

To determine the relative influences of LU and GEOS landscape variables on stream chemistry and physical habitat, we used partial RDA, where one landscape variable type was held constant and variation due to the other landscape set was examined independently. Using this approach, total variation in a multivariate data set can be decomposed in a manner analogous to analysis of variance (23, 24). For this analysis, we attributed variation in the reach variables (habitat and chemistry) to four separate components:

- The variation in reach variables that LU variables explained independently of GEOS variables.
- The variation in reach variables that GEOS variables explained independently of LU variables.
- The variation in reach variables that both GEOS and LU variables shared. This shared variation could have been due to both the dependence of one type of variable on the other as well as noncausal relationships (e.g., the types of soil found in a watershed determine in large part the types of agriculture that can be practiced).
- The variation in reach variables that were unexplainable. This may have been attributable to sampling error, stochastic variation, or other variables not sampled.

Results

Regional Characteristics

Land Use

Land use within the study region was dominated by row-crop agriculture (see Table 3). Individual watersheds ranged from 14 to 99 percent in agricultural land uses, with the East basin watersheds exhibiting the greatest proportion of agricultural land use and the Flint having the lowest proportion of agricultural land use. The Chippewa/Pine and Kawkawlin watersheds exhibited the greatest diversity of land use and cover types within the study region.

Wetlands represented a minor land use component with most watersheds having between 0 and 15 percent land area. The Cass and Kawkawlin basins had the greatest proportion of wetlands, with a median of 6.8 percent for individual watersheds.

Macroinvertebrates

Considerable variation existed among the major basins with respect to the 15 macroinvertebrate community metrics during summer (see Table 4). Metric values for the Flint, Shiawassee, and Chippewa watersheds were similar. Sites within the Kawkawlin and East basins differed considerably from the Flint, Shiawassee, and East basins in several of the metrics. The Kawkawlin watershed was notable for high shredder and filterer proportions and a low proportion of detritivores. The East basin also had a high proportion of shredders. Both the East and Kawkawlin basins had lower proportions of strictly erosional taxa and higher proportions of depositional taxa than the other major basins.

Taxa at the East and Kawkawlin basins also exhibited lower oxygen tolerance than at other major basins. In addition, their Hilsenhoff Biotic Index (HBI) scores (which are sensitive to oxygen availability) were higher than other basins, and they had the lowest EPT (Ephemeroptera, Plecoptera, Trichoptera) richness. Total richness at Kawkawlin was relatively high, however. Richness was highest in the Chippewa/Pine watershed and lowest in the East basin.

In general, fall patterns of macroinvertebrate metrics resembled those of summer. The Kawkawlin and East basins had high HBI scores, low EPT scores, low proportions of erosional taxa, and high proportions of depositional taxa. The proportion of predators was exceptionally high in the Kawkawlin basin due to the abundance and trophic classification of one chironomid genus.

Table 3. Summary of Landscape Metrics in Six Major Basins of the Saginaw River Drainage

Landscape Variables	East Basin	Cass	Flint	Shiawassee	Chippewa/Pine	Kawkawlin
n	**8**	**7**	**8**	**5**	**15**	**3**
Row crops	86.4 79.7-98.0	58.4 45.1-73.3	38.0 25.6-65.5	43.5 26.7-71.8	48.2 13.9-91.3	26.4 18.5-74.9
Other agricultural land	0.2 0.0-1.2	0.6 0.4-1.4	3.1 0.6-4.1	0.4 0.1-2.0	4.0 0.3-6.6	1.9 0.2-2.3
Urban	0.8 0.1-2.0	1.9 0.4-3.7	8.56 1.4-23.2	10.3 2.4-15.0	2.1 0.8-4.0	1.2 1.0-6.8
Deciduous forest	6.4 0.5-12.1	16.9 5.8-33.4	17.5 10.2-20.5	19.2 16.4-30.4	23.2 3.2-42.1	56.8 13.2-64.5
Mixed hardwood	0.03 0.0-0.3	0.3 0.01-1.7	2.6 0.0-3.8	0.7 0.4-0.9	5.4 0.1-8.5	0.3 0.1-0.3
Range: Herb	1.0 0.1-3.6	6.7 2.1-12.7	13.0 6.6-14.5	12.6 0.9-17.3	6.9 1.0-9.0	2.5 0.7-3.1
Range: Shrub	1.3 0.0-2.3	4.7 3.2-7.7	6.5 2.6-10.5	7.3 2.2-10.1	5.6 0.7-9.0	3.0 2.7-3.2
Forested wetlands	0.1 0.0-0.5	0.1 0.0-0.3	2.3 0.3-3.2	0.7 0.0-3.7	1.0 0.0-2.5	1.8 0.1-2.2
Non-forested wetlands	1.7 0.0-5.3	6.7 1.9-14.5	3.9 0.4-5.2	3.6 0.7-5.0	4.3 0.1-7.9	5.0 1.3-5.9
Slope (degrees)	0.15 0.07	0.29 0.12	0.41 0.11	0.27 0.10	0.35 0.10	0.14 .03
Elevation (meters)	206.5 19.8	239.8 6.7	277.2 22.5	252.6 47.1	278.7 34.9	203.1 9.8
Patch heterogeneity	241.5 102.0	711.4 250.6	950.4 228.6	762.7 223.6	703.6 185.9	519.6 156.7
Watershed area (hectares)	14,968.9 11,132.0	38,117.7 58,321.9	28,926.1 19,236.7	46,530.0 50,798.0	53,704.2 44,872.8	22,240.2 3,848.0

Land use/cover variables (agricultural land through nonforested wetlands) are reported as median and range; landscape structure variables (slope through watershed area) are reported as mean and standard deviation. Land use/cover represents proportional areas of each watershed.

Identification of Important Reach-Scale Variables

Chemistry

RDA showed that chemical variables explained 26 percent of the variation in macroinvertebrate data in summer and 33 percent in fall. The most important variables in summer were TN and TSS (see Table 5). Fall macroinvertebrate communities were influenced by a greater number of variables, including NH_3, TP, ALK, and TSS.

Physical Habitat

The 13 physical habitat variables explained 37 percent of the macroinvertebrate data in summer and 46 percent of the macroinvertebrate data in fall. In summer, the percentage of deep pools and canopy extent along with channel dimensions, such as bank full width (BFW) and bank full depth (BFD), were the most important variables. In fall, the percentage of fines and deep pools as well as canopy extent were among the most important variables (see Table 6).

Landscape Influences on Surface Water Chemistry

In summer, the landscape data explained 55 percent of the variation in chemical variables. The proportion attributable to LU was larger than that attributable to GEOS (see Figure 2). The two data types shared 12 percent of

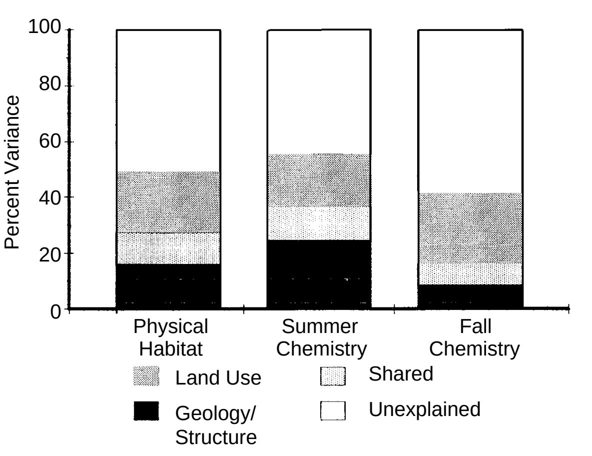


Figure 2. Results of variance decomposition from partial RDA.

Table 4. Mean and Standard Deviation of Macroinvertebrate Metrics Calculated for Summer Collection Periods for Six Major Basins of the Saginaw River Drainage

	East Basin	Cass	Flint	Shiawassee	Chippewa/Pine	Kawkawlin
n	**8**	**7**	**8**	**5**	**15**	**3**
Chironomidae	59.1	57.9	45.9	32.2	45.5	67.1
	35.3	20.3	27.9	28.5	29.6	18.6
Omnivores	19.4	19.1	18.1	14.4	21.5	22.0
	13.7	7.1	13.9	3.8	9.8	16.7
Detritivores	57.1	69.7	75.3	79.9	70.4	29.0
	34.1	9.7	16.0	6.4	9.8	26.1
Shredders	30.3	18.7	10.6	7.7	14.4	51.0
	33.9	5.1	6.0	6.8	15.6	26.9
Gatherers	59.8	18.7	64.3	65.0	65.8	39.8
	32.9	5.1	12.8	14.1	17.2	30.1
Filterers	27.4	23.4	22.4	18.9	17.6	39.9
	38.1	18.5	14.1	12.8	15.1	35.7
Grazers	32.2	13.6	26.2	40.1	25.5	25.4
	34.4	16.2	21.0	22.7	21.1	22.9
Predators	1.5	1.2	1.5	1.0	1.4	1.9
	2.2	1.2	2.0	1.0	0.8	0.8
2 Dominants	64.5	54.3	50.3	54.2	51.9	60.0
	25.5	6.1	15.2	9.5	11.6	21.7
Total abundance	2077	650	574	325	497	433
	4951	739	622	91	230	297
HBI	7.1	5.6	5.6	6.0	5.1	8.1
	1.4	?	0.8	0.5	1.1	0.8
Erosional taxa	25.9	36.1	35.5	38.9	36.1	14.9
	12.4	5.5	9.5	14.7	11.0	5.3
Depositional	35.5	23.7	27.5	27.0	25.4	52.4
taxa	13.2	9.6	11.5	6.6	10.7	6.5
Species	17.2	18.3	22.1	20.6	26.6	23.3
richness	4.5	9.6	8.2	4.7	3.0	4.9
EPT taxa	5.0	5.7	7.3	8.0	10.0	3.3
richness	2.7	2.8	3.0	2.3	3.7	0.5

the variation. The relationship between LU variables and chemistry was significant ($p < 0.05$), and the relationship between GEOS variables and chemistry was not significant ($p > 0.05$).

In fall, variation explained by LU was proportionally less than during summer (see Figure 2). GEOS landscape variables explained 25 percent of the total variation while LU variables accounted for less than 10 percent of the total variation. LU and GEOS variables shared approximately 8 percent of the variation. In contrast with summer, GEOS variables were significant and LU variables were not significant when examined with the Monte Carlo test ($p < 0.05$).

The importance of GEOS and LU variables in explaining variation in the chemistry variables, as well as the total amount of variation explained, differed considerably among the chemistry variables. Figure 3 shows only summer data. For example, the landscape variables explained almost 80 percent of the total variation in TN (see Figure 3). The largest proportion of variance was explained by the shared influences of GEOS and LU. Alkalinity, which was also well predicted, however, was much more influenced by variation attributable to LU variables. In comparison, LU variables explained less than 45 percent of the total variance in TP, and the majority of this variance was attributable to GEOS.

To further examine the influence of specific landscape variables, we compared the various axes in the significant partial ordinations (see Figure 4). In summer, when we observed a significant effect of LU variables on water chemistry, forested land covers and nonrow-crop agriculture had their greatest influence on TSS and ALK. LU heterogeneity and shrub vegetative cover most strongly influenced both TN and NH_3. In fall, when GEOS variables had a significant relationship to water chemistry, ALK and NH_3 were more influenced by peat land soils and watershed size. The proportion of sand and gravel soils, as well as clays, explained much of the variation in TN and TSS.

Table 5. **Chemistry Variables That Had a Correlation (r) of at Least 0.30 With One of the RDA Axes With the Summer or Fall Ordinations; a Monte Carlo Analysis Indicated That Both Summer and Fall Ordinations Were Significant ($p < 0.05$)**

	RDA 1		RDA 2		RDA 3	
Variable	**Summer**	**Fall**	**Summer**	**Fall**	**Summer**	**Fall**
TN	0.44	0.03	-0.04	-0.07	0.05	-0.14
NH_3	0.05	0.04	0.09	0.52	0.07	-0.09
TP	0.12	0.40	0.07	0.26	0.25	0.14
PO_4	-0.04	0.19	0.01	-0.08	0.26	0.34
TSS	0.01	-0.31	0.43	0.36	0.22	-0.13
ALK	-0.08	-0.36	-0.16	0.09	0.15	0.2

Table 6. **Physical Habitat Variables That Had Correlations (r) Over 0.3 With the Ordination Axes; Results of the Monte Carlo Simulation Indicated That the Fall but Not the Summer Ordinations Were Significant ($p < 0.05$)**

	RDA 1		RDA 2		RDA 3	
Variable	**Summer**	**Fall**	**Summer**	**Fall**	**Summer**	**Fall**
Percentage of fines	0.27	0.47	0.04	0.28	0.18	-0.09
Percentage of shallows	-0.06	0.17	-0.28	-0.11	0.32	-0.38
Wood	0.09	0.34	-0.1	0.02	0.10	-0.17
Percentage of deep pools	0.50	0.54	0.16	0.09	-0.17	0.26
Erosion	0.16	0.30	-0.3	-0.31	0.01	0.27
Maximum depth	0.08	-0.14	0.08	0.16	-0.39	0.13
Canopy extent	0.75	0.21	-0.47	-0.42	0.21	-0.25
BFW	-0.10	-0.23	0.52	0.36	0.07	0.03
BFD	-0.09	-0.13	0.32	0.02	-0.08	0.22
Flood ratio	0.03	-0.30	0.07	0.23	-0.41	0.02

Landscape Influences on Physical Habitat

GEOS landscape variables attributed for the largest portion (22 percent) of the explained variation in physical habitat variables (see Figure 2). LU variables accounted for 16 percent of the explained variance. The partial ordination for GEOS but not LU was significant as the Monte Carlo procedure determined.

As noted with chemistry variables, there were considerable differences in the ability of the landscape variables to predict individual habitat characteristics. Landscape variables were best at predicting BFW and least powerful for predicting the percentage of deep pools (see Figure 5). BFW was influenced predominantly by GEOS and only minimally by LU. The most influential GEOS variables for BFW related to watershed area (see Figure 6). Woody debris was predominantly influenced by LU variables. The most influential LU variables for woody debris related to forested wetlands. Flood ratio was intermediate to these examples. Both sets of landscape variables shared the largest proportion of explained variance for this parameter.

Discussion

Our studies demonstrate the distinct influences landscape features have on stream macroinvertebrate communities through modifying surface water chemistry and stream habitat. Land use most strongly influences stream chemistry during summer months when surface runoff and soil leaching are greatest. In addition, fertilizer application in row-crop agriculture is highest in the first part of the growing season. The strong relationship between some aspects of land use and stream water chemistry were similar to those observed in other studies (4, 6, 7). The specific mechanism by which stream chemistry influences macroinvertebrates is not clear. The addition of nutrients can significantly affect stream productivity (25-27); however, light often limits primary production in agricultural areas (28-30). Nutrients may

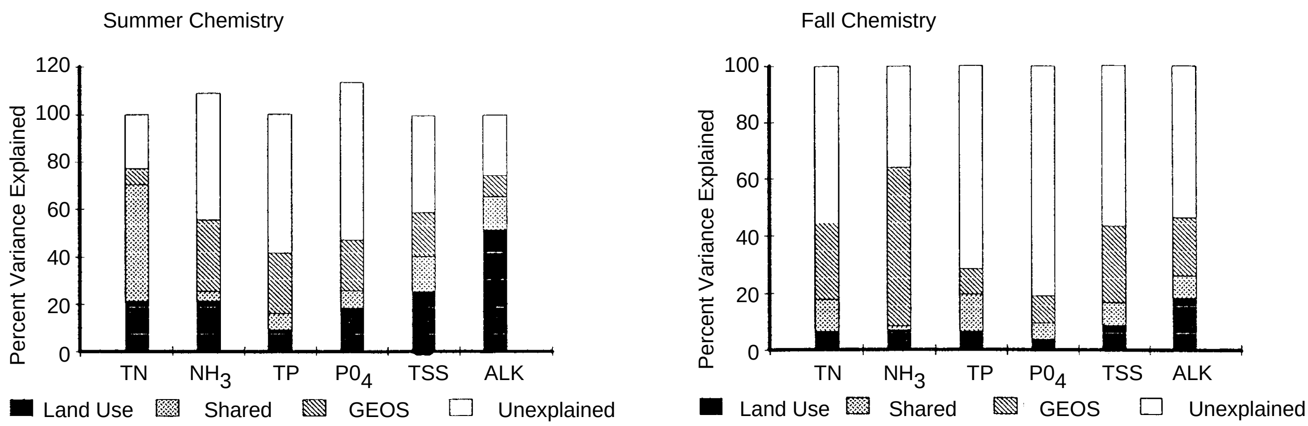


Figure 3. Results of variance decomposition for chemical parameters from partial RDA.

also act as indicators of other agricultural chemicals (e.g., herbicides, pesticides) and carbon sources, which we did not measure in our study, but their inputs are moderated by agricultural land use.

The influence of landscape features on stream habitat may be even more complex than stream chemistry can show. Low-gradient midwestern streams in the United States are under a complex set of controls related not only to subtle topographic gradients but also to interactions among land use, soil type, riparian vegetation, and historic anthropogenic modifications. Ditching, channelization, and hydrologic changes associated with large-scale land-cover changes, such as logging and wetland drainage, have a significant effect on current stream physical conditions.

Our analysis did not quantify many of these effects. Nonetheless, some land use influences seem clear. First, we found that the extent of permanent vegetation in a watershed played a central role in land use impacts on stream habitats. All of the key factors that described variation in the habitat variables included land covers with permanent vegetation. In addition, our initial land-cover analysis found that row-crop agriculture highly negatively correlated with almost all permanently vegetated land covers. Second, although wetlands represented a relatively small proportion of total watershed area (less than 5 percent), they had a relatively high influence on habitat features. Both forested and non-forested wetlands were important to habitat features such as woody debris and some aspects of channel dimensions.

We found that one of the most important landscape influences on stream habitat was watershed area, which explained much of the variation in channel dimensions. These habitat parameters, in turn, were among the most important for explaining variation among macroinvertebrate communities. Stream size, including channel dimension, is an important determinant of several macroinvertebrate community characteristics that stem from longitudinal phenomena associated with stream ecosystems (31, 32). Because these relationships are fixed, they are not amenable to manipulation by management. Consequently, if the monitoring objective is to

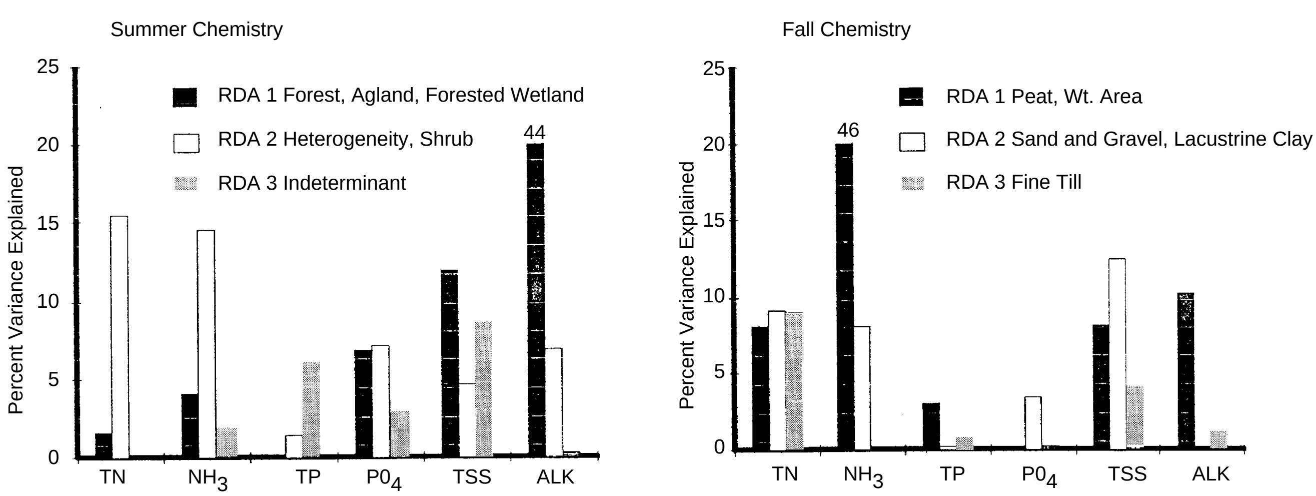


Figure 4. Factors influencing chemical parameters in summer (land use) and fall (geology/structure).

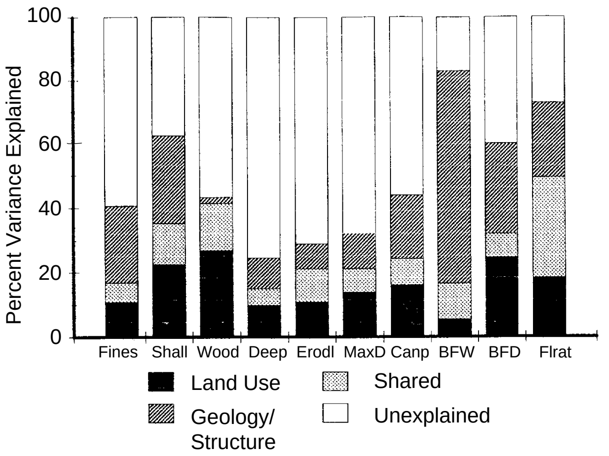


Figure 5. Results of variance decomposition from partial RDA for physical habitat.

detect changes resulting from land use, then specific macroinvertebrate community characteristics that respond to changes in channel dimensions are not good biological indicators.

Our study indicates that land use and geology/structure have similar magnitudes of influence on reach environments and biotic communities. This suggests that although some community and water quality characteristics are controlled by landscape features that are difficult to modify or manage, other aspects may be altered through specific activities designed to increase or decrease land uses for a desired effect.

The methods described in this study show promise for examining landscape-stream interactions. Much of the variation we found in both stream reach environments and macroinvertebrate communities, however, remained unexplained. Undoubtedly, we can attribute much of this unexplained variation to the abbreviated sampling procedures that large-scale field studies such as this dictate. Questions of scale and position, however, may also be responsible for much unexplained variation.

The minimum mapping unit used in our study (1 hectare) may be insufficient to detect local land use practices that influence streams. For example, riparian zones along many streams may only be a few meters in width and yet exert considerable control over nutrient inputs (33). Higher spatial resolution may improve predictive power. In addition, the position of certain landscape elements may determine consequent effects on stream conditions. Johnston et al. (7) found that wetland position within a watershed influenced the ability of the wetland to modify stream chemistry. With respect to biologic communities, the abundance of some taxa may relate to the presence of upstream refugia, which can provide sources of colonists (34) and other biogeographic effects.

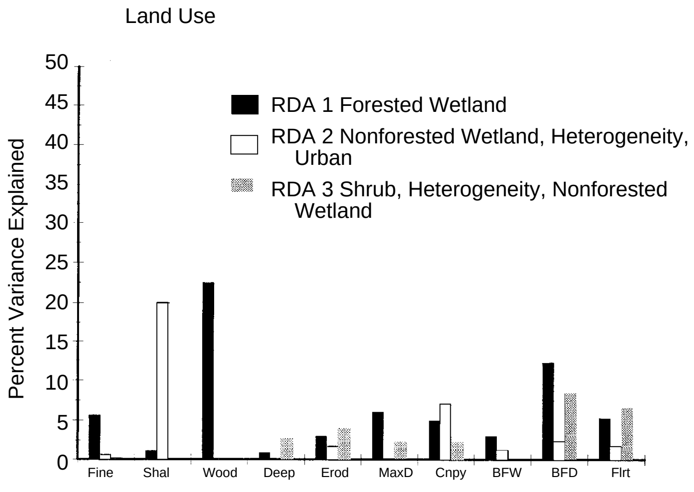
The further development of empirical relationships between landscape parameters and streams will provide important insights into understanding the spatial and temporal scales needed for modeling and monitoring large watersheds. Government and commercial geographic databases that can be easily adapted to GIS technology should become more readily available. As they do, the ability to identify relationships across scales, and consequently predict stream community composition from watershed-scale attributes, will facilitate the assessment of ecological risk associated with land use changes.

Acknowledgments

Jim Wessman, Frank Kutka, and Tim Aunan developed much of the data reported in this study. John Arthur and Tom Roush collected most of the chemistry and macroinvertebrate data and provided invaluable logistical support. Cajo ter Braak provided helpful insights to statistical analysis. This research was funded by the U.S. Environmental Protection Agency through CR-818373-01-0.

References

1. Minshall, G.W. 1988. Stream ecosystem theory: A global perspective. J. North Am. Benthological Soc. 7:263-288.
2. Naiman, R.J. 1992. New perspectives for watershed management: Balancing long-term sustainability with cumulative environmental change. In: Naiman, R.J., ed. Watershed management. New York, NY: Springer-Verlag. pp. 3-11.
3. Likens, G.E., F.H. Bormann, N.M. Johnson, D.W. Fisher, and R.S. Pierce. 1970. Effects of forest cutting and herbicide treatment on nutrient budgets in the Hubbard Brook Watershed ecosystem. Ecol. Monogr. 40:23-47.
4. Correll, D.L., E.D. Friebele, and J. Miklas. 1977. Nutrients in land runoff from Rhode River watersheds in 1975 and 1976. In: Nonpoint source studies on Chesapeake Bay, Vol. II. Publication No. 55. Chesapeake Bay Center for Environmental Studies/Smithsonian Institution, Edgewater, MD.
5. U.S. EPA. 1976. The influence of land use on stream nutrient levels. EPA/600/3-76/014. Washington, DC.
6. Osborne, L.L., and M.J. Wiley. 1988. Empirical relationships between land use/cover patterns and stream water quality in an agricultural watershed. J. Environ. Mgmt. 26:9-27.
7. Johnston, C.A., N.E. Detenbeck, and G.J. Niemi. 1990. The cumulative effect of wetlands on stream water quality and quantity: A landscape approach. Biogeochemistry 10:105-141.
8. Detenbeck, N.E., C.A. Johnston, and G.J. Niemi. 1993. Wetland effects on lake water quality in the Minneapolis/St. Paul metropolitan area. Landscape Ecol. 39-61.
9. Duda, A.M., D.R. Lenat, and D.L. Penrose. 1982. Water quality in urban streams: What can we expect? JWPCF 54:1,139-1,147.



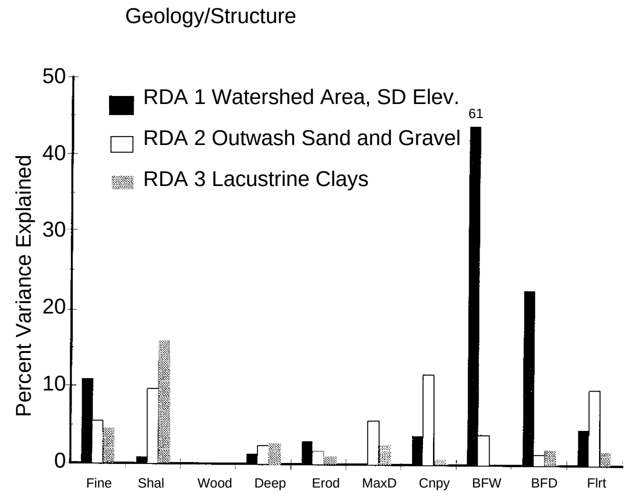


Figure 6. Factors influencing physical habitat parameters.

10. Jones, R.C., and C.C. Clark. 1987. Impact of watershed urbanization on stream insect communities. Water Resources Bull. 23:1,047-1,055.

11. Bisson, P.A., R.E. Bilby, M.D. Bryant, C.A. Dolloff, G.B. Grette, R.A. House, M.L. Murphy, K.V. Koski, and J.R. Sedell. 1987. Large woody debris in forested streams in the Pacific Northwest: Past, present, and future. In: Salo, E.O., and T.W. Cundy, eds. Streamside management: Forestry and fishery interactions. Contribution 57. Institute of Forest Resources, University of Washington, Seattle, WA. pp. 143-190.

12. Everest, F.H., R.L. Beschta, J.C. Scrivener, K.V. Koski, J.R. Sedell, and C.J. Cedeholm. 1987. Fine sediment and salmonid production: A paradox. In: Salo, E.O., and T.W. Cundy, eds. Streamside management: Forestry and fishery interactions. Contribution No. 57. Institute of Forest Resources, University of Washington, Seattle, WA. pp. 98-142.

13. Richards, C., and G.E. Host. 1994. Using GIS to examine the influence of land use on stream habitats and biota. Water Resources Bull. 30:729-738.

14. Rosenberg, D.M., and V.R. Resh. 1993. Freshwater biomonitoring and benthic macroinvertebrates. New York, NY: Chapman and Hall.

15. Richards, C., and G.W. Minshall. 1992. Spatial and temporal trends in stream macroinvertebrate species assemblages: The influence of watershed disturbance. Hydrobiologia 241:173-184.

16. Ohio EPA. 1987. Biological criteria for the protection of aquatic life, Vol. II. User's manual for biological assessment of Ohio surface waters. Columbus, OH.

17. Barbour, M.T., J.L. Plafkin, B.P. Bradley, C.G. Graves, and R.W. Wisseman. 1992. Evaluation of EPA's rapid bioassessment benthic metrics: Metric redundancy and variability among reference stream sites. Environ. Toxicol. Chem. 111:437-449.

18. Karr, J.R., and B.L. Kearns. 1992. Components of biological integrity: Their definition and use in development of an invertebrate IBI. In: Simon, T.P., and W.S. Davis, eds. Proceedings of the 1991 Midwest Pollution Control Biologists Meeting: Environmental indicators: Measurement and assessment endpoints. EPA/905/R-92/003. U.S. EPA Region 5, Environmental Sciences Division, Chicago, IL. pp. 1-16.

19. APHA. 1989. Standard methods for the examination of water and waste water. American Public Health Association.

20. Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer. 1976. A land use and land cover classification system for use with remote sensor data. Geological Survey Professional Paper 964. U.S. Government Printing Office, Washington, DC.

21. ter Braak, C.J., and I.C. Prentice. 1988. A theory of gradient analysis. Advances in Ecol. Res. 8:271-317.

22. Jongmann, R.H., C.J. ter Braak, and O.F. van Tongeren. 1987. Data analysis in community and landscape ecology. Wageningen, The Netherlands: Pudoc.

23. Borchard, D., P.L. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. Ecology 73:1,045-1,055.

24. ter Braak, C.J. 1990. Update notes: CANOCO Version 3.0. Agricultural Mathematics Group, Wageningen, The Netherlands.

25. Petersen B.J., J.E. Hobbie, T.J. Corliss, and K. Kriet. 1983. A continuous-flow periphyton bioassay: Tests of nutrient limitation in a tundra stream. Limnology and Oceanography 28:583-591.

26. Grimm, N.B., and S.G. Fisher. 1989. Stability of periphyton and macroinvertebrates to disturbance by flash flooding. Ecol. Monogr. 52:93-110.

27. Pringle, C.M. 1987. Effects of water and substratum nutrient supplies on lotic periphyton growth: An integrated bioassay. Can. J. Fish. Aquat. 44:619-629.

28. Munn, M.D., L.L. Osborne, and M.J. Wiley. 1989. Factors influencing periphyton growth in agricultural streams of central Illinois. Hydrobiologia 174:89-97.

29. Bushong, S.J., and R.W. Bachmann. 1989. In situ enrichment experiments with periphyton in agricultural streams. Hydrobiologia 178:1-10.

30. Wiley, M.J., L.L. Osborne, and R.W. Larimore. 1990. Longitudinal structure of an agricultural prairie river system and its relationship to current stream ecosystem theory. Can. J. Fish. Aquat. 47:373-384.

31. Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell, and C.E. Cushing. 1980. The river continuum concept. Can. J. Fish. Aquat. 37:130-137.

32. Statzner, B., J.A. Gore, and V.H. Resh. 1988. Hydraulic stream ecology: Observed patterns and potential applications. J. North Am. Benthological Soc. 7:307-360.

33. Osborne, L.L., and D.A. Kovacic. 1993. Riparian vegetated buffer strips in water-quality restoration and stream management. Freshwater Biol. In press.

34. Sedell, J.R., G.H. Reeves, F.R. Hauer, J.A. Standford, and C.P. Hawkins. 1990. Role of refugia in recovery from disturbances: Modern fragmented and disconnected river systems. Environ. Mgmt. 14:711-724.

Nonpoint Source Water Quality Impacts in an Urbanizing Watershed

Peter Coffin, Andrea Dorlester, and Julius Fabos
University of Massachusetts at Amherst, Amherst, Massachusetts

Abstract

As part of the larger Narragansett Bay Estuary Project, the University of Massachusetts Cooperative Extension Service contracted with the university's METLAND research team to develop a geographic information system (GIS) database, generate watershed-wide maps, perform analyses, and develop a modeling procedure. The objective was to educate local officials about the impacts of development on water quality and to help local boards minimize the effect of nonpoint sources of pollution.

Because the receiving waters of the Narragansett Bay are located far downstream in Rhode Island, the upstream communities in Massachusetts are reluctant to enact measures to improve water resources outside of their jurisdictions. A GIS was used to create awareness of existing downstream problems and to show the upstream communities how development will ultimately affect water resources in their own backyards.

To nurture this awareness, a "buildout" analysis was conducted for an entire upstream subwatershed, the Mumford River watershed, containing parts of four towns, and roughly 50 square miles. This buildout was coupled with a loading model using Schueler's Simple Method to illustrate the potential impacts of future development, and encourage local boards to minimize future nonpoint sources of pollution.

GIS proved its usefulness by developing customized maps for each town, by generating several "what if" scenarios showing the impacts of different zoning changes, by facilitating long-range planning for small towns without professional staff, and by encouraging a regional perspective on development issues. The entire planning process was most successful in creating a series of partnerships that will continue after the grant expires. The university shared coverages with the state GIS agency, creating new coverages not previously available, specifically soils, ownership, and zoning. Small towns learned about the potential of the new technology. Students gained from hands-on experience with real-world problems. State agencies saw their efforts understood at the local level, especially as they reorganize on a basin approach and begin to implement a total mass daily loading (TMDL) procedure to coordinate permitted discharges and withdrawals.

As greater emphasis is placed on controlling nonpoint sources of pollution, more attention needs to be focused on local boards, who control land use decisions in New England.

Introduction

Project Description

Narragansett Bay is a vital resource for southern New England. The health of its waters is critical to the regional economy, supporting fisheries, tourism, and quality of life. Increased development along the bay's shorelines and throughout its drainage basin threatens the quality of these waters, however. The U.S. Environmental Protection Agency (EPA) recognized the threats to this important water body and designated the Narragansett Bay under its National Estuary Program in 1985.

Completing a Comprehensive Conservation and Management Plan (CCMP) for Narragansett Bay took 7 years. The CCMP identified seven priority areas for source reduction or control, including the reduction of agricultural and other nonpoint sources of pollution. The nonpoint source strategy identified United States Department of Agriculture (USDA) agencies, conservation districts, and other public and private organizations as having principal roles in nonpoint source management.

Whereas the vast majority of Narragansett Bay lies within the boundaries of Rhode Island, a significant portion of its pollution load originates in Massachusetts. Recognizing that the watershed extends beyond state boundaries, the USDA provided 3 years of funding to Cooperative Extension and the Soil Conservation Service (SCS) in both Massachusetts and Rhode Island to coordinate their efforts in an innovative attempt to reduce the impact of nonpoint sources of pollution on Narragansett Bay. While water quality is a relatively new focus for Cooperative Extension, it fits well with the historic mission of extending the knowledge base of the land-grant colleges out into the community, and providing training and capacity building for local officials and community organizations.

With such a large area of concern, the management team decided to focus on a smaller subwatershed area in each state for the first 2 years. The strategy was first to develop a program for the mitigation of nonpoint source pollution on the smaller scale of a watershed of roughly 50 square miles, then take the lessons learned and apply the most appropriate efforts throughout the larger watershed. By using similar strategies in Rhode Island and Massachusetts, but choosing subwatersheds that differ in terms of location relative to the receiving water, size, staffing, and sophistication, the two states gained from each other's experience, sharing the successful techniques and avoiding each other's mistakes.

For its pilot study, Rhode Island chose Aquidneck Island, home of Newport, Portsmouth, and Middletown, with a special focus on protecting surface water supply reservoirs. Massachusetts chose an upstream watershed in the Blackstone Valley, somewhat rural in character, but rapidly undergoing a transformation to suburbia.

Watershed Description

The Blackstone River drops 451 feet in its 48-mile journey from Worcester, Massachusetts, to Pawtucket, Rhode Island. In the 19th century, this drop of roughly 10 feet per mile was ideally suited to providing power to mills during the early years of the industrial revolution. By the Civil War, every available mill site was developed, earning the Blackstone River the name "The Hardest Working River."

The Blackstone has a long history of pollution. First, the textile industry, then steel, wire, and metal finishing industries used the river for power, in their manufacturing process, and for waste disposal.

In Massachusetts, the Blackstone River is the major source of many pollutants to Narragansett Bay. Based on total precipitation event loading calculations, the Blackstone River is the principal source of solids, cadmium, copper, lead, nitrate, orthophosphate, and PCBs to the bay (1). The Blackstone River has an average flow of 577 million gallons per day or 23.2 percent of the freshwater input to the bay.

The watershed area in Massachusetts equals 335 square miles; with a population of 255,682, this results in a density of 763 people per square mile. The Blackstone Valley has 9,000 acres in agricultural use, with more land in hay (4,500 acres) than crops (3,700 acres) to support its 4,400 animals.

Based on aerial mapping flown in 1987, the Blackstone Valley has lost 5 percent of its cropland, 9 percent of its pasture, and 21 percent of its orchards since 1971. The valley remains more than 60 percent forested, but that represents a decrease of 5 percent. The forest and farmlands were lost to development as low density housing grew by 45 percent, commercial use grew by 15 percent, and transportation grew by 54 percent. Waste disposal grew 52 percent to 582 acres, and mining, which in this region represents gravel pits, grew 22 percent to 1,100 acres.

Watershed soils consist mainly of compact glacial till on rolling topography, with 3 to 15 percent slopes. The river and stream valleys are underlain by glacial-derived sand and gravel outwash, which provide drinking water to all towns in the area except Worcester and support the large gravel pits. The high clay content in the till soils of the uplands makes for a high water table, which is beneficial for growing corn but causes problems for septic systems.

Following a preliminary study of the subwatersheds, the Mumford River in the Blackstone Valley was selected as the focus watershed based on its size, location, land use, and existing water quality (see Figure 1). The Mumford River watershed has an area of 57 square miles, with a length of 13 miles, and lies within the towns of Douglas, Northbridge, Sutton, and Uxbridge. These towns share the attributes of small, rural communities undergoing rapid development, with no professional planning staff (see Figure 2). According to the 1990 Census, Douglas grew 46 percent in 10 years to 5,438; Uxbridge experienced 24 percent growth to 10,415; Sutton increased 17 percent to 6,824; and Northbridge grew 9 percent to 13,371.

Project Strategy

Because the generation of nonpoint sources of pollution is so closely tied to land use, and because local boards composed of citizen volunteers have principal control over land use in New England, the key focus of this program is to train local boards to recognize and begin managing the threat that nonpoint sources of pollution pose to water quality. Local planning boards, conservation commissions, and boards of health address land use issues and can regulate and shape existing and proposed development. By developing a program to train local officials, Cooperative Extension can focus its outreach where it will have the greatest impact in both the short and long term. Local boards have the strongest opportunity to comment on how land is to be used as it undergoes development. Therefore, this project focused on preventing future deterioration as opposed to fixing

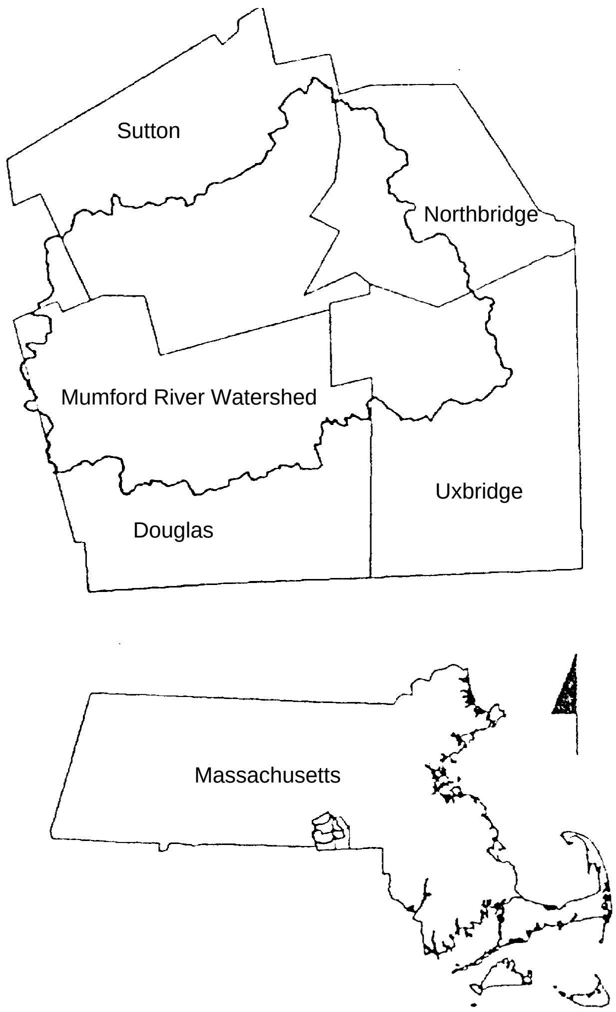


Figure 1. Map of Mumford River watershed study area.

existing problems. This is especially appropriate in a rapidly urbanizing setting.

Both Massachusetts and Rhode Island chose to utilize GIS technology because of its ability to store, analyze, transform, and display geographic, or spatial information. Its database management and analytical capabilities make it a useful tool for pollution load modeling and buildout scenario development, while its mapping capabilities make it an excellent tool for sharing information with local officials. This paper documents a case study on how GIS technology was used to apply a watershed-wide pollution loading model and to develop buildout scenarios for demonstrating to local officials the potential impacts of future development on water quality.

This project used GIS in four different applications:

- *Printing customized, large-scale maps:* This most basic application of a GIS proved the most useful for local officials. It was a revelation for some officials to see how their current zoning related to actual land use. In one town, these maps inspired a change in zoning to protect the area of a future water supply reservoir. These maps helped officials see how their towns fit into the regional picture and how their zoning and land use affected the adjoining towns, and vice-versa.
- *Performing "buildout" analysis:* A "buildout" analysis demonstrates the consequences of existing zoning. It assumes that all land that can be developed will be developed at some future date. In essence, it is a spreadsheet that divides the land available for development in each zone by the required lot size, subtracting

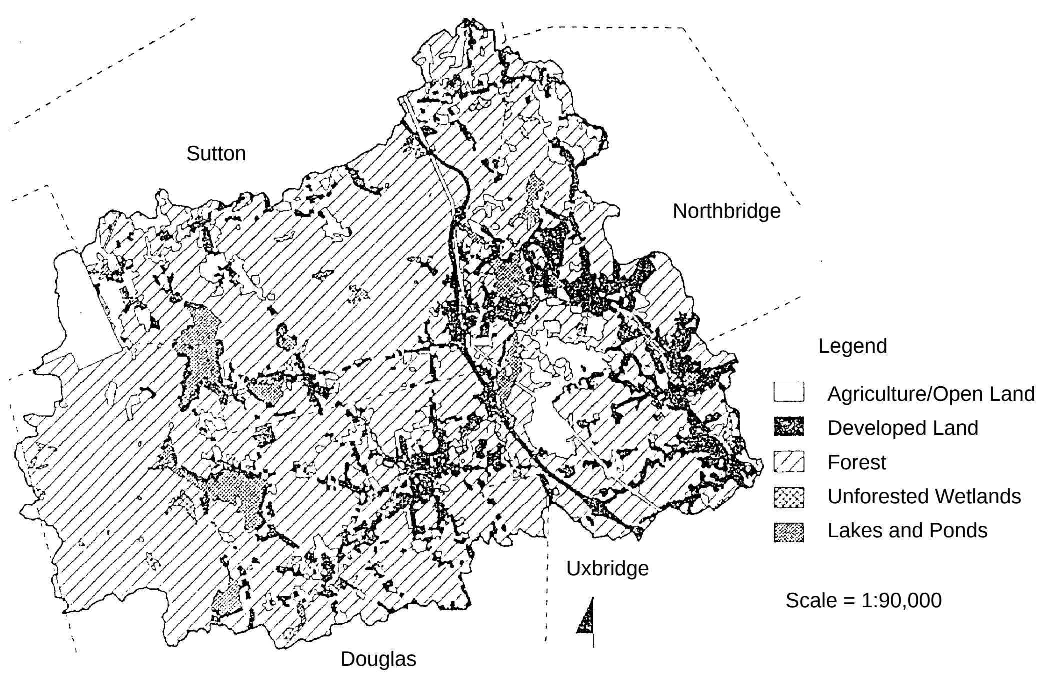


Figure 2. Land use/land cover map of Mumford River watershed.

a certain percentage for the road network and steep slopes. It is best used to evaluate different development scenarios, substituting different zoning requirements.

- *Applying a watershed-wide pollutant loading model:* GIS provided the input needed to apply the "Simple Method" for estimating existing and potential pollutant loads. Future pollution loading was estimated using a buildout with existing zoning and again assuming the implementation of cluster zoning. The Simple Method was compared in one subwatershed with the Galveston Bay Method, which accounts for the hydrologic class of the soils.
- *Promoting planning for a greenway:* Land use maps were overlaid with parcel ownership to show the existing network of preserved open space and to identify those parcels of land having significant wildlife habitat and recreational value. In one town, these maps were used to gain funding for planning a river walk.

Database Development

The most daunting aspect of using a GIS is the prospect of spending a great deal of time and money creating a useful database. Fortunately for Massachusetts, many of the basic coverages needed for regional planning are housed in a state agency, MASS GIS, and are available for a small processing fee. These coverages include most of what appears on the standard United States Geological Survey (USGS) map: roads, streams, town boundaries, as well as watershed boundaries and land use data generated from the interpretation of aerial photographs. The university entered into an agreement whereby we gained access to this data at no charge, in return for sharing the new coverages that the project would generate.

New coverages needed for the study included: zoning, soils, sewer and water lines, and land ownership, or parcels taken from the assessor's maps. The soils maps were obtained from the SCS, digitized by hand, then the scale was converted with a computer program, "rubber-sheeting," to achieve a uniform scale of 1:25,000. All other new coverages were transferred onto a USGS topographical map at a scale of 1:25,000, then digitized directly into the computer. We obtained elevation data, but the triangulation process used to convert elevation data to slopes would require so much time and memory that, for our purpose, deriving a slope map from the four classes identified on the soils map was sufficient.

While GIS computer programs are powerful enough to perform most overlay and analysis functions necessary in nonpoint source pollution load modeling, database development and accuracy issues can limit the effectiveness of such modeling. The choice of which model to use is a function of which data are available for input. Physics-based distributed models are more precise but require detailed input parameters, beyond the scope of this project. The extent of our database limited us to lumped-parameter empirical models. We chose two such models, the Simple Method and the Galveston Bay Method.

GIS Applications

The Simple Method

Schueler (2) developed the Simple Method, one of the simplest lumped-parameter empirical models. The input data necessary to compute pollutant loading with the Simple Method are land use, land area, and mean annual rainfall. Land use determines which event mean concentration (EMC) values and percentage of imperviousness to use in the computation. The amount of rainfall runoff is assumed to be a function of the imperviousness of various land uses. More densely developed areas have more impervious surfaces, such as rooftops and paving, which cause stormwater to run off the land rather than be absorbed into the soil. The Simple Method can generate rough figures for annual pollutant loading within a watershed and can effectively show relative increases in pollutant levels as land is developed.

The formula used in the Simple Method is as follows:

$$L = [(P)\,(Pj)\,(Rv)/12] \quad * \quad (C) \quad * \quad (A) \quad * \quad (2.72)$$

$$(\text{load}) \quad = \quad (\text{runoff}) \quad * \quad (\text{EMC}) \quad * \quad (\text{area})$$

where:

L = pounds of pollutant load per year
P = rainfall depth (inches) over the desired time interval (1 year)
Pj = percentage of storms that are large enough to produce runoff (90 percent)
Rv = fraction of rainfall that is converted into runoff ($Rv = 0.05 + 0.009\ (I)$, where I represents the percentage of site imperviousness)
C = flow-weighted mean concentration (EMC) of the targeted pollutant in runoff (milligrams per liter)
A = area (in acres) of the study region

The Simple Method can be applied using a hand-held calculator or a computer spreadsheet program. For this project, the calculations were performed entirely within the ARC/INFO GIS environment, where the input data were stored. Results were exported to the Excel spreadsheet program for presentation purposes.

The application of the Simple Method consists of three major steps.

Step 1: Aggregate Land Uses and Obtain Area Figures for Land Use Categories Within Each Subbasin

The land use coverage in our database has 21 categories. For the purpose of applying the Simple Method, these were aggregated into the following six major categories, based on development density: undeveloped forest and other open land, large-lot single-family residential, medium-density residential, high-density residential, commercial, and industrial. The aggregated land use categories were matched with study basins from the Nationwide Urban Runoff Program (NURP) for the purpose of assigning EMC values.

Step 2: Enter Percentage Imperviousness and Event Mean Concentrations for Each Land Use Type

The TABLES module of ARC/INFO was used to assign percentage of imperviousness and EMC values to individual land use polygons within the watershed's subbasins. The estimated percentage of imperviousness was obtained from Schueler's guide to using the Simple Method (2). EMC values for three pollutants—phosphorous, nitrogen, and lead—were taken from selected NURP study basins and were assigned to the aggregated land uses within the watershed.

Step 3: Input the Simple Method's Mathematical Loading Formula, Calculate Loading Results for Each Distinct Land Use Area, and Sum Results by Watershed Subbasin

Finally, the pollutant load was calculated for each distinct land use area within the Mumford River watershed by inputting the loading formula through the TABLES module of ARC/INFO. The mean annual rainfall figure was assumed to be that of Worcester, Massachusetts, or 47.6 inches. After calculating loading figures for phosphorous, nitrogen, and lead for each distinct land use area, these numbers were summed for each watershed subbasin, using the ARC/INFO frequency table reporting capability.

The Galveston Bay Method

As an experiment, we applied the Galveston Bay Method to one of the subbasins to compare results with the Simple Method. The slightly more sophisticated Galveston Bay model considers soil drainage characteristics in addition to land use/imperviousness to determine rainfall runoff. This method is similar to the Simple Method, in that amount of rainfall runoff and EMCs for particular land uses are multiplied by land area to determine total pollutant load (3). Runoff in this method, however, is calculated using the USDA SCS's TR 55 runoff curve model. The SCS model calculates runoff as a function of both land use and soil type. Runoff equals total rainfall minus interception by vegetation, depression storage, infiltration before runoff begins, and continued infiltration after runoff begins (4).

The formula used with the Galveston Bay Method has a structure similar to that of the Simple Method and is as follows:

$$L = \frac{P - 0.2\,[(1000/CN) - 10]^2}{P + 0.8\,[(1000/CN) - 10]} * EMC * A$$

$$(\text{load}) = \quad (\text{runoff}) \quad * \quad (\text{EMC}) \quad * \quad (\text{area})$$

where:

L = milligrams of pollutant load per year
P = mean annual rainfall amount
CN = runoff curve number, which is a function of soil type and land use
EMC = event mean concentration
A = area (in acres) of the study region

The application of the Galveston Bay Method consists of four major steps.

Step 1: Aggregate Land Uses and Obtain Area Figures for Land Uses Categories. Aggregate Soils According to Drainage Classes

Land use types were aggregated into the same six major categories as the Simple Method in order to match EMC values and to allow for later comparison of the two pollution loading methods. Soils were aggregated according to drainage classes for use with the USDA SCS TR 55 runoff formula. The SCS identifies four classes of soils according to their drainage capacity:

Class A = excessively to well-drained sands or gravelly sands.
Class B = well to moderately drained, moderately coarse soils.
Class C = moderately to poorly drained fine soils.
Class D = very poorly drained clays or soils with a high water table.

Step 2: Overlay Soils Data With Land Use Data and Clip This New Coverage Within the Subbasin

The ARC/INFO GIS overlay capability was used to overlay land use and soils maps for the Mumford River watershed on top of each other. This created new, distinct areas of different land use and soils combinations. Because we were only applying this model in one subbasin, the subbasin boundary was used in conjunction with the ARC/INFO "clip" command to cut out (like a cookie cutter) that portion of the watershed within the subbasin.

Step 3: Assign Runoff Curve Numbers and EMC Values to Each New Land Use/Soils Polygon Within the Subbasin

EMC values were assigned to each distinct land use/soils area in the same manner as they were assigned to land use areas using the Simple Method.

Runoff curve numbers were assigned to each distinct land use/soils area within the subbasin according to values established by the USDA SCS.

Step 4: Calculate Loading Results for Each Distinct Land Use/Soils Area and Sum Results for the Entire Subbasin

Finally, the pollutant load was calculated for each distinct land use/soils area within the subbasin by inputting the loading formula through the TABLES module of ARC/INFO. After calculating loading figures for phosphorous, nitrogen, and lead for each distinct land use/soils area, these figures were then summed for the subbasin using the ARC/INFO frequency table reporting capability. Results of this modeling were converted from milligrams per acre per year to pounds per acre per year to facilitate later comparisons.

Buildout Scenarios

For planning purposes, GIS is most useful in its ability to quickly generate alternative scenarios. When these development scenarios are coupled with a pollutant load model as described above, alternative scenarios can be evaluated according to their impact on water quality. This project generated two different scenarios for each of the four towns in the watershed: a maximum buildout with existing zoning and a maximum buildout with clustered development.

Maximum Buildout

A maximum buildout scenario was used to show the worst case for development according to current zoning regulations (see Figure 3). The result of this buildout is expressed both in the number of new residential units to be built and in the area of land to be converted from undeveloped to residential and other urban uses.

Step 1: Eliminate From Consideration All Land That Is Already Developed

Step 2: Eliminate From Consideration All Land That Is Under Water

Step 3: Eliminate From Consideration All Land That Is Protected From Development

These protected lands included cemeteries, parks, and all land permanently restricted from development.

Step 4: Reduce the Remaining Amount of Land by 20 Percent To Account for New Roadways and Extremely Steep Slopes

The remaining land was considered to have "developable" status. Wetlands were included in this category because while a house probably would not be built on a wetland, wetlands can and often do constitute portions of the required lot size of large residential lots.

Step 5: Overlay the Land Use Coverage With Zoning and Minimum Lot Size Information

This created new land use areas as a function of zoning. All forests and fields were converted to a developed status.

Step 6: Divide Net Developable Land Area Within Each Zone by Minimum Lot Size Allowed To Obtain the Number of New Units

Results from the buildout are expressed in the number of new units. Results can also be shown spatially by shading in areas on the map according to future density

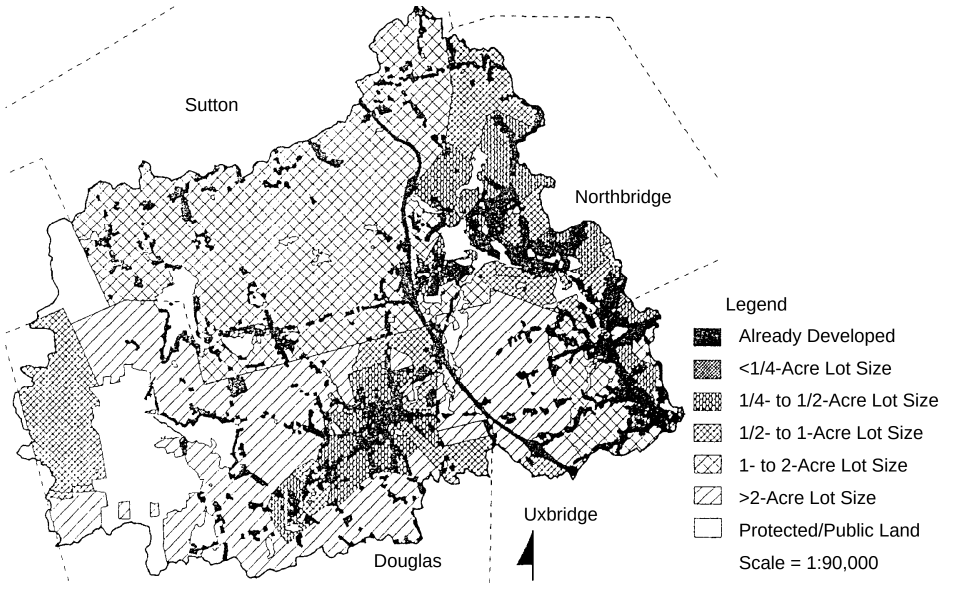


Figure 3. Maximum buildout scenario within the Mumford River watershed.

of development (darker shades for higher density, lighter shades for lower density).

Clustered Buildout

Another alternative development scenario was generated assuming the implementation of clustered development. All areas zoned for lots larger than 1 acre were changed to cluster zones, where three-fourths of the land area remains undeveloped, and the remaining one-fourth of the land area is developed at a density of 1/2-acre lot size. With clustering, an area zoned for 2-acre house lots still supports the same number of new units, but three-quarters of the land area remains open space for passive recreation, protected wildlife habitat, and as a buffer zone to filter runoff.

Step 1: Select All Land Available for Development Zoned for 1-Acre Lots or Larger

Step 2: Multiply Selected Land by 0.75 and Add to the Category of Protected Land

Step 3: Multiply Selected Land by 0.25 and Change the Minimum Lot Size to One-Half an Acre

Step 4: Divide Step 3 by 20 Percent To Allow for New Roads and Steep Slopes

Step 5: Divide Step 4 by 21,780 (One-Half an Acre) To Determine Number of New Housing Units

Results

Lumped-parameter empirical models were chosen for this project and were applied to watershed subbasins ranging in size from 1 to 20 square miles and having an average of 4 square miles. The application of the Simple Method to existing land use conditions allowed for a comparison of the Mumford River watershed's subbasins for the purpose of identifying the subbasins that contribute the highest levels of pollutants per acre per year. The development of a maximum buildout scenario identified those areas within the watershed that will sustain the greatest amount of new growth. The application of the Simple Method to this maximum buildout scenario revealed that pollutant levels in surface water runoff would increase substantially for all subbasins in the watershed. This finding supports the theory of a positive relationship between development and increased pollutant levels from surface water runoff.

The development of a customized buildout scenario for future development identified those areas that are currently zoned for large-lot residential "sprawl" and that can support higher development density under cluster zoning, while protecting a significant amount of open space that can support a variety of beneficial uses. The application of the Simple Method to the customized buildout scenario revealed that the use of cluster development can reduce future levels of water pollution, especially from nutrients (see Figures 4 and 5).

Results determined by applying the Galveston Bay Method to one subbasin were compared with those obtained using the Simple Method. The predicted pollutant loading from current conditions differed significantly

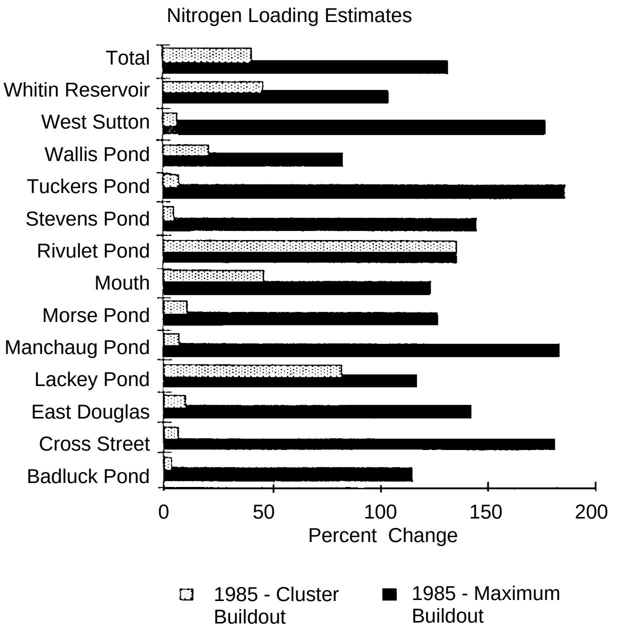


Figure 4. Chart showing difference in simple method results for nitrogen loading between maximum and customized buildout scenarios.

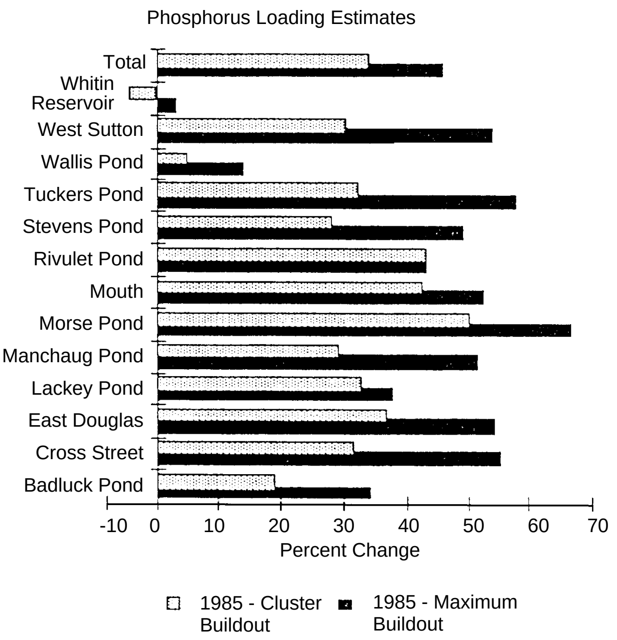


Figure 5. Chart showing difference in simple method results for phosphorus loading between maximum and customized buildout scenarios.

between the two methods; the Simple Method consistently predicted five times the amounts generated by the Galveston Bay Method.

When the two methods were applied to both the maximum and customized buildout scenarios, however, the percentage growth of predicted pollutant loadings was remarkably similar for both methods; the Simple Method consistently predicted loadings 10 to 15 percent greater than the Galveston Bay Method. This indicates that while the Galveston Bay Method may provide more accurate results in predicting actual pollutant loading, the Simple Method is adequate enough for evaluating and comparing different development scenarios (see Figures 6 and 7).

Discussion

As states begin to implement a TMDL approach to regulating water quality, they face the quandary of how to determine the extent of nonpoint source pollution in the rivers. The crudest method is to subtract from the total load those quantities generated by point sources and call all the rest nonpoint source. While this is appropriate in some settings, it is unacceptable in a watershed with a long history of pollution because a significant source of pollution is the resuspension of historical sediments stirred up by storms. The situation demands the development of a model to predict the loading from nonpoint sources. Only a computer can handle the multiple factors that interact to generate nonpoint sources of pollution.

As greater emphasis is placed on watershed planning, the abilities of a GIS to input, store, manipulate, analyze, and display geographic information become indispensable. As the scientific community improves its knowledge base for determining the critical factors influencing nonpoint source pollution, GIS technology is improving in its ability to store and handle large amounts of data.

While a detailed, physics-based distributed model would be more accurate than the lumped-parameter models used for this project, they are difficult to apply at the watershed scale. The real limiting factor is the provision of all the data coverages needed to apply complex models. Lumped-parameter models, such as the Simple Method and the Galveston Bay Method, are ineffective for accurately predicting pollutant loads, but they are suitable for comparing and evaluating alternative development scenarios.

Time, and the development community, will not wait until all the answers are known. Local officials continue to approve development with no thought to the impacts on water quality. These officials need to be informed about the implications of haphazard growth. A GIS, with its ability to generate customized maps and quickly evaluate alternative development scenarios, is a powerful tool to help local officials visualize how the decisions they make on paper today will have an impact on the land tomorrow.

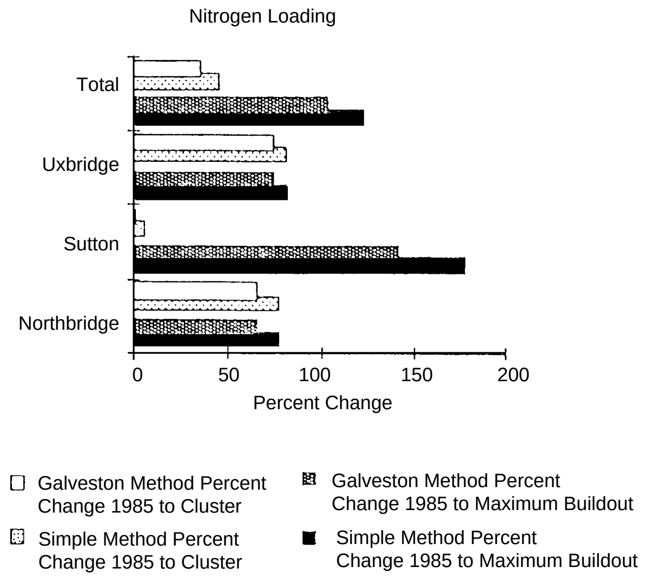


Figure 6. Chart showing difference between Simple Method results and Galveston Bay Method results for nitrogen loading.

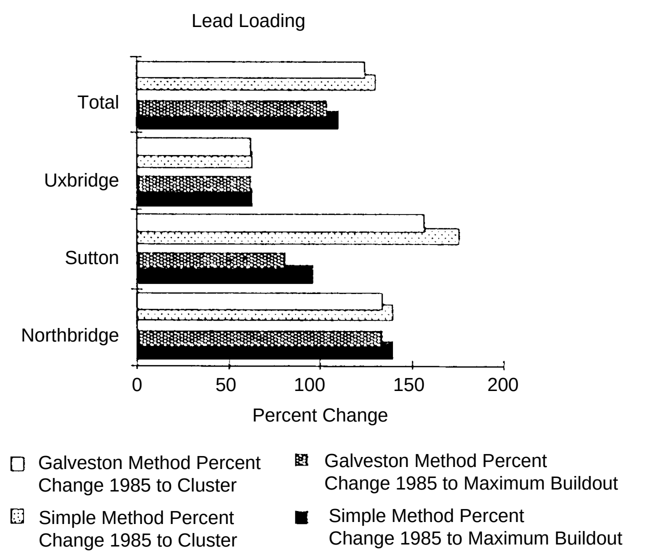


Figure 7. Chart showing difference between Simple Method results and Galveston Bay Method results for lead loading.

References

1. Metcalf & Eddy, Inc. 1991. Assessment of pollutions in Narragansett Bay. Draft report to the Narragansett Bay Project.

2. Schueler, T. 1987. Controlling urban runoff: A practical manual for planning and designing urban BMPs. Washington, DC: Metropolitan Washington Council of Governments.

3. Newell, C., et al. 1992. Characterization of nonpoint sources and loadings to Galveston Bay, Vol. I, Technical Report: Galveston Bay National Estuary Program (March).

4. U.S. Department of Agriculture (USDA) Soil Conservation Service. 1975. Urban hydrology for small watersheds. Technical Release 55. Springfield, VA: U.S. Department of Agriculture.

Additional Reading

Arnold, C.L., et al. 1993. The use of Geographic Information System images as a tool to educate local officials about the land use/ water quality connection. Proceedings of Watershed Conference, Alexandria, VA.

Chesebrough, E. 1993. Massachusetts nonpoint source management plan. Massachusetts Department of Environmental Protection, Office of Watershed Management (October).

Joubert, L., et al. 1993. Municipal training for water quality protection. Contribution #2845. College of Resource Development, University of Rhode Island/Rhode Island Agricultural Experiment Station.

Massachusetts Geographic Information System. 1993. MassGIS Datalayer descriptions and guide to user services. Boston, MA: Executive Office of Environmental Affairs.

McCann, A., et al. 1994. Training municipal decision-makers in the use of Geographic Information Systems for water resource protection. Contribution #2927. College of Resource Development, University of Rhode Island/Rhode Island Agricultural Experiment Station.

Narragansett Bay Project. 1992. Comprehensive conservation and management plan: A summary (January).

U.S. EPA. 1983. Results of the Nationwide Urban Runoff Program (NURP), Vols. I and II. Final report prepared by Water Planning Division. NTIS PB84185537. Springfield, VA: National Technical Information Service.

U.S. EPA. 1992. Compendium of watershed-scale models for TMDL development. EPA/841/R-92/002. Washington, DC (June).

A GIS for the Ohio River Basin

Walter M. Grayman
W.M. Grayman Consulting Engineers, Cincinnati, Ohio

Sudhir R. Kshirsagar
Global Quality Corporation, Cincinnati, Ohio

Richard M. Males
RMM Technical Services, Inc., Cincinnati, Ohio

James A. Goodrich
Risk Reduction Engineering Laboratory, Office of Research and Development,
U.S. Environmental Protection Agency, Cincinnati, Ohio

Jason P. Heath
Ohio River Valley Water Sanitation Commission, Cincinnati, Ohio

Abstract

Much of the information used in the management of water quality in a river basin has a geographic or spatial component associated with it. As a result, spatially based computer models and database systems can be part of an effective water quality management and evaluation process. The Ohio River Valley Water Sanitation Commission (ORSANCO) is an interstate water pollution control agency serving the Ohio River and its eight member states. The U.S. Environmental Protection Agency (EPA) entered into a cooperative agreement with ORSANCO to develop and apply spatially based computer models and database systems in the Ohio River basin.

Three computer-based technologies have been applied and integrated: geographic information systems (GIS), water quality/hydraulic modeling, and database management.

GIS serves as a mechanism for storing, using, and displaying spatial data. The ARC/INFO GIS, EPA's agencywide standard, was used in the study, which assembled databases of land and stream information for the Ohio River basin. GIS represented streams in hydrologic catalog units along the Ohio River mainstem using EPA's new, detailed RF3-level Reach File System. The full Ohio River basin was represented using the less detailed RF1-level reach file. Modeling provides a way to examine the impacts of human-induced and natural events within the basin and to explore alternative strategies for mitigating these events.

Hydraulic information from the U.S. Army Corps of Engineers' FLOWSED model enabled EPA's WASP4 water quality model to be embedded in a menu-driven spill management system to facilitate modeling of the Ohio River mainstem under emergency spill conditions. A steady-state water quality modeling component was also developed under the ARC/INFO GIS to trace the movement and degradation of pollutants through any reaches in the RF1 representation of the full Ohio River basin.

Database management technology relates to the storage, analysis, and display of data. A detailed database of information on dischargers to the Ohio River mainstem was assembled under the PARADOX database management system using EPA's permit compliance system as the primary data source. Though these three technologies have been widely used in the field of water quality management, integration of these tools into a holistic mechanism provided the primary challenge of this study.

EPA's Risk Reduction Engineering Laboratory in Cincinnati, Ohio, developed this project summary to announce key findings of the research project, which is fully documented in a separate report of the same title.

Introduction

During the past 25 years, computers have been actively used in water quality management, demonstrating their potential to assist in a wide range of analysis and display tasks. Technologies such as geographic information systems (GIS), database management systems (DBMS), and mathematical modeling have been applied in the water quality management field and have proven to be effective tools. For computers to achieve their full potential, however, they must become integrated into the normal programmatic efforts of agencies and organizations in the planning, regulation, and operational areas of water quality management.

Recognizing this need for routine use of computer-based tools, the Ohio River Valley Water Sanitation Commission (ORSANCO) and the Risk Reduction Engineering Laboratory (RREL) of the U.S. Environmental Protection Agency (EPA) commenced a study in 1990. The goals of the study included the adaptation, development, and application of modeling and spatial database management (DBM) tools that could assist ORSANCO in its prescribed water quality management objectives. These goals were consistent with EPA's ongoing programs involving the use of GIS and modeling technology. The study's goals also coincided with EPA's Drinking Water Research Division's work over the past decade, which applied similar technology to study the vulnerability of water supplies on the Ohio and Mississippi Rivers to upstream discharges.

Methodology Overview

To address the goals of this project, three basic technologies have been applied and integrated: GIS, water quality/hydraulic modeling, and DBM. GIS serves as a mechanism for storing, using, and displaying spatial data. Modeling provides a way to examine the impacts of human-induced and natural events within the basin and to explore alternative strategies for mitigating these events. DBM technology relates to the storage, analysis, and display of data. Though these three technologies have been widely used in the field of water quality management, integration of these tools into a holistic mechanism provided the primary challenge of this study.

GIS Technology

The guiding principle in developing the GIS capability was to maximize the use of existing GIS technology and spatial databases. The study used ARC/INFO GIS, EPA's agencywide standard. Remote access of ARC/INFO on a VAX minicomputer facilitated the initial work. Subsequently, both PC ARC/INFO and a workstation-based system were obtained.

EPA has developed an extensive spatial database related to water quality and demographic parameters. This served as the primary source of spatial data for the study. Following is a summary of spatial data used in this study:

- State and county boundaries.
- City locations and characteristics.
- Water supply locations and characteristics.
- Locations and characteristics of dischargers to water bodies.
- Toxic loadings to air, water, and land.
- Dam locations and characteristics.
- Stream reaches and characteristics.

The primary organizing concept for the water-related information was EPA's Reach File System (1). This system provides a common mechanism within EPA and other agencies for identifying surface water segments, relating water resources data, and traversing the nation's surface water in hydrologic order within a computer environment. A hierarchical hydrologic code uniquely identifies each reach. Information available on each reach includes topological identification of adjacent reaches, characteristic information such as length and stream name, and stream flow and velocity estimates. The original reach file (designated as RF1) was developed in the early 1980s and included approximately 70,000 reaches nationwide. The most recent version (RF3) includes over 3,000,000 reaches nationwide.

As part of this project, an RF1-level database was established for the entire Ohio River basin. The RF3 reach file was implemented for the Ohio River mainstem and lower portions of tributaries. River miles along the Ohio River were digitized and established as an ARC/INFO coverage to provide a linkage between the reach file and river mile indexing used by ORSANCO and other agencies along the river. Figure 1 shows the RF1 reach file representation of the Ohio River basin along with state boundaries.

The study incorporated several EPA sources of information on dischargers to water bodies. The industrial facility discharger (IFD) file contains locational and characteristic data for National Pollutant Discharge Elimination System (NPDES) permitted discharges. Detailed permit limits and monitoring information was accessed from the permit compliance system (PCS). The toxic release inventory (TRI) system includes annual loading of selected chemicals to water, land, air, and sewer for selected industries based on quantity discharged. All water data are referenced to the NPDES permit number, which is spatially located by reach and river mile, and by latitude and longitude.

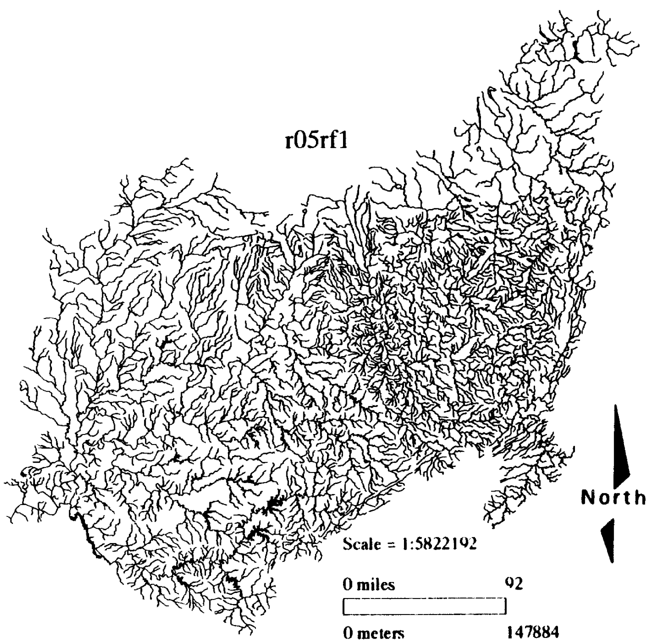


Figure 1. RF1 reaches in the Ohio River basin.

Spill Modeling

An important role that ORSANCO fills on the Ohio River relates to the monitoring and prediction of the fate of pollutant spills. Typically, ORSANCO serves as the overall communications link between states during such emergency conditions. ORSANCO coordinates and participates in monitoring and serves as the information center in gathering data and issuing predictions about the movement of spills in the river. In the past, a series of time-of-travel nomographs, based on National Weather Service flow forecasts, Corps of Engineers flow-velocity relationships, and previous experience, were used to predict the movement of spills. This project combined a hydraulic model with a water quality model to serve as a more robust method for making such predictions.

The U.S. Army Corps of Engineers' FLOWSED model was selected as the means of predicting daily flow quantities and water levels along the mainstem and portions of major tributaries near their confluence with the Ohio River (2). The Ohio River Division of the Corps of Engineers applies FLOWSED daily as part of its reservoir operations program. The Corps can generate 5-day forecasts of stage and flow for 400 mainstem and tributary segments, and ORSANCO can access the results via phone lines.

EPA's WASP4 water quality model was selected for use in the project (3). WASP4 is a dynamic compartment model that can be used to analyze a variety of water quality problems in a diverse set of water bodies. Because the primary use of the model in this project is quick response under emergency situations, only the toxic chemical portion of the model with first order decay is being used. The FLOWSED and WASP4 models have been combined into a user-friendly spatial decision support system framework described later in this project summary.

Discharger Database Management System

EPA's PCS and historical records maintained by ORSANCO furnish a rich source of data on discharge information for the Ohio River. To organize these data and make them available for analysis, a database was developed using the PARADOX DBM system.

The database was established using a relational structure with a series of related tables (two-dimensional flat files). Individual tables contain information on facilities, outfalls, permit limits, monitoring data, and codes used in the other tables. The NPDES permit number is used as the primary key in each data table. A mechanism for downloading and reformatting data from the national PCS database has been developed along with custom forms for viewing and editing data, and custom reports for preparing hard copy summaries. Latitude and longitude values for each facility can provide the locational mechanism for use of this data in conjunction with GIS.

Integration of GIS/Modeling/Database Technologies

A major objective of this study was the integration of GIS, modeling, and DBMS technologies into a holistic tool for use by ORSANCO. Several integration mechanisms were implemented as summarized below.

Steady-State Spill Tracing

The NETWORK component of the ARC/INFO GIS provides a steady-state, transportation-oriented routing capability. This capability was used in an arc macro language (AML) program to construct a routing procedure for determining downstream concentrations and travel times. The pollutant may be treated as a conservative element or represented by a first order exponential decay function. This capability has been implemented for use with the RF1 reach file representation of the full Ohio River basin. The user may select from six flow regimens: average flow, low flow, and four multiples of average flow ranging from one-tenth to 10 times average flow. This system gives ORSANCO the ability to estimate the arrival time of a spill from any RF1 tributary to the Ohio River mainstem.

Spill Management System

A PC-based spatial decision support system (SDSS) was built as a spill management system to be a quick response tool for analyzing and displaying the results of pollutant spills into the Ohio River. The schematic in Figure 2 illustrates the components in this computerized spill management system. The system is implemented in the C language using a commercial menuing system and a series of graphic display routines developed at EPA. Custom, written routines have been used to read the output from the U.S. Army Corps of Engineers' FLOWSED model, to generate input files for EPA's WASP4 model, to create output reports and output plots, and to provide an animated representation of the concentration profiles moving down the river. Figure 3 presents an example of a graphic output the system generated. Additionally, the system generates a file in

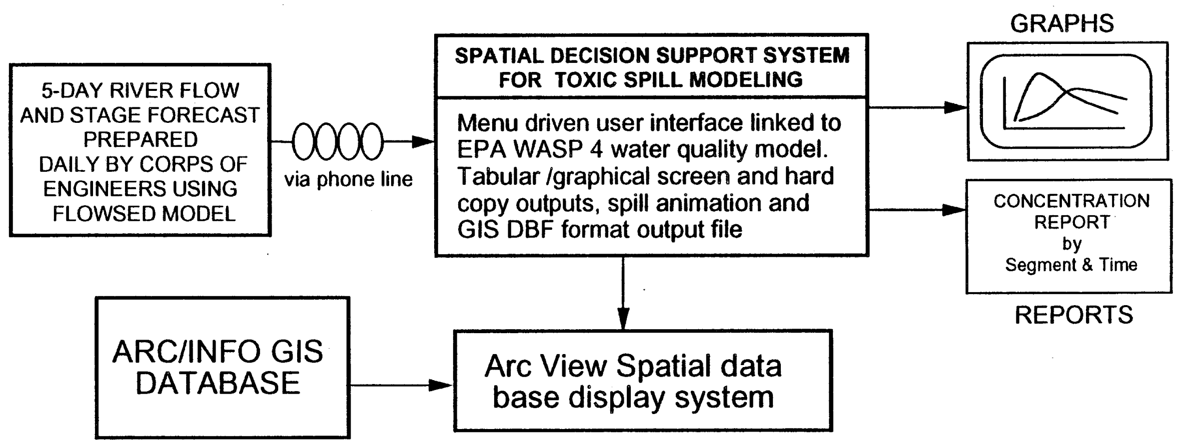


Figure 2. Schematic representation of spill modeling system process.

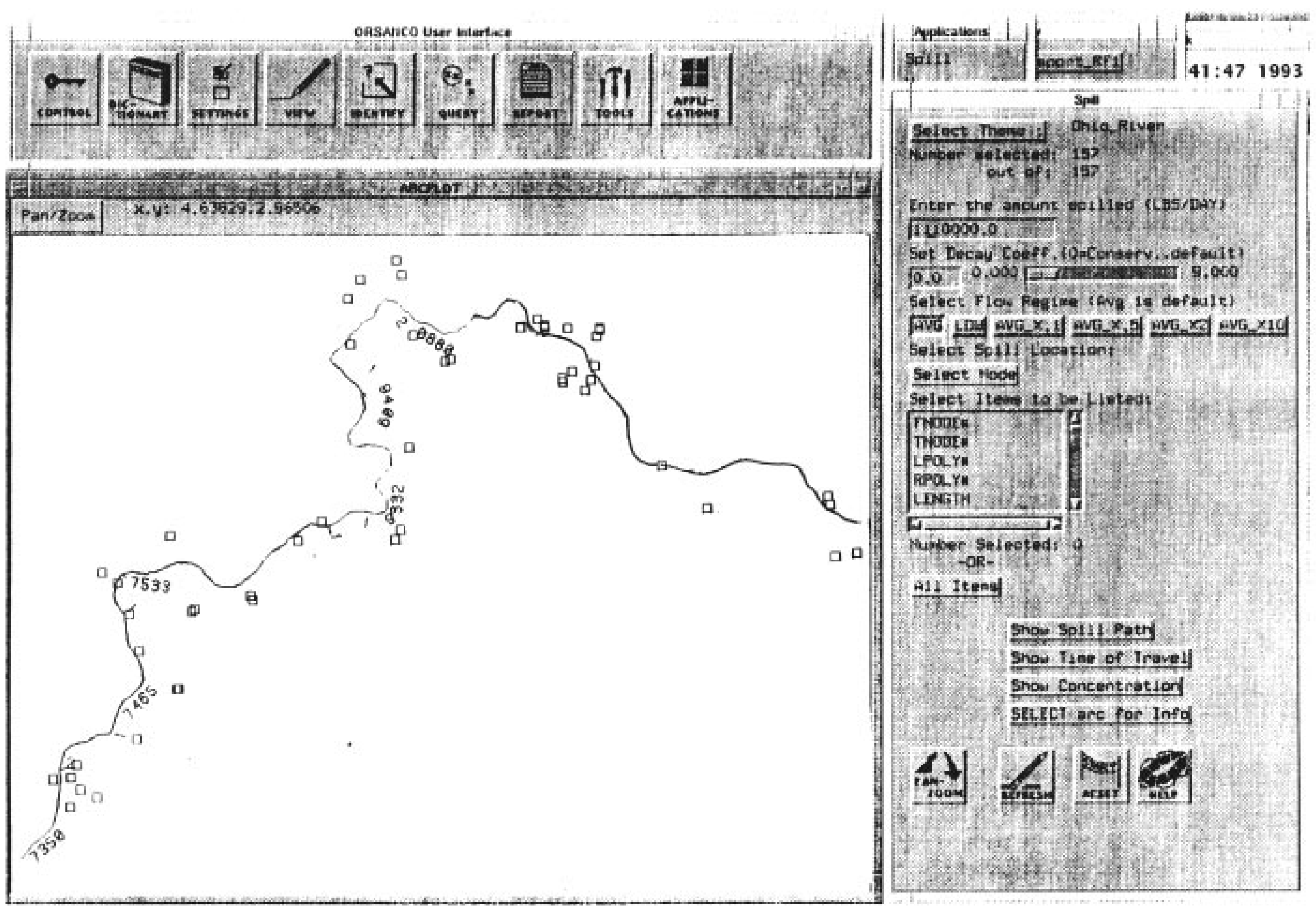


Figure 3. Graphic output from the basinwide network spill model.

DBF format that may be read by ARC/VIEW (the companion software to ARC/INFO for user-friendly viewing of spatial data).

Hardware Platform

Within the study, the initial hardware platform was a combination of local PCs (in Cincinnati) and a remote access terminal to a VAX computer located at EPA's National Computer Center in Research Triangle Park, North Carolina. The final platform, and the one on which the completed system was installed, comprised a UNIX-based Data General workstation and a PC workstation. The full hardware configuration is shown schematically in Figure 4.

Conclusions

The application of computer-based display, analysis, and modeling tools in conjunction with GIS technology proved to be an effective strategy for water quality management. This study used an existing GIS package and DBMS in conjunction with existing water quality and hydraulic models. The study focused primarily on assembling available spatial and relational databases and integrating the systems to provide a usable, effective tool.

References

1. Horn, R.C., and W.M. Grayman. 1993. Water-quality modeling with EPA reach file system. J. Water Res. Planning and Mgmt. 119(2):262-274.
2. Johnson, B.H. 1982. Development of a numerical modeling capability for the computation of unsteady flow on the Ohio River and its major tributaries. Vicksburg, MS: U.S. Army Engineer WES.
3. Ambrose, R.B., T.A. Wool, J.P. Connolly, and R.W. Schanz. 1988. WASP4, a hydrodynamic and water quality model: Model theory, users' manual, and programmers' guide. Athens, GA: EPA Environmental Research Laboratory. NTIS PB88185095.

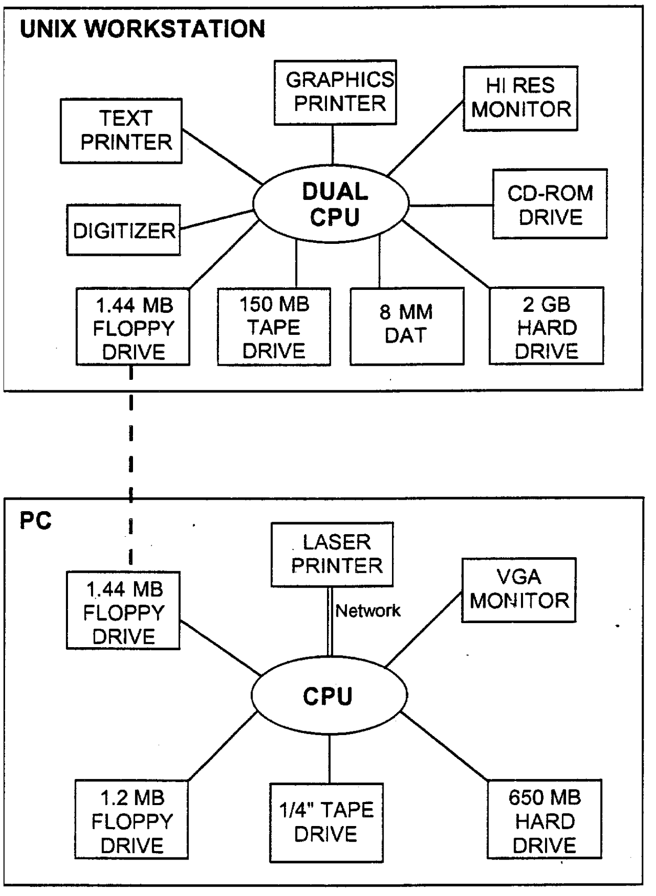


Figure 4. Hardware configuration.

Nonpoint Source Pesticide Pollution of the Pequa Creek Watershed, Lancaster County, Pennsylvania: An Approach Linking Probabilistic Transport Modeling and GIS

Robert T. Paulsen
The Paulsen Group, Bowie, Maryland

Allan Moose
Southampton College, Southampton, New York

Abstract

The U.S. Environmental Protection Agency (EPA) has mandated that each state prepare a state management plan (SMP) to manage pesticide residues in the state's environment. One aspect of an SMP involves identifying specific soils and sites that may be vulnerable to the transport of pesticides into water resources. A recently developed system identifies vulnerable areas by coupling probabilistic modeling that uses the Pesticide Root Zone Model (PRZM) with a desktop geographic information system (GIS-MAPINFO). A limited test of this system succeeded in identifying and mapping individual soil series in a watershed that were shown to have transported atrazine to surface and ground water.

During this project, various digital data sources were evaluated for availability and ease of use, including:

- STATSGO.
- U.S. Geological Survey (USGS) digital line graphs (DLGs).
- National Oceanic and Atmospheric Administration (NOAA) climate data.

This study documents hands-on hints and tricks for importing and using these data.

From 1977 to 1979, the USGS measured the movement of atrazine off fields of application into water resources in the Pequa Creek basin in Lancaster County, Pennsylvania (1). Atrazine in surface water appeared at levels exceeding 20 parts per billion in storm flow and above the 3 parts per billion maximum contaminant level (MCL) during base flow from Big Beaver Creek, a tributary to Pequa Creek. Each soil series in the subbasin was digitized into a GIS. PRZM allowed simulation of runoff, erosion, and leaching of atrazine (applied at 2.24 kilograms per hectare in conventionally tilled corn) for each soil. This process included simulating each soil under different slopes for an 11-year period from 1970 to 1980. Interpreting the results for each soil series determined the probability distribution of atrazine in kilograms per hectare for each mode of transport. GIS used these data to thematically map each soil series for atrazine loss.

The results of this demonstration project suggest that the Manor silt loam, with slopes varying from 6 percent to 20 percent, had a high potential to transport atrazine residues to surface water. This type of analysis could suggest that this soil series be:

- Farmed using conservation tillage.
- Managed to install grass waterways or buffer strips to stop runoff.
- Set aside from production to protect water resources.

Digital databases were available for the study area, but many technical problems were encountered in using the data. Researchers embarking on these types of modeling and GIS projects should prepare themselves for significant expenditures of time and finances.

Introduction

A significant volume of published literature documents pesticide residues in ground water, and the volume of investigations of residues in surface water is expanding. The growing acceptance of immunoassay techniques for the determination of pesticide residues in water has given the field of pesticide monitoring an accurate and economical analytic methodology. This will result in an increase in monitoring capability at the federal, state, local, and university levels. These increases in monitoring capability have documented and will continue to document the occurrence of pesticides in water resources as the result of past transport through the soil profile. The U.S. Environmental Protection Agency (EPA) has mandated that each state prepare a state management plan (SMP) to manage pesticide residues in the state's environment. Lacking, however, is a reliable pesticide screening technique to indicate which soils, on a countywide scale, may be sensitive to the transport of a specific pesticide to deep within the soil profile or to the surface water resources. These assessments would greatly supplement the usability and validity of SMPs.

Electronic databases such as State Soils Geographic (STATSGO), Data Base Analyzer and Parameter Estimator (DBAPE), or the SOIL5 subsets found in Nitrate Leaching and Economic Analysis Package (NLEAP) provide easy access to detailed soil data and model input estimator subroutines, thereby simplifying data entry to numerical models. Two groupings define soils: soil series and soil associations. Soil series are the individual soil taxa found in a field. Soil associations represent groups of soil series, usually three or four soil series occurring together in an area, and are mapped as a single unit on a county scale. Mapping of most soil associations across the United States is complete, with open access to the county scale maps. A digital soils mapping data set called SSURGO contains many of the soil series maps for the United States. Climatologic databases also provide easy access to long-term data from the National Oceanic and Atmospheric Administration (NOAA) weather stations, allowing a user the opportunity to input realistic climate data to pesticide transport models.

Many numerical pesticide transport models, such as the Pesticide Root Zone Model (PRZM), Ground-Water Leaching Effects of Agricultural Management Systems (GLEAMS), and Leaching Estimation and Chemistry Model (LEACHM), can produce transport estimates for specific pesticides in specific soils. Each model has its own strengths and weaknesses, and detailing these characteristics is beyond the scope of this paper. Several authors, however, have described comparisons between models (e.g., Smith et al. [2], Mueller et al. [3], and Pennell et al. [4]). These numerical models all generally require extensive site-specific soil, agronomic, and climatologic databases. The results from these models are extremely detailed. Their pesticide transport estimates, however, are only valid for those locations for which site-specific data are sufficient to allow calibration of the model. Applying such site-calibrated model results to larger scales (county scales) is inappropriate.

In one procedure, a user could identify soils and use transport models that may have a limited ability to retard pesticides from reaching water resources. This type of modeling has recently been called probabilistic modeling (5). The concept behind this procedure is to use an existing transport model, such as PRZM, and vary certain input parameters (e.g., slope, organic content, pesticide K_{OC}) to produce a probability of a given output being equaled or exceeded. For example, a PRZM model could be created for a soil series with an average organic content of 1 percent and a slope of 8 percent in the eastern corn belt. The model would use the 30 years of historical climate data for a nearby station. The model would vary the organic content and surface slope within given ranges for the soil series and run 1,000 simulations. The analysis could then entail plotting the results (i.e., monthly runoff loads, erosion loads, and leaching through the root zone) in a frequency diagram and generating probability curves. This analysis would allow the user to estimate the anticipated pesticide losses, runoff, erosion, and leaching for any given soil in the county. The soils with greater probabilities for pesticide loss could be identified and mapped using GIS.

Recent advances in computing speed and efficiency have reduced the amount of time and expense needed to run numerical pesticide transport models. This makes it possible, in a relatively short amount of time, to quantitatively model not just one soil series, but the hundreds of major soil series that occur in an entire state (e.g., 357 major soil series combined into 464 different soil associations in Wisconsin). This type of model can be very useful to the development of SMPs as well as to a variety of users, including pesticide registrants, bulk pesticide handlers, custom appliers, county agricultural extension agents, and individual growers.

Objective

The objective of this study was to use probabilistic modeling analyses and a geographic information system (GIS) to determine which soil series in a watershed may contribute to nonpoint source pollution through runoff of agricultural chemicals. Specifically, the study aimed to locate a watershed with historical atrazine runoff, map the soils, and perform transport modeling using historical precipitation. The results of this procedure would determine which soil series had a high potential to contribute to the nonpoint source pollution of the watershed. Once the transport modeling was completed, a GIS would help map the distribution of the sensitive soil series. The mapping would act as a base for implementing best management practices (BMPs) to reduce nonpoint source pollution.

Background

Ward (1) described the water quality in the Pequa Creek basin in Lancaster County, Pennsylvania, for the years 1977 through 1979. Flow from Pequa Creek (154-square-

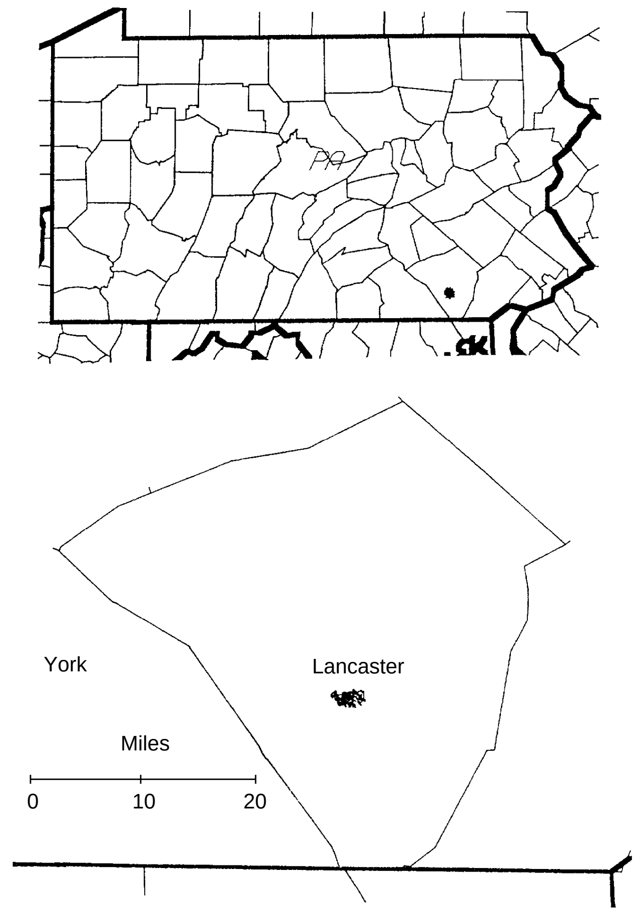


Figure 1. Location of Pequa Creek basin, Lancaster County, Pennsylvania.

mile drainage area) eventually discharges into the Chesapeake Bay (see Figure 1). The data collection efforts (6) documented the occurrence of atrazine (2-chloro-4-ethylamino-6-isopropylamino-s-triazine), a commonly used herbicide for weed control in corn-growing regions, and other agrichemicals in both base-flow and storm-flow conditions of the Pequa Creek. A subbasin of Pequa Creek, Big Beaver Creek, had the greatest reported atrazine concentrations during the sampling period. The maximum reported atrazine concentrations at the Big Beaver Creek sampling station, near Refton, Pennsylvania, were 0.30 parts per billion during base-flow conditions and 24.0 parts per billion during storm-flow conditions. The Big Beaver Creek basin is 20.4 square miles in area, and agriculture constituted about 66 percent of the land use in 1979. Corn was grown on 26.6 percent of the agricultural lands in this subbasin. The average rainfall for this basin is about 37 inches annually (1).

As noted, agriculture represented the major land use in the area. The primary agricultural soils in Lancaster County are silt loams (Typic Hapludults and Hapludalfs) in texture with slopes that range from 0 percent to 8 percent (7). Upon inspection of the air-photo soil series maps found in the county soil survey, however, agricultural crops grew on lands with slopes of up to and exceeding 15 percent, with soils such as the Manor silt loam, Pequa silt loam, and Chester silt loam (7).

Natural soil organic contents in the agricultural soils range from 0.1 percent to 2.0 percent. Water contents of the agricultural soils range from 10 percent to nearly 30 percent. The soil Erosion Factor (K) for the surface layer ranges from 0.17 for the relatively stable Ungers Series to 0.43 for the Pequa Series. The greater the value, the greater the susceptibility to sheet and rill erosion. The soil Erosion Factor (T in tons/acre/year) for the entire soil profile ranges from 2 for the relatively stable Clarksburg Series to 5 for the Elk Series.

Methods

Determining which soil series in the Big Beaver watershed may contribute to nonpoint source pollution through runoff of agricultural chemicals entailed performing a combination of probabilistic modeling analyses and a GIS data manipulation.

Physiographic and Soil Series Boundaries

The orientation of Pequa Creek, Big Beaver Creek, and other surface water bodies was digitized directly from the 1:50,000-scale county topographic map for Lancaster County (8). This map also provided the basis for digitizing the Pequa Creek drainage divide, location of urban areas, and roadways. The MAPINFO GIS allows for the creation of boundary files by tracing the boundary off the topographic map with a digitizing tablet configured to the latitude and longitude coordinates of three points on the map. The latitude and longitude are displayed while the boundary is being traced, allowing the user to verify the accuracy of the boundary against known coordinates on the map. GIS contains a self-checking boundary closure program to ensure that the polygons are closed and that the boundary contains no extraneous line segments. These boundary data are already available from the USGS in a digitized format, digital line graphs (DLGs). Because digitizing is an easy task, however, and to minimize costs, the project used manual digitizing rather than purchase the data.

The roadways and urban areas were digitized to allow use of standard control points, such as road intersections and benchmarks, to configure the U.S. Department of Agriculture (USDA) Soil Conservation Service air-photo-based, 1:15,840-scale soil series maps for the Big Beaver Creek watershed. Known land grid coordinates were noted on the air-photo maps (7). We concluded, however, that using known reference points, such as roadways and towns, allowed for a better configuration of the digitizing tablet to the air photos and eliminated concerns over scale distortions sometimes common in air photos.

After configuring the air photos to the digitizing tablet, a 2-square-mile area around the surface water sampling points was digitized. The next step entailed digitizing all the mapped soil series units within this area. The locations of crop areas, as plowed fields, were noted. In noting forested areas, it became apparent that only minor acreages were not in agricultural production. Those mapped units were generally the Manor and Pequa Series soils with slopes exceeding 25 percent.

Pesticide Transport Modeling

The PRZM pesticide transport model helped to quantify the ability of several soil series to retard the transport of atrazine through the root zone as leachate, dissolved in surface runoff and adsorbed on sediment that moved during erosion. The PRZM model performed in an uncalibrated or screening model mode. The input values for soil properties came from both the EPA DBAPE database and the Lancaster County Soil Survey (7). The modeled soil profile was 150 centimeters thick and divided into 5-centimeter compartments. The soil half-life of atrazine was set at 57 days in accordance with values that the PRZM manual listed (9). The primary soil property that varied in this demonstration project was surface slope. All other parameters, such as soil organic content, moisture content, and bulk density, appeared as midpoint values for the ranges listed in DBAPE.

The agronomic scenario that the model simulated was for corn grown continuously for 10 years using conventional tillage practices and planted on May 7 of each year. Atrazine was surface applied at a rate of 2 pounds per acre (2.24 kilograms per hectare) on May 1 of each year. For climatic input, the model used the historical precipitation regimen from 1970 through 1980, as measured at the Harrisburg, Pennsylvania, station.

PRZM simulations were made for each of the following soil series:

- Chester
- Conestoga
- Elk
- Glenelg
- Glenville
- Hollinger
- Letort
- Manor
- Pequa

Monthly values were calculated for leachate, runoff, and erosion per hectare. Unfortunately, no data for the Pequa Series were available in the DBAPE database; therefore, this portion of the analyses omitted it. In addition, analyses of the Manor Soil Series included more detailed probabilistic modeling where the surface slope held constant (6-percent slope) and the surface soil organic content varied to include the high, average, and low organic contents as listed in DBAPE. PRZM also calculated the volume of water as evapotranspiration, runoff, and recharge through the root zone.

Results

The results of this study should demonstrate the application of transport modeling to the possible protection of water resources. Regulatory decision-makers should not consider these results in their current form because such decisions would require a much more rigorous simulation strategy to increase the level of confidence in the data. As a demonstration study, however, the results do show the usefulness of this approach. Table 1 contains the cumulative frequency data for the simulated atrazine residues in runoff, erosion, and leaching that occurred under 30 years of historical precipitation. The data cover the 12 soils mentioned, with the surface slope held constant at 6 percent.

Atrazine in Runoff and Erosion

The results of this analysis suggest that the Haggerstown Series had the greatest potential for yielding atrazine in runoff; approximately 50 percent of the simulated monthly atrazine in runoff values equaled or exceeded 0.0001 kilogram per hectare. Conversely, the Elk Series yielded the least atrazine to runoff; 50 percent of the runoff data were at residue levels of 1×10^{-6} kilograms per hectare. Within the Big Beaver Creek subbasin, the Manor Series had the greatest potential to yield atrazine in runoff.

As with the runoff data, the Haggerstown Series had the greatest potential to yield atrazine in eroded sediments, and the Elk Series yielded the least atrazine in erosion. Within the Big Beaver Creek subbasin, the Manor Series had the greatest erosion potential regarding atrazine.

GIS Analyses

After entering the results from the transport modeling into a database, GIS could produce maps showing the location of soils with high runoff potentials. Figure 2a shows the orientation of soil series around the surface water sampling points in Big Beaver Creek. Figure 2b represents the same scene but fills in the soils with high runoff potential. Using this type of analysis can help areas that may be sources of nonpoint source runoff contamination. Once identified, these soils can be targeted for alternative management practices that may reduce the amount of runoff and the degree of nonpoint source contamination.

Table 1. Cumulative Frequency of Simulated Atrazine Residues in Runoff for 12 Major Soils in Lancaster County, Pennsylvania (values are percentage of data)

Load (kilograms per hectare)	Soil Series in Lancaster County, Pennsylvania											
	Bucks	Chester	Clymer	Connestoga	Elk	Glenelg	Haggerstown	Hollinger	Lansdale	Letort	Manor	Ungers
IE-10	83.33	84.85	84.09	84.08	78.79	84.85	93.94	83.33	88.64	85.61	86.36	84.09
IE-9	78.79	80.30	80.30	81.82	75.76	80.30	93.94	81.06	85.61	84.85	84.61	84.09
IE-8	74.24	76.52	75.00	78.03	69.70	76.52	93.94	77.27	80.30	84.09	84.85	81.82
IE-7	66.67	68.94	66.67	72.73	59.09	68.18	91.67	68.94	75.00	83.33	75.52	75.76
IE-6	56.06	46.97	56.06	61.36	50.00	56.06	84.85	59.09	65.15	76.52	66.67	67.42
IE-5	47.78	35.61	46.21	50.00	43.18	47.73	74.24	47.73	51.52	67.42	46.97	53.03
IE-4	36.36	21.97	32.58	35.61	31.82	34.85	53.79	34.85	40.15	46.97	31.82	39.39
IE-3	24.24	15.15	21.21	21.21	21.21	21.21	37.12	21.21	24.24	31.82	16.67	24.24
0.00	15.91	6.82	14.39	15.12	15.15	13.64	19.70	15.15	15.91	17.24	4.55	15.15
0.01	6.82	0.00	6.06	6.06	6.82	3.79	7.58	6.06	6.82	4.55	0.00	4.55
0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table reads as follows: Given the Elk Series, the first value reads that 78.79 percent of the simulated data were greater than or equal to IE-10 kilograms per hectare. Similarly, within the same column, 6.82 percent of the data were greater than 0.01 kilograms per hectare but less than 0.1 kilograms per hectare.

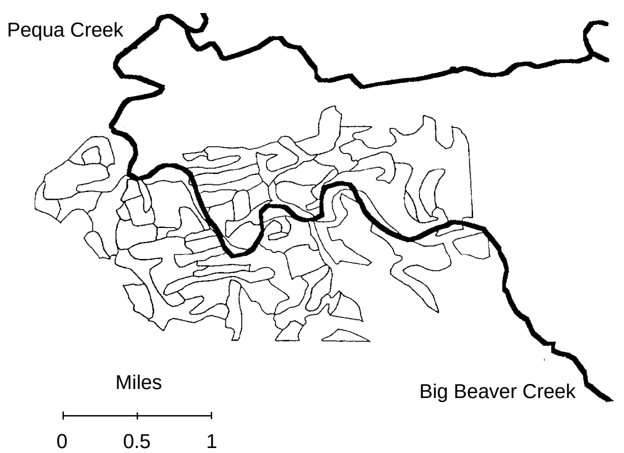


Figure 2a. Soil series in the Big Beaver Creek basin.

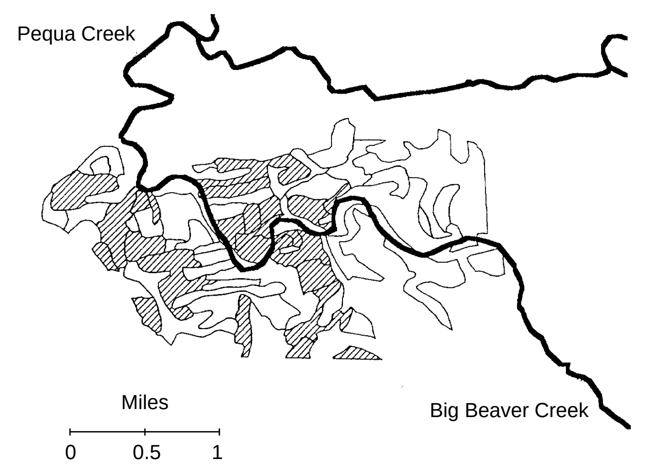


Figure 2b. Soils sensitive to atrazine runoff in the Big Beaver Creek basin.

Detailed Modeling of the Manor Series

Performing an introductory probabilistic modeling exercise allowed further investigation of the potential of the Manor Series to release atrazine into runoff. The existing 30-year climate data and the stated agronomic data were retained from the previous modeling. The organic carbon content of the surface soil layer, however, was allowed to vary between the published low, average, and maximum values found in the DBAPE database. This exercise followed the principles set forth by Laskowski et al. (5) and others who describe probabilistic modeling approaches. In essence, by varying input parameters within known endpoints, the probabilistic approach can generate a distribution of pesticide residue values that statistically reflects the anticipated residues. Parameters to vary may include:

- Organic carbon content
- Surface slope
- Kd (distribution coefficient)

- Moisture content

By allowing input variables to vary according to a normal distribution, this approach thereby eliminates some of the uncertainty associated with pesticide transport modeling. The probabilistic modeling approach requires the creation of a significant database by performing many runs (e.g., 1,000 model runs that generate 12,000 monthly values for each soil).

This study included a limited probabilistic modeling exercise. Table 2 lists the results for the mean atrazine residues in runoff, erosion, and leaching for the Manor Soil Series during the:

- Entire year
- Growing season
- Winter months

The surface slope was held constant at 6 percent, but the soil organic carbon content varied within the published range. The means for all months show limited variation in mean residues. Runoff was by far the major source of atrazine. Erosion and leaching values were on similar scales (trace amounts).

Table 2. Summary Statistics for Detailed Modeling of the Manor Soil Series in the Pequa Creek Watershed (statistics based on 1,080 values)

	Mean Atrazine Residue (kilograms per hectare)		
Percent Organic Carbon[a]	**Runoff**	**Erosion**	**Leaching**
All Months			
Low	0.01381	0.00024	0.00028
Average	0.01229	0.00042	0.00012
High	0.01105	0.00057	< 0.000006
Growing Season			
Low	0.03290	0.00058	0.00017
Average	0.02881	0.00100	< 0.000008
High	0.02540	0.00130	< 0.000004
Winter Months			
Low	0.00014	< 0.000002	0.00036
Average	0.00045	< 0.000001	0.00014
High	0.00075	< 0.000003	< 0.000007

[a]Data taken from DBAPE soils database as low, midpoint, and maximum reported organic contents.

The greatest runoff and erosion values occurred during the growing season. The greatest leaching, however, occurred during the winter months. These results support the general observations that surface residues run off during spring and summer but that as the crops grow and evapotranspiration increases, recharge to ground water decreases, subsequently limiting pesticide transport to ground water. Conversely, during the winter months, the surface soil pesticide residues generally decrease because of exposure to months of photolysis, hydrolysis, and biodegradation. Subsurface residues have been protected from degradation, however, and increased ground-water recharge, due to great reductions in evapotranspiration, transports the residues through the soil column.

This limited exercise provided a valuable learning experience regarding probabilistic modeling. As computing techniques and hardware advance, the cost in time and money for each simulation should decrease dramatically. Although researchers tend not to have great faith in pesticide transport modeling, the advances in this field will reduce uncertainty and instill greater confidence in the modeling process.

GIS Pitfalls

GIS is a powerful tool and has great promise for use in environmental problem-solving. Several points or pitfalls, however, hinder broad acceptance of GIS. As with most new technologies, cost is the overriding concern in using GIS. Although technical staff and project scientists understand the power of GIS and the effort that data preparation requires, management and corporate staff often do not see the benefits for the costs. Many managers assume that current GIS systems resemble those seen on “Star Trek,” and when reality becomes apparent, managers tend to discard GIS as too costly and complex. Several points need consideration when contemplating the use of GIS. Although various products exist, this discussion focuses on ARC/INFO and MAPINFO products.

Hardware

Computer hardware is plentiful if the available budget can support a purchase. Many high- powered GIS packages (e.g., GRASS, ARC/INFO, INTERGRAPH, IDRISI) run best on mainframes or minicomputers. Most technical staff, however, only have access to PC machines. Corporate purchasing departments more readily expend funds for PC technology because they will eventually find use for these machines even if they are not used for GIS. A recent ARC/INFO advertisement (August 1994) lists costs for SUN SPARC minicomputer systems with ARC/INFO software at $12,000 to $15,000 depending on configuration.

Minicomputers and mainframes require specialized staff to configure and maintain the hardware. Today, many staff level personnel can open and augment their PC machines with a minimum of external support. GIS performance reflects the tradeoff in hardware, particularly when a considerable amount of data manipulation is required. For example, if linking discrete depth soil series data to STATSGO soil associations is necessary, then a minicomputer system may be best. The postprocessed data could, however, be exported to a format that will run on PC-based systems. If the user wants to import and manipulate remote sensing imaging (e.g., SPOT or Landsat data), then minicomputers are recommended. If the user wants to display already edited images and preprocessed GIS data, then PC-based computing may be sufficient. The ultimate use of GIS drives the hardware selection.

Software

A great number of GIS software packages are available to meet almost any level of use and expertise. Software runs under both UNIX and DOS/Windows (denoted as DOS for the remainder of this paper) operating systems. The UNIX-based software tends to be more powerful and flexible than the DOS-based software. UNIX-based packages require more specialized staff to optimize GIS, however.

UNIX-based software packages include GRASS, ARC/INFO, INTERGRAPH, and IDRISI. Costs vary from public domain charges for GRASS and IDRISI to vendor supplied ARC/INFO and INTERGRAPH, which can cost several thousand dollars each.

DOS versions of ARC/INFO (e.g., PC ARC/INFO, ARCAD, ARC/VIEW) are also available and provide the user with various levels of data editing and manipulation abilities. Generally, PC ARC/INFO is the same as the UNIX version, varying in speed of processing. ARCAD is a GIS engine that uses AutoCAD for drawing and displaying, giving the user most of the abilities of the UNIX-based version. ARC/VIEW I was primarily a display and simple analysis tool. It allowed the user to view, display, and manipulate existing arc data but did not support image editing. Currently, ARC/VIEW II provides more support for image editing and data manipulation. Costs range from about $3,000 for PC ARC/INFO and ARCAD (AutoCAD also costs about $2,000) to around $500 for the ARC/VIEW products.

MAPINFO is a DOS-based GIS that was designed for marketing and demographic applications. Several researchers, however, have used MAPINFO for environmental applications. The most outstanding feature of MAPINFO is that it easily imports data layers as it reads dBASE type files directly. MAPINFO V3 also reads database files and recreates them as *.TAB files. In contrast to the “coverage and entity” concepts of the ARC/INFO line of programs, MAPINFO reads latitude and longitude coordinates and displays the results. This simplifies data management because many researchers who have already created custom databases can easily import those data as long as latitude and longitudes coordinates are present.

As with the ARC/INFO line of programs, many common data layers can be purchased for use in MAPINFO. These layers can be expensively priced, costing approximately $1,000 per county for roadway, census, and demographic data. One major lapse is the poor library of environmental layers, USGS topography, hydrography, soil boundaries, or climate stations. MAPINFO does sell a module that allows users to convert to and from ARC/INFO coverages so that common data layers can be established. Experience shows, however, that conversion programs do not always work as advertised. For example, large boundary files (STATSGO data for Indiana) do not readily convert from ARC/INFO to MAPINFO. Third-party vendors may be needed to convert data for use in MAPINFO.

One very important factor supporting the use of MAPINFO is that it has a business application slant; therefore, it is slightly easier to convince corporate management to invest in GIS because marketing and sales data (territories) can be relatively easily overlain onto environmental data.

Finally, some packages that are add-ons to spreadsheet programs tend not to be powerful or versatile enough for use in environmental GIS work. These software packages may be valuable as an introduction to GIS techniques, however.

Data Availability and Format

After compiling the hardware and software into GIS, the next step entails accessing data layers such as:

- State and county boundaries
- Land use covers
- Water boundaries

Currently, USGS DLGs for hydrography, land use, transportation, and cultural features are available for minimal costs. Shareware programs can convert the USGS DLG3 formats into DXF (data transfer files) for import to GIS packages. These data require conversion to DXF or ARC coverage type formats for use in either ARC/INFO or MAPINFO.

The USDA Soil Conservation Service produces digital data for soil types (STATSGO and SSURGO) that users can import to ARC/INFO relatively easily. The STATSGO data cost approximately $1,000 per state and are available for most states. The detailed soil series maps, SSURGO, cost approximately $500 per county and are not available for every county in the United States. Many data layers are available for direct use by GRASS. As of yet, however, no convenient conversion utilities exist to move GRASS data to ARC/INFO or MAPINFO. The

U.S. Fish and Wildlife Service now distributes data layers from the National Wetlands Inventory on the Internet (enterprise.nwi.fws.gov).

Other data sources available through private vendors are listed in the MAPINFO and ARC/INFO user guides and in any issue of *GIS World*. The user should be prepared to absorb significant costs if purchasing all the required data layers.

Conclusion

This study shows that the technology and software exist for a water resource manager to couple pesticide transport modeling with GIS to identify areas or individual soils that may contribute to nonpoint source pollution of water resources. The study used PC-based computing system and software. Soils maps and hydrographic maps can be easily digitized for limited cost. A skilled scientist or technician, without being a GIS expert, can run GIS to answer specific questions. The technologies this study demonstrated may be extremely valuable to managers responsible for producing SMPs.

The pesticide transport modeling performed during this study was intended for illustrative purposes. More detailed analyses, and additional simulations, would be necessary to use these data for regulatory actions or land use management. The study did, however, succeed in identifying the Manor Soil Series, with slopes exceeding 6 percent, within the Big Beaver Creek subbasin of the Pequa Creek basin, as a potential source for atrazine in runoff.

References

1. Ward, J.R. 1987. Surface-water quality in Pequa Creek Basin, Pennsylvania, 1977-1979. U.S. Geological Survey. Water Resources Investigation Report 85-4250.
2. Smith, M.C., A.B. Bottcher, K.L. Campbell, and D.L. Thomas. 1991. Field testing and comparison of the PRZM and GLEAMS models. Trans. Amer. Soc. of Agric. Eng. 34(3):838-847.
3. Mueller, T.C., R.E. Jones, P.B. Bush, and P.A. Banks. 1992. Comparison of PRZM and GLEAMS computer model predictions with field data for alachlor, metribuzin, and norflurazon leaching. Environ. Toxicol. and Chem. 11:427-436.
4. Pennell, K.D., A.G. Hornsby, R.E. Jessup, and P.S.C. Rao. 1990. Evaluation of five simulation models for predicting aldicarb and bromide behavior under field conditions. Water Resour. Res. 26(11):2,679-2,693.
5. Laskowski, D.A., P.M. Tillotson, D.D. Fontaine, and E.J. Martin. 1990. Probability modeling. Phil. Trans. Royal Soc. of London 329:383-389.
6. Ward, J.R., and D.A. Eckhardt. 1979. Nonpoint-source discharges in Pequa Creek Basin, Pennsylvania, 1977. U.S. Geological Survey. Water Resources Investigation Report 79-88.
7. Custer, B.H. 1985. Soil survey of Lancaster County, Pennsylvania. Washington, DC: U.S. Department of Agriculture, Soil Conservation Service.
8. U.S. Geological Survey. 1977. Lancaster County, Pennsylvania: County map series (topographic) 1:50,000 scale.
9. Carsel, R.F., C.N. Smith, L.A. Mulkey, D.J. Dean, and P. Jowise. 1984. Users manual for Pesticide Root Zone Model (PRZM), Release I. EPA/600/3-84/109.

Integration of GIS With the Agricultural Nonpoint Source Pollution Model: The Effect of Resolution and Soils Data Sources on Model Input and Output

Suzanne R. Perlitsh
State University of New York, Syracuse, New York

Abstract

The assessment of agricultural nonpoint source pollution has been facilitated by linking data contained in a geographic information system (GIS) with hydrologic models. One such model is the Agricultural Nonpoint Source (AGNPS) Pollution Model, which simulates runoff, nutrients, and sediment from agricultural watersheds. Vector-based (ARC/INFO) and raster-based (IDRISI) GIS systems were used to generate AGNPS input parameters.

The objectives of this project were to generate AGNPS input parameters in GIS format from GIS data at different resolutions and different levels of detail (soil survey soils data versus soils data currently available in digital format from the United States Department of Agriculture). Differences in the AGNPS model sediment output based on the variations in GIS-generated AGNPS model input were evaluated.

The study also evaluated the influence of cell size resolution and soils data on sediment generated within each cell in the watershed (SGW), sediment yield from each cell in the watershed (SY), sediment yield at the watershed outlet, and peak flow. Model output was validated by comparison with measured values at the watershed outlet for a monitored storm event. Results of this study indicate that the use of different resolution GIS data and different soils data sources to assemble AGNPS input parameters affects AGNPS model output. Higher resolution data do not necessarily provide better results. Such comparisons could affect decision-making regarding the level and type of data analysis necessary to generate sufficient information.

Introduction

Agricultural runoff is a major contributor to nonpoint source pollution. Fifty-seven percent of the pollution in impaired lakes and 64 percent of the pollution in impaired rivers of the United States can be attributed to agricultural nonpoint source pollution (1). Sediment is one of the most common agricultural nonpoint source pollutants and is the largest pollutant by volume in the United States (2). More than 3 billion tons of sediment enter surface waters of the United States each year as a result of agricultural practices (1).

Accurate assessment of the effects of agricultural activities on water quality within a watershed is vital for responsible watershed management and depends on our ability to quantify the spatial variability of the watershed and the complex interactions of hydrologic processes (3). Computer models have been developed to simulate these hydrologic processes to provide estimates of nonpoint source pollutant loads. Adequate simulation of a watershed's spatial variability helps provide the best representation of hydrologic processes within the watershed.

Preservation of spatial variability within hydrologic models can be accomplished using a distributed parameter model. The distributed parameter model is more advantageous than lumped parameter models, which generalize watershed characteristics, because distributed parameter models provide more accurate simulations of the systems they model (4). One of these models is the Agricultural Nonpoint Source (AGNPS) Pollution Model. AGNPS is a distributed process model because it produces information regarding hydrologic processes at grid cells within the watershed, thus enabling preservation of the spatial variation within the watershed. Distributed parameter models integrate well with GIS because GIS can replicate the grid used in a distributed parameter model. Manual compilation of AGNPS input parameters required to evaluate small areas at low resolution (large grid cells) is relatively easy. Manually assembling data to evaluate larger areas at finer resolutions becomes tedious, however. The integration of GIS data with the AGNPS model facilitates data assembly and manipulation (5).

Several researchers have integrated AGNPS with GIS (4-11). Smaller cell sizes within distributed parameter models are thought to best represent spatial variability within a watershed (5, 10). Certain AGNPS input parameters show sensitivity to changes in grid cell size, affecting sediment yield output (11). The use of GIS to generate input parameters for the AGNPS model enables analysis of watersheds at higher resolutions than would be practical using manual methods (5).

Research Hypotheses

The project investigated the following research hypotheses:

- AGNPS output at the highest resolution will better approximate sediment yield at the watershed outlet.
- AGNPS output for sediment generated within each cell in the watershed at highest resolution will best reflect the watershed processes.
- AGNPS output generated from the more detailed soil survey data will better estimate watershed processes.

The project also investigated other questions: will certain AGNPS input parameters (cell land slope, soil erodibility [K], the cropping factor [C], and the U.S. Department of Agriculture [USDA] Soil Conservation Service [SCS] curve number [CN] show sensitivity to changes in grid cell size? How does slope affect model output? Does a qualitatively significant difference exist between model input parameters and output calculated from data sets generated at different resolutions with different levels of detail in soils output?

Objectives/Tasks

The research in this project included analyses of:

- Certain AGNPS input parameters generated at different resolutions (10- x 10-, 30- x 30-, 60- x 60-, and 90- x 90-meter resolutions).
- AGNPS output for sediment yield (SY) and sediment generated within each cell (SGW) at different resolutions (center cells of 10- x 10-, 30- x 30-, 60- x 60-, and 90- x 90-meter resolutions).
- AGNPS output generated from different levels of detail in the soils input data (soil survey versus STATSGO data sources).

Significance

Version 4.03 of AGNPS was released in June 1994. Version 4.03 allows for evaluation of 32,767 cells. This version allows for cell sizes from 0.01 to 1,000 acres (approximately 6.36- x 6.36-meter resolution to 2,012- x 2,012-meter resolution). Previous versions of AGNPS limited the number of cells to 3,200 and the cell size resolution to 0.4 hectares (or 63.25 x 63.25 meters) (12). Reviewed literature provides no evidence that AGNPS has been used to evaluate a watershed at 10- x 10-meter resolution.

The soils data in this study were compiled at two different levels of detail. Soils data at the 1:20,000 soil survey level were generated in digital format. This level of detail was compared with soils data at the State Soil Geographic (STATSGO) database level with a scale of 1:250,000. Reviewed literature mentions no previous studies comparing AGNPS output with input generated from these two different levels of detail in soils input.

Technology for collecting and processing geographic data is continuously improving. Currently, the United States Geological Survey (USGS) 1:24,000 digital elevation models (DEMs) are available at 30- x 30-meter resolution. New satellite technology will enable DEM data to be available at 10- x 10-meter resolution, or higher. Certain satellites currently provide land cover data at 10- to 30-meter resolution (13).

An important objective of this project was to determine whether higher resolution data provide different results when routed through AGNPS. Does spatial data at higher resolutions provide better information? This paper describes the results of an analysis of AGNPS output based on different levels of both resolution and soils detail in GIS data input sources.

Materials

The Study Area

An ongoing effort is underway to clean up Onondaga Lake, Onondaga County, New York. To accomplish this effort, areas contributing agricultural nonpoint source pollution to the lake are being evaluated.

The Onondaga Lake watershed is approximately 287.5 square miles, with 40 subwatersheds. The subwatersheds in the agricultural portion of the Onondaga Lake watershed (south of Syracuse, New York) have been isolated for study of their potential nonpoint source contributions to Onondaga Lake (see Figure 1). The study area watershed (1.84 square miles, 1,177.5 acres) is one of these agricultural subwatersheds. GIS data were collected within the Otisco Valley quadrangle (USGS 1:2,400), which includes the southern portion of the Onondaga Lake watershed. Elevations in the study watershed range from 1,820 feet to 1,203 feet, with an average elevation of 1,510 feet. The watershed perimeter is approximately 6.5 miles (34,505 feet). The streams in the watershed flow from south to north to Rattlesnake Gulf, with a stream length of approximately 3.08 miles (16,265.4 feet). The stem fall of the main stream stem is quite steep at 283 feet per mile. The drainage density of the watershed is 1.67 miles of stream per square mile. Land use in the watershed is predominantly agricultural (82.8 percent).

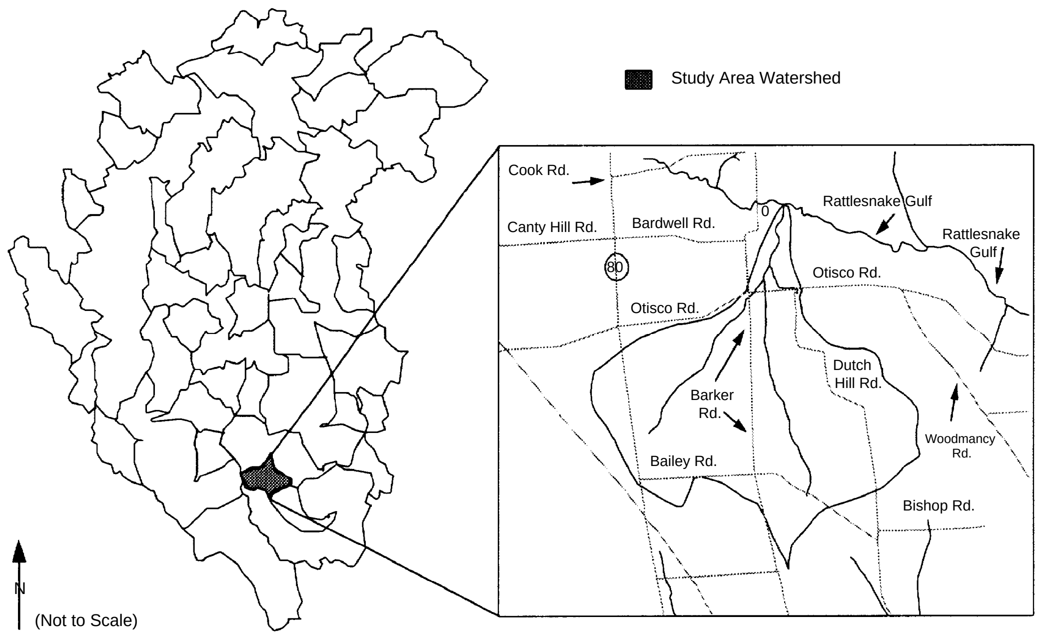


Figure 1. Onondaga Lake watershed and study area (not to scale).

The AGNPS Model

AGNPS was developed to analyze and provide estimates of runoff water quality, specifically to evaluate sediments and nutrients in runoff from agricultural watersheds for a specific storm event (11). To use AGNPS, a watershed is divided into cells of equal area. Calculations for each of the model output values are made within each cell based on the watershed data contained in each cell. Approximately 1,000 people in 46 different countries use the AGNPS model. Users include students, university professors, government agencies, lake associations, and environmental engineers.[1] AGNPS was developed in 1987 by the Agricultural Research Service (ARS) in cooperation with the Minnesota Pollution Control Agency (MPC) and the SCS. The model runs on an IBM-compatible personal computer.

Data Sources

The GIS packages of ARC/INFO Version 3.4D (14) and IDRISI Version 4.1 (15) were used to prepare input parameters for the AGNPS model. AGNPS input parameters were derived from three base maps—land use, a DEM, and soils. Table 1 shows the 22 input parameters that AGNPS required, and the base source for the data.

The land use map was obtained from a classified ERDAS image (resolution of 28 x 28 meters). The image was converted to IDRISI, brought into ARC/INFO, and regridded based on the resolution requirements of each data set. USGS could not provide a DEM for the study area, so the DEMs were interpolated from points digitized in ARC/INFO based on the Clarke method (16). The DEMs were interpolated in IDRISI on 10- x 10-, 30- x 30-, 60- x 60-, and 90- x 90-meter resolution surfaces.

Soil survey data were obtained from Onondaga County Soil Survey air photographs. An orthophoto of the 7.5 minute quadrangle was obtained from the USGS and was used with a zoom transfer scope to ortho-correct the soil survey data. The corrected soil polygons were then digitized in ARC/INFO. The Otisco Valley quadrangle comprises 79 soils mapping units. Thirty-eight different mapping units occur in the study area watershed.

The USDA SCS (now the Natural Resource Conservation Service [NARCS]) provides digital soils data from its STATSGO database. The mapping scale of STATSGO data is 1:250,000, thus it is best suited for broad planning and management uses. The number of soil polygons per quadrangle is between 100 and 400, and the minimum area mapped is 1,544 acres. The STATSGO soil data used in this project were obtained from the Onondaga County Soil Conservation Service. Approximately seven STATSGO soil groups were identified for the Otisco Valley quadrangle. Only one STATSGO soil type occurs in the study area watershed (Honeyoe silt loam).

Methods

AGNPS input parameters that showed sensitivity to changes in grid cell size in previous studies were compared between the resolutions. The AGNPS model was run eight times using precipitation values from the actual

[1] Personal communication from AGNPS Technical Support, September 1994.

Table 1. AGNPS Input Parameters

#	AGNPS Parameter	Root Data Source	General Derivation of Input
1.	Cell number	Watershed map	Program written to determine #
2.	Cell division	Not applicable	No cell division, assumed 0
3.	Receiving cell number	Aspect map from DEM	Program written to determine #
4.	Receiving cell division	Not applicable	No cell division, assumed 0
5.	Flow direction	Aspect map from DEM	Reclassed 1-8 from azimuth map
6.	SCS curve number	Land use and soils coverage	SML written to determine CN
7.	Land slope percentage	Slope map from DEM	Provided in slope percentage from IDRISI
8.	Slope shape factor	Algorithm	Assume uniform slope
9.	Average slope length	Table of values	Obtained from SCS
10.	Manning's n coefficient	Literature values	Attached to land use database
11.	USLE K factor	SCS and soil survey	Attached to soils database
12.	USLE C factor	Literature values	Attached to land use database
13.	USLE P factor	Literature values	Attached to land use database
14.	Surface condition constant	Land use coverage	Attached to land use database
15.	Chemical oxygen demand	Land use coverage	Attached to land use database
16.	Soil texture	Soil survey	Attached to soils database
17.	Fertilizer indicator	Land use coverage	Assumed for agricultural land class
18.	Pesticide indicator	Land use coverage	Assumed for agricultural land class
19.	Point source indicator	USGS 1:24,000 map	Points in ARC/INFO and IDRISI
20.	Additional erosion	Field survey, known gullies	Assume no additional erosion
21.	Impoundment indicator	1:24,000 map, field survey	Assume no impoundments
22.	Channel indicator	Streams coverage	Assume no significant channel

storm that was monitored. Each time, the model was run using an input file created with the different input data sources as follows:

1. 30- x 30-meter resolution-soil survey data.
2. 30- x 30-meter resolution-STATSGO data.
3. 60- x 60-meter resolution-soil survey data.
4. 60- x 60-meter resolution-STATSGO data.
5. 90- x 90-meter resolution-soil survey data.
6. 90- x 90-meter resolution-STATSGO data.
7. Center cells of the 10- x 10-meter resolution—soil survey data.
8. Center cells of the 10- x 10-meter resolution—STASGO data.

As grid cell size increases, the time required to assemble data as well as the space required to store the data files increase. If a cell size resolution is cut in half, the number of cells in that coverage quadruples. In the study area watershed, increasing grid cell size from 90- x 90-meter resolution to 60- x 60-meter resolution created 778 more cells within the watershed. Moving from 60- x 60-meter resolution to 30- x 30-meter resolution added 3,646 cells to the watershed, and moving from 30- x 30-meter resolution to 10- x 10-meter resolution added 40,115 cells to the watershed data set (see Figure 2). Due to the 32,767-cell limitation of AGNPS Version 4.03, AGNPS output for SY and SGW at the 10- x 10-meter resolution (which contains 45,104 cells) could not be obtained. Input parameters at 10- x 10-meter resolution, however, could be compared with input parameters at 30- x 30-meter resolution.

Methodology of Data Analysis

The input parameter "maps" were converted to IDRISI files and combined in a format that could be routed through the AGNPS model. AGNPS model output for soil generated within each cell and for sediment yield was assembled. The 30- x 30-meter resolution maps were compared with the 60- x 60-meter resolution maps; the 60- x 60-meter resolution maps were compared with the 90- x 90-meter resolution maps; and the 30- x 30-meter resolution maps were compared with the 10- x 10-meter resolution maps.

A method for comparing maps with different grid sizes was developed so that maps of different resolutions

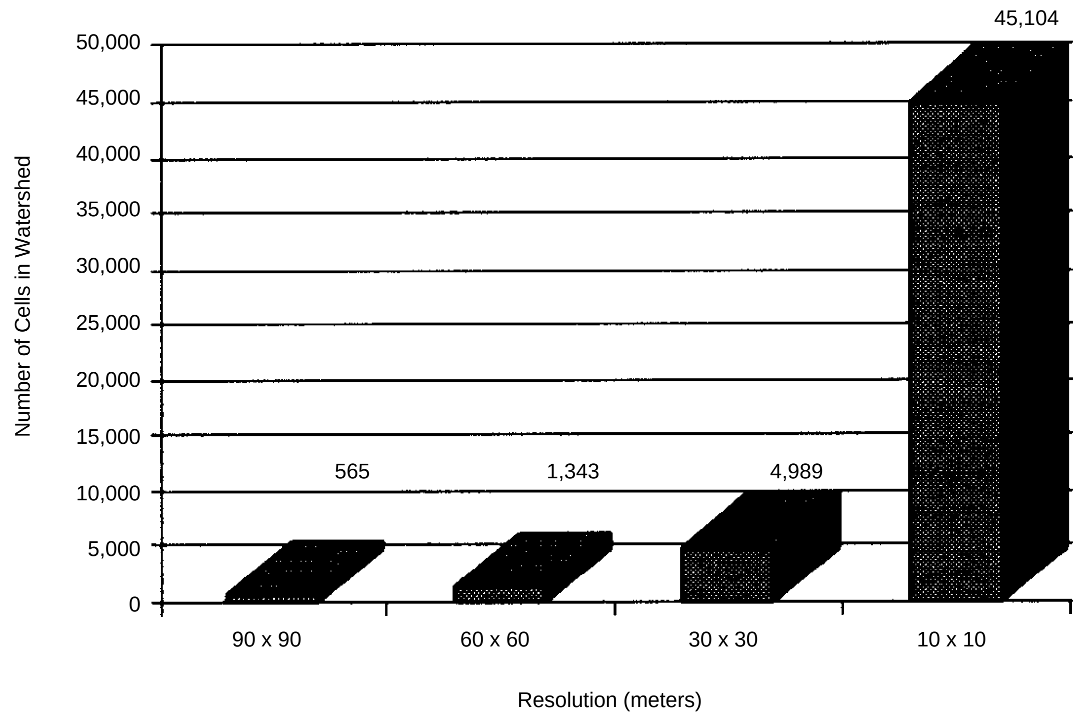


Figure 2. Effect of resolution on study area data set.

could be compared using the root mean square (RMS) statistic. This method was selected to provide a means for statistical analysis that could accommodate the spatial variability of this data set. The 60- x 60-meter resolution maps were “expanded” in IDRISI. This duplicated each grid cell so that the original 60- x 60-meter resolution map was equivalent to the 30- x 30-meter resolution map. The IDRISI command “RESAMPLE” was used to bring the 60- x 60-meter resolution map data onto the same grid size as the 90- x 90-meter resolution map. The center cells of the 10- x 10-meter resolution map were selected for comparison with the 30- x 30-meter grid cells. This comparison is based on the fact that the center cell of the 10-meter resolution is the cell that best corresponds to the entire cell of 30-meter resolution (see Figure 3).

Once maps were registered on comparable grids, the RMS difference between the maps of differing resolution was used to compare the difference between the maps

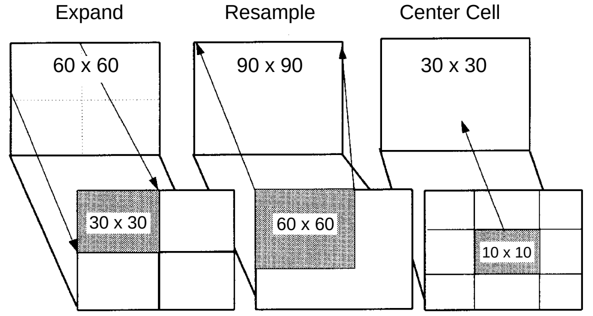


Figure 3. Method for comparing cells of differing resolution.

and the effect of cell size resolution. The RMS statistic is a measure of the variability of measurements about their true values. The RMS is estimated by comparing values in one grid system with the values in the comparison grid system. The difference between corresponding values in each grid system is squared and summed. The sum is then divided by the number of measurements in the sample to obtain a mean square deviation. Finally, the square root of the mean square deviation is calculated. The RMS difference quantifies the discrepancy between two data sets.

$$\mathrm{RMS} = \sqrt{\sum_{i=1}^{n} \frac{(\mathrm{grid1} - \mathrm{grid2})^2}{n}}$$

Results

The Storm Event

The storm that was monitored for the purposes of this field validation occurred on August 28, 1994, at approximately 8:45 p.m.; the storm duration was approximately 1.5 hours. It was a high intensity, short duration thunderstorm (see Figure 4). A global flow probe (Model FP101) was used to obtain discharge velocity measurements in feet per second. These values were then multiplied to obtain cubic feet per second. The peak discharge occurred at 11:30 p.m. on August 28, 1994, with a flow discharge of 11.15 cubic feet per second. The average runoff for the period was 4.458 cubic feet per second, or 2.42 cubic feet per second per square mile. The runoff volume per day was 0.09 inches.

Total sediment yield was derived from the analysis of total solids measured in field samples throughout the 24-hour storm period. The samples were processed to evaluate total suspended solids (TSS) using the vacuum filtration procedure (17). A total of 1.204 tons of suspended sediment was predicted at the watershed outlet from field data samples. A LaMotte field nutrient test kit was used to measure nitrate and phosphate concentrations in the stream. Nutrient values in this watershed for this storm event were so small (phosphorous below 0.1 parts per million and nitrogen 0.3 parts per million), they were not selected as parameters to be used in evaluating and validating the AGNPS model. The AGNPS predicted nutrient output for the storm was not measurable (0.00 parts per million). The low levels of nitrogen and phosphorous in the stream channel during the storm event can be attributed to the time of year in which the stream was monitored. At the time of field validation, agricultural activities were not operating.

Results at the Watershed Outlet: Peak Flow

The peak flow values that AGNPS calculated are largest at the highest resolution and decrease as cell size increases. The peak discharge from the watershed during the monitored storm event was 11.15 cubic feet per second. Comparisons of the AGNPS predicted peak flow and the actual field-validated peak flow showed that the 30- x 30-meter resolution cells best approximate the peak flow of the watershed for the sampled storm event. As grid cells increase from 30 x 30 to 60 x 60 and from 60 x 60 to 90 x 90, the peak flow is underestimated. As grid cells decrease from 30 x 30 to 10 x 10, the peak flow is grossly overestimated (see Table 2).

Results at the Watershed Outlet: Sediment Yield

Sediment yield at the watershed outlet was determined to be 1.204 tons. In all of the resolutions, the amount of sediment deposited at the watershed outlet cell increased as cell resolution increased (see Table 2). For this particular watershed in this particular storm, the AGNPS model overestimated the sediment yield predicted at the watershed outlet at the 10- x 10-, 30- x 30-, and 60- x 60-meter resolutions and underestimated sediment yield at the 90- x 90-meter resolution. Table 2 includes the information that AGNPS predicted for the cell designated as the watershed outlet within each resolution. (The results reported include output from the center cells of the 10- x 10-meter resolution data set, routed through the AGNPS model. Although these values are reported, the results from this data set cannot be assumed to approximate the sediment output that would result had the entire 10- x 10-meter resolution data set been simulated.)

Soil Survey Versus STATSGO Data

The Kappa statistic (14,18) was used as an indicator of similarity to describe the differences between the AGNPS output for SY and SGW generated from STATSGO and soil survey data. Results (see Table 3) indicate that no significant difference exists between the output derived from the STATSGO and soil survey data inputs within

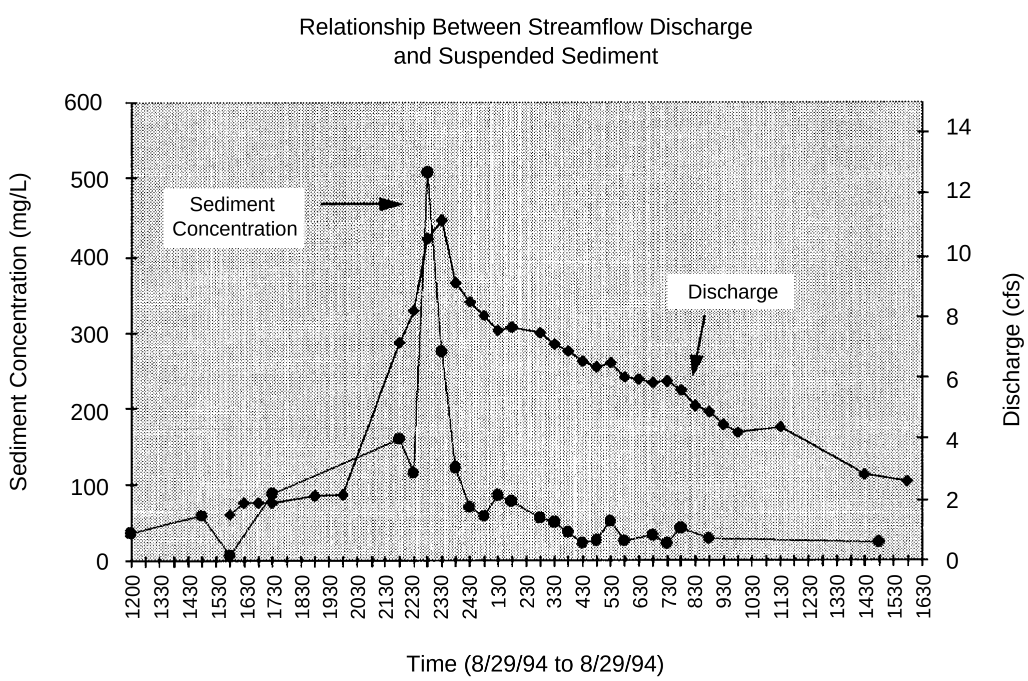


Figure 4. Storm event hydrograph and pollutograph.

Table 2. AGNPS Results at the Watershed Outlet Versus Actual Field Values

Resolution (meters)	Predicted Soil Survey Data	Difference From Actual	Predicted STATSGO Data	Difference From Actual
10 x 10 (center cells)				
Peak runoff rate (cfs)[a]	29.88	+18.73	29.88	+18.73
Total sediment yield (tons)	4.48	+3.49	4.69	+3.28
30 x 30				
Peak runoff rate (cfs)	11.27	+0.12	11.27	+0.12
Total sediment yield (tons)	2.84	+1.64	3.11	+1.91
60 x 60				
Peak runoff rate (cfs)	9.96	-1.19	9.96	-1.19
Total sediment yield (tons)	2.11	+0.91	2.12	+0.92
90 x 90				
Peak runoff rate (cfs)	8.87	-2.28	8.87	-2.28
Total sediment yield (tons)	0.86	-0.34	0.87	-0.33

[a]cfs: cubic feet per second.

Table 3. Kappa Coefficient of Similarity

Resolution (meters)	Kappa
10 x 10 (center)	
Soil survey versus STATSGO SY	0.9866
Soil survey versus STATSGO SGW	0.9703
30 x 30	
Soil survey versus STATSGO SY	0.9859
Soil survey versus STATSGO SGW	0.9785
60 x 60	
Soil survey versus STATSGO SY	0.8960
Soil survey versus STATSGO SGW	0.7542
90 x 90	
Soil survey versus STATSGO SY	0.8743
Soil survey versus STATSGO SGW	0.6574

the same resolutions. This may be due to the homogeneity of the soil textures in both soils data sets (both dominated by silty soils).

Effects of Resolution on SGW and SY

AGNPS output for SGW and SY was evaluated for every cell within the watershed. The RMS difference was applied to determine the relative effect of input data resolution on SY and SGW output (see Table 4 and Figure 5).

Table 4. RMS for AGNPS Sediment Loss Output

Description	SGW Pounds	SY Pounds
Soil Survey Data Constant		
10 centers to 30-meter resolution	168.38	344.22
30- to 60-meter resolution	125.43	739.75
60- to 90-meter resolution	93.50	498.01
30- to 90-meter resolution	29.91	711.49
STATSGO Data Constant		
10 centers to 30-meter resolution	164.45	312.48
30- to 60-meter resolution	125.37	782.07
60- to 90-meter resolution	97.07	501.81
30- to 90-meter resolution	122.04	747.51

Moving to higher cell resolutions increasingly affects sediment generated within each cell; the largest difference in sediment generated within each cell occurs as cell resolution increases from 30 x 30 to 10 x 10 meters. Sediment generated within each cell is least affected by moving from 60 x 60 to 90 x 90 meters. Sediment yield per cell is most affected when cell resolution increases from 60 x 60 to 30 x 30 meters and least affected by increasing resolution from 30 x 30 to 10 x 10 meters.

These results prompted an assessment of the methods used to compare resolutions, to determine whether the effect on sediment yield between the 30- x 30- and 60- x 60-meter resolutions could result from the method used in comparing the resolutions (expansion of the 60- x 60-meter resolution). The procedure of comparison between the 30- x 30- and the 60- x 60-meter resolutions was repeated; however, rather than expanding the 60 x 60 data file, the 60 x 60 data file was "resampled" onto the 30- x 30-meter resolution grid, then the files were compared. The RMS results (see Table 5) show that both methods for comparing data between the resolutions provide essentially the same results. The effect of resolution on sediment yield per cell is, in fact, greatest as resolution increases from 60 x 60 to 30 x 30 meters.

Results: AGNPS Input Parameters of Concern

Previous AGNPS analyses have shown sediment yield (and sediment-associated nutrient yields) to be most affected by AGNPS inputs for cell land slope, the soil erodibility factor (K), the Universal Soil Loss Equation's

Table 5. Comparison of RMS for 30 x 30 to 60 x 60

Method	RMS SGW Pounds	RMS SY Pounds
Expansion of 60 x 60 to 30 x 30	125.43	739.75
Resampling 60 x 60 to 30 x 30	125.37	739.30
Difference	0.06	0.45

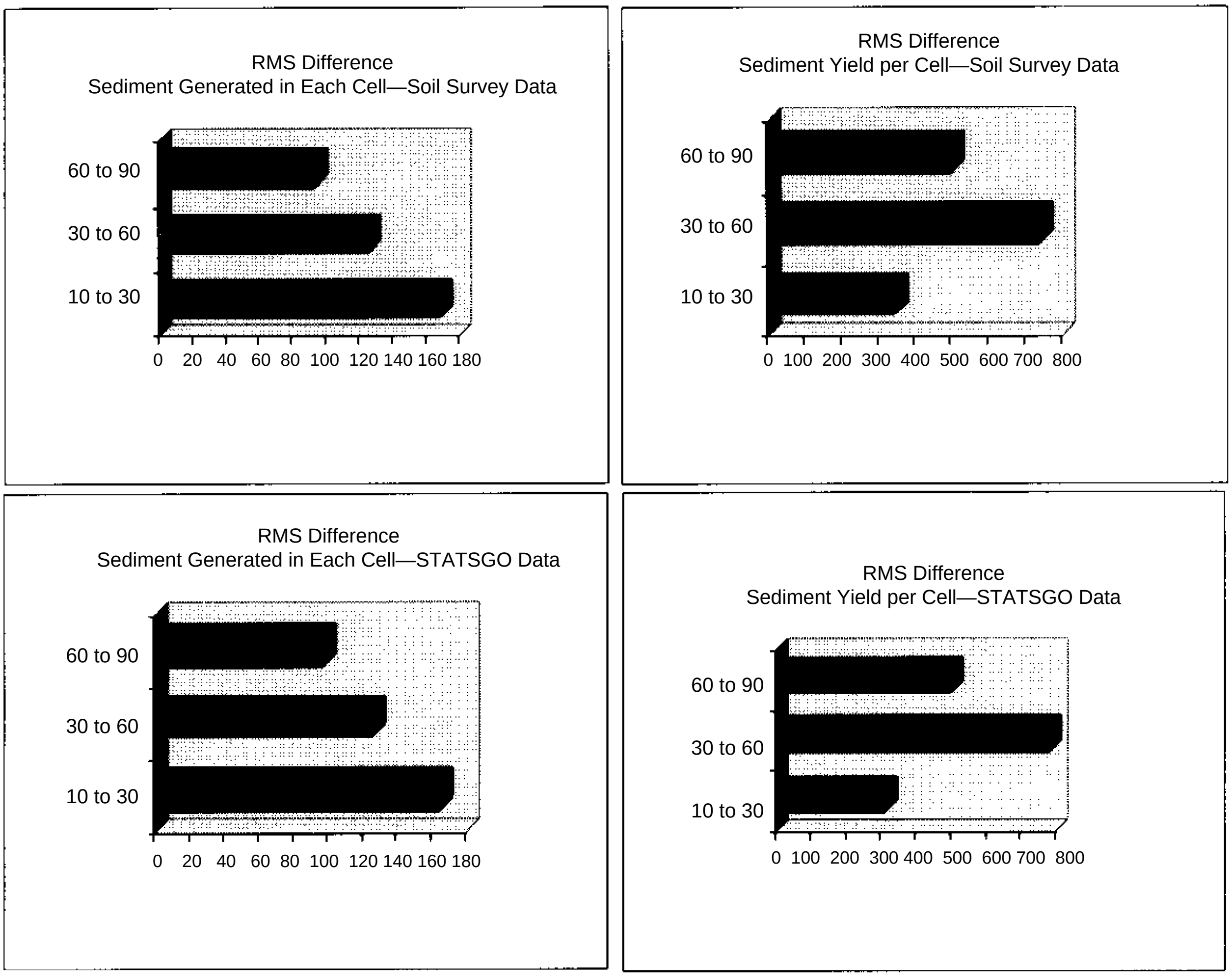


Figure 5. RMS for AGNPS sediment output.

(USLE) cropping management factor (C) and the SCS curve number (CN). To address the concerns regarding the influence of these parameters on sediment yield, the RMS differences (see Table 6 and Figure 6) and general statistics (see Table 7) for these parameters were computed.

Discussion

When evaluating the RMS as an indicator of the effect of resolution on input parameters and output sediment values, looking at the overall trend between resolutions, rather than focusing on specific values, is important. The RMS statistics for the soil erodibility factor (K), the cropping management factor (C), and the SCS curve number (CN) are least affected by a decrease in cell size resolution from 10 x 10 meters to 30 x 30 meters. These parameters are most affected when cell size resolution decreases from 30- x 30-meter to 60- x 60-meter resolution. As resolution decreases further from 60 x 60 meters to 90 x 90 meters, the effect on RMS decreases. The small-large-smaller trend in the RMS for these parameters is the same trend seen in the RMS for sediment yield throughout the watershed. The sediment yield within each cell therefore seems to be most affected by these input parameters. The general statistics for each of these parameters of concern show that very little difference exists in the values within each resolution, with the exception of slope. Slope values are higher at the higher resolutions and decrease as resolution increases. This is related to the method in which the GIS calculates slope.

Table 6. RMS Difference: AGNPS Input Parameters of Concern

Parameters of Concern	10 to 30	30 to 60	60 to 90
K value	0.0058	0.054	0.051
Cropping factor	0.03	0.173	0.015
SCS curve number	2.11	11.40	8.40
Slope	8.17	5.59	3.68

The RMS statistics comparing resolutions for sediment generated within each cell follow the same trend as the RMS statistics for slope percentage. As resolution increases, so do the discrepancies between the compared

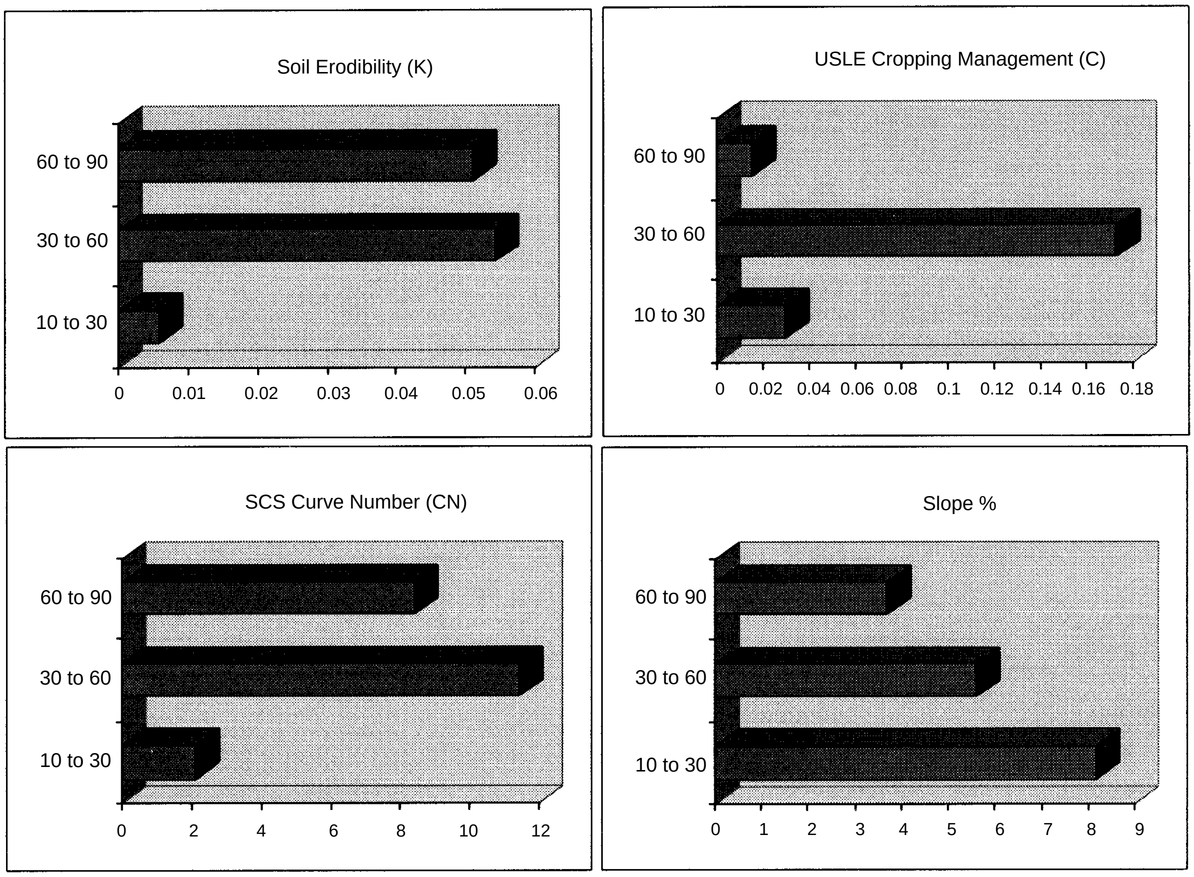


Figure 6. RMS difference for AGNPS input parameter of concern.

data sets. This trend between resolutions indicates that slope values influence sediment generated within each cell in the watershed.

The results from the 10- x 10-meter resolution data set were obtained by selecting the center cells of the 10- x 10-meter resolution data set and routing the data from this set through AGNPS using the flow pathways developed for the 30- x 30-meter resolution. The results do not provide the same information as would be provided had the entire 10- x 10-meter data set been routed through AGNPS. The RMS values obtained from comparisons of the 10- x 10-meter resolution input parameters with the 30- x 30-meter resolution input data reveal that little difference exists between the data in these resolutions. Comparison of the center cell 10- x 10-meter data set output with field monitored data shows that the 10 x 10 center cell data overestimates both peak flow and sediment yielded at the watershed outlet. This can be attributed to the larger slope values in this resolution.

Conclusion

This study used GIS to generate data files for application to the AGNPS model. The objectives of this project were to evaluate the effect of different levels of detail used in generating the input files on selected input and output parameters. The results show that, for a watershed with characteristics equivalent to those of the study area watershed, differences exist in model output based on the cell size resolution of the watershed.

The selected cell size resolution directly affects slope values. The influence of the slope parameter dominates AGNPS predictions for sediment generated within each cell and sediment yield at the watershed outlet in the study area watershed. The indicated parameters of concern have the most influence on sediment yield for each cell in the watershed. The greatest variation in the indicated parameters of concern and thus the sediment yield output occurs between the 30- x 30-meter and 60- x 60-meter resolutions. AGNPS estimates for sediment yield in files generated from STATSGO data were larger than sediment yields from files generated with soil survey soils data in the 30- x 30-, 60- x 60-, and 90- x 90-meter resolutions. For this watershed, however, no significant difference existed between data generated from soil survey and STATSGO data sources as indicated by the kappa coefficient of similarity.

Results predicted by the AGNPS model at the watershed outlet were compared with results from an actual storm monitored at the watershed outlet. The 30- x 30-meter resolution data set provided the most accurate

Table 7. Statistics for AGNPS Input Parameters of Concern

Value	10 x 10	30 x 30	60 x 60	90 x 90
K Value (units of K)				
Average	0.2989	0.2989	0.2882	0.2889
Maximum	0.49	0.49	0.37	0.49
Minimum	0.17	0.17	0.17	0.17
Standard deviation	0.0453	0.0453	0.0513	0.0456
C Factor (units of C)				
Average	0.0306	0.0306	0.0295	0.0295
Maximum	0.076	0.076	0.076	0.076
Minimum	0	0	0	0
Standard deviation	0.0212	0.0211	0.0213	0.0208
CN (units of CN)				
Average	71.004	71.003	70.75	70.68
Maximum	100	100	100	100
Minimum	55	55	55	55
Standard deviation	9.13	9.08	8.97	8.84
Slope (%)				
Average	34.13	30.153	27.67	26.38
Maximum	567	224	152	99
Minimum	0	0	0	0
Standard deviation	33.75	23.45	18.97	16.44

prediction for peak flow at the watershed outlet. AGNPS output in the 10- x 10-meter center, 30- x 30-meter, and 60- x 60-meter resolutions overestimated the actual sediment yield recorded at the watershed outlet for the validated storm event.

For the study area watershed, cell size resolution of 30 x 30 meters seems appropriate based on the accurate AGNPS model prediction for peak flow when validated with the field-monitored storm. The steep slopes created in the 10- x 10-meter resolution data set may lead to an overestimation of sediment output, rendering data at this resolution unreliable. At this time, the 10- x 10-meter resolution is both impractical and infeasible for use with the AGNPS model.

AGNPS output at the highest resolution does not provide a better approximation of sediment yield at the watershed outlet. AGNPS output for sediment generated within each cell in the watershed at the highest resolution does not accurately simulate the watershed processes. AGNPS output generated from the more detailed soil survey data is not significantly different from data generated by the STATSGO digital soils database.

This study raises the following questions:

- What level of detail is both practical and acceptable for policy-making and decision-making?
- What constitutes a cost-effective analysis?

Ultimately, these questions are best answered on a case-by-case basis and should be determined based on the size of the study area and on how the results of the analysis will be used (i.e., to make a direct land use decision or for broader planning). For broad planning analyses on large watersheds, the benefit of digitizing the soil survey data is outweighed by the cost in time and effort to generate this detailed database. STATSGO data may be sufficient. If a direct land management decision is being made for a small area such as a farm within a watershed, however, the analysis should use the most detailed soils data.

Recommendations for Future Work

The original intent of this study was to use the capabilities of AGNPS Version 4.03 to evaluate a watershed using data generated at a high cell size resolution—10 x 10 meters. AGNPS h

ad never been used to evaluate data at such a high resolution. As discovered during this project, the newest version of AGNPS is not, at this time, capable of handling a data set that has more than 32,767 cells (19). Once this limitation with the AGNPS model is remedied, the entire 10- x 10-meter resolution data set should be routed through the model so that definite conclusions regarding the applicability of such a detailed data set can be made.

References

1. U.S. Department of Agriculture (USDA). 1991. Riparian forest buffers: Function and design for protection and enhancement of water resources. N1-PR-07-91.
2. Ashraf, M.S., and D.K. Borah. 1992. Modeling pollutant transport in runoff and sediment. Trans. Amer. Soc. Agric. Eng. 35:1,789-1,797.
3. Barten, P., and K. Stave. 1993. Characterization of streamflow and sediment source areas for the Little Beaver Kill Watershed, Ulster County, NY (July).
4. Vieux, B., and S. Needham. 1993. Nonpoint-pollution model sensitivity to grid cell size. J. Water Resour. Planning and Mgmt. 119(2).
5. Kumar, V. 1993. Geographic information system application for nonpoint source pollution management. Logan, UT: Utah State University.
6. Yoon, J., L. Padmanabhan, and L. Woodbury. 1993. Linking Agricultural Nonpoint Source Pollution Model (AGNPS) to a geographic information system. Proceedings of the AWRA Conference on Geographic Information Systems and Water Resources (March).
7. Haddock, G., and P. Jankowski. 1993. Integrating nonpoint source pollution modeling with a geographic information system. Department of Geography, University of Idaho.

8. Mitchell, J., B. Engel, R. Srinivasan, and S. Wang. 1993. Validation of AGNPS for small watersheds using an integrated AGNPS/GIS system. Proceedings of the AWRA Conference on Geographic Information Systems and Water Resources (March).

9. He, C., J. Riggs, and Y. Kang. 1993. Integration of geographic information systems and a computer model to evaluate impacts of agricultural runoff on water quality. Proceedings of the AWRA Conference on Geographic Information Systems and Water Resources (March).

10. Freezor, D., M. Hirschi, and B. Lesikar. 1989. Effect of cell size on AGNPS prediction. Paper No. #89-2184. ASAE/CSAE Meeting, St. Joseph, MI.

11. Young, R., C. Onstad, D. Bosch, and W. Anderson. 1989. AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds. J. Soil and Water Conserv. 44(2).

12. U.S. Department of Agriculture (USDA). 1987. Agricultural Research Service, AGNPS: Agricultural nonpoint-source pollution model, a watershed analysis tool. Conservation Research Report No. 35.

13. Star, J., and J. Estes. 1990. Geographic information systems: An introduction. Englewood Cliffs, NJ: Prentice Hall.

14. ESRI. 1987. PC ARC/INFO user's guides. Redlands, CA: Environmental Systems Research Institute, Inc.

15. Eastman, R. 1993. IDRISI technical reference. Worcester, MA: Clark University.

16. Clarke, K.C. 1990. Analytical and computer cartography. Englewood Cliffs, NJ: Prentice Hall.

17. Clesceri, L., A. Greenberg, and R. Trussell. 1989. Standard methods for examination of water and wastewater, 17th ed. pp. 2-75 to 2-76.

18. Rosenfield, G.H., and K. Fitzpatrick-Lins. 1986. A coefficient of agreement as a measure of thematic classification accuracy. Photogrammetric Engin. and Remote Sensing 52(2).

19. U.S. Department of Agriculture, Soil Conservation Service. 1995. AGNPS Newsletter.

XGRCWP, a Knowledge- and GIS-Based System for Selection, Evaluation, and Design of Water Quality Control Practices in Agricultural Watersheds

Runxuan Zhao, Michael A. Foster, Paul D. Robillard, and David W. Lehning
Penn State University, University Park, Pennsylvania

Abstract

The Expert GIS Rural Clean Water Program (XGRCWP) integrates a geographic information system (GIS), a relational database, simulation models, and hypertext mark language documents to form an advisory system that selects, evaluates, sites, and designs nonpoint source pollution control systems in agricultural watersheds. Its major features include:

- Customized GIS functions to obtain spatial and attribute data and feed them to a rule-based expert system for selecting feasible control practices.
- A user interface for examining the field-specific conditions and recommended control practices on the screen by clicking on the displayed field boundary map.
- A direct linkage between the GIS spatial data and the relational attribute data, which allows users to examine data on the screen interactively.
- A graphic user interface to GIS functions, which enables users to perform routine watershed analyses.
- Linkage to hypertext reference modules viewable by Mosaic Internet document browser.
- Dynamic access to other models such as the Agricultural Nonpoint Source Simulation Model.

The software environment of XGRCWP is GRASS 4.1 and X-Windows on SUN OS 4.3.1. Its major functions have been tested for the Sycamore Creek watershed in Ingham County, Michigan.

Introduction

In 1981, the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Agriculture (USDA) initiated the Rural Clean Water Program (RCWP) in 21 agricultural watersheds. This program represents the most intensive water quality monitoring and implementation and evaluation of nutrient, sediment, and pesticide reduction practices ever undertaken in the United States (1). More than a decade of research efforts has resulted in a wealth of experiences and lessons on selection, siting, and evaluation of nonpoint source control practices.

The storehouse of knowledge gained from RCWP is of little use, however, unless it is properly integrated and packaged in an easily accessible form. Technology transfer of this knowledge is therefore critically important. To integrate and synthesize the lessons learned from RCWP, Penn State University initiated an RCWP expert project. The hypertext-based version of the RCWP expert system, completed in 1993, can select and evaluate nonpoint source control systems at a single site. Although the hypertext-based version is still suitable for users who do not have access to geographic information systems (GIS) data, it is inadequate for the comprehensive selection and evaluation of control systems on a watershed basis. It does not provide the user the spatial reference of a site and requires the user's subjective judgment for the model input.

The Expert GIS Rural Clean Water Program (XGRCWP) is the UNIX and X-Window version of the RCWP expert system, which integrates GIS and the RCWP expert system to provide decision support at multiple spatial scales from single fields to subwatersheds to the watershed scale. This paper presents the major features of XGRCWP, including design of the expert system, interface to GIS functions, and linkages to a relational database and simulation models.

Overview

XGRCWP comprises five major components (see Figure 1):

- An expert system for recommending control practices based on site-specific information.

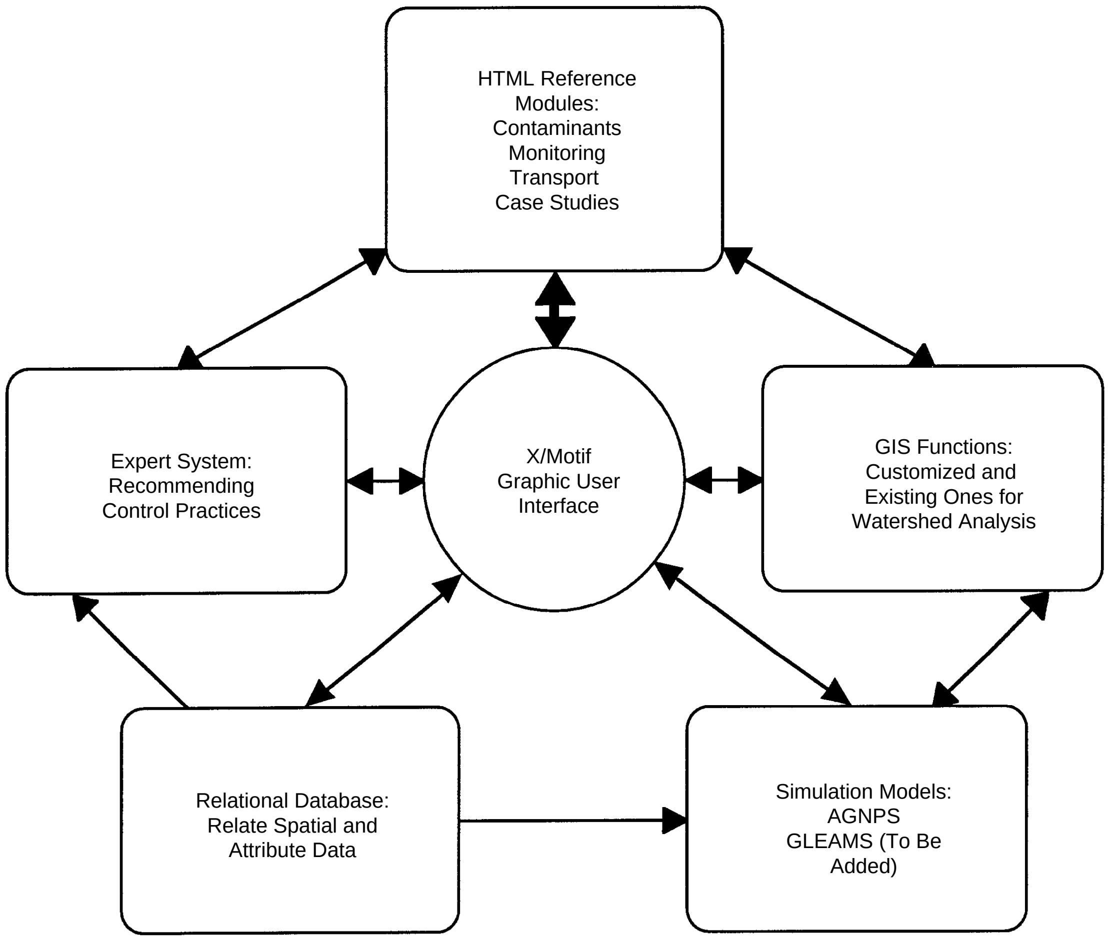


Figure 1. Major components of XGRCWP and their relationships.

- Custom and existing GIS functions for watershed analysis and estimation of contaminant loading potential.

- Linkage to fields, soils, and land use databases.

- Linkage to the Agricultural Nonpoint Source Simulation Model (AGNPS) (2).

- Hypertext mark language (HTML) reference modules.

The X/Motif graphic user interface (GUI) integrates the five components and allows the user to navigate flexibly among them. The components are also internally connected in different ways. For example, the expert system can use the customized GIS functions to retrieve site-specific information from Geographical Resource Analysis Support Systems (GRASS) (3) data layers and INFORMIX relational database tables. In addition, the expert recommendations of control practices can be displayed and examined using GRASS functions. Finally, the GIS functions can help generate input to the AGNPS model, and its output can be converted to GIS format for additional analyses.

Design of the Expert System

The objective of the expert system is to recommend feasible control systems (i.e., complementary sets of control practices to reduce nonpoint source pollution based on site-specific conditions). One distinct feature of this system is the combination of two modes of data acquisition: direct user input and GIS functions. XGRCWP also has two modes for deriving the expert recommendations: batch or interactive. This section discusses these aspects of the expert system as well as its knowledge base.

Rules for Control Practice Selection

The knowledge base of the expert system includes the following six site-specific characteristics:

- Contaminant of interest and its adsorption characteristic.
- Potential level of contaminant loading (low, medium, or high).
- Potential level of contaminant leaching (low, medium, or high).
- Soil hydrologic group (A, B, C, or D).

- Time of year (during or outside the growing season).
- Type of land use (cropland, animal waste, or critical area).

The user first chooses a contaminant of interest from a list consisting of four kinds of pesticides (strongly, moderately, or weakly adsorbed, and nonadsorbed) and eight other contaminants (ammonia, bacteria, sediment, total nitrogen, total phosphorus, nitrate, orthophosphorus, and viruses). The values of other characteristics, some of which vary with the contaminant specified, can then be input either directly by the user or by custom GRASS functions as discussed in the section of this paper on data acquisition.

The RCWP used 14 general categories of control practices (see Table 1). Many suitable conditions were established for each general category. For example, conservation tillage is recommended to reduce runoff for cropland under conditions otherwise favoring loss through sediment transport, such as a contaminant strongly adsorbed to the soil (e.g., total phosphorus), the nongrowing season, and soils with a relatively high runoff potential (soil group C or D) (see Figure 2). Each general category includes several specific control practices. When a general practice category is recommended, the user must decide which specific practice within that general category to evaluate further by consulting the nonpoint source database (NPSDB) for the reported research data about this practice or by running the AGNPS simulation model.

Data Acquisition

The expert system recommends one or more control systems based on site-specific conditions that are either directly input by the user or calculated by the customized GRASS functions. The user always specifies the contaminant of interest and the season, while a GRASS function (R.HYDRO-GRP) always determines the soil hydrologic group of each field. For the other factors (loading potential, leaching potential, and application class), however, the user has two alternative ways to decide input values. For example, after the user selects a contaminant of interest, the program displays the contaminant loading potential window (see Figure 3). The potential level of the selected contaminant can be indicated if the user knows it. Otherwise, the user can let the GRASS functions derive loading potential from existing field data.

Table 1. The Best Management Practices Used in the Rural Clean Water Program

Source control practices	Nutrient Management (NUTR) Pesticide Management (PEST)
Structural control practices	Animal waste systems (AWS) Diversion systems (DIV) Sediment retention and water control (SED) Terrace systems (TERR) Waterway systems (WATW)
Vegetative control practices	Conservation tillage (CT) Critical area treatment (CAT) Cropland protection systems (CPS) Grazing land protection (GLP) Permanent vegetative cover (PVC) Stream protection (SP) Stripcropping (SCR)

The direct input option can also be used to help the user address “what-if” questions. When the user selects the GIS functions to determine the loading potential, XGRCWP makes a series of calls to appropriate customized GRASS functions according to the current contaminant of interest. For example, if the contaminant is total nitrogen, the functions R.MANURE, R.FERT, and R.B.CONCENTRATION are called to estimate total nitrogen from manure, fertilizer, and soil base concentration, respectively. Another GRASS function, R.NP.LOADING, is then called to translate the quantitative measure of loading potential into the qualitative classification (low, medium, or high) as input to the expert systems. These GRASS functions generate the inputs by searching and converting the data from INFORMIX relational data tables that are associated with the GRASS spatially referenced data, such as field boundary and soil map. Table 2 lists the customized GRASS functions developed for data acquisition.

Control System Recommendation

XGRCWP derives the expert recommendations for control systems in two ways: in a batch mode for every field in a watershed and in an interactive mode for a user-specified field.

In batch mode, an existing GRASS function, R.INFER, is used to create a raster data layer for each general practice category of control practice according to a rule-set prepared for that general category. For example, the contents and formats for the conservation tillage practice are documented in Table 3. The raster data layer for representing the conservation tillage recommendations (CT.rec) is generated by running R.INFER with the appropriate rule. The category value of CT.rec is 1 at each point in the data layer where the conservation tillage is recommended, or 0 otherwise. The R.INFER function is similarly called for other general practice categories. Additional GRASS functions can then display or further analyze the resulting map layers. The batch mode provides the user the overall picture with a watershed-wide view of feasible control systems.

In the interactive mode, the field boundary map is displayed and the user can specify any field of interest by clicking the mouse on it. The recommendations and the site-specific conditions of the field are displayed on the right half of the screen. The recommended control practices are also displayed in a popup window for further

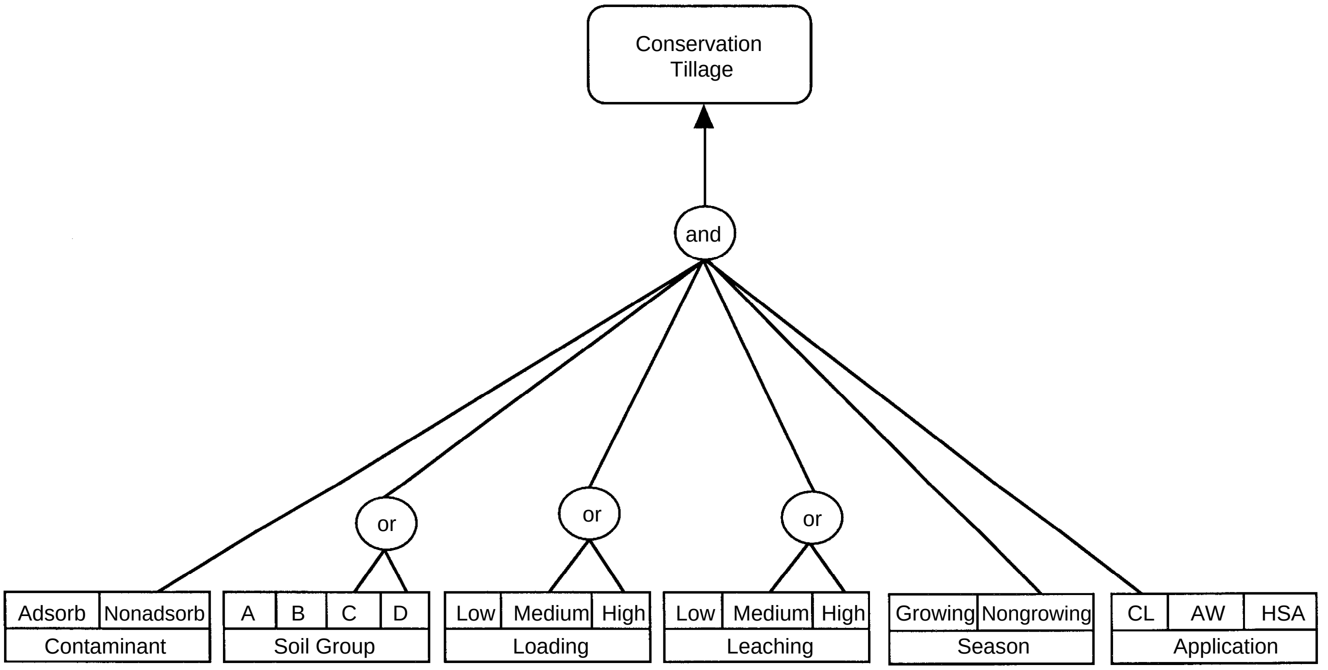


Figure 2. Dependency network (AND-OR diagram) for site-specific recommendation of conservation tillage.

examination, such as the specific practices within each general category, the feasible control systems for nonpoint source pollution control, and research data on the practices. The interactive mode is implemented through the integration of a Bourne shell script, structured query language (SQL) commands, a customized GRASS function (R.RCWP.EXPERT), and GRASS display functions with the Motif GUI. Interactive mode is intended for detailed consideration of a specific farm.

Interface to GIS Functions

XGRCWP provides a GUI to most of the customized GRASS functions and some of GRASS's existing functions (see Figure 4). This interface shields the user from complex syntax so the user can focus on the subject matter. The GUI makes it easier for the user to perform routine operations such as estimation of contaminant loading, identification of critical areas, erosion and runoff calculation, and other watershed analysis tasks. It also helps the user make full, effective use of all custom and some existing GRASS functions.

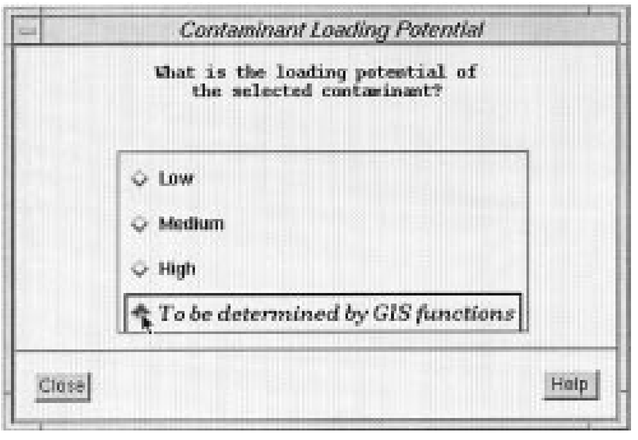


Figure 3. The popup window for the potential level of contaminant loading.

Linkages to Database and Other Models

Data Structure

The GRASS functions used to generate inputs for the expert system use the same soils and fields relational databases as the Water Quality Model/GRASS Interface under development by the Soil Conservation Service (SCS) (4). XGRCWP and our custom GRASS functions were tested for the Sycamore Creek watershed, Ingham County, Michigan. In this data structure, spatial data (e.g., field boundaries, watershed boundaries, soils map unit boundaries, and elevation data) are saved as GRASS raster data layers while attribute data (e.g., crop information, fertilizing schedule, soil information) are stored in INFORMIX relational database tables. Each field or soil map unit is assigned a unique identification (ID) number. The field attribute (INFORMIX) data also contain this ID number. The linkage between the GRASS raster map and the INFORMIX data is accomplished with a GRASS category label (see Figure 5).

Linkage to Database

To allow the interactive examination of field data from GRASS raster layers and the associated relational da-

Table 2. **Summary of the Customized GRASS Functions Developed by Nonpoint Source Agricultural Engineering Research Group at Penn State University To Generate Inputs for the RCWP Expert System**

Name	Descriptions
R.FERT	Produces raster maps of total nitrogen or total phosphorous from the scheduled fertilizer applications for different crops by dynamically retrieving information from a GRASS data layer and INFORMIX data tables
R.MANURE	Calculates the total manure on each farm according to animal numbers and types (e.g., dairy cow, beef cow, horses, swine), allocates manure to the fields on a farm by a user specified strategy (uniformly spreading or inverse distance weighted distributing method), and finally estimates nitrogen and phosphorous loading from manure application rate, conversion factor, percentages of transportation losses, and volatile losses
R.B.CONCENTRATION	Estimates nitrogen and phosphorous concentration in parts per million within different types of soils according to the organic matter contents
R.NP.LOADING	Classifies the loading potential of nitrogen or phosphorous into three categories (low, medium, and high) based on the actual loading from fertilizer and manure and the N or P concentration in soils
R.EROSION	Obtains a relative measure of soil erosion severity by dividing the amount of erosion by the tolerance values of the soils and then reclassifying them into three categories (low, medium, and high)
R.LEACHING.P	Estimates leaching index from soil hydrologic group and annual and seasonal precipitation and classifies it into three categories (low, medium, and high)
R.HYDRO-GRP	Retrieves soil hydrologic group from the INFORMIX database and reclassifies the soil map into soil hydrologic groups

tabase tables, XGRCWP calls our custom function, D.WHAT.FIELD.SH, a Bourne shell script that dynamically links GRASS raster layers and the INFORMIX database tables. When the user clicks on a field, for example, this function extracts field-specific information from INFORMIX tables such as field information, fertilization schedule, crop operation schedule, and soil information. The D.WHAT.FIELD.SH function then displays all related soils and fields information for the given field. It also marks the field boundary map to remind the user which fields have already been examined.

Linkage to Reference Modules

At any stage of the selection, evaluation, siting, and design procedure for control practices, the user can consult reference modules that provide information, guidance, and data about contaminant properties, transport variables, and examples of applications from RCWP projects. Four reference modules are available in the Macintosh version of RCWP expert system: contaminants, monitoring, transport, and case studies. We are currently converting these reference modules into Mosaic-viewable HTML documents so that they can be accessed from XGRCWP. Mosaic is a public domain, Internet-aware document browser that is available for X-Windows, Macintosh, and Microsoft Windows.

All four modules use graphics to demonstrate design procedures and contaminant control processes. The contaminant module provides information about 11 categories of contaminants cited in RCWP projects and their impacts on surface and ground-water resources. The monitoring module describes different aspects of water quality sampling and analysis systems. The transport module describes contaminant pathways in surface and ground water. The case studies module presents detailed examples from key RCWP projects. These examples cover both practice selection and implementation aspects of control systems. The reference modules serve as a complementary component of XGRCWP.

Linkage to AGNPS

AGNPS is a distributed-parameter, storm event-based model that estimates runoff, sedimentation, and nutrient loss in surface runoff within agricultural watersheds (2). The prototype version of the Water Quality Model/GRASS Interface developed by SCS conveniently generates an AGNPS input file for all cells in a watershed from the spatial and relational soils and fields databases. The UNIX version of AGNPS can then use this input file. XGRCWP can call AGNPS directly from its X-Window interface and convert standard AGNPS model outputs for all cells in the watershed into GRASS raster format for display and analysis.

Discussion

The literature on software systems for managing nonpoint source pollution in agricultural watersheds is diverse and rapidly growing. With few exceptions (5-7), these decision support systems are purely model-based, GIS-based (8), or hybrid systems with models running within a GIS framework (9-14). The addition of expert system components can overcome some of the difficulties in primarily model-based systems:

- Overly intensive input data requirements.
- Inability to handle missing or incomplete data.
- Requirements that all inputs be numerically expressed.

Table 3. The Rule File for Recommending Conservation Tillage

IFNOTMAP app.class 3	!application class is not high-source area
ANDIFMAP contam.feature 2	!contaminant is nonadsorbance
ANDIFMAP leaching.p 1	!leaching potential is low
ANDIFMAP soil.g 1 2	!soil groups are A or D
ANDIFMAP contam.load 1	!contaminant loading is low
ANDIFMAP season 2	!nongrowing season
THENMAPHYP 1 yes, CT	
is recommended	
!	
IFNOTMAP app.class 3	!application class is not high-source area
ANDIFMAP contam.feature 2	!contaminant is nonadsorbance
ANDNOTMAP leaching.p 3	!leaching potential is not high
ANDNOTMAP soil.g 1	!soil groups are not A
ANDNOTMAP contam.load 3	!contaminant loading is not high
ANDIFMAP season 1	!growing season
THENMAPHYP 1 yes, CT	
is recommended	
!	
IFNOTMAP app.class 3	!application class is not high-source area
ANDIFMAP contam.feature 1	!contaminant is strong adsorbance
ANDNOTMAP contam.load 3	!contaminant loading is not high
ANDIFMAP season 1	!growing season
THENMAPHYP 1 yes, CT	
is recommended	
!	
IFNOTMAP app.class 3	!application class is not high-source area
ANDIFMAP contam.feature 1	!contaminant is strong adsorbance
ANDNOTMAP leaching.p 3	!leaching potential is not high
ANDIFMAP soil.g 3 4	!soil groups are C or D
ANDNOTMAP contam.load 3	!contaminant loading is not high
ANDIFMAP season 2	!nongrowing season
THENMAPHYP 1 yes, CT	
is recommended	

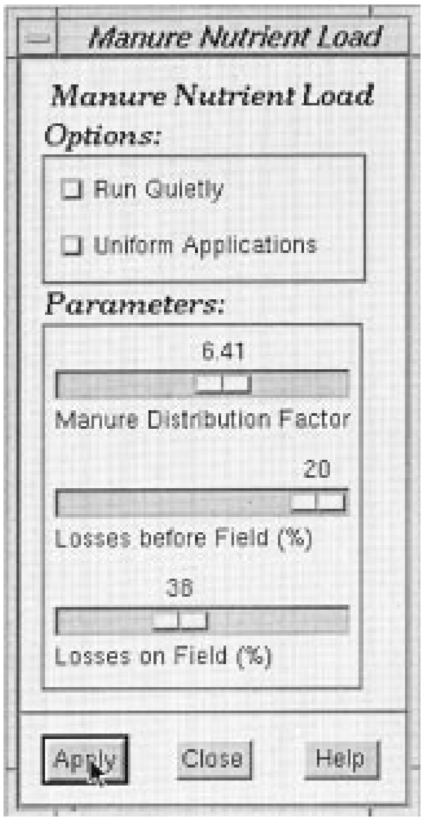


Figure 4. The GUI to the R.MANURE function.

- High degree of expertise needed to structure model input and explain model output relative to the user's problem context.

The expert system component of XGRCWP also reduces the number of model runs needed for decision support through preliminary, rule-based screening of control systems at each site of interest in the watershed.

Conclusions

XGRCWP incorporates several kinds of expertise for the user's benefit:

- Subject matter expertise in siting and selecting nonpoint source control systems in agricultural watersheds.
- Expertise in configuring AGNPS model input from the soils and fields databases.
- Expertise in interpreting, explaining, and visualizing expert system and model input.

The integration of the expert system and the GRASS GIS makes input to the expert system easier and more objective. It enhances the expert system's capability for recommending effective control practices at the field level to achieve watershed contaminant loading objectives. XGRCWP is designed as an open structured

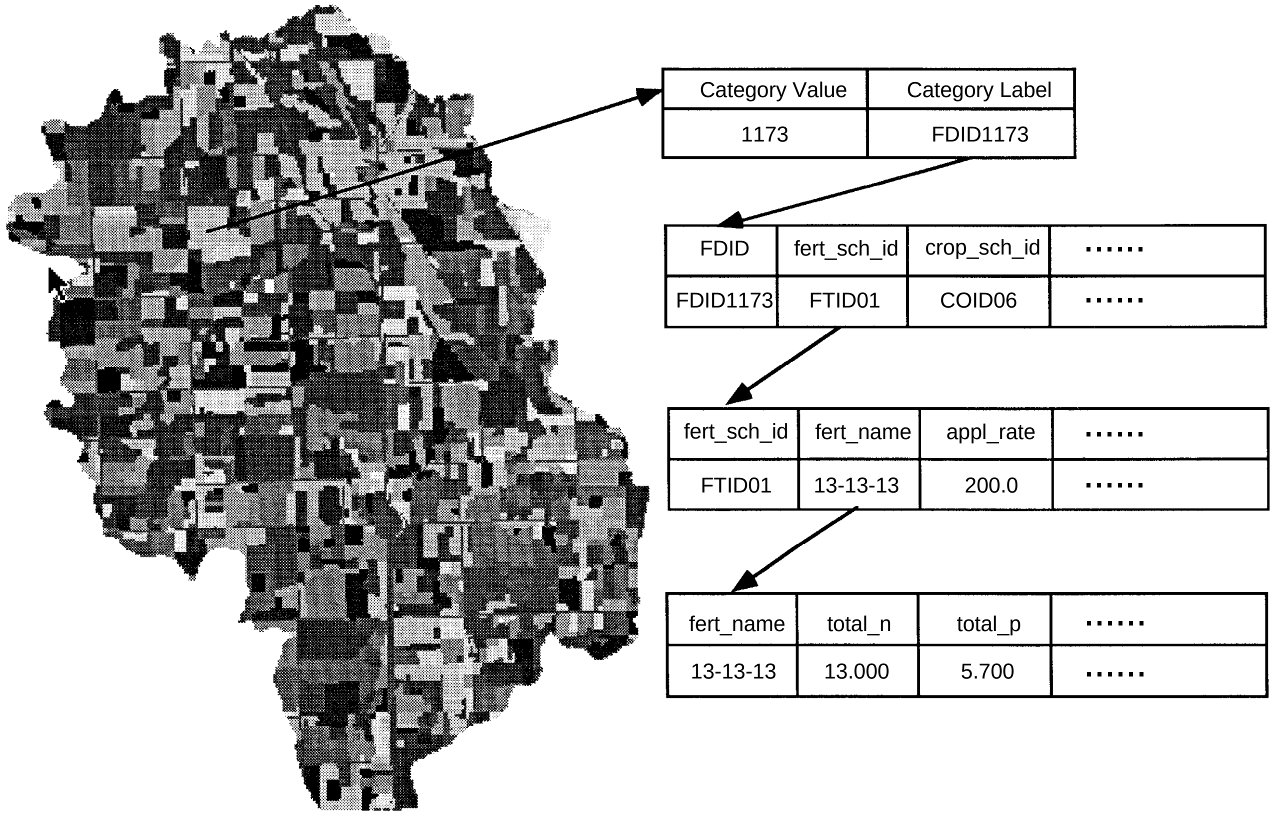


Figure 5. Data structure of Sycamore Creek watershed, Ingham County, Michigan.

program and has great potential to be improved easily and continually according to users' feedback. Ongoing efforts to enhance the program include:

- Developing more rules that incorporate topographical factors such as slope and slope length for the expert system so that more site-specific control practices can be recommended.
- Adding dynamic hypertext-based help and reference to the program.
- Establishing intelligent linkages among the expert system, the GIS functions, and other simulation or design models for nonpoint source control practices.

Acknowledgments

We thank Vicki Anderson, Ruth Shaffer, and Brent Stinson of the Michigan USDA SCS and John Suppnick of the Michigan Department of Natural Resources for their assistance in working with the underlying databases in the Sycamore Creek watershed. This project is supported by U.S. EPA Grant No. X-818240.

References

1. U.S. EPA. 1992. Extending the RCWP knowledge base to future nonpoint source control projects. In: Proceedings of the National RCWP Symposium. EPA/625/R-92/006. Washington, DC.
2. Young, R.A., C.A. Onstad, D.D. Bosch, and W.P. Anderson. 1989. AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds. J. Soil and Water Conserv. 44(2):168-173.
3. USACERL. 1993. GRASS 4.1 users reference manual. Champaign, IL: U.S. Army Corps of Engineers Construction Engineering Research Laboratory.
4. USDA. 1993. Users' guide to water quality model/GRASS interface. Fort Collins, CO: U.S. Department of Agriculture, Soil Conservation Service, Technology Information Systems Division.
5. Barnwell, T.O., L.C. Brown, and W. Marek. 1989. Application of expert systems technology in water quality modeling. Wat. Sci. Tech. 21:1,045-1,056.
6. Ford, D.A., A.P. Kruzic, and R.L. Doneker. 1993. Using GLEAMS to Evaluate the Agricultural Waste Application Rule-Based Decision Support (AWARDS) computer program. Wat. Sci. Tech. 28(3-5):625-634.
7. Yakowitz, D.S., J.J. Stone, L.J. Lane, P. Heilman, J. Masterson, J. Abolt, and B. Imam. 1993. A decision support system for evaluating the effects of alternative farm management systems on water and economics. Wat. Sci. Tech. 28(3-5):47-54.

8. Hamlett, J.M., D.A. Miller, R.L. Day, G.A. Peterson, G.M. Baumer, and J. Russo. 1992. Statewide GIS-based ranking of watersheds for agricultural pollution prevention. J. Soil Water Cons. 47(3):399-404.

9. Srinivasan, R., and B.A. Engel. 1994. A spatial decision support system for assessing agricultural nonpoint source pollution. Water Resources Bull. 30(3):441-452.

10. Srinivasan, R., and J.G. Arnold. 1994. Integration of a basin-scale water quality model with GIS. Water Resources Bull. 30(3):453-462.

11. Engel, B.A., R. Srinivasan, J. Arnold, and S.J. Brown. 1993. Nonpoint source (NPS) pollution modeling using models integrated with geographic information systems (GIS). Wat. Sci. Tech. 28(3-5):625-690.

12. He, C., J.F. Riggs, and Y. Kang. 1993. Integration of geographic information systems and a computer model to evaluate impacts of agricultural runoff on water quality. Water Resources Bull. 29(6):891-900.

13. Kiker, G.A., G.M. Cambell, and J. Zhang. 1992. CREAMS-WT linked with GIS to simulate phosphorus loading. ASAE Paper No. 92-9016. St. Joseph, MI: American Society of Agricultural Engineers.

14. Tim, U.S., S. Mostaghimi, and V.O. Shanholtz. 1992. Identification of critical nonpoint pollution source areas using geographic information systems and water quality modeling. Water Resources Bull. 28:877-887.

Integration of EPA Mainframe Graphics and GIS in a UNIX Workstation Environment To Solve Environmental Problems

William B. Samuels
Science Applications International Corporation, McLean, Virginia

Phillip Taylor
Tetra Tech, Inc., Fairfax, Virginia

Paul Evenhouse
Martin Marietta, Inc., Research Triangle Park, North Carolina

Robert King
Assessment and Watershed Protection Division, U.S. Environmental Protection Agency, Washington, DC

Abstract

The Assessment and Watershed Protection Division of the Office of Wetlands, Oceans, and Watersheds has developed water quality analysis software on the U.S. Environmental Protection Agency (EPA) mainframe computer. This software integrates national on-line environmental databases and produces maps, tables, graphics, and reports that display information such as water quality trends, discharge monitoring reports, permit limits, and design flow analyses.

In the past, this graphic software was available only to users connected to the mainframe with IBM graphics terminals or PCs with graphics emulation software. Recently, software has been developed that can be used to: 1) access the EPA mainframe from a UNIX workstation via the Internet, 2) execute the Water Quality Analysis System (WQAS) procedures, 3) display WQAS graphics in an X-Window on the workstation, and 4) download data in a geographic information system (GIS) format from the mainframe. At the same time, this workstation can execute ARC/INFO and ARC/VIEW applications in other X-Windows. This capability allows analysts to have the power of GIS, the mainframe databases (e.g., Permits Compliance System [PCS], STORET, Reach File, Industrial Facilities Discharge File, Daily Flow File, Toxic Chemical Release Inventory), and the retrieval/analysis/display software (Environmental Data Display Manager, Mapping and Data Display Manager, Reach Pollutant Assessment [RPA], PCS-STORET Interface, UNIRAS) available to them on one desktop. This capability extends the tool set available to GIS analysts for environmental problem-solving.

This paper discusses application of these tools and databases to several problems, including EPA's watershed-based approach to permitting, and the RPA, an automated method to identify priority pollutants in watersheds.

Introduction

The purpose of this paper is to explore new geographic information systems (GIS) data integration tools that are applicable to a wide range of environmental problems, including the U.S. Environmental Protection Agency's (EPA's) watershed-based approach to permitting and the Reach Pollutant Assessment (RPA), an automated method to identify priority pollutants in watersheds. The ultimate goal is to make these tools and databases accessible to a wide range of users.

Understanding aquatic resource-based water quality management depends on access to and integration of diverse information from many sources. To date, the techniques to perform this integration, and thus yield meaningful analyses supporting environmental decision-making, are neither fully developed nor documented. New tools and information resources are now available, but not used to their full potential, for more valuable water quality and watershed analyses. EPA headquarters is responsible for ensuring that integrated data management tools are available for water quality analyses and data reporting as well as making national data systems more useful. EPA will accomplish this by upgrading and crosslinking systems, developing interactive data retrieval and analysis mechanisms, and providing easy downloading of data to client workstations.

The Assessment and Watershed Protection Division (AWPD) of the Office of Wetlands, Oceans, and Watersheds (OWOW) has developed water quality analysis software on the EPA mainframe computer (1). This software integrates national on-line environmental databases and produces maps, tables, graphics, and reports that display information such as water quality trends, discharge monitoring reports, permit limits, and design flow analyses. In the past, this graphic software was available only to users connected to the mainframe with IBM graphics terminals or PCs with graphics emulation software. Recently, software has been developed that can be used to:

- Access the EPA mainframe from a UNIX workstation via the Internet.
- Execute the Water Quality Analysis System (WQAS) procedures.
- Display WQAS graphics in an X-Window on the workstation.
- Download data in a GIS format from the mainframe.

At the same time, this workstation can execute ARC/INFO and ARC/VIEW applications in other X-Windows. This capability allows analysts to have the power of GIS, the mainframe databases (e.g., Permits Compliance System [PCS], STORET, Reach File, Industrial Facilities Discharge File, Daily Flow File, Toxic Chemical Release Inventory), and the retrieval/analysis/display software (Environmental Data Display Manager, Mapping and Data Display Manager, RPA, PCS-STORET Interface, UNIRAS) available to them on one desktop. This extends the tool set available to GIS analysts for environmental problem-solving. This paper discusses how these tools and databases have been applied to two examples: 1) a watershed-based approach to permitting and 2) the RPA, an automated procedure for identifying watersheds with priority pollutants.

Mainframe Databases and Tools

The EPA IBM ES9000 mainframe computer, located in Research Triangle Park, North Carolina, contains a large volume of digital water quality and environmental data available on-line through a number of data retrieval and display tools (see Figure 1). Other documents describe these databases and tools in detail (2, 3).

This effort focused on showing how these databases and tools can complement GIS activities. In some cases, data can be directly downloaded to a workstation in GIS format. An example of this is accessing EPA's Reach File (Version 1) from GRIDS (4). In other cases, databases are accessed by mainframe tools, the data are processed, and a GIS data set is produced that can be downloaded to a workstation. An example of this case is the RPA (RPA3) tool that integrates data from the Reach File, STORET, and PCS to identify priority pollutants in watersheds (5).

The mainframe can be accessed through several paths: Internet, PC dialup, or dedicated line into a terminal controller (see Figure 2). In the applications presented here, the Internet connectivity is emphasized because this is the mechanism that makes these databases and tools available to GIS analysts at their workstations. Figure 3 shows the hardware and software requirements for Internet access to the water quality data integration tools. The basic components are a UNIX workstation with an X-Window manager, the X3270 software, and Internet connectivity. The X3270 software is required to emulate an IBM 3270 full-screen terminal. This software is publicly available through EPA's National Computer Center User Support. In addition, an account on the mainframe computer is required. Once this account is established, an additional software module, GDDMXD, is required to map IBM host-based graphics to the workstation's X-Window. The GDDMXD software resides on the mainframe and is loaded when the user logs in. Once the hardware and software are set up, a single UNIX workstation can provide access to mainframe and workstation tools and databases on one desktop (see Figure 4).

Applications

To illustrate how these mainframe and workstation tools/databases can work together to solve environmental problems, we present two applications. The first shows a watershed-based approach to permitting; the second describes the RPA.

Watershed-Based Approach to Permitting

The watershed approach is a process to synchronize water quality monitoring, inspections, and permitting to support water quality protection activities on a geographic basis. It is a coordinated and integrated method to link science, permits, and other pollution control and prevention activities to meet state water quality standards. Numerous local, state, and federal agencies have recognized watershed approaches as the best way to manage natural resources effectively and efficiently. Establishing a schedule for data collection, permit issuance, and other elements of this approach affords the opportunity to coordinate and integrate other natural resource management efforts to make better use of

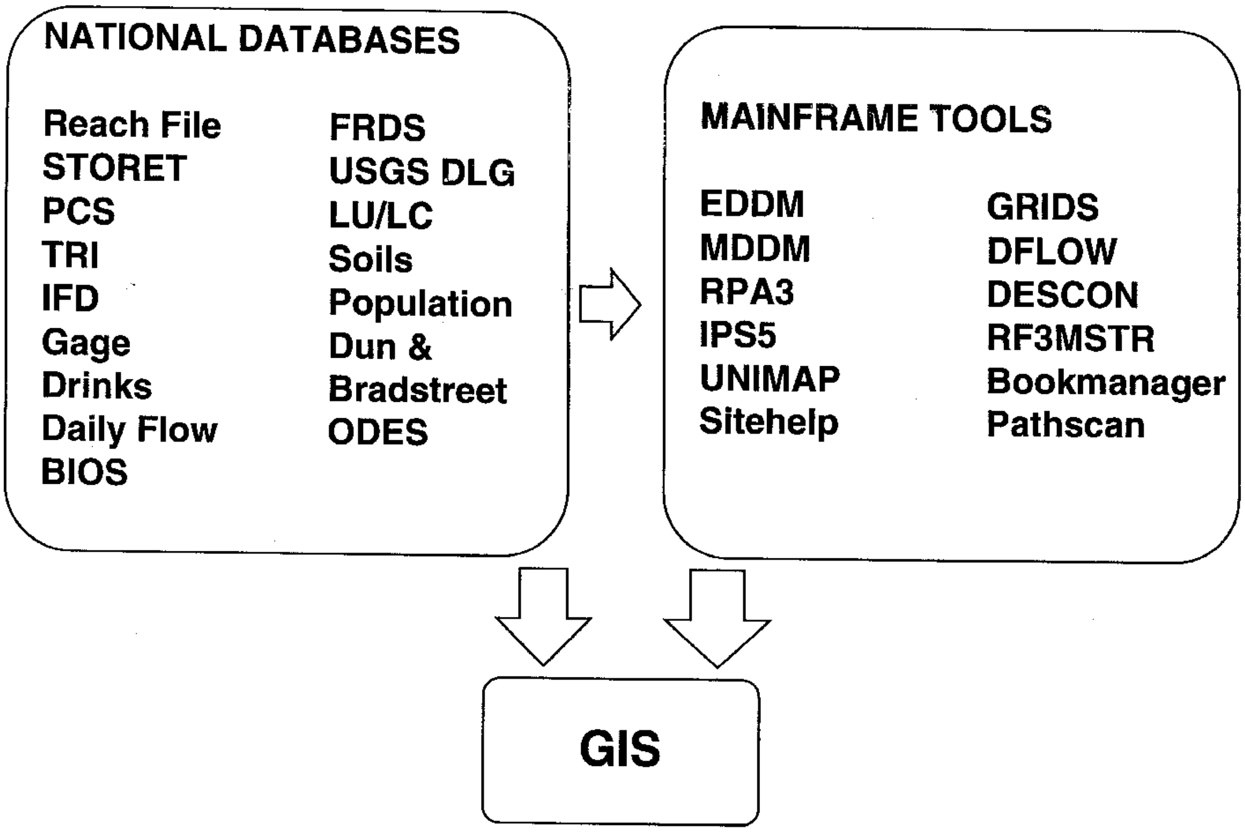


Figure 1. EPA mainframe databases/tools and linkage to GIS.

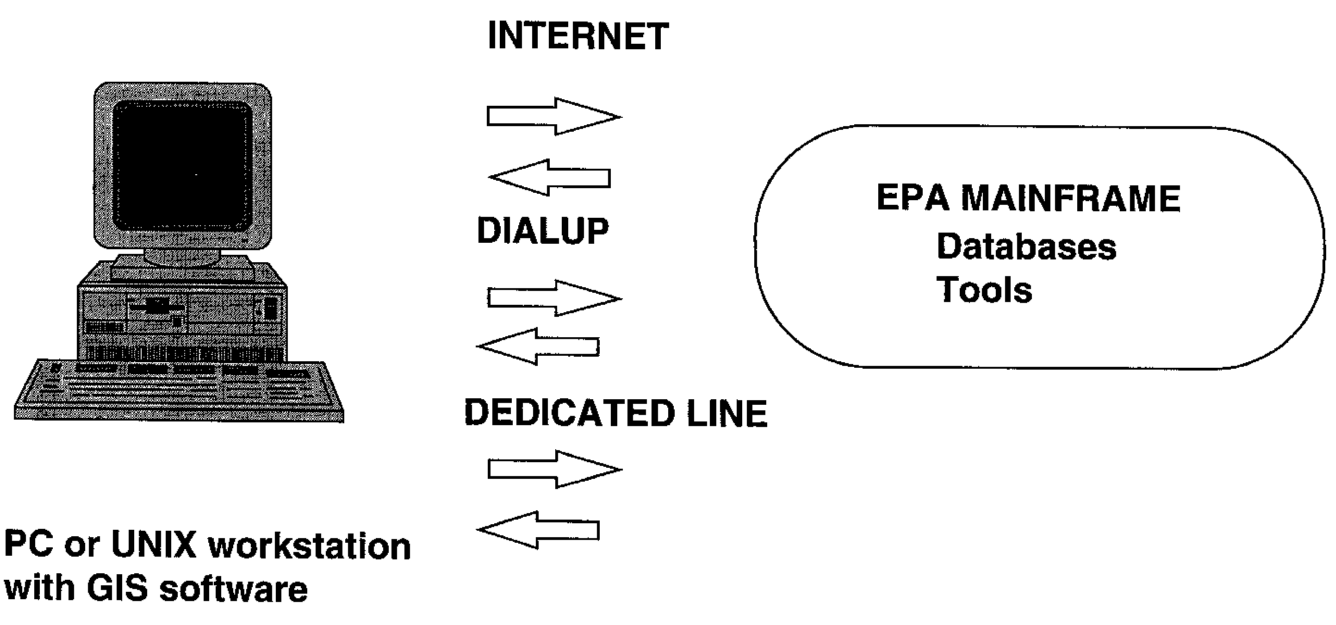


Figure 2. Access to the EPA mainframe databases and tools.

LOCAL WORKSTATION

HARDWARE

UNIX workstation (with X-Windows) on Internet (i.e., DG Aviion)

SOFTWARE

X3270 Software - creates an X-Window, which emulates an IBM 3270 full-screen session (provided by EPA)

EPA MAINFRAME ACCOUNT

GDDMXD software provided on the mainframe to map GDDM graphics into X-Windows

Figure 3. Hardware and software requirements for accessing water quality data integration tools via INTERNET.

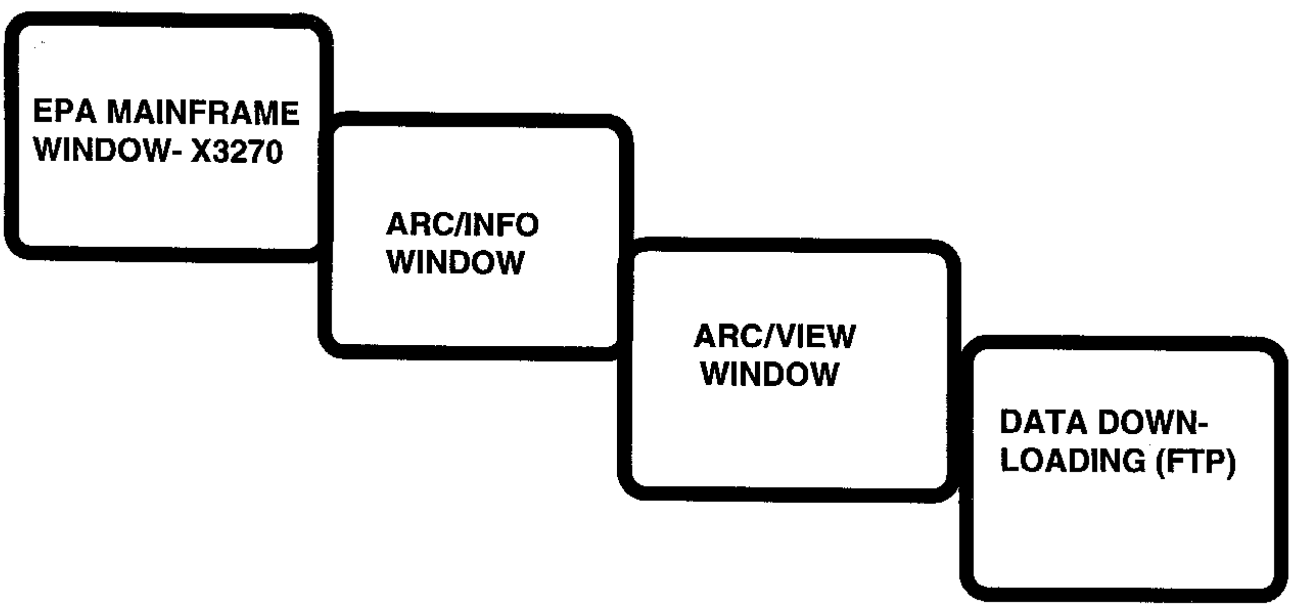


Figure 4. A UNIX workstation environment provides access to mainframe and workstation tools and databases on one desktop.

limited local, state, and federal financial and human resources (6).

This application illustrates how the watershed approach used GIS and EPA mainframe databases and tools. As an example, a four-step approach (see Figure 5) has been developed and applied to an impaired watershed (Saluda River basin) in South Carolina. Steps one and two identified watersheds of concern through their nonattainment of designated uses (see Figure 6) and highlighted the cause of nonattainment, in this case pathogens (see Figure 7). The data sets used were U.S. Geological Survey (USGS) hydrologic unit boundaries, Soil Conservation Service (SCS) watershed boundaries, and data from the EPA waterbody system, which were indexed to the SCS watersheds.[1] These data sets were integrated into an ARC/INFO arc macro language (AML) to allow users to pose queries and prioritize watersheds for further investigation.

Once priorities were set, the third step was to evaluate, in detail, the sources and causes of nonattainment. The Saluda River basin, which had pathogens as its cause of nonattainment, was selected for further analyses. In this step, the mainframe tools supplement the workstation GIS capabilities illustrated so far. A STORET retrieval was performed for ambient water quality stations monitoring for fecal coliforms. The STORET stations were partitioned into three categories (low, medium, and high) according to the state fecal coliform standard (7), which reads:

> not to exceed a geometric mean of 200/100 mL, based on five consecutive samples during any 30 day period; nor shall more than 10% of the total samples during any 30 day period exceed 400/100 mL

The categories in Figure 8 correspond to the standard as follows:

low: $<$ 200/100 milliliters

200/100 milliliters $\leq$ medium $<$ 400/100 milliliters

high: $\geq$ 400/100 milliliters

Figure 8 illustrates the use of ARC/VIEW to visualize the location of fecal coliform "hot spots" in the Saluda River basin. Figure 9 focuses on one SCS watershed (03050109-040) where pathogens cause nonattainment. The locations of industrial and municipal dischargers are plotted, and facilities with fecal coliform limits and their respective permit expiration dates (captured from the PCS) are shown in a table included with the figure. The GIS capabilities used to generate Figure 9

[1] Clifford, J. 1994. Personal communication with Jack Clifford, U.S. EPA, Washington, DC.

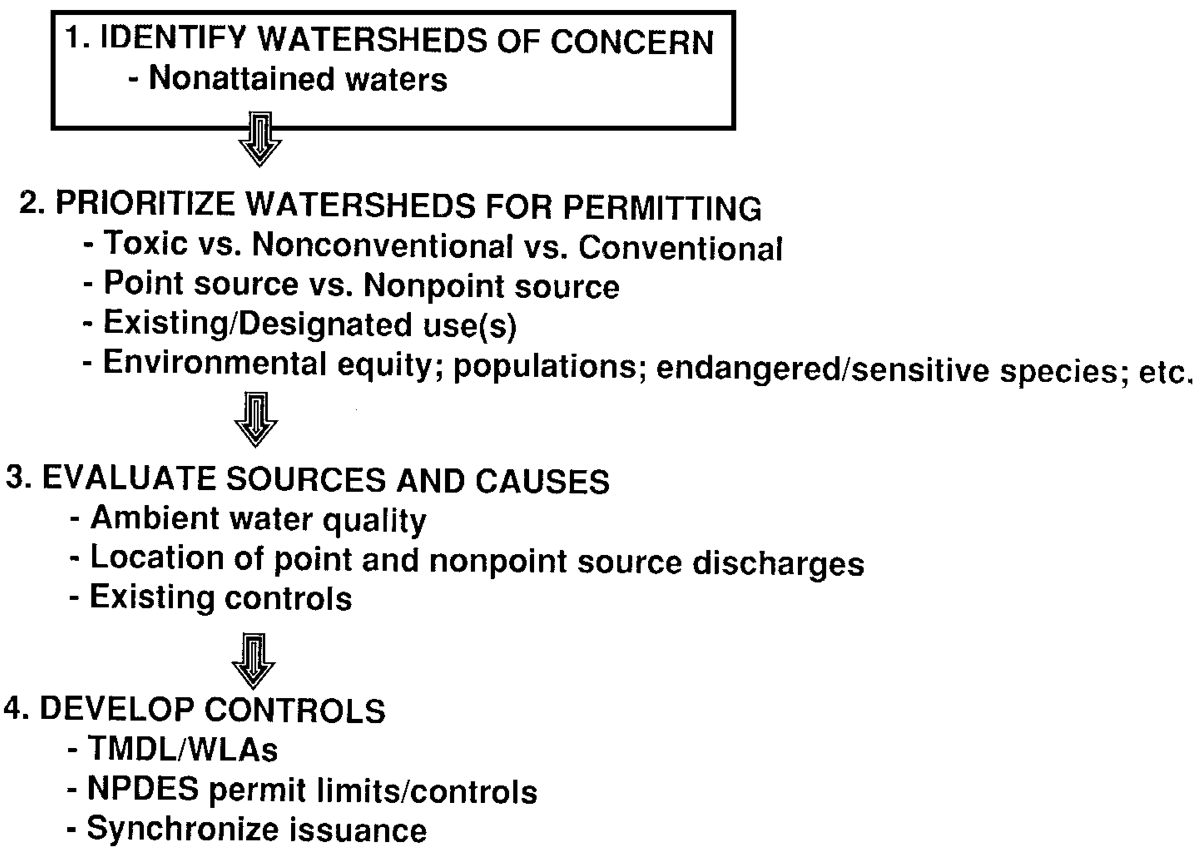


Figure 5. Four steps illustrating an example approach to permitting on a watershed basis.

Step 1: Identify Watersheds of Concern

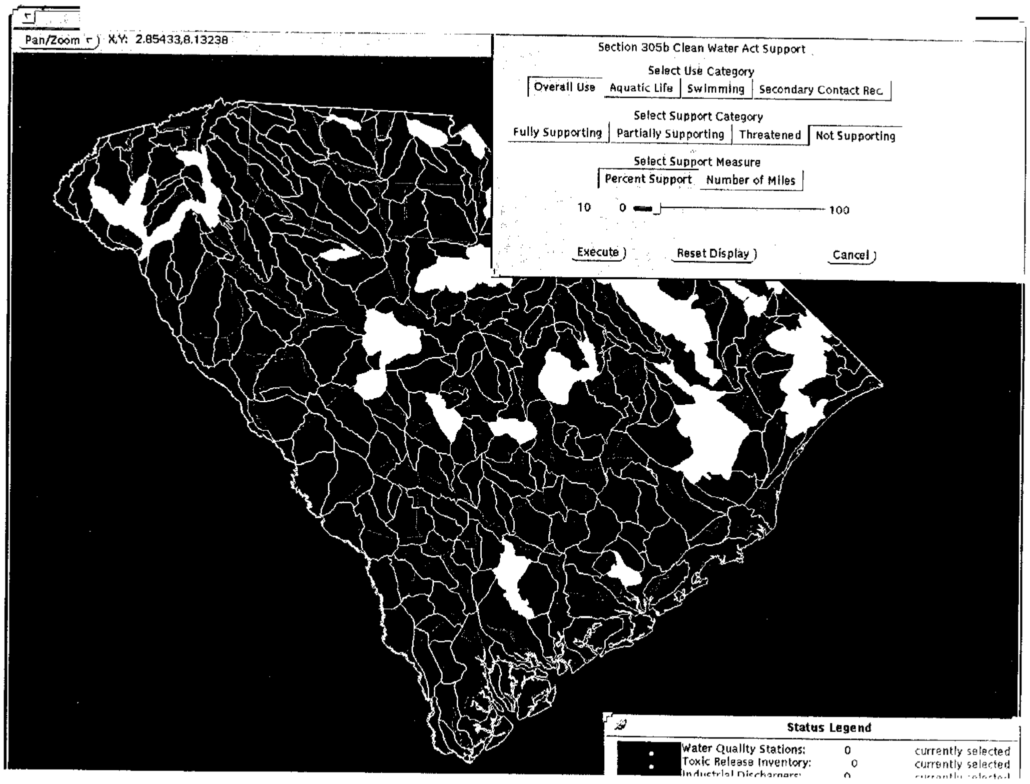


Figure 6. **Identification of watersheds of concern—watersheds where at least 10 percent of the reaches are not fully supporting overall designated use (data and ARC/INFO AMLs provided by Jack Clifford).**[2]

Step 2: Evaluate Priorities

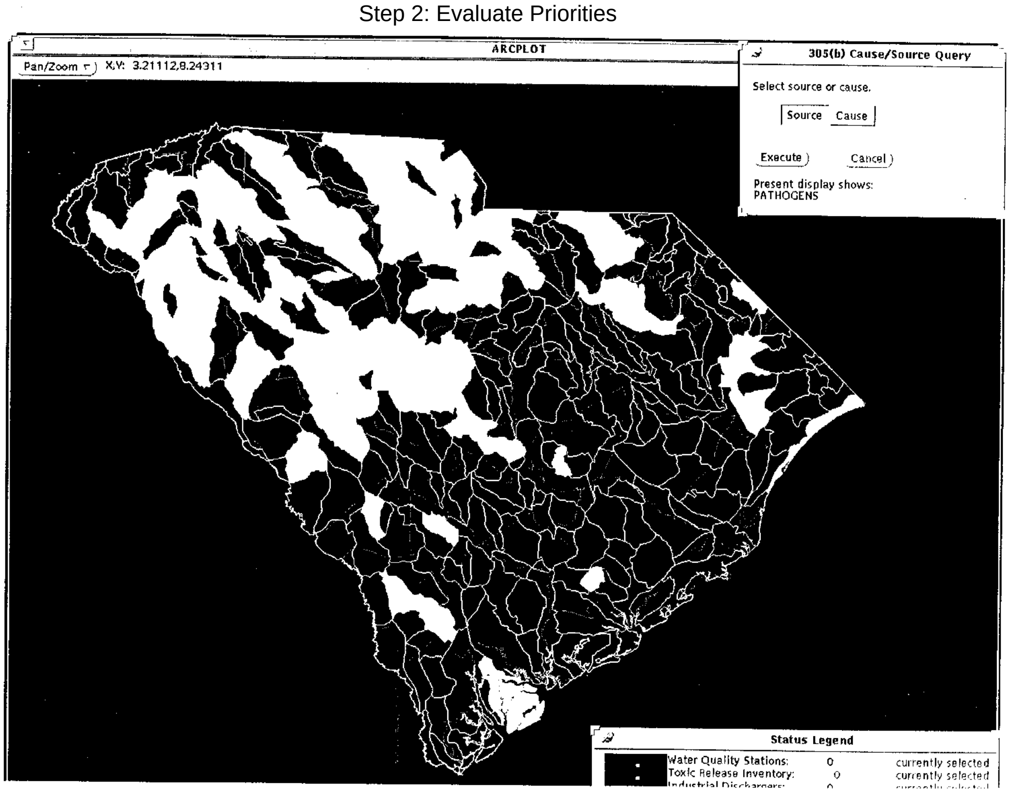


Figure 7. **Watersheds where the cause of nonattainment is pathogens (data and ARC/INFO AMLs provided by Jack Clifford).**[3]

[2] See note 1.
[3] See note 1.

Step 3: Evaluate Ambient Water Quality—Low, Medium, and High Levels of Fecal Coliforms

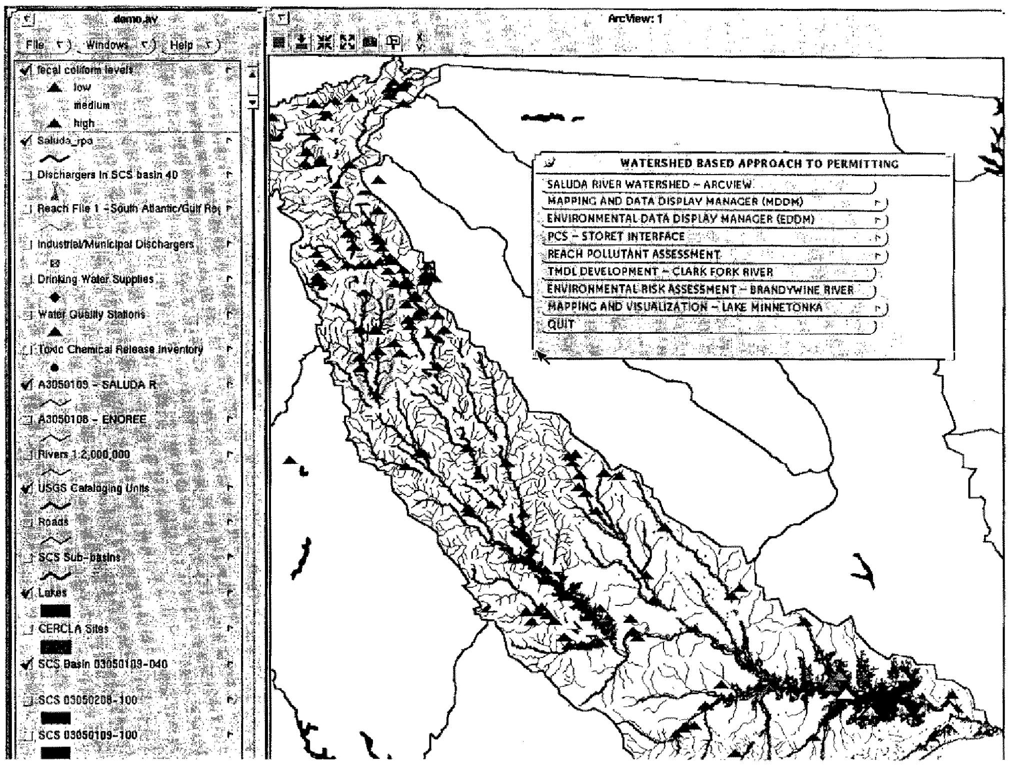


Figure 8. Map of the Saluda River basin showing the location of STORET monitoring stations and fecal coliform levels.

Step 3: Existing Controls—Dischargers That Have Limits for Fecal Coliforms

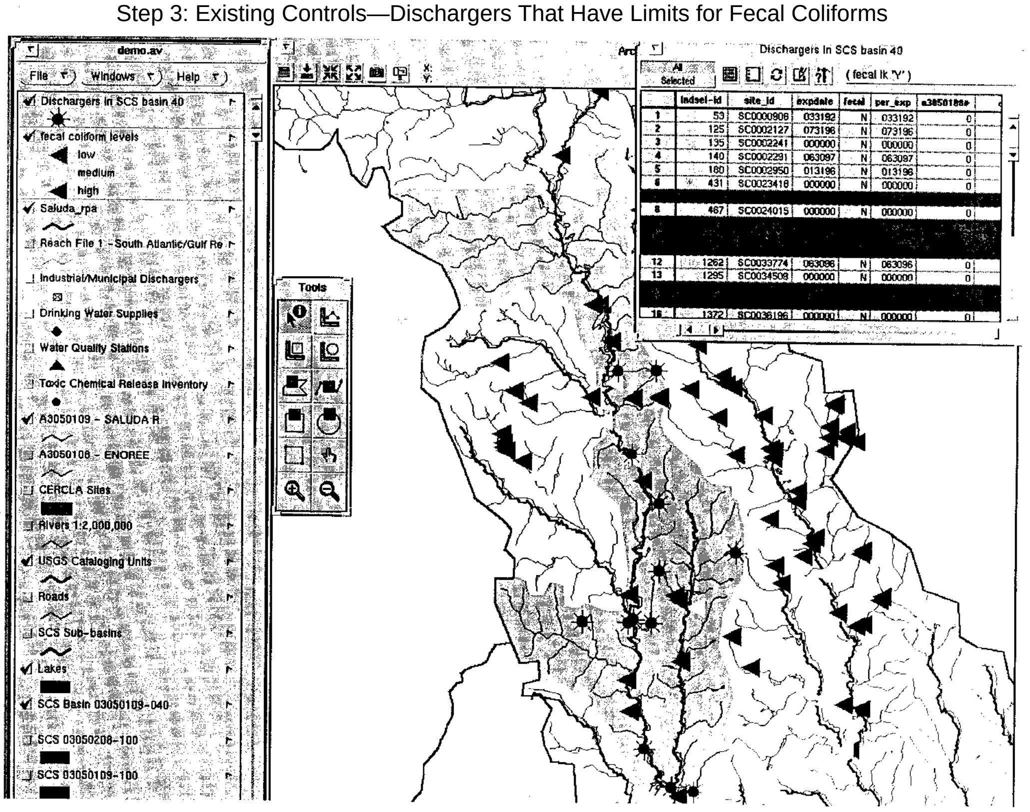


Figure 9. Focus on SCS watershed 03050109-040 (shaded in gray). Locations of STORET stations, municipal and industrial dischargers are also shown. The table in the upper right highlights dischargers with fecal coliform limits and their respective permit expiration dates.

show that permit issuance is not synchronous, which is a key element in the watershed approach.

Within this local workstation GIS environment, the attributes associated with the STORET monitoring stations and the PCS dischargers are limited with respect to the large amount of time series sampling data that exists in these databases. Figures 10 through 12 illustrate how an EPA mainframe procedure, the Environmental Data Display Manager (EDDM), can be accessed from an X-Window on the GIS workstation to query the entire STORET and PCS databases and thus provide additional data analysis and display capabilities to the GIS workstation. In Figures 10 and 11, a water quality inventory was performed for a STORET station, and a time series plot of fecal coliform levels is displayed. In Figure 12, the limits and discharge monitoring report (DMR) data were accessed from PCS for a sewage treatment plant. Excursions beyond the PCS limits for fecal coliforms are easily visualized in the plot.

The fourth and final step in the watershed approach was to develop controls for achieving water quality standards. This might include the development of total maximum daily loads (TMDL), waste load allocations (WLA), and the synchronization of permit issuance. Another mainframe tool, the PCS-STORET INTERFACE (referred to as IPS5 on the mainframe), can be used to access and compute design flows for TMDL development (see Figure 13) and to find all facilities discharging to a particular reach (see Figure 14), an initial step in the synchronization of permit issuance.

RPA

The RPA is a procedure on the EPA mainframe that automates identification of reaches where priority pollutants have been detected. It can be run for a user-selected state or USGS hydrologic unit.

Section 304(l) of the Clean Water Act (CWA) identifies water bodies impaired by the presence of toxic substances,

Step 3: Using EDDM To Evaluate Ambient Upstream Fecal Coliform Levels—Regulation of Upstream Dischargers Is Necessary

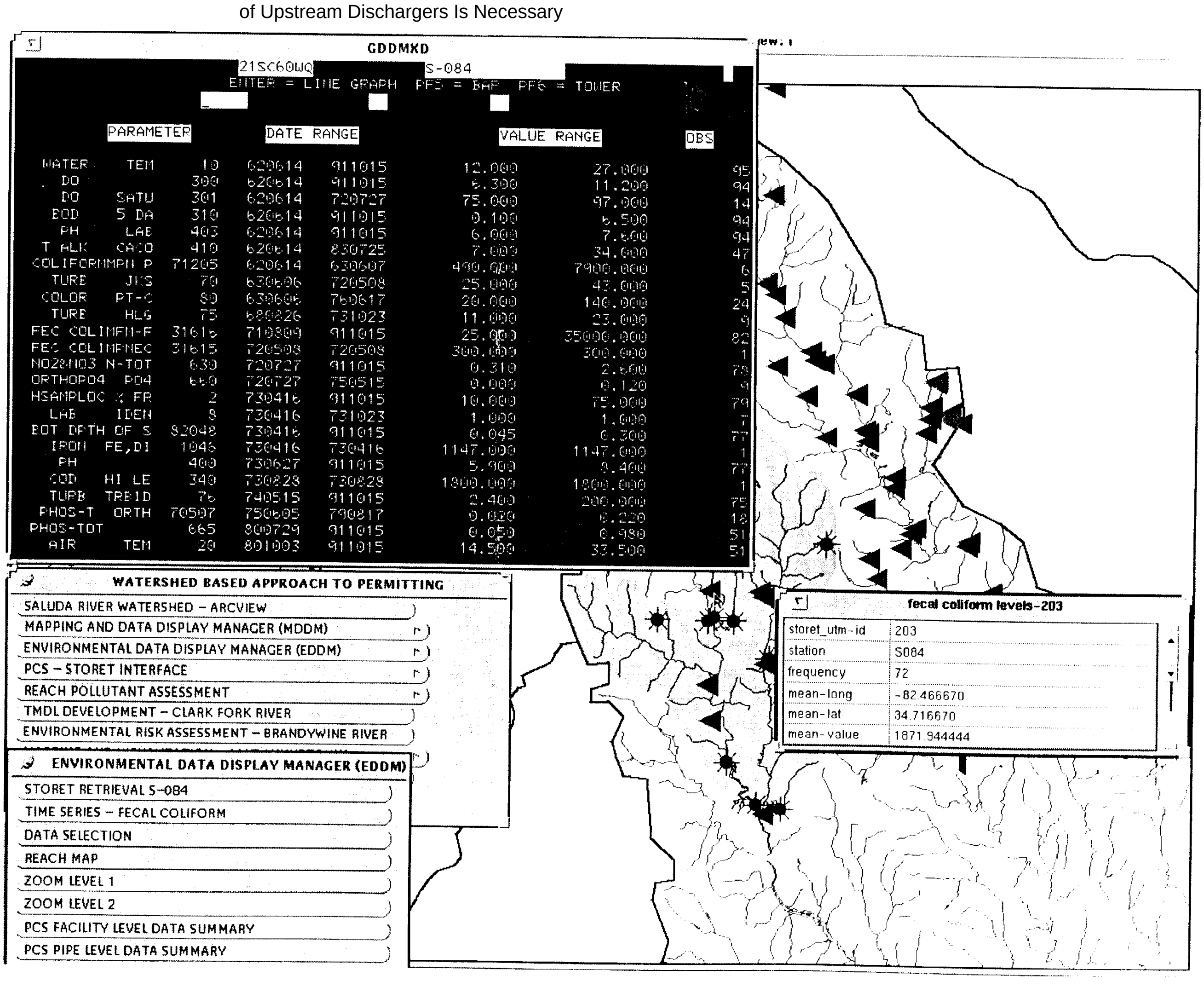


Figure 10. Using EDDM to perform a water quality inventory for STORET station 21SC06WQ S-084 in the Saluda River basin, South Carolina.

Step 3: Using EDDM To Visualize Trends in Ambient Coliform Levels Upstream of a Discharger

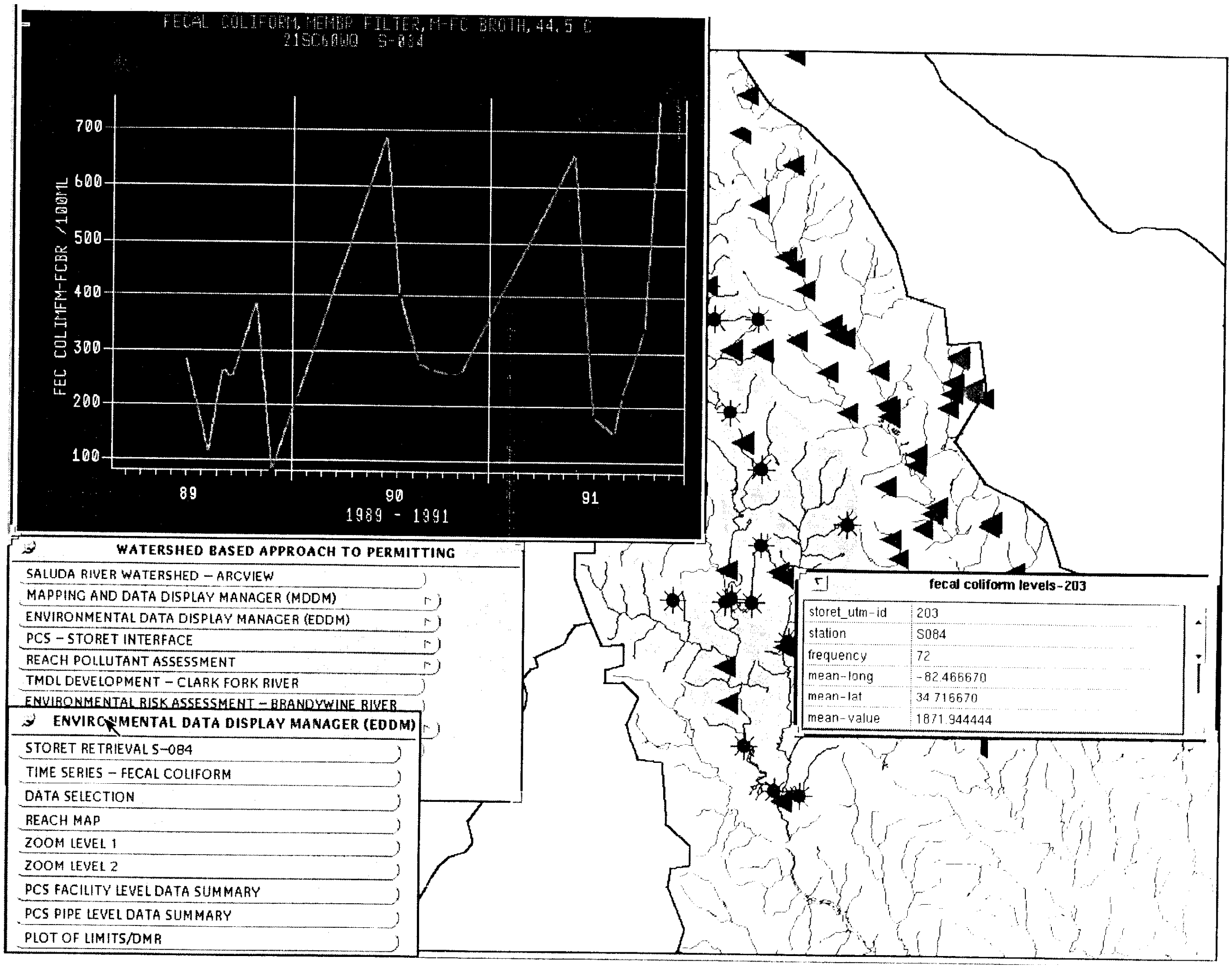


Figure 11. From the EDDM water quality inventory table, the fecal coliform parameter (31616) was selected for a time series plot.

identifying point source dischargers of these substances and developing individual control strategies for these dischargers. To meet these requirements, the EPA Office of Wetlands, Oceans, and Watersheds prepared guidance identifying criteria to be used in reviewing state reports.

The RPA was designed to address the requirements under criterion 7 of Section 304(l): identification of state waters with likely presence of priority toxic pollutants. This assessment was accomplished by identifying and summarizing reaches with point source dischargers of priority pollutants and water quality stations with priority pollutant data.

Information on the state's waters is summarized using the USGS hydrologic unit naming convention and the Reach Structure File (Version 1). Numerous databases were accessed and analyzed, including the Reach Structure and Reach Trace File (Version 1), industrial facilities discharge (IFD) file, STORET parameter file, PCS, and the STORET water quality file. The IFD file and PCS provided the facility information. Comparing information from both data sources identified active facilities and generated a complete list of facilities by their assigned reach numbers. Water quality data from STORET were summarized on reaches with priority pollutant monitoring data. Stations were retrieved with the following restrictions:

- Stations located within the state or hydrologic unit of interest.
- Ambient monitoring stations located on streams, lakes, or estuaries.
- Stations sampled for at least one priority pollutant in either water, sediment, or fish tissue on or after January 1, 1982.

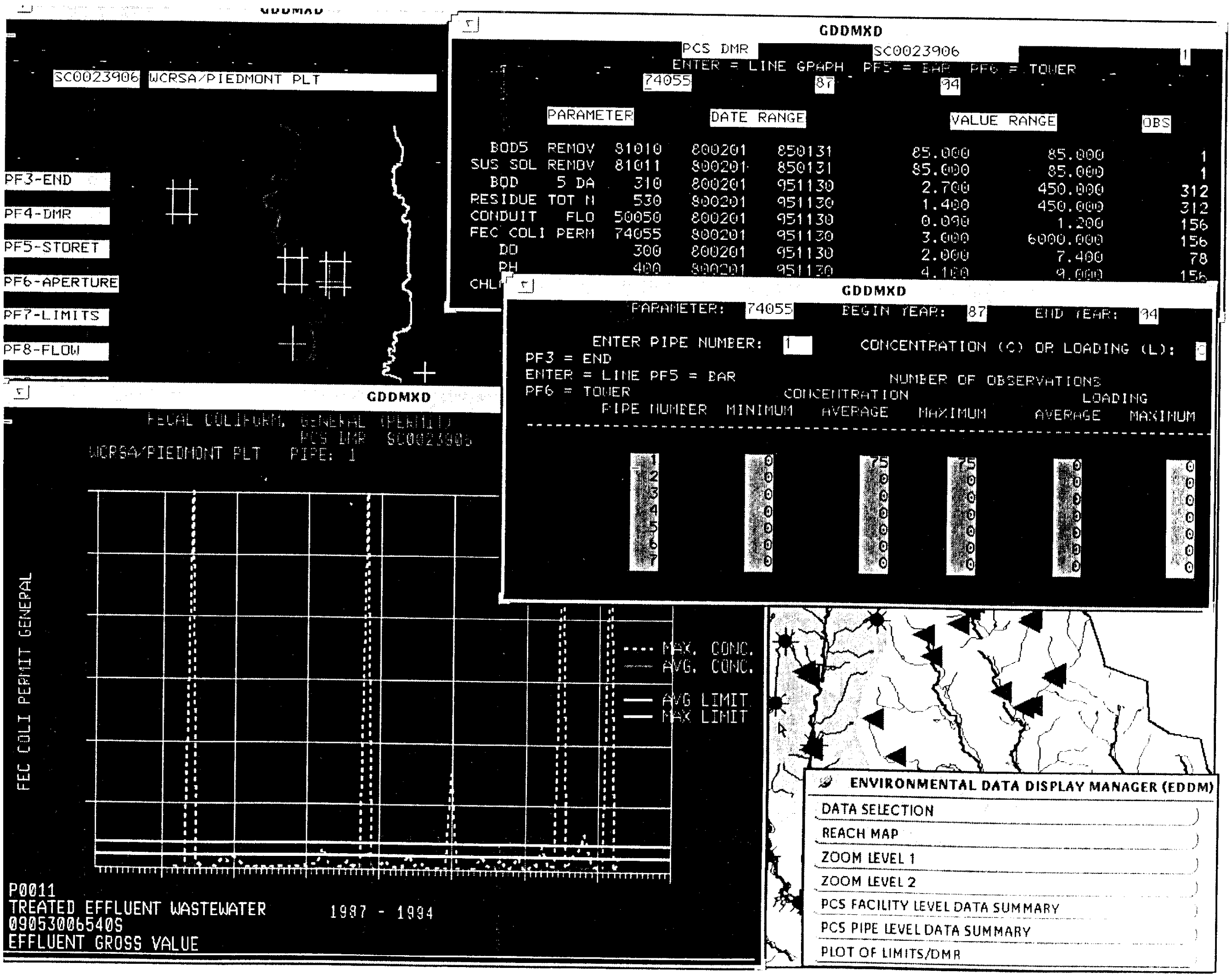


Figure 12. Using EDDM, PCS data is accessed for discharger SC0023906, Piedmont sewage treatment plant. The windows show plant location (upper left), facility and pipe summary data (upper right) and time series plot (lower left).

Recently, the RPA program was modified to output files compatible with GIS. An example of this is shown in Figure 15. The data in the table portion of this figure are written to two files as follows:

- For each reach in the hydrologic unit, the geographic coordinates are written to a file in ARC/INFO GENERATE format.

- The attributes associated with each reach (e.g., name, length, number of water quality stations) are written to a delimited ASCII file.

A third file is also automatically generated. This file is an AML that GENERATES the line coverage of reaches, defines and populates the INFO table of attributes, then joins the attributes in the INFO table to the line coverage of reaches. Once these three files are created, they are downloaded to the GIS workstation (via ftp) and processed by ARC/INFO. In ARC/VIEW, the user can identify a reach and determine:

- The number of water quality stations with priority pollutant monitoring data.

- The number and type of industrial facilities with priority pollutant discharge.

- The number of publicly owned treatment works (POTW) with and without indirect dischargers (see Figure 15).

In addition to this reach summary data, other tables (cross-linked to reaches, water quality stations, and dischargers) are produced that summarize the data by pollutant (see Figure 16). In this figure, each pollutant is cross-linked to the reach where it was detected, and the source of detection is also identified (i.e., water column, sediment, fish tissue, NPDES permit limit, Form 2(c) submittal) or

Step 4: Using the PCS/STORET INTERFACE for NPDES Permit Development—Analysis of Receiving Stream Flow Data

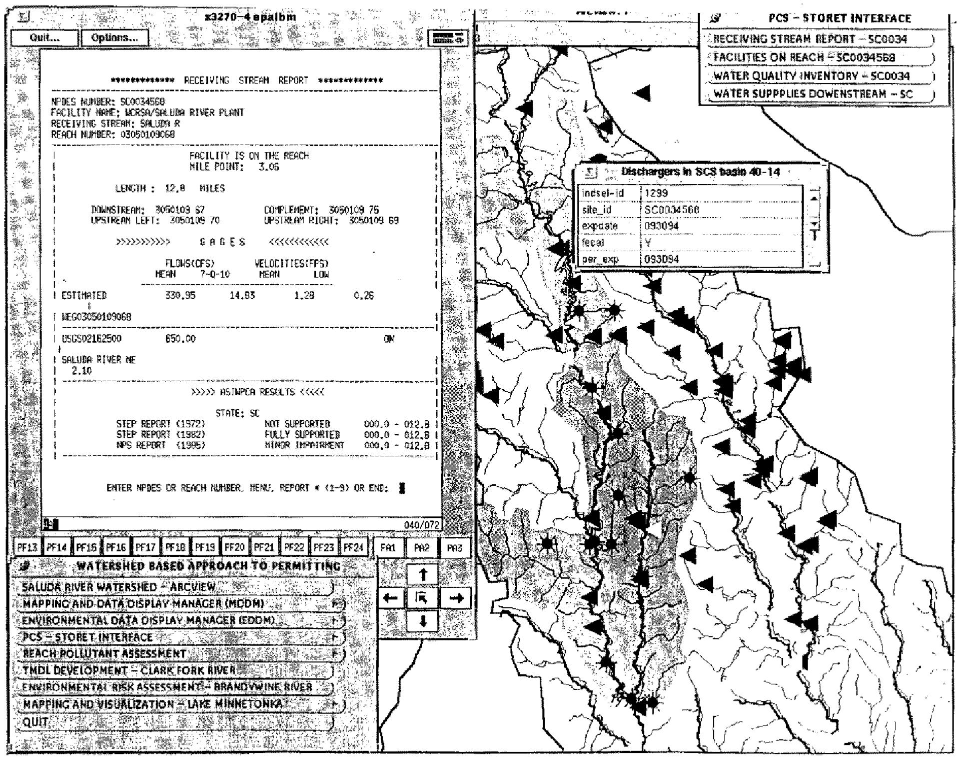


Figure 13. Using the PCS-STORET INTERFACE to access and compute design flows for a specific reach.

Step 4: Using the PCS/STORET INTERFACE To Determine Other Facilities on a Reach for Wasteload Allocation Purposes

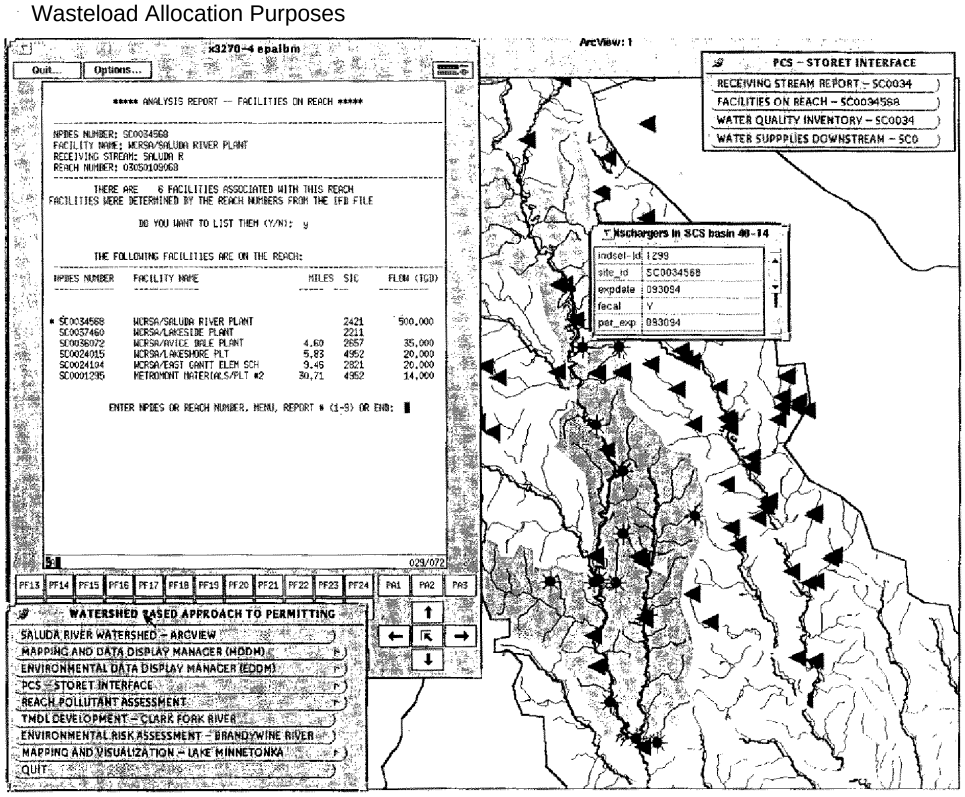


Figure 14. Using the PCS-STORET INTERFACE to list all facilities that discharge to a specific reach.

predicted to be in the discharger's effluent based on the standard industrial classification (SIC) code. More detailed information is also generated. For example, Figure 17 shows a detailed report for priority pollutants detected in the water column (similar reports are generated for sediment and fish tissue). Each pollutant is cross-linked to a reach and the specific monitoring station where it was detected. Basic summary statistics are also presented.

Figure 18 shows a detailed report for pollutants detected in the NPDES permit limit. In this figure, each pollutant is cross-linked to a reach and the specific NPDES discharger containing a permit limit. In addition, each discharger is identified as a major, minor, or POTW.

Figures 15 through 17 show the RPA output in the foreground and coverages displayed by ARC/VIEW in the background. Inspection of these figures shows that the Bush River is a priority pollutant reach containing seven industrial facilities, four of which discharge priority pollutants (one pulp and paper mill, three textile factories). Further examination of the data shows that cadmium has been detected in the water column and is also contained in the NPDES permit limit. It is also predicted to be in the discharge effluent based on SIC code. In the water column, cadmium was measured at 10 µg/L at two stations sampled in 1988.

Finally, there is a limit for cadmium in NPDES facility SC0024490 (Newberry plant), a POTW on the Bush River. The RPA output, linked to GIS, can be used as a screening and targeting tool for identifying specific reaches within watersheds where toxic priority pollutants cause water quality degradation.

Summary and Conclusions

The proliferation of GIS workstations, the expansion of the Internet, and the development of X-Window-based graphics emulation software (X3270 and GDDMXD) has afforded analysts the opportunity to use the powerful analytical capabilities of GIS and the EPA mainframe databases and tools together on one desktop. Thus, a user performing a GIS watershed analysis can also have immediate and complete access to national on-line databases such as STORET and PCS by opening up a "window" to the EPA mainframe. This allows detailed queries to be performed that supplement the data already being analyzed at the local workstation. This capability allows users to easily visualize additional data without having to spend effort in retrieval, downloading, transforming, and reformatting to make it useful. By enhancing existing mainframe programs to create output in GIS format, the time spent importing data to the GIS is reduced and more time can be spent on analysis. An example of this capability is the RPA program.

Reach Pollutant Assessment: Priority Pollutant Detection in the Bush River

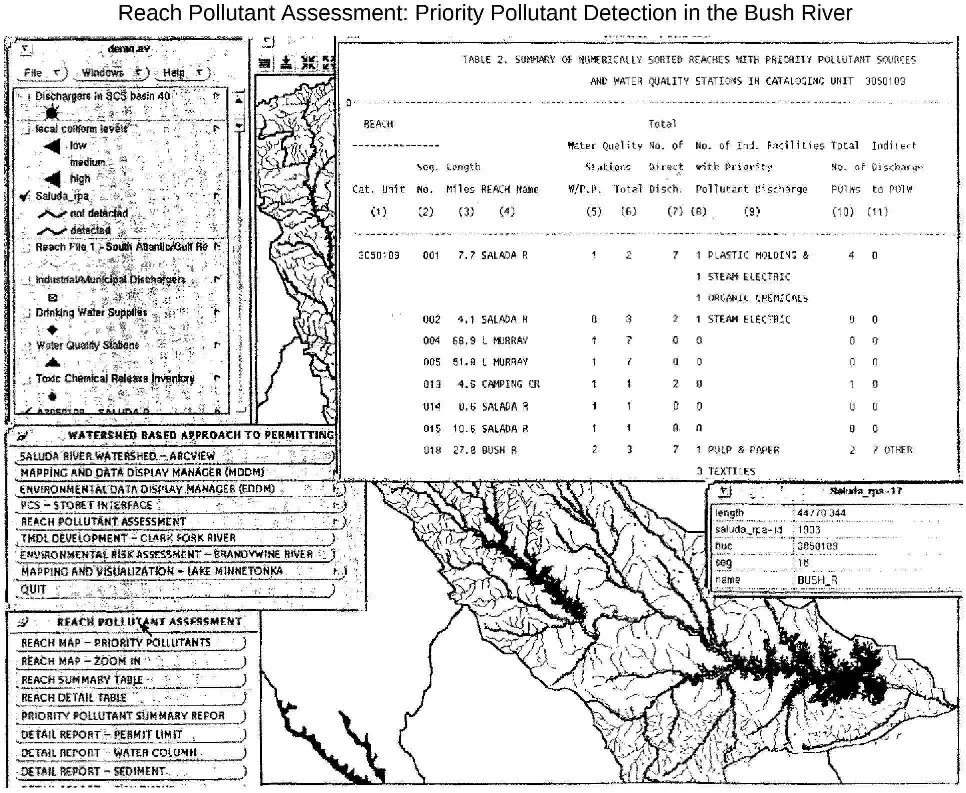
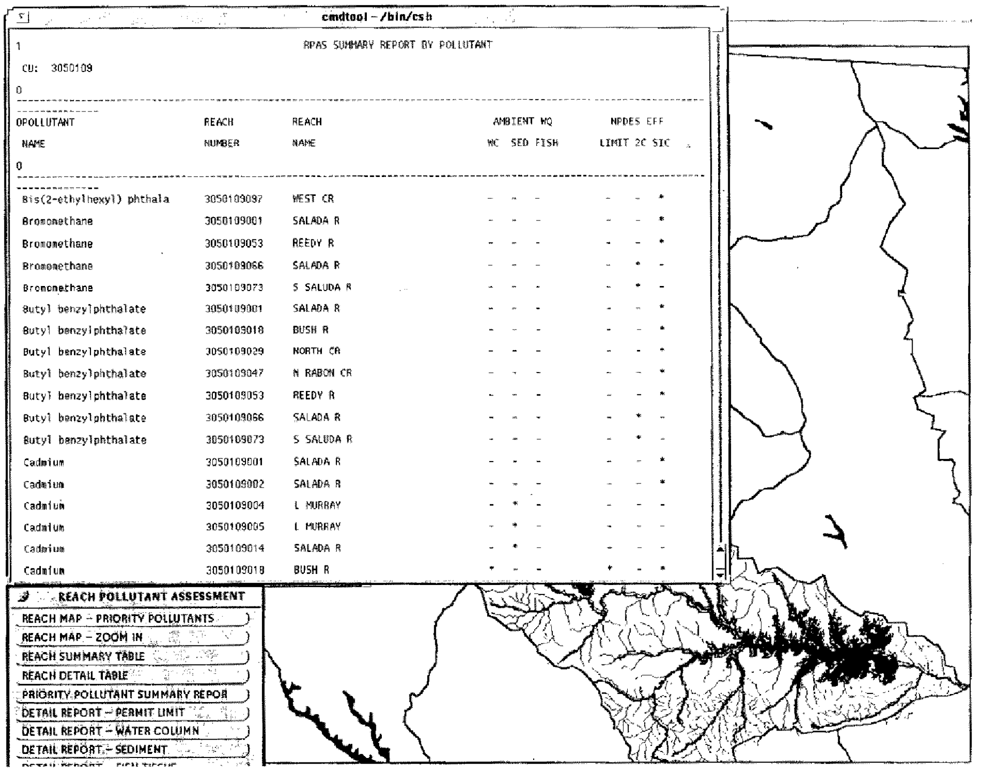


Figure 15. Using the RPA procedure to identify specific reaches with priority pollutants.

Reach Pollutant Assessment: Cadmium in the Bush River



1 RPAS SUMMARY REPORT BY POLLUTANT

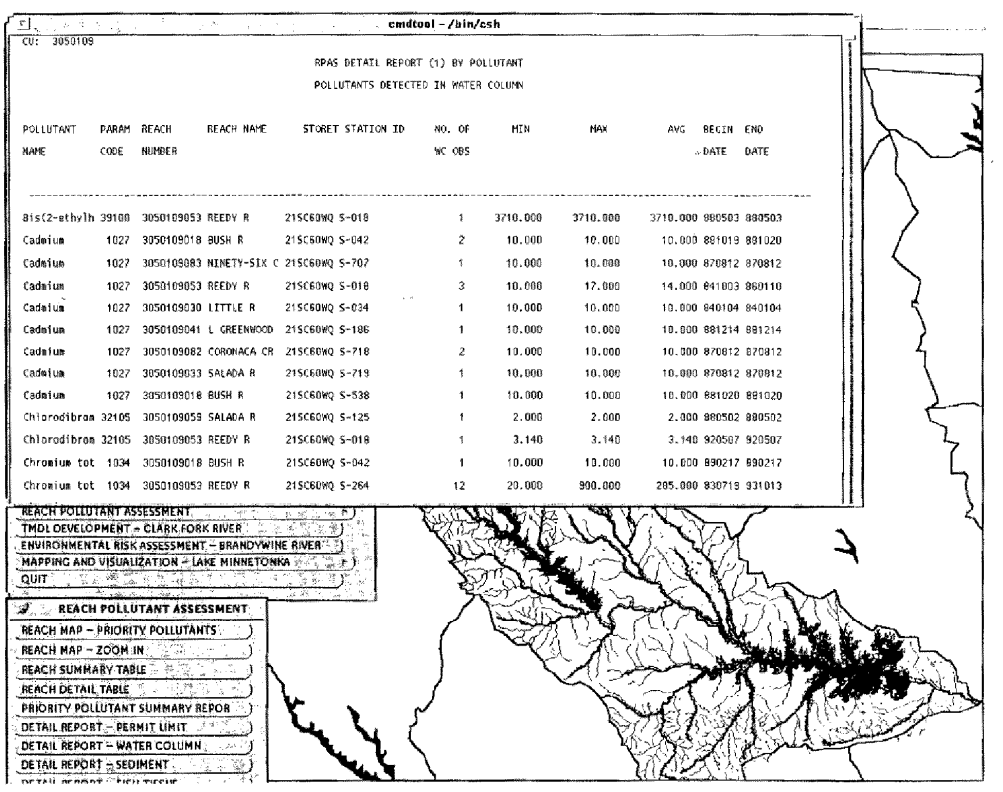
CU: 3050109

0

OPOLLUTANT NAME	REACH NUMBER	REACH NAME	AMBIENT WQ WC	AMBIENT WQ SED	AMBIENT WQ FISH	NPDES EFF LIMIT	NPDES EFF 2C	NPDES EFF SIC
Bis(2-ethylhexyl) phthala	3050109097	WEST CR	-	-	-	-	-	*
Bromomethane	3050109001	SALADA R	-	-	-	-	-	*
Bromomethane	3050109053	REEDY R	-	-	-	-	-	*
Bromomethane	3050109066	SALADA R	-	-	-	-	*	-
Bromomethane	3050109073	S SALUDA R	-	-	-	-	*	-
Butyl benzylphthalate	3050109001	SALADA R	-	-	-	-	-	*
Butyl benzylphthalate	3050109018	BUSH R	-	-	-	-	-	*
Butyl benzylphthalate	3050109029	NORTH CR	-	-	-	-	-	*
Butyl benzylphthalate	3050109047	N RABON CR	-	-	-	-	-	*
Butyl benzylphthalate	3050109053	REEDY R	-	-	-	-	-	*
Butyl benzylphthalate	3050109066	SALADA R	-	-	-	-	*	-
Butyl benzylphthalate	3050109073	S SALUDA R	-	-	-	-	*	-
Cadmium	3050109001	SALADA R	-	-	-	-	-	*
Cadmium	3050109002	SALADA R	-	-	-	-	-	*
Cadmium	3050109004	L MURRAY	-	*	-	-	-	-
Cadmium	3050109005	L MURRAY	-	*	-	-	-	-
Cadmium	3050109014	SALADA R	-	*	-	-	-	-
Cadmium	3050109018	BUSH R	*	-	-	*	-	*

Figure 16. RPA summary report by pollutant.

Reach Pollutant Assessment: Bush River: Cadmium in the Water Column



CU: 3050109

RPAS DETAIL REPORT (1) BY POLLUTANT

POLLUTANTS DETECTED IN WATER COLUMN

POLLUTANT NAME	PARAM CODE	REACH NUMBER	REACH NAME	STORET STATION ID	NO. OF WC OBS	MIN	MAX	AVG	BEGIN DATE	END DATE
Bis(2-ethylh	39100	3050109053	REEDY R	21SC60WQ S-018	1	3710.000	3710.000	3710.000	880503	880503
Cadmium	1027	3050109018	BUSH R	21SC60WQ S-042	2	10.000	10.000	10.000	881019	881020
Cadmium	1027	3050109083	NINETY-SIX C	21SC60WQ S-707	1	10.000	10.000	10.000	870812	870812
Cadmium	1027	3050109053	REEDY R	21SC60WQ S-018	3	10.000	17.000	14.000	841003	860110
Cadmium	1027	3050109030	LITTLE R	21SC60WQ S-034	1	10.000	10.000	10.000	840104	840104
Cadmium	1027	3050109041	L GREENWOOD	21SC60WQ S-186	1	10.000	10.000	10.000	881214	881214
Cadmium	1027	3050109082	CORONACA CR	21SC60WQ S-718	2	10.000	10.000	10.000	870812	870812
Cadmium	1027	3050109033	SALADA R	21SC60WQ S-719	1	10.000	10.000	10.000	870812	870812
Cadmium	1027	3050109018	BUSH R	21SC60WQ S-538	1	10.000	10.000	10.000	881020	881020
Chlorodibrom	32105	3050109059	SALADA R	21SC60WQ S-125	1	2.000	2.000	2.000	880502	880502
Chlorodibrom	32105	3050109053	REEDY R	21SC60WQ S-018	1	3.140	3.140	3.140	920507	920507
Chromium tot	1034	3050109018	BUSH R	21SC60WQ S-042	1	10.000	10.000	10.000	890217	890217
Chromium tot	1034	3050109053	REEDY R	21SC60WQ S-264	12	20.000	900.000	285.000	830719	931013

Figure 17. RPA detail report: pollutants detected in the water column.

Reach Pollutant Assessment: Permitted Industries for Cadmium—Bush River

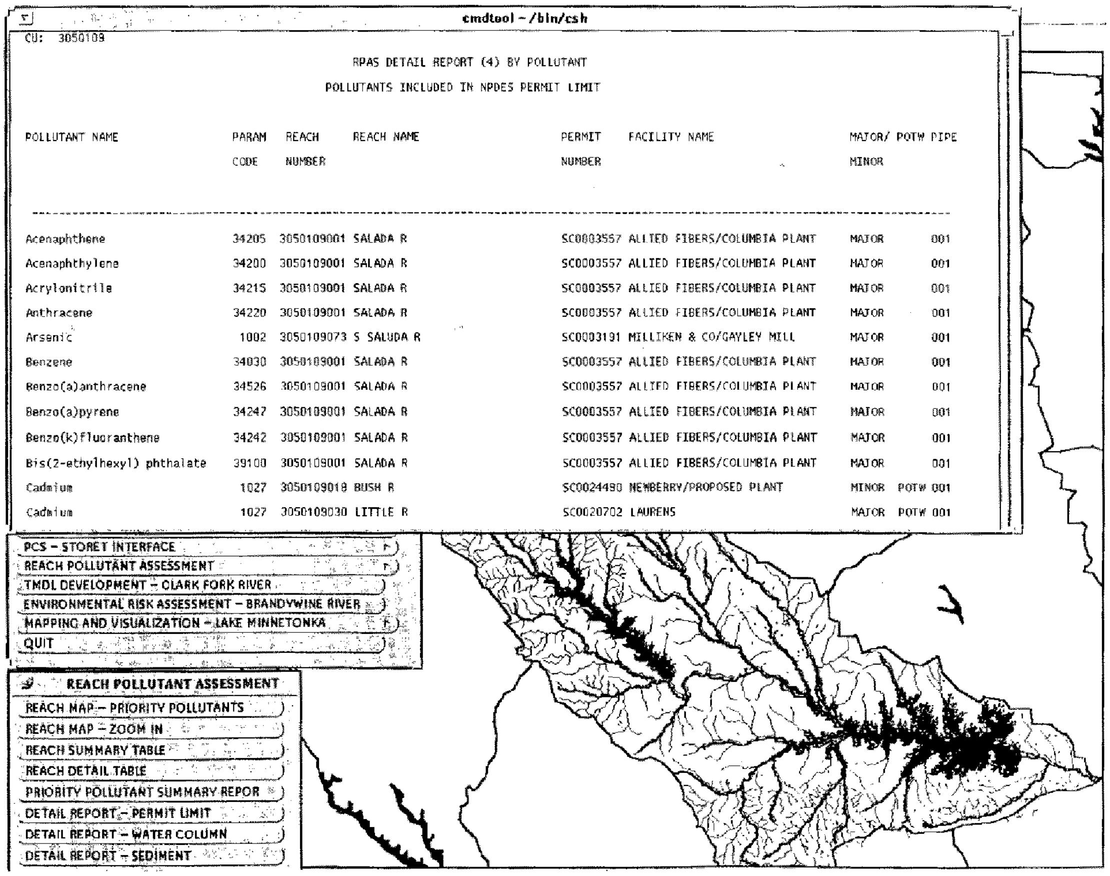


Figure 18. RPA detail report: pollutants included in the NPDES permit limit.

Two examples were presented as illustrations of how GIS and the mainframe databases and tools can work together.

In the first example, EPA's watershed-based approach to permitting, a four-step approach, was outlined, showing how a combination of local GIS functions and remote mainframe databases and tools were used in each step of the process. The end result was the targeting and prioritizing of watersheds of concern, and a detailed look at where and why water quality standards were not being met.

In the second example, the RPA program along with GIS was used to identify and map toxic priority pollutants and cross-link them to reaches, media (water column, sediment, fish tissue), NPDES dischargers, and monitoring stations. This analysis focused on what the toxic pollutant problems are and where they occur.

Acknowledgments

The authors wish to thank Jack Clifford, AWPD, OWOW, EPA, for providing data and ARC/INFO AMLs that were used in the South Carolina watershed example. The authors also extend their appreciation to Duane More and Tom Lewis (both with Martin Marietta), who provided support in setting up the X3270 and GDDMXD software.

References

1. Taylor, P.L., P. Evenhouse, L. Hoelman, T. DeWald, W.B. Samuels, and O. Hansen. 1988. STORET—Water quality analysis system. Presented at the State/EPA Data Management Conference, Raleigh, NC (June).
2. U.S. EPA. 1992. Office of Water environmental and program information systems compendium, information resources management: Tools for making water program decisions. EPA/800/B-92/001. April.
3. Samuels, W.B., P.L. Taylor, P.B. Evenhouse, T.R. Bondelid, P.C. Eggers, and S.A. Hanson. 1991. The environmental display manager: A tool for water quality data integration. Water Resources Bull. 27(6):939-956.
4. U.S. EPA. 1993. Geographic Resources Information and Data System (GRIDS): User guide. Office of Information Resources Management, National GIS Program.
5. Samuels, W.B. 1990. Reach pollutant assessment user's guide. McLean, VA: Science Applications International Corporation.
6. Washington State Department of Ecology (WDEC). 1993. Watershed approach to water quality management. F-WQ-93-029. October.
7. State of South Carolina. 1992. South Carolina code of regulations, Chapter 61, Regulation 68 and 69. Amended April 24, 1992.

Wetlands Applications

Wetlands Mapping and Assessment in Coastal North Carolina: A GIS-Based Approach

Lori Sutter and James Wuenscher
Division of Coastal Management, State of North Carolina Department of Environment, Health, and Natural Resources, Raleigh, North Carolina

Introduction

The coastal area of North Carolina covers 20 counties and over 9,000 square miles of land area, about 20 percent of the state (see Figure 1). It also includes over 87 percent of the state's surface water. The North Carolina Coastal Management Program (NC CMP) is responsible for managing this area to meet the goals set forth in the Coastal Area Management Act (CAMA) (North Carolina General Statute [NCGS] 113A, Article 7). These goals provide a broad mandate to protect the overall environmental quality of the coastal area and to guide growth and development in a manner "consistent with the capability of the land and water for development, use, or preservation based on ecological considerations" (NCGS 113A-102(b)(2)).

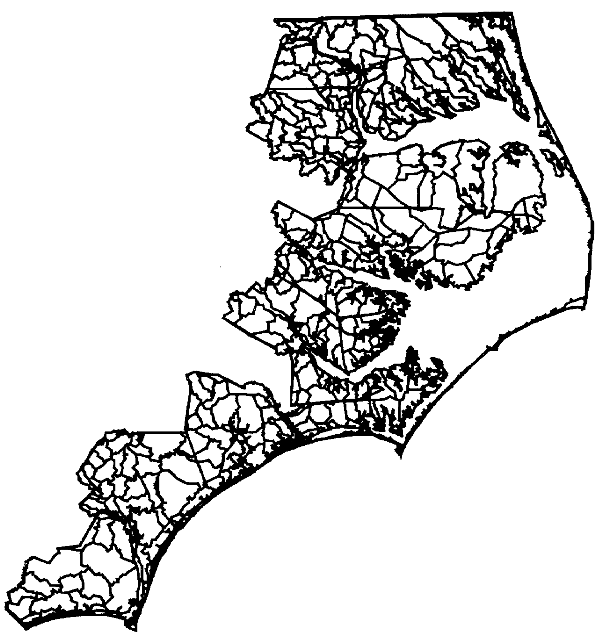


Figure 1. HU and county boundaries in the North Carolina coastal area.

Much of the North Carolina coastal area consists of wetlands, which, in many areas, constitute nearly 50 percent of the landscape. These wetlands are of great ecological importance, in part because they occupy so much of the area and are significant components of virtually all coastal ecosystems, and in part because of their relationships to coastal water quality, estuarine productivity, wildlife habitat, and the overall character of the coastal area.

Historically, close to 50 percent of the original wetlands of the coastal area have been drained and converted to other land uses (1-3). Although agricultural conversion, the largest historical contributor to wetlands loss, has largely stopped, wetlands continue to disappear as they are drained or filled for development. Conflicts between economic development and wetlands protection continue to be a major concern, with many coastal communities considering wetlands protection to be a major barrier to needed economic development.

Because wetlands are such a dominant part of the coastal landscape and are vitally important to many aspects of the area's ecology, their management and protection is a major concern of the NC CMP. The State Dredge and Fill Act (NCGS 113-229) and the CAMA regulatory program stringently protect tidal wetlands, or "coastal wetlands" as law and administrative rules call them. Coastal wetlands are designated areas of environmental concern (AECs), with the management objective "to give highest priority to the protection and management of coastal wetlands so as to safeguard and perpetuate their biological, social, economic and aesthetic values; and to coordinate and establish a management system capable of conserving and utilizing coastal wetlands as a natural resource essential to the functioning of the entire estuarine system" (15A NCAC 7H .0205).

North Carolina law does not, however, specifically protect nontidal freshwater wetlands. State protection of freshwater wetlands is limited to the regulatory authority provided under federal laws for state agency review of federal permits; in this case, §404 permits granted by the U.S. Army Corps of Engineers. Under §401 of the Federal Water Pollution Control Act (33 USC 1341), a Water Quality Certification from the North Carolina Division of Environmental Management (DEM) is required for a 404 permit to discharge fill material into wetlands. Section 307 of the federal Coastal Zone Management Act (CZMA - 16 USC 1451 et seg.) also requires that 404 permits be consistent with the enforceable rules and policies of the NC CMP. The standards for consistency are the use standards for AECs and wetlands policies stated in the applicable local land use plan. Other than AECs, the NC CMP has no consistent policies regarding wetlands. A few local land use plans include policies to protect freshwater wetlands, but most do not.

Wetlands Conservation Plan

In 1991, the CZMA §309 Assessment of the NC CMP revealed NC CMP's weakness in protecting nontidal wetlands (4). The assessment demonstrated that both opponents and proponents of wetlands protection considered the current system inadequate. Economic development interests found the 404 regulatory program to be unpredictable and inconsistent, often resulting in the loss of needed economic growth in coastal counties. Environmental interests felt that the program allowed the continued loss of ecologically important wetlands. As a result, the assessment identified wetlands management and protection as one of the primary program areas in need of enhancement.

The North Carolina Division of Coastal Management (DCM) developed a 5-year strategy (5) for improving wetlands protection and management in the coastal area using funds provided under the Coastal Zone Enhancement Grants Program established by 1990 amendments to §309 of the federal CZMA. The Office of Ocean and Coastal Resources Management (OCRM) in the National Oceanographic and Atmospheric Administration (NOAA), U.S. Department of Commerce administers the §309 program. Funds provided under this program, particularly Project of Special Merit awards for fiscal years 1992 and 1993, supported the work reported in this paper. A grant from the U.S. Environmental Protection Agency (EPA) for a Wetlands Advance Identification (ADID) project in Carteret County, North Carolina, also funded this work.

The key element of DCM's strategy for improving wetlands protection is the development of a wetlands conservation plan for the North Carolina coastal area. The plan has several components:

- Wetlands mapping inventory
- Functional assessment of wetlands
- Wetlands restoration
- Coordination with wetlands regulatory agencies
- Coastal area wetlands policies
- Local land use planning

The obvious first step in developing a wetlands conservation plan is to describe the wetlands resource. An extensive geographic information system (GIS) wetlands mapping program is helping to accomplish this first step by producing a GIS coverage of wetlands by wetland type for the entire coastal area. The GIS coverage allows generation of paper maps for areas within any boundaries available in GIS format. This is the subject of the first part of this report.

One weakness of the 404 program is that, for individual permits, it attempts to apply the same rules and procedures equally to all wetlands, regardless of the wetland type and location in the landscape. This approach can result in permits being granted for fill of wetlands of high ecological significance or permits being denied to protect wetlands of little significance. Neither outcome is desirable because the result may be the loss of either vital wetland functions or beneficial economic activity. This is an unsatisfactory way to manage wetland resources in an area such as the North Carolina coast, where:

- A high proportion of the land is wetlands.
- Many of the wetlands are vital to the area's environmental quality.
- Economic stimulation is sorely needed.

To help overcome this weakness in the current wetland regulatory framework, the Wetlands Conservation Plan includes an assessment of the ecological significance of all wetlands to determine which are the most important in maintaining the environmental integrity of the area. This will result in a designation of each wetland polygon in the GIS coverage as being of high, medium, or low functional significance in the watershed in which it exists. The procedure by which this occurs is the subject of the second part of this report.

The remaining components of the Wetlands Conservation Plan comprise the means by which the results of the wetland mapping and functional assessment steps will be used to improve wetland protection and management. Close coordination with other state and federal agencies involved in wetlands protection and management has been an important component of the entire effort. Agency representatives have been involved in developing the methods used, and the agencies will receive copies of the resulting maps for use in their own planning and decision-making. Policies for protection of wetlands of varying functional significance will be proposed to the Coastal Resources Commission to serve as the basis for consistency review of 404 permit applications. Wetland maps and functional assessment results will also be provided to local governments for use in local land use planning, and DCM will work with local governments to increase local involvement in the wetlands regulatory structure.

While the wetland maps themselves are useful for land use planning and helping to find suitable development sites, simply knowing where the wetlands are located is insufficient information for many purposes. Any area for which a 404 permit application is in process has been officially delineated as a wetland by the Corps of Engineers. The value of wetland maps to the regulatory review agencies at this stage is limited to determining the relationship of the site to other wetlands in the area. While, ideally, all wetlands should be avoided in planning development, avoiding wetlands completely in the coastal area is difficult, and avoiding all wetlands in any extensive development is virtually impossible.

The results of the functional assessment will provide additional information about the ecological significance of wetlands. This information will be valuable to wetland regulatory review agencies in determining the importance to an area's environmental integrity of protecting a particular site for which a permit to fill has been requested. It will also enable development projects to be planned so as to avoid, at all reasonable costs, the most ecologically important wetlands. An accurate functional assessment of wetland significance, then, is the most valuable component of the Wetlands Conservation Plan.

Wetlands Mapping Inventory

An important, initial step in developing a comprehensive plan for wetlands protection is to understand the extent and location of wetlands in the coastal area. When developing mapping methods, DCM quickly realized that the more than 9,000-square-mile coastal area was too large for any mapping effort in the field (see Figure 1). To complete this task in an accelerated timeframe, DCM needed to use existing data compatible with GIS. Reviewing the existing data revealed that most are not applicable for one of two reasons: (1) available wetlands data are based on older photography, and (2) more recent data are not classified with the intent of wetlands mapping. These data types, used independently, are inappropriate for use in a coastal area wetlands conservation plan. In addition, the classification schemes used in the existing methods are too complex or not focused on wetlands.

While several data sets were believed to be inappropriate if used exclusively for wetlands mapping in coastal North Carolina, each contained useful components. DCM elected to combine three primary layers of data and extract the most pertinent information from each layer. DCM selected the National Wetlands Inventory (NWI) because its primary purpose is to map wetlands. Unfortunately, these maps were based on photography from the early 1980s in coastal North Carolina, and many changes have occurred in the landscape since that time. NWI also omitted some managed wet pine areas from its maps; DCM wished to include these areas because they are important to the ecology of the North Carolina coastal area. DCM also selected detailed soils lines for use in its mapping efforts. While soils alone should not be used to identify wetlands, soils can be very useful in identifying marginal areas. Finally, DCM also employed thematic mapper (TM) satellite imagery in its methods. This data layer was not developed as a wetlands inventory; however, the imagery is more recent than the soils and NWIs. DCM desired to incorporate the benefits of each of these data sources into its mapping techniques.

The information provided by this mapping exercise will be useful to county and municipal planners in helping guide growth away from environmentally sensitive areas. For this reason, DCM elected to pursue mapping on a county by county basis. In addition, a single county allowed DCM to focus methodology development to a limited geographic area to refine its methods. Carteret County was selected as a methods development laboratory because data were available for the area and because Carteret has a large number of representative wetlands.

Data Descriptions

The U.S. Fish & Wildlife Service produces the NWI for all wetlands in the country. For the coastal North Carolina area, these vector data were developed from 1:58,000-scale color infrared photography taken during the winters of 1981, 1982, and 1983. Photointerpreters delineated wetland polygons on clear stabilene mylar taped over the photographs. After an initial scan of the photographs to identify questions or problem signatures, the photointerpreters reviewed areas in the field. They performed approximately one-half to one full day of field verification per quadrangle (quad) (6). Features were compared with U.S. Geological Survey (USGS) topographic maps for consistency. Following completion of the 'draft' paper maps, the Regional Coordinator reviewed the data. After approval as a final map, each quad was digitized. Initially, the North Carolina Center for Geographic Information and Analysis (CGIA) digitized the coastal North Carolina NWI maps, and later, the NWI Headquarters in St. Petersburg, Florida, who subcontracted the task, digitized them. Digital maps were obtained initially from 1/4-inch tape transfer and later from direct access to NWI via the Internet.

CGIA provided digital, detailed soil lines, which also are vector data based on 1:24,000 quads. County soil scientists delineated soil boundaries on aerial photographs based on slope, topography, vegetative cover, and other characteristics. This process occurs in any soil survey. After appropriate personnel approved the lines, a qualified soil scientist recompiled them onto orthophoto quads. CGIA scanned or manually digitized these lines. The coverage incorporated databases describing soil characteristics, which were then released for use.

The Landsat Thematic Mapper (TM) imagery was classified as part of the Albemarle-Pamlico Estuarine Study (APES). To provide complete coverage for the southernmost region of DCM's jurisdiction (Onslow, Pender, Brunswick, and New Hanover Counties), DCM contracted with CGIA and the North Carolina State University (NCSU) Computer Graphics Center to have that area processed identically to the APES region. These data provide a raster-based coverage of approximately 30-meter pixel resolution. Some of the imagery was taken at high tide, which precludes some near-water wetlands from appearing in certain areas. Using ERDAS, imagery specialists grouped similar spectral signatures into one of 20 classes. DCM used these data in two formats: filtered and unfiltered. The unfiltered information was vectorized with the ARC/INFO GRIDPOLY command. To remove some of the background noise in the coverage, it was filtered using ERDAS 'scan' with a Majority filter of 5 by 5 pixels, then vectorized with the ARC/INFO GRIDPOLY command.

Methods

Within each county, mapping is based on 1:24,000 USGS quads. After completion, each quad is assembled into a countywide coverage, which eventually is assembled into a coastal area coverage. The initial step in the mapping process is to ensure completion of the base layers described previously. Reviewing for errors at early stages prevents confusion in correction later in the process; therefore, the importance of the preliminary techniques cannot be overemphasized. The NWI data are first inspected to ensure complete coverage. If parts of the quad are missing, the error is investigated and corrected. Omissions may be areas of severe cloud cover on the photography or areas neglected during the digitization process. Next, the coverage is reviewed for missing label points. Any omissions are corrected based on the finalized version of the published NWI paper map. Appropriate NWI staff are contacted for the necessary information. At this time, labels are verified for typographical misentry. If not corrected, these errors could lead to confusion later in the mapping process.

Once the label errors are detected and corrected, the polygons are reviewed for completion. Verifying every line in the areas of coastal North Carolina densely populated with wetlands is impossible, but the lines are reviewed for completeness. NWI staff again must provide necessary information for any omissions. When the map is approved, technicians ensure projection of the quad to the State Plane Coordinate System. If this has not been completed, the ARC/INFO PROJECT command is employed.

The soils information is prepared in a similar manner to the NWIs, with questions being directed to the county soil scientist. Prior to the steps described previously, soils must be verified for completeness. Because soils are mapped by county boundaries and DCM maps by quad, some files must be joined in quads that intersect county boundaries. At this time, the quad must be checked for differing abbreviations between counties. Discrepancies are handled on a case-by-case basis. When an abbreviation describes different soils in different counties, a temporary abbreviation is created for one of the counties. If a single soil is described by two abbreviations across counties, both abbreviations are incorporated into the classification scheme.

The Landsat data do not require additional verification. Review of this layer is often helpful, however, to ensure that the geographic boundaries match. Cases where landforms do not appear to match require investigation of the discrepancies. If the area is misregistered, this layer might be omitted from the analyses. To date, no area has been mapped without this imagery.

The hydrogeomorphology of a wetland is unique in defining the wetland's function (7). Because these maps serve as the base for additional wetland projects (as described later in this report), an accurate determination of this characteristic is essential. Prior to the overlay procedure, technicians add a new item, hydrogeomorphic (HGM), to the NWI coverage. Because DCM considers both vegetation and landscape position in its classification (discussed later), riverine, headwater, and depressional wetland polygons are assigned an HGM of 'r,' 'h,' or 'd,' respectively. The digital line graphs (DLGs) of hydrography are essential in this step of the procedure.

All wetlands that are adjacent to streams or rivers are considered in the riverine HGM class and are designated as riverine polygons. This class should include all bottomland hardwood swamps and some swamp forests. It rarely includes any of the interfluvial wetland types. If it does, it is a small section of a large interfluvial flatwood from which a small stream emerges. Only the polygons adjacent to the stream are considered riverine. Headwaters are defined as linear areas adjacent to riverine areas that do not have a stream designated on the hydrography data layer. Because these unique systems form the transition between flatwoods and riverine wetlands, they are treated specially. Finally, polygons that exist on interfluvial divides are designated as flats or depressional wetlands. This class should not include any wetlands along streams.

The complete data coverages are overlaid to create a new, integrated coverage that often approaches 100,000 polygons. Each polygon has many characteristics assigned to it, including the Cowardin classification assigned by the NWI, the soil series provided by the detailed soil lines, the unfiltered land use/land cover code, the filtered land use/land cover code from the Landsat TM imagery, and the HGM classification assigned in the previous step.

Based on these characteristics, each polygon is assigned to one of DCM's classes through an automated ARC/INFO model using an arc macro language (AML). Personnel from the NWI and the North Carolina Department of Environment, Health, and Natural Resources Division of Soil and Water Resources have reviewed the classification of the Cowardin types into DCM wetland types. The classes that DCM currently recognizes are upland, salt/brackish marsh, estuarine shrub scrub, estuarine forest, maritime forest, pocosin, bottomland hardwood, swamp forest, headwater swamp, hardwood flatwoods, piney flatwoods, and managed pinelands. DCM also classifies soils as hydric or nonhydric based on List A of the U.S. Soil Conservation Service (SCS) List of Hydric Soils.

The base of the map is the NWI polygon coverage. Some NWI polygons are omitted from the DCM maps because they are temporarily flooded, but on nonhydric soils or because recent TM imagery indicates these areas are currently bare ground. The managed pineland wetland group on DCM maps includes areas that NWI considers uplands, identified as pine monocultures on the imagery, and that occur on hydric soil.

In addition, DCM also provides a modifier to some of these polygons. DCM notes if NWI has determined that the area has been drained or ditched. Areas designated as wetlands at the time of the NWI photography that currently appear as bare ground on the TM imagery are designated as 'cleared' on the maps. Many of the cleared areas would no longer be considered jurisdictional wetlands. These modifiers are useful indicators of the impacts wetlands sustain from human activities.

Initiation of an interactive session follows completion of the automated procedure. This session considers landscape characteristics that are not easily described to a computer model in correcting the classification. This is especially important in distinguishing bottomland hardwood swamps from hardwood flats. Both contain deciduous, broad leaf species of trees and can be temporarily flooded. The hydrology of these systems, however, is completely different. All bottomland hardwood swamps, for example, must be adjacent to a river where they receive seasonal floodwaters from the channel. Conversely, hardwood flatwoods should be located on interfluvial flats and not adjacent to any streams. Water is not introduced into hardwood flatwoods via a channel; rather, precipitation and ground water provide the water for this system. Polygons that are adjacent to rivers or estuaries but do not have a distinct channel designated in the hydrography coverage are considered headwater swamps.

During the course of methodology development, staff members visited at least 371 sites in the field. As staff members encountered new Cowardin classes, they would verify that the polygons were being placed into the correct DCM categories. If they determined that a particular Cowardin class was systematically misidentified, they updated the algorithm for automation. While this method does not provide for a usable accuracy assessment, it allowed development of the most accurate methods.

The accuracy of these data is unknown at this time. An accuracy assessment of the data is anticipated in the near future. This assessment will allow map users to understand the strengths and limitations of the data. It also will provide an overall summary of data error.

Functional Assessment of Wetlands

Certain initial considerations shaped the approach and methods used in developing a wetlands functional assessment procedure. The procedure needed to fit within the context and objectives of the Wetlands Conservation Plan for the North Carolina coastal area as described above. This context, and the opportunities and limitations it imposed, had considerable influence on the specific procedure developed.

Because we are dealing with a large geographic area with many wetlands, we recognized from the outset that we needed a method we could apply to large land areas without site visits to each individual wetland. This ruled out the many site-specific functional assessment methods that were applied in other contexts. Almost of necessity, a GIS-based approach was chosen. That meant we would have to use information available in GIS format and make use of GIS analytical techniques. The wetland mapping on which the functional assessment is based was performed using GIS, so the basic digital data were available.

The primary objective was to produce information about the relative ecological importance of wetlands that would be useful for planning and overall management of wetlands rather than to serve as the basis for regulatory decisions. While we could not visit every wetland, the goal was to predict the functional assessment value that a detailed, site-specific method would determine. We wanted to be able to predict in advance what the wetland regulatory agencies would determine as a wetland's significance so that the resulting maps would identify those wetlands where a 404 permit would be difficult or impossible to obtain. The resulting information would then be useful in determining where not to plan development. This would benefit potential permit applicants by preventing ill-advised plans that would be unlikely to receive permits and simultaneously serve to protect the most ecologically important wetlands. The result of the procedure, then, is not a substitute for a site visit in making regulatory decisions, but a predictor of what a site visit would determine.

A primary consideration was that the procedure be ecologically sound and scientifically valid, based on the best information available about the functions of wetlands. It needed to be based on fundamental principles of wetlands and landscape ecology rather than on arbitrary or subjective decisions.

Finally, the procedure was to be watershed-based. This requirement was primarily because consideration of a wetland's role in its watershed is the soundest basis for determining its ecological significance, but also because the other components of the Wetlands Conservation Plan, including wetland mapping and restoration planning, are based on watershed units. The watersheds being used are 5,000- to 50,000-acre hydrologic units (HUs) delineated by the SCS as illustrated in Figure 1. The North Carolina coastal area comprises 348 of these HUs. Watershed units of any size, however, could be used without changing the validity of the watershed-based considerations used in the procedure.

These initial considerations result in a summary definition of the functional assessment procedure. It is a GIS-based, landscape scale procedure for predicting the relative ecological significance of wetlands throughout a region using fundamental ecological principles to determine the functions of wetlands within their watersheds.

The functional assessment procedure is meant to be used with GIS data for regional application. It is not a field-oriented, site-specific method that involves visiting individual wetlands and recording information. A GIS-based procedure is the only practical approach for dealing with a large geographic area with many wetlands in a limited amount of time.

This GIS-based approach can make information on wetland functional significance available for broad regions in advance of specific development plans. The information is then available for planning to help avoid impacts to the most ecologically important wetlands. In this sense, the North Carolina procedure is unlike other functional assessment techniques that are designed for use in a regulatory context or that require field data for each wetland.

Data Requirements

Because the procedure uses GIS analysis, it requires digital information in GIS format. GIS data layers used in the procedure include:

- Wetland boundaries and types (the topic of the first section of this report).
- Soils maps.
- Land use/land cover.
- Hydrography.
- Watershed boundaries.
- Threatened and endangered species occurrences.
- Estuarine primary nursery areas.
- Water quality classifications.

In the North Carolina coastal area, these data layers either already existed and were available from the CGIA or were developed as part of the Wetlands Conservation Plan. Because other projects funded most of the data acquisition and digitization, developing the necessary GIS databases was not a major cost.

The soils coverage consists of digitized, detailed county soils maps produced by SCS and digitized by CGIA. The soils coverage allows identification of the soil series underlying a wetland, and the properties of the series are used to determine soil capacity for facilitating the wetland's performance of various functions.

The land use/land cover data layer was produced for the APES from interpretation of satellite TM imagery (8). It is used to determine land cover and uses surrounding each wetland and in the watershed.

The basic hydrography coverage consists of 1:24,000-scale USGS DLGs. Because the functional assessment procedure uses stream order as an indicator of watershed position, stream order according to the Strahler system was determined manually and added to the DLG attribute files.

As described previously, the watersheds used in the procedure are relatively small HUs delineated by SCS. DCM contracted with CGIA to have these boundaries digitized for the coastal area. During the digitization process, the watershed boundaries were rectified to USGS and DEM boundaries of larger subbasins to ensure that the HUs could be combined into larger watershed units.

A data layer produced by the North Carolina Natural Heritage Program is used to identify threatened and endangered species occurrences. The North Carolina Division of Marine Fisheries maintains the coverage of primary nursery areas, and the Division of Environmental Management developed a map of water quality classifications that was digitized by CGIA.

The ways in which these data layers are used to determine values for various parameters in the functional assessment procedure are described later in this report. The GIS procedures have been automated using ARC/ INFO

AML on a Sun workstation. The AML programs are available from DCM to anyone planning to use the procedure elsewhere.

Because the assessment procedure was designed for GIS analysis, the choice and expression of individual parameters have been shaped to some extent by the GIS data available and the capabilities and limitations of ARC/INFO techniques and AML automation. DCM was fortunate to have a relatively large amount of GIS data readily available. For use in other areas, the procedure could be modified to use different GIS coverages. At least the first five databases listed above, however, are essential to its basic propositions.

Classification Considerations

The HGM classification system for wetlands (7) classifies wetlands into categories based on landscape position (geomorphic setting), water sources, and hydrodynamics (direction of water flow and strength of water movement). It is being increasingly used as the basis for wetland classification and functional assessment systems. HGM classification focuses on the abiotic features of wetlands rather than on the species composition of wetland vegetation as do most traditional wetland classification schemes.

Several features of the HGM classification system make it a useful starting point for an assessment of wetland functions. Because the HGM system is based on geomorphic, physical, and chemical properties of wetlands, it aggregates wetlands with similar functions into classes. The HGM class of a wetland, in itself, indicates much about the ecosystem functions of the wetland. The HGM approach also forces consideration of factors external to the wetland site, such as water source. This helps relate the wetland to the larger landscape of which it is a part and puts consideration of the wetland's functions in a landscape and watershed context.

Three HGM classes are used as the starting point for the North Carolina functional assessment procedure. All wetlands are first classified as one of the following:

- Riverine
- Headwater
- Depressional

Riverine wetlands are those in which hydrology is determined or heavily influenced by proximity to a perennial stream of any size or order. Overbank flow from the stream exerts considerable influence on their hydrology. Headwater wetlands exist in the uppermost reaches of local watersheds upstream of perennial streams. Headwater systems may contain channels with intermittent flow, but the sources of water entering them are precipitation, overland runoff, and ground-water discharge rather than overbank flow from a stream. Depressional wetlands, including wet flats and pocosins, generally are not in direct proximity to surface water. While they may be either isolated from or hydrologically connected to surface water, the hydrology of depressional wetlands is determined by ground-water discharge, overland runoff, and precipitation.

The functions of wetlands in these different HGM classes differ significantly. Riverine wetlands regularly receive overbank flow from flooding streams and, thus, perform the functions of removing sediment and pollutants that may be present in the stream water and providing temporary floodwater storage. Headwater and depressional wetlands cannot perform these functions because they do not receive overbank flow. Headwater wetlands occur at landscape interfaces where ground water and surface runoff coalesce to form streams. Headwater wetlands provide a buffer between uplands and stream flow so they can perform significant water quality and hydrology functions. While depressional wetlands do not perform buffer functions, they often store large amounts of precipitation or surface runoff waters that otherwise would more rapidly enter streams. Wetlands in all HGM classes can perform important habitat functions.

Because the wetlands in these different HGM classes are functionally different, their functional significance is assessed using different, though similar, procedures. If the same procedure were used for all HGM classes, depressional wetlands would always be considered of lower functional significance simply because they are not in a landscape position to perform some of the water quality and hydrologic functions of riverine and headwater wetlands.

In addition to HGM classes, wetland types identified by dominant vegetation are used at several points in the functional assessment. This reflects a recognition that the biologic properties of a wetland site considered together with its hydrogeomorphic properties can provide a more detailed indication of its functions than either taken alone. The HGM class of a wetland, as a broad functional indicator, determines which assessment procedure to use. Within each HGM class and corresponding assessment procedure, wetland type determines the level or extent of specific parameters.

The wetland types used are those typical of the North Carolina coastal area. They result from a clumping of the Cowardin classes used on NWI maps into fewer types with more intuitively obvious type names (e.g., swamp forest, pocosin), as described previously. These wetland types are used in the wetland maps that form the starting point for the functional assessment.

Wetland types are used in the procedure as indicators of functional characteristics. Correlations between wetland type and wetland functions were determined from statistical analysis of field data from nearly 400 sites. At each site, the presence or absence of a list of functional indicators was recorded. Dr. Mark Brinson of East Carolina University developed the functional indicators lists, in part. Dr. Brinson served as primary scientific consultant in developing the HGM classification system and the field sampling methodology.

Wetland types differ in other areas, so their inclusion in this procedure limits its use in its current form to the southeastern coastal plain. Adaptation of the procedure for use in other areas would require either extensive field sampling as was performed in coastal North Carolina or a more arbitrary clumping of wetland types based on best professional judgment. Other methods of wetland classification could be used, provided wetlands are classified in such a way that functional characteristics of the wetland types are constant and can be determined by field sampling, literature values, and/or professional judgment. The procedure could be applied directly to NWI polygons if these are the only wetland map base available.

In addition to wetland type, several other parameters are used as indicators of the existence or level of specific wetland functions. These include both site-specific parameters, such as wetland size and soil characteristics, and landscape considerations, such as watershed position, water sources, land uses, and landscape patterns. GIS analysis determines values for these parameters based on the data layers discussed above. They could be determined manually, but the process would be very labor intensive.

Unlike assessment procedures that depend solely on information that can be collected within a wetland, this procedure relies heavily on factors external to the wetland site itself. Relationships between a wetland and the landscape within which it exists are integral considerations in determining wetland functional significance. Characteristics of the landscape surrounding a wetland are often more important determinants of its functional significance than are the characteristics of the wetland itself. Of the 39 parameters evaluated in the procedure, 21 are landscape characteristics, and 18 are internal characteristics of the wetland itself.

While we believe this emphasis on a wetland's landscape context is a more ecologically sound approach to functional assessment than site-specific methods, it requires a great deal more information than could be collected within the wetland itself. The procedure is based on GIS data and analysis, not only to make it suitable for regional application, but because GIS provides the most practical way to analyze the spatial relationships of landscape elements and their properties.

Structure of the Assessment Procedure

The assessment procedure uses a hierarchical structure that rates individual parameters and successively combines them to determine the wetland's overall functional significance. The complete hierarchical structure is illustrated in Figure 2. It consists of four levels:

- Overall functional significance of the wetland.
- Specific functions and risk of wetland loss.
- Subfunctions.
- Parameters evaluated to determine the level and extent of functions.

The objective of functional assessment is to determine an individual wetland's ecological significance in its watershed and the larger landscape. The highest hierarchical level, or end result of applying the procedure, then, is the wetland's overall functional significance.

The second hierarchical level includes the four primary factors that are considered in determining the wetland's functional significance (see Figure 3). The overall ecological significance of a wetland is determined by the degree to which it performs, or has the capacity to perform, specific functions. The broadest grouping of wetland functions includes water quality functions, hydrologic functions, and habitat functions. The nature of the landscape and the water characteristics of the watershed in which a wetland functions also determine ecological significance to some extent. These factors determine the potential risk to watershed and landscape integrity if the wetland functions were lost. Including a "risk factor" as a basic consideration in functional assessment also provides a means of considering cumulative impacts and the practicality of replacing lost functions through mitigation in determining a wetland's overall significance.

Each primary function of wetlands is actually a combination of separate, more specific subfunctions. Water quality subfunctions include the removal of nonpoint source pollutants from surface runoff and the removal of suspended or dissolved pollutants from flooding streams. Hydrology subfunctions include storage of precipitation and surface runoff, storage of floodwater from streams, and shoreline stabilization. Habitat subfunctions include providing habitat for both terrestrial species and aquatic life. Several considerations that, while not truly wetland functions, are called subfunctions for parallelism also determine risk factor. The subfunction levels of the assessment procedure are illustrated in Figures 4 through 7.

Properties of the wetland and its surrounding landscape determine the extent to which a wetland performs these different subfunctions. The assessment procedure refers to these properties as "parameters." Parameters

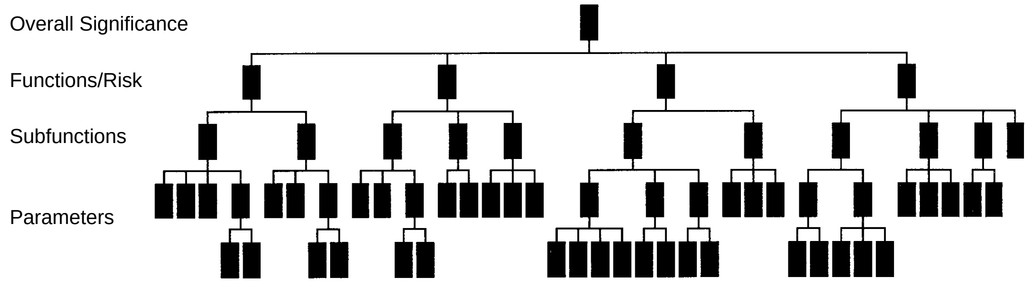


Figure 2. Overall hierarchical structure of the functional assessment procedure.

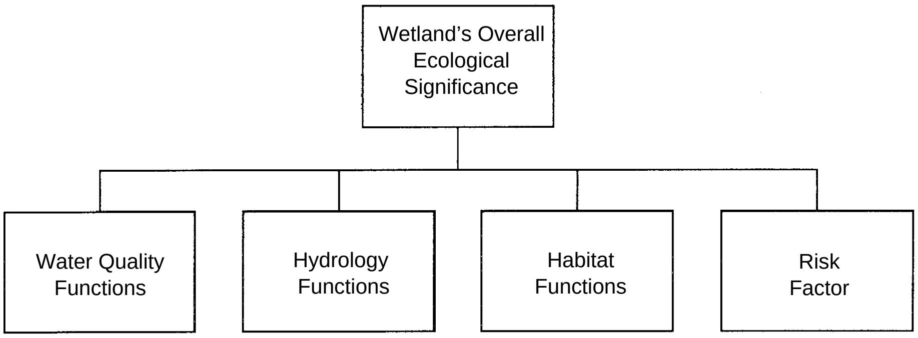


Figure 3. Assessment level two: Primary wetland functions and risk factor.

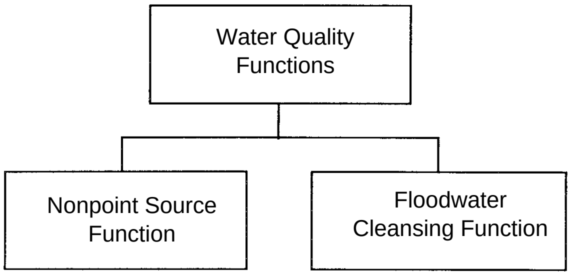


Figure 4. Water quality subfunctions.

make up the levels in the hierarchical structure that are actually evaluated based on fundamental ecological considerations. Parameter values, in turn, are combined to produce ratings for the subfunctions. Future reports will explain in detail all parameters evaluated in the assessment procedure and document them for scientific validity. This paper discusses only the parameters under the nonpoint source removal subfunction of the water quality function for illustration (see Figure 8).

The first parameter determining a wetland's significance in removing nonpoint source pollutants from surface runoff water is whether the water contains sediment, nutrients, or toxic pollutants in significant quantities. This is evaluated in the "proximity to sources" parameter based on the land uses surrounding the wetland. If agricultural fields or developed areas from which pollutants are likely to enter surface runoff largely surround the wetland, the wetland's potential for removing nonpoint source pollutants is high. If, on the other hand, natural vegetation from which runoff water is likely to be largely unpolluted mostly surrounds the wetland, its potential for removing significant pollutants is low.

Proximity to sources is an "opportunity" parameter. That is, it determines whether a wetland has the opportunity to remove pollutants from surface runoff by considering how likely the runoff water is to be polluted. The other parameters for this subfunction are "capacity" parameters that measure the wetland's ability to perform the function if the opportunity is present. Opportunity and capacity parameters are treated differently in determining a wetland's overall significance to prevent a wetland from being rated lower simply because present opportunity does not exist. This is discussed in more detail below.

The second parameter considered in determining a wetland's significance in nonpoint source removal is its proximity to a surface water body. If runoff entering a wetland would otherwise directly enter surface water, the wetland's significance as a filter is greater than if the wetland is far removed from surface water. In that case, pollutants in runoff could either settle out or be removed by other means before they enter surface water as pollutants.

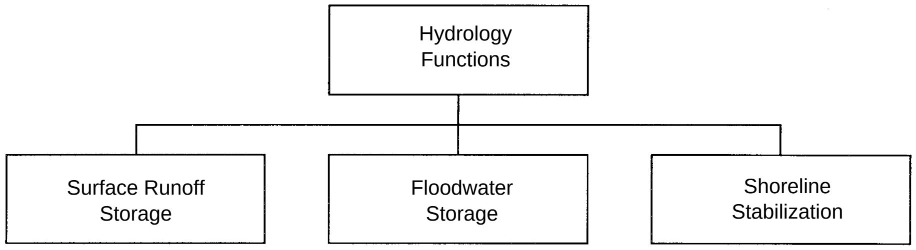


Figure 5. Hydrology subfunctions.

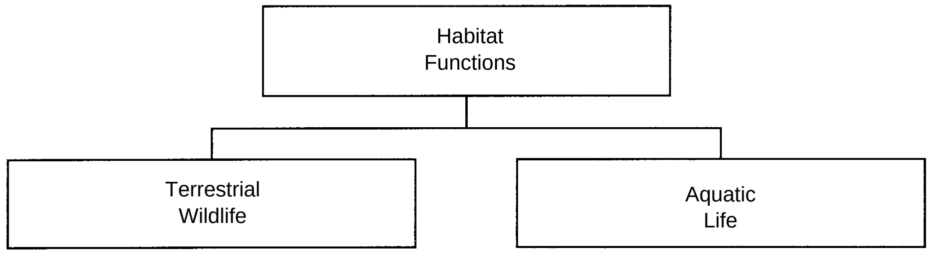


Figure 6. Habitat subfunctions.

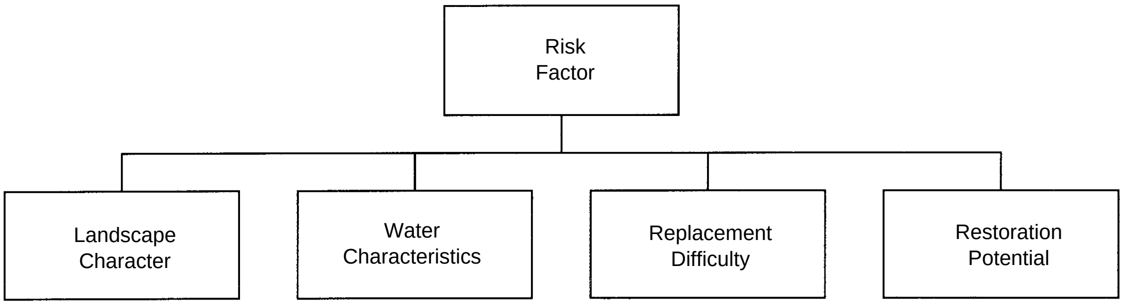


Figure 7. Risk factor subfunctions.

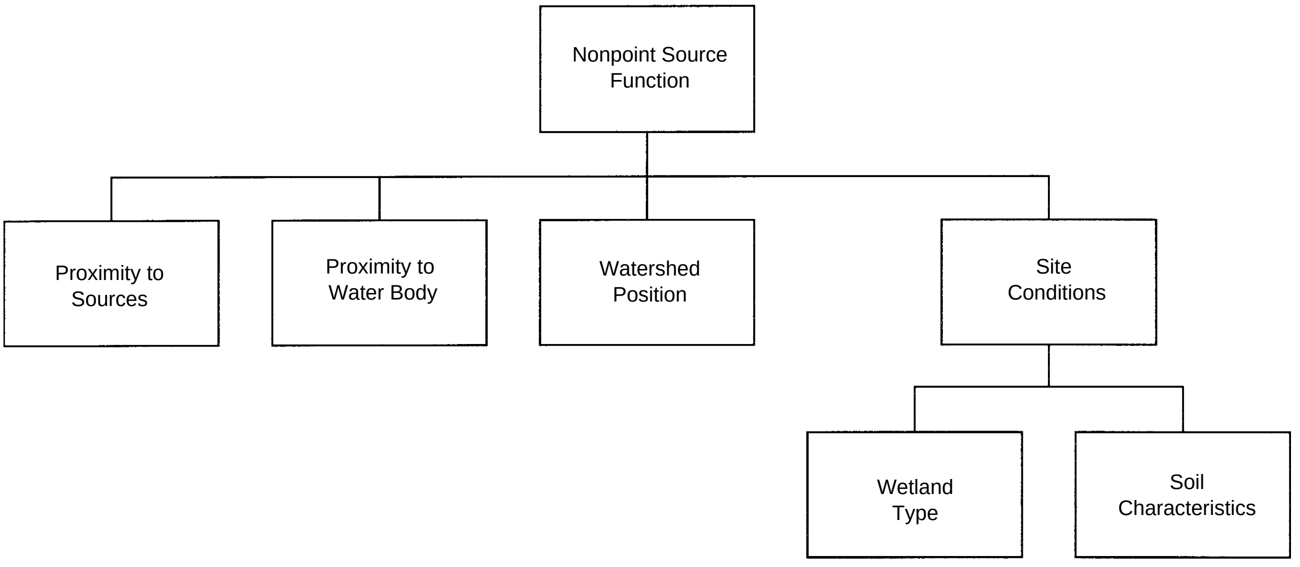


Figure 8. Parameters evaluated under nonpoint source pollutant removal subfunction.

The third parameter is the position of the wetland in its watershed. Several studies have documented that headwater wetlands are most effective in removing non-point source pollutants (9-11). Thus, the higher in its watershed a wetland is located, the higher is its significance in nonpoint source removal.

Two subparameters, wetland type and soil characteristics, determine the value of the fourth parameter, site conditions. By virtue of their typical microtopography, hydrology, and vegetative structure, some wetland types more effectively retain and filter surface runoff than do other types. Some soil series are more effective than others in retaining and chemically transforming pollutants. Each subparameter is rated, and their combined values produce a rating for the site conditions parameter.

A similar evaluation of specific parameters is performed to derive significance ratings for other wetland subfunctions. In all cases, GIS analysis determines parameter values based on the data layers described above. Some parameters, such as wetland type in the nonpoint source illustration, are surrogates or indicators of other wetland properties that actually determine the wetland's functional capacity. The limitations of GIS data and techniques necessitate the use of indicator parameters.

Evaluation Procedure

The objective of the assessment procedure is to determine an individual wetland's ecological significance in the watershed in which it exists. Ecological significance is divided into three broad classes (high, medium, and low) rather than attempting to derive a specific numerical "score." This is partly because of the procedure's initial application in an EPA ADID project performed by DCM in Carteret County, North Carolina. Standard ADID procedure is to classify wetlands into three groups:

- Areas generally unsuitable for the discharge of dredged or fill material.
- Areas that require a project-by-project determination.
- Possible future disposal sites for dredged or fill material.

These groups correspond to the H, M, and L used in the assessment procedure.

The approach of classifying wetlands into three broad functional significance classes is also used, however, because it is feasible with our current understanding of wetland function. Attempting to assign a specific value along a numeric continuum of functional significance greatly exaggerates the precision with which we can realistically apply current knowledge. The three significance classes used in the assessment procedure provide the information necessary to meet the procedure's objectives without going beyond the realm of reasonable scientific validity.

As explained above, the basic evaluation is performed at the parameter level. An H, M, or L value is assigned to each parameter as it relates to the performance of the wetland subfunction being considered. For example, if the soils underlying a wetland have properties that are highly conducive to the function being considered, the soil characteristics parameter is rated H; if soil properties are less conducive to performing the function, the parameter is rated M; and if soil properties are not at all conducive to the function, the parameter is rated L. All individual parameters under a given subfunction receive similar ratings.

The individual parameter ratings are then combined to give an H, M, or L rating for each subfunction. The subfunction ratings are combined into a rating of the wetland's significance in performing each of the primary wetland functions. Finally, the ratings for primary functions are combined into an overall rating of the wetland's functional significance.

The process of successively combining ratings up the structural hierarchy is the most complex aspect of the assessment procedure. The combining, as well as the evaluation of individual parameters, is based on fundamental ecological principles about how wetlands and landscapes function. Because the ecological processes themselves interact in complex ways, combining ratings is much more complex than a simple summation of individual ratings. Some parameters are normally more important than others in determining the level at which a wetland performs a specific function and, thus, must be weighed more heavily in determining the combined value. In some cases, different combinations of individual parameter ratings result in the same level of functional significance. Each possible combination of parameters must then be considered.

The automated version of the assessment procedure maintains all individual parameter ratings and combinations in a database. Because the combining process is complex, the reason a wetland receives an overall H, M, or L rating may not be intuitively obvious. The database makes it possible to trace through the parameter, subfunction, and primary function ratings that result in a wetland's overall rating.

This database also allows consideration of specific wetland functions individually. For example, in a watershed targeted for nonpoint source pollution reduction, one management objective may be to give the highest level of protection to wetlands most important in performing this function. The database allows examination of each wetland for its significance in nonpoint source removal and production of a map of wetlands rated according to their significance for this single function.

Individual function ratings in the database can also be used to improve planning, impact assessment, and mitigation for development projects that affect wetlands. If alternative sites are available, such as alternative corridors for a highway, the alternative with the least impact on the wetland function considered most important in the watershed can be identified. Rather than simply minimizing acres of wetland impact, the objective would be to minimize impacts to the most important wetland functions. Environmental assessment of wetland impacts can identify specific functions to be lost. Mitigation can be improved by giving priority to sites with the highest potential for performing the same functions.

Future reports will explain detailed procedures for evaluating individual parameters and combining them into functional ratings. This paper illustrates only the water quality nonpoint source removal subfunction. The rating system for this subfunction is summarized in Figure 9.

Four parameters are evaluated to determine the significance of the nonpoint source removal subfunction. Because (d), the site conditions parameter, has subparameters below it, it is first evaluated using a relatively simple procedure. If conditions typical of the wetland type and characteristics of the underlying soil are both highly conducive to removal of pollutants in runoff water entering the wetland, the site conditions parameter is H. If either the wetland type or the soil is not at all conducive to pollutant removal and the other subparameter is no more than somewhat conducive, the site conditions parameter is L. Any other combination results in an M.

Parameters

(a) Proximity to Sources

(b) Proximity to Surface Water

(c) Watershed Position

(d) Site Conditions

  (1) Wetland Type

  (2) Soil Characteristics

Evaluation Procedures

Site Conditions

H Both Parameters H

M Other combinations

L One parameter L and neither H

NPS Subfunction

H (a) & (b) H and (d) at least M *or*
(c) & (d) H and (b) at least M

M Other combinations

L Two of (b), (c), & (d) L

Figure 9. Parameters evaluated under nonpoint source pollutant removal subfunction.

Following evaluation of all parameters, they are combined to evaluate the significance of the wetland in removing nonpoint source pollutants. Two combinations result in the wetland being evaluated as highly significant in performing this function. First, if the wetland is adjacent to both a significant source of polluted runoff (a = H) and a permanent surface water body into which the runoff would flow if the wetland were not there (b = H), *and* has site conditions that are at least reasonably efficient in catching, holding, and removing pollutants from the runoff (d at least M), it receives an H. Alternatively, even if the wetland is not adjacent to a pollutant source, it receives an H if it is in the headwaters of the watershed (c = H), site conditions are highly conducive to pollutant removal (d = H), and it is at least close to an intermittent stream (b at least M).

On the other hand, if any two of parameters (b), (c), and (d) are evaluated L, the significance of the wetland for nonpoint source pollutant removal is **L**ow. That is, the wetland is evaluated as L for this function if any of the following conditions exist:

- The wetland is not close to surface water (b = L) and downstream in the watershed (c = L).
- The wetland is not close to surface water (b = L), and its site conditions are poor for pollutant removal (d = L).
- The wetland is downstream in the watershed (c = L) and has poor site conditions (d = L).

Any combination of parameter evaluations other than those resulting in an H or L results in the wetland being evaluated as of moderate significance for removing nonpoint source pollutants. This example is typical of evaluation procedures used for all subfunctions. More often than not, the evaluation procedures are complex and multifarious in their reasoning and application. Hopefully, though, they are scientifically valid based on current knowledge of wetland ecology.

Opportunity and Capacity

The concepts of opportunity and capacity for a wetland to perform a given function were briefly discussed above. For a wetland to actually perform a function, it must have both the opportunity and the capacity for the function. In terms of the nonpoint source example, a source of potentially polluted runoff must enter the wetland to provide an opportunity, and the wetland must have the internal capacity to hold the runoff and remove the pollutants before releasing the water. Factors external to the wetland usually determine the opportunity to perform a function, while properties of the wetland itself along with its landscape position determine the capacity to perform the function.

Because the assessment procedure is a landscape scale procedure that evaluates the functions a wetland performs in relation to its surroundings, essentially every subfunction includes opportunity parameters. A functional assessment that is too heavily dependent on opportunity parameters, however, is static and rapidly becomes invalid as land uses change. A wetland that is bordered by natural forest today can be bordered by a young pine plantation or a subdivision under construction by next year. The fact that a wetland does not have the opportunity to perform certain functions today does not mean that it will not have the opportunity in the future. If an assessment of wetland significance is to remain valid over time in a landscape subject to change, opportunity parameters alone cannot be determinative.

The evaluation procedure for the nonpoint source subfunction explained above is an example of how the assessment procedure handles this situation. The opportunity for a wetland to receive polluted runoff water from surrounding lands (a = H) *can* result in an evaluation of H for this subfunction if other properties are also present, but it does not *have* to be present for a wetland to be evaluated H. Other parameters (c and d = H, and b at least M) that give a wetland a high capacity to remove nonpoint source pollutants can also result in an H. Conversely, lack of present opportunity (a = L) does not result in an evaluation of low significance for this function. At least two of the other parameters must be L for the wetland to be evaluated as L.

These conventions hold throughout the procedure. A present high opportunity to perform a function can result in an evaluation of high significance for the function, but high capacity can also result in an H evaluation even if present opportunity is lacking. Lack of present opportunity alone never results in an evaluation of low significance for a function. High opportunity is treated essentially as a "bonus" consideration that can result in a higher evaluation for a wetland than its capacity alone would indicate but that will never result in a lower evaluation because of its absence.

Overriding Considerations

Several considerations are of such importance in the North Carolina coastal area that their presence alone will result in a wetland evaluation of high significance. These parameters are evaluated first as either true or false, and if one or more of them is true, the rest of the evaluation procedure is not performed.

The first overriding consideration is whether the wetland is a salt or brackish marsh meeting the definition of "coastal wetland" as set forth in North Carolina statutes (NCGS 113-229(n)(3)) and rule (NCAC 7H .0205(a)). Coastal wetlands in North Carolina are designated by law as highly significant. Consequently, the assessment procedure evaluates them automatically as H and includes no considerations for differentiating among the functional significance of these wetland types.

The second overriding consideration is whether the wetland is adjacent to an officially designated primary nursery area (PNA). All designated PNAs are included in "areas of environmental concern" in the NC CMP and are protected by a specific set of regulations. They are areas where initial postlarval development of finfish and crustaceans takes place and, thus, are critical to estuarine fish and shellfish populations. Wetlands adjacent to PNAs are highly important in maintaining water quality and appropriate salinity gradients in these critical areas and are automatically evaluated as of high functional significance.

The third overriding consideration is whether the wetland contains threatened or endangered species. If a known threatened or endangered plant or animal species on either federal or state lists is present, the wetland is evaluated as highly significant. The determination is based on information obtained from the North Carolina Natural Heritage Program.

The fourth overriding consideration is whether the wetland includes all or part of a critical natural area as designated by the North Carolina Natural Heritage Program. If so, the site is considered of high significance. GIS data layers maintained by the Natural Heritage Program also help make this determination.

Verification

Throughout the development and initial application of the assessment procedure, we have checked and verified its validity. Parameter evaluations and combination procedures are based on the best wetland science available in the scientific literature. The validity and accuracy of the GIS databases used to apply the procedure have been verified to the extent possible. Following sections of this report fully document any assumptions made about wetland ecology, GIS data, or GIS analytical techniques.

An advisory panel of wetland scientists familiar with the wetlands of coastal North Carolina and representatives of several state and federal wetland-related agencies reviewed every step of the procedure's development. While their review does not represent an endorsement of the procedure or its results by the agencies or individuals included, it does indicate the level of peer review the procedure has received.

During development of the procedure, field visits were made to nearly 400 wetland sites to gather data on functional indicators. On these same site visits, a field-based functional assessment procedure, the Wetland Rating System developed by the North Carolina Division of Environmental Management, was applied. This provides the basis for a field verification of the assessment procedure.

Discussion

As we continue to understand more about the role of wetlands in maintaining a healthy environment, the usefulness of wetlands locational data continues to grow in importance. Spatial data can assist county planners in guiding development away from environmentally sensitive areas. Landowners now have the capability to look at a map and realize very quickly that wetlands exist on a given area of land. In addition, economic development councils can use this information to plan development in areas attractive to a particular industry. If a new business or industry wishes to locate in an area positioned such that the wetlands permitting process could be avoided, maps showing lands void of wetlands could be a significant tool to the economic development council. The representation of these wetlands' ecological significance dramatically increases the utility for these data.

While paper maps can be distributed to all interested parties, digital data also are available to public agencies who have GIS capabilities. In Carteret County, for example, a publicly installed workstation will be made available with these data installed. The county government will be able to view wetlands in the context of cadastral boundaries that already are on GIS. Information about sensitive resources made available prior to any development will, hopefully, lead development away from environmentally sensitive areas.

References

1. Hefner, J.M., and J.D. Brown. 1985. Wetland trends in the southeastern United States. Wetlands 4:1-12.
2. Dahl, T.E. 1990. Wetlands losses in the United States 1780s to 1980s. U.S. Department of the Interior, Fish and Wildlife Service, Washington, DC.
3. DEM. 1991. Original extent, status, and trends of wetlands in North Carolina. Report No. 91-01. North Carolina Department of Environment, Health, and Natural Resources, Division of Environmental Management, Raleigh, NC.
4. DCM. 1992. Final assessment of the North Carolina coastal management program. Report to the Office of Ocean and Coastal Resource Management, NOAA, U.S. Department of Commerce, performed under the Coastal Zone Enhancement Grants Program (January 10). Section 309, CZMA, North Carolina Division of Coastal Management, Raleigh, NC.
5. DCM. 1992. Final strategy for achieving enhancements to the North Carolina coastal management program. Proposal to the Office of Ocean and Coastal Resource Management, NOAA, U.S. Department of Commerce, performed under the Coastal Zone Enhancement Grants Program (March 25). Section 309, CZMA, North Carolina Division of Coastal Management, Raleigh, NC.
6. Hefner, J.M., and K.K. Moorhead. 1991. Mapping pocosins and associated wetlands in North Carolina. Wetlands 2(Special Issue):377-389.
7. Brinson, M.M. 1993. A hydrogeomorphic classification for wetlands. U.S. Army Corps of Engineers Waterways Experiment Station Report WRP-DE-4. Washington, DC.
8. Khorram, S., H. Cheshire, K. Siderelis, and Z. Nagy. 1992. Mapping and GIS development of land use/land cover categories for the Albemarle-Pamlico drainage basin. Report No. 91-08-NC. Raleigh, NC: North Carolina Department of Environment, Health, and Natural Resources.
9. Leopold, L.B. 1974. Water: A primer. San Francisco, CA: W.H. Freeman and Co.
10. Whigham, D.F., C. Chitterling, and B. Palmer. 1988. Impacts of freshwater wetlands on water quality: A landscape perspective. Environ. Mgmt. 12(5):663-671.
11. Novitski, R.P. 1979. The hydrologic characteristics of Wisconsin wetlands and their influence on floods, streamflow, and sediment. In: Greeson, P.E., J.R. Clark, and J.E. Clark, eds. Wetland functions and values: The state of our understanding. Lake Buena Vista, FL: American Water Resources Association.

Decision Support System for Multiobjective Riparian/Wetland Corridor Planning

Margaret A. Fast and Tina K. Rajala
Kansas Water Office, Topeka, Kansas

Kansas has numerous programs that affect riparian corridors and associated wetlands. These programs include planning, monitoring, assistance, research, and regulatory activities. Although administration of these programs often overlaps, integration of program objectives into a holistic, multiobjective approach to resource planning and management has been lacking. A large amount of resource data was routinely collected and compiled, but no effective way had been developed to integrate these data into the decision-making process.

The Kansas Water Office (KWO) was awarded a grant in September 1992 from the U.S. Environmental Protection Agency (EPA) to develop a geographic information system (GIS) decision support system (DSS) that would enable the state to augment its ability to manage riparian/wetland corridors. The project used GIS to differentiate between reaches of a stream corridor to evaluate their environmental sensitivity. The Neosho River basin, one of 12 major hydrologic basins in Kansas, was used as a pilot to demonstrate the feasibility of the concept.

The KWO will use the DSS to help target sensitive areas in the Neosho basin for further planning activities. The project will also benefit other state agencies in their riparian/wetland corridor efforts. The implementation of planning objectives may involve local units of government and, ultimately, private landowners.

Major phases of the project included:

- A needs assessment study
- A feasibility analysis
- A system design
- Construction of the DSS for the Neosho River basin
- A final evaluation of the DSS capabilities

An interagency project advisory group (IPAG), consisting of representatives from eight agencies directly or indirectly involved in riparian and wetland protection activities, was formed to assist in project design and evaluation.

Major steps involved in designing the DSS included:

- Selection and GIS development of databases used for riparian corridor evaluation.
- Creation of riparian corridor segments.
- Development of an analysis methodology to apply to corridor segments.
- Evaluation of the DSS.

Databases Selected for Decision Support System Development

Many types of data were reviewed for the DSS. Several were not used due to the costs associated with geographically referencing the data, given the current data format.

The databases listed in Table 1 are available in the DSS.

During the system design phase of the project, the IPAG identified the need to develop a pilot study area for the DSS. The IPAG had difficulty understanding how a DSS would use geographically referenced data sets (coverages). Before committing to a design for the development of a basinwide system, the IPAG decided first to develop a pilot study area, with a specific focus (application), that could be on-line and demonstrated early. This would allow time for further refinement of the scope of work and identification of coverages to be developed prior to basinwide development of the DSS. For the pilot study application, the IPAG chose to assess the value and vulnerability of the riparian areas in two 11-digit hydrologic unit code (HUC11) watersheds to allow the user to evaluate a corridor segment and compare between segments and to prioritize or target segments for further planning activities.

As development of data layers progressed for the pilot, the IPAG quickly determined that the DSS project

Table 1. DSS Database List

DSS Name	Data Description	Source
Boundary	Neosho River basin boundary	Soil Conservation Service (SCS) HUC11 drainage basins; 1:100,000-scale[a]
Buffer	Riparian corridor	Original buffer on mainstem Neosho and Cottonwood; 147 corridor segments split on tributary confluences
Channels	Stream channelization	Division of Water Resources (DWR) legal description of locations
Con_ease	Conservation easements	Locations of important natural resources that could be purchased by the state from willing landowners for conservation protection
Contam	Water contamination	Kansas Department of Health and Environment (KDHE) contamination locations[a]
Corridor	Riparian corridor	Final riparian corridor; 63 corridor segments developed from HUC11 boundaries
County	County boundaries	Kansas Geological Survey (KGS) cartographic database; 1:24,000 scale[a]
Dams	Dam structures	DWR legal descriptions of locations
Dwrapp	Water appropriations	DWR legal descriptions of locations[a]
Gages	United States Geological Survey (USGS) stream gaging stations	USGS latitude-longitude descriptions; GIS cover developed by USGS
Geology	Surface geology	KGS 1:500,000-scale[a]
Huc11	11-digit hydrologic unit boundaries	SCS HUC11 drainage basins; 1:100,000-scale[a]
Hydr100k	Hydrology	USGS 1:100,000-scale digital hydrology[a]
Kats	Kansas water quality action targeting system	KDHE target valuable and vulnerable scores by HUC11 drainage basin
Landc	Land cover	1:100,000-scale developed from satellite imagery by the Kansas Applied Remote Sensing Program, University of Kansas[a]
Lc_stats	Land cover statistics	Summary statistics on land cover by corridor segment
MDS	Minimum desirable stream flow monitoring gages	Subset of USGS gaging stations
NPS	Nonpoint source pollution	Target watersheds identified in the Kansas Water Plan
Perenial	Perennial hydrology	Reselected perennial streams from 1:100,000 USGS digital hydrology
Pop	Population	Urban land cover (from landc) with 1980 and 1990 Census population data
PPL	Populated places	Geographic names information system (GNIS) entries for Kansas; GIS cover developed by USGS
Publand	Public lands	State and federally owned land digitized from 1:100,000-scale USGS quad maps
Roads	Roads	USGS 1:100,000-scale digital roads[a]
Sections	Section corners	KGS cartographic database; 1:24,000-scale[a]
Streamev	1981 stream evaluation	U.S. Fish and Wildlife stream evaluation study; Kansas Department of Wildlife and Parks (KDWP) provided data on paper maps
T_and_e	Threatened and endangered species	Stream locations of state and federal identified threatened and endangered species; KDWP provided data on paper maps
Temussel	Threatened and endangered species	Locations of state endangered floater mussels; KDWP provided data on paper maps
Tigrcity	City boundaries	U.S. Census 1:100,000-scale TIGER line data; boundaries only, areas not named (use with PPL)
Twp	Townships	KGS cartographic database; 1:24,000-scale[a]
Watrfowl	Water fowl locations	KDWP locations and counts of annual waterfowl migration; data developed from paper maps (Restrict public distribution of data per KDWP request.)
Wq_eff	Water quality: effluent	KDHE sampling sites; GIS cover developed by KDHE
Wq_grnd	Water quality: ground water	KDHE sampling sites; GIS cover developed by KDHE
Wq_lake	Water quality: lake	KDHE sampling sites; GIS cover developed by KDHE
Wq_strm	Water quality: stream	KDHE sampling sites; GIS cover developed by KDHE

[a] Data available at the Kansas Data Access and Support Center (DASC).

parameters would have to be limited to the riparian corridor along the mainstem of the Neosho and Cottonwood Rivers. The costs associated with developing riparian corridor segments for all perennial waters in the Neosho basin was far greater than the available funding.

Creation of Riparian Corridor Segments

A buffer width of one-half mile (one-quarter mile from each stream bank) for the mainstem of the Neosho and Cottonwood Rivers was used to produce the riparian corridor. If more time and funding had been available, riparian corridors for all perennial streams in the Neosho basin could have been developed. The development of this second view of data, organized by the HUC11 watershed, would then have been useful for individual watershed analysis because all perennial streams in the watershed could be analyzed.

The intersection of the HUC11 basin boundaries segmented the corridor. In several instances, small sliver polygons were produced where the HUC11 boundary paralleled the river within the 1/4-mile corridor. The sliver polygons were dissolved into the majority HUC11. In other words, this project assumed that the 1/4-mile corridor buffer was more accurate and useful than the 1:100,000-scale HUC11 boundary.

Many of the HUC11 boundaries that the Soil Conservation Service (SCS) developed actually follow the course of the Kansas streams, rather than intersect them. When this occurred along the Neosho and Cottonwood Rivers, we found that the resulting opposing corridor segments did not always balance with an equivalent length. Also, some HUC11 boundaries would first follow the river, then cross the river. This resulted in corridor segments that encompass both sides of the river for a portion of the segment and follow only one side of the river for another portion of the segment. To address these situations, the KWO arbitrarily added intersections to create equivalent left and right bank corridor segments and to create corridor segments that encompassed either one side of the river or both sides of the river.

Once the corridor segments were finalized and numbered, the corridor segment identification number (corrseg-id) was attached to the other GIS covers. This allows the reselection of data for a given corridor segment, using Boolean expressions in the DSS.

Development of an Analysis Methodology: Land Use

The IPAG determined that one of the most significant factors associated with the quality of the riparian corridor is land cover. Land cover was analyzed for the riparian corridor segments; the GIS cover lc_stats contains summary statistics for each corridor segment. The calculations discussed in the following paragraphs identify the data found in the lc_stats cover. Due to the size of the land cover data set in the Neosho River basin, the DSS includes only the land cover within the riparian corridor.

One way of identifying corridor segments in need of protection or remedial action is to determine the ratio of the number of acres in the corridor segment that contain the preferred riparian land cover types (grasses, woods, and water) to the number of acres that contain the least preferred types of land cover (crops and urban areas). The corridor segments can then be ranked according to that ratio.

Other calculations are useful:

- *bad_pct:* percentage of the corridor segment that contains crop and urban land cover types.
- *bad_tbad:* percentage of all crop and urban land cover for the entire riparian corridor that resides in the corridor segment.
- *'type'_pct:* percentage of the corridor segment that is crop, grass, wood, water, and urban. 'Type' refers to each of the five land cover types; lc_statsuses: a separate value for each (e.g., crop_pct).
- *'type'_t'type':* percentage of each type of land cover for the entire riparian corridor that resides in the corridor segment (e.g., crop_tcrop).
- *'type'_acres:* total acreage of each type of land cover in the corridor segment (e.g., crop_acres).
- *good_acres:* total acreage of grass, wood, and water in the corridor segment.
- *bad_acres:* total acreage of crop and urban in the corridor segment.

Another significant benefit of the DSS is the ability to see where the land cover types are in relation to the river. As an example, the ability to identify corridor segments that have crop land extending to the river on both banks is useful because they are the segments most vulnerable to bank erosion. Those segments can then be targeted for further remedial activities planning.

Decision Support System Requirements

The DSS data sets were developed and analyzed using ARC/INFO on a UNIX-based workstation. The final covers were then exported and transferred to a microcomputer for use in ARC/VIEW. Hardcopy prints are printed to a Tektronix Phaser III color wax printer with 18 Mb of RAM, running in Postscript mode.

The DSS data sets total 26 Mb. ARC/VIEW version 1 requires 8 Mb of RAM to load the program. To run the DSS efficiently, a 486DX-66 with 16 Mb of RAM is preferred. The DSS is slower on a 486DX-33 with 8 Mb of RAM. It was not tested on any other PC configuration, so a configuration in between the two may be satisfactory.

Processing GIS Data

Reselecting the perennial streams in the Neosho basin and further identifying the mainstem of the Neosho and Cottonwood Rivers using United States Geological Survey (USGS) 1:100,000-scale hydrography can be time consuming. Perhaps the River Reach III covers should replace that data in the future.

Attaching census data to the urban land cover polygons, as was done for the pop cover, is not recommended. Use of the TIGER line files and cover would give a more accurate distribution of the population. Because less than 5 percent of the riparian corridor had urban land cover, the KWO did not use the pop cover in its evaluation. Several summary covers of TIGER and census data will soon be available from DASC.

Clipping the other ARC/INFO covers to the Neosho basin and attaching the corrseg-id, using the identity command, was unremarkable.

Processing Non-GIS Digital Data

Channels and dams were in digital format but were not in ARC/INFO format. The files were processed using the LeoBase conversion software from the KGS, then generated into ARC/INFO covers. Some records were lost in the conversion. The LeoBase program fails to convert, or incorrectly converts, legal descriptions for sections that do not have four section corners (e.g., northeastern Kansas). The Division of Water Resources is in the process of attaching latitude-longitude to the point locations. Processing these data should take only a few hours at most.

Processing Nondigital Data

Several covers were developed on contract from paper maps or legal descriptions. They were: conservation easements (con_ease), public lands (publand), stream evaluation (streamev), threatened and endangered species (t_and_e and temussel), and water fowl (watrfowl). Most of the data for these covers were drafted on a 1:100,000-scale USGS quad map and digitized. The stream evaluation data were developed using a scanned paper map of the coded streams as a backdrop for the 1:100,000-scale hydrography; the digital streams were reselected and coded.

In summary, KWO's GIS personnel needed approximately 275 hours to develop the riparian corridor segments, process the land cover data and summary statistics, export the covers, transfer and import the covers for ARC/VIEW, and assist in the development and presentation of the DSS demo. Contract personnel spent approximately 183 hours developing GIS covers for the DSS. This does not include the time spent identifying the perennial and mainstem hydrology in the USGS 1:100,000-scale hydrology.

Final Evaluation of the Decision Support System

In its final evaluation of the system, the IPAG determined the system to be useful and an excellent start at consolidating a variety of data that have application for riparian corridor/wetland issues. Many IPAG members found ways to use the DSS in their own programs. Additional comments on the system evaluation are as follows:

- Concern about the lack of complete wetland data. The land cover data available could not identify wetland areas.
- Need for more detailed woodland data. Again, the resolution of the land cover data precluded detailed identification of woodland areas. The Kansas Biological Survey (KBS), the KWO, and EPA are now pursuing options to develop more detailed land cover data, including wetlands and woodlands.
- The lack of information on the tributaries did not allow full basin analysis, which would be desirable. This issue is addressed in the "construction" discussion above.
- Desirability of expanding the project with elevation and temporal data.
- Lack of definition of the floodplain. Federal Emergency Management Agency (FEMA) floodplain data are not easily incorporated into a GIS. Other options, including satellite imagery of the flood of 1993, will be evaluated.
- Project development requires extensive communication between program people and GIS technicians. This can be a daunting task due to the technical vocabularies involved and the many other ongoing activities of the participants.
- Consideration of the requirements for transferring the project to other potential users. GIS applications generally use large databases. User microcomputers may not have the CPU, RAM, and storage capacity necessary for the DSS application and often have a limited number of options for data transfer.
- Concern about costs and time associated with the expansion of the DSS to other basins in the state. This project was focused on one of the 12 major hydrologic regions in Kansas. Funding options, project scope, and system refinements based on the physical characteristics of the other basins need to be pursued.

The KWO learned that clearly defining a single DSS application at the outset of the project is critical. The KWO originally believed that the DSS could be developed with general descriptions of the broad range of program applications, utilized by multiple agencies, that could benefit from the DSS. Each participating agency could bring its programs and needs to the IPAG for discussion; the resulting DSS would then serve those multiple programs and needs. Instead, the ambiguity of the objective confused the IPAG. Once the IPAG chose to focus on a single application, the assessment of riparian corridor value and vulnerability to target priority areas for further planning activities, the IPAG became more confident in its advisory role. Upon completion of the project, the IPAG members could readily identify how the DSS could be enhanced, modified, or directly used in their own programs.

Design of GIS Analysis To Compare Wetland Impacts on Runoff in Upstream Basins of the Mississippi and Volga Rivers

Tatiana B. Nawrocki
Natural Resources Research Institute, University of Minnesota, Duluth, Minnesota

Introduction

The attention given in hydrologic studies to wetlands differs significantly between the United States and Russia at the present time. Fundamental theories and mathematical models are developed in both countries to describe hydrologic processes and impacts of watershed conditions on surface runoff. In the United States, however, theoretical investigations are directly pointed at wetlands and are supported by large-scale field studies and advanced technological capabilities to manage spatially distributed information. Unlike in Russia, in the United States, special scientific symposia are devoted to wetland hydrology, where major tasks for hydraulic and hydrologic research needs are formulated. Among these tasks are the understanding and assessment of relationships between various hydrologic modifications and wetland functions, especially wetland flood conveyance and water quality protection functions (1). Watershed-scale comprehensive field studies of wetland functions are underway, for example, at constructed experimental wetlands in the Des Plaines River basin in Illinois (2). A new long-term goal—strategic restoration of wetlands and associated natural systems—has been formulated (3).

The intensive efforts of many U.S. scientists yielded numerous results and attracted more attention to the complicated nature of wetlands processes. Wetlands were evaluated as runoff retention basins, and it was found that, in northwestern states, up to 12 inches of water could be accumulated per wetland acre (4, 5). Over time, piecemeal loss and degradation of wetlands in many areas of the United States have seriously depleted wetland resources. Researchers also discovered that adverse impacts from wetland degradation could appear indirectly with little obvious spatial or temporal connections to sources. As described by Johnston (6):

> Cumulative impacts, the incremental effect of an impact added to other past, present, and reasonably foreseeable future impacts, has been an area of increasing concern. . . . Impacts can accumulate over time or over space and be direct or indirect. An indirect impact occurs at a location remote from the wetland it affects, such as the discharge of pollutants into a river at a point upstream of a wetland system.

The process of solving environmental problems related to wetlands is increasingly complex. Analyzing diversified data over increasingly broad areas becomes essential for making competent decisions.

Comparing wetland hydrologic functions in headwaters of the Mississippi River (United States) and the Volga River (Russia) could provide additional information about how alternative management strategies affect runoff, peak flow, and water quality under changing climates. A macro-scale “field experiment” in both of these naturally similar areas is already under way. Wetland conservation as opposed to drainage is now the prevailing policy in the upper Mississippi basin. In Russia, however, economic problems have prevented this type of policy from becoming a priority. Instead, peat mining, reservoir construction on lowlands, and drainage for farming and private gardening are common.

This project, which is being implemented at the Natural Resources Research Institute (NRRI), University of Minnesota, Duluth, has the following goals:

- Developing a multilayered hierarchical base of geographic information system (GIS) data for headwater watersheds of the Mississippi and Volga Rivers.
- Developing a comparative analysis of wetland impacts on the hydrology of the rivers.
- Studying the relationships between natural and human-induced factors on wetland functions under climate change and variable strategies of wetland conservation.

- Defining criteria and thresholds for wetland system stability with regard to flood risk and water quality.
- Outlining recommendations for wetland management in the headwaters.

The methodology for comparative assessments involves statistical analysis, hydrologic models, GIS, and remote sensing. Representative watersheds will be studied in more detail, and procedures for scaling information from the local to the regional level will be developed.

Input Data

In recent years, U.S. governmental and state agencies, as well as a number of private companies, have expended considerable efforts to compile the available data in GIS format for multidisciplinary analysis of watershed problems. Among the major sources of information essential for studies of wetland hydrologic functions are:

- The National Wetlands Inventory, conducted by the U.S. Fish and Wildlife Service.
- Digital elevation maps (DEMs) developed by the U.S. Geological Survey (USGS).
- Major and minor watershed boundaries, outlined for Minnesota by the Minnesota Pollution Control Agency (MPCA).
- The water quality sampling network from the U.S. Environmental Protection Agency (EPA).
- The digital chart of the world (DCW), issued by the Environmental Science Research Institute (ESRI) in scale 1:1,000,000.

These and other sources, listed in Table 1, were used to compile the map illustrations for this paper.

Almost no similar data in GIS form could be found for the territory of the former USSR, however. Any specific data (e.g., detailed maps, hydrology records, water quality sampling data) are generally in paper files dispersed among many agencies and are hard to obtain. The forms of information storage and means of its analysis are out of date, and most maps exist in single or few copies in paper files. In Russia, the time lag grows between the dynamic changes in the environment and the traditional pattern and inertia of management structures.

The GIS situation in Russia developed some interesting paradoxes. During the first few decades of space programs, certain state agencies accumulated an outstanding bank of world image data. When economic hardships hurt the previously privileged space industry, numerous joint ventures with foreign companies were created to distribute images on the world market for hard currency. These data are hardly available for domestic uses, however. The current domestic price for image data is 20,000 to 25,000 rubles for a black-and-white picture of an area 60 by 60 kilometers, or 60,000 to 70,000 rubles for the same image on a computer disk. With the present level of funding for scientific research, the price is too high. Security regulations still restrict access to later data, showing land use changes.

Another paradox is scientists' attitude toward their data. Abandoned by the state, agencies and institutes are reluctant to share their specific data in multidisciplinary projects. Data files are now a commodity. Accomplishing an overlay and integration of special data coverages, which is essential to any watershed GIS study, is almost impossible.

The third paradox is the attitudes of local, regional, and central authorities toward GIS. Many authorities are still ignorant about the potential of this technology. Those who are knowledgeable prefer not to promote GIS for watershed-related tasks because it involves land use analysis. With the onset of land privatization, the best pieces of property (e.g., the waterfront lots adjacent to drinking water reservoirs in the Moscow region) are rapidly allocated to the most powerful landowners. Thus, limiting access to this kind of information is deemed safer.

Experts in Russia have not yet applied GIS to wetland hydrology studies because GIS is still a very new and mostly unfamiliar technology. This makes the current study unique both for its results and for its application of GIS methodology.

Closer review of data sources indicates that most of the input data for the project is available, though dispersed among many agencies. Table 1 is a preliminary list of data and data sources.

Project Design

The project addresses the following questions:

- How do the extent and positioning of wetlands in the headwaters of large rivers affect runoff and peak flow?
- What are the spatial relationships between wetland and other land uses regarding flood risk and water quality under variable climate conditions?
- What is the role of wetlands for diffuse pollution prevention and sediment deposition control under alternative management?
- What determines major criteria for wetland conservation in headwaters, ensuring environmentally sustainable development under multiobjective land and water resource uses?

Table 1. **Data Sources for a GIS Study of Wetlands in the Basins of the Mississippi (United States) and the Volga (Russia)**

Data **Level 1**	**United States**	**Russia**
Base maps	DCW[a]	DCW[a]
Stream network	DCW[a]	DCW[a]
Urban and rural areas	DCW[a]	DCW[a]
Wetlands, unclassified	NWI[a]	DCW[a]
Forests	DCW[a]	MGU[b]
Agricultural lands	LMD[b,c]	MGU[b]
Level 2		
Watershed boundaries	MPCA[a]	RWRC[c]
Digital elevation maps	USGS[a]	NA
Digital orthophotos	USGS, LMIC[a]	CD, RPI[b,c]
Soils	SCS[b,c]	RPI[c]
Hydrologic records	USGS[b]	CHM[b,c]
Water quality records	EPA/MPCA[a,b]	CHM, RCP, RPI[c]
Land uses	LMIC, LSAT[a]	LSAT, CD, RPI[a-c]
Wetlands, classified	NWI[a]	RPI[c]

[a]Data available in GIS ARC/INFO format.
[b]Databases; needed conversion to ARC/INFO.
[c]Data available in paper files; needed digitizing.

Key:

CD = Commercial distributors
CHM = Russian Committee on Hydrometeorology
DCW = ESRI digital chart of the world
EPA = U.S. Environmental Protection Agency
LSAT = Satellite image data
LMD = Published literature and map data
LMIC = Minnesota Land Management Information Center
MGU = Moscow State University
MPCA = Minnesota Pollution Control Agency
NA = Data not available
NWI = U.S. National Wetland Inventory
RCP = Russian State Committee on Natural Resources and Conservation
RPI = Miscellaneous planning agencies and research institutes
RWRC = Russian Water Resources Committee
SCS = U.S. Soil Conservation Service
USGS = U.S. Geological Survey

Tasks established to address these questions include:

- Developing a multilayered hierarchical base of GIS data for headwater watersheds of the Mississippi and Volga basins.
- Performing a comparative analysis and simulation of wetland impacts on hydrology and water quality at representative watersheds.
- Deriving the relationships between natural and human-induced factors and wetland functions under climate change with regard to variable strategies of wetland conservation.
- Defining the criteria and thresholds for wetland system stability with regard to flood risk and water quality parameters.
- Outlining recommendations for land and water resources management and wetland positioning in the headwaters.

GIS is the essential tool for manipulating and integrating the many types of spatial data on water resources, soils, vegetation, land use, economics, and the environment. GIS compiles many sources (e.g., maps, field notes, remote sensing, statistical data) into a consistent, interpretable database used for specific scientific goals and development decisions. The user can run GIS ARC/INFO software on workstation and PC platforms and apply the hierarchical approach to GIS data management, developed earlier (7). At the task level, data resolution and corresponding modeling tools vary.

Level 1 contains the basic reference information for large regions (e.g., the Upper Volga and the Minnesota portion of the Upper Mississippi River basins). It covers an area of several hundred thousand square kilometers, with a map scale approaching 1:1,000,000. Landsat thematic map data and the DCW (8) are used as sources of data at this level. Vogelmann et al. (9) demonstrated the methodology for detection of freshwater wetlands using remote sensing data based on maximum likelihood supervised classification. A graphic data file on GIS focuses on basic physical characteristics such as stream network, geology, soils, wetland classification, and other major land uses. A complementary tabular database or attribute file contains information on stream flow, water quality, pollution sources, and wetland impacts on material fluxes. At this level, the general physiographic and statistical information is accumulated and analyzed, territories are classified, and major problems and typical case study watersheds are defined. This information is compiled from literature, cartographic data in paper and digitized form, statistics, and space image data.

In Level 2, the more detailed GIS analysis and scenario-based modeling is implemented at the watershed scale with a map scale of approximately 1:25,000. The watershed demonstration focuses on alternative approaches to priority-setting in wetland management, climate impact analysis, and resulting interactions with landform, soils, biosphere, and runoff. The sources of data are special, topographic maps and air photo interpretation.

Simulation studies of water balance and fluxes among the various reservoirs are implemented at Level 2. Developing procedures for scaling information from the local to regional level is the important task at this level. GIS assists in handling the input parameter library and analyzing the output. GIS studies, involving area measurements and distribution analysis, evaluate cumulative impacts on runoff and its quality from the loss of wetland area, caused by drainage or filling, under stationary and changing climate.

Wetland functions are considered under two sets of scenarios. Management scenarios compare different wetland and farming allocations, conservation practices, and agricultural chemical use. Climate scenarios assume rainfall and temperature changes under global warming. Scenario-based simulation is applied in the analysis of watershed runoff, wetland moisture regime, soil erosion, and water quality processes.

Methodology

GIS database structure is related to the selected methodology. GIS serves as a linking tool for input-output data analysis and transfer between models, used at different levels and stages of studies.

Scientists in both the United States and Russia developed statistical methods to obtain quantitative relationships between stream flow and wetland area in the river basins. Johnston (6) summarized the U.S. findings:

> Empirical equations for predicting streamflow, developed by U.S. Geological Survey in Wisconsin and Minnesota, indicate that flood flow is proportional to the negative exponent of wetlands and lakes ratio on a watershed (10, 11). This means that relative flood flow is decreased greatly by having some wetlands in a watershed, but a watershed with a large proportion of wetlands does not reduce flood flow much more than a watershed with an intermediate proportion of wetlands. For example, predicted flood flow was 50 percent lower in Wisconsin watersheds with 5 percent lakes or wetlands than it was in watersheds with no lakes or wetlands, but increasing the proportion of lakes and wetlands to 40 percent decreased relative flood flow by only an additional 30 percent (12).

Other estimates agree that wetland encroachment on a watershed of less than 25 percent generally has a minimum influence on peak flow (5, 13, 14).

Johnston and colleagues (15) applied these equations to watersheds in central Minnesota. They found that a watershed with 1.6 percent lakes and wetlands had a flow per unit watershed area that was 10 times the flow predicted for a watershed with 10 percent lakes and wetlands, while watersheds with 10 to 50 percent lakes and wetlands had about the same flood flow per unit area.

Statistical analysis indicates that peak discharge increases with decreasing wetland area within the drainage basin. The regression equation defines the approximation for northwestern Minnesota (16):

$$Q_{AM} = 58.4\, A_W^{0.677}\, (L_S)^{-0.506}$$

$$L_S = 100(A_L + A_M)/A_W + 1$$

where:

Q_{AM} = arithmetic mean of the annual series, cubic feet per second.

A_W = watershed area, square miles.

A_L = lake area within the watershed, square miles.

A_M = marsh area within the watershed, square miles.

A similar statistical approach was developed for peak flow determination in Russia. Maximum flow discharge from snow melt is calculated for the central European zone as (17):

$$Q_m = K_0 * h_p * S_1 * S_2 * S_3 / (A + 1)^n$$

where:

Q_m = flow discharge, cubic meters per second.

K_0 = coefficient of flood concurrence, K0 = 0.006 for plain river basins.

h_p = calculated flood runoff for given probability, millimeters.

A = drainage area, square kilometers.

n = coefficient, n = 0.17.

S_1 = lake storage coefficient, if lake area is less than 1 percent of A, then S1 = 1.

S_2 = pond and reservoir storage coefficient, S_2 = 0.9 with ponds and S_2 = 1 without ponds.

S_3 = combined wetland and forest storage coefficient.

S_3 = 1 - 0.8 lg (0.05 Sf + 0.1 S_w + 1), where S_f and S_w are forest and wetland area, percentage to total drainage area.

Calculated flood runoff for a given probability, h_p, is determined based on average flood runoff h, millimeters, coefficient of variation C_v, and tabulated parameter F, as follows:

$$h_p = (1 + F * C_v)\ h$$

$$h = K_t\ h_k$$

h_k = 100 millimeters for Moscow region.

K_t = land surface coefficient, Kt = 0.9 for plains and sandy soils, K_t = 1.1 for hills and clay soils.

The studies mentioned above generally agree with an assumption that the incremental loss of wetland area would have a small effect on flood flow from watersheds with 10 percent up to 40 to 50 percent wetlands, but a large effect on flood flow from watersheds with less than 10 percent wetlands.

The existence of similar thresholds was found in relation to wetlands abilities to intercept pollutants. As Johnston stated (6),

> The same 10 percent threshold was identified by Oberts (18) for suspended solids, a measure of water quality function. Stream-water draining watersheds having 10 to 20 percent wetlands had about the same loading of suspended solids, so the contribution of suspended solids was relatively constant per unit area of watershed. However, the watersheds with less than 10 percent wetlands had loading rates per unit area that were as much as 100 times greater than the loading rates from the watersheds with more than 10 percent wetlands.

GIS could be especially helpful in determining the impacts on downstream water quality of the spatial positioning of wetlands within watersheds. Studies prove that the location of wetlands can affect their cumulative function with regard to water quality (6). In an earlier work (15), Johnston developed an index of wetland location and applied it to a landscape-level GIS study of urban and rural stream watersheds in central Minnesota. The index is formulated as:

$$PWP = \sum_{i=1}^{j} A_i i / \sum_{i=1}^{j} A_i$$

where:

PWP = relative wetland position.

j = stream order (19) of water quality sampling station.

i = stream order of wetland.

Ai = area of *i*th order wetlands.

Calculated values for the index ranged from 0 (i.e., all wetlands were on streams of the same order as that of the sampling station) to 2.6 (i.e., average wetland position was 2.6 stream orders upstream of the sampling station). Watersheds with wetlands located close to sampling stations had significantly better water quality (i.e., lower concentrations of inorganic suspended solids, fecal coliform, and nitrate; lower flow weighted concentrations of ammonium and total phosphorus) than watersheds with wetlands located far from sampling stations (6).

The review of methodological approaches, as shown above, indicates that parameters describing wetland extent, positioning, and land surface characteristics are of universal significance for any comprehensive watershed-scale wetland study.

In the current study, GIS is used at Level 1 to evaluate wetland area per watershed and to develop input parameters for relative wetland position assessment. The comparison and selection procedures for case study watersheds in the Volga and Mississippi basins are based on these values. The parameters, derived from GIS, are as follows:

1. Total watershed area.
2. Lake, pond, and reservoir area.
3. Forest area.
4. Wetland area.
5. Ratio of wetland area to total watershed area.
6. Wetland area by subwatersheds of different order.
7. Relative wetland extent by subwatersheds of different order.
8. Land surface coefficients.

Parameters listed in groups 1 through 4 are obtained directly from GIS attribute tables as values of “area” items for the respective land cover polygons. Parameters 5 through 7 require calculations relating values of area items for different polygon coverages. Land surface coefficients (group 8) could be determined indirectly based on basic soil, land cover, and topography data. Most U.S. methodologies use hydrologic soil groups, based on soil permeability, rates of infiltration, and Soil Conservation Service (SCS) runoff curve numbers (20). Some Russian methodologies have adopted similar empirical land surface coefficients. For example, for central European Russia, this value varies from $K_t = 0.9$ for plains and sandy soils, to $K_t = 1.1$ for hills and clay soils (17).

Level 2 of analysis applies two hydrologic simulation models:

- The Agricultural Watershed Runoff and Water Quality Model (Agricultural Nonpoint Source Pollution Model [AGNPS]), developed by the Agricultural Research Service of the U.S. Department of Agriculture, contains explicit procedures to evaluate the impacts of management practices and landscape feature positioning on watershed runoff. AGNPS is a cell-based runoff model that estimates water volume, peak flow, eroded and delivered sediment, chemical oxygen demand, and nutrient export from watersheds (20-22).
- The Forest Runoff Watershed Model (FRWM) combines analytical and numerical methods for solving hydro- and thermodynamics equations (23). This model considers snow melt constituent in runoff in more detail than does AGNPS. Hydrologic simulation is based on physical process descriptions for snow cover dynamics, freezing and thawing of soil, soil moisture dynamics in frozen and thawed soils, interception of liquid and solid precipitations by vegetation, surface runoff, ground-water aquifers, and channeled streams.

Both models use a similar set of watershed input data, derived from GIS (e.g., elevations, slopes, channel slopes, stream network configuration, soil texture, land cover). The methodology, linking GIS with hydrologic models, was already tested in the wetland study project at the Voyageurs National Park in Minnesota. The ARC/INFO GRID module was used to derive watershed variables for input to AGNPS. GIS then presented and interpreted the scenario-based results of the simulation (24). The typical stages of such an analysis and interpretation for a watershed-scale area are presented in Figures 1 through 5.

Case Study Watersheds

The areas where wetland impacts on runoff are evaluated are located in Minnesota (United States) and Moscow and adjacent regions (Russia) (see Figures 6 through 15). They have mixed urban, rural, recreational, and forest land uses. Both regions have a variety of development pressures. The relative effects of different alterations in watershed management are distinguished and quantified. GIS provides metrics for comparative assessments and analysis of related variables for both areas.

Table 2 and Figures 7 through 14 present a general overview of wetland extent in both areas. The case study subwatersheds used for more detailed analysis will include tributaries of the second and third order. At this stage, several watersheds are considered for more detailed analysis. The limitations imposed by data availability as well as by project resources could affect the final selection. Table 2 serves, therefore, as a preliminary overview of several areas that could potentially be adopted for more detailed studies.

GIS analysis shows that in the Upper Volga, wetlands extent very much depends on allocation of populated areas. The heavily urbanized Moscow metropolitan area affects a large territory of many thousands of square kilometers. The ratio of wetlands as a percentage of total land area is one-tenth of that in the neighboring Tver area, which has the same size but a smaller population (see Figure 11). In areas of intensive agriculture (e.g., the Pronya basin located southeast of Moscow), almost all wetlands were drained and have not existed for several decades.

In Minnesota, despite the growing urbanization (e.g., the Twin Cities area [7,330 square kilometers]), about half of the presettlement wetlands still remain (25); wetlands occupy 442 square kilometers, or over 6 percent of the land area; and shallow lakes constitute an additional 114 square kilometers (1.56 percent). Some watersheds within the Twin Cities metropolitan area have a high wetland percentage, such as 18.9 percent in the Lamberts Creek watershed. Intensive studies with GIS application of landscape feature functioning and wetland impacts on stream flow and water quality demonstrated an innovative approach and made detailed databases available for this area (15).

Preliminary comparative analysis indicates that two pairs of case study watersheds could be initially selected for further studies in the Mississippi and Volga basins:

- Upstream watersheds with wetlands area of 15 to 20 percent (Tver region in Russia and Cass and adjacent counties in Minnesota).
- Tributary watersheds downstream with wetlands area of 1 to 2 percent (the Istra basin in Russia and subwatersheds of the Minnesota River basin, located in Sibley, Scott, and adjacent counties in Minnesota).

Case study watersheds in the Mississippi and Volga basins are situated on gently rolling plains in mixed forest zones with southern portions extending into the forest/steppe and prairies. The Quaternary sediments are of glacial, glaciofluvial, lacustrine, and alluvial origin. Wetlands have hydric soils with various degrees of gley process development and/or peat accumulation, varied by wetland type and soil moisture regimen (28). The annual precipitation is 500 to 600 millimeters with similar

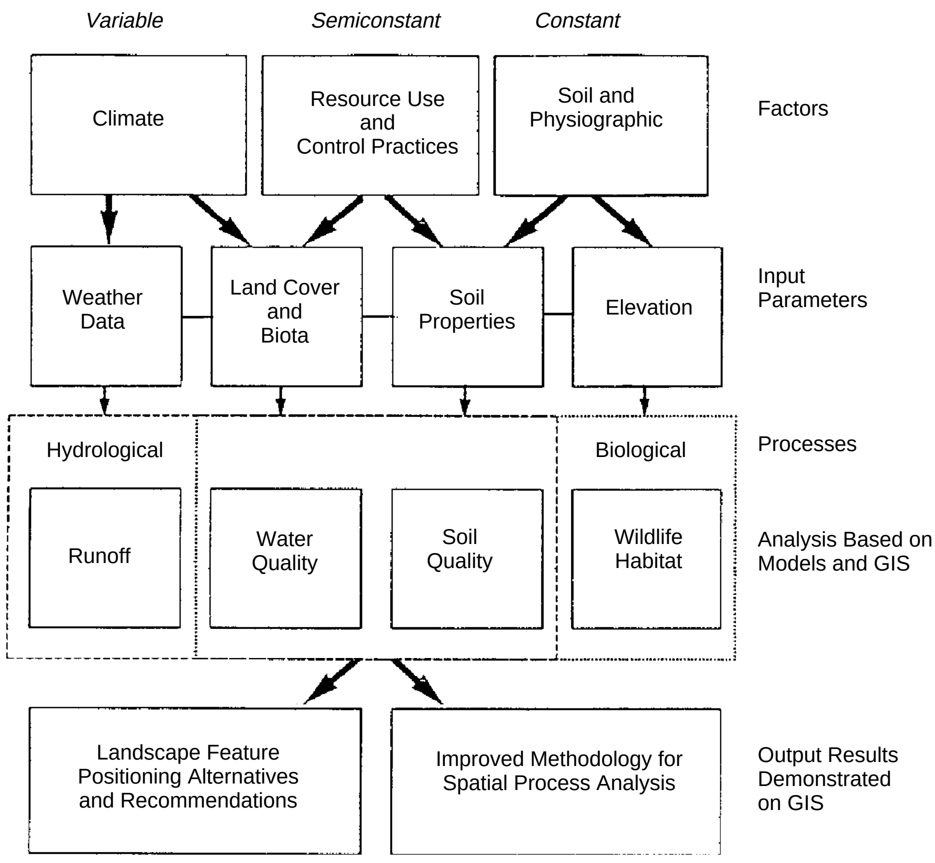


Figure 1. Conceptual framework of linking GIS and models for environmental management.

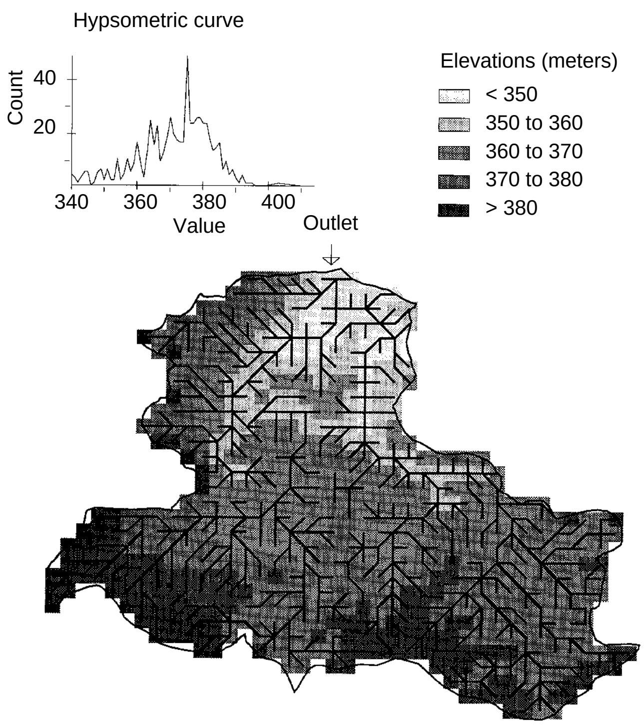


Figure 2. Stream network configuration derived from GIS elevation map.

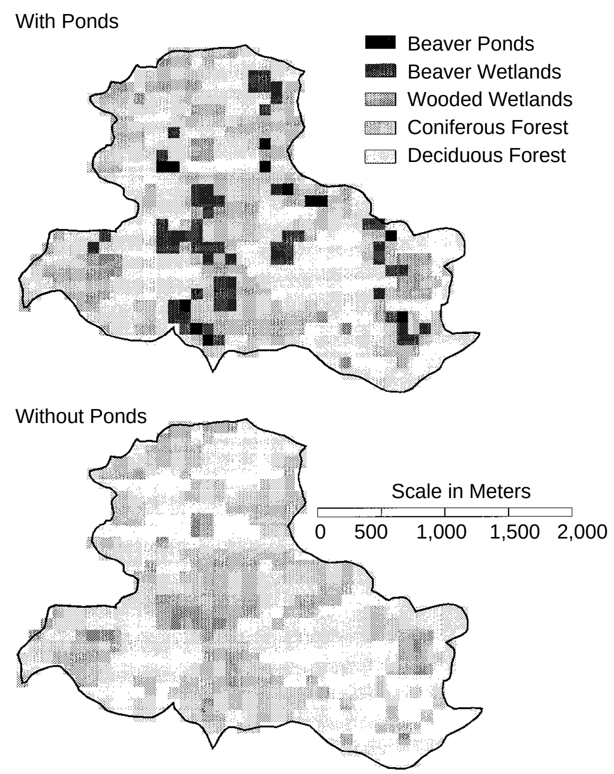


Figure 3. Scenarios of land use.

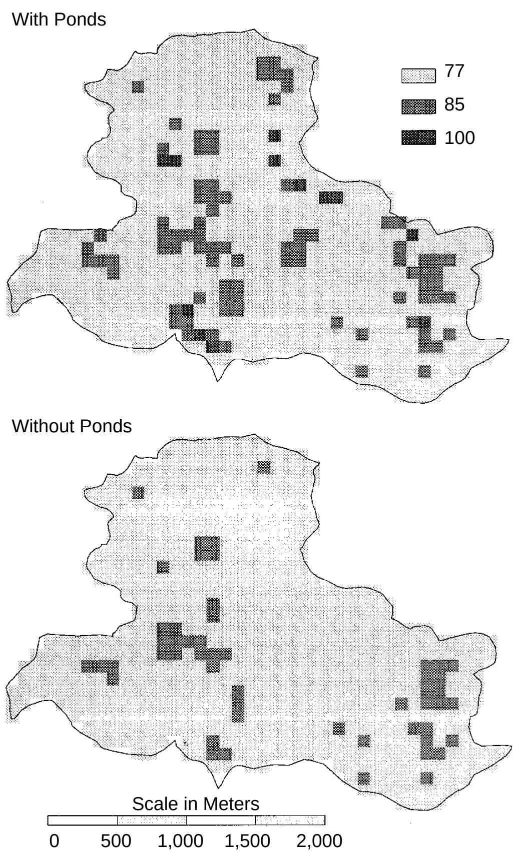


Figure 4. Scenario-related hydrologic curve numbers.

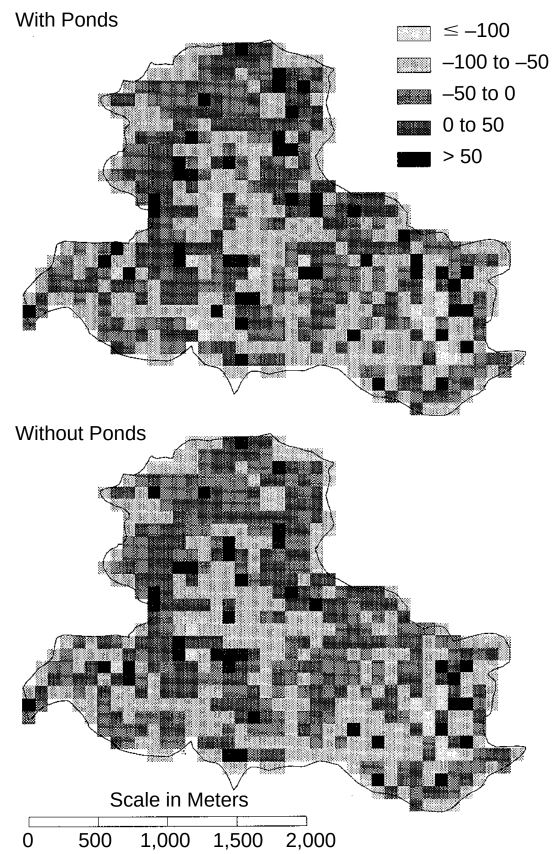


Figure 5. Patterns of sediment transfer between cells (%), + deposition, – erosion.

Table 2. Comparative Data on Wetland Extent in Minnesota and in the Upper Volga Basin (8, 25-27)

Region	Total Area (Square Kilometers)	Wetland Area (Square Kilometers)	Percentage
United States			
Minnesota	205,940.30	30,500.00	14.80
Beltrami Co.	7,923.04	3,909.23	49.34
Cass Co.	6,256.38	1,505.29	24.06
Hubbard Co.	2,624.81	283.22	10.79
Le Sueur Co.	1,204.82	28.31	2.35
Hennepin Co.	1,588.07	36.37	2.29
Sibley Co.	1,555.38	24.26	1.56
Wright Co.	1,852.97	24.27	1.31
Scott Co.	982.56	8.06	0.82
Lambert Creek	9.51	3.69	18.90
Russia			
Tver region	10,000.00	1,169.08	16.90
Moscow region	10,000.00	165.39	1.70
Istra basin	1,827.38	24.07	1.32
Pronya basin	10,200.00	[a]	[a]

[a]Wetland area is insignificant and not identified by available maps.

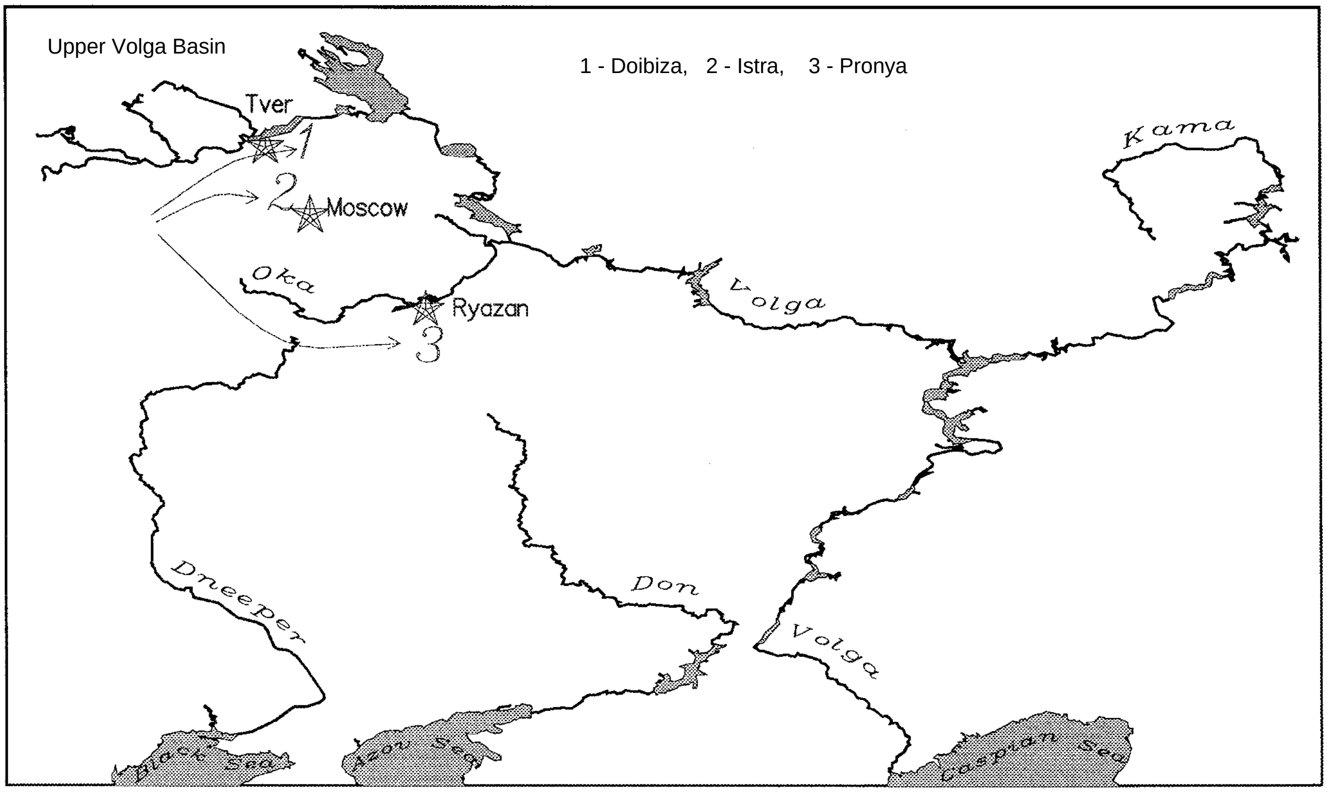


Figure 6. Location of study areas in the Volga basin.

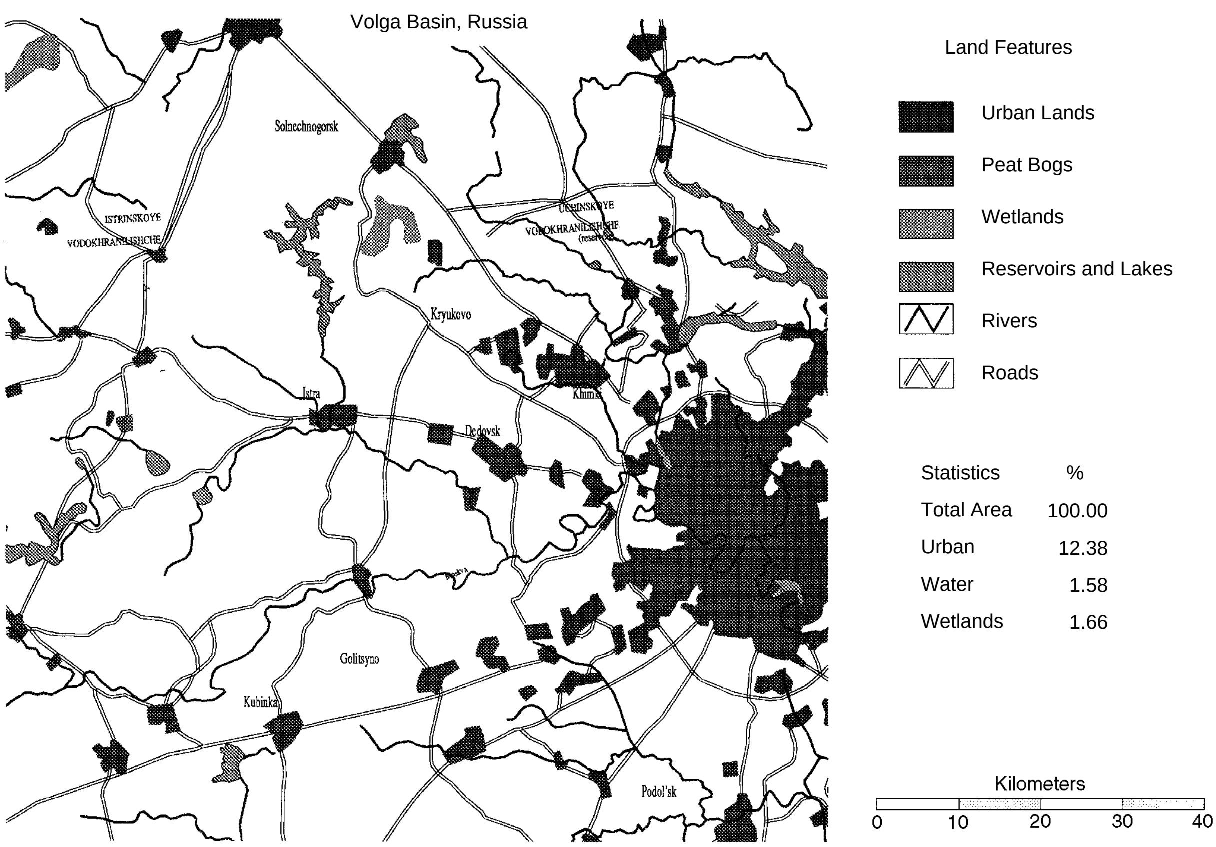


Figure 7. Wetlands and urban lands in the Moscow region.

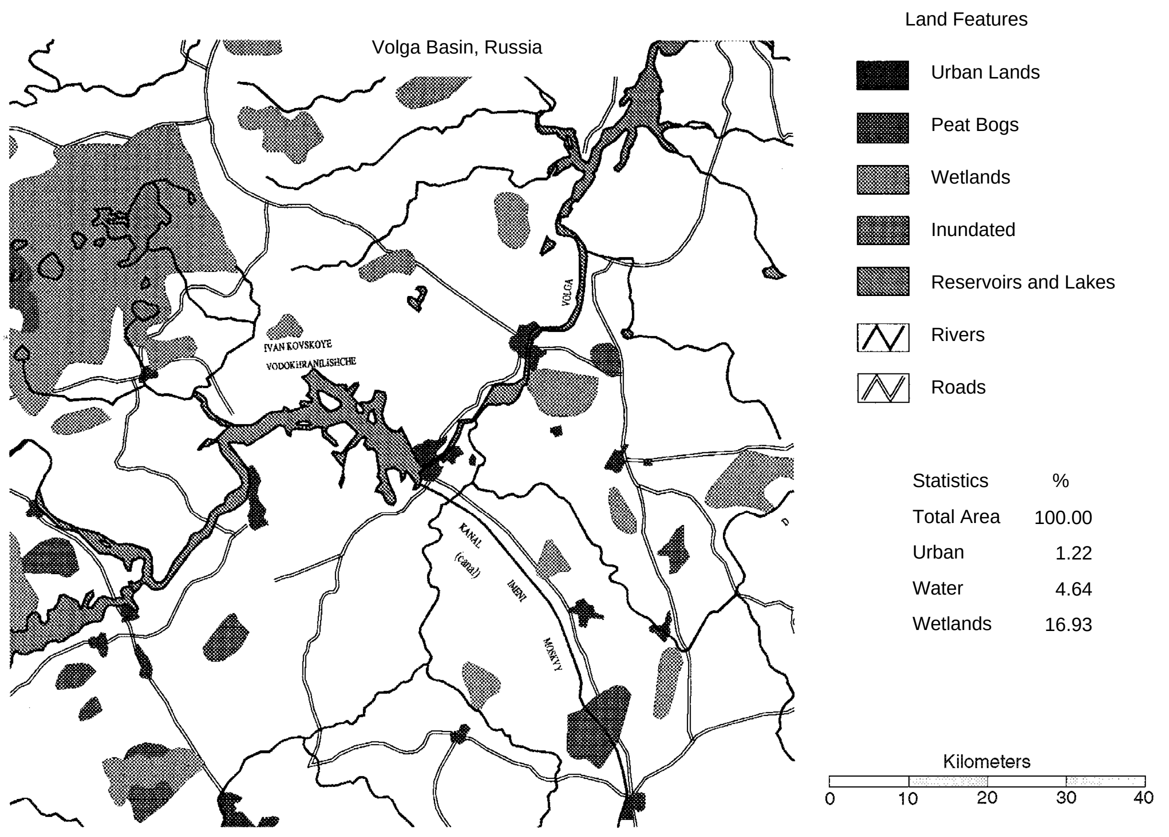


Figure 8. Wetlands and urban lands in the Tver region.

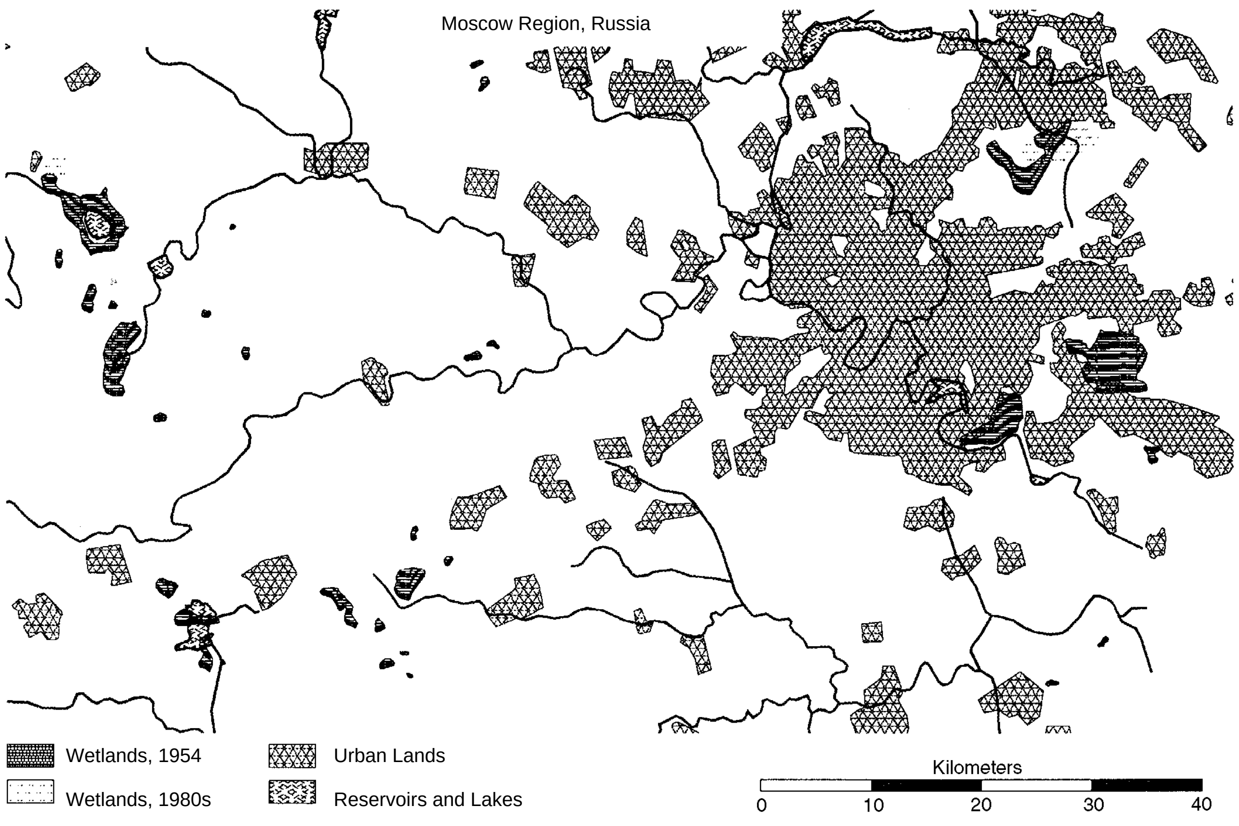


Figure 9. Wetland decline since 1954, Moscow region.

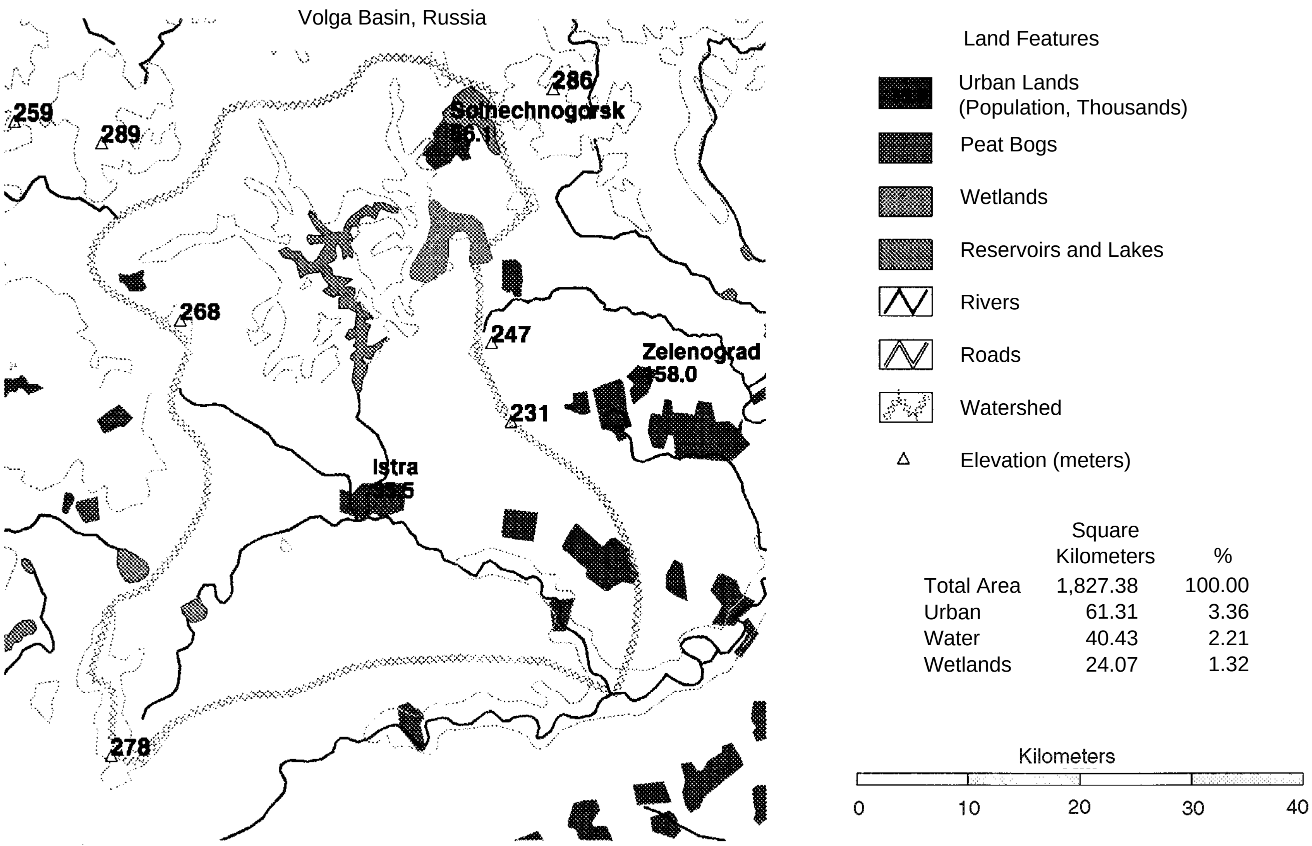


Figure 10. Istra watershed in the Moscow region.

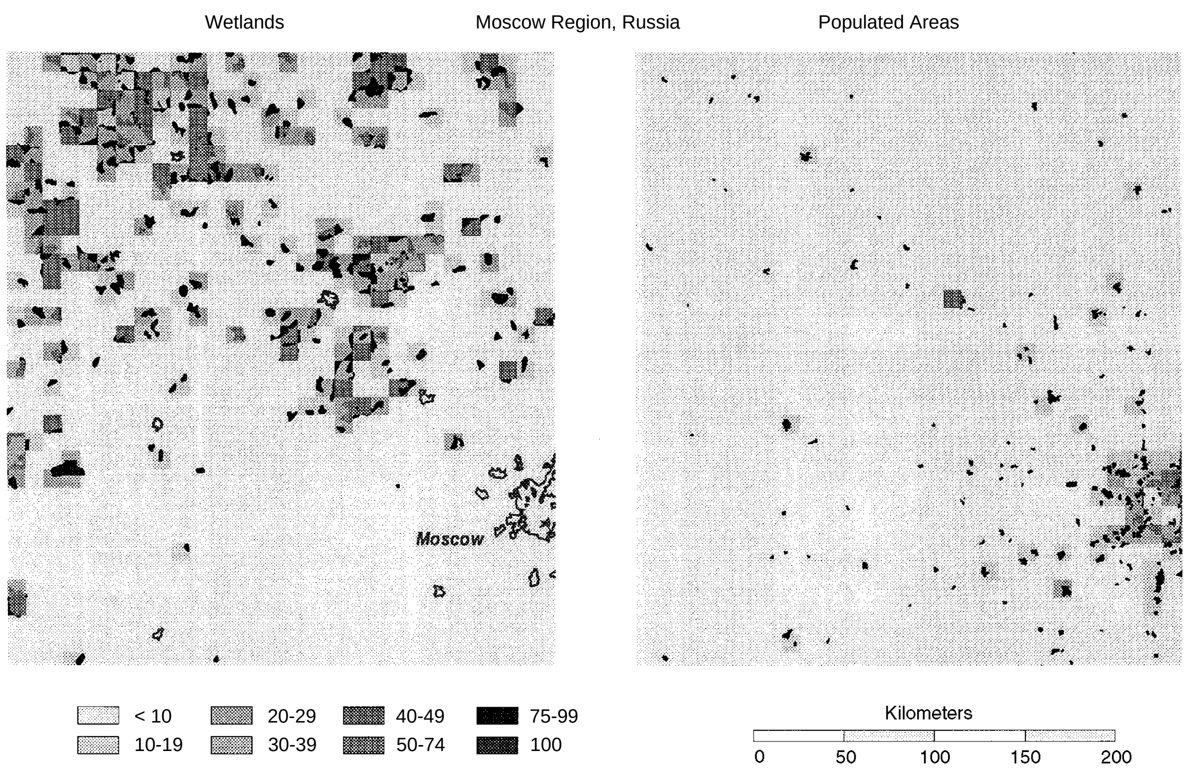


Figure 11. Land cover, percentage of total in Moscow and Tver regions.

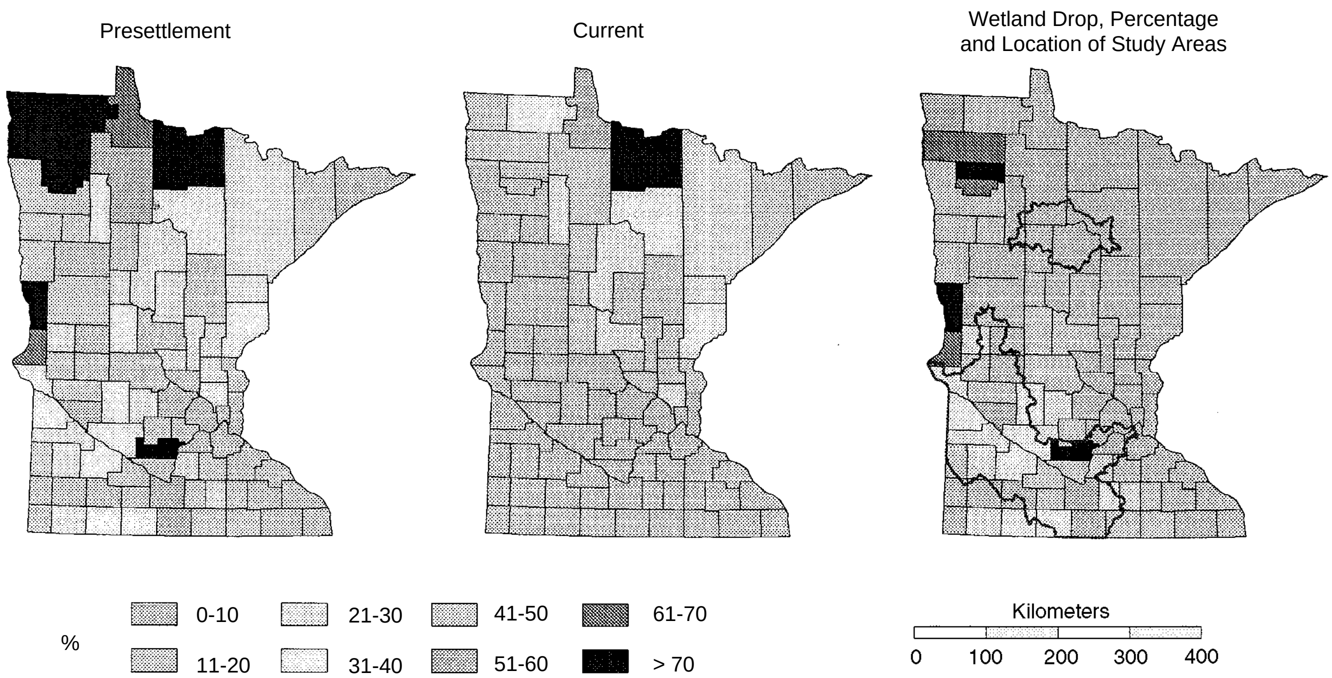


Figure 12. Wetlands in Minnesota, percentage of total area.

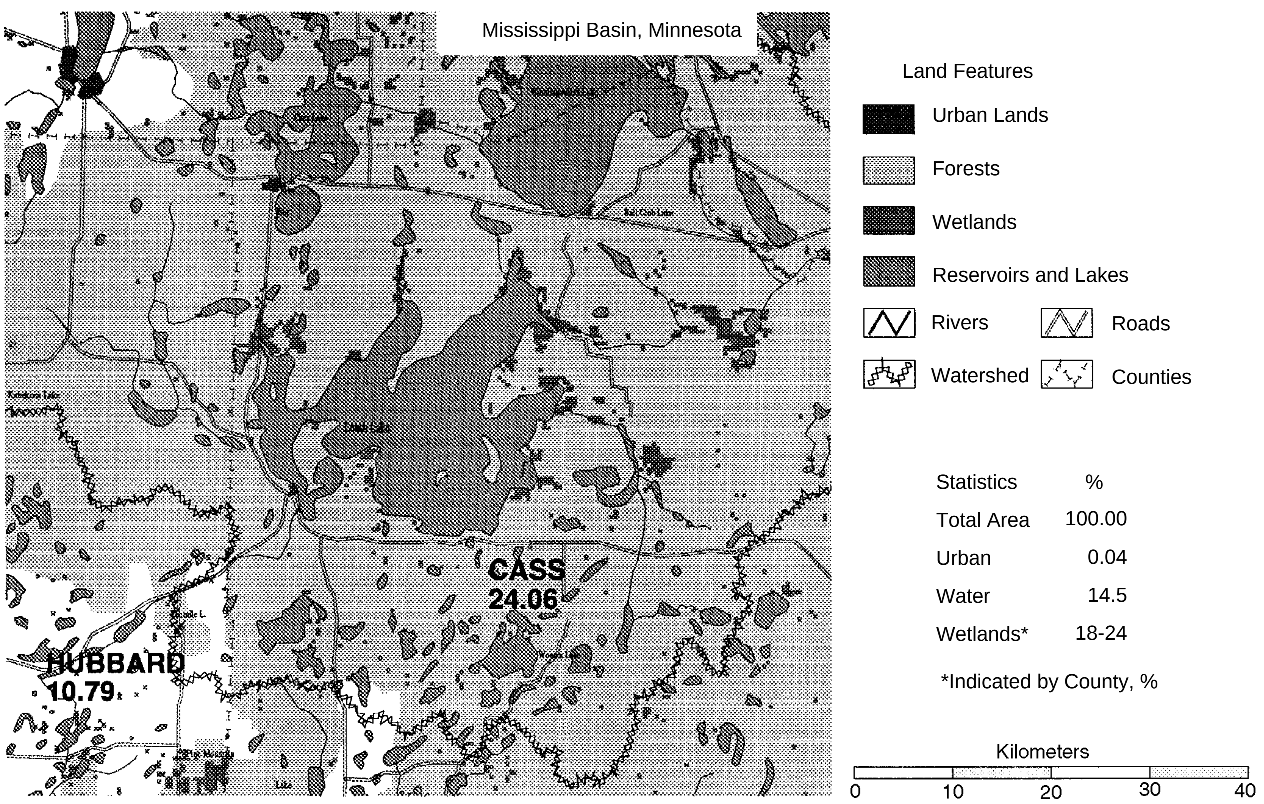


Figure 13. Wetlands and urban lands in Cass County area.

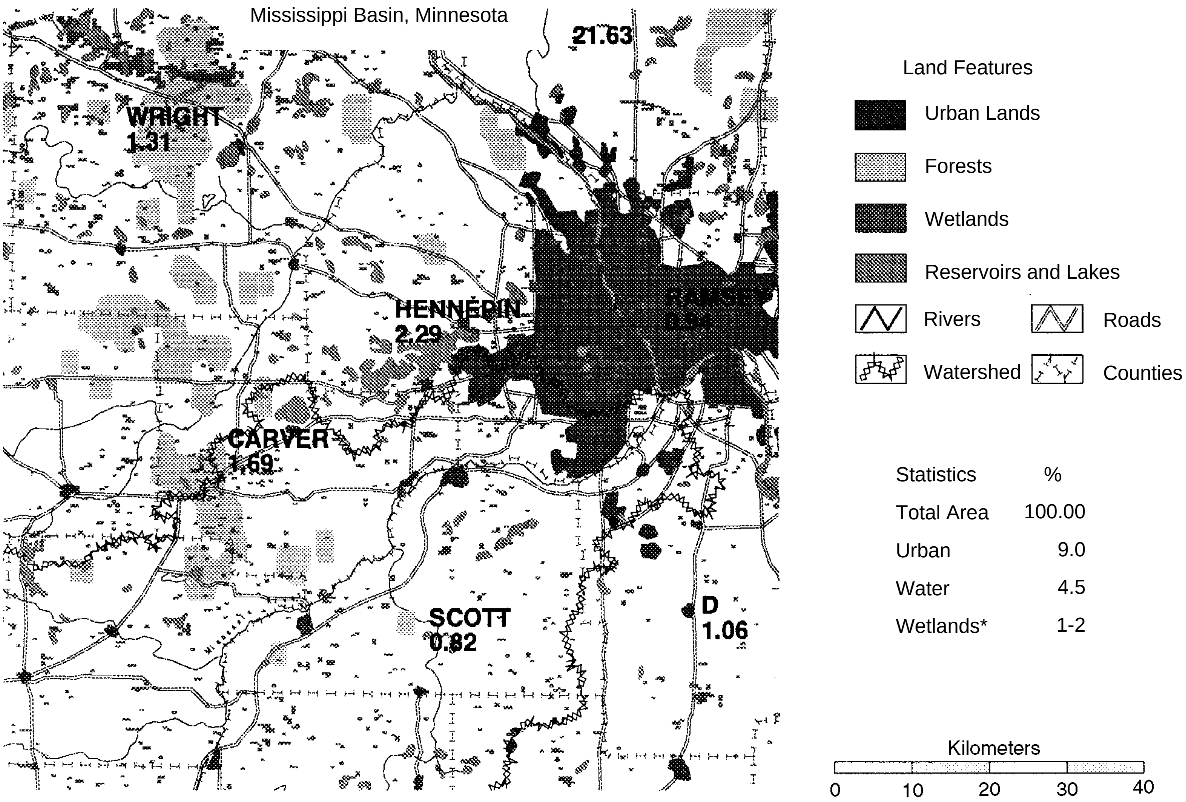


Figure 14. Wetlands and urban lands in Twin Cities area.

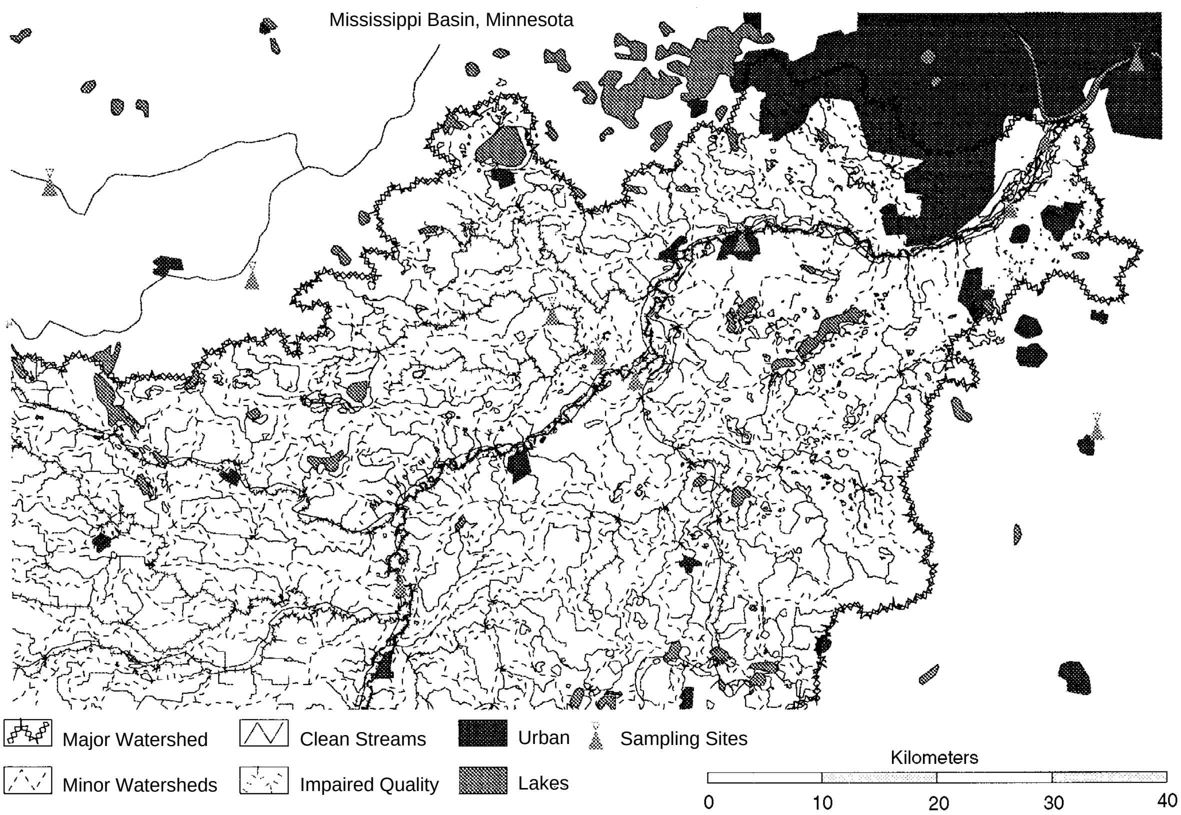


Figure 15. Minnesota River watershed in Twin Cities area.

seasonal distribution in both areas. Average runoff ranges from 150 to 250 millimeters (7, 29).

Conclusion

The current status of the project indicates that most of the input data is available, though dispersed among many agencies. In both the United States and Russia, research methodologies have been developed and applied to study landscape feature impacts on runoff quantity and quality based on simulation and statistical analysis. The comparative analysis of hydrologic and diffuse pollution processes on watersheds in the Upper Mississippi and Upper Volga basins will allow derivation of metrics of wetland loss relative to impacts on runoff and water quality.

The applications of GIS to watershed hydrology are currently much more advanced in the United States than in Russia. Initiatives emerging in the United States, however, could considerably promote GIS use in Russia. Such promotion is beneficial for several reasons. First, this kind of cooperative political activity is in full compliance with the 1992 Freedom Support Act, approved by the U.S. Congress. Second, support of GIS as a new information technology will create a favorable infrastructure in many bilateral economic fields and businesses. Third, a better meshing of the GIS systems in the two countries will lead to further international cooperation in responding to global changes.

Project implementation also helps meet the goal of providing a basis for sound environmental, technical, and economic decision-making on the use of natural resources. This knowledge is essential in developing practical guidelines for sustainable economic development through applied research and technologies.

Acknowledgments

The Water Quality Division of MPCA provided valuable assistance with GIS data for Minnesota. The National Science Foundation (NSF/EAR-9404701) contributed research support. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

1. Wetland management: Hydraulic and hydrologic research needs. 1987. In: Wetland hydrology: Proceedings of the National Wetland Symposium, Chicago, IL (September 16-18). Association of State Wetland Managers, Inc. pp. 331-333.

2. Sather, J.H. 1992. Intensive studies of wetland functions: 1990-1991 research summary of the Des Plaines River Wetland demonstration project. Technical Paper No. 2. Wetland Research, Inc.

3. Hey, D.L. 1985. Wetlands: A strategic national resource. National Wetlands Newsletter 7(1):1-2.

4. Kloet, L. 1971. Effects of drainage on runoff and flooding within the Pembina River basins, North Dakota and Manitoba. Unpublished report by the Bureau of Sport, Fisheries, and Wildlife (December).

5. Simon, B.D., L.J. Stoerzer, and R.W. Watson. 1987. Evaluating wetlands for flood storage. In: Wetland hydrology: Proceedings of the National Wetland Symposium, Chicago, IL (September 16-18). Association of State Wetland Managers, Inc. pp. 104-112.

6. Johnston, C.A. 1994. Cumulative impacts to wetlands. Wetlands 14(1):49-55.

7. Higgins J.M., T.B. Nawrocki, and N.A. Nazarov. 1993. Hierarchical approach to integrated watershed management: Joint TVA/Russian demonstration project. Proceedings of AWWA CONSERV'93 Conference, Sessions 4B-1 through 7C-3, Las Vegas, NV. pp. 1,177-1,197.

8. Environmental Science Research Institute. 1993. Digital chart of the world.

9. Vogelmann, J.E., F.R. Rubin, and D.G. Justice. 1991. Use of Landsat thematic mapper data for fresh water wetlands detection in the Merrimack River watershed, New Hampshire (unpublished).

10. Conger, D.H. 1971. Estimating magnitude and frequency of floods in Wisconsin. Open File Report. Madison, WI: U.S. Geological Survey.

11. Jacques, J.E., and D.L. Lorenz. 1988. Techniques for estimating the magnitude and frequency of floods in Minnesota. Water Resources Investigation Report 87-4170. St. Paul, MN: U.S. Geological Survey.

12. Novitzki, R.P. 1979. Hydrologic characteristics of Wisconsin's wetlands and their influence on floods, stream flow, and sediment. In: Greeson, P.E., J.R. Clark, and J.E. Clark, eds. Wetland functions and values: The state of our understanding. Minneapolis, MN: American Water Resources Association. pp. 377-388.

13. Larson, L.A. 1985. Wetlands and flooding: Assessing hydrologic functions. Proceedings of the National Wetland Assessment Symposium, Portland, ME (June 17-20). pp. 43-45.

14. Ogawa, H., and J.W. Male. 1986. Simulating the flood mitigation role of wetlands. J. Water Res. Planning and Mgmt. 112(1).

15. Johnston, C.A., N.E. Detenbeck, and G.J. Niemi. 1990. The cumulative effect of wetlands on stream water quality and quantity: A landscape approach. Biogeochemistry 10:105-141.

16. Moore, I.D., and C.L. Larson. 1979. Effects of drainage projects on surface runoff from small depressional watersheds in the North Central Region. WRRC Bulletin 99. Water Resources Research Center, University of Minnesota.

17. Maslov, B.S., I.V. Minaev, and K.V. Guber. 1989. Reclamation manual. Rosagropromizdat (in Russian).

18. Oberts, G.L. 1981. Impacts of wetlands on watershed water quality. In: Richardson, B., ed. Selected proceedings of the midwest conference on wetland values and management. Navarre, MN: Freshwater Society. pp. 213-226.

19. Horton, R.E. 1945. Erosion development of streams. Geol. Soc. Amer. Bull. 56:281-283.

20. Chow, V. 1964. Handbook of applied hydrology. New York, NY: McGraw-Hill.

21. Wishmeier W.H., and D.D. Smith. 1978. Predicting rainfall erosion losses: A guide to conservation planning. Agriculture Handbook No. 537. U.S. Department of Agriculture.

22. Young, R.A., C.A. Onstad, D.D. Bosch, and W.P. Anderson. 1987. AGNPS: Agricultural nonpoint-source pollution model. U.S. Department of Agriculture Conservation Research Report 35.

23. Nazarov, N.A. 1988. Model formation of the flood hydrograph of Northern Plain Rivers. Water Resour. 15(4):305-315.

24. Nawrocki, T., C. Johnston, and J. Salés. 1994. GIS and modeling in ecological studies: Analysis of Beaver Pond impacts on runoff and its quality. Voyageurs National Park, Minnesota, case study. NRRI Technical Report 94/01 (February).

25. Anderson, J.P., and W.J. Craig. 1984. Growing energy crops on Minnesota wetlands: The land use perspective. Minneapolis, MN: Center for Urban and Regional Affairs, University of Minnesota. pp. 1-88.

26. Mitsch, W.J., and J.G. Gosselink. 1986. Wetlands. New York, NY: Van Nostrand Reinhold.

27. Brown, R.G. 1987. Effects of wetland channelization on storm water runoff in Lamberts Creek, Ramsey County, Minnesota. In: Proceedings of the National Wetland Symposium on Wetland Hydrology, Chicago, IL (September 16-18). pp. 130-136.

28. SCS. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. Agriculture Handbook No. 436. Soil Conservation Service, U.S. Department of Agriculture.

29. USGS. 1988. Water resources data: Minnesota, water year 1988. Vol. 2, Upper Mississippi and Missouri River basins. U.S. Geological Survey Data MN-88-1.

Water Quality Applications

Vulnerability Assessment of Missouri Drinking Water to Chemical Contamination

Christopher J. Barnett, Steven J. Vance, and Christopher L. Fulcher
Center for Agricultural, Resource, and Environmental Systems, University of Missouri, Columbus, Missouri

Introduction

In 1991, the Missouri Department of Natural Resources (MDNR) implemented the Vulnerability Assessment of Missouri Drinking Water to Chemical Contamination project. MDNR's Public Drinking Water Program (PDWP) contracted with the Center for Agricultural, Resource, and Environmental Systems (CARES) to conduct this assessment. They designed the project to determine which, if any, public water supplies are threatened by chemicals being tested under the Safe Drinking Water Act.

Under Phase II of the Safe Drinking Water Act, the United States Environmental Protection Agency (EPA) required that all public drinking water systems be routinely monitored for 79 contaminants beginning January 1, 1993. If a selected chemical parameter is not detected in an area that would affect a water supply (where "detected" is defined as used, stored, manufactured, disposed of, or transported regardless of amount), then the water supply need not be tested for that chemical. Instead, that system would be granted a use waiver, meaning that the state would not test for that chemical. EPA grants use waivers for 43 of the 79 contaminants. Use waivers can result in considerable cost savings.

Because use waivers are granted based on the spatial relationship between drinking water sources and contaminant sources, accurate positional data needed to be collected for those items. A geographic information system (GIS) was used to store and analyze this information in a spatial context.

Water Sources

Water sources, as defined for this study, are the points where water is drawn from a river, lake, or aquifer for use in a public water supply. Our efforts focused primarily on the development of the water source layers for the GIS. These layers, containing wellheads, impoundment intakes, and river intakes, were created in house or obtained from state and federal agencies. MDNR regional office personnel inspected these water source layers in the spring of 1993. Since these personnel routinely inspect Missouri public drinking water supplies, their knowledge of these locations is exceptional. The updated water source information was mapped on 1:24,000-scale USGS topographic quadrangles at the regional offices, then entered into the GIS. MDNR's PDWP provided available attribute information, which was associated with these layers. The layers offer the most accurate and current information available. Only the community (e.g., cities, subdivisions, mobile home parks) and the nontransient, noncommunity (e.g., schools, large businesses) water supply systems were considered for water source mapping. This study did not consider private wells.

The information is stored in the GIS in the form of geographic data sets or layers. The wellhead layer contains 2,327 public wells and their attributes (e.g., well depth, casing type). The majority of the wellheads are located in the Ozarks and Southeast Lowlands. Naturally poor ground-water quality prohibits a heavy reliance on ground water for drinking water in other areas of the state. The surface water impoundment layer contains 105 points representing the intake locations for systems that rely on lake water. Additionally, the drainage basin and lake area are mapped for these systems. The majority of the systems that rely on lake water are located in northern and western Missouri. The final layer represents the systems that use river water. The majority of the 50 intakes are located on the Mississippi and Missouri Rivers and on the major streams in the Grand and Osage River basins.

Contaminant Sources

Contaminant sources, as defined for this study, are the points or areas where existing databases indicate the presence of a chemical contaminant. Incorporation of contaminant data into the GIS proved to be the most difficult task. These data usually contained very precise information about what contaminants were found at a site and who was responsible, but the quality of the locational information was often poor.

Ninety-three state and federal databases were reviewed for contaminant information before performing the final use waiver analysis. The contaminant information was broken into two separate types, contaminant sites and pesticide dealerships. The contaminant sites were locations at which certain chemicals were known to exist. The pesticide dealerships were dealerships licensed to distribute restricted use pesticides. Information about contaminant sites was extracted from the databases and entered into Microsoft Excel, a spreadsheet program. The small amount of data with coordinate (latitude/longitude) or map information was readily converted to the GIS. The majority of the contaminant records, however, contained only address information, often appearing as a rural route address or post office box number.

While the water source locations were being verified, personnel at the MDNR regional offices reviewed the contaminant site records. The regional office personnel were familiar with their respective territories and could assist CARES personnel in locating the contaminant sites. The Missouri Department of Agriculture pesticide use investigators provided additional information about the locations of contaminant sites. All contaminant source information was also mapped on the 1:24,000-scale USGS topographic quadrangles and transferred to the GIS.

Of more than 2,800 contaminant sites found in these databases, 88 percent were geographically located and used in the study. At this time, the contaminant site layer contains 2,493 points representing the information collected on the 43 chemical contaminants required by MDNR. Each point contains a seven-digit chemical code indicating the chemical it represents and serving as a link to the chemical contaminant files. The contaminant sites tend to be concentrated more in urban areas than rural areas. Even though this layer is being continually updated, the basic distribution of contaminant sites remains the same.

A second contaminant source layer represents Missouri's licensed pesticide dealers. This information is included to indicate potential contamination even though specific chemicals at dealership locations are not known. At this time, we have been able to locate 1,344 dealerships out of 1,650. Two types of dealerships are included in the layer, active dealers and inactive dealers. Of the active dealerships in 1991, 91 percent were found and entered into the GIS. Of the inactive dealerships, 79 percent were located.

Spatial Analysis

The final parameters for the use waiver analysis were developed from EPA and MDNR guidelines and account for the capabilities of the GIS. These parameters were designed to present a conservative list of the systems that needed to be tested for the possible presence of studied chemicals. Parameters for the wellhead analysis are as follows:

- A ¼-, ½-, and 1-mile radius around each wellhead was searched for contaminant sites and pesticide dealerships (see Figure 1). Any contaminant sources found within those radii were reported to PDWP. (PDWP requested that the results of the three radius analyses be reported, but the ½-mile radius was used to determine the issue of the use waiver.)
- Any wellheads found within a contaminant area were denied a use waiver for that contaminant.
- Each highway and railroad within 500 feet of a wellhead was recorded. This indicates the threat posed by the transport of chemicals near wellheads.
- Additionally, the percentage of the county planted in corn, soybeans, wheat, sorghum, tobacco, cotton, and rice was listed for each well to indicate the threat posed by agricultural chemical use within that county.

The parameters for the systems relying on lake water are as follows:

- Any contaminant sources found within a surface water impoundment drainage basin caused the associated intake(s) to fail use waiver analysis for those contaminants.

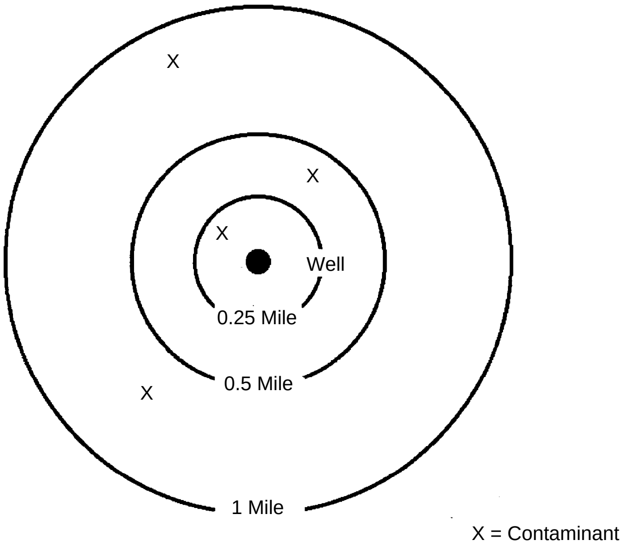


Figure 1. Use waiver search radius distances.

- Any area of contamination overlapping a drainage basin caused the associated intake to fail use waiver analysis for that contaminant.
- Transportation corridors passing through a drainage basin were noted to indicate the threat posed by transport of chemicals within the basin.
- The percentage of the county planted in the seven crops mentioned above was listed to indicate agricultural chemical use within the drainage basin.

Many of the rivers that supply water to systems in Missouri have their headwaters outside the state. To fully evaluate the potential for contamination within those drainage basins, we would have to collect data for large areas outside of the state. For example, the Mississippi and Missouri River drainage basins cover large portions of the United States. Because collecting data for those areas would be impractical, we have recommended to MDNR that use waivers not be granted to river supplies.

The following provides details on how the analysis was performed. The GIS searches around each wellhead for each radius and notes which contaminant sites affect which wellheads. If a contaminant falls within that radius, we recommend that the wellhead be monitored. In this example, the well is affected by one contaminant within the ¼-mile radius, two within the ½-mile radius, and four within the 1-mile radius.

Results

The results of the use waiver analysis indicate which systems may be affected by the use of a chemical near a water source. Several results show the substantial savings realized from our analysis. For example, the analysis showed that only five wells serving four public drinking water systems were potentially affected by dioxin and should be monitored. By not testing the remaining systems for dioxin, the state can realize a considerable cost savings, as the test for dioxin is the most expensive test to perform.

The final wellhead system analysis shows that the ½-mile buffer analysis affected a total of 447 wellheads in 241 systems. That is, a chemical site or pesticide dealership was found within ½ mile of 447 public wellheads. A result form was generated for each of the 1,340 systems in the state listing each well or intake and the potential threat posed by nearby contaminant sources.

The cost of testing all wellhead systems for all 43 contaminants without issuing use waivers is more than $15 million (see Table 1). According to our analysis, CARES estimates that only $1.8 million need be spent to monitor vulnerable wells. Therefore, the state can save more than $13.5 million in monitoring costs.

Table 1. Estimated Cost Savings for Public Drinking Water Systems

Method	Estimated Total Cost	Estimated Mean Cost per System	Estimated Total Cost Savings
No use waiver	$15,533,100	$12,200	$0
With use waiver	$1,813,900	$1,400	$13,719,200

Summary and Recommendations

To date, the investment the state made in the vulnerability assessment project has provided many benefits. The state saved several million dollars in testing costs and developed several spatial and nonspatial databases that will have many uses. In addition, the project established a basic framework for future assessments, which EPA requires on a regular basis.

The basic data required for use waiver analysis are the locations of water sources and the locations of potential contamination sources. CARES determined that the available data did not contain the information necessary to map these locations or that the data were of questionable quality. Many layers required update and correction. Considerable effort was necessary to improve existing locational information for both water source layers and chemical contaminant files. Local knowledge of an area was heavily relied upon to determine accurate locations, particularly contaminant sites. The vast majority of these sites contained only the address as the geographic reference. An address is not a coordinate system; it does not indicate a fixed location on a map. Because the location of any chemical detection site is of vital importance, state and federal agencies that collect these data need to record more complete geographic information. Ideally, a global positioning system could be employed to generate coordinates. Realistically, the recording of legal descriptions or directions from an easily located point would substantially improve the quality of the current databases.

In many cases, data resided in digital format; however, due to regulations or lack of agency cooperation, they could only be distributed in paper format. Reentering data from paper format into digital format required considerable time and expense. Interagency cooperation should be emphasized to reduce unnecessary data entry.

Reach File 3 Hydrologic Network and the Development of GIS Water Quality Tools

Stephen Bevington
Water Quality Section, Division of Environmental Management, North Carolina Department of Environment, Health, and Natural Resources, Raleigh, North Carolina

Introduction

The application of geographic information system (GIS) tools to water quality management is limited by the lack of geographically referenced data describing the surface water environment. Ongoing efforts at the local, state, and federal level are producing a multitude of GIS data coverages describing land use/cover and relevant water quality data files. As these data coverages become available, water quality managers will need to develop new analysis techniques to take advantage of the vast amount of geographically referenced data. A key step in the development of analytical tools for water quality management will be the development and maintenance of a coverage describing the structure and hydrology of surface waters.

Reach File 3 (RF3) is one potential source of surface water maps and topology for the development of a GIS-based water quality analysis tool. This paper describes a pilot project designed to examine the suitability of RF3 as a network system for the collection, integration, and analysis of water quality data.

To be considered an appropriate water quality analysis tool, RF3 should provide the following functions:

- Present a working environment that allows users to explore geographic relationships between surface water features, landmark features, and data coverages.
- Allow users to select specific stream segments, including all points upstream and downstream of a given point.
- Provide tools to assist users in partitioning water quality databases into hydrologically meaningful subsets.

Reach File 3

RF3 is a hydrographic database of the surface waters of the United States. The database contains 3 million river reaches mapped at 1:100,000 scale. The source for RF3 arcs were digital line graphs (DLGs).

Attribute data for RF3 arcs include the major-minor DLG pairs, stream name, water-body type, stream order, and a unique identifying reach number. The unique reach numbers are structured in such a way as to provide a logical hydrologic framework. Reach numbers can be used to sort the database for all reaches in any specified watershed or locate all upstream or downstream reaches.

The U.S. Environmental Protection Agency (EPA) originally designed RF3 as a tabular data set. It evolved into a GIS data coverage, and EPA and the U.S. Geological Survey (USGS) will likely maintain it as a surface water mapping standard. At present, RF3 as a GIS data layer is not widely used for water quality applications.

RF3 Pilot Study: Upper Yadkin River Basin

The Upper Yadkin River basin (USGS h03040101) was selected to test RF3 water quality applications (see Figures 1 and 2). The Upper Yadkin was chosen because of the availability of water quality and stream flow data layers in that area. Also, the Upper Yadkin RF3 file contained arcs depicting lakes and double-line rivers as well as simple stream networks. These two-dimensional water features present interesting complications to network routing and path-finding.

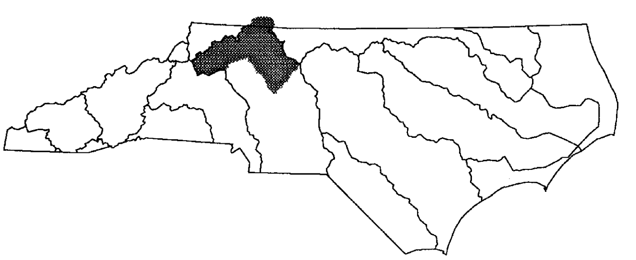


Figure 1. The Upper Yadkin River watershed, North Carolina and Virginia.

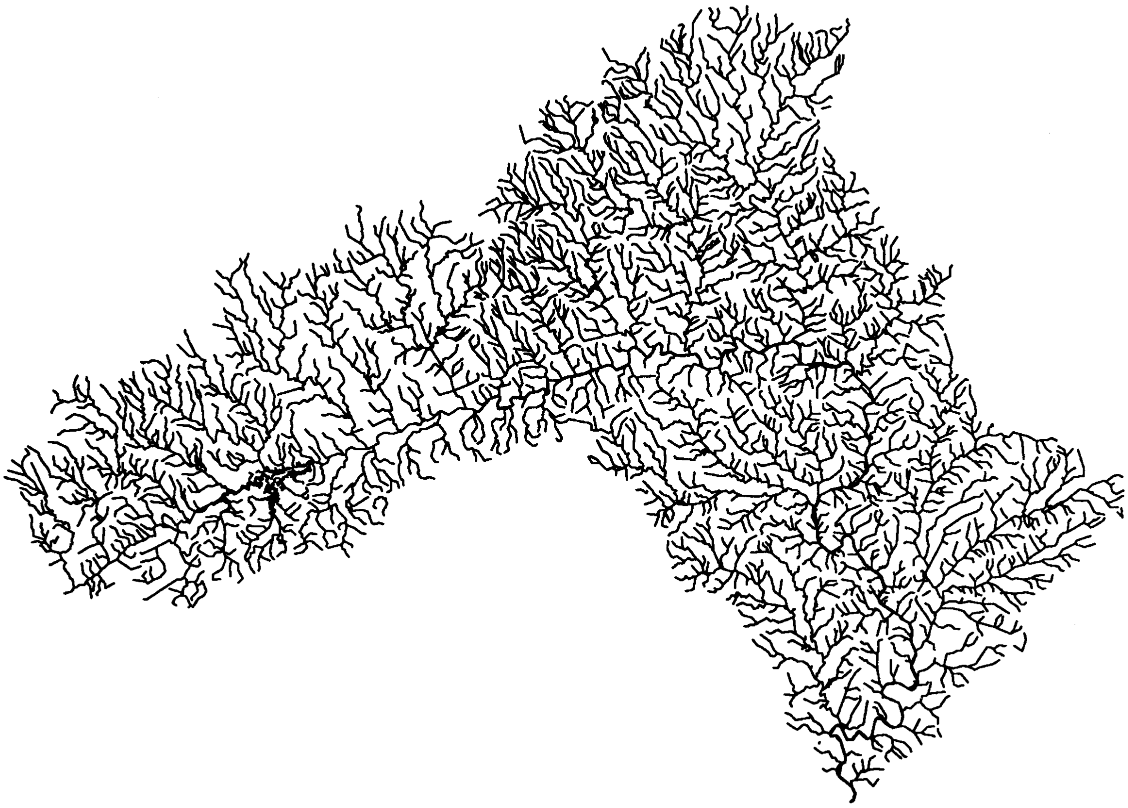


Figure 2. RF3 hydrography for the Upper Yadkin River basin.

Two forms of point source data were used in the study: National Pollutant Discharge Elimination System (NPDES) wastewater discharge points and USGS gages. The NPDES coverage includes data on the permit limits such as daily flow, dissolved oxygen, biochemical oxygen demand (BOD), and ammonia. The USGS gage coverage includes data on several flow statistics for each USGS gage in the basin. Both data layers contain information about the location of the site and stream with which it is associated.

Coverages of counties and cities were also made available for geographic orientation.

Preparing the Network

The original RF3 file received from the USGS had several topological issues that needed to be addressed before RF3 could function as a stream network. First, not all arcs were connected to each other (see Figure 3). The ARC/INFO command TRACE was used to select all connected arcs. This revealed three major blocks of connected arcs and many isolated arcs. The three major blocks were easily connected in ARCEDIT by extending the main tributary links between the blocks. Processing of the isolated arcs was not pursued for this study. Complete processing of arcs for this RF3 basin would not be difficult or time consuming, with the possible exception of the many arcs surrounding the lake. A functional network encompassing a high percentage of the arcs was not difficult to achieve, however.

The second network issue concerned the direction of the arcs. RF3 has all arcs oriented toward the top of the watershed, with the exception of one side of double-line streams. Arcs that make up double-line streams are oriented up one side of the double-line section and down the other (see Figure 3). Clearly, this complicates routing. To allow for accurate downstream routing, arcs on the downward-facing side of the stream were flipped using ARCEDIT. With all arcs in the network facing upstream, most hydrologic routes can be traced. Given the network system alone, upstream routing from double-line streams does not function properly, ignoring all tributaries on one side of the double-line stream.

Double-Line Stream Routing

Many possible solutions exist for the problems caused by double-line streams. Some involve improving the network (e.g., by adding center-line arcs down the middle of double-line streams). This would involve not only adding arcs but establishing conductivity with all tributaries. This option will involve significant topological changes to RF3. To maintain compatibility with other

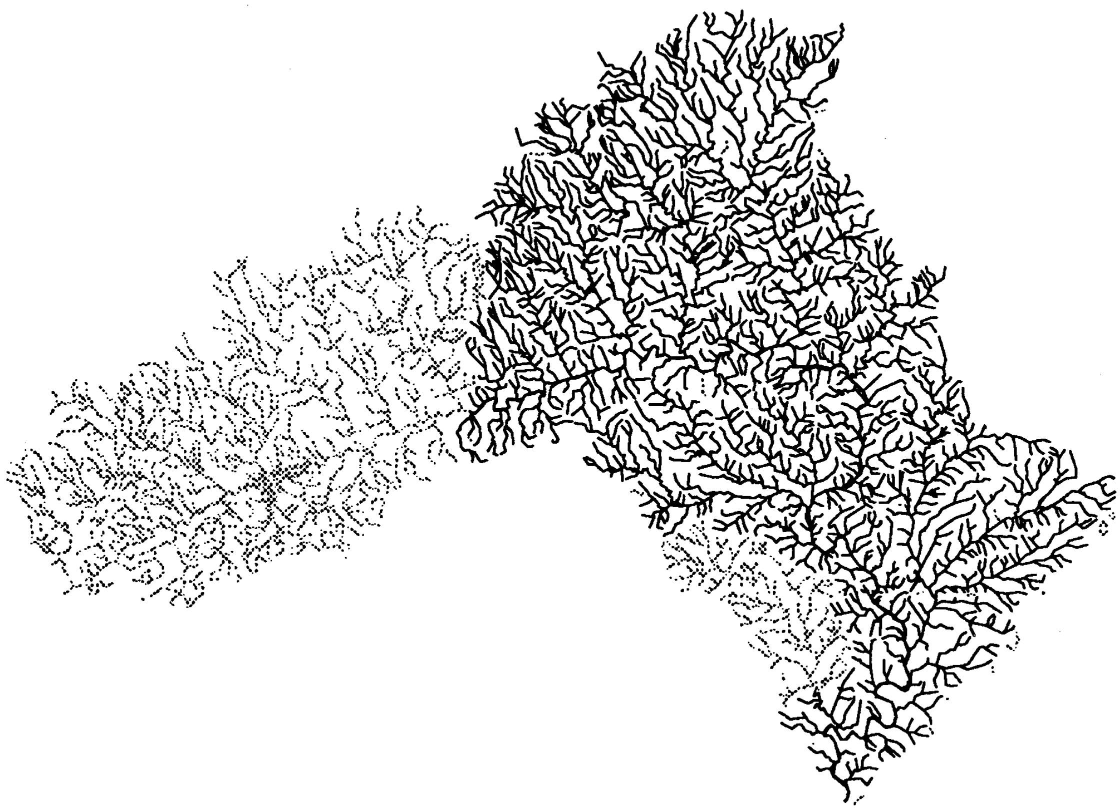


Figure 3. Original conductivity of RF3 hydrography.

RF3 and DLG sources, this option should be considered only as part of a major RF3 upgrade.

At the other end of the technological spectrum, one could simply instruct users to watch for double-line streams and select arcs from both sides of the river. Users may have trouble with this option, however, if they are not working at an appropriate scale to easily differentiate between double- and single-line streams.

A third option is to program an arc macro language (AML) to check for double-line streams and run upstream traces from both sides of the stream. The difficulty in this method is to find the appropriate starting place on both banks. The algorithm developed to do this goes as follows:

- Select stream segment and trace upstream. (Results in incomplete trace.)
- Find the minimum segment and mile of selected double-line streams.
- Unselect all double-line streams below minimum segment and mile.
- Add to selection all non-double-line streams.
- Trace from original point both upstream and downstream. (Results in completed upstream trace.)

Results and Conclusions

AMLs and menus were written that can perform upstream and downstream traces on the RF3 stream network and select data points within 500 feet of the stream. Lists of attributes can be returned to the screen. This system is easy to use and can be used to quickly identify general watersheds and water quality data points. An AML can be used to trace upstream from a double-line stream given only one point on the stream (see Figure 4). The success of these methods suggests that two-dimensional surface water features can be successfully integrated into RF3 water quality analyses.

This system could be further developed to support polygon analysis using the ARC command BUFFER. Other developments could include the procedures to write selected attributes to files and increased flexibility for the screen environment and outputs.

This pilot project demonstrates only a few of the potential applications of RF3 to water quality. Success in this pilot project suggests that RF3 is a potentially valuable water quality analysis tool. It may also be a valuable tool for demonstrating the results of water quality analyses to managers or the public.

Because RF3 will require some processing before network algorithms can be run, it is important to plan for the integration of RF3 into other GIS tools and data coverages.

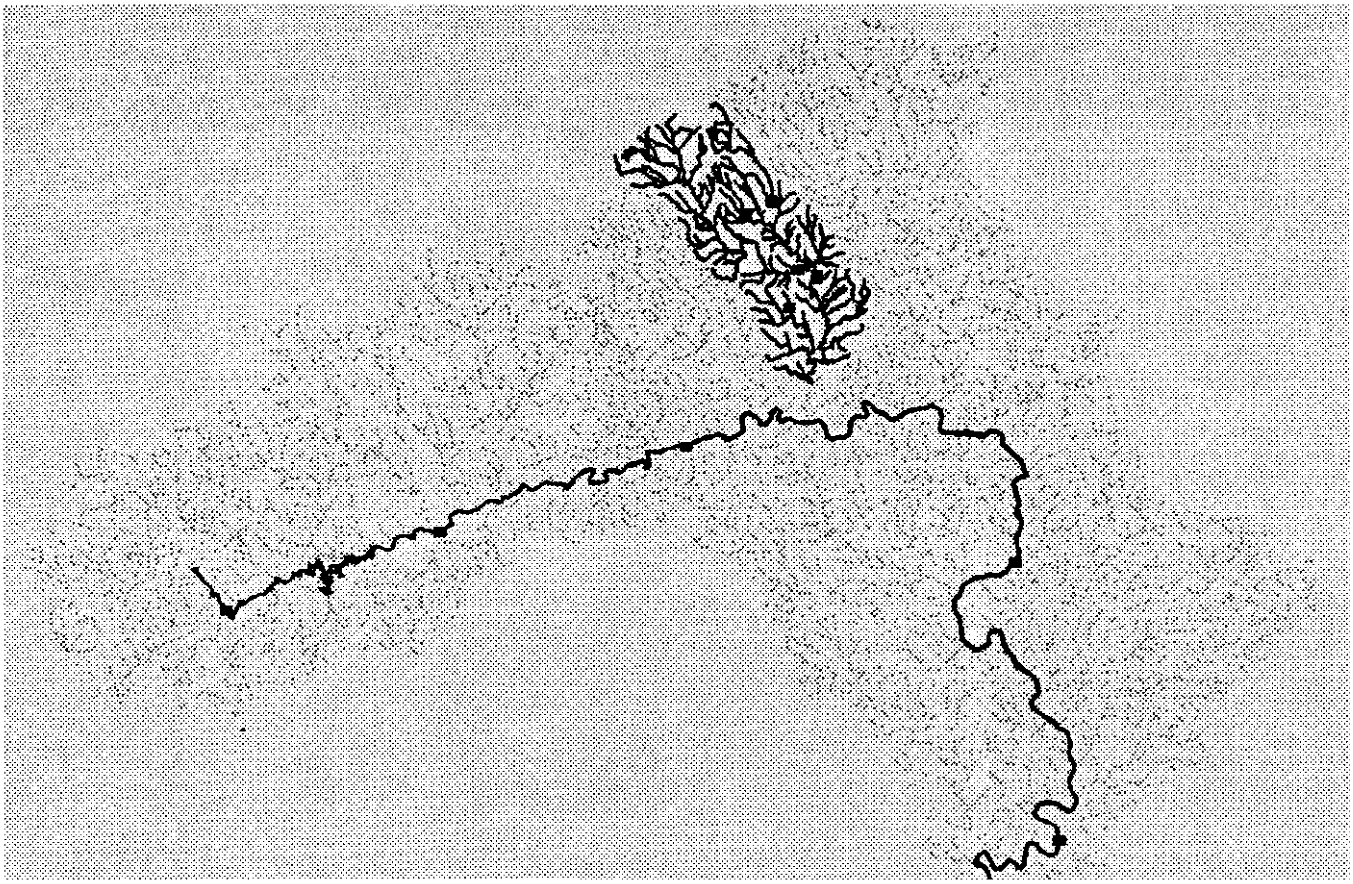


Figure 4. Upstream and downstream traces of RF3 hydrography.

Ongoing efforts to update RF3 may address some of these problems. If RF3 is to be developed into a productive water quality management tool, it is important to proceed in a way that is compatible with ongoing efforts to update RF3 and the development of new data sources.

EPA's Reach Indexing Project—Using GIS To Improve Water Quality Assessment

Jack Clifford
Office of Wetlands, Oceans, and Watersheds, U.S. Environmental Protection Agency, Washington, DC

William D. Wheaton and Ross J. Curry
Research Triangle Institute, Research Triangle Park, North Carolina

Abstract

The Waterbody System (WBS), which the U.S. Environmental Protection Agency (EPA) originally developed to support preparation of the report to Congress that Section 305(b) of the Clean Water Act requires, is a potentially significant source of information on the use support status and the causes and sources of impairment of U.S. waters. Demand is growing for geographically referenced water quality assessment data for use in interagency data integration, joint analysis of environmental problems, establishing program priorities, and planning and management of water quality on an ecosystem or watershed basis.

Because location of the waterbody assessment units is key to analyzing their spatial relationships, EPA has particularly emphasized anchoring water bodies to the River Reach File (RF3). The reach file provides a nationwide database of hydrologically linked stream reaches and unique reach identifiers, based on the 1:100,000 U.S. Geological Survey (USGS) hydrography layer.

EPA began the reach indexing project to give states an incentive to link their water bodies to RF3 and to ensure increased consistency in the approaches to reach indexing. After a successful 1992 pilot effort in South Carolina, an expanded program began this year. Working with Virginia, a route system data model was developed and proved successful in conjunction with state use of PC Reach File (PCRF), a PC program that relates water bodies to the reach file. ARC/INFO provides an extensive set of commands and tools for developing and analyzing route systems and for using dynamic segmentation.

One important advantage of the route system is that it avoids the necessity of breaking arcs; this is an important consideration in using RF3 as the base coverage in a geographic information system (GIS). Using dynamic segmentation to organize, display, and analyze water quality assessment information also simplifies use of the existing waterbody system data. Because of the variability in delineation of water bodies, however, other states used a number of different approaches. Working with these states has defined a range of issues that must be addressed in developing a consistent set of locational features for geospatial analysis.

Wider use of these data also depends upon increased consistency in waterbody assessments within and between states. Several factors complicate the goal of attaining this consistency in assessment data:

- The choice of beneficial use as the base for assessment of water quality condition.
- The historical emphasis on providing flexible tools to states.
- The lack of robust standards for assessment of water quality condition.

This paper explores possible resolutions to the problem of building a national database from data collected by independent entities.

Section 305(b) of the Clean Water Act and the Waterbody System

Background of Section 305(b)

Since 1975, Section 305(b) of the Federal Water Pollution Act, commonly known as the Clean Water Act (CWA), has required states to submit a report on the quality of their waters to the U.S. Environmental Protection Agency (EPA) administrator every 2 years. The administrator must transmit these reports, along with an analysis of them, to Congress.

State assessments are based on the extent to which the waters meet state water quality standards as measured against the state's designated beneficial uses. For each use, the state establishes a set of water quality criteria or requirements that must be met if the use is to be realized. The CWA provides the primary authority to states to set their own standards but requires that all state beneficial uses and their criteria comply with the 'fishable and swimmable' goals of the CWA.

Assessments and the Role of Guidelines

EPA issues guidelines to coordinate state assessments, standardize assessment methods and terminology, and encourage states to assess support of specific beneficial uses (e.g., aquatic life support, drinking water supply, primary contact recreation, fish consumption). For each use, EPA asks that the state categorize its assessment of use support into five classes:

- *Fully supporting:* meets designated use criteria.
- *Threatened:* may not support uses in the future unless action is taken.
- *Partially supporting:* fails to meet designated use criteria at times.
- *Not supporting:* frequently fails to meet designated use criteria.
- *Not attainable:* use support not achievable.

In the preferred assessment method, the state compares monitoring data with numeric criteria for each designated use. If monitoring data are not available, however, the state may use qualitative information to determine use support levels.

In cases of impaired use support (partially or not supporting), the state lists the sources (e.g., municipal point source, agriculture, combined sewer overflows) and causes (e.g., nutrients, pesticides, metals) of the use support problems. Not all impaired waters are characterized. Determining specific sources and causes requires data that frequently are not available.

States generally do not assess all of their waters each biennium. Most states assess a subset of their total waters every 2 years. A state's perception of its greatest water quality problems frequently determines this subset. To this extent, assessments are skewed toward waters with the most pollution and may, if viewed as representative of overall water quality, overstate pollution problems.

Assessment Data Characteristics

Each state determines use support for its own set of beneficial uses. Despite EPA's encouragement to use standardized use categories, the wide variation in state-designated beneficial uses makes comparing state uses an inherent problem. This affects the validity of aggregation and use of data across state boundaries. Comparably categorizing waters into use support categories also poses a problem; different states apply the qualitative criteria for use support levels in very different ways. Further limiting the utility of Section 305(b) data is that data are aggregated at the state level and questions about the use support status of individual streams cannot be resolved without additional information. While some states report on individual waters in their Section 305(b) reports, EPA's Waterbody System (WBS) is the primary database for assessment information on specific waters.

State monitoring and assessment activities are also highly variable. States base assessments on monitoring data or more subjective evaluation. The evaluation category particularly differs among states.

Waterbody System

The WBS is a database and a set of analytical tools for collecting, querying, and reporting on state 305(b) information. It includes information on use support and the causes and sources of impairment for water bodies, identification and locational information, and a variety of other program status information.

As pointed out earlier, although some states discuss the status of specific waters in their 305(b) reports, many do not. The WBS is generally much more specific than the 305(b) reports. It provides the basic assessment information to track the status of individual waters in time and, if georeferenced, to locate assessment information in space. By allowing the integration of water quality data with other related data, the WBS provides a framework for improving assessments.

WBS has significant potential for management planning and priority setting and can serve as the foundation for watershed- and ecosystem-based analysis, planning, and management. In this respect, it can play a vital role in setting up watershed-based permitting of point sources. The primary function of WBS is to define where our water quality problems do and do not exist. WBS is increasingly used to meet the identification requirement for waters requiring a total maximum daily load (TMDL) allocation. It can serve as the initial step in the detailed allocation analysis included in the TMDL process. In addition, WBS is an important component of EPA participation in joint studies and analyses. For instance, EPA is currently participating with the Soil Conservation Service (SCS) in a joint project to identify waters that are impaired due to agricultural nonpoint source (NPS) pollution. WBS can also anchor efforts to provide improved public access at the state and national levels to information on the status of their waters.

It is important to recognize that use of WBS is voluntary. Of the 54 states, territories, river basin commissions, and Indian tribes that submitted 305(b) reports, approximately

30 used the WBS in the 1992 cycle. While submissions for the 1994 cycle are not complete, we anticipate about the same level of participation. This represents about a 60-percent rate of participation in WBS, which may be the limit for a voluntary system. This severely limits use of WBS assessment data for regional and national level analysis. If data at the national level are required, mandatory data elements, formats, and standards may be necessary.

EPA is currently attempting to achieve consistency through agreement with other state and federal agencies. The recent work of the Interagency Task Force on Monitoring offers hope for eventual consensus on the need for nationally consistent assessment data and mutually agreed upon standards for collection, storage, and transfer. The Spatial Data Transfer Standards already govern spatial data, allowing movement of data between dissimilar platforms. The Federal Geographic Data Committee provides leadership in coalescing data integration at the federal level; it provides a model for government and private sector efforts. This level of cooperation, however, has not always been present in water assessment data management. Assuming that national and regional assessment data are needed, if consistency is elusive through cooperative efforts, regulations may be necessary. Developing a national database may not be feasible without a mutual commitment by EPA and the states to using common assessment standards.

WBS was originally developed as a dBASE program in 1987. It has undergone several revisions since then, and the current Version 3.1 is written in Foxpro 2.0. The WBS software provides standard data entry, edit, query, and report generation functions. WBS has grown substantially in the years since its inception, primarily in response to the expressed needs of WBS users and EPA program offices. The program's memory requirements and the size of the program and data files, however, are of growing concern to state WBS users and the WBS program manager. Because of the wide range of WBS user capabilities and equipment, users must be equipped to support an array of hardware from high capacity Pentium computers to rudimentary 286 machines with 640 Kb of memory and small hard disks. This range makes memory problems inevitable for some users.

While WBS contains over 208 fields, exclusive of those in lookup tables, approximately 30 fields in four files provide the core data needed to comply with 305(b) requirements. These fields contain:

- Identification information for the water body.
- The date the assessment was completed.
- The status of use support for beneficial uses.
- The causes and sources of any use impairment in the water body.

The uses WBS considers are both state-designated uses and a set of nationally consistent uses (e.g., overall use, aquatic life support, recreation) specified in the 305(b) guidelines. The other essential piece of information is the geographic location of the water body, which the remainder of this paper discusses in detail.

Significant differences exist in the analytical base as well as in assessments. EPA provided little initial guidance on defining water bodies; therefore, states vary widely in their configurations of water bodies. Water bodies are supposed to represent waters of relatively homogeneous water quality conditions, but state interpretation of this guidance has resulted in major differences in waterbody definition.

Initially, many states developed linear water bodies, and these were often very small. The large number of water bodies delineated, however, created significant difficulties in managing the assessment workload and were not ideal in the context of the growing need for watershed information. Some states, such as Ohio, developed their own river mile systems.

As discussed below, some states indexed their water bodies to earlier versions of the reach file, and therefore, the density of the streams these water bodies include is fairly sparse. Recently, many states have redefined their water bodies on the basis of small watersheds (SCS basins, either 11-digit or 14-digit hydrologic unit codes [HUCs]).

Locating water bodies geographically is a necessary prerequisite to assessing water quality on a watershed or ecosystem basis. The WBS has always included several locational fields, including county name and FIPS, river basin, and ecoregion. These fields have not been uniformly populated, however. One of the WBS files includes fields for the River Reach File (RF3) reaches included in the water body. While a few states had indexed their water bodies to older versions of the reach file (RF1 and RF2), however, no state had indexed to RF3 until 1992.

In 1992, EPA initiated a demonstration of geographic information system (GIS) technology in conjunction with the South Carolina Department of Health and Environmental Control. This project involved:

- Indexing South Carolina's water bodies to RF3.
- Developing a set of arc macro languages (AMLs) for query and analysis.
- Producing coverages of water quality monitoring stations and discharge points.
- Using GIS tools in exploring ways to improve water quality assessments.

South Carolina has defined its water bodies as SCS basins.

The results have been very encouraging. First, South Carolina took the initial coverages and decided they needed much more specificity in their use support determinations and their mapping of the causes and sources of impairment. As a result, they mapped these features down to the reach level. Next, they decided that they needed better locational information, so they used global positioning satellite receivers to identify accurate locations for discharges and monitoring stations. They then used GIS query and analysis techniques to relate their monitoring and discharge data to their water quality criteria. South Carolina is using GIS to actively identify water quality problems and improve their assessments.

In 1993, EPA worked cooperatively with several states to index their water bodies to the reach file. Virginia, the next state to be indexed, demonstrated the successful use of PC Reach File (PCRF) software (described later in this paper) for indexing water bodies to the reach file. Ohio and Kansas also are essentially complete. Each of these states required a somewhat different approach than Virginia. The need for flexibility in dealing with states on reach indexing issues is essential. Existing waterbody delineations often represent considerable investment; therefore, EPA must provide the capability to link the state's existing assessment data to the reach file in order to encourage state buy-in.

Figure 1 shows the results of Ohio's indexing of a typical cataloging unit (CU). Figure 2 reflects part of the output of the Kansas work. We can link use support, cause, and source data to each of these water bodies now. In the future, we hope to map these attributes at a higher level of resolution, down to the reach segment level. GIS has proven to be a useful assessment tool. With higher resolution, it should prove to be even more helpful in identifying water quality problems, picking up data anomalies, and assessing management actions, strategies, and policies. This entire process has taught us much and has strengthened enthusiasm for place-based management.

The Reach Indexing Project—Georeferencing the Waterbody System

Purpose and Overview

The reach indexing project is designed to locate water bodies using RF3 as an electronic base map of hydrography and to code RF3 reaches with the specific waterbody identifier (WBID). After linking water bodies to their spatial representation, they can be queried and displayed with assessment data located in WBS files.

Reach indexing includes several steps. First, the state must supply waterbody locations and WBIDs. The next step entails developing a set of procedures for indexing. Finally, the coded RF3 data must be produced.

Input data to the indexing process includes:

- A list of valid WBIDs. In most cases, the state has already input these identification numbers to the WBS.
- Information about the location of each water body. Locational information may be found in marked-up paper maps showing waterbody locations or electronic files from WBS containing waterbody indexing

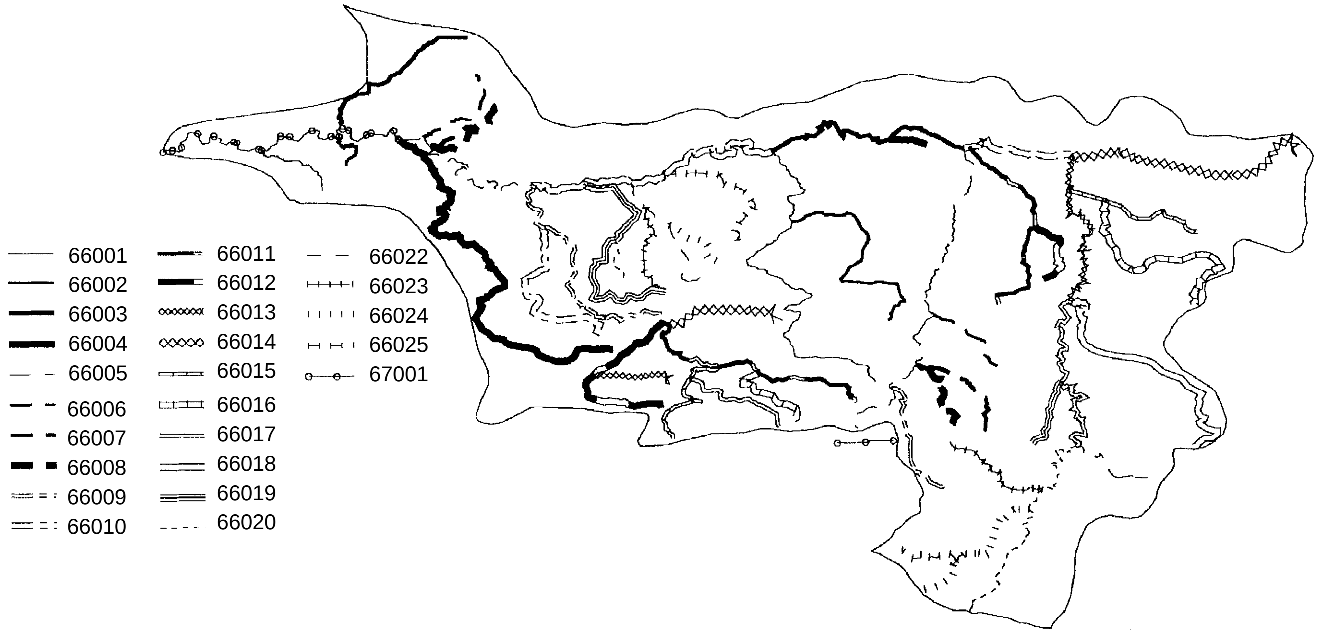


Figure 1. State of Ohio water bodies in CU 04100008.

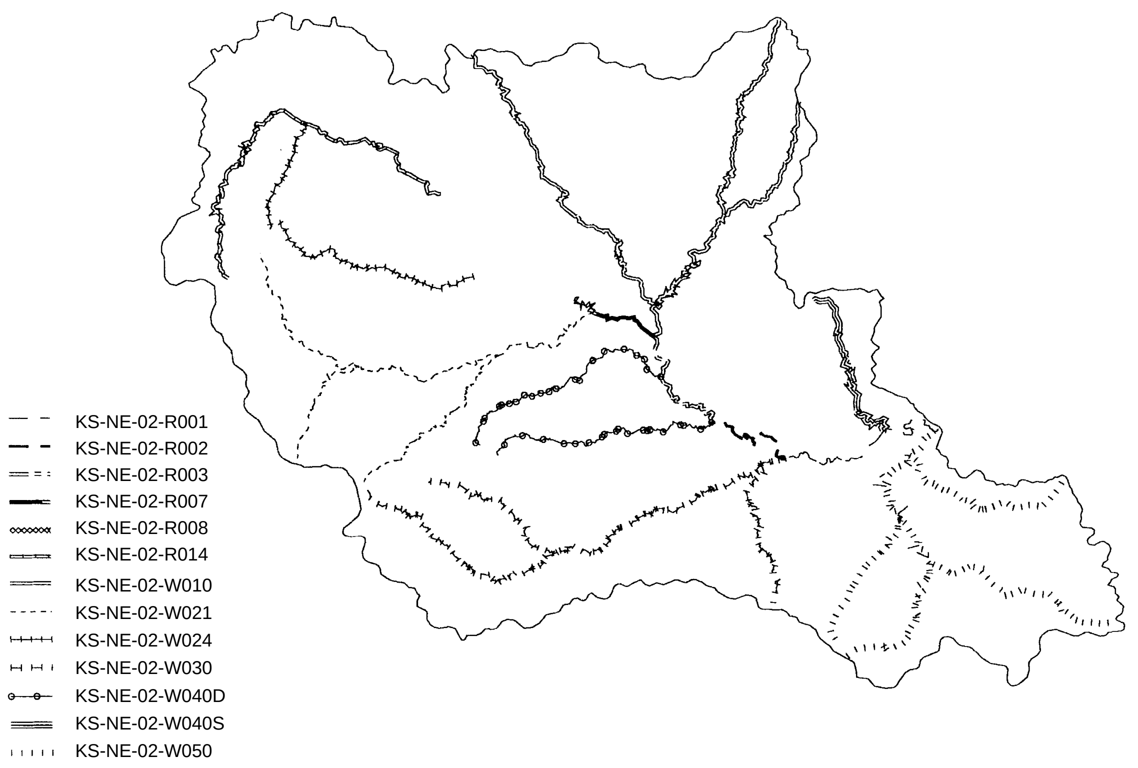


Figure 2. State of Kansas water bodies in CU 11070202.

expressions (discussed later), or it may be embedded in the WBID itself.

- A complete set of RF3 data for the state being indexed.

Depending on the type of information the state supplies, procedures used to index water bodies can be almost fully automated, semiautomated, or completely manual.

The final result of the indexing processes is a set of RF3 coverages that contain a WBID attribute. This product allows querying and displaying of assessment data, which is collected and stored by water body, in a GIS environment.

The Reach File Database

The reach file is a hydrographic database of the surface waters of the continental United States. Elements within the database represent stream segments. The elements were created for several purposes:

- To perform hydrologic routing for modeling programs.
- To identify upstream and downstream connectivity.
- To provide a method to uniquely identify any particular point associated with surface waters.

The unique reach identifier has succeeded in associating other EPA national databases, such as STORET, to surface waters. Any point within these databases can be associated with and identified by a specific location on any surface water element, such as a reservoir, lake, stream, wide river, or coastline.

There are three versions of the reach file. The first was created in 1982 and contained 68,000 reaches. The second version, released in 1988, doubled the size of Version 1. The third version (RF3) includes over 3,000,000 individual reach components.

The base geography of RF3 is derived from U.S. Geological Survey (USGS) hydrographic data (1:100,000 scale) stored in digital line graph (DLG) format. Unlike DLG data, which are partitioned by quad sheet boundaries, RF3 data are partitioned by CU. A CU is a geographic area that represents part or all of a surface drainage basin, a combination of drainage basins, or a distinct hydrologic feature. The USGS uses CUs for cataloging and indexing water-data acquisition activities.

The continental United States comprises over 2,100 CUs. CUs are fairly small; for example, 45 units fall partially or completely within the state of Virginia (see Figure 3).

RF3 is a powerful data source used in hydrologic applications for many reasons, including the following:

- RF3 has spatial network connectivity that topological upstream/downstream modeling tools use.

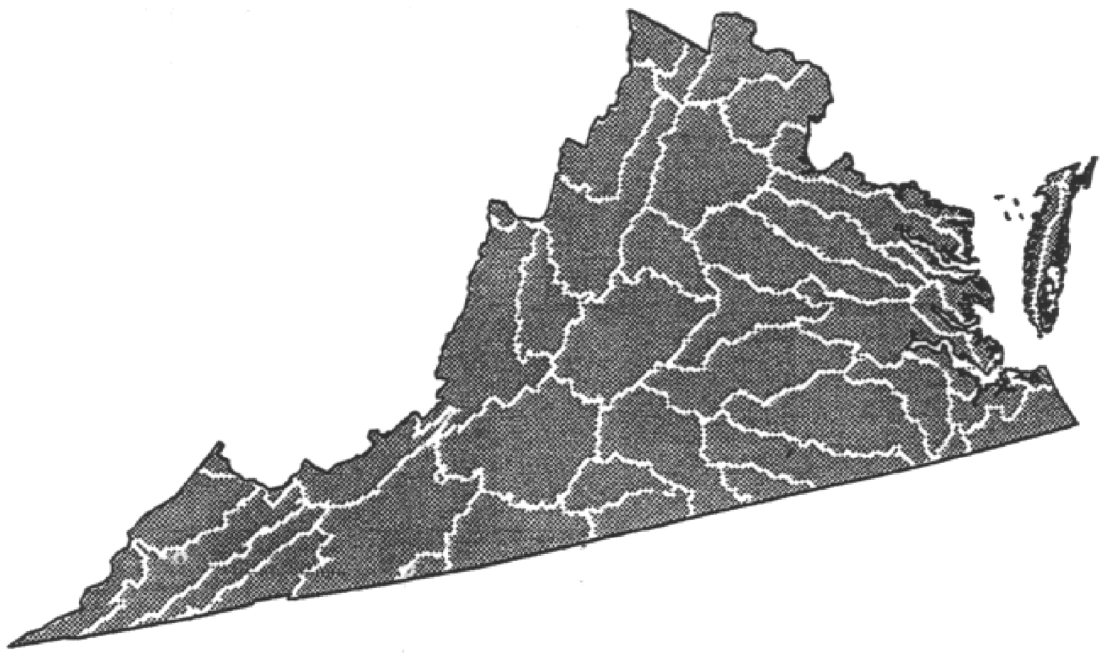


Figure 3. Cataloging units in Virginia.

- RF3 has attributes that describe connectivity, which offers the ability to accomplish upstream/downstream navigation analytically (without topological networking).
- RF3 has a simple and consistent unique numbering system for every stream reach in the United States.
- RF3 has built-in river mileage attributes that describe upstream/downstream distances along river reaches.

Use of RF3 in the Indexing Process

When importing Reach File data from EPA's mainframe computer, an arc attribute table (AAT) is automatically built for each RF3 coverage. The AAT contains the standard AAT fields, plus the items found in Table 1.

Table 1. Fields Found in Arc Attribute Table

12070104-ID	CU	SEG	MI	UP	DOWN
1	12070104	1	0.00	-1	0
2	12070104	1	1.30	-1	0
3	12070104	1	2.10	-1	0
4	12070104	2	0.00	-1	0
5	12070104	3	0.00	-1	0
6	12070104	3	1.15	-1	0
7	12070104	4	0.00	-1	0

The CU item stores the USGS CU number of this piece of RF3. Every arc in the coverage has the same value for CU.

The SEG item stores the number of the stream segment to which the particular arc is assigned. SEG numbers start at 1 and increase incrementally by 1 to 'N' for each CU. A SEG could represent all the arcs of a mainstream, the arcs of a tributary, or piece of a mainstream or tributary. SEG numbers were defined in the production of RF3.

MI stores the marker index for each particular arc. The MI resembles a mile posting along a stream. In reality, the MI field does not truly measure mileage along the RF3 stream network. It does, however, represent a method of producing a unique identifier (in combination with the CU number and the SEG number) for every reach in the United States (see Figure 4).

Together CU, SEG, and MI uniquely identify every arc in RF3 nationwide. These three items are combined in the redefined item called RF3RCHID. This provides a powerful scheme for consistently identifying locations along streams everywhere in the country.

Along with the AAT file, a second attribute file is automatically created for RF3 coverages. This file is always named COVER.DS3. The DS3 file stores a wealth of information about arcs in the coverage. Some of the important fields in the DS3 file contain:

- Upstream and downstream connectivity for navigating along reaches.
- Codes to describe the type of reach (e.g., stream, lake boundary, wide river).
- DLG major and minor attributes.

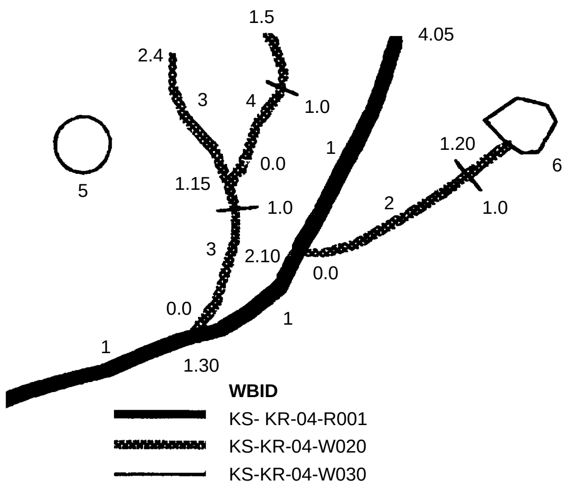


Figure 4. RF3, SEG, and MI data elements.

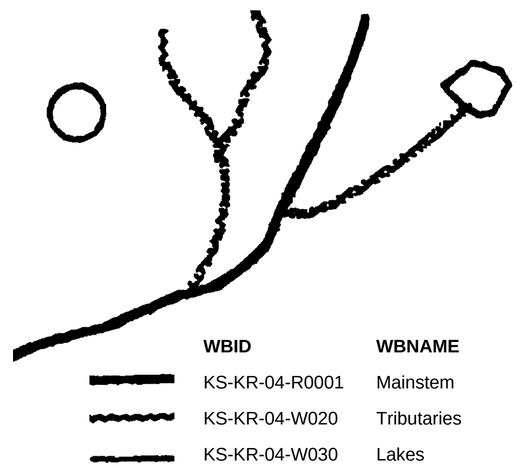


Figure 5. Potential definitions of water bodies.

Waterbody Locations

Because states define water bodies, they provide the only information on waterbody location. South Carolina was indexed to RF3 in 1992, followed by Virginia. Virginia indexed its water bodies using the PCRF program instead of indexing in a GIS environment with ARC/INFO.

PCRF is a PC-based system that indexes water bodies and locates other assessment data from WBS. PCRF stores the definitions of water bodies (including their location) in a file that is linked to other WBS database files that contain information about the assessment status and quality of the waters.

A water body is a set of one or more hydrologic features, such as streams, lakes, or shorelines, that have similar hydrologic characteristics. Water bodies are the basic units that states use to report water quality for CWA 305(b) requirements. Depending on the state's assessment goals and resources, water bodies can be defined in several ways, including (see Figure 5):

- All streams within a watershed
- All lakes and ponds within a watershed
- Sets of streams with similar water quality conditions

PCRF stores locational data for a water body with a unique WBID. WBS uses this WBID as a common field to relate the water body's definition and location to descriptive data about the water body's assessment status and quality. The two most important files used in PCRF are the SCRF1 and SCRF2 files.

The SCRF1 file simply lists the unique water bodies by state. Table 2 offers an example. The most relevant data for reach indexing in this file are the WBID, WBNAME, and WBTYPE, as defined by the state. The WBID, as stated, is a unique identifier for each water body the state has defined. The WBNAME stores a verbal description of the water body. Finally, the WBTYPE defines the type of water body; for example, R is for river, L is for lake.

The SCRF2 file contains an explicit definition of each water body. Because of the complexity involved in defining water bodies, this file may include more than one record for each water body. The SCRF2 file can be considered a waterbody definition *language* because it contains specific codes, attributes, and keys that can be converted into specific reaches on the RF3 data, if read properly (see Table 3). The WBBEGIN and WBEND fields contain explicit CU, SEG, and MI attributes to define the location of the starting point and ending point for the water body. The WBDIR field contains an attribute that describes whether to go upstream or downstream from the WBBEGIN to the WBEND. In addition, a blank WBEND field denotes that the water body should include all upstream or downstream reaches (depending on the WBDIR) of the WBBEGIN reach.

Virginia used PCRF to create an SCRF2 file that contains reach indexing expressions for all of their defined water bodies. ARC/INFO macros were then written to process this file and expand the expressions into the set of specific arcs that compose each water body. The macros will be described in more detail later.

Table 2. Example of SCRF1 File Data

WBID	WBNAME	WBTYPE	WBSIZE
KS-KR-04-R001	KANSAS RIVER	R	15.20
KS-KR-04-W020	LOWER WAKARUSA RIVER	R	61.60
KS-KR-04-W030	MUD CR	R	39.43
KS-KR-04-W040	CAPTAIL CR	R	15.63

Table 3. Example of SCRF2 File Data

WBID	WBDIR	WBBEGIN	WBEND	RFORGFLAG
KS-KR-04-R001	U	10270104001 0.00	1027010400115.20	2
KS-KR-0R-W020	D	10270104005 10.80	10270104005 0.00	2
KS-KR-04-W030	U	10270104059 12.05	10270104007 8.10	3
KS-KR-04-W040	U	10270104038 0.00		3

States that have not already generated indexing expressions in PCRF must provide locations in some other way. The most basic method is for the state to supply a set of 1:100,000 USGS quad sheets that they have marked up with locations of each water body. The maps can be used in conjunction with a digitizer to manually select the appropriate RF3 reaches and code them with the WBID.

The state of Ohio created a GIS database of its river reaches several years ago. The GIS coverage is representational in nature. The stream reaches are 'stick-figures' only. Generally, they fall along the paths of the actual streams, but they are schematic in nature and do not show the true shape of streams. The GIS layer, however, contains the attributes of Ohio's stream reach numbering system, which is used to identify water bodies as well. Ohio's river reach coverage contains information on the locations of water bodies and is being manually conflated to transfer the WBIDs to RF3. The conflation process will be covered later in this paper.

The state of Kansas had previously defined its water bodies on RF2, the precursor to RF3. Kansas defined some indexes by a set of RF3 SEG numbers in a CU and some by the RF3 reaches in a small watershed polygon within a CU. The locations were, in effect, defined within the WBID itself.

Indexing Procedures

Procedures developed for performing waterbody indexing include automated, semiautomated, and manual systems.

Automated Indexing Procedures

As stated, Virginia used PCRF to perform the indexing operation. The state delivered an SCRF2 file containing indexing expressions for all of its water bodies. AML programs were created to read the SCRF2 file and select the reaches specified by each indexing expression. The selected sets of reaches were then coded with the appropriate WBID. The macros were designed to run on one RF3 CU at a time, so the operator specified runs of up to 10 CUs at a time. The macros had to accommodate indexing expressions that included:

- Select reaches upstream of a specified location.
- Select reaches on a reach-by-reach basis.
- Select reaches within a given polygon area.
- Select shorelines of lakes or ponds given latitude and longitude coordinates.
- Select reach downstream from a specified location.

Kansas water bodies were also indexed through an automated process. Kansas supplied an ARC/INFO coverage of small watershed polygons (sub-CU polygons) containing a watershed identifier. The state's WBID contained all other information necessary to determine the RF3 CU and the set of reaches making up each water body. An example of a Kansas WBID is KS-KR-02-W030. This is explained by the following:

- KS refers to the state. All WBIDs in Kansas begin with KS.
- The second component (in this case KR) is an abbreviation of the basin in which the water body falls. KR indicates that this water body is in the Kansas-Lower Republican River basin.
- The third component contains the last two digits of the eight-digit CU number. Although basins comprise several CUs, the last two digits of each CU in a basin are unique; therefore, between the basin (e.g., KR) designation and the last two digits of the CU (e.g., 02), the complete eight-digit CU number in which the water body falls is defined.
- The next letter (in this case W) denotes whether the water body is defined by a watershed polygon (W), an RF2 SEG (R), or a lake or pond shoreline (L).
- Finally, the WBID ends with the number of the polygon (in this case 030) that contains the reaches for the water body in the watershed coverage.

The completed macros could index the entire state in a single run provided that all the WBIDs were contained in single file.

In all cases, Kansas has indexed to RF2 reaches. Only RF3 reaches originally created in RF2 production, therefore, are coded with a WBID.

Manual Indexing Procedures

Because Ohio already has a coverage of river reach codes, WBIDs from this coverage had to be transferred to the RF3 reaches they represent. This entailed using a manual conflation process. The operator displayed a CU of RF3 along with the Ohio river reach system for the same area. In a simple process of 'pointing and clicking,' the operator first selected an Ohio river reach arc, then the RF3 arcs that seemed to coincide. As each RF3 arc was selected, it was coded with the WBID of the previously selected Ohio river reach arc.

Other states that have no means of describing water bodies in electronic files may have to mark up paper maps to show waterbody locations. These maps can then be used in a manual process of selecting RF3 reach and coding them with WBIDs either in ARC/INFO or in PCRF.

Using the Route System Data Model To Store Water Bodies

Because water bodies can be defined as noncontiguous sets of arcs and portions of arcs, a robust linear database model is necessary to model these entities. ARC/INFO's route system data model seems well suited for this application. The route system data model allows one to group any set of arcs or portions of arcs into routes. Each route is managed as a feature in itself. Attributes of water bodies are stored in a route attribute table (RAT) and relate to all the arcs defined as the water body. Figure 6 helps illustrate the route system model.

Each route comprises one or more arcs or sections of arcs. ARC/INFO manages the relationship between arcs and routes in the section table (SEC). The structure of the SEC, which is an INFO table, is defined in Table 4. Table 5 reflects how the sections that make up the above routes would appear.

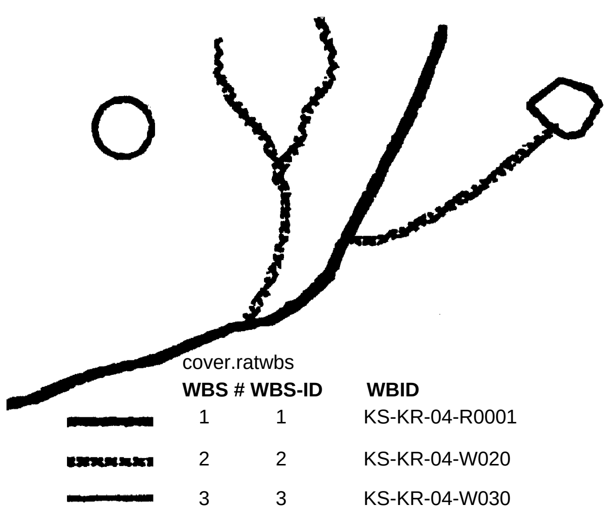


Figure 6. The route attribute table containing waterbody data.

Table 4. Definition of Structure of SEC INFO Table

ROUTELINK#	The route upon which the section falls
ARCLINK#	The arc upon which the section falls
F-MEAS	The measurement value at the beginning of the section
T-MEAS	The measurement value at the end of the section
F-POS	The percentage of the distance along the arc at which the section begins
T-POS	The percentage of the distance along the arc at which the section ends
SEC#	The internal identifier of the section
SEC-ID	The user identifier of the section

Table 5. How Sections Appear in SEC INFO Table

ROUTE-LINK#	ARCLINK#	F-MEAS	T-MEAS	F-POS	T-POS	SEC#	SEC-ID
1	1	0	1.30	0	100	1	1
1	2	1.30	2.10	0	100	2	2
1	3	2.10	4.05	0	100	3	3
2	4	0	1.20	0	100	4	4
3	5	0	1.15	0	100	5	5
3	6	1.15	2.4	0	100	6	6
3	7	0	2.5	0	100	7	7

Representing Water Bodies as Routes

ARC/INFO offers several ways of grouping sets of arcs into discrete routes. One can use ARCEDIT to select a set of arcs to group them into a route, or ARCSECTION or MEASUREROUTE in ARC to group arcs into routes. The method described here uses the MEASUREROUTE command. This method requires that the AAT or a related table has an attribute containing the identifier of the route to which an arc should be assigned. In the application the authors employed, they converted the SCRF2 file into an INFO table containing, for each arc in the coverage, the RF3RCHID of the arc and the WBID to which the arc should be assigned. The WBID item is used to group arcs into routes. One route exists for each unique WBID. Table 6 illustrates the table used in the MEASUREROUTE command method. This table is related to the AAT of the RF3 coverage by the RF3RCHID.

An RAT is automatically created for the coverage, which now can be related to other WBS assessment files for display and query. Figure 4 illustrates the RAT. The most important characteristic of the file is that it has only one record for each water body. This simplifies the display and query of water bodies based on water quality data.

Using EVENTS for Subwaterbody Attributes

Water bodies, as states define them, often constitute a gross aggregation of the water in an area. States often have more specific data about particular stretches of streams within a water body. A system is needed to

Table 6. Table Used in MEASUREROUTE Command Method To Group Arcs Into Routes

$RECNO	RF3RCHID	WBID
1	10270104 1 0.00	KS-KR-04-R0001
2	10270104 1 1.30	KS-KR-04-R0001
3	10270104 1 2.10	KS-KR-04-R0001
4	10270104 2 0.00	KS-KR-04-W020
5	10270104 3 0.00	KS-KR-04-W020
6	10270104 3 1.15	KS-KR-04-W020
7	10270104 4 0.00	KS-KR-04-W020
8	10270104 5 0.00	KS-KR-04-W030
9	10270104 6 0.00	KS-KR-04-W030

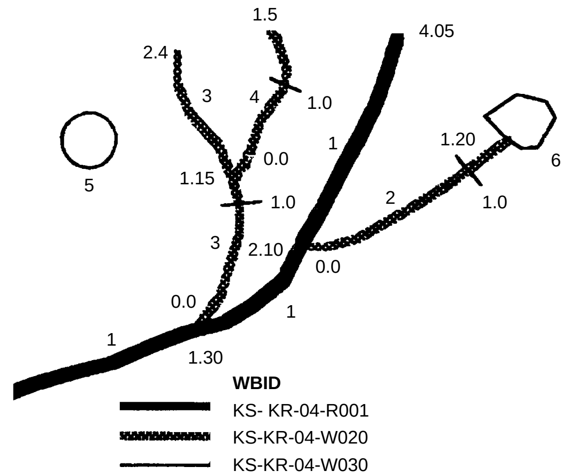


Figure 7. Measurements along SEGs.

query and display data at the subwaterbody level. ARC/INFO's dynamic segmentation tools and event tables are useful for this application. Once water bodies have been defined and reporting methods have been set up based on those water bodies, the task of redefining them is cumbersome.

Event tables can help to keep these waterbody definitions yet still offer the ability to store, manage, and track data at the subwatershed level. Event tables are simple INFO files that relate to route systems on coverages. This concept and data structure can work in conjunction with the predefined waterbody system. We have already seen how a route system called WBS is created in RF3 to group arcs into waterbody routes. This works quite well when displaying water bodies and querying their attributes. A route system based on the WBID cannot, however, act as an underlying base for subwaterbody events because the measures used to create the WBS route system are not unique for a particular route. For example, in the route depicted in Figure 7, three locations are defined as being on WBID KS-KR-04-W020 and having measure 1.0.

The mileage measurements along SEG, however, are always unique (see Figure 8). To use EVENTS, therefore, a second route system must be created based on the RF3 SEG attribute, which provides a unique code for each CU.

The ARCSECTION command, instead of the MEASUREROUTE command, is used to create the SEG route system. This is because the measurement items (MI on the AAT and UPMI on the DS3) already store the summed measures along particular SEGs. Table 7 lists the contents of the resulting RAT table.

Because the name of the route system is SEG, the SEG# and SEG-ID are the names of the internal and user IDs. The SEG item contains the actual SEG number in the RF3 coverage. Because the SEG numbers for each RF3 CU coverage start at 1 and increase incrementally by 1, the SEG item looks much like the SEG-ID and SEG#.

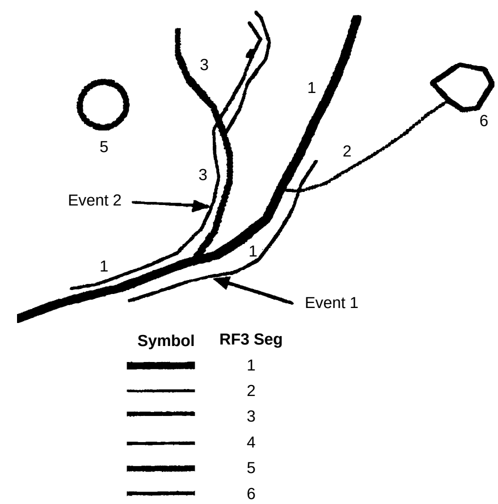


Figure 8. Events located on RF3 data.

Table 7. Route Attribute Table

SEG-ID	SEG#	SEG
1	1	1
2	2	2
3	3	3

Event tables contain a key item, the WBID or SEG, to relate them to the appropriate route system (see Figure 8). They also contain locational information on where to locate the events on the route (either WBID to indicate the water body on the WBS route or SEG to identify the route in the SEG route system). Separate event tables can then relate use support, causes, and sources as linear events. FROM and TO store the starting and ending measures for each event. Using event tables allows us to apply many useful cartographic effects (e.g., hatching, offsets, text, strip maps). Events can be queried both in INFO and graphically. Event data can help in producing overlays of two or more event tables.

An event table can display use support information (see Table 8). WBS users can update their event tables using RF3 maps supplied by EPA without having proficiency in ARC/INFO. ARC/VIEW2 is expected to support events and route systems. This will give users powerful tools for spatial query of assessment data. Developing event tables would also display and query data on the causes and sources of use impairment. These events can be offset and displayed to show the areas of interaction. More permanently, preparing line-on-line overlays can show intersections and unions.

An alternative approach is to use an EVENT-ID as a unique identifier for each event. The SEG field stores the number of the route (SEG) upon which the event occurs. FROM and TO store the beginning and end measures along the route upon which the event occurs. WBID contains the identifier of the water body upon which the event occurs (see Table 9). An event can occur within a single SEG, across two or more SEGs, within a single water body, or across two or more water bodies.

Additional attribute tables can be created to store descriptive attributes for each event. These tables would resemble the SCRF5 and SCRF6 files except that instead of using the WBID to relate to a water body, a field called 'EVENT-ID' would link the use, cause, and source data to a particular event (see Table 10).

Both approaches offer some advantages. In either case, they allow us to map our water quality assessment data and communicate it in a meaningful and useful way.

Table 8. Event Table That Reflects Use Support Information

SEG	FROM	TO	WBID	USE	USE SUPPORT
1	0.80	1.30	KS-KR-04-R0001	21	Fully
1	1.30	2.10	KS-KR-04-R0001	21	Partial
1	2.10	2.31	KS-KR-04-R0001	21	Not supported
1	0.50	1.30	KS-KR-04-R0001	40	Threatened
3	0.00	1.15	KS-KR-04-W030	21	Fully
4	0.00	2.5	KS-KR-04-W040	40	Not supported

Table 9. Using EVENT-ID as a Unique Event Identifier

EVENT-ID	SEG	FROM	TO	WBID
1	1	0.80	1.30	KS-KR-04-R0001
1	1	1.30	2.10	KS-KR-04-R0001
2	4	0.00	2.5	KS-KR-04-W040

Table 10. Using EVENT-ID To Link Use, Cause, and Source Data to an Event

EVENT-ID	ASCAUSE	ASSOURC
1	900	1200
1	-9	1100
1	0500	1100
2	1200	9000
2	0900	8100

Environmental Management Applications

Ecological Land Units, GIS, and Remote Sensing: Gap Analysis in the Central Appalachians

Ree Brannon, Charles B. Yuill, and Sue A. Perry
West Virginia University, Morgantown, West Virginia

Abstract

The gap analysis team in West Virginia is assessing the state's natural communities as part of a nationwide, comprehensive planning effort. Underrepresented or unrepresented habitats represent gaps in the present network of conservation lands and conservation activities. After identifying these gaps, we can assess whether our current management direction will maintain natural diversity and will prevent additional species from being classified as threatened or endangered.

The relationship between vegetation and ecological variables serves as the basis for classifying ecological land units. To characterize the ecological land units, many layers of physical data can be integrated in a geographic information system (GIS). Satellite imagery and videography map existing conditions over the state. The existing vegetation is classified to reflect physiognomic and floristic elements to correlate with vertebrate and butterfly habitat requirements. This correlation of vegetation and wildlife habitat creates mappable habitat types. Analysis of these habitat types with landownership data indicates where the species-rich areas occur in the landscape and whether the most species-rich areas are protected.

Introduction

To respond to the urgency of habitat loss and its effect on species diversity, scientists must implement a methodology for rapid assessment and documentation of natural communities at a scale pertinent for regional management activities.

Geographic information systems (GIS) and remote sensing support the development of ecological land classifications over large regions. GIS-based mapping of ecological land classes allows users to combine and display environmental variables for spatial modeling and refinement of ecological land units (1).

The Gap Analysis Project is a comprehensive planning effort for land conservation in the United States. The objective of the Gap Analysis Project is to identify species, species-rich areas, and vegetation types underrepresented or unrepresented in existing biodiversity management areas. Unprotected communities are the gaps in the conservation strategy. The Gap Analysis Project is not merely identifying communities with the largest number of species; its ultimate goal is to identify clusters of habitats that link the greatest variety of unique species.

Local areas with considerable diversity of habitat or topography usually have richer faunas and floras (2). Nature reserves, which incorporate a variety of habitats, may be the best guarantee for long-term protection of biodiversity. By protecting species-rich regions, we can reduce the enormous financial and scientific resources needed to recover species on the brink of extinction.

The West Virginia Gap Analysis Project began in 1991. The objective is to map existing vegetation and to use that as the foundation to model potential distribution of vertebrate and butterfly species. High cost precludes intensive field inventory and monitoring of wildlife. Therefore, habitat modeling is critical to predict wildlife species composition and potential ranges over the various landscape types of West Virginia. Lastly, the vegetation map will provide a record of the existing habitat to use in monitoring changes due to human activities and natural disturbances.

Pilot Study Area

Distribution of wildlife and plant communities will be modeled for the entire state. Initially, we will focus on a smaller pilot study area. This region includes approximately 50,000 hectares in the central Appalachian Mountains and spans several physiographic provinces and vegetative communities. Generally, soils in the pilot study area are of two kinds: acidic soils that develop a clay horizon from extensive leaching over time and younger soils that are found on steep slopes and where environmental conditions, such as cold climate, limit soil development (3, 4).

The vegetation types include spruce-fir, oak-pine, high elevation bogs, northern hardwoods, Appalachian oak, mixed-mesophytic, open heath barrens, and grass balds. The mixed-mesophytic, the most diverse in West Virginia, lies primarily west of this area, but localized stands do occur in the lower elevations. Cover types within the Appalachian oak and mixed-mesophytic types are not discrete and will be difficult to delineate.

The pilot study area includes a variety of land uses, such as residential, commercial, industrial, mining, and agriculture. Portions of the Monongahela National Forest in this area are the Fernow Experimental Forest, and the Otter Creek and Laurel Forks Wilderness Areas.

Methods

The following discussion describes the methods formulated and data compiled for the West Virginia Gap Analysis Project.

Describe Ecological Land Units With Existing Vegetation

Davis and Dozier (1) note that a landscape can be partitioned by ecological variables, which contributes to an ecological land classification. This process is applied frequently to analysis and mapping of natural resources. Davis and Dozier classified vegetation in California based on the documented associations of vegetation with terrain variables. They based this approach on the assumptions that the arrangement of natural landscape features is spatially ordered by an ecological interdependence among terrain variables and that actual vegetation is a reliable indicator of these ecological conditions. Similar documentation exists for the distribution of vegetation types in West Virginia, and the gap analysis team is proceeding along a similar course.

West Virginia lies in two major provinces, the Eastern Broadleaf Forest and the Central Appalachian Broadleaf Forest-Coniferous Forest-Meadow Provinces (5). Within these provinces are several broad vegetation types. The gradient diagram in Figure 1 (6) illustrates the range of these types. The vertical axis represents elevation in feet. Three vegetation types emerge distinctly along the elevation gradient. The horizontal axis is not quite as explicit. This gradient spans moist, protected slopes to dry, exposed ridgetops, and the vegetation types are much less distinct. Drier oak and pine types occur almost exclusively on exposed ridgetops. The vegetation types along the horizontal axis are the mixed mesophytic forest association of the Allegheny and Cumberland Mountains and can have 20 to 25 overstory and understory species per hectare in North America (7).

The distribution of the vegetation along gradients such as elevation and soil moisture lends itself to a GIS analysis. Physical data such as elevation and soil moisture regime can be incorporated into a GIS. These layers of information can be manipulated graphically or mathematically to model the spatial distribution of vegetation types or to provide useful ancillary data for classification of satellite imagery (see Figure 2). Much of this data is digital and can be used to substantially reduce the time required to develop a database. To standardize output, members of the national gap team have specified ARC/INFO as the software to generate final products.

Classify Satellite Imagery To Create a Map of Current Distribution

Remote sensing provides an effective means to classify forests, and the Gap Analysis Project has successfully used it in the western United States (1, 8, 9).

The Gap Analysis Project is using Landsat Thematic Mapper imagery in all states to standardize the baseline information. The hypothesis is that the spectral data from the imagery is related to the distribution of the ecological land units and land use across the landscape. The data include all spectral bands, except thermal, for the entire state. The West Virginia project is using two seasons of data, spring and fall. Temporal changes, which record phenologic variation in the deciduous species, enhance classification accuracy. The spectral resolution is 30-meter pixels. This is equivalent to approximately 1/6 hectare (1/2 acre). Our final product will be a series of 1:100,000 maps. The minimum mapping unit is 40 acres.

The mountainous terrain in the Appalachian Mountains offers disadvantages and advantages for using remotely sensed data. Irregular topography can cause inconsistencies in the spectral data that diminish the classification accuracy. Similar cover types may have different spectral signatures; for instance, if one stand is in sunlight and the other is shaded. Also, phenology can vary due to microclimatic influences. Conversely, topographic features influence the distribution of vegetation types, and ancillary data, such as digital elevation models (DEMs), enhance classification results of the imagery. The West Virginia gap analysis team selected the following strategy for image classification.

1. Stratify the imagery using ecological units based on a hierarchical scheme. Bauer et al. (10) found that an initial stratification of physiographic regions was necessary to reduce the effect of broadscale environmental factors caused by changes over latitude. Therefore, stratification enhanced the

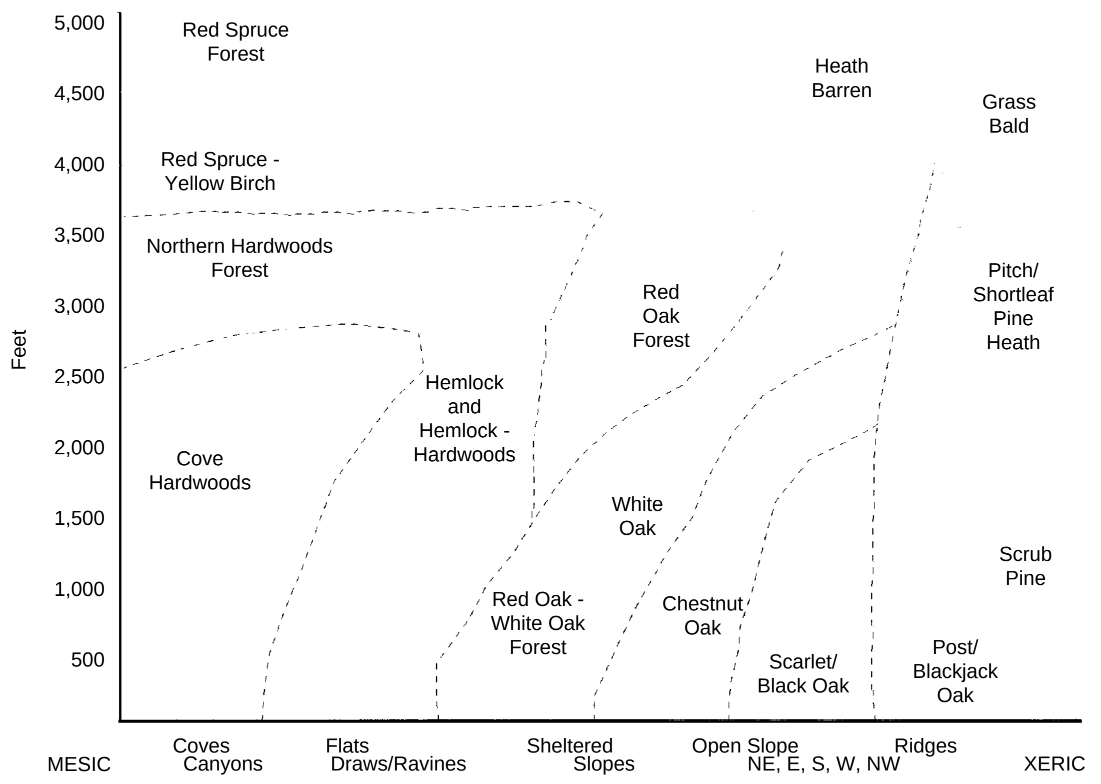


Figure 1. Environmental gradients for vegetation.

efficiency of the training data. An interagency committee, including ecologists from the Monongahela National Forest, West Virginia Division of Forestry, and geologists from the State Geologic Survey, generated a draft map of physiographic regions. They delineated sections based on geomorphology and climate. Sections were divided into subsections: those most typical of the section or those that are transitional, or irregular, to the section. Figure 3 is a draft map of these sections and subsections in West Virginia.

2. Classify stratified imagery using the ancillary data. High resolution imagery has not been used widely in the eastern United States, where forests are not homogeneous stands of relatively few tree species as they are in the West. Researchers who classified eastern forests from satellite imagery attempted to find the most distinctive spectral band combinations for discriminating cover types (10-13). One recent technique (14) uses a nonparametric approach that combines all spectral and informational categories to classify imagery. We are testing a variety of methods such as nonparametric processes, traditional clustering techniques, and use of derived vegetation indexes to find the most successful method.

3. Assess accuracy with random plots from videography. Videography will be acquired in the spring of 1995. Aerial transects, which extend the length of the state, will be flown with approximately 30-kilometer spacing. By regulating flight altitudes, the resolution per frame can be captured at 1 kilometer per frame. About 7,000 frames will be collected, which make up a 3 percent sample of the state. About 2 to 3 percent of the videography frames will be field verified. With this strategy, we will test the effectiveness of using videography, instead of intensive field plots, to verify classification of satellite imagery. Areas of special interest, which the systematic transects may not capture, will require separate flights. The bulk of the videography will provide training data for supervised image classification. The remaining frames will be used to assess the accuracy of the classification.

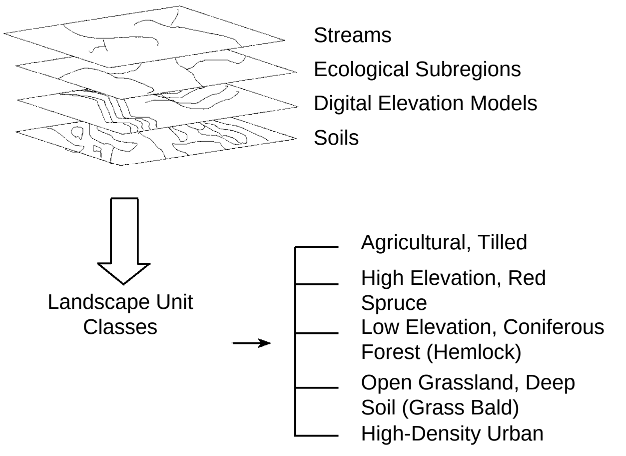


Figure 2. GIS and the development of ecological land units.

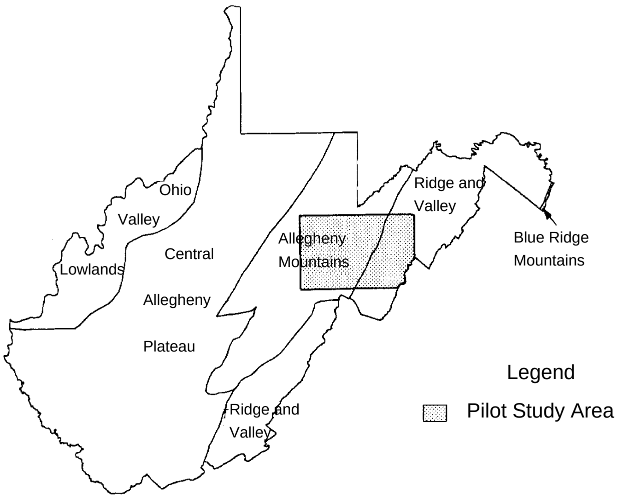


Figure 3. Physiographic regions of West Virginia and pilot study area.

In summary, 100 percent of the state will be classified using the Thematic Mapper imagery. Aerial videography, covering approximately 3 percent of the state, will help to verify the image classification, and 2 to 3 percent of the videography will be verified from transects on the ground.

4. Determine sources of data for image classification. Due to the increasing interest in GIS, digital data are more readily available from a variety of sources, such as the federal government, state agencies, and private companies. Acquisition of available data sets can substantially reduce the time and cost of database development. Users must bear in mind, however, that databases are developed with differing objectives and techniques, so one must consider scale and standards of production when deciding which data sets are appropriate for project design. The West Virginia gap team determined that the following GIS coverages are important for image classification.

The U.S. Geological Survey's (USGS's) graphic information retrieval and analysis system (GIRAS) earmarked land use/land cover data. The classification was done several years ago, and although these data are not current, they provide excellent information concerning urban and agricultural land use. Land-cover categories represent Level II classifications from Anderson's (15) system. The maps are produced at a 1:250,000 scale, so they require few GIS operations to piece together a regional coverage of land use.

The Southern Forest Experiment Station mapped U.S. forestland using advanced very high resolution radiometer (AVHRR) satellite imagery (16). This imagery is relatively current (1991 to 1992), but the resolution is coarse at 1 kilometer per pixel (100 hectares or 247 acres). The classes are based on Forest Inventory and Analysis plots established by the U.S. Department of Agriculture (USDA) Forest Service and Küchler's (17) potential natural vegetation types. We are using the maps to depict broad changes in forest type over a region, such as the state of West Virginia. This coverage does not show land use.

The eastern region of the Nature Conservancy has completed a draft of the classification of the terrestrial community alliances (18). The classification hierarchy is that prescribed by the national gap team, and as such, reflects physiognomic and floristic characteristics necessary for correlating vegetation structure and floristic composition with vertebrate habitat requirements. The descriptions include the range of alliances and characteristic species of the overstory, understory, and herbaceous layer. This provides information on associated species not detectable by image classification.

The National Wetlands Inventory data are available digitally. Maps have been digitized at a scale of 1:24,000, and the classification scheme is from Cowardin et al. (19). Coverages come with attribute data for each polygon, arc, or point as needed. A labor-intensive effort is required to join the maps in a GIS for an area the size of West Virginia, but the information will be invaluable for masking water and forested wetlands on the satellite imagery. The U.S. Fish and Wildlife Service includes detailed instructions for converting the data to coverages.

Field data, much of it already digital, has been acquired from many sources. Commercial timber companies provided data for timber stand composition and age groups. The USDA Forest Service ecologist conducted transects throughout the Monongahela National Forest to characterize ecological land units. Forest inventory and analysis plots are also available. We acquired these data to verify videography classification.

The USGS has digital data of terrain elevations. The West Virginia gap analysis team acquired 3 arc-second DEMs as an additional band in the satellite imagery. We will use these data to generate coverages of slope, aspect, and elevation classes to further stratify the physiographic regions of the state. This will increase the accuracy of the classification.

The Soil Conservation Service created a statewide database called STATSGO, produced at a scale of 1:250,000. For West Virginia, the map of the soil mapping units consists of approximately 450 polygons. Each mapping unit is an aggregation of soil components that occupy a certain percentage of the mapping unit area. The database is extensive and includes information on soil attributes such as soil taxonomy, soil chemistry, soil structure, and interpretations for erodability and wildlife habitat. For an attribute such as soil temperature, each component has an individual value so that each mapping unit may have several different values for soil temperature. Attribute information is difficult to query in ARC/INFO, where there is a one-to-many relationship between polygons and database entries (for instance, each mapping unit, or polygon, has several soil components). We found that exporting the attribute information from ARC/INFO to another software package such as Excel was easier. The values can be aggregated by attribute and then imported into ARC/INFO to produce individual coverages such as soil texture, soil depth, or soil group. STATSGO data can provide useful information for the physical variables that influence vegetation, such as soil moisture and nutrient availability.

To review, the project researchers will first identify the physical parameters that govern the distribution of ecological land units and the existing vegetation in the state. Then, the team will gather applicable ancillary data of physical data in GIS to support the image classification. Once the imagery has been classified, the wildlife models can be incorporated.

Integrate Terrestrial Vertebrate Models

Once classification is completed for the state, the gap analysis team will integrate terrestrial vertebrate models. Concurrent with the image classification, the team will develop a species profile for each vertebrate species known to occur in the state. These profiles, when completed, will be condensed into rule-based models for associated species that can be linked to the ecological land units (see Figure 4). This step will create habitat types. After integration, the team will generate maps that display species richness of vertebrates for each habitat type (see Figure 5). These maps link spatial data to the species database. This enables users to identify areas in the landscape that combine habitats with the greatest number of unique species. When a coverage of landownership is overlaid on this map, land managers or conservation groups can take a proactive stance to seek protection of critical habitats. Additional analyses that users can perform are displays of the potential distribution of vertebrate groups, such as upland salamanders. Another analysis would be to report the species that occur in the fewest habitats and that would be most vulnerable to landscape changes. Clearly, GIS provides a powerful environment for quick and efficient retrieval of spatial data for management decisions.

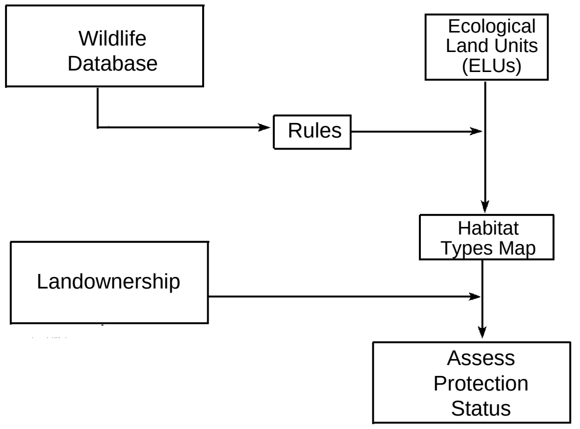


Figure 4. Informational flow chart for wildlife data.

Summary

To summarize, the West Virginia gap analysis team is assessing the natural communities in the state as a part of the national comprehensive planning effort. We need to conduct the assessment rapidly, compiling existing information and integrating these data with GIS and remotely sensed data. Ecological land units are classified according to the relationship between vegetation and ecological variables. Satellite imagery is used to map existing conditions over the state. The existing vegetation is classified to reflect physiognomic and floristic elements to correlate with vertebrate and butterfly habitat requirements.

The gap analysis team is using many widely available data sets such as DEMs, land use/land cover, wetlands inventory, and soils data. While these can reduce the time and cost of developing an ecological database, they do present implications for project design and accuracy. When the user combines maps of different scales, accuracy is constrained by the map with the smallest scale. Additionally, data sets may be constructed with objectives for an intended use that is not compatible with project needs. The classification of AVHRR data reflects forest types but not land use, so another source may be required for these data.

The Gap Analysis Project is not a substitute for intensive biological studies at a fine scale. It is merely a quick assessment at a broad scale to provide information on existing conditions. While accumulating data and modeling potential wildlife distributions, we will identify where inventory data may be lacking. Additional work must be done to verify wildlife models and the classification of vegetation, but this preliminary analysis will be a valuable framework that will direct future studies of biological diversity. Finally, this effort will provide a data set that can be used to monitor changes to land cover and land use.

Acknowledgments

This study is being funded by the Cooperative Research Units Center, National Biological Survey, United States Department of the Interior. This manuscript is published as Scientific Article No. 2499 of the West Virginia Agriculture and Forestry Experiment Station.

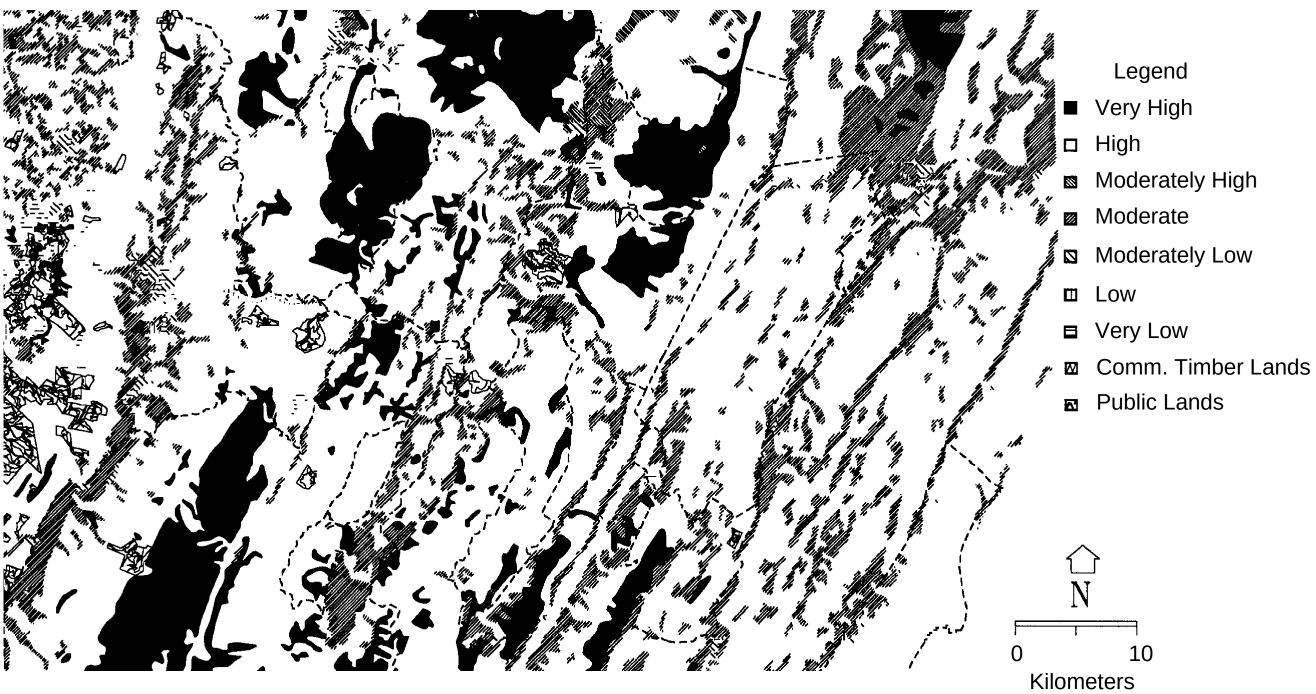


Figure 5. Relative species richness.

References

1. Davis, F.W., and J. Dozier. 1990. Information analysis of a spatial database for ecological land classification. Photogramm. Eng. and Remote Sensing 56:605-613.

2. Scott, J.M., B. Csuti, K. Smith, J.E. Estes, and S. Caicco. 1991. Gap analysis of species richness and vegetation cover: an integrated biodiversity conservation strategy. In: Kohm, K., ed. Balancing on the brink of extinction: the Endangered Species Act and lessons for the future. Washington, DC: Island Press. pp. 282-297.

3. Stephenson, S. 1993. Upland forests of West Virginia. Parsons, WV: McClain Printing Co.

4. Buol, S.W., F.D. Hole, and R.J. McCracken. 1980. Soil genesis and classification, 2nd ed. Ames, IA: Iowa State University Press.

5. Bailey, R.G. 1994. Ecoregions and subecoregions of the United States (map). U.S. Department of Agriculture, Forest Service. Revised February 23.

6. Fortney, R. 1994. Vegetation pattern of central Appalachian Mountains. (Unpublished).

7. Barbour, M.G., J.H. Burk, and W.D. Pitts. 1987. Terrestrial plant ecology, 2nd ed. Menlo Park, CA: Benjamin/Cummings Publishing Company.

8. Davis, F.W., D.M. Stoms, J.E. Estes, J. Scepah, and M. Scott. 1990. An information systems approach to the preservation of biological diversity. Int. J. Geogr. Info. Syst. 4:55-78.

9. Scott, J.M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T.C. Edwards, Jr., J. Ulliman, and R.G. Wright. 1993. Gap analysis: A geographic approach to protection of biological diversity. Wildl. Monogr. 123:1-41.

10. Bauer, M.E., T.E. Burk, A.R. Ek, P.R. Coppin, S.D. Lime, T.A. Walsh, D.K. Walters, W. Befort, and D.F. Heinzen. 1994. Satellite inventory of Minnesota forest resources. Photogramm. Eng. and Remote Sensing 60:287-298.

11. Benson, A.S., and S.D. DeGloria. 1985. Interpretation of Landsat-4 thematic mapper and multispectral scanner data for forest surveys. Photogramm. Eng. and Remote Sensing 51:1,281-1,289.

12. Hopkins, P.F., A.L. Maclean, and T.M. Lillesand. 1988. Assessment of thematic mapper imagery for forestry application under lake states conditions. Photogramm. Eng. and Remote Sensing 54:61-68.

13. Schriever, J.R., and R.G. Congalton. 1993. Mapping forest cover-types in New Hampshire using multitemporal Landsat Thematic Mapper data. Papers presented at ASCM/ASPRS Annual Convention and Exposition, New Orleans, LA (February). pp. 333-342.

14. Cetin, H., T.A. Warner, and D.W. Levandowski. 1993. Data classification, visualization, and enhancement using n-dimensional probability density functions (nPDF): AVIRIS, TIMS, TM, and geophysical applications. Photogramm. Eng. and Remote Sensing 59:1,755-1,764.

15. Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer. 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 965. Washington, DC: U.S. Geological Survey.

16. Zhu, Z., and D.L. Evans. 1992. Mapping midsouth forest distributions: AVHRR satellite data and GIS help meet RPA mandate. J. Forestry 27-30.

17. Küchler, A.W. 1964. Potential natural vegetation of the conterminous United States (map and manual). American Geographical Society Special Publication 36. New York, NY: American Geographical Society.

18. Sneddon, L., M. Anderson, and K. Metzler. 1994. A classification and description of terrestrial community alliances in the Nature Conservancy's eastern region: First approximation. Washington, DC: U.S. Department of the Interior Fish and Wildlife Service.

19. Cowardin, L.M., V. Carter, F.C. Golte, and T.E. LaRoe. 1979. Classification of wetlands and deepwater habitats of the United States. Report No. FWS/OBS-79/31. Washington, DC: U.S. Department of the Interior Fish and Wildlife Service.

A GIS Strategy for Lake Management Issues

Michael F. Troge
University of Wisconsin, Stevens Point, Wisconsin

Abstract

Lake management plans are crucial to the sustained life of a lake as it experiences pressures from human as well as environmental activities. As proven in the past, geographic information systems (GIS) can meet the needs of most if not all environmental entities. Applying GIS to lakes and lake management, however, is a fairly new concept because most previous work focused on the terrestrial realm. Future studies must address problems relating to dimension, but adopting certain methods (i.e., cross-sectional coverages) can help lake managers plan for critical lake issues. By using sufficiently planned coverages, lake quality data management coverages can increase storage and/or analysis efficiency. After evaluating certain management criteria, a lake management plan can be derived and set up as a coverage. These criteria can then correspond collectively to form management zones within a lake. Each of these zones has its own set of management goals to which all lake users must strictly adhere.

Introduction

The importance of maintaining lake quality has long concerned recreationalists and ecologists. The multifaceted interrelationships of the lake environment, however, usually make proper assessment and analysis of lake quality information difficult. Over the last decade, assessment has become easier due to the increased use and acceptance of geographic information systems (GIS). This computer-based tool has allowed successful integration of water quality variables into a comprehensible format.

One area of the environmental sciences that has neglected GIS is lake management. This paper presents an alternative method for using a traditional two-dimensional GIS for viewing, querying, and displaying three-dimensional information—in this case, lake quality information and lake management criteria. Lakes, unlike geologic entities, offer a three-dimensional realm that humans can fully penetrate without a great amount of effort. Lakes also contain a complete aquatic environment of physical, chemical, and biological entities that humans can effectively observe and analyze. This paper does not discuss the issue of dimension; however, future studies, primarily those relating to the creation of three-dimensional GIS, should address this issue.

GIS allows incorporation of a multitude of environmental variables (e.g., water chemistry, geologic strata) into a synergism of the many coexisting variables of the lake environment. The ability of a GIS to "capture, manipulate, process, and display spatial or georeferenced data" is now well known and accepted (1). Surprisingly though, GIS is rarely used for lake management databases and associated water quality analysis.

The few examples that exist include Schoolmaster's Texas Water Development Board System (2), which examined water use on a county basis, and RAISON GIS (3), which is an expert computer system implementing proper application of hydrologic principles to a particular lake. Many other systems are simply database collection storehouses of lake information, such as the Galveston Bay National Estuary Program (4).

One notable example is the LAKEMAP program (5). This extensive and comprehensive GIS spans the entire United States covering approximately 800,000 lake sampling sites from the U.S. Environmental Protection Agency's (EPA's) STORET system. LAKEMAP uses both a database management and mapping display system, allowing retrieval of information for specific sites or aggregation of regional areas. This GIS is unique because it examined the creation of standards that could be used across the country in database development and the presentation of that data.

Using GIS in a lake management study inspires many questions because of the lack of existing research and the absence of any true standards. For example, how should one create a lake quality database for general purpose management? Is visualizing the integration of several variables within the lake ecology possible? Can one examine temporal changes in pH? Many technical and logistical GIS questions therefore existed when the Legend Lake study began.

Background of the Legend Lake Study

Legend Lake (see Figure 1) is located in Menominee Reservation, which is in Menominee County in northeastern Wisconsin. Legend Lake is a 1,230-acre impoundment comprising eight natural drainage lakes that a single stream once connected. In the late 1960s, a plan was introduced to convert this area into an impoundment/recreational area, and construction soon began. The ecology and hydrology had not been seriously evaluated since the development was finalized, hence the Legend Lake project was designed in cooperation with EPA, Wisconsin Department of Natural Resources (WDNR), Menominee Reservation and County, Legend Lake District, and the Legend Lake Property Owners Association. This intensive study spanned the qualitative and quantitative aspects of surface water, ground water, sediment, and aquatic plants, as well as human influences on the Legend Lake watershed and surrounding areas.

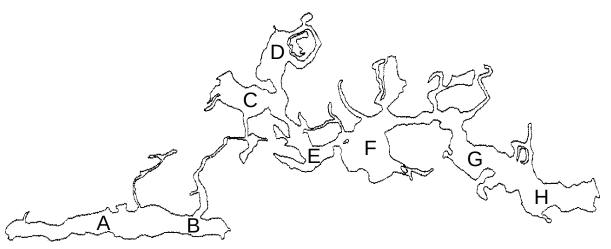


Figure 1. Legend Lake with basin identifiers.

Initially, questions needed to be answered regarding aquatic plants and sediment and their influences on lake management strategies. Because the scope of the Legend Lake study included subjects such as surface- and ground-water chemistry, land use and development, septic system impacts, and recreational stress, all these factors needed to be considered in determining optimum management strategies. In addition, the study addressed the question of GIS's ability to alleviate some technical aspects of deriving and presenting a lake management plan. Given these questions, the goal was to integrate collected data into a GIS database to create a prototype standard for future lake studies, as well as to present new techniques for visualization and analysis of lake quality data.

GIS/Database Creation

Several techniques were used to best manage the data for a three-dimensional system: cross-sectional coverages (described later), data summary coverages, and multidate coverages. The latter two coverage types are traditional coverages that contain general lake information excluding water column data. These techniques work fairly well for data storage and visualization; however, they were not sufficient for determining the geographic areas *within* the lake that required intensive management decisions as opposed to areas that needed little attention.

During formulation of a new coverage design, the study focused on the littoral zone, which is usually defined as that area of the lake with a depth of 15 feet or less. Most management concerns deal with the littoral zone because most recreational and ecological activities occur in this zone. One of the most pressing issues in lake management is aquatic plant growth. The fact that most aquatic plant growth is confined to the littoral zone reinforced the decision to use the littoral zone as the primary sink for potential management decisions.

To curtail the dimensional problem, depth was basically ignored. This allowed for easier delineation of areas within the lake. This, in turn, facilitated classifying areas into management zones to implement varying degrees of activity, ranging from casual to intensive efforts. Thus, management recommendations for a particular zone were the same at a depth of 2 feet as at a depth of 12 feet. This greatly reduced complication of the model and centered effort on the areal extent of the lake. It also facilitated visualization of management decisions by professionals and lay people.

Problems can arise when combining depths for management considerations, as this technique did. For example, a littoral zone that contains a gentle slope usually does not receive the same attention as a littoral zone with a very abrupt slope because the littoral zone with the gentle slope contains more area. Thus, if a lake manager recommends restricting boat traffic in a management zone (a section of the littoral zone) that has a gentle slope to promote wildlife habitat, the amount of area available for boaters would decrease significantly. In this situation, primary activity would be most crucial in shallow areas where wildlife or waterfowl predominate rather than in more open water areas where boaters predominate. Situations like these may require the creation of management subzones when setting up a lake management GIS and database to increase efficiency of lake area use and increase support by lake users.

Many criteria can affect the decisions made for a management zone. For instance, if the lake manager recognizes excessive, unhealthy weed growth in a particular zone, the lake manager may recommend extensive weed harvesting to neutralize the situation. An adjacent zone may have very little weed growth and may not require weed harvesting. Criteria such as these must be recognized before constructing the GIS. Table 1 lists some common criteria to consider. Generally, management criteria include anything that is influent on the

Table 1. Criteria To Consider When Creating Lake Management Plans

Environmental Variables	Artificial Variables
Aquatic plants	Adjacent land use
Sediment	Septic systems
Ground water	Development
Surface water	Construction sites
Wildlife/waterfowl	Fuel leaks
Fisheries	Shoreland zoning
Climate	Population density
Wind	Primary lake use (recreational)
Geology/geography	Nutrient loading
Adjacent natural land cover	Visitor use
Natural nutrient loading	
Hydrologic characteristics	

shoreline and lake itself and any influence the lake has on the shoreline or adjacent shorelands.

A primary concern when accessing lake data from a computer database is being sure to query the correct lake. For example, searching a database for all data on "Sand Lake" would be a legitimate action, except that the database may include close to 200 lakes with the name of Sand Lake. This is one of the main reasons why the WDNR developed a system known as the Water Mileage System (6). Based on logical criteria, each water body (e.g., lakes, streams, sloughs) receives a unique six- or seven-digit number called the waterbody number. Thus, if the number 197900, assigned to Sand Lake near Legend Lake, is the query subject, then the output should include all data for this particular Sand Lake. Because having a unique identifier for each specific entity in a GIS database is ideal, the waterbody number was used, and all sampling performed on this lake will be linked with this number.

Examples of Types of Coverage

Management Zone Coverages

Figure 2 and Table 2 together show how a potential lake management GIS and plan might work. The lake manager can easily manage and frequently update this system if necessary, or the system can serve as a long-term plan to consult for all decision-making. A plan of this sort specifically emphasizes areas that need intensive management over areas that may need frequent monitoring. It provides specific instructions for plan implementation, leaving little guess-work for the manager. This technique is also visually informative to the lake user because the user can easily discern areas of concern. This example is hypothetical, but a plan is being formulated based on the information collected during the Legend Lake study.

Cross-Sectional Coverages

Another technique currently included in the Legend Lake study entails the z dimension. Cross-sectional views of each lake basin, derived from 1992 lake contour maps, provided a more detailed description of the lake bottom. These cross sections were then digitized and transformed into GIS coverages. For each lake basin, 22 tests were conducted on the deepest part of the lake at several different depth intervals along the water column. These data provided valuable information on the way various chemical and biological attributes react to depth. Using GIS, a point could represent each depth where data were collected. These points could actually act as labels for polygons based on depth.

For example, if performing a series of analyses on a lake (maximum depth of 10 feet) at 3-foot, 5-foot, and 8-foot depth intervals, the labels on the cross-sectional coverage would be placed at these respective depths. Thus, labels would be positioned at depths of 3, 5, and 8 feet. Because these labels represent cross-sectional polygons, the 3-foot depth label may represent a polygon with boundaries at 0 feet and 4 feet. The 5-foot depth label may represent a polygon with boundaries at 4 feet and 6 feet, and the 8-foot depth label may represent a polygon with boundaries at 6 feet and the lake bottom (10 feet). Those who are familiar with lake ecology understand that no clear-cut boundaries distinguish where chemical values jump from one measurement to another without a gradual transition. All users of a lake quality GIS must be made aware of these types of inaccuracies (see Figure 3).

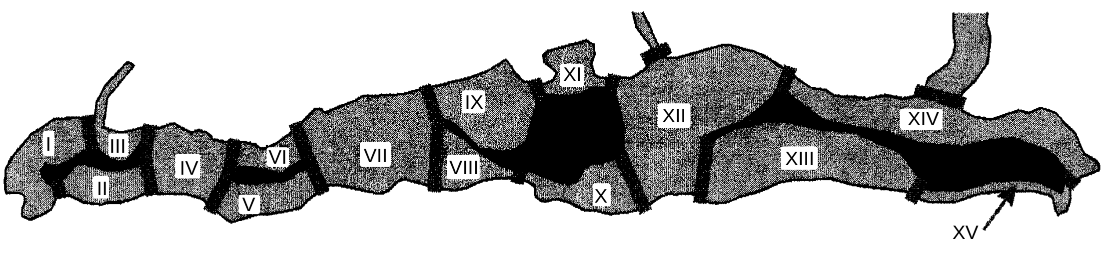


Figure 2. Hypothetical management zones for a section of Legend Lake that correspond to the management plans in Table 2; black areas indicate depths greater than 15 feet.

Table 2. **Hypothetical Management Plans for a Section of Legend Lake (see Figure 2)**

Management Zone	Management Plan (brief explanations)
I	High recreation area, high plant growth, frequent harvesting; frequently monitor water quality
II	Moderate recreation; manage for fish habitat
III	Moderate recreation; manage for fish habitat
IV	Open water, high recreation/possible fish habitat; consider subzoning
V	High-grade wildlife habitat; restrict human contact
VI	Moderate recreation; manage for fish habitat
VII	Open water, high recreation, increased shoreland development; frequently monitor water quality
VIII	Adjacent to high recreation area, possible fish habitat; manage for fish and aquatic habitat
IX	Adjacent to high recreation area, shoreland development; monitor water quality
X	Increased development; frequently monitor water quality
XI	[a]Prime wildlife/waterfowl habitat adjacent to high recreation area; restrict human presence (hot spot)
XII	Open water, high recreation area; frequently monitor water quality
XIII	Open water, high recreation area, possible fisheries and wildlife habitat; consider subzoning
XIV	Excessive aquatic plant growth (species listed) choking out preferred species; potential wildlife/waterfowl habitat, fisheries potential; continual harvesting; restrict human presence
XV	High-grade slope with little plant growth, potential for increased sedimentation; monitor shoreland development

[a]Critical area between good habitat and high recreation area; monitor extensively.

Figure 3 shows a cross section from one of the larger basins, Basin F, in the Legend Lake system (see Figure 1). The shades of gray represent ranges of temperature, with the lightest being the coldest and the darkest being the warmest. The thermocline can be located roughly in the middle. Each colored section represents a depth range where certain chemical attributes were collected. Using ARC/VIEW, the user can choose these areas with a pointer (mouse) and gain access to the database that contains all the sample results for this depth range.

Conclusion

These ideas are still preliminary as the Legend Lake study analysis concludes. Clear-cut discussions and recommendations will become available at a later date, although some observations can be made at this point. First, future research should focus on creating and developing a three-dimensional GIS, not to be confused with three-dimensional or cartographic models. Ideally, lake management plans should consider depth. This paper did not include depth because of the dimensional factor. Depth was not compatible with our two-dimensional GIS, and the presentation quality was not sufficient to relay our results in the form of management zones. We must develop a three-dimensional GIS to address the problems of the three-dimensional environment in which we live.

Lastly, maps are the best form for communicating this information to professionals and the public. Maps can also easily confuse people, however. Proper cartographic techniques are a necessity (7). Significant effort must be devoted to map creation to ensure a successful plan and successful relationships between lake managers and lake users. GIS and map-making are closely related. Both the planning stages and the database development phase of the lake quality GIS should emphasize this point. At an early stage of the process, management criteria should be determined, and all players or potential players must be included. A poorly planned project can lead to a failed GIS.

Creativity may offer new ideas in map development. For instance, animation (8) has some unique traits. Trend analysis using animation may produce the best visual results. Techniques such as these augment our methods of communication, and some are very revolutionary. Remember, however, that cartographic principles still must apply.

Acknowledgments

I would like to thank Dr. Keith Rice of the University of Wisconsin, Stevens Point, for his review efforts as well as Dr. Byron Shaw and Steven Weber for their technical support on lake management issues. Also, I would like

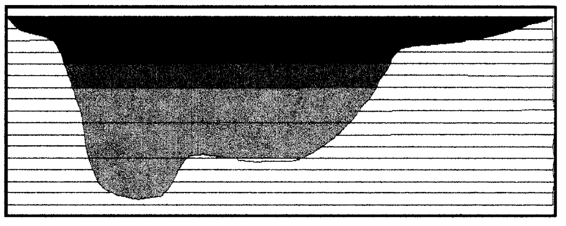


Figure 3. Cross section from one of the larger basins, Basin F, in the Legend Lake system.

to thank members of the University of Wisconsin, Stevens Point, Geography/Geology Department for their support and for the use of their equipment.

References

1. Fedra, K. 1993. Models, GIS, and expert systems: Integrated water resources models. Proceedings of the HydroGIS '93: Application of Geographic Information Systems in Hydrology and Water Resources Conference, Vienna. International Association of Hydrological Sciences Publication No. 211.
2. Schoolmaster, F.A., and P.G. Marr. 1992. Geographic information systems as a tool in water use data management. Water Resources Bull. 28(2):331-336.
3. Lam, D.C.L., and D.A. Swayne. 1993. An expert system approach of integrating hydrological database, models and GIS: Application of the RAISON system. Proceedings of the HydroGIS '93: Application of Geographic Information Systems in Hydrology and Water Resources Conference, Vienna. International Association of Hydrological Sciences Publication No. 211.
4. Rifai, H.S., C.J. Newell, and P.B. Bedient. 1993. GIS enhances water quality modeling. GIS World 6(8):52-55.
5. Samuels, W.B. 1993. LAKEMAP: A 2-D and 3-D mapping system for visualizing water quality data in lakes. Water Resources Bull. 29(6):917-922.
6. Fago, D. 1988. Retrieval and analysis system used in Wisconsin's statewide fish distribution survey, 2nd ed. Report 148. Wisconsin Department of Natural Resources Bureau of Research.
7. Ahner, A.L. 1993. Modern cartography plays essential role in GIS. GIS World 6(10):48-50.
8. Loucks, D.P. 1993. Interactive multimedia, GIS, and water resources simulation. Proceedings of the HydroGIS '93: Application of Geographic Information Systems in Hydrology and Water Resources Conference, Vienna. International Association of Hydrological Sciences Publication No. 211.

A Watershed-Oriented Database for Regional Cumulative Impact Assessment and Land Use Planning

Steven J. Stichter
Division of Coastal Management, State of North Carolina Department of Environment, Health, and Natural Resources, Raleigh, North Carolina

Introduction

In 1974, North Carolina passed the Coastal Area Management Act (CAMA) to guide growth and development in the state's coastal zone. Today, the Division of Coastal Management (DCM), under the direction of the governor-appointed Coastal Resources Council, implements CAMA. DCM's jurisdiction covers the 20 counties that border either the Atlantic Ocean or the Albemarle-Pamlico estuary.

This coastal region comprises a diverse set of human, animal, and plant communities. A broad array of coastal plain ecosystems occurs in this area, from the barrier dunes and maritime forests of the outer banks to cedar swamps and large pocosin complexes of interior areas. This area includes some of the state's fastest growing counties and some that are losing population. Urban centers such as Wilmington do exist, but the region remains primarily rural.

In recognition of the 20th anniversary of the passage of CAMA, the governor designated 1994 as the "Year of the Coast." Associated celebrations, panels, and studies highlighted the unique features of the North Carolina coast, successes of coastal management in the state, and unresolved problems and concerns. Problems remain despite protection efforts by various agencies. For instance:

- Fish landings have dropped dramatically of late.
- Shellfish Sanitation recently closed a set of shellfish beds located in outstanding resource waters.
- Shellfish statistics show that the quality of the state's most productive coastal waters continues to decline.

Because coastal North Carolina as a whole is growing more rapidly than any other section of the state, pressures on coastal resources can only continue to increase.

Declining water quality and associated sensitive habitats, resources, and animal populations have prompted several state agencies to develop new approaches to environmental protection that incorporate a broader, natural systems perspective. The North Carolina Division of Environmental Management is developing river basin plans to guide point and nonpoint water pollution control efforts. The DCM has begun work to assess and manage the cumulative and secondary impacts of development and other land-based activities by using coastal watersheds as the basis for analysis. The goal of this work is to expand the regulatory and planning programs in order to better address cumulative impacts. This paper describes the approach that DCM has developed for cumulative impacts management, with special emphasis on the use of a geographic information system (GIS). The project described here is scheduled for completion by the fall of 1996.

Cumulative Impacts Management

The concept of cumulative impacts management (1) is not new to North Carolina's coastal program. CAMA requires the consideration of cumulative impacts when evaluating development permits within defined areas of environmental concern. A permit must be denied if "the proposed development would contribute to cumulative effects that would be inconsistent with the guidelines. . . ." Cumulative effects are defined as "impacts attributable to the collective effects of a number of projects and include the effects of additional projects similar to the requested permit in areas available for development in the vicinity" (2). Despite this directive, few permitting actions have been denied because of cumulative effects; the existence of limited impact data and a dearth of viable analysis approaches have restricted application of this rule.

Since the passage of the National Environmental Policy Act (NEPA) in 1969, many attempts have been made to define and assess cumulative impacts. The Council on Environmental Quality developed the most familiar definition in its guidelines for NEPA implementation. It defines cumulative impact as:

> . . . the impact on the environment which results from the incremental impact of the action when added to other past, present, and reasonably foreseeable future actions, regardless of what agency or person undertakes such other actions. Cumulative impacts can result from individually minor but collectively significant actions taking place over a period of time (3).

Although this definition focuses the discussion of cumulative impacts, it provides little guidance on how to carry out such an analysis. Selecting both an appropriate time frame for the assessment (how far into the past and future to carry the analysis) and appropriate boundaries for the study (municipal or county boundaries, watersheds, ecoregions) are but two of the questions that require answers to successfully investigate cumulative impacts. Such decisions become even more complex when incorporating the limits imposed by available data and existing management structures.

Rigorous cumulative impact analysis is a difficult proposition. It requires identification of all sources of degradation that affect a given resource. The next step involves assigning relative significance to each of these sources along with any impacts that result from additive or synergistic interaction between sources. Assessment of the impacts of a pier on surrounding sea grasses, for instance, must include not only impacts related to the structure, such as shading and wave or current changes, but also such ambient impacts as natural wave and wind effects, upland runoff, and varying salinity.

All these investigations require the availability or collection of baseline environmental data at an appropriate spatial and temporal scale. Quantifying all the sources and causal pathways that affect a resource is extremely complicated in all but the simplest of systems. Assigning proof of significant impact is difficult unless the cause is clear and direct.

Because of the difficulties associated with assigning cause in cumulative impact analysis, especially in a regional review, DCM has chosen a different approach. It is focusing instead on locating areas at high risk to cumulative impacts. Impacts management studies and responses can then target the areas at greatest risk of degradation. Changing the scale of analysis from the site to the region requires applying some simplifying assumptions. The first assumes that any existing resource degradation results from the cumulative impact of all sources within the system boundaries. Locating such areas is relatively straightforward because most natural resource fields have developed measurements and indicators for locating degraded resources. The second assumption claims that a sufficiently intensive concentration of activities within a limited area will result in cumulative impacts on the affected system. Determining a threshold beyond which impacts cause degradation is much harder than locating already degraded resources because the level of such a threshold depends upon both the strength or spatial concentration of the impacts and the sensitivity of the resource.

Working from these simplifying assumptions, DCM's first step in assessing regional cumulative impacts is to identify areas within coastal North Carolina that exhibit symptoms of resource degradation, contain a concentration of activities that affect resources, or contain a concentration of sensitive resources. The use of categories of resources and impacts have helped to focus this search. These eight cumulative impact, high-risk area categories are:

- Impaired water quality
- High potential for water quality impairment
- Sensitive ground-water resources
- Impaired air quality (present or potential)
- Historical rapid growth
- Anticipated high growth
- High-value resources
- Productive and aesthetic resources

This set of categories is presently under public review. The next step is to develop indicators of the presence of impacts or resources appropriate to each of these categories. These indicators, when applied to a database of information about the study area, will help identify those locations at high risk as defined by the eight categories.

The regional cumulative impacts assessment approach that DCM developed is a hybrid of various assessment techniques. The overall approach is grounded in the theory and methods of site-specific cumulative impact assessment. Determination of high-risk categories and appropriate indicators and indexes is closely associated with both relative risk assessment procedures and geographic targeting. By focusing on known causes and effects of cumulative impacts on terrestrial and aquatic natural resources instead of attempting to quantify all impact pathways, available data and analysis techniques can help assess relative risk of cumulative impacts.

A Watershed Database for Cumulative Impact Assessment

High-risk categories and indicators of degradation or sensitivity are useless without information on the location of sensitive resources and impact sites. Consequently, a comprehensive database of information about coastal

North Carolina is central to cumulative impacts management in this area. The form of any database determines what types of questions to ask it; the selection of boundaries has been central to this study.

County boundaries constitute the most typical reporting unit in DCM operations. Counties determine the boundaries of DCM's jurisdiction, and the great majority of statistics used in planning and assessment are available primarily or solely by county. County size, heterogeneity, and the small number of counties available for comparison, however, have made county boundaries inappropriate for this project. Because the study focuses on impacts on natural resources, clearly the most appropriate boundaries would relate more directly to those resources.

Although using a single set of boundaries may not be appropriate for assessing impacts on all resources, management constraints limit the choice to one boundary type. Because the primary resources of concern are water based, watersheds were considered most appropriate. Surface waters receive the integrated effects of activities within a watershed; such boundaries fit intuitively with the concept of cumulative impact assessment. The number of water-related resources of concern also supported this choice. This analysis used small watersheds (5,000 to 50,000 acres) delineated in 1993 by the Soil Conservation Service for the entire state of North Carolina.

The Population, Development, and Resources Information System (PDRIS), which was designed for this project, is a PC-based, watershed-oriented database that contains the following information about the coastal area:

- Natural resources
- Population and housing
- Agricultural activities
- Economic activities
- Development activities

Table 1 includes a list of database fields. The presence and extent (or absence) of each of the features that this database represents will be available for each coastal watershed. The small watershed orientation of this study is only possible because of the availability of GIS; the volume and complexity of the watershed boundaries preclude any other assessment tool. In fact, 348 of these watersheds fall wholly or partially in the 20-county region. Figure 1 shows a map of these small watersheds. This map indicates county boundaries and shorelines in solid lines and the watershed boundaries in gray.

Data Needs and GIS Analysis

Over the past 5 years, North Carolina has actively collected a large amount of natural resource and base map information in GIS form. Research and funding associated with the Albemarle/Pamlico Estuarine Study (APES), a national estuary program study, spurred much of this data development in the coastal area. The state maintains a central repository for geographic data at the North Carolina Center for Geographic Information and Analysis (CGIA). Table 2 lists the general types of information available from the state database. The availability of data in GIS form is but one criterion for selecting a data set for use in this analysis. To be useful, the scale and accuracy of the data must be appropriate to the analysis.

Data Scale

The majority of data in the state's GIS database was collected at a scale of 1:100,000. Broader use and interest will probably urge the development of data layers at finer scales. A recently released layer of closed shellfish waters, for instance, was created at 1:24,000 scale. This prompted an update of the associated shoreline coverage to the same base scale. A handful of state departments and divisions, including Coastal Management, now use global positioning systems to collect even more precise locational information. This scale suits DCM's regional cumulative impacts scan, which is based on summary values for entire watersheds. More detailed intrawatershed planning and analysis would require finer scale data. A scale of 1:24,000 delineated the watershed boundaries in this project.

Mixing these 1:24,000 boundaries with the 1:100,000 data sets, however, can cause problems. For instance, a number of watersheds were designated for the large open water areas in Albemarle and Pamlico sounds. Although these should comprise exclusively water, overlay analysis of these watershed boundaries on the TIGER-derived census boundaries (1:100,000 scale) resulted in the assignment of small population counts to some of these watersheds. Individually locating and correcting such discrepancies is necessary.

Database Accuracy

Data layer accuracy problems are difficult to identify and assess. Because other agencies developed the majority of data used in this project, these source agencies must be relied upon for accuracy assessment of the source data. CGIA, steward of the state GIS database, adheres to National Map Accuracy Standards for all GIS data that it maintains. CGIA delivers metadata reports with any data; these reports include the source agency, collection date, and scale for the information used to derive the GIS layer. Descriptions of data lineage (collection and processing procedures), completeness, and positional

Table 1. Population, Development, and Resource Information System: Database Fields

Agriculture: Livestock and Poultry
Beef feedlots (< 300 head, > 300 head)
Dairy farms (< 70 head, > 70 head)
Hog farms (< 200 head, > 200 head)
Horse stables (< 200 head, > 200 head)
Poultry farms (< 15,000 birds, > 15,000 birds)

Agriculture: Farming
Land in farms (acres, % of HU)
Land with best mgmt. practices (acres, % of HU)
Land w/o best mgmt. practices (acres, % of HU)
Land in conservation tillage (acres, % of HU)
Land w/o conservation tillage (acres, % of HU)
Harvested cropland (acres, % of HU)
Hay crops (acres, % of HU)
Irrigated land (acres, % of HU)
Pasture land (acres, % of HU)
Row crops (acres, % of HU)

Primary
Estuarine waters (acres, % of HU)
Freshwater lakes
HU name
Receiving HU
Receiving water body
Primary water body
Secondary water body
Shoreline
Waterways w/vegetated buffers (miles, % of HU)
Population 1970
Population 1980
Population 1990
Population growth 1970 to 1980
Population growth 1980 to 1990
Counties
Total HU size
Land area (acres, % of HU)
Water area (acres, % of HU)
Stream length (miles)
Stream order (miles, % of stream length)

Development
Building permits—all residential
Building permits—amusement/recreation
Building permits—multifamily residential
Building permits—one-family residential
Building permits—hotels and motels
Building permits—retail
Building permits—industrial
Highway mileage:
- Total (miles)
- Primary (miles, % of total)
- Secondary (miles, % of total)
- Paved (miles, % of total)
- Unpaved (miles, % of total)

Rail lines (miles)
Increase of primary & secondary roads (miles, %)
Increase of paved vs. unpaved roads (miles, %)

Economic
Ag-related business (number, employees, income)
Farms (number, employees, income)
Fisheries business (number, employees, income)
Forestry/wood-using business (number, employees, income)
Lodging establishments (number, employees, income)
Manufacturing establishments (number, employees, income)
Marinas (number, employees, income)
Mining establishments (number, employees, income)
Recreation business (number, employees, income)
Restaurants (number, employees, income)
Retail establishments (number, employees, income)

Ground Water
Ground-water contamination incidents
Ground-water class (acres, % of HU)
Ground-water contamination area (acres, % of HU)
Ground-water capacity use areas (acres, % of HU)

Land and Estuarine Resources
Anadromous fish streams (miles, % of streams)
Coastal reserve waters (acres, % of HU)
Coastal reserve lands (acres, % of HU)
Federal ownership:
- National parks (acres, % of HU)
- National forests (acres, % of HU)
- Military reservations (acres, % of HU)
- USFWS refuges (acres, % of HU)
- Federal ownership—other (acres, % of HU)

State ownership:
- Game lands (acres, % of HU)
- State parks (acres, % of HU)
- State forests (acres, % of HU)
- State ownership—other (acres, % of HU)

Natural heritage inventory sites (count)
Primary nursery areas (acres, % of water area)
Private preservation (acres, % of HU)
Secondary nursery areas (acres, % of water area)
Threatened/endangered species habitat
Water supply watersheds (acres, % of HU)

Land Use
Total wetland area (acres, % of HU)
High-value wetlands (acres, % of HU)
Medium-value wetlands (acres, % of HU)
Low-value wetlands (acres, % of HU)
Predominant land cover

Population and Housing
Average seasonal population
Peak seasonal population
Units without indoor plumbing
Units with septic tanks
Units on central water systems
Units on central sewer
Units with wells

Permits
Air emission permits—PSD
Air emission permits—toxic
CAMA minor permits
CAMA general permits
CAMA major permits
CAMA exemptions
CWA Sect. 404/10 permits
Landfill permits—municipal
Landfill permits—industrial
Nondischarge permits
NPDES permits—industrial
NPDES permits—other
NPDES permits—POTW
Stormwater discharge permits
Sedimentation control plans
Septic tank permits

Shellfish
Shellfish waters (acres, % of HU)
Shellfish closures—permanent (acres, % of HU)
Shellfish closures—temporary (acres, % of HU)

Water Quality—Open Water
Class B waters (acres, % of water area)
Class C waters (acres, % of water area)
HQW waters (acres, % of water area)
NSW waters (acres, % of water area)
ORW waters (acres, % of water area)
Swamp waters (acres, % of water area)
SA waters (acres, % of water area)
SB waters (acres, % of water area)
SC waters (acres, % of water area)
WS-I waters (acres, % of water area)
WS-II waters (acres, % of water area)
WS-III waters (acres, % of water area)

Table 1. Population, Development, and Resource Information System: Database Fields (Continued)

Water Quality—Streams
Class B streams (miles, % of streams)
Class C streams (miles, % of streams)
HQW streams (miles, % of streams)
NSW streams (miles, % of streams)
ORW streams (miles, % of streams)
Swamp water streams (miles, % of streams)
SA streams (miles, % of streams)
SB streams (miles, % of streams)
SC streams (miles, % of streams)
WS-I streams (miles, % of streams)
WS-II streams (miles, % of streams)
WS-III streams (miles, % of streams)

Key
HU = hydrologic unit
PSD = point source discharges
POTW = publicly owned treatment work
NPDES = National Pollutant Discharge Elimination System
HQW = high-quality waters
NSW = nutrient-sensitive waters
ORW = outstanding resource waters
SA = saltwater classification A
SB = saltwater classification B
SC = saltwater classification C
WS1 = water supply classification 1
WS2: water supply classification 2
WS3: water supply classification 3

Water Quality—Use Support
Algal blooms (count, extent/severity)
Fish kills (count, extent/severity)
Streams fully supporting (miles, % of streams)
Streams support threatened (miles, % of streams)
Streams partially supporting (miles, % of streams)
Streams nonsupporting (miles, % of streams)
Waters fully supporting (acres, % of water area)
Waters support threatened (acres, % of water area)
Water partially supporting (acres, % of water area)
Waters nonsupporting (acres, % of water area)

accuracy are not available from these standard metadata reports, however.

DCM's cumulative impacts analysis also incorporates information not available from the state GIS database. Some of this information, such as business locations, is available from private data providers. Other information, especially agricultural statistics, does not presently exist in GIS form. Non-GIS formats include county statistics, voluntary compliance databases with self-reported coordinates, and other tabular databases. Typically, little quality control has been performed on any coordinate information. When the data originate from other state agencies, DCM is often the first user of the data outside of the source agency.

GIS Analysis Procedures

This study involves no sophisticated GIS analysis procedures. GIS helps to generate summary statistics by watershed for each of the database features. GIS drawing and query operations allow analysis of database accuracy. If the data are acceptable, the next step requires overlaying the watershed boundaries on the feature and assigning the appropriate watershed codes to all features that fall within the study area. Statistics can then be generated on the number of points, length of lines, acreage of polygons, or a total of any other numeric field in the feature coverage. Finally, the resulting summary file is converted to the format that PDRIS requires. A macro has been developed to complete this analysis. This macro generates a page-size reference map, performs the overlay, generates the watershed summary statistics, and converts the statistics to the final PC format.

Extra steps are necessary to analyze any information that does not already exist as a GIS coverage. Typically, these are tabular summaries associated with a specific boundary layer, such as county or U.S. Census statistics. These cases entail overlaying the watershed boundaries on the reporting unit boundaries; the data are distributed to the watershed in direct proportion to the percentage of the unit that falls into the watershed.

For instance, if a census tract falls 30 percent into watershed A and 70 percent into watershed B, 30 percent of the total tract population will be assigned to watershed A and the remainder to B. After performing all assignments, summary statistics are again generated by watershed. This procedure assumes that the distribution of the feature is even across each reporting unit. Rarely is this a valid assumption, but when the units are considerably smaller than the watersheds, as is the case with census tracts and blocks, this assumption introduces only limited errors. Watershed estimates based solely on county statistics, however, can be grossly inaccurate. When working with county information, therefore, using covariate information that ties more precisely to specific locations is necessary. Cropland location derived from the LANDSAT land cover layer, for instance, can be used to better distribute county-level agricultural statistics.

Database and Analysis Documentation

Data documentation is essential to this project. Given the large number of fields in the final database and the

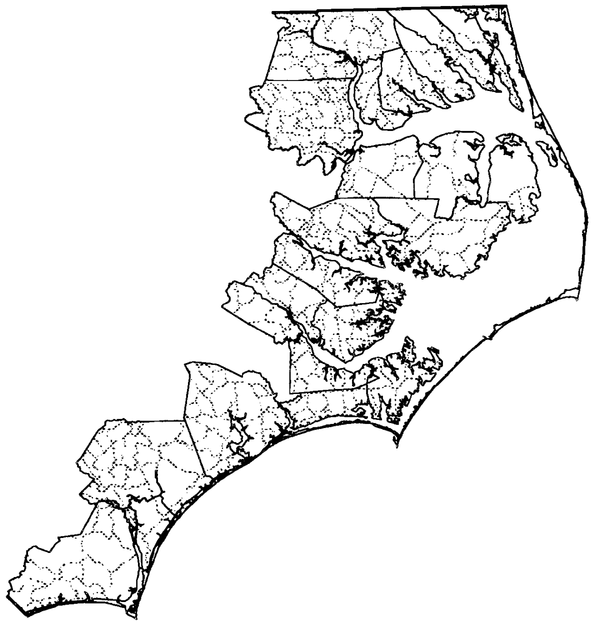


Figure 1. Watersheds in the North Carolina coastal area.

correspondingly large number of data types and sources, such documentation is key to understanding the quality of the individual database components as well as easing future database additions and updates. Because the results of this cumulative impact analysis exercise will be used to extend DCM's resource management efforts, documenting data sources and analyses will be critical if any decisions made based on this information are disputed.

A metadata database has been developed to document PDRIS data sources and analysis procedures. For each database entry, fields exist for a description and contact, collection methodology, and geographic extent of the source data. Data selection, overlay, and conversion procedures are also documented, along with any assumptions made in the analysis. In addition, recording data source, analysis procedure, and the date facilitates future database updates. Fields also record accuracy assessments for positional and attribute accuracy, logical consistency, and completeness.

The restrictions listed above regarding source data accuracy assessment, however, have limited their use. Once an entry is made to the PDRIS, all project team members receive metadata reports along with a reference map for a final review of completeness of the source data, data selection, and analysis logic. Figure 2 shows an example of a blank metadata worksheet.

Status of the Cumulative Impacts Assessment

DCM is presently gathering, verifying, and analyzing information for entry into the PDRIS. Although each source data layer was checked for accuracy before use, the logical consistency of each of the database entries relative to the other components also needs addressing. One example of such database inconsistencies is the watersheds that are covered entirely by water but also, according to the database, support a resident population. These inconsistencies could result from problems related to scale, differing category definitions, data inaccuracies, or errors in the GIS conversion or analysis at DCM. Database precision is essential for an accurate analysis and for general support of DCM's cumulative

Table 2. A Sample of North Carolina GIS Database Contents

Type	Examples	Coverage
Natural resource	Fishery nursery areas	Coastal North Carolina
	Natural heritage sites	Statewide
	Detailed soils	Varied
	Closed shellfish areas	Coastal North Carolina
	Water quality use classes	Statewide
	Detailed wetlands maps	Varied
Base data	Hydrography (24K, 100K)	Statewide
	Roads/transportation	Statewide
	County and city boundaries	Statewide
	LANDSAT-derived land cover	Coastal North Carolina
Ownership	Federal and state ownership	Statewide
Permits, waste sites	NPDES permit site	Statewide
	Landfills, hazardous waste, Superfund	Statewide
Cultural, population	TIGER boundaries, census information	Statewide
	Historic register sites, districts	Statewide

General Information:

Field					
Description					
Database			Units		
Definition	0 0				

Source Data Description:

Contact
Data
Scale
Sample Method
Geographic Extent

Database Entry:

Procedures
Assumptions

Accuracy Assessment:

	Overall	Positional	Attribute	Logical Consistency	Completeness
Rating					
Logic Test					
Error Comment					
Value Range					

Update Procedure:

Next Update		Source
Procedure		

Figure 2. Population/Development database: Data dictionary

impacts approach. Because the watershed database produced for this project will be widely available, errors and inconsistencies will undermine support for the rest of the project. Careful documentation of data sources, limitations, and analysis assumptions and procedures will provide useful support should problems or concerns arise.

Once database development is sufficiently complete (the database encompasses much dynamic data and could be constantly updated), indexes describing each of the cumulative impact, high-risk areas must be finalized. Applying these indexes to the database will allow identification of the watersheds at highest risk to cumulative impacts. Discussions held concurrently with index development will determine which management responses are appropriate to each high-risk category. Possibilities include strengthened land use planning requirements, new permit standards, or the designation of a new type of environmental critical area.

Although the data-intensive approach that DCM has chosen relies heavily on a GIS, the greatest challenges in this project do not lie in the GIS analysis. Applying this watershed-based analysis to existing political jurisdictions will be a more difficult undertaking. A convincing demonstration of the importance of including a natural systems perspective into a development permitting system, land use plan, or even economic development strategy, will ultimately contribute more to environmental protection in coastal North Carolina than any individual regulation that emanates from this project.

Summary

Twenty years after the passage of CAMA, DCM has developed a framework for a consistent approach to the problem of cumulative impacts of development. The approach and PDRIS database combine existing natural resource management techniques to locate areas of the coast at greatest risk of serious impairment from cumulative impacts. The availability of natural resource data at an acceptable scale (1:100,000) eases the development of the database essential to this analysis. The simultaneous development of a set of comprehensive small watershed boundaries for the state, along with the initial planning of this project, provided the final critical component to DCM's approach.

Perhaps more importantly, both DCM and individual local governments will have a large volume of information on natural units, which will provide an important, new perspective on the problems and prospects for local governmental action.

This project will not solve all problems related to cumulative impacts. The PDRIS will provide little support for site-specific or within-watershed cumulative impacts analysis; such an analysis at fine scales requires a much more precise database. By providing a broader-scale framework for this discussion, however, DCM's regional cumulative impacts study will hopefully further discussion, understanding, and management of cumulative and secondary impacts on natural systems.

References

1. Wuenscher, J. 1994. Managing cumulative impacts in the North Carolina coastal area. Report of the Strategic Plan for Improving Coastal Management in North Carolina. North Carolina Division of Coastal Management.
2. North Carolina General Statutes (NCGS) 113A-120.
3. 40 CFR §1508.7.

A GIS Demonstration for Greenbelt Land Use Analysis

Joanna J. Becker
Environmental Planning Services, Santa Rosa, California

Purpose

The goal of this project was to demonstrate what analyses could be undertaken with a GIS program without substantial GIS training or time input. The demonstration attempted to show how planning staff and decision-makers could easily and usefully employ GIS. It was not intended as a complete study of all possible variables. Only available data were used. Diverse techniques were presented while keeping the content as simple and relevant as possible. The project was designed as a demonstration using regional scale data and was combined with another parcel-specific demonstration that showed urban GIS applications.

The demonstration showed the following modeling techniques:

- Buffer zones
- Combination of variables (overlays)
- Weighting of values
- Absolute value variables
- Reclassification of final values

Site Location

The San Luis Obispo watershed comprises an area of approximately 84 square miles. The watershed drains into the Pacific Ocean at Avila, California. The major creek in the watershed, San Luis Obispo Creek, is a perennial creek, but many of its tributaries have only seasonal flow. Agriculture and grazing are the major land uses in the watershed, although a significant number of areas have been developed. Growth of these areas is moderate to limited but has a pronounced effect on the watershed. The watershed also supports a large amount of riparian and other natural vegetation. Figure 1 demonstrates the distribution of land cover/land use within the watershed.

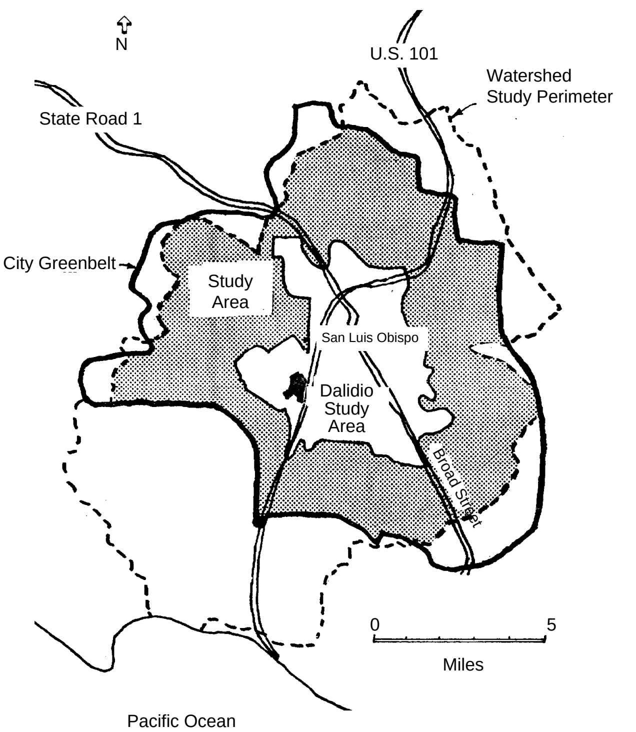


Figure 1. A generalized map of the San Luis Obispo area showing the location of the two ARC/INFO GIS study areas: (1) the parcel analysis in the Dalideo area and (2) the regional analysis in the city greenbelt stippled area surrounding the city. All boundaries are approximate and are for schematic purposes only (1, 2).

Background

The City Council approved the open space element of the San Luis Obispo General Plan in January 1994 and identified a greenbelt area that extended from the Urban Reserve Line approximately to the boundaries of the San Luis Obispo Creek watershed. The intent of this greenbelt area is to provide a buffer to the city and to preserve the agricultural and natural resources of the area.

The data for the watershed were already available through the work of the Landscape Architecture Department of California Polytechnic State University at San Luis Obispo for a study of San Luis Obispo Creek (1), and the city's greenbelt area lay approximately within those boundaries (see Figure 1). Variables were selected that could be extracted from the available data.

The data for the creek study were initially entered into workstation ARC/INFO in a polygon format. They were then transferred to MacGIS, a PC raster program, for simplicity of use. The final product was then transferred back to ARC/INFO as a grid format and viewed in a PC version of ARC/VIEW using DOS files.

In interpreting the overlay of values, the assumption was made that the occurrence of high values for the most variables would result in the most suitable land for that land use. This was presented as a range of values derived from the total values divided into three approximately equal groups of high, medium, and low. In addition to providing a composite analysis, however, any one of the data sets can be queried separately such that, for example, slopes greater than 20 percent could be identified or two layers such as storie index and distance from roads could be compared.

Criteria

Note that the ratings of high, medium, and low are based on available data, and the rating of low implies no suitable use. In addition, these ratings do not imply that categories rated low could not be used for a particular land use but rather that other land uses might be more appropriate. For example, open space use was rated low for flatter slopes but only because this category would likely be more suitable for agriculture.

The demonstration used an existing cell size of the data on the MacGIS program of 75 meters per side, which is assigned during the initial conversion process. Therefore, buffer zones are in aggregates of 75 meters. This size cell does not allow for minute analysis but reduces the size of the files, which may become extensive in raster format.

In presenting the final analysis, land contained within the urban reserve limit line has been excluded.

Procedure

Initially, eight variables from the available data were deemed suitable for this analysis:

- Slope
- Storie index (indicating soil fertility)
- Distance from major roads
- Distance from creeks
- Erosion hazard[1]
- Oak woodlands
- Land use compatibility
- Grasslands[2]

After selecting the six variables, the categories were recoded to conform to a rating of high, medium, or low.

Each land use was then evaluated separately for every variable except for combining the variables of slope and distance from creeks for rangeland analysis. In this case, composite values were assigned to the two variables, then recoded to produce a high, medium, or low rating.

After obtaining maps for each of the variables according to land use, the maps were compiled to indicate the density of overlays for each land use category. In assessing the suitability of land for the three land uses, the values of all the variables, except land use, were aggregated and a rating system developed. In addition, a double weight was assigned to the storie index in evaluating agriculture (because this is a primary index for considering prime agricultural land). If less than 75 meters, the distance from creeks also received added weight in consideration of open space preservation (because this is likely to ensure the least erosion and pollution to the waterways). The weighting then altered the scores as follows:

Land Use	Attributes	Number of Values
Agriculture	5	6
Rangeland	4	4
Open space	5	6

After the values had been assigned for each ranking, further recoding established three categories of high, medium, and low for each land use.

The land use buffer was added to this recoded aggregate map and resulted in an additional three values due to the interaction of the buffer with each category. These additional values were recoded according to each land use to produce a final map with three values.

The Urban Reserve Area was overlaid on the final map to exclude urban areas.

Assumptions

In determining what properties would be most suitable for each land use, the following assumptions were made.

[1] After reviewing the material, erosion hazard was eliminated because it was similar to the Soil Conservation Service (SCS) storie index data while the identification of native grasslands fell only within the area currently designated as open space land, so it was not included.

[2] See above note.

Open Space

This land is desirable to preserve as open space because of the existence of scenic or significant natural resources. It could also be land that is inappropriate for other uses due to the presence of such factors as steep slopes or poor soils. A distinction is made in the final map between land that is designated open space for recreational uses, such as parks, and land preserved for habitat or species protection. Open space adjacent to an urban area would be rated high if public accessibility was desirable but low if its purpose was resource protection or preservation. Separate maps based on two types of proposed uses present the contrast in analysis.

The analysis of the variables thus was rated as follows:

A. Land with steep slopes and therefore less suited for other purposes.

B. Land that has oak woodland vegetation resources.

C. Riparian land.

D. Low storie index indicating a lack of suitability for agriculture or rangeland.

E. At least approximately one-eighth of a mile from a major road to avoid negative impact on wildlife.

F. Either approximately one-quarter of a mile from urban areas if designated to protect resources or adjacent to urban areas if designated to serve as parks and recreation.

Agriculture

This land use includes all forms of agricultural activity; obviously, its suitability for specific crops and practices would vary. The determination of suitability would need to be made on a site-specific basis.

For the purpose of general agricultural suitability, the highest land suitability for agriculture was a rating of the variables as follows:

A. Land that does not have oak woodland.

B. Land that is not close to perennial creeks (to avoid fertilizer/pesticide runoff contamination).

C. The flattest slopes.

D. The highest storie index.

E. Proximity to a major road (considered an advantage for trucking and farm equipment access).

Rangeland

Some types of livestock can graze under most conditions, but for purposes of this analysis, land more suited for either open space or agricultural designation was rated above that of rangeland. The major limitation to suitability of land for rangeland was a combination of steep slopes and proximity to creeks. A rating of medium for the other variables was considered the most desirable for rangeland purposes.

Details of the Variables

Storie Index

In determining the most suitable uses according to soil fertility, the SCS storie index rating was used, with a modification of the categories to three to accommodate the ratings of high, medium, and low that were used throughout the analysis. Therefore, the first two SCS categories of excellent and good were combined into Category 1. Categories fair and poor were combined into Category 2, and categories very poor, nonagricultural, urban, and mines were combined to compose Category 3.

Subsequently, recoding was undertaken to prioritize these categories according to land use:

Agriculture	High for Category 1
Rangeland	High for Category 2
Open space	High for Category 3

Roads

The five principal arterials of the watershed were used in this analysis:

- Highway 1
- Highway 227
- Los Osos Valley Road
- U.S. 101
- Avila Valley/San Luis

A buffer on each side of the road was created, which was then recoded into the three categories of more than 75 meters, 75 meters to 150 meters, and greater than 150 meters. Each land use was then evaluated according to these criteria, with agriculture deemed the most suitable closest to the roads and open space the least suitable.

Slope

The slope categories in the San Luis Obispo watershed data set were divided into a number of classes, which were recoded into the most appropriate grouping of three classes for the analysis of the three land uses. The existing categories were not altered for the purpose of the demonstration, so they do not necessarily represent the most ideal slopes for the particular uses. A separate category of less than 10 percent slopes was provided for agriculture because most agricultural practices require flat land.

Slopes between 10 and 21.5 percent would be limited to such activities as orchards or vineyards.

Agriculture:	< 10 percent	High (H)
	10 to 21.5 percent	Medium (M)
	> 21.5 percent	Low (L)
Open space:	< 10 percent	L
	10 to 46 percent	M
	> 46 percent	H

No slope analysis was undertaken for rangeland because this category was combined with that for streams (see Streams section).

Urban Adjacency

Urban adjacency was treated as an overlay of the aggregate map of the other variables because it is an absolute value. That is, this variable has no ranking. Land is either within or outside the buffer. The suitability of each land use adjacent to urban areas was determined, then the aggregate map was adjusted according to a comparison of the recoded aggregate values with the designated ranking of land use suitability.

The first step was to recode the existing data on land uses (interpreted from 1989 aerial photography obtained from the United States Department of Agriculture [USDA], Agricultural Stabilization and Conservation Service, Atascadero, California) into urban/commercial areas and nonurban/noncommercial areas. A buffer zone of approximately one-quarter of a mile was then applied around the urban/commercial areas and an analysis undertaken of suitability for the three land uses to be located within this buffer.

In making the analysis, the following assumptions were made:

Land Use	Ranking	Reason
Agriculture	L	Conflict with dust, noise, pesticides, and urban use
Rangeland	M	Fire hazard of open grassland near buildings
Open space	H	Most suitable if used as parks/recreation
	L	Least suitable if designated for habitat/wildlife protection

Therefore, the analysis provided for two planning alternatives for open space, with the scenarios presented as separate maps overlaid on the aggregate map for the other rangeland variables.

In interpreting this map, the combined values were rated according to the above criteria, with any values lower than the desired ranking receiving a low value no matter what the aggregate value had been. This produced the following results:

- **Agriculture:** A rating of low for any land within the buffer zone.
- **Rangeland**: A rating of medium for any land within the buffer zone except that rated low for the aggregate map.
- **Open Space:** A rating equivalent to the rating of the aggregate map if the land was to be used for parks/recreation, or a rating of low for any land within the buffer zone if the land was to be used for habitat protection.

Streams

The original data for streams (from 1:24,000-scale USGS maps), which included both intermittent and perennial creeks, were used for rangeland analysis. A buffer of 150 meters was then applied to these streams, and these data were combined with the slope analysis. This combination was important in evaluating the erosion hazard and resulting stream pollution caused by nitrogen waste and hoof disturbance. The complete stream complex for the watershed area was therefore evaluated using the following matrix:

Stream Buffer	Slope		
	< 21.5 percent	21.5 to 46 percent	> 46 percent
< 75 meters	1/L	2/M	3/L
75 to 150 meters	4/M	5/H	6/M
> 300 meters	7/H	8/H	9/M

The stream data were then modified for agricultural and open space land use analysis to indicate only the perennial creeks as defined by the California Department of Fish and Game. Buffers were created for this as follows:

Stream Buffer	Open Space	Agriculture
< 150 meters	H	L
150 to 300 meters	M	M
> 300 meters	L	H

Oak Woodlands

The recoding of oak woodland data for agriculture was different than for the other two land uses because the presence of oak woodlands is not conducive to agriculture:

No oak woodlands	H
< 10 percent	M
> 10 percent	L

The suitability of oak woodlands for open space and rangeland was ranked as follows:

	Open Space	Rangeland
< 33 percent	L	M
33 to 75 percent	M	H
> 75 percent - high	H	L

References

1. Hallock, B.G., L.S. Bowker, W.D. Bremer, and D.N. Long. 1994. Nutrient objectives and best management practices for San Luis Obispo Creek. San Luis Obispo, CA: Coastal Resources Institute, California Polytechnic State University.

2. Community Development Department of San Luis Obispo. 1994. Open space element for the city of San Luis Obispo general plan. San Luis Obispo, CA.

GIS as a Tool for Predicting Urban Growth Patterns and Risks From Accidental Release of Industrial Toxins

Samuel V. Noe
University of Cincinnati, Cincinnati, Ohio

Introduction

The catastrophic Bhopal incident demonstrated to the world what could happen when industry and population are geographically incompatible. Many believe that the large urban population “should not have been there.” A recent publication, “New York Under a Cloud,” presents a frightening map of New York State that indicates potential areas of serious population exposure to accidental releases of chemicals stored by area industries and municipalities.

Conventional urban planning and administrative practices at the local level do not adequately provide for the minimization of these risks. Local jurisdictions on the fringes of metropolitan areas may be particularly ill-equipped to respond and plan effectively. Their elected officials, supported by minimal professional staffs and unaware of specific potential risks, may be more interested in soliciting new industrial development along with the tax base it brings. They therefore create industrial zones without restricting facilities that may generate hazardous substances and without recognizing the possibility that underground aquifers, which are current or potential sources of drinking water, may underlie these zones. Jurisdictions often permit facilities that could routinely or accidentally release toxic substances into the air without due regard for prevailing wind patterns or existing or projected urbanized areas that may be affected.

Although the available data and methodology have some gaps, much of the knowledge required to provide adequate protection from these risks exists. We know how to identify the hazardous substances that these sites may produce or store and how to calculate the types and levels of risks associated with them. We can accurately map the locations of streams, underground aquifers, and their catchment areas. Although with less precision, we also can indicate the areas more likely to receive the outfall of airborne and waterborne pollutants. At the other end of the equation, we can construct models for projecting patterns of urban growth. Those responsible for planning, however, have not made the connections between these techniques. Hazardous facilities sites are thus still permitted in areas that place existing urban residents and their drinking water supplies at risk, and new urban development grows in areas polluted by existing hazardous substance sites.

Clearly, this situation displays a need for the coordinated application of scientific risk assessment techniques and new approaches to regulating urban development. Equally critical, however, is the need to give greater attention to formulating appropriate public policy measures at the local and state levels for dealing with the complex disputes that surround these issues.

Project Background

The project this paper describes addressed these needs. It was undertaken by a team of faculty from the University of Cincinnati’s School of Planning, Department of Environmental Health, and the College of Law. The study team focused on the accidental release of hazardous materials both into the air and into the ground-water supply. The team’s purpose was to develop an integrated approach to scientific risk assessment, environmental analysis, urban planning, and policy analysis to address conflicts between:

- Expected patterns of suburban residential growth.
- The need to safeguard existing and new residential areas, and their water supplies, from toxic chemical pollution.
- The promotion of industrial development on the peripheries of urban regions, which often leads to the proliferation of hazardous substances sites.
- The need for effective regulation of these sites in complicated multijurisdictional environments.

This project examined these issues within a 100-square-mile area on the northern edge of metropolitan Cincinnati. The study area is not yet completely urbanized but lies in the path of urbanization. It contains a significant number of industrial or storage facilities that house supplies of hazardous materials. A major aquifer serving as a public water supply source passes under the area. Approximately 17 local jurisdictions fall within the study area: two counties, six townships, and nine municipalities. The area encompasses an intricate mix of agricultural, residential, and commercial land uses. In addition, several major industrial concentrations, as well as a number of jurisdictions, are aggressively soliciting new industrial employment. Because of its proximity to most major employment sites in southwest Ohio and to a variety of large retail complexes, the area is experiencing rapid residential development.

Projecting Areas of Future Development

The study team used PC ARC/INFO geographic information systems (GIS) to project the locations of future residential and industrial growth in the study area, to show the locations of areas at various degrees of risk from either airborne or waterborne industrial toxins, and to reveal the potential areas of population exposure resulting from the overlap of these areas.

Although this paper does not describe the models used to project residential and industrial growth in the study area, it does include the criteria used to make projections. The criteria we used to project residential growth were:

- Travel times to major employment concentrations in the region.
- Proximity to interstate highways, interchanges, and main trunk sewers.
- Avoidance of areas composed of steep slopes and flood plains.
- Land currently zoned for agriculture or housing.

The criteria for projecting industrial areas were:

- Relatively flat, not in a flood plain, and zoned industrial.
- Proximity to existing industrial development, main trunk sewers, and interstate highways.
- Relative aggressiveness of local jurisdictions in attracting industrial development.

Identifying Areas at Risk From Airborne Releases

The study team determined the model project areas at risk from airborne releases by using information available from the Ohio Environmental Protection Agency. This information, which included the location, identity, and quantity of hazardous materials recently stored in the study area, was available as a result of reporting requirements mandated by several federal statutes. The study team assumed that a similar ratio reflecting sites containing hazardous materials to the area of industrially developed land would continue into the future. Based on this assumption, the study team could randomly project new potential release sources.

The algorithm and associated software used for calculating the plume size of aerial dispersion of hazardous chemicals was Aerial Locations of Hazardous Atmospheres (ALOHA). The National Oceanic and Atmospheric Administration (NOAA) developed this system, which is in wide use by government and industry for the preparation of emergency contingency plans. This software is available for use on any Macintosh computer, or any IBM-compatible with an Intel 80286 (or better) CPU. This software employs three classes of variables in calculating the plume dispersion for a specific chemical:

- Chemical variables
- Meteorological options
- Source strength options

Chemical variables include both physical properties of the chemical and parameters that define the human health effects of the chemical. In the latter case, the two variables are the threshold limit value (TLV) and the immediately dangerous to life and health (IDLH) value. The TLV is a measure of chronic toxicity of the chemical in humans. It represents the maximum concentration of the chemical in air to which a human can safely be exposed for 8 hours per day on a daily basis. The IDLH is a measure of acute toxicity of the chemical in humans. The IDLH represents the maximum concentration to which a human can be exposed for a short time and not experience death or some other severe endpoint.

Meteorological options describe the ambient atmospheric conditions into which the chemical disperses. ALOHA has the capability of downloading real-time data from NOAA satellites. This case, however, employed average meteorological conditions for the study area over the course of a year. The variables this study used were atmospheric inversion height (or no inversion), wind velocity, air temperature, ground roughness (rural or urban), and stability class (a combined variable describing cloud cover and incoming solar radiation).

The source options quantify the amount of chemical being released and describe how the chemical is released (instantaneous or continuous).

The ALOHA model provides a procedure for showing the IDLH and TLV risk zones from a single accidental release of a single chemical, given specified conditions of atmospheric stability, wind direction, and air temperature. Obviously, climatic conditions change daily, so areas surrounding a single industrial site experience different degrees of risk depending on the variability of these conditions. Moreover, a single site that can potentially release more than one chemical poses a higher risk to surrounding areas. Finally, when release plumes from two or more sites that are located relatively close to each other overlap, risks also increase. To account for these factors of climatic variation and overlapping chemical release plumes, the study team constructed the model described below.

In discussing these procedures, bear in mind that the risk factors are relative. Sufficient data are not available to estimate the absolute probability of an accidental release of industrial toxins into the air. Consequently, no absolute risk levels can be estimated.

1. We acquired NOAA climatic statistics for a full year for the weather station nearest the study. NOAA tabulates eight wind directions (N, NE, E, SE, S, SW, W, and NW). We sorted the average daily wind speeds and average daily temperatures according to the daily prevailing wind direction. Thus, for each of the eight wind directions, it was possible to determine the number of days in the year that the prevailing wind comes from that direction, as well as the average daily wind speed and temperature.

2. We then input the temperature and wind speed data, derived as explained above, into the ALOHA model to prepare plots of the IDLH and TLV zones, or plumes, for each of the eight wind directions. Individually, the plumes emanated downwind from the source of the release. In our study, IDLH plumes varied from 0.17 miles to greater than 10 miles. TLV plumes varied from 0.52 miles to greater than 10 miles.

   When the plumes for the eight different wind directions were combined, the results were translated into a pattern of wedges representing various plume lengths in each of the eight different wind directions. These risk levels vary according to the number of days per year the wind blows in each direction from the source of the release. Plumes blowing in different directions vary in length according to average temperature and wind velocities for the days the wind blew in each direction. The numbers in the wedges represent risk factors assigned as indicated in Table 1.

Table 1. Assigned Risk Factors

Frequency of Wind	Risk Factor
0-25 days/year	1
26-50 days/year	2
51-75 days/year	3
76-100 days/year	4

3. When a single industrial location employed more than one chemical, the IDLH and TLV risk patterns for each were overlaid on one other. Where the overlap occurred, we added the risk factors together.

4. Finally, when the risk patterns from two or more sites overlapped, we added together the risk factors assigned to overlapping areas. The overlap capabilities of GIS allowed us to easily draw and combine the risk patterns, superimposed on a map of the industrial sites under consideration.

When we compared the eight existing combined IDLH and TLV risk patterns with the risk patterns that appear when six projected new sources of releases and projected areas of residential growth are added, the changes are rather dramatic. Of course, we must note that the projected sources of potential releases are hypothetical and their locations selected at random. In reality, precise locations of areas at greater risk cannot be predicted with any degree of certainty. We can reasonably deduce from the maps, however, that any substantial increase in industries storing or using toxic chemicals that might be accidentally released into the air can compound risk levels much more than might be expected. The maps we produced indicated five risk levels.

The study results also showed that an industry capable of generating accidental releases of airborne toxins will very likely place at risk residents not only of the same community but those in neighboring jurisdictions as well. This is particularly true on the fringes of metropolitan areas where highly fragmented political boundaries exist. This fact complicates tremendously the ability of each jurisdiction to protect its citizens and suggests a need for a more comprehensive approach to regulation than conventional land use zoning measures that each locality administers.

Identifying Areas at Risk From Accidental Releases Into Ground-Water Supplies

As the introduction of this paper indicates, local government usually manages conventional land use planning, while air quality is largely a state or federal responsibility. Thus, decisions regarding the regulation of industrial location may not account for the types of industrial operations proposed, the possible use of hazardous materials, and possible risks to local residents from accidental release of toxins into the air. The same problem applies to the location of industries that have the potential for accidental release of hazardous materials into ground-water supplies. In our study, we outlined a technique for predicting where local residents may be placed at risk by drinking water from sources vulnerable to contamination by industrial toxins.

As stated earlier, the study area includes a major aquifer. We noted industrial and related facilities that use and store hazardous materials, and that are located over or immediately adjacent to the area aquifers. We distinguished between sites where a previous spill had been reported, those where a waste well is located, and all other locations where a hazardous material is used or stored. The results clearly indicated a significant potential for contamination.

To predict the number and locations of additional industrial toxin sources that might appear over a 10-year period following 1990 (the base year of the study), we employed a procedure similar to that used in the air pollution section of the study. We assumed that during this period, the number of such sites would increase by 35 percent. The 35-percent increase represented an arbitrarily selected figure approximately midway between estimates of 25- and 50-percent industry growth in the study area. We used this increment to project additional waste well sites and sites that would experience a spill sometime during the decade, as well as the total number of new sites.

We projected the locations of the additional sites by overlaying a 5,000-foot by 5,000-foot grid on the study area and assigning the new sites to grid cells using a random number generator. We considered only cells lying completely or largely over the aquifers and also falling within the area of projected industrial land use. Each cell was assigned a relative contamination risk factor based on the number of projected sites it contained, with a multiplier of 3 applied to sites with a waste well and a multiplier of 4 applied to sites assumed to have had spills. We also assigned existing sites to grid cells and scored them in the same manner.

To obtain more information, we used a simplified version of DRASTIC maps prepared by the Ohio Department of Natural Resources. DRASTIC is an acronym for **D**epth to water, net **R**echarge, **A**quifer media, **S**oil media, **T**opography, **I**mpact of the vadose zone, and hydraulic **C**onductivity of the aquifer. These factors contribute to an index of the relative vulnerability of the aquifer to pollution. Different shades on the map represented the relative vulnerability of sections of the aquifers to ground-water pollution. We assigned two points to aquifers with a DRASTIC pollution potential index less than 180 and four points to aquifers with a higher index. This allowed us to use the GIS to combine the DRASTIC vulnerability map with the maps that showed risks of pollution from the hazardous materials sites. The resultant maps showed the existing and projected potential risk of pollution in different areas of the aquifers.

The next step in the study related the above information to public water companies that extract water from different parts of the aquifers. Thus, we were able to associate risk levels with well locations as well as with the areas served by water companies whose supplies may be at risk. In this case, the projected risks of polluted water supplies were identical to existing risk levels because no public water wells happened to be located in areas where additional potential sources of pollution were projected.

Of course, wide areas in the study area had no public water supply. Residents in these areas may be at risk depending on where they dig private wells. The maps we generated showing areas at risk nonetheless provide useful guides to potentially hazardous areas.

As in the construction of all such models, we needed to make a number of assumptions, simplifications, and value judgments. In this case, these included the projected number of new toxic sites and point scores assigned to hazardous materials sites and the various areas in the DRASTIC maps. Also, for the sake of simplicity, we projected no new well sites in preparing one of the maps. This additional element should probably be included, assuming local water companies could provide projected well locations. Use of a GIS, however, makes it possible to explore the implications of adjustment of any of these factors.

We must note two more significant omissions from the model that the GIS cannot factor in. One was our inability to identify from the data specific chemicals that each sight might release and the relative effects of each. We would have required more complicated techniques for dealing with these variables and for projecting the travel and dilution of plumes of contaminants in an aquifer. In the interest of providing a simple, if relatively crude, model capable of replication by a local planning agency with a simple GIS, we elected not to propose use of more sophisticated techniques.

Another obvious omission was the consideration of water pumping and treatment measures that might mitigate risks of water contaminated by accidental release of industrial toxins. Perhaps, with knowledge of the specific contaminants found in the water at any given time, such mitigation might be possible. We elected not to consider this factor for reasons of simplification but also because this study aimed to provide planning agencies the means to identify potential risks to local residents and to prevent or minimize them through better land use planning and regulatory measures.

Some Final Notes

As stated earlier, the purpose of the study was to propose a technique that planning agencies could use to identify:

- The locations of existing and projected patterns of residential and industrial development in a multijurisdictional suburban area.

- The locations of existing and projected industrial, storage, and disposal sites of hazardous materials.
- Residential areas that might be placed at risk by the accidental release of these hazardous materials into the air or into ground-water supplies.
- The relative levels of risks resulting from potential exposure to more than one hazardous material at a single site, or from multiple sites in the vicinity.

The technique we used in the study permitted projection of *relative* risk levels. Projecting absolute risk levels is impossible without data on the actual incidence of accidental releases of toxins over time in this or in similar areas. A related question is whether it is possible to meaningfully indicate the combined risk from airborne releases of industrial toxins and drinking water contamination to a particular residential area. The issue of weighting relative risk factors is central here. Could we, for example, weight the risk levels of exposure to airborne releases three—or possibly four—times higher than water contamination risks? Or can we say that the risks may be the same, but the danger from airborne releases is three or four times greater?

Obviously, these would be futile exercises, especially because the point scores in the separate mapping studies were arbitrarily assigned. The public should know, however, which present or projected residential areas carry some level of risk from both types of exposure. Thus, after calculating point scores to derive relative risk levels from waterborne pollutants, we multiplied the scores by 4 to bring their maximum ranges into the same order of magnitude. Otherwise, the effect of waterborne pollutants would not be apparent. Consequently, we created two maps to show the combined existing and projected risks; therefore, we highlighted the combined risks that residents face from both airborne and waterborne hazardous materials.

The maps of airborne releases used in these combinations showed TLV. Continuing exposure within the TLV areas over an extended period can also have adverse health effects. This study focused only on accidental releases, however, and a single release is unlikely to sustain continuous exposure. Of course, residential areas at risk from several sites might approach conditions of sustained exposure. This situation is more analogous to prolonged exposure to contaminated drinking water supplies. We did not combine the maps of IDLH airborne releases with the maps of areas at risk from groundwater contamination because they are not analogous conditions. Nonetheless, the IDLH risk maps in themselves reveal the conditions that local planning officials should most seriously consider.

Replicating the procedures outlined in this study should be technically and financially feasible for local planning agencies. Armed with the results of such an investigation, their next step should be to establish the planning and regulatory measures that would minimize both existing and projected levels of risk to area residents. The attorney on our team has outlined a range of possible measures, but detailing them would be the subject of another paper.

Integration of GIS and Hydrologic Models for Nutrient Management Planning

Clyde W. Fraisse, Kenneth L. Campbell, James W. Jones,
William G. Boggess, and Babak Negahban
Agricultural Engineering Department, University of Florida, Gainesville, Florida

Introduction

Recent evidence that agriculture in general, and animal waste in particular, may be an important factor in surface and ground-water quality degradation has increased the interest in nutrient management research. The presence of nitrogen and phosphorus in surface water bodies and ground water is a significant water quality problem in many parts of the world. Some forms of nitrogen and phosphorus, such as nitrate N and soluble P, are readily available to plants. If these forms are released into surface waters, eutrophic conditions that severely impair water quality may result. Advanced eutrophication (pH variations, oxygen fluctuations or lack of it in lower zones, organic substance accumulation) can cause physical and chemical changes that may interfere with recreational use and aesthetic appreciation of water. In addition, possible taste and odor problems caused by algae can make water less suitable or desirable for water supply and human consumption (1).

Increases in nutrient loadings to water resources have recently been observed in the southeastern United States, where well-drained sandy soils with low nutrient retention capacity and high water table conditions are found in most coastal areas. Those increases were associated statistically with nutrient sources such as agricultural fertilizers and dense animal populations (2, 3). Repetitive occurrences of extensive blooms of blue-green algae that threatened the overall health of Lake Okeechobee, located in southern Florida, were attributed to an increase in nutrient loadings to the lake. The South Florida Water Management District (SFWMD) reported an increase of phosphorus concentrations in the lake water from an annual average of 0.049 milligrams per liter in 1973 and 1974 to a peak of 0.122 milligrams per liter in 1988 (4).

Most water quality problems concerning phosphorus result from transport with sediment in surface runoff into receiving waters. Continuous high loadings from animal waste on sandy soils with low retention capacity, however, may contribute significant quantities of labile phosphorus to subsurface drainage. Ground-water aquifers may also become polluted due to recharge of high loadings of nitrogen. Drinking water with nitrate N concentrations higher than 10 milligrams per liter may lead to methemoglobinemia in infants. Ground-water monitoring of the Middle Suwannee River area in Florida has shown high concentrations of nitrate nitrogen near intensive agricultural operations. The U.S. Geological Survey has intensively monitored dairy and poultry farms and has found high nitrate levels below these operations compared with nearby control wells (5).

Animal waste management has always been a part of farming, but historically has been relatively easy due to the buffering capacity of the land. In fact, land application of animal waste at acceptable rates can provide crops with an adequate level of nutrients, help reduce soil erosion, and improve water holding capacity. As the animal industry attempts to meet the food requirements of a growing population, however, it applies new technologies that reduce the number of producers, but create larger, more concentrated operations. That, in addition to the decreasing amount of land available for waste application, has increased the potential for water quality degradation.

Successful planning of an animal waste management system requires the ability to simulate the impact of waste production, storage, treatment, and use on water resources. It must address the overall nutrient management for the operation, including other nutrient sources such as supplemental fertilizer applications. Livestock operations are highly variable in their physical facilities, management systems, and the soil, drainage, and climatic conditions that affect the risk of water pollution from animal wastes (6). Linkage between geographic information systems (GIS) and hydrologic models offers an excellent way to represent spatial features of the fields being simulated and to improve results. In addition, a GIS containing a relational database is an excellent way to store, retrieve, and format the spatial and tabular data required to run a simulation model.

This paper examines some of the issues related to the integration of hydrologic/water quality models and GIS programs. In addition, the paper discusses the approaches used in the Lake Okeechobee Agricultural Decision Support System (LOADSS), which was recently developed to evaluate the effectiveness of different phosphorus control practices (PCPs) in the Lake Okeechobee basin. The paper also details a dairy model, designed to simulate and evaluate the impacts of alternative waste management policies for dairy operations, that is currently under development.

Hydrologic Models and GIS

By using models, we can better understand or explain natural phenomena and, under some conditions, make predictions in a deterministic or probabilistic sense (7).

A hydrologic model is a mathematical representation of the transport of water and its constituents on some part of the land surface or subsurface environment. Hydrologic models can be used as planning tools for determining management practices that minimize nutrient loadings from an agricultural activity to water resources. The results obtained depend on an accurate representation of the environment through which water flows and of the spatial distribution of rainfall characteristics. These models have successfully dealt with time, but they are often spatially aggregated or lumped-parameter models.

Recently, hydrologists have turned their attention to GIS for assistance in studying the movement of water and its constituents in the hydrologic cycle. GIS programs are computer-based tools to capture, manipulate, process, and display spatial or georeferenced data. They contain geometry data (coordinates and topological information) and attribute data (i.e., information describing the properties of geometrical objects such as points, lines, and areas) (8). A GIS can represent the spatial variation of a given field property by using a cell grid structure in which the area is partitioned into regular grid cells (raster GIS) or by using a set of points, lines, and polygons (vector GIS).

A close connection obviously exists between GIS and hydrologic models, and integrating them produces tremendous benefits. Parameter determination is currently one of the most active hydrology-related areas in GIS. Parameters such as land surface slope, channel length, land use, and soil properties of a watershed are being extracted from both raster and vector GIS programs, with a focus on raster-based systems. The spatial nature of GIS also provides an ideal structure for modeling. A GIS can be a substantial time saver that allows different modeling approaches to be tried, sparing manual encoding of parameters. Further, it can provide a tool for examining the spatial information from various user-defined perspectives (9). It enables the user to selectively analyze the data pertinent to the situation and try alternative approaches to analysis. GIS has been particularly successful in addressing environmental problems.

Approaches for Integrating GIS and Models

A significant amount of work has been done to integrate raster and vector GIS with hydrologic/water quality models. Several strategies and approaches for the integration have been tried. Initial work tended to use simpler models such as DRASTIC (10) and the Agricultural Pollution Potential Index (11). In these cases, the models were implemented within the GIS themselves. These studies attempted to develop GIS-based screening methods to rank nonpoint pollution potential. The use of more complex models requires that the GIS be used to retrieve, and possibly format, the model data. The model itself is implemented separately and communicates with GIS via data files. Goodchild (12) refers to this mode as “loose coupling,” implying that the GIS and modeling software are coupled sufficiently to allow the transfer of data and perhaps also of results, in the reverse direction. Fedra (8) refers to this level of integration as “shallow coupling” (see Figure 1). Only the file formats and the corresponding input and output routines, usually of the model, must be adapted. Liao and Tim (13) describe an application of this type, in which an interface was developed to generate topographic data automatically and simplify the data input process for the Agricultural Nonpoint Source (AGNPS) Pollution Model (14), a water quality model.

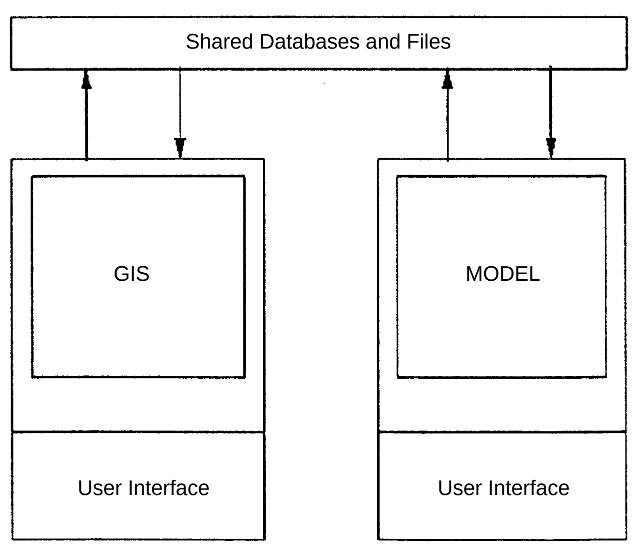


Figure 1. Loose or shallow coupling through common files (8).

Higher forms of connection use a common interface and transparent file or information sharing and transfer between the respective components (see Figure 2). The dairy model, currently under development, is an application of this kind. It will link the Ground-Water Loading Effects of Agricultural Management Systems (GLEAMS) (15) model and GIS to evaluate potential leaching and runoff of both nitrogen and phosphorus.

LOADSS is an extension of this type of application because it includes an optimization module that enables the system to select the best PCPs at the regional scale, based on goals and constraints defined by the user.

Both applications use ARC/INFO's arc macro language (AML), a high-level application language built into the GIS. A subset of functions of a full-featured GIS, such as creation of maps (including model output) and tabular reports, as well as model-related analysis, are embedded in the applications, giving the system great flexibility and performance. Fedra (8) describes a deeper level of integration that would merge the two previous approaches, such that the model becomes one of the analytical functions of a GIS, or the GIS becomes yet another option to generate and manipulate parameters, input and state variables, and model output, and to provide additional display options. In this case, software components would share memory rather than files.

The choice between integrating a water quality model with a raster or vector GIS depends on the importance of spatial interactions in the process being studied and the nature of the model itself. Some water quality models, such as GLEAMS, are field-scale models that provide edge-of-the-field values for surface runoff and erosion as well as deep percolation of water and its constituents. In this case, spatial interactions between adjacent fields are ignored and a vector GIS can be used to describe the system. Moreover, important factors in the simulation process, such as land use and management practices, are normally field attributes and thus, are better represented in a vector structure.

Other factors playing an important role in the hydrologic process, such as field slope, aspect, and specific catchment area, are hard to estimate in vector systems, however. A raster-based GIS is better suited for handling watershed models in which the routing process is important and spatial interactions are considered. For those, several algorithms for estimating important terrain attributes are often incorporated in commercially available raster-based GIS programs.

LOADSS

LOADSS was developed to help address problems created by phosphorus runoff into Lake Okeechobee. It was designed to allow regional planners to alter land uses and management practices in the Lake Okeechobee basin, then view the environmental and economic effects resulting from the changes. The Lake Okeechobee basin coverage incorporates information about land uses, soil associations, weather regions, management practices, hydrologic features, and political boundaries for approximately 1.5 million acres of land and consists of close to 8,000 polygons.

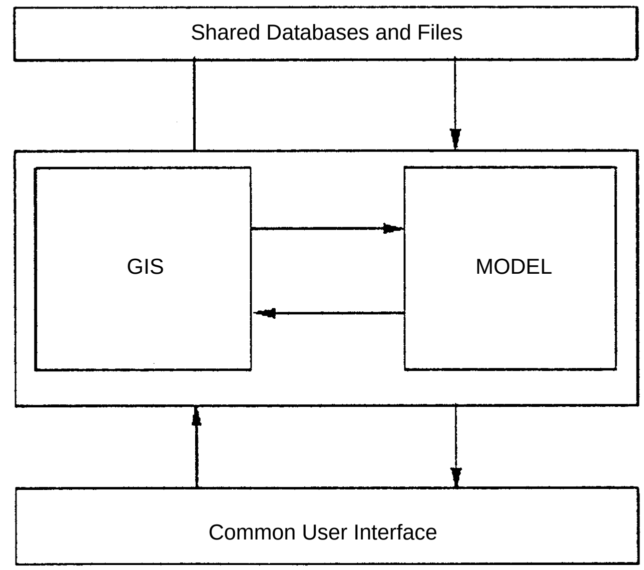


Figure 2. Deep coupling in a common framework (18).

The SFWMD, responsible for managing Lake Okeechobee, has initiated numerous projects to develop effective control practices for reducing the level of phosphorus in agricultural runoff as part of the Lake Okeechobee Surface Water Improvement and Management (SWIM) Plan. These projects, numbering more than 30, were designed to develop information on the control and management of phosphorus within the lake basin and to determine the costs and effectiveness of selected management options. Three types of control options are being studied:

- Nonpoint source controls, such as pasture management.
- Point source controls, such as sewage treatment.
- Basin-scale controls, such as aquifer storage and retrieval.

After completing most of these research efforts, the need arose for a comprehensive management tool that could integrate the results for all three classes of PCPs. In response to these needs, design and implementation of a decision support system was initiated with the following objectives (16):

- Organize spatial and nonspatial knowledge about soils, weather, land use, hydrography of the lake basin, and PCPs under a GIS environment.

- Develop and implement algorithms for modeling nonpoint source, point source, and basin-scale PCPs.
- Develop and implement mechanisms for evaluating the performance of the entire Lake Okeechobee basin under different combinations of PCPs applied to the basin.
- Design and develop a user interface that would facilitate use of the system by noncomputer experts.

The goal in developing LOADSS was to create an information system that would integrate available information to help regional planners make decisions. LOADSS can generate reports and maps concerning regional land attributes, call external hydrologic simulation models, and display actual water quality and quantity sampling station data. LOADSS is a collection of different components:

- The regional scale GIS-based model used to develop and manipulate regional plans for reducing phosphorus loading to Lake Okeechobee.
- The Interactive Dairy Model (IDM) used to develop field-level management plans for dairies and run the Field Hydrologic and Nutrient Transport Model (FHANTM) simulation model for nutrient transport modeling.
- An optimization module that enables the system to select the best PCPs at the regional scale (currently under development).

Although these components can run independently, they are fully integrated in the LOADSS package and can exchange information where necessary. A design schematic of LOADSS is given in Figure 3.

Regional-Scale GIS-Based Model

LOADSS serves both as a decision support system for regional planning and as a graphic user interface for controlling the different components. One consideration in the design of LOADSS was the size of the database that was being manipulated. Because the land use database consisted of nearly 8,000 polygons, running the simulation models interactively would not be a feasible option. Thus, the CREAMS-WT (17) runoff model was prerun for different levels of inputs and management for each land use, soil association, and weather region (18).

Depending on the land use and its relative importance as a contributor of phosphorus to the lake, anywhere from one (background levels of inputs to land uses like barren land) to 25 (dairies, beef pastures) levels of inputs were selected. Each set of inputs to a particular land use was given a separate PCP identification code. A CREAMS-WT simulation was performed for each PCP, on each soil association and weather region. This resulted in approximately 2,600 simulation runs. Annual average results were computed for use in LOADSS. CREAMS-WT provides an average annual estimate of phosphorus runoff from each polygon. Phosphorus assimilation along flow paths to Lake Okeechobee are estimated as an exponential decay function of distance traveled through canals and wetlands (4).

The imports, exports, and economics of each PCP are based on a per production unit basis. Depending on the type of polygon, the production unit can be acres (e.g., pastures, forests), number of cows (dairies), or millions of gallons of effluent (waste treatment plants and sugar mills). Developing a regional plan in LOADSS involves assigning a PCP identification code to each one of the polygons in the Lake Okeechobee basin. Accessing the results of a regional plan involves multiplying the production unit of each polygon by its appropriate database import, export, or economic attribute and summing the resulting values over all polygons in the Lake Okeechobee basin. LOADSS runs in the ARC/INFO Version 6.1.1 GIS software on SUN SPARC stations.

Interactive Dairy Model

Although the LOADSS level of detail is adequate for regional planning, a more detailed model was necessary to analyze individual dairies in the Lake Okeechobee basin, as dairies were one of the large, concentrated sources of phosphorus runoff into the lake. Thus, the IDM was developed and incorporated into LOADSS. IDM utilizes FHANTM to simulate phosphorus movement in dairy fields. FHANTM is a modification of DRAINMOD (19) with added functions to handle overland flow routing, dynamic seepage boundary, and soluble phosphorus algorithms for P input, mass balance, and transport (20).

Unlike in LOADSS, FHANTM is run interactively, as IDM requires. Furthermore, in LOADSS, the user can only select from a number of predefined PCPs, while in IDM, the user has access to more than 100 input and management variables, all of which can take a range of values. This allows for the development and evaluation of detailed dairy management plans that otherwise would be impossible at a regional scale. While LOADSS only provides average annual results, IDM displays daily time series simulation results. IDM uses the same assimilation algorithm and can produce the same phosphorus budget maps and reports as LOADSS.

Optimization Module

A variety of factors must be considered in planning nutrient management programs. Production and environmental goals need to be balanced, and these goals are often incompatible. Performing this exercise on a regional scale, comprising many fields for which a variety of land uses and management options can be assigned, is a tremendously time-consuming, if not impossible,

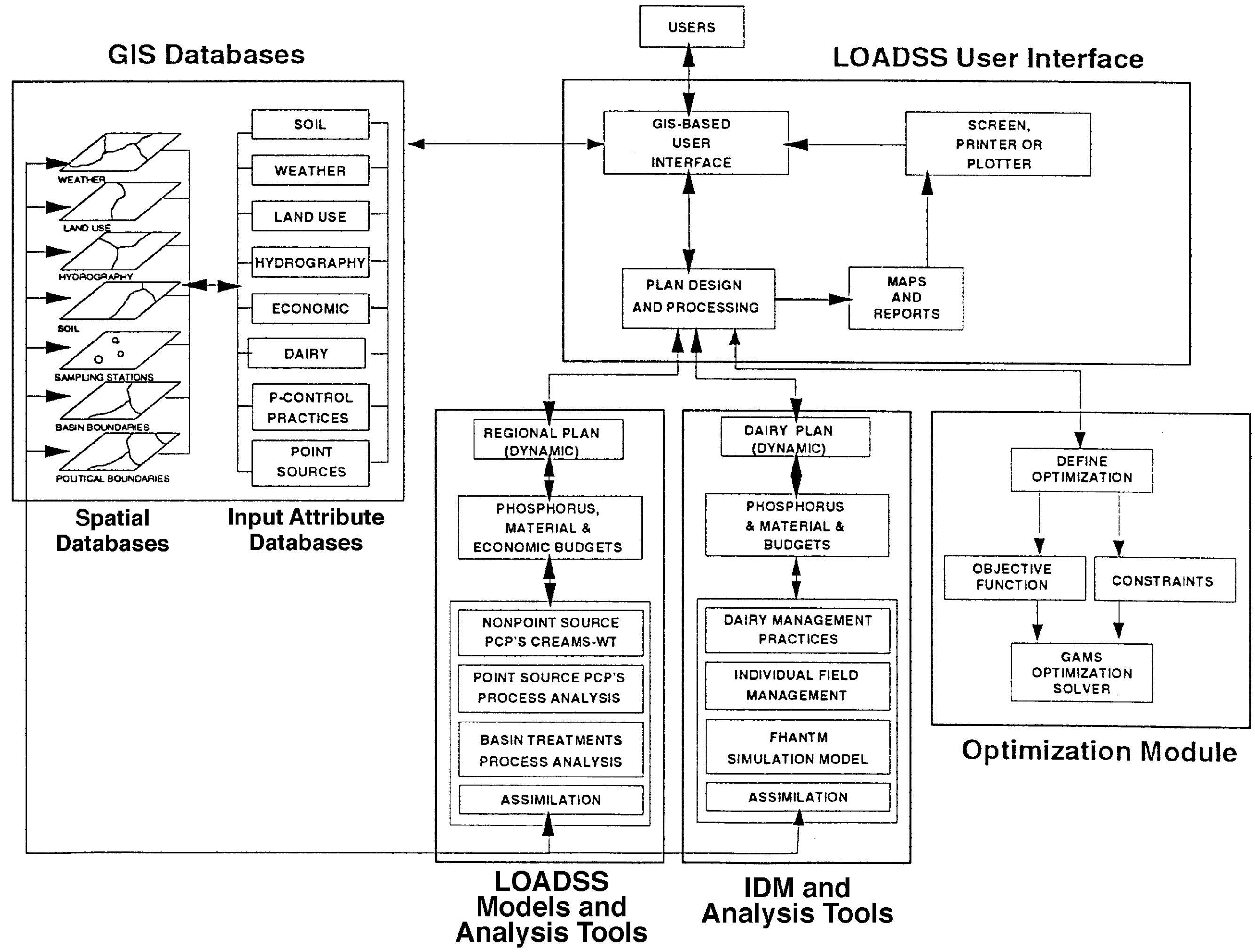


Figure 3. LOADSS design schematic (16).

task. The optimization component of LOADSS, currently under development, will determine the best combination of agricultural, environmental, and regulatory practices that protects and maintains the health of Lake Okeechobee and also maintains the economic viability of the region. The optimization process will provide another method for modifying the PCPs assigned to individual fields. Different optimization solution methods, such as linear programming and integer linear programming, will be available for solving the optimization problem that the user defines.

Dairy Simulation Model

The dairy model was expected to be fully functional by the end of 1994. It is designed to be an additional tool for answering questions related to the environmental costs and impacts of dairy operations. A design schematic of the dairy model is given in Figure 4. It differs from the LOADSS/IDM model in the following aspects:

- It is designed to be generic so that any dairy represented by a coverage for which relevant data, such as topography, soil characteristics, weather, and field boundaries, are available can be simulated.
- The GLEAMS water quality model will be used for simulating nutrient transport of nitrogen and phosphorus.
- The user will be able to assign a larger variety of crops and crop management practices to the individual fields, including crop rotation.

GLEAMS (15) is a field-scale water quality model that includes hydrology, erosion/sediment yield, pesticide, and nutrient transport submodels. GLEAMS was developed to use the management oriented CREAMS (21) model and incorporate more descriptive pesticide subroutines and more extensive treatment of the flow of water and chemicals in the root zone layers. The water is routed through computational soil layers to simulate the percolation through the root zone, but the volume of percolation in each layer is saved for later routing in the pesticide component. A minimum of three and a maximum of 12 layers with variable thickness may be used. Soil parameter values are provided by soil horizon, and

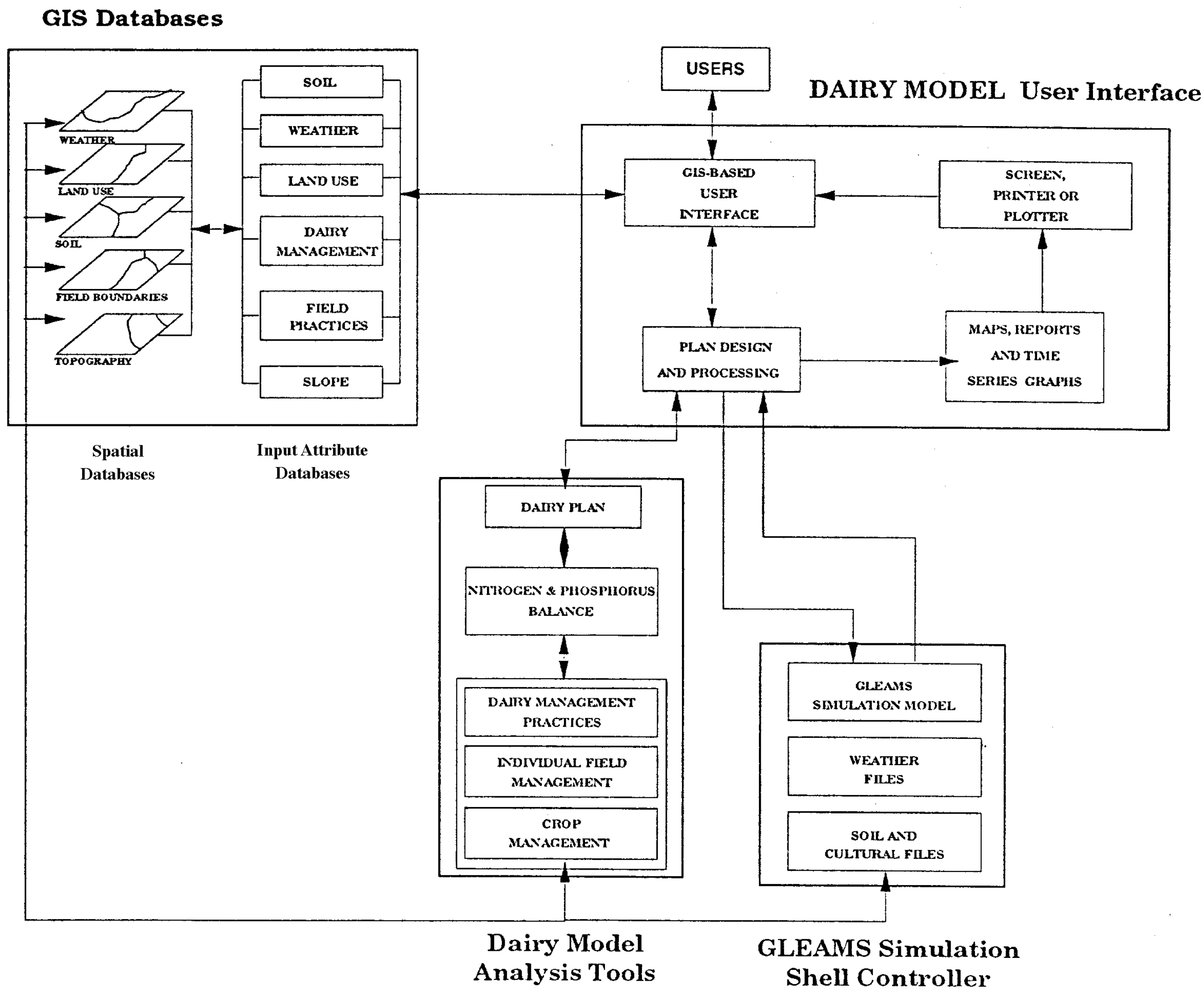


Figure 4. Dairy model design schematic.

the crop root zone may have up to five horizons. The values for parameters, such as porosity, water retention properties, and organic content, are automatically fitted into the proper computational layers. Two options are provided in the model to estimate potential evapotranspiration, the Priestly-Taylor method (22) and the Penman-Monteith method (23). The nutrient component of the model simulates land application of animal wastes by creating appropriate nitrogen and phosphorus pools for mineralization. It considers ammonia volatilization from surface-applied animal waste by using a relationship developed by Reddy et al. (24).

The graphic interface is designed to help the user plan a balanced nutrient management program for the dairy being simulated. First, total nutrient production and accounting are estimated, based on information related to the dairy management such as herd size, confinement system, waste characterization, and handling. Figure 5 shows the general structure of the graphic interface and a first version of the menu used to estimate the total amount of nitrogen and phosphorus available for assignment to the various fields. Nutrient losses during waste storage and treatment vary widely depending on the method of collection, storage, and treatment. Climate can also have a great effect on the losses. Covering all possible methods of storage and treatment is practically impossible, especially in an application that is designed to be generic and applied in any part of the country. A simplification was made to overcome this problem: the user must provide the percentage of original nitrogen and phosphorus that is retained after waste storage and treatment. The menus designed to enter information related to the management of fields and crops are given in Figure 6.

For each field, a sequence of crops can be defined in the Field Management Table, and for each crop, the sequence of practices or field operations is defined in the Crop Management Table. Every time a waste application operation is defined or a field is used as pasture for a certain period, the corresponding amount of nutri-

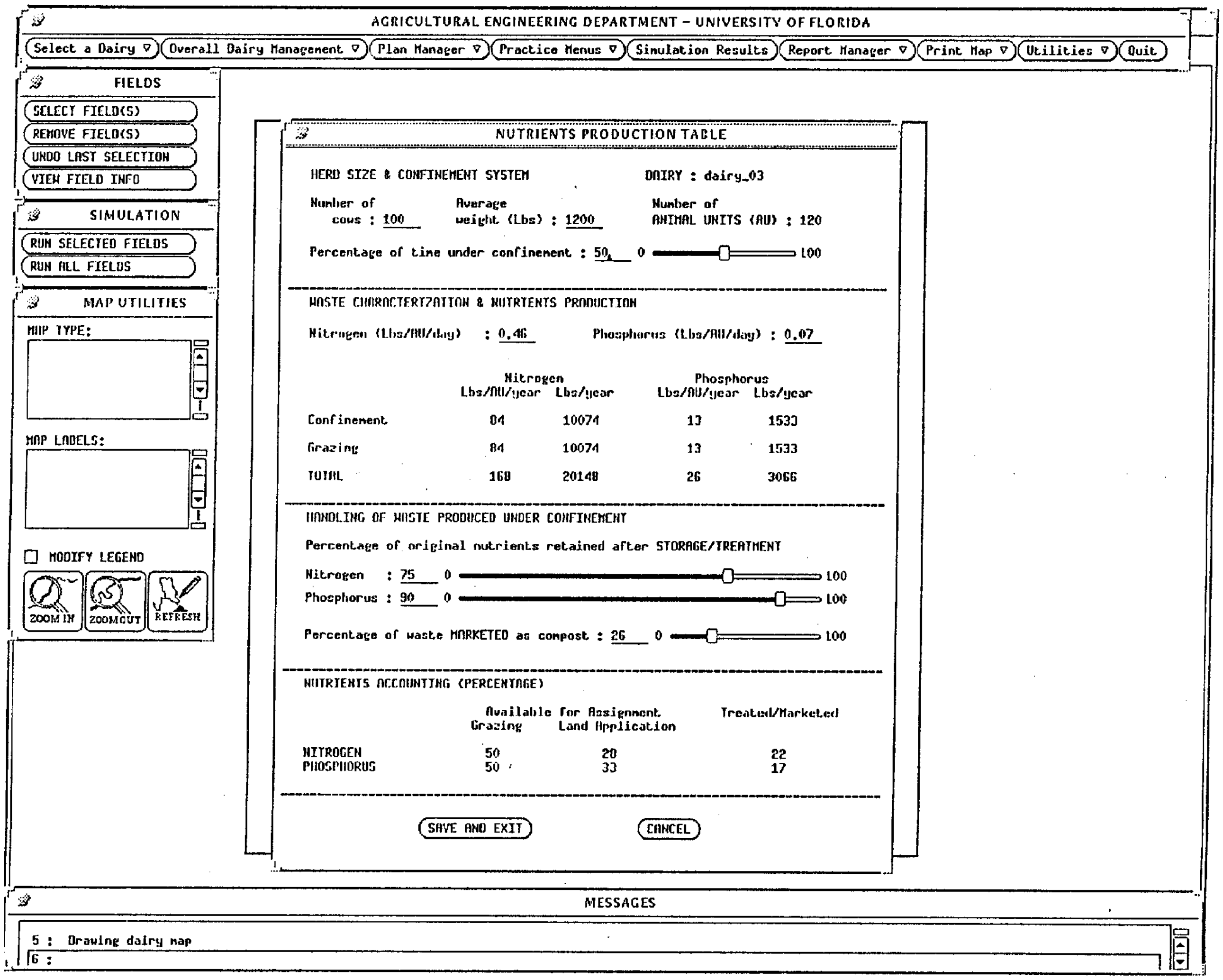


Figure 5. The nutrient production table of the dairy model interface.

ents will decrease from the amount available for assignment and the total available for future assignment will be updated. Once the total amount of nutrients is assigned, the model can be run for the several fields in the dairy and the results evaluated in terms of nutrient loadings to the edge of fields and ground water. Alternative plans can be designed and saved for comparison and selection of best management options. The best solutions in terms of reducing nutrient loadings to surface and ground water must also consider economic aspects. A producer's decision about competing waste management practices is ultimately economically motivated. Thus, the system will eventually include a tool for economic analysis of alternative management options.

Summary and Conclusions

The search for solutions to the many problems concerning nutrient management that affect water resources implies a continued demand for the development of modeling systems that can be used to analyze, in a holistic approach, the impact of alternative management policies.

The development of LOADSS exemplifies how the integration of hydrologic models and a GIS can be used for analyzing nutrient control practices at different scales. The addition of optimization algorithms further enhances the ability of policy- and decision-makers to analyze the impact of alternative management practices and land uses at the regional level.

The first part of LOADSS (Version 2.2) that includes the CREAMS-WT regional-scale model and the IDM components is fully functional and currently available at the SFWMD. Preliminary results show that LOADSS behaves consistently with measured data at the lake basin scale. Some of this, however, is due to offsetting errors in model behavior at the subbasin scale, particularly in subbasins that are adjacent to or very far from the lake. Currently, projects are underway to further verify and calibrate the model at the subbasin level to improve its

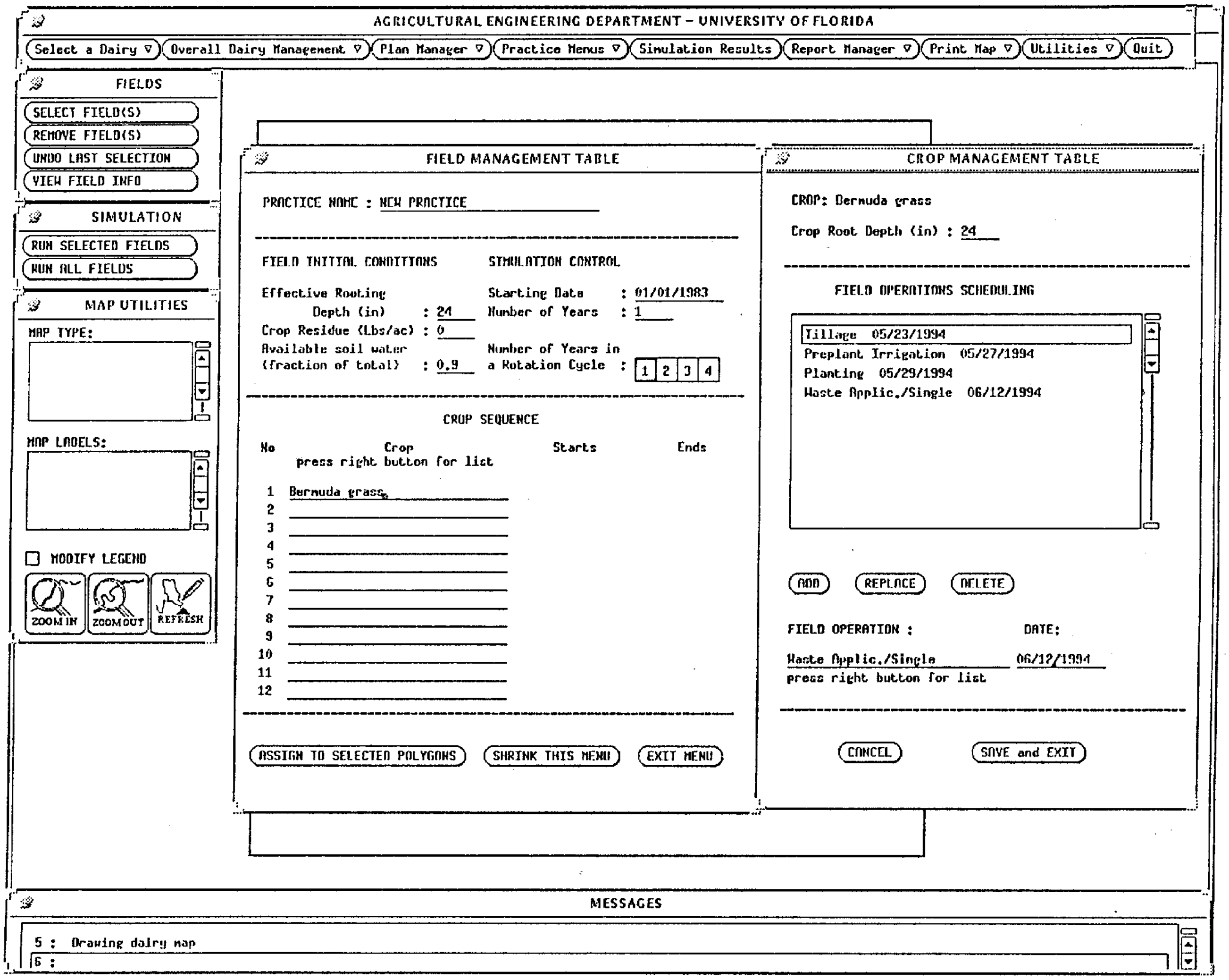


Figure 6. Field and crop management tables of the dairy model interface.

performance at smaller scales (16). Initial results of the optimization component are currently being evaluated.

The dairy model represents a different approach in integrating water quality models and GIS in that it is designed to be generic and focused mainly on the farm level. It is primarily designed to help policy- and decision-makers analyze the effects of alternative dairy waste management practices on the farm level. The framework can easily be adapted to handle different types of animal wastes (such as beef cattle and poultry) and to simulate the impact of other crop management practices such as pesticide applications.

References

1. Novotny, V., and G. Chesters. 1981. Handbook of nonpoint pollution: Sources and management. New York, NY: Van Nostrand Reinhold Company.

2. Council on Environmental Quality (CEQ). 1989. Environmental trends. Washington, DC: U.S. Government Printing Office.

3. U.S. Geological Survey (USGS). 1987. Analysis and interpretation of water quality trends in major U.S. rivers, 1974-81. Water Supply Paper 2307. Washington, DC: U.S. Government Printing Office.

4. SFWMD. 1989. Interim Surface Water Improvement and Management (SWIM) plan for Lake Okeechobee. West Palm Beach, FL: South Florida Water Management District.

5. Reck, W.R. 1994. GLEAMS modeling of BMPs to reduce nitrate leaching in Middle Suwannee River area. In: Proceedings of Second Conference on Environmentally Sound Agriculture, Orlando, FL. American Society of Agricultural Engineers. pp. 361-367.

6. Fedkiw, J. 1992. Impacts of animal wastes on water quality: A perspective from the USDA. In: Proceedings of the National Workshop on Livestock, Poultry, and Aquaculture Waste Management, Kansas City, MI (1991). American Society of Agricultural Engineers (ASAE) Pub. 03-92. pp. 52-62.

7. Woolhiser, D.A., and D.L. Brakensiek. 1982. Hydrologic system synthesis. In: Haan, C.T., H.P. Johnson, and D.L. Brakensiek, eds. Hydrologic modeling of small watersheds. ASAE Monograph No. 5. St. Joseph, MI: American Society of Agricultural Engineers. pp. 3-16.

8. Fedra, K. 1993. GIS and environmental modeling. In: Goodchild, M., B. Parks, and L. Steyaert, eds. Environmental modeling with GIS. New York, NY: Oxford University Press. pp. 35-50.

9. Tim, U.S., M. Milner, and J. Majure. 1992. Geographic information systems/simulation model linkage: Processes, problems, and opportunities. ASAE Paper No. 92-3610. St. Joseph, MI: American Society of Agricultural Engineers.

10. Whittemore, D.O., J.W. Merchant, J. Whistler, C.E. McElwee, and J.J. Woods. 1987. Ground-water protection planning using the ERDAS geographic information system: Automation of DRASTIC and time-related capture zones. In: Proceedings of the National Water Well Association (NWWA FOCUS) Conference on Midwestern Ground-Water Issues, Dublin, OH. pp. 359-374.

11. Petersen, G.W., J.M. Hamlett, G.M. Baumer, D.A. Miller, R.L. Day, and J.M. Russo. 1991. Evaluation of agricultural nonpoint pollution potential in Pennsylvania using a geographic information system. ER9105. University Park, PA: Environmental Resources Research Institute.

12. Goodchild, M. 1993. The state of GIS for environmental problem-solving. In: Goodchild, M., B. Parks, and L. Steyaert, eds. Environmental modeling with GIS. New York, NY: Oxford University Press. pp. 8-15.

13. Liao, H.H., and U.S. Tim. 1992. Integration of geographic information system (GIS) and hydrologic/water quality modeling: An interface. ASAE Paper No. 92-3612. St. Joseph, MI: American Society of Agricultural Engineers.

14. Young, R.A., C.A. Onstad, D.D. Bosch, and W.P. Anderson. 1989. AGNPS: A nonpoint source pollution model for evaluation of agricultural watersheds. J. Soil and Water Conserv. 44(2):168-173.

15. Leonard, R.A., W.G. Knisel, and D.A. Still. 1987. GLEAMS: Ground-water loading effects of agricultural management systems. Trans. of the Amer. Soc. of Agricul. Eng. (ASAE) 30(5):1,403-1,418.

16. Negahban, B., C. Fonyo, W. Boggess, K. Campbell, J. Jones, G. Kiker, E. Hamouda, E. Flaig, and H. Lal. 1993. A GIS-based decision support system for regional environmental planning. In: Proceedings of the Conference on Application of Advanced Information Technologies: Effective Management of Natural Resources, Spokane, WA. American Society of Agricultural Engineers (ASAE). pp. 169-178.

17. Heatwole, C.D., K.L. Campbell, and A.B. Bottcher. 1987. Modified CREAMS hydrology model for coastal plain flatwoods. Trans. of the Amer. Soc. Agricul. Eng. (ASAE) 30(4):1,014-1,022.

18. Kiker, G.A., K.L. Campbell, and J. Zhang. 1992. CREAMS-WT linked with GIS to simulate phosphorus loading. ASAE Paper No. 92-9016. St. Joseph, MI: American Society of Agricultural Engineers.

19. Skaggs, R.W. 1980. DRAINMOD reference report. U.S. Department of Agriculture, Soil Conservation Service. Fort Worth, TX: South National Technical Center.

20. Tremwel, T.K., and K.L. Campbell. 1992. FHANTM, a modified drainmod: Sensitivity and verification results. ASAE Paper No. 92-2045. St. Joseph, MI: American Society of Agricultural Engineers.

21. Knisel, W.G. 1980. CREAMS: A field-scale model for chemicals, runoff, and erosion from agricultural management systems. Conserv. Res. Rep. No. 26. Washington, DC: U.S. Department of Agriculture.

22. Priestly, C.H.B., and R.J. Taylor. 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly Weather Rev. 100:81-92.

23. Jensen, M.E., R.D. Burman, and R.G. Allen, eds. 1990. Evapotranspiration and irrigation water requirements. Manuals and reports on engineering practice. Amer. Soc. Civil Eng. No. 70.

24. Reddy, K.R., R. Khaleel, M.R. Overcash, and P.W. Westerman. 1979. A nonpoint source model for land areas receiving animal waste: II. Ammonia volatilization. Trans. of the Amer. Soc. Agric. Eng. (ASAE) 22(6):1,398-1,405.

Other GIS Applications

Expedition of Water-Surface-Profile Computations Using GIS

Ralph J. Haefner, K. Scott Jackson, and James M. Sherwood
Water Resources Division, U.S. Geological Survey, Columbus, Ohio

Abstract

Water-surface profiles computed by use of a step-backwater model such as Water Surface PROfile (WSPRO) are frequently used in insurance studies, highway design, and development planning to delineate flood boundaries. The WSPRO model requires input of horizontal and vertical coordinate data that define cross-sectional river-channel geometry. Cross-sectional and other hydraulic data are manually coded into the WSPRO model, a labor-intensive procedure. For each cross section, output from the model assists in approximating the flood boundaries and high-water elevations of floods with specific recurrence intervals (for example, 100-year or 500-year). The flood-boundary locations along a series of cross sections are connected to delineate the flood-prone areas for the selected recurrence intervals.

To expedite the data collection and coding tasks required for modeling, the geographic information system (GIS), ARC/INFO, was used to manipulate and process digital data supplied in AutoCAD drawing interchange file (DXF) format. The DXF files, which were derived from aerial photographs, included 2-foot elevation data along topographic contours with +0.5-foot resolution and the outlines of stream channels. Cross-section lines, located according to standard step-backwater criteria, were digitized across the valleys. A three-dimensional surface was generated from the 2-foot contours by use of the GIS software, and the digitized section lines were overlain on this surface. GIS calculated the intersections of contour lines and cross-section lines, which provided most of the required cross-sectional geometry data for input to the WSPRO model.

Most of the data collection and coding processes were automated, significantly reducing labor costs and human error. In addition, maps at various scales can be easily produced as needed after digitizing the flood-prone areas from the WSPRO model into GIS.

Introduction and Problem Statement

Losses due to flood damage generally cost the American public hundreds of millions of dollars annually. In 1968, the National Flood Insurance Act established the National Flood Insurance Program (NFIP) to help reduce the cost to the public and provide a framework to help reduce future losses. The Federal Emergency Management Agency (FEMA) administers the NFIP. As listed in Mrazik and Kinberg (1), the major objectives of the NFIP are to:

- Make nationwide flood insurance available to all communities subject to periodic flooding.
- Guide future development, where practical, away from flood-prone areas.
- Encourage state and local governments to make appropriate land use adjustments to restrict development of land that is subject to flood damage.
- Establish a cooperative program involving the federal government and the private insurance industry.
- Encourage lending institutions, as a matter of national policy, to assist in furthering program objectives.
- Authorize the continuing studies of flood hazards.

Studies of flood-prone areas typically involve using step-backwater computer algorithms (digital models) to estimate river water-surface profile elevations and flood-inundation patterns along the topography of the river and its overbanks. FEMA recognizes the U.S. Geological Survey's (USGS's) step-backwater model, Water Surface PROfile (WSPRO), as a suitable computer model for use in flood insurance studies (2, 3). Basic data input for step-backwater models includes:

- Estimates of flood discharge and initial water-surface elevations.
- Stream cross-sectional geometry.
- Roughness coefficients for cross sections.

- Contracted opening geometry if bridges or culverts are located along the study reach.

Obtaining meaningful model results typically requires numerous stream cross sections referenced to a common elevation datum along a stream reach. The data-collection efforts to obtain these cross-sectional data require costly, labor-intensive fieldwork. Study efforts along lengthy stream reaches may, however, involve the generation of a contour map using aerial photogrammetric mapping techniques. Processing the spatial data may still require extensive labor to extract the cross-sectional data needed for the WSPRO model.

The development of geographic information systems (GIS) technology has greatly enhanced analyses of spatial data such as topography. In an effort to improve the quality of mapping and delineation of flood-prone areas in Summit County, Ohio, the USGS developed a method of using a GIS as a pre- and postprocessor of the input and output data for the WSPRO model. This paper describes the steps the USGS used to develop this interface and discusses some difficulties encountered during the process.

Approach

Several steps were taken that resulted in the delineation of a flood-prone area in Summit County, Ohio. These steps are shown in a flow chart (see Figure 1) and described below.

Data were obtained for this study via aerial photography during April 1990. These data include mappable features at the given scale including topography at 2-foot contour intervals, stream boundaries, roads, and buildings. The data are estimated to be vertically accurate to +0.5 feet. The data were put into AutoCAD and were prepared for delivery to the USGS on 3.5-inch floppy disks in AutoCAD drawing interchange file (DXF) ASCII format. ARC/INFO was used to convert the DXF file into two separate data layers containing only the topography and traces of stream banks within the study area.

A three-dimensional surface was generated from the topographic data using the ARC/INFO software package Triangulated Irregular Network (TIN). Cross-section lines were digitized over the topography data layer. The cross sections were placed according to standard step-backwater criteria (4) and were generally:

- Perpendicular to stream flow
- At major breaks in streambed profiles
- At minimum and maximum cross-sectional areas
- At major changes in stream conveyance
- Spaced about one cross-section width apart

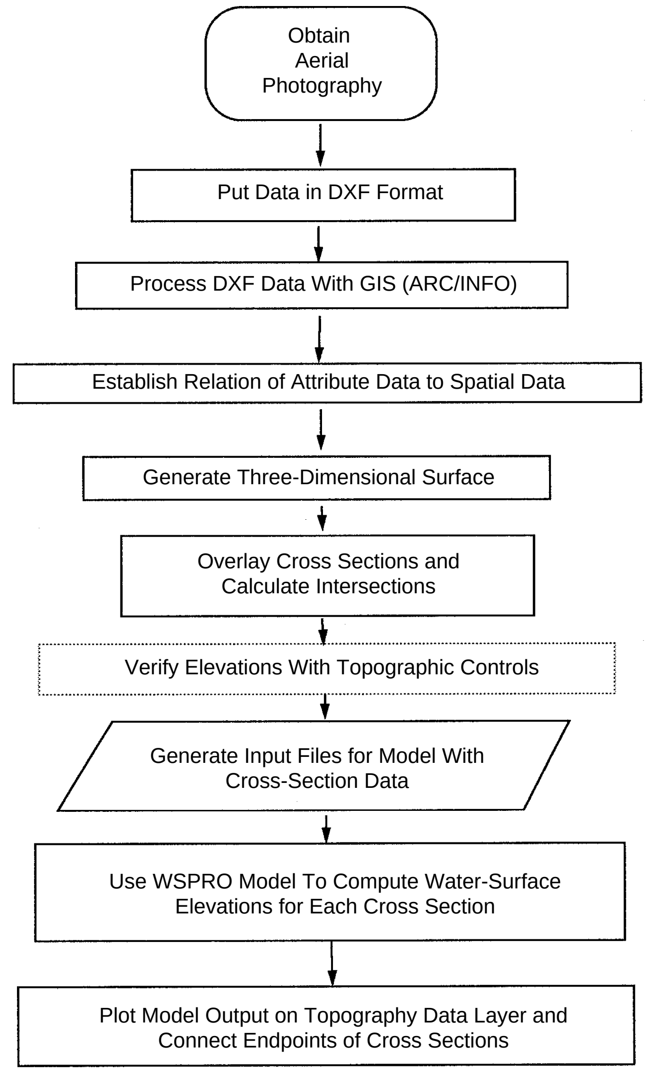


Figure 1. Flowchart of data conversion and processing for use in the Water Surface PROfile.

The cross-section lines were then overlaid on the three-dimensional surface of topography, and GIS calculated the intersections of the contour lines and cross sections. The locations and elevations of these intersections were output as an ASCII file and slightly modified for input into the WSPRO model.

These GIS data were used along with the aforementioned required data as input to the WSPRO model. Input for the model included estimates of the 100-year flood discharge (5), stream cross-sectional geometry (supplied by this work), and estimates of roughness coefficients for cross sections. The WSPRO model was then run, providing output in the form of water-surface elevations at specific distances along section lines corresponding to the simulated elevation of a 100-year flood.

Points corresponding to the flood elevations along the cross-section lines were plotted on the topography data layer and were connected manually (to delineate flood boundaries) by interpolating the elevations with respect to adjacent contours. A polygon of the flood "surface" was generated and drawn on a map (see Figure 2).

Results

The supplied topographic data were of sufficient quality and resolution to substitute for field-surveyed elevations; however, field surveys to verify the elevations along the cross sections would augment this quality control process (see Figure 1). Typically, a crew of two individuals may take up to 4 days to survey and reduce the field data for the study area chosen for this study. Because aerial photography is commonly substituted for land surveying, the most significant effort and source of error may come from manually extracting elevations and distances along cross sections for input into the WSPRO

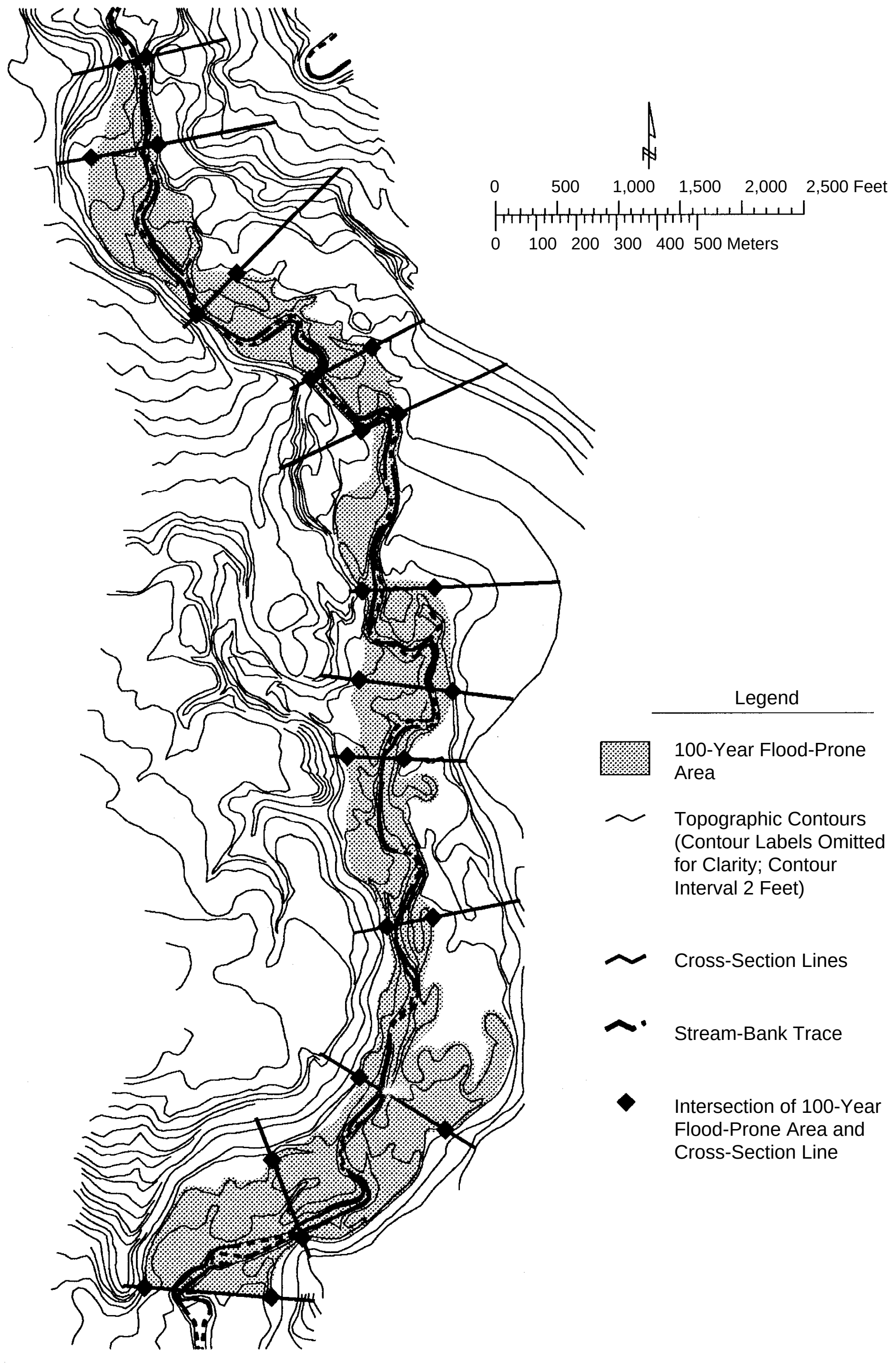


Figure 2. Watershed showing delineation of 100-year flood-prone area.

model. Initial development of the method to use GIS for this analysis took approximately 1 week to refine; however, future analyses would probably only take one person 1 day to perform. This represents a significant cost savings. Additionally, reducing the amount of human-induced error can substantially improve the reliability and accuracy of the computer-generated flood-prone area data.

Because topography, stream traces, and other features are supplied in the DXF file, these data can easily be brought into GIS. Maps can be made that show these features in relation to the predicted flood-prone area. Maps showing a variety of features can be produced at any scale, with accuracy limited only by the accuracy of the source scale. Additionally, GIS can calculate the intersection of map features that may lie within the flood-prone area, such as buildings that may contain hazardous materials. GIS can also overlay land use data layers within the flood-prone area to define areas that should not be developed or that have already been overdeveloped in accordance with the aforementioned NFIP objectives.

FEMA now requests that future flood-study mapping be completed using GIS format, a common goal that both the USGS and FEMA are working toward. These data are important to land planners, flood-plain regulators, and insurance companies that rely on accurate estimates of flood-prone areas. By increasing the accessibility of the data by using GIS, we can substantially improve our ability to analyze spatial data efficiently.

Problems Encountered

Problems using data supplied in DXF format in conjunction with GIS resulted primarily from the fact that the DXF data were prepared for the purpose of making a topographic map, not a GIS data layer. The contour lines were segmented; that is, where ends of segments met, they were not physically connected to form a topologically viable data layer. The data layer needed to be edited because GIS requires topology for spatial-data processing. Additionally, in areas where the topographic gradient was particularly steep, contour lines were omitted. In both cases, an attempt was made to allow GIS to establish a physical connection of contour lines, but subsequent manual interpolation was also required. This may have introduced error into the data set. If future work requires the use of DXF data, the request for data should specifically state that all topographic contours be continuous.

AutoCAD stores data differently from GIS, so a relationship needed to be established between the data file containing elevations and the data file associated with the lines that make up the topography data layer. Several lines from the DXF file did not have any data associated with them, thus necessitating the addition of contour elevation data by context with the adjacent contours that did have data. This step may also have introduced errors, but quality-control measures to verify the topographic contours and contour elevations could help to minimize these errors.

Output from the WSPRO model is in the form of a series of points along cross sections that were connected by manual interpolation. This step also may introduce some error, but the same process must be performed when not using GIS.

Conclusions

This report documents an example of how GIS can be used to facilitate step-backwater modeling of flood-prone areas. The results of the study show that significant savings may be expected in the form of reduced labor requirements. Furthermore, FEMA now requires the use of GIS to conduct flood-study mapping, thus providing a means to conduct additional spatial analyses more efficiently. As aerial photography and GIS technology improve, although additional sources of error may arise, the overall accuracy, reliability, and reproducibility of the model input and results should also improve.

References

1. Mrazik, B.R., and H.A. Kinberg. 1991. National flood insurance program: Twenty years of progress toward decreasing nationwide flood losses. Water Supply Paper 2375. U.S. Geological Survey.
2. Shearman, J.O., W.H. Kirby, V.R. Schneider, and H.N. Flippo. 1986. Bridge waterways analysis model. Research Report FHWA/RD-86/108. Federal Highway Administration.
3. Shearman, J.O. 1990. Users manual for WSPRO, a computer model for water-surface profile computations. FHWA-IP-89-027. Federal Highway Administration.
4. Davidian, J. 1984. Computation of water-surface profiles in open channels: Techniques of water-resources investigations of the United States Geological Survey. In: Applications of Hydraulics, Vol. 3.
5. Koltun, G.F., and J.W. Roberts. 1990. Techniques for estimating flood-peak discharges of rural unregulated streams in Ohio. Investigations Report 89-4126. U.S. Geological Survey, Water Resources.

Small Is Beautiful: GIS and Small Native American Reservations—Approach, Problems, Pitfalls, and Advantages

Jeff Besougloff
Upper Sioux and Lower Sioux Indian Communities, Redwood Falls, Minnesota

Background

The Lower Sioux Indian Reservation

The Lower Sioux Indian Reservation covers 1,743 acres in southwestern Minnesota bordering the Minnesota River. The land base consists of several hundred acres of prime, flat agricultural land, a large wetlands slough complex, prairie pothole wetlands, bottom land wetlands, small lakes, and approximately 250 acres of timber and brush. The elevation ranges from Minnesota River level to the adjacent bluffs several hundred feet higher.

This rural reservation contains a moderate amount of infrastructure, including paved and dirt roads, 12-acre sewage lagoon serving a moderately sized casino, community water system composed of a tower and small treatment plant for the 90 mostly single-family dwelling homes, convenience store/gas station/gift shop, community center, small two-room schoolhouse, pottery works with gift shop, warehouse, and church. The casino-fueled economic boost to the community recently resulted in improvements to infrastructure and plans for additional projects.

Office of the Environment

The tribal government was formed under the Indian Reorganization Act of 1934. The governing body is an elected five-person tribal community council that administers several government departments and is responsible for all government activities.

Under a U.S. Environmental Protection Agency (EPA) Region 5 multimedia grant, the Upper Sioux and Lower Sioux Office of the Environment (OE) was formed in late 1992. This unique joint venture between two tribes and EPA envisioned moving the tribal governments into compliance with major federal environmental legislation. At the present time, only the Lower Sioux are developing a tribal geographic information system (GIS). Therefore, this article is solely applicable to this community, although adoption of an Upper Sioux Community GIS would likely follow a similar lifeline.

Environmental Regulation in Indian Country

Reservations are subject to a bewildering array of environmental regulations. Numerous meetings, publications, projects, and court decisions are devoted to determining what law does or does not apply on any particular reservation. In very general terms, the following can be stated: state environmental regulations do not apply, federal regulations do apply, and tribal regulations may apply. From a tribal environmental employee's point of view, numerous environmental regulations (whether federal or tribal) do exist that apply to reservation activities and land, and they require compliance.

The Problem and the Solution

The OE's responsibility is to bring the reservation into compliance with the 14 major pieces of environmental legislation administered through EPA and directly applicable to tribes. The OE finds itself responsible for any and all other applicable environmental regulations and all other less-regulated environmental media. The OE currently has a staff of one.

In addition to the responsibility of moving the tribes into compliance with federal environmental regulations, the OE also develops environmental infrastructure, institutes environmental programs, and performs grant writing. Lower Sioux programs currently include Clean Water Act (CWA) and Safe Drinking Water Act (SDWA) compliance, solid waste planning including development and institution of a household recycling program, wetlands regulations compliance, wetlands mapping and restoration, National Environmental Policy Act (NEPA) compliance and site assessments, basic hydrological data gathering and mapping, radon testing and mitigation, environmental education as necessary, SARA

Title III compliance and planning, and a variety of related tasks.

Contracts or grants are currently being administered under several Bureau of Indian Affairs (BIA) programs, U.S. Geological Survey (USGS) and U.S. Army Corps of Engineers (COE) matching funds programs, two EPA programs, one Federal Emergency Management Agency (FEMA) program, one Administration of Native Americans (ANA) program, one Great Lakes Intertribal Council (GLITC) program, and a cooperative project with the National Tribal Environmental Council (NTEC).

Needless to say, responsibilities of the OE are limited by staff hours rather than need. As distressing as the reservation's unaddressed environmental needs are, equally distressing (prior to GIS development) was the helter skelter manner in which the OE digested the data and information flowing into the office. Because of its broad responsibilities and the administrative problems being encountered, the OE began to investigate developing a tribal GIS.

The Lower Sioux GIS

System Selection

The Lower Sioux GIS system is a networked PC station through the Bureau of Indian Affairs Geographic Data Service Center's (BIA GDSC's) two Sun MP690 SparcServers in Golden, Colorado. The GDSC is the hub of BIA's GIS and remote sensing program, known as the Indian Integrated Resources Information Program (IIRIP). The purpose of the IIRIP is twofold: first, make GIS and remote sensing technology available to tribes and BIA personnel; second, transfer these technologies to tribal organizations.

Database development and management functions, technical support, development of simplified user interfaces, remote sensing interpretation, and implementation of equipment directives are performed by approximately 30 GDSC employees for approximately 230 GDSC users. User technical support is also available through BIA field offices, each of which has a designated GIS coordinator. Simplified user interfaces for specialized programs have been developed including the Lightening Display System and the Land Title Mapping System. Quality control is provided for non-BIA-produced data that will be inputted.

The GDSC has standardized on the ARC/INFO family of software produced by Environmental Systems Research Institute (ESRI). GDSC has developed a number of hardware/software configuration options depending on tribal needs and financial resources and based upon GDSC experience. The OE happily relies on this experience to avoid the familiar horror stories related to equipment and software incompatibility.

Based upon GDSC configuration advice, the initial GIS setup will be on the OE's existing Compaq PC using Tektronix Terminal Emulation software (EM4105) and a Multitech modem (MT932BA). The system can use the OE's Hewlett Packard (HP) Deskjet 500, although a significant upgrade, possibly to an HP Paint or HP Excel Paint, is soon expected. Initial startup hardware and software costs are minimal in this configuration. Costs for the above equipment and introductory training are less than $5,000.

GIS Users at Lower Sioux

Initial setup and data loading will be in the OE, and the OE employee will receive introductory training on the system. Because the OE is formed through a cooperative agreement between two tribes, the Upper Sioux and the Lower Sioux, the OE is centrally located between the reservations. The system will probably be relocated to the Lower Sioux Community Center within 1 year. A tribal government employee will receive advanced GIS training and be available for all tribal government departments and businesses.

Funding

In addition to tribal contributions, funding has come through several sources and joint agreements with the tribe and BIA, EPA, and ANA.

Training

The GDSC supplies no-cost training to tribes. The *Geographic Data Service Center 1995 Training Catalog* (no federal document number available) offers eight formal courses repeatedly throughout the year, a 5-week intern program, and a cooperative student program. Courses are held at the GDSC or by request at BIA field offices and tribal locations.

The GDSC also produces the monthly *The Service Center Review* (ISSN 1073-6190), a helpful compilation of current issues, available resources, system bugs, and other items of interest to GDSC users.

Data Collection and Input

Data collection can be divided into three categories: aerial photography, portable global positioning system (GPS) data, and ARC/INFO export files created under contracted studies.

Aerial Photography

Surface features and topography will be obtained using aerial photography reduced to GIS format, then downloaded to the GDSC. Coverages will consist of 62 categories of features on a scale of 1 inch = 100 feet with 2-foot contour intervals.

Global Positioning Systems

Use of a portable Trimble, Inc., GPS Pathfinder Pro XL submeter GPS mapping system purchased with assistance from an ANA grant will allow updating of surface features and addition of nonsurface features as necessary. It will also facilitate input of attribute data.

The GPS will also be used during field work by USGS on the Lower Sioux hydrological mapping project to obtain data that otherwise would not be put into ARC/INFO export file for any reason (i.e., it might not be directly related to the project at hand or outside the agreed upon data to be converted to ARC/INFO export file form but nevertheless is of importance to the OE). The alternative is that this type of information never makes it into the GIS and is lost.

ARC/INFO Export Files

Fortunately, most federal agencies that supply funding to tribes for environmental work are well versed in GIS applications and the need for GIS-ready data. The OE now requires all information and mapping to be delivered as an ARC/INFO export file with data registered to a real world coordinate system. Downloading of this data to the GDSC mainframes allows for direct input of the data. The OE has contemplated, but not acted on, conversion of existing data for the GIS. This is an expensive and time consuming process that must be weighed in comparison with recollecting the data. Ironically, the lack of reservation data therefore becomes a benefit because time consuming and expensive data conversion is unnecessary.

The Intertribal GIS Council

Information gathering, networking, and addressing uniquely tribal problems were some of the accomplishments at the first annual meeting of the Intertribal GIS Council (IGC) held in June 1994. Vendors as well as BIA regional office and GDSC representatives answered questions and presented panels. This annual conference is likely to become a major benefit to the tribe as it continually develops the GIS.

The Future

As the tribal government becomes more familiar with the GIS, its uses, and advantages, recognized governmental needs will likely drive the development of further coverages. The OE also expects to access existing governmental data of importance to the tribe in an effort to expand the GIS database and is actively seeking sources of such information.

Philosophical Caveat

Albert Einstein stated that, “The significant problems we face cannot be solved at the same level of thinking we were at when we created them.” Some assume that GIS is the next level of reasoning in the environmental profession because we can accomplish tasks more quickly, more efficiently, with more variables accounted for, and beyond what we could have hoped to accomplish prior to GIS.

Essentially, what we have gained is speed and the capacity to include additional data, which is not what Einstein was referring to when he spoke of the next level. Wisdom, in the sense of a higher level of understanding, is the necessary ingredient to the solution of current environmental problems; in other words, movement beyond the paradigm that created the problem. GIS may be the tool that pushes the environmental professional to the next level of wisdom by presenting the data and information in a manner that allows the user to stand back and see more clearly on a higher plane. But that level can be found only within the environmental professional himself or herself and not within GIS.

A GIS-Based Approach to Characterizing Chemical Compounds in Soil and Modeling of Remedial System Design

Leslie L. Chau, Charles R. Comstock, and R. Frank Keyser
ICF Kaiser Engineers, Inc., Oakland, California

Introduction: The Problem

The cost-effectiveness of implementing a computerized geographic information system (GIS) for environmental subsurface characterization should be based on long-term remedial objectives. A GIS project was developed to characterize soil contamination and to provide design parameters for a soil vapor extraction remedial system, as part of a $120-million remediation and “land sale” project in California. The primary purposes of the GIS were to efficiently combine and evaluate (model) disparate data sets, provide “new” and more useful information to aid in short-term engineering decisions, and support the development of long-term cleanup goals.

The project had a major change in scope early on, and the schedule was expedited to allow for the development of “land sale” options and for actual site redevelopment at the earliest opportunity. Characterization of chemically affected soil would have been compromised given the above circumstances without an ambitious undertaking of concurrently developing and implementing a GIS with three-dimensional (3-D) geostatistical and predictive modeling capabilities.

The GIS Approach

Computer solutions included the use of a cross-platform (DOS and UNIX) GIS to quickly and systematically incorporate spatial and chemical data sets and to provide a distributed data processing and analysis environment (see Figure 1). Networked, DOS-based relational databases were used to compile and disseminate data for the numerous investigatory and engineering tasks. UNIX-based computer aided design (CAD) and modeling applications received data from databases, performed quantitative analyses, and provided 3-D computer graphics. Given the aggressive project schedule, exclusive use of one platform would not be realistic due mainly to the limited data modeling capacity and 3-D graphics in DOS systems. On the other hand, the high startup and operating costs of several UNIX workstations would render their exclusive use much less cost-effective.

The hardware and software configurations were integrated in a client/server Intergraph InterPro 6400 with 48 megabytes of memory. It is largely a 3-D CAD system with add-on modules of geologic mapping and 3-D visual models capable of consolidating both environmental and engineering parameters for analysis (see Figure 1). Textual environmental and geologic data were extracted by SQL queries from relational databases and were transferred to mapping and modeling modules via PC/TCP cross-platform linkage.

The GIS assisted in making short- and long-term decisions regarding health-risk-based regulatory strategy and engineering feasibility. Use of spatial statistical and predictive models was part of a GIS-based decision-making loop (see Figure 2). The process supported concurrent activities in:

- Data collection: field program.
- Numerical models of remedial system configurations.
- Development of cleanup goals from health risk assessments.
- Remedial design with CAD capability.

Site Background

In early 1993, the remedial investigation of the operable unit for soil at a former aircraft manufacturing facility in southern California was thought to be ready for remedial alternatives feasibility study. ICF Kaiser Engineers, Inc., was awarded the contract to perform feasibility studies on applicable soil cleanup technologies and to subsequently design and manage the installation and early operation of the selected technologies. After $700,000 was spent evaluating data collected by previous consultants, it was decided that an additional $5 million worth

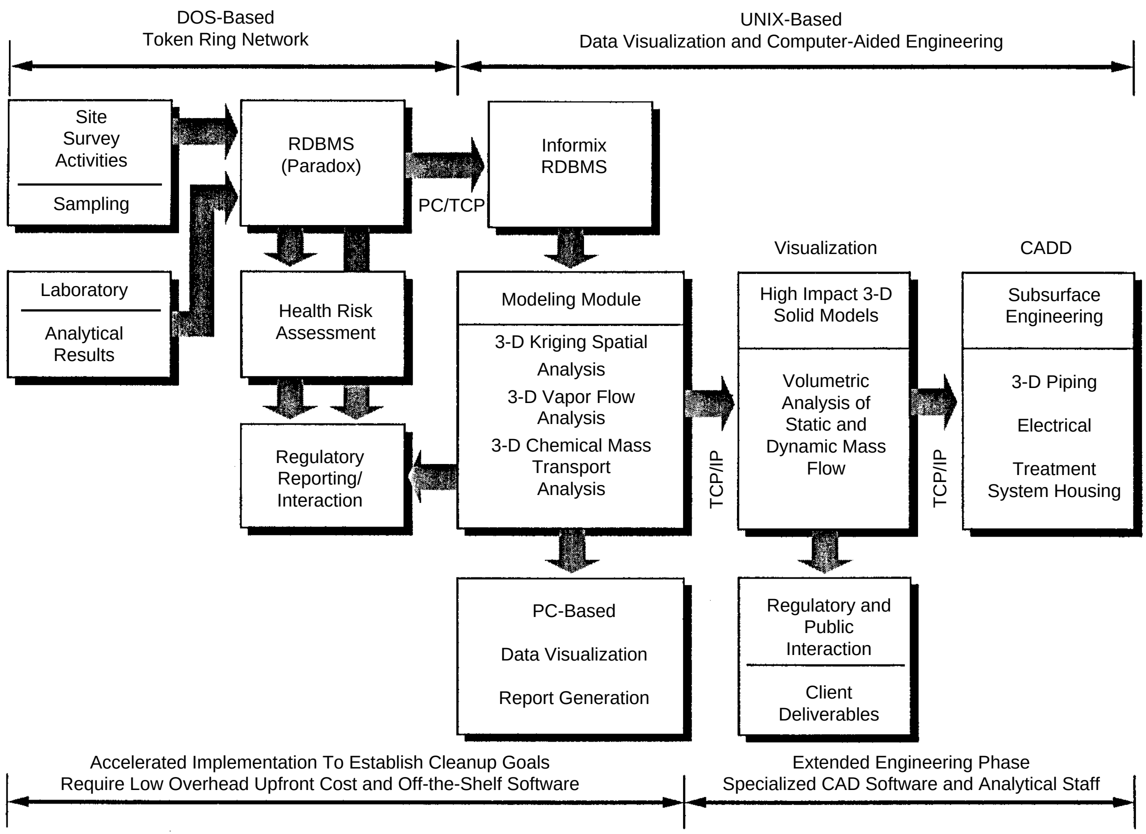


Figure 1. Multiplatform GIS project.

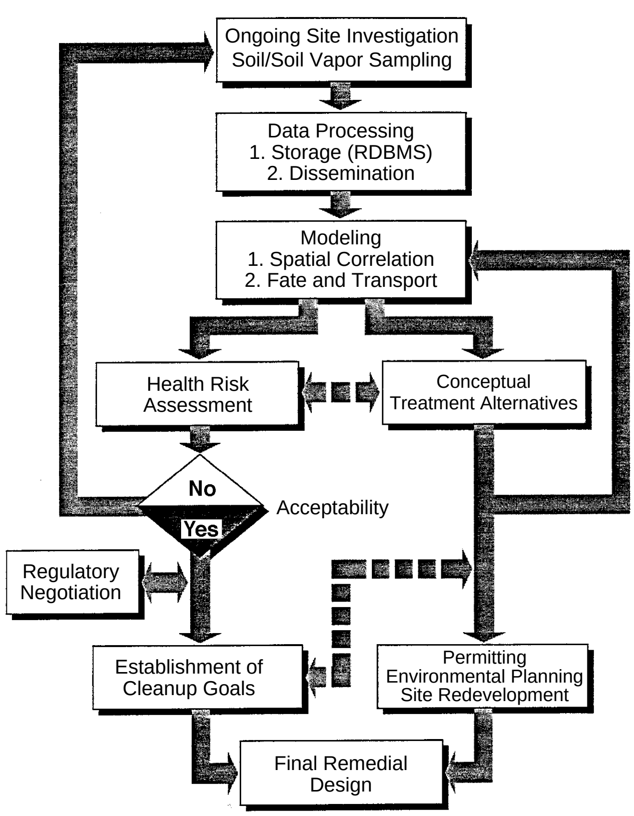


Figure 2. GIS-assisted decision tree.

of field activities were required to more definitively estimate the volume of chemically affected soil and the nature and extent of contamination at the facility. Because of the data gaps, the selection and design of alternatives could not be addressed with a high degree of certainty. Hence, computer assisted data processing was crucial to speed up the feasibility study, accelerate downstream work, and reduce the overall project schedule to the minimum.

The site is environmentally complex, covering an area of approximately 120 acres. As a result of nearly half a century of aircraft production and development, soil beneath the facility is affected by fuel and heavy oil hydrocarbons (TPH) commingled with volatile compounds, mainly perchloroethylene (PCE) and trichloroethylene (TCE) (see Figure 3). Ground water at 170 feet below ground surface is affected by TCE and PCE, but it is not part of the drinking water aquifer. The facility has been demolished, and shallow contaminated soil has been excavated and back-filled to an interim grade.

Methodology

Health-Risk-Based Cleanup Goals

Central to determining the volume and kinds of data to be collected was the question of whether chemicals in soil represented potentially unacceptable risks to human

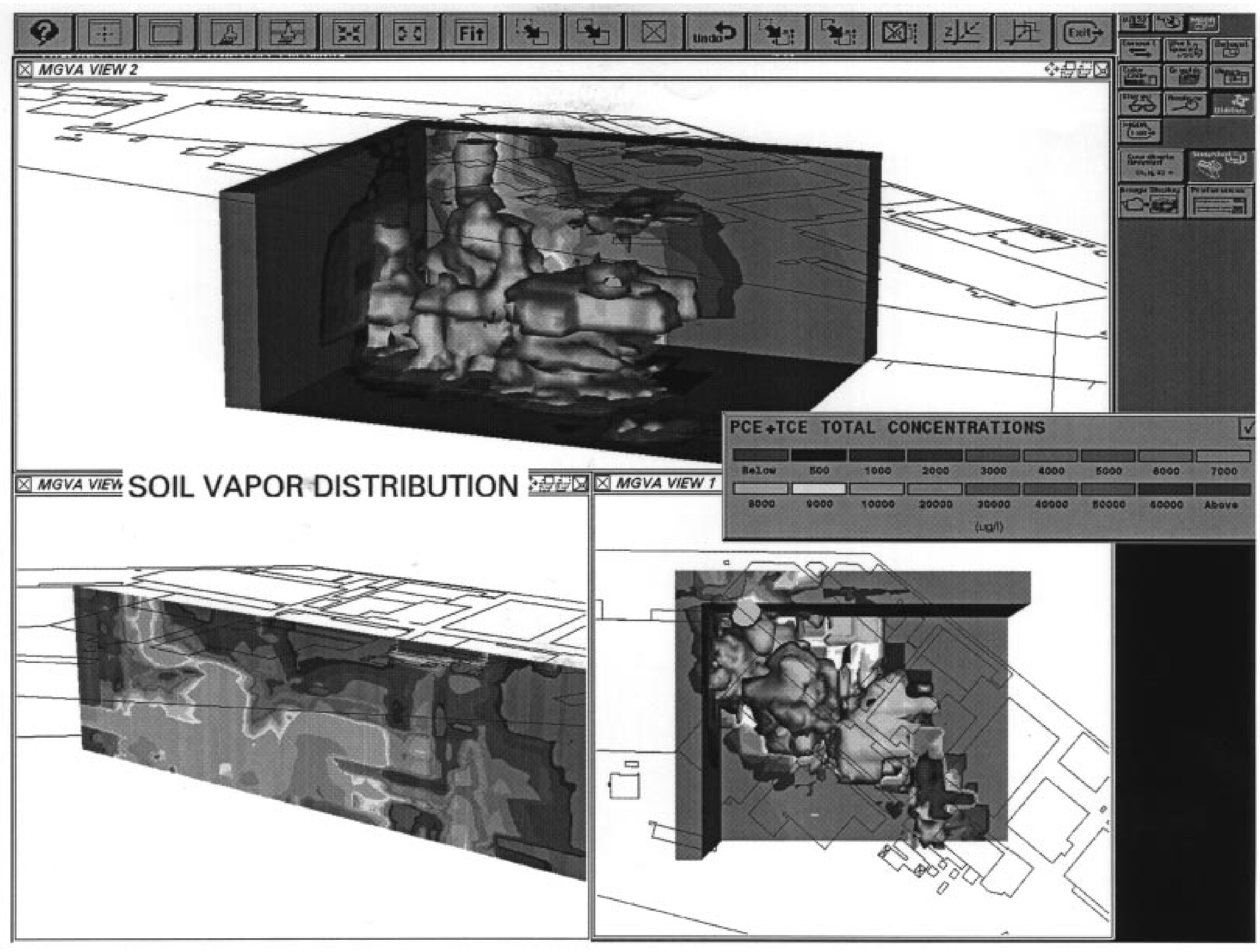


Figure 3. Aircraft manufacturing facility in California. Outline of demolished buildings located at the 120-acre site are shown as surface features for reference. A geostatistical model of a 3-D kriged VOC soil vapor cloud in the subsurface was simulated with Intergraph's MGVA. Views displayed are: 3-D isometric, vertical section of chemical isoplaths, and a nearly plan view. Digital simulation also illustrates VOCs affecting ground water in a dispersive nature at a depth of nearly 170 feet bgs (shown at bottom of isometric view).

health and to the environment, with the former being of particular concern to construction workers onsite during redevelopment.

Because site redevelopment was scheduled to begin in the near term, data collection and GIS analysis concentrated on shallow depths (top 20 feet), with decreasing sample density at greater depths. A health-risk-based cleanup goal (HBCG) approach to collecting more data was to establish cleanup goals for near-term remediation of the shallow soils as well as for long-term remedial measures of contaminated soils at greater depths. Further, various regulatory agencies had to approve the estimated cleanup goals in a short time. Ongoing site demolition and excavation schedules encouraged the aggressive regulatory negotiations. The shallow cleanup goals for volatile organic compounds (VOCs) and TPH determined the volume of contaminated soil to be removed. At greater depths, data gaps were minimized to more definitively characterize the nature of TPH and VOC contaminations and to facilitate the implementation of long-term remedial objectives (i.e., in situ soil vapor extraction).

In situ soil vapor and soil sampling composed the field program, which provided data to map the subsurface distribution of volatile organic compounds, including TCE and PCE. Only in situ soil sampling was used for characterizing TPH. The ratio of soil vapor to soil samples was 4:1. No previous soil vapor information was available. ICF Kaiser has been refining the technique of comparing results from paired soil vapor and soil samples in past and similar projects. Hydraulic probes were used instead of drilling to acquire soil vapor samples at shallow depths. This minimized waste and cost in the field program significantly.

Risk Assessment and Spatial Analysis

Human health risk analyses were conducted for the entire site, and risk factors were contoured and overlaid on maps of past usage and known soil contamination areas. Before the risk modeling could proceed, chemical and lithological data gathered in the past 7 years and those acquired by ICF Kaiser populated the environmental relational databases. Approximately 522 soil vapor probes were located in 100-square-foot spacings with additional probes in areas requiring better plume definitions. The database contains approximately 15,000 xyz-records of soil and soil gas laboratory analytical results. This information in text and graphics form, combined with site infrastructures and building outlines with attributes of "past usage," were stored as map layers, making up the GIS nucleus. Accuracy of site maps was verified with aerial photographs when available. Data types combined for computerized evaluation included known locations of contaminated soil, contaminated ground water, soil types, and site features. Composite risk maps of the above data were analyzed for data gaps at discrete depth intervals. This analysis was performed while the field program was in progress and hence gave guidance to optimize the locations of additional data points and to minimize the number of samples taken.

The MGLA/MGLM mapping module and the MSM terrain modeling module tracked the earth excavation and removal of contaminated soil. Excavation was largely part of site demolition. It also expedited the removal of TPH contaminated soils, however, because no other short-term means of remediation are available for these substances. Tracking of removed soils was essential because concurrent field activities were occurring in site demolition, data gathering, and risk modeling.

The GIS coordinated all three. Geologists and surveyors provided terrain data from daily excavation activities, which were transcribed into database formats. Maps illustrated the locations of excavated soil and removed chemicals in soil at various depths. Although TCE and PCE were of foremost concern as health risks, all compounds and some metals identified in soil were screened for unacceptable risk. Terrain modeling (mapping) as part of health risk assessment may seem unusual, but results of estimated cleanup levels and accurate locations of left-in-place contamination, mostly soils at greater depths, were critical to the cost-effectiveness and proper design of long-term remedial systems.

Characterization of Subsurface VOCs

In situ soil vapor extraction (SVE) of total volatile compounds in dense nonaqueous, liquid, gaseous, and adsorbed solid forms in the subsurface produced favorable results that have been well documented in recent years. ICF Kaiser proposed a very large-scale SVE system (see Figure 4), perhaps the largest yet, as long-term remedial technology for this former aircraft manufacturing site. The primary design problem was speculating on air flow capacity and operating time of the complex system components. The SVE system comprises three fundamental elements:

- Front-end, in situ subsurface vents (totaling 193 corings).
- Applied vacuum and air transport manifolds linking the subsurface vents to the treatment compound (distance of one-quarter mile with over 100 manifolds).
- A multivessel activated carbon treatment system.

To size the pipes, carbon vessels, and vacuum required to achieve a certain rate of VOC removal, the total mass and nature of sorption had to be known. Due to the schedule-driven nature of this project, the SVE design accounted for the time needed to accomplish the cleanup goals.

To estimate the extent and total mass of VOCs in the subsurface, soil vapor data were input to a 3-D kriging algorithm (1) to produce a concentration continuum model (see Figure 3). This solid model of predicted total VOC concentrations took the form of a uniformly spaced 3-D grid-block that completely encased the site. Cell sizes ranged from 10 to 20 cubic feet, depending on the model run, number of data clusters, density of data points in areas of clustered data, and the standard deviation of variances for estimated values in all cells. The Fortran program estimated a concentration value for each cell based on the nearest field sample(s).

The validity of such "block kriging" models can be judged by the size of the variances, smoothness, and agreement with nearby field data. Because volume is a known quantity in kriging, the total mass can be calculated by incorporating soil bulk density or porosity, both of which were less than abundant for this investigation. Renderings of kriged results in 2-D plan view contour maps, cross-sectional maps, and 3-D "vapor cloud" (see Figure 3) were included in client reports and used in regulatory presentations and public forums.

Remedial Design Layout

Final Extension of a Fully Integrated GIS

With the total mass and extent of VOCs derived from 3-D kriged results, the applied vacuum at individual vent heads and the cumulative pressure (negative) necessary to extract and transport VOC vapors from the subsurface to the treatment system can be estimated. We performed 3-D air flow analysis by use of finite difference fluid flow models and chemical transport models. The Fortran codes used to approximate compressible flow and chemical transport were AIR3D (2) and VT3D (3), respectively. Air flow simulations focused on maximizing vacuums at the shallow depths down to 20 feet to expedite remediation of contaminated soils that were

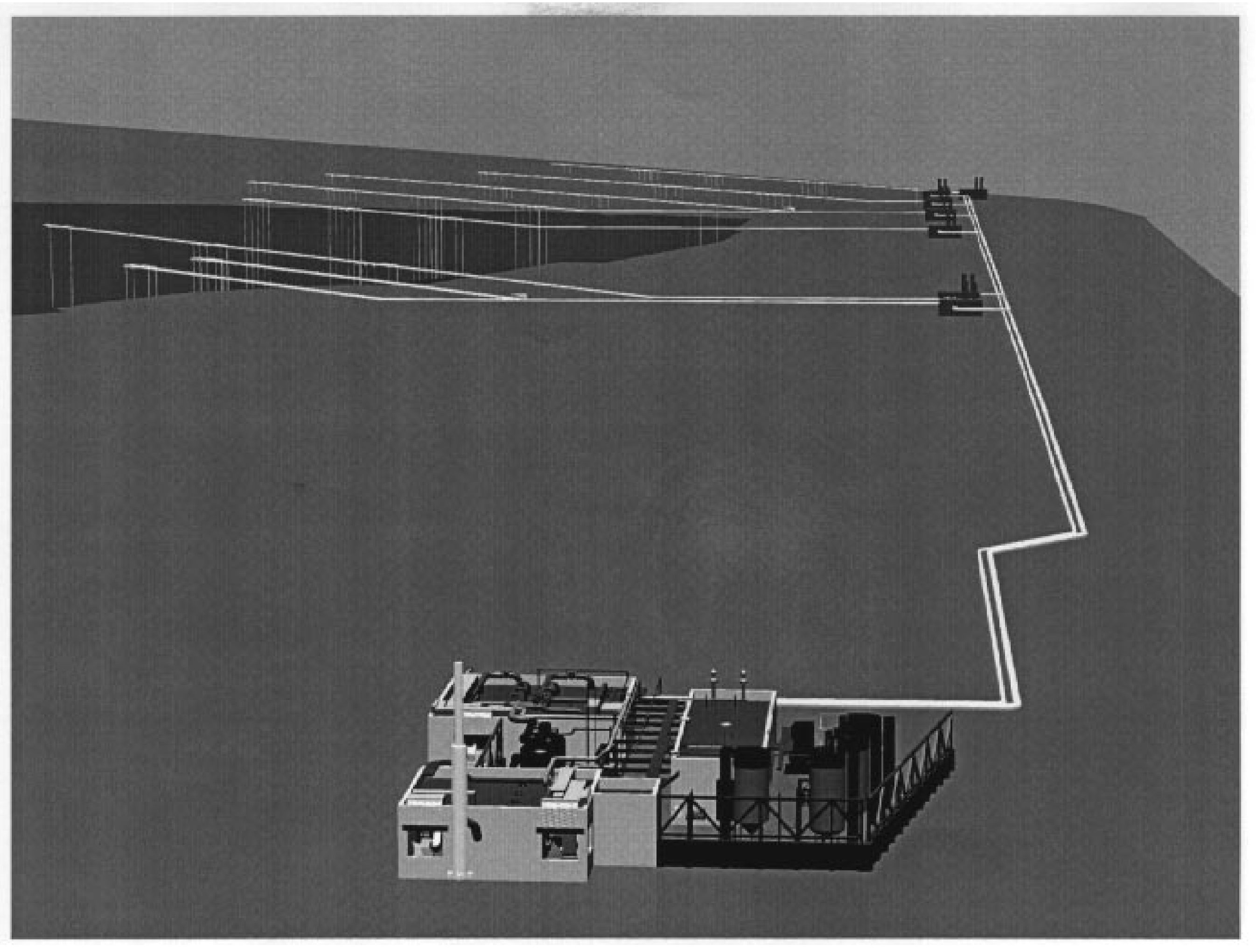


Figure 4. A rendering by the Intergraph 3-D Plant Design System of an in situ soil vapor extraction and treatment system. The cutaway section located near the upper left portion of the figure exposes some of the 193 subsurface extraction vents bottoming at 120 feet (bgs). These vents are located in a cluster for long-term extraction of the VOC vapor cloud presented in Figure 3. Vents are connected to a system of parallel airflow manifolds (right side of figure), which runs one-quarter of a mile to the treatment compound (foreground of figure).

not removed during site demolition and excavation. The lower depths were also included in each simulation. Transient mass transport models incorporate flow fields, given by flow models, and predicted cleanup times based on established HBCG cleanup goals. As VOC concentrations in an operating SVE system fall below cleanup levels in the top 20 feet, thus minimizing human risk, available vacuums thereafter will be diverted to vents at lower depths to be part of long-term extraction scenarios. Models suggested that cleanup for the top 20 feet can be accomplished within 1 year.

Numerical models prescribed vacuum levels at each vent head, which is the aboveground segment of a subsurface SVE vent. The 193 vents are connected to a system of parallel manifolds (see Figure 4) that transport vapor to the treatment system. With the vacuums known at vent heads, the size of manifolds and capacity of vacuum blowers can be determined and integrated into the overall system design. With 3-D Plant Design module as part of the Intergraph CAD/GIS, manifold layouts and treatment compound can be modeled in 3-D and easily checked for pipe routing interferences. The final layout of the SVE system was overlaid onto contour maps of total VOC concentrations to check on accuracy and completeness of vent locations and manifold layouts.

Conclusion

Maximized Visual and Analytical Responses

One goal of this project was to expedite regulatory negotiations and gain early acceptance of cleanup goals. The computerized data processing and visualization contributed generously to the rapid understanding of modeling results by expert regulators and the lay public. Likewise, the GIS facilitated the response to regulatory comments. Positive comments first came from the client's in-house review of model results and the high impact 3-D color rendering of kriged VOC distributions in the subsurface (see Figure 3).

Analytically, benefits were derived from the efficiency of electronic data access and the ability to "predict" the presence of contaminant in areas with sparse field data. The process of kriging involves the linear interpolation and extrapolation of existing data. The resultant contaminant distribution is a "conservative" model that provided the best fit with field data and validated conceptualized subsurface conditions. Further, models provided conservative estimates of mass and extent of PCE and TCE contaminations. Kriging also provided information on the uncertainty of the predicted chemical distribution, which is extremely useful for regulatory discussion and system design. The efficiency of computer models allowed investigators to perform numerous model runs with varied boundary parameters, such as cell size and search radii, in the kriging process.

Accurate mapping of excavated soil and the removal of most TPH source areas provided the incentive to critically assess the feasibility of a no-action remedial scenario for these substances at greater depths. With removal of many TPH source areas, 1-D finite difference models (4) were used to assess the mobility of TPH in NAPL and adsorbed residual phase. Specifically, models assessed the likelihood of largely residual-phase TPH affecting ground water and migrating upward to affect indoor air volumes via gaseous diffusion. Results were extremely favorable; models predicted negligible likelihood of TPH affecting ground water or indoor air volumes. Combined with GIS graphic evidence of specific areas of excavated soil and the absence of TPH sources, regulatory agencies accepted the model results, and the no-action remedial alternative for TPH was approved.

References

1. Deutsch, C.V., and A.G. Journel. 1992. Geostatistical software library and user's guide. New York, NY: Oxford University Press.
2. U.S. Department of the Interior Geological Survey. 1993. AIR3D: An adaptation of the ground-water flow code MODFLOW to simulate three-dimensional air flow in the unsaturated zone. Books and open file reports. Denver, CO.
3. Zheng, C. 1994. VT3D: Numerical model for VOC removal from unsaturated soil (draft). Bethesda, MD: S.S. Papadopulos and Associates, Inc.
4. Rosenbloom, J., P. Mock, P. Lawson, J. Brown, and H.J. Turin. 1993. Application of VLEACH to vadose zone transport of VOCs at an Arizona Superfund site. Groundwater monitoring and remediation (summer). pp. 159-169.

Polygon Development Improvement Techniques for Hazardous Waste Environmental Impact Analysis

David A. Padgett
Austin Peay State University, Clarksville, Tennessee

Introduction

Recently, concern has arisen regarding the effect Superfund sites have on surrounding communities and, specifically, the distribution of those impacts on target populations. In designing geographic information systems (GIS) applications for analyzing potential impacts of hazardous wastes or waste sites on adjacent neighborhoods, many challenges may be encountered. GIS database design requires addressing questions of time, space, and scale.

The U.S. Environmental Protection Agency (EPA) and other federal agencies have conducted studies that indicate that certain sectors of the population may be more vulnerable to exposure to toxics than others. To date, federal departments have enlisted in several GIS-based research projects that attempt to delineate "geographic hot spots" of toxic contamination. Such GIS applications at hazardous waste sites have typically used polygons to represent data from census tracts and/or municipal boundaries. In most cases, however, census tract and other boundaries do not necessarily jibe with community and neighborhood boundaries; therefore, the polygons representing characteristic data for target populations may not be consistent with the actual status of those populations.

The objective of this paper is to demonstrate GIS methods for producing, to the greatest degree possible, socioeconomically and culturally homogenous polygons for impact analysis of specific sensitive populations and/or communities. The paper presents case studies of community/neighborhood characterization problems encountered in developing polygons during previous field investigations involving lead (Pb) contamination, toxic release inventory (TRI) sites, and solid/hazardous waste sites. The paper attempts to demonstrate effective solutions and suggestions for improving polygon development, including GIS data manipulations and software applications. In addition, the paper provides geographic and groundtruthing field methods to support and enhance the accuracy of remotely obtained information. Finally, the discussion includes community and geographic hot spot analyses for potential public health impacts.

Background

In 1992, EPA established the Environmental Equity Workgroup. Its members included personnel from the Offices of Toxic Substances and Civil Rights, as well as Policy, Planning and Evaluation. The workgroup conducted an extensive study on environmental equity issues. Their report offered several recommendations for improving federal agency efforts in protecting minority and low-income populations and recognized a need for more spatial and demographic data. The final report, titled *Environmental Equity: Reducing Risk for All Communities* (1), was released in February 1992 and concluded that "there is (sic) limited data on environmental health effects by race; there are differences by race and income in potential and actual exposures to some pollutants." In response to the above findings, the workgroup offered the following recommendations (1):

> EPA should establish and maintain information which provides an objective basis for assessment of risks by income and race, commencing with developing a research and data collection plan.
>
> It (EPA) should revise its risk assessment procedures to ensure . . . better characterization of risk across population, communities or geographic areas. In some cases it may be important to know whether there are any population groups at disproportionately high risk.

The Agency for Toxic Substances and Disease Registry (ATSDR) formed a community health branch to specifically examine the potential health impact of hazardous waste sites upon people living in surrounding communities. The new branch's personnel direct ATSDR's minority health initiative, which focuses upon health threats to minority populations, including those from environmental contaminants.

In addition, EPA established the Office of Environmental Equity. The office's mission includes analyzing environmental impacts upon minority populations, providing technical assistance to disadvantaged communities, and establishing environmental initiatives at minority academic institutions (MAIs). The office serves as a clearinghouse of environmental data and information for groups and individuals involved in environmental equity activities.

In 1993, Representative John Lewis (Democrat-Georgia) introduced the Environmental Justice Act to Congress. The act requires EPA and the Department of Health and Human Services (DHHS) to establish the geographic units for determining environmental high-impact areas (EHIAs), which are the 100 geographic areas found to have the highest volumes of toxic chemical releases.

GIS Applications

GIS could potentially help address the above data and information needs of EPA. The Agency specifically acknowledges this in other recommendations (1):

> EPA could further develop its enforcement prioritization policy to target high-risk populations. Under this scheme, the most exposed and highly susceptible populations in each region would be targeted for enforcement actions. Geographic Information System technology could be used to identify high-risk populations.

Several recent environmental studies have employed computer applications and spatial data. Goldman (2) used GIS in a major study that graphically displayed counties having high percentages of African-Americans, hazardous wastes, and diseases. Mohai and Bryant (3) applied a linear regression model to show a positive correlation between increasing proportions of minority populations and decreasing distances from hazardous waste sites in Detroit. Lavalle and Coyle (4) conducted an extensive analysis of computer databases that hold hazardous waste law enforcement information for the past 10 years. They found inequity in enforcement and remedial actions in white communities versus nonwhite communities.

EPA has enlisted GIS for community environmental impact projects at Regions II and III. EPA's Office of Health Research (OHR) is investigating methods for linking demographic data with TRI information to evaluate the relationship between levels of hazardous waste releases and exposure risks in minority communities. EPA has also developed the TRI "risk screening" process, which employs TRI, U.S. Census data, and GIS to identify TRI releases that may pose significant risk to human health or the environment (5).

Both EPA (6) and the North Carolina Department of Environment, Health, and Natural Resources (7) have recently completed GIS-based environmental investigations. The EPA study involved GIS analyses of TRI chemical releases in the southeastern United States. The report included numerous GIS-produced maps that show locations where TRI releases may be affecting densely populated areas and sensitive ecosystems. The North Carolina study applied GIS in searching for sources of lead-poisoning in children. Findings indicated a positive spatial correlation between high lead-contamination risk communities and those having certain socioeconomic characteristics, such as low income, above-average African-American population percentages, and above-average percentages of residents receiving public assistance.

The ATSDR recently implemented a study using GIS to evaluate and analyze the demographic characteristics of populations near National Priorities List (NPL) sites. According to the ATSDR, "As a result of our pilot tests, we have determined GIS to be the best methodology for identifying potentially impacted minority populations" (8).

Limitations of GIS

While GIS may be a viable tool for investigating environmental inequity, it is not an absolute solution. Issues involving hazardous waste impact assessments tend to be very complex without the added dimension of racism or discrimination. Efforts to determine a causal relationship between the presence of minority communities and environmental hazards must consider the questions of time, scale, and place. Unfortunately, in many instances, GIS applications may be unable to adequately illustrate these three pertinent issues resulting in skewed or altogether incorrect conclusions.

With respect to scale, among the immediate concerns when applying GIS is selecting appropriate sizes for polygons. As indicated above, EPA is in the process of determining the scale for EHIAs. A polygon may be a county or a census tract. Figure 1 illustrates problems

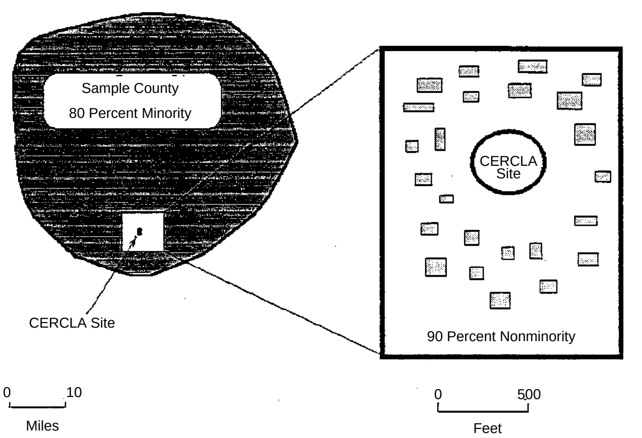


Figure 1. Problems of scale in GIS polygon design.

of scale associated with polygon size selection. At the county-size scale, a case of environmental inequity apparently exists with the presence of a Superfund site in the sample county because 80 percent of the county's population belongs to a minority ethnic group. A closer look, however, reveals that the population residing in the immediate vicinity of the waste site is predominantly nonminority.

Useful as GIS may be, its output in some cases may display static conditions without considering human movements over time. Figure 2 displays a common situation associated with the "filtering" phenomenon, in which a nonminority population moves out of an area while increasing numbers of an ethnic minority group moves into it. The figure shows that in 1950, a nonminority community surrounded a TRI site (i.e., an active industrial site releasing toxic substances). By 1990, the demographics of the neighborhood had changed along with the status of the TRI site, which is now an abandoned Superfund site. A GIS database probably would contain only information on the community from 1990. Such an instance could suggest that some form of environmental injustice exists given the presence of the Superfund site within the minority community. Accounting for the dynamics of time and human movement, however, would show that the waste site preceded the minority population and that, in actuality, the minority community moved toward the site. This conflicts with the prior notion that unsavory forces placed the site in the minority community.

Figure 3, a schematic of polygons used in an investigation into sources of lead-poisoning in children, also displays the limitations of GIS with respect to time and human dynamics, but at a lesser time interval. The study area is divided into low- and high-risk areas for lead contamination. The locations at which children with lead poisoning were found, however, do not correspond with the areal risk factors. In this case, the GIS is limited in its ability to follow human movements on a daily basis. For instance, a parent with a child who exhibits unhealthy blood-lead concentrations may report the child's home address as someplace within the low-risk polygon. The child may attend school in the high-risk area, however. The children's points of contact with lead may not necessarily correspond with their home addresses, resulting in an inaccurate graphic display.

With respect to place, GIS may be limited in its ability to determine the specific borders of a socioeconomically and demographically homogenous human population. Tosta (9) and Coombes et al. (10) discuss the dilemmas associated with neighborhood boundary delineation in GIS applications. Figure 4 displays a schematic of a census tract. The GIS database may list the tract's per

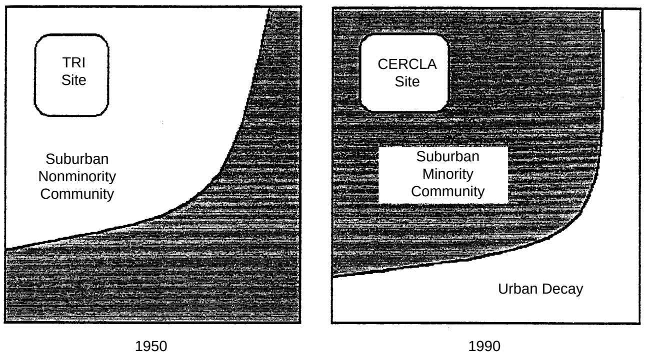


Figure 2. Example of changing community demographics with time near a hazardous waste facility.

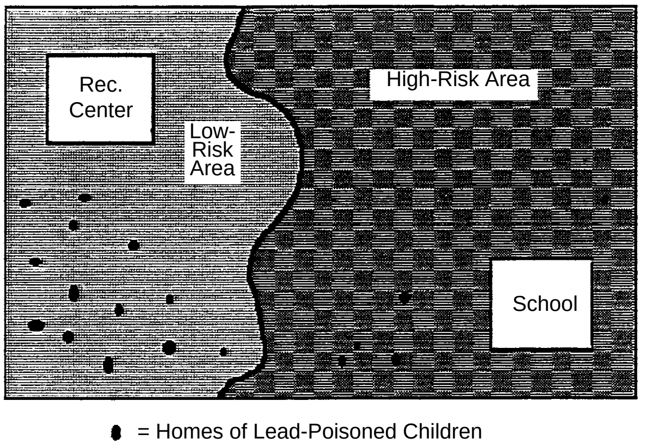
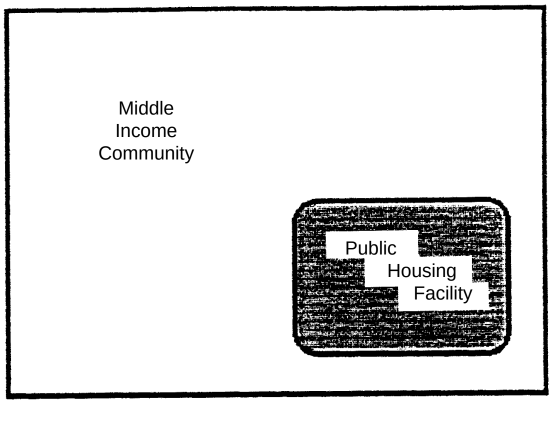


Figure 3. Lack of correspondence between locations of lead-poisoned children and high-contamination risk areas due to daily dynamics of human movements.



Census Tract Schematic

Figure 4. Example of significant neighborhood-type variation within a single polygon, possibly resulting in skewed socioeconomic data.

capita income as relatively low and may list the tract as a low-income neighborhood. Further investigation may find, however, that two very different socioeconomic communities exist within the tract, one middle-class and the other a public housing facility. Frequently, middle-income, African-American communities have low-income housing projects built adjacent to them. With respect to the polygon in Figure 4, if an investigator wanted to research health impacts of toxic wastes for low-income households, using this polygon and others like it would inaccurately depict communities within them.

An additional problem in community definition is determining exactly what defines a minority community. The most common indicator for discerning a minority community would be the existence of a clear majority of some minority group as in Polygon A of Figure 5, or where the minority group makes up more than 50 percent of the population as in Polygon B. In some instances, however, communities have received minority status without the presence of the conditions in Polygons A and B.

Previous investigations show a number of measures used to identify minority communities and census tracts. Greenpeace conducted a 1990 environmental justice study that determined a community's status as minority based upon the relationship between a community's percentage of minority population and the selected minority group's national percentage (11).

Polygon C in Figure 5 depicts Greenpeace's minority community definition. Taking the target ethnic group in Polygon C as African-Americans and the hypothetical extent of Polygon C as the United States (African-Americans make up approximately 12 percent of the total U.S. population), one may determine that Subpolygon C is a minority polygon or community because its minority percentage is over twice that of the national percentage or extent of the large population in Polygon C. The condition that Polygon C illustrates is also evident in a study

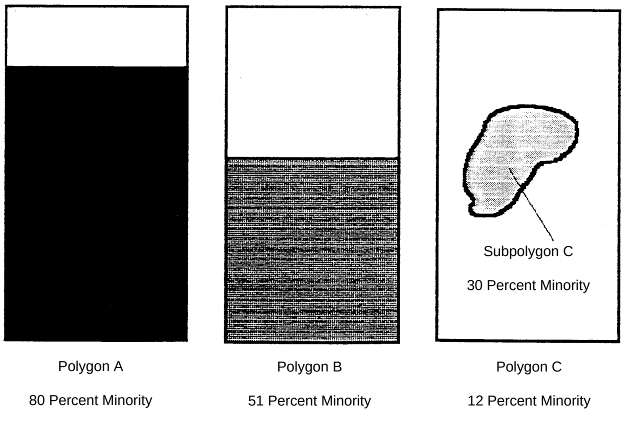


Figure 5. Problems in defining minority polygons.

by Mohai and Bryant (3). The authors claim that environmental inequity exists in Detroit where they found that on average, within a 1-mile radius of the city's hazardous waste facilities, 48 percent of the population is nonwhite.

Solutions With GIS and Supporting Technological Methods

From the above discussion, GIS may appear very limited for use in environmental community impact investigations, but GIS can actually be an extremely effective tool if employed with appropriate supporting technology.

Preliminary Groundtruthing

To design GIS databases that reflect the true nature of target groups and human dynamics, preliminary groundtruthing may be necessary. In some cases, investigators make gross interpretations of suspected environmental inequity without actually visiting the study area. Without groundtruthing prior to final database development, questions of time, scale, and human dynamics may be left unanswered. The integrity of databases produced this way and the antecedent conclusions based upon them may fall into question. Thus, because the nature of environmental and human health impact studies is complex, some on-the-ground work should precede or at least accompany database construction efforts.

Cause-Effect Analyses

Epidemiological studies are increasingly employing GIS. This use is important with respect to environmental investigations because in many cases, proof of a correlation between a waste site and community health problems may be necessary. Croner et al. (12) describe statistics-supported GIS applications for linking "cancer hot spots" with pollution sources. Without conclusive evidence that waste sites and other environmental hazards negatively affect health in socioeconomically disadvantaged populations, claims of environmental injustice may be difficult if not impossible to prove.

In historical analyses of waste facility sitings, GIS may be useful, along with the support of gravity models, in investigating whether the sitings followed the prescribed logic for such siting decisions. Noble (13) wrote that the costliest aspect of waste facilities management is transportation; therefore, siting decisions should favor locations in closest proximity to a selected community's centroid of waste production. Using a GIS gravity model with data for household wastes produced within a given locality, the center of gravity of the volumes of wastes produced could be located. Historical analyses of past siting decisions may find that past siting decisions defied logic. Instead of finding waste sites placed in environmentally safe locations as close as possible to areas of greatest refuse generation, analyses may find instead that sites have been placed farther away in disadvantaged communities. Both the taxpaying and potentially affected residents would pay for such unsavory siting practices.

Conclusion

The potential for technological applications, including GIS, in this arena is great. Increased involvement by technologically trained environmental professionals is imminent. Their future involvement, however, must focus on the scientific soundness of investigative methods, data integrity, and the equitable participation of potentially affected citizens in any subsequent decision-making processes.

References

1. U.S. EPA. 1992. Environmental equity: Reducing the risk for all communities. EPA/230/DR-92/002. Washington, DC.
2. Goldman, B. 1991. The truth about where you live. New York, NY: Times Books.
3. Mohai, P., and B. Bryant. 1992. Environmental racism: Reviewing the evidence. In: Bryant, B., and P. Mohai, eds. Race and the incidence of environmental hazards: A time for discourse. Boulder, CO: Westview Press.
4. Lavalle, M., and M. Coyle. 1992. Unequal protection: The racial divide in environmental law. The Environmental Professional 153:S1-S12.
5. U.S. EPA. 1990. Toxics in the community: National and local perspectives. EPA/560/4-90/017. Washington, DC.
6. Stockwell, J.R., J.W. Sorensen, J.W. Eckert, Jr., and E.M. Carreras. 1993. The U.S. EPA geographic information system for mapping environmental releases of toxic chemical release inventory TRI chemicals. Risk Analysis 132:155-164.
7. North Carolina Department of Environment, Health, and Natural Resources. 1993. Lead database evaluation and GIS modeling project. PR-242592. Raleigh, NC: National Institute for Environmental Health Sciences.
8. Harris, C.H., and R.C. Williams. 1992. Research directions: The Public Health Service looks at hazards to minorities. EPA J. 181:40-41.
9. Tosta, N. 1993. Sensing space. Geo. Info. Sys. 39:26-32.
10. Coombes, M., S. Openshaw, C. Wong, and S. Raybould. 1993. Community boundary definition: A GIS design specification. Regional Studies 3:280-86.
11. Bullard, R.D. 1993. Environmental equity: Examining the evidence of environmental racism. Land Use Forum (Winter). pp. 6-11.
12. Croner, C.M., L.W. Pickle, D.R. Wolf, and A.A. White. 1992. A GIS approach to hypothesis generation in epidemiology. In: American Society for Photogrammetry and Remote Sensing/American Congress on Surveying and Mapping/RT 92, Washington, DC (August). Technical papers, Vol. 3: GIS and cartography. pp. 275-283.
13. Noble, G. 1992. Siting landfills and LULUs. Lancaster, PA: Technomic Publishing Company, Inc.

Comparing Experiences in the British and U.S Virgin Islands in Implementing GIS for Environmental Problem-Solving

Louis Potter
Office of the Chief Minister of the Government, British Virgin Islands

Bruce Potter
Island Resources Foundation, St. Thomas, U.S. Virgin Islands

The British and U.S. Virgin Islands: Comparisons and Contrasts

British Virgin Islands

Three miles to the north and east of the U.S. Virgin Islands (USVI) lie the British Virgin Islands (BVI), a group of 36 islands and cays with a total area of 60 square miles. The four largest are Tortola (24 square miles), Anegada (14 square miles), Virgin Gorda (8.5 square miles), and Jost Van Dyke (4.5 square miles).

Geologically, the BVI belong to the Greater Antilles, and like the USVI and Puerto Rico, rise from the Virgin Banks (or Puerto Rican Shelf). Rocks of the BVI, except Anegada, consist of thick, steeply inclined, metamorphosed volcanic and sedimentary stratified series of Cretaceous age, with dioritic and pegamitic intrusions. Anegada is a 30-foot-high emergent coral limestone platform, apparently from the Pleistocene age.

The BVI are a crown colony of the United Kingdom (UK) with a total population in 1991 of 16,108. Most of the population resides on Tortola (13,225 inhabitants). Between 1980 and 1991, population increased 46.6 percent. The BVI have internal self-government with an elected council headed by a chief minister. The UK appoints a governor to represent the queen and to manage defense, internal security, external affairs, civil service, and court administration.

The BVI economy is based mainly on tourism and servicing international business. Sailing and diving are pivotal features in BVI tourism. The average number of tourists per capita in the BVI is 221, compared with 119 for the USVI (1). The BVI 1990 per capita income was recorded at $10,125. Prior to the 1970s, the BVI economy was based on subsistence agriculture and remittances from British Virgin Islanders who worked in the USVI.

U.S. Virgin Islands

The USVI are an unincorporated territory of the United States, purchased from Denmark in 1917. The total population in 1990 was about 101,000, divided among St. Thomas (48,000), St. Croix (50,000), and St. John (3,500). The tourism-dominated economy of the Virgin Islands generated a per capita Gross Territorial Product in 1990 of approximately $11,000—the highest in the Caribbean (2).

Geographically, geologically, and topographically, St. Thomas (28 square miles) and St. John (20 square miles) are similar; they are both largely volcanic, have deeply indented coastlines, and lie on the Puerto Rican Shelf. St. Thomas and St. John are close to the BVI. St. Croix is a relatively large (84 square miles) and mostly limestone island that lies on its own submarine ridge, which rises more than 4,000 feet from the bottom of the Caribbean Sea. St. Thomas and St. John are about 5 miles apart, and St. Croix is 48 miles south of them.

During the height of tourism development (from the late 1950s through the mid-1970s), the USVI experienced average annual compound population growth rates of over 6 percent, as well as a doubling in real incomes. This unprecedented paroxysm of growth is still being assimilated by a population that differs greatly from the 30,000 people who lived in a predominantly agricultural USVI in 1950. In 1990, the USVI received a daily average of 37 visitors per square kilometer. This compares with a visitor load of 23 visitors per day per square kilometer in the BVI, which is also a high-density tourist destination (1).

Background to GIS Implementation Activities

British Virgin Islands

The idea of geographic information systems (GIS) applications in the BVI first arose with a presentation about a proposed project for St. Lucia, made by Dr. Jan Vermeiren of the Organization of American States at the Caribbean Conference of Planners in Kingston, Jamaica, in 1984 (3). The Town and Country Planning Department recognized that it could use GISs analytical and display properties to make presentations to the chief minister and the BVI Executive Council. This proposal fell on fertile ground, given a relatively long-held tradition of support for the Town and Country Planning Department by the United Nations Development Programme (UNDP) and the British Development Division, dating back to the early 1970s.

Subsequent to this inspiration, the Town and Country Planning Department requested budget authority to develop a land use database. This database would include buildings, property boundaries, and constructed and natural features of importance. The Finance Department hoped this project would help combat growing competition to the postal services by independent package delivery services. They renamed the project the National Addressing System, and the legislature provided $200,000 to provide a physical address for each property in the territory.

The Town and Country Planning Department conducted a pilot project, focusing on Road Town, the capital. The pilot project demonstrated that the hard copy land ownership or cadastral maps that the Survey Department was then using were inadequate for accurately accounting for properties, even in the BVI's most developed urban areas. Therefore, the Town and Country Planning Department expanded the scope of the National Addressing System project to identify options for increasing the accuracy of property ownership records, including maps.

Because the Town and Country Planning Department had little experience with computerized land information systems, departmental managers sought support from the UNDP office in Barbados. The UNDP had previously assisted with the department's development control applications database. This mode of operation, in which the BVI purchase services provided through the UNDP, which acts as a "vetting agent" for consultants and other technical assistance, continues to this day. An expert from the United Nations Community and Housing Services (UNCHS) Nairobi office provided the first such consultation by exploring how the BVI might implement a GIS. Over the next few years, three other experts provided input, which the Town and Country Planning Department gradually integrated into a picture of how to use a GIS within the technical and financial limitations of a small island government.

In the meantime, external conditions were improving the chances for the success of the BVI program. British Virgin Islanders were receiving formal and informal training in computer applications in general, and specifically in AutoCAD drafting systems, which increasing numbers of local architects and engineers are using. In addition, the power of rugged microcomputer systems that could withstand the harsh operating conditions of the Virgin Islands was also improving, and local dealers were increasing their skills in support of such systems.

U.S. Virgin Islands

In 1988, a proposal to develop a locally supported GIS was being discussed in detail in the USVI (4). This project resulted in a formal application from the government of the USVI for financial assistance from technical assistance funds provided by the Office of Territorial and Insular Affairs of the U.S. Department of the Interior. The grant was awarded in March 1991. The proposed project combined existing information from several USVI government agencies to produce GIS overlays, as shown in Table 1 (5).

In addition, according to the grant application, (which was written by the Virgin Islands government and may not have represented U.S. Geological Survey's [USGS's] intentions) the National Mapping Division (NMD) of the USGS agreed to digitize the eight USGS 1:24,000-scale quad sheets ("quadrangles") that cover the USVI, including the following categories:

- Roads and trails
  - Power transmission lines
- Hydrography
  - Stream networks
  - Shorelines
  - Wetlands
  - Mangroves
  - Reefs

Table 1. GIS Overlays

Agency	GIS Overlays
DPNR	Zoning Flood plain Subdivision
WAPA	Water distribution Aquifer profiles Electrical distribution
VITEMA and emergency services	Critical routes Critical facilities
DPW	Sewer line network Transportation Flood plain

DPNR = Department of Planning and Natural Resources
WAPA = [Virgin Islands] Water and Power Authority
VITEMA = Virgin Islands Emergency Management Agency
DPW = Department of Public Works

- Topographic contours
- Political boundaries

(These categories were subsequently adjusted based on discussion between the USGS, WAPA, and government agencies. Documentation for these new coverages is not available, but the general idea holds, although no VITEMA data were used, and the USGS did not digitize the WAPA power system.)

The underlying notion was that the whole would be bigger than the parts—each organization would bring its own corporate data and maps so all could share in the digitized product.

Five objectives were identified for the project:

- Provide the USGS in St. Thomas with a complete microcomputer (GIS) workstation.
- Develop a digitized database from USGS topographic maps.
- Contract for the digitizing.
- Acquire digital data to load into GIS.
- Enter data not in digital format.

The proposal required a 50-percent local cost contribution to the project, including a matching suite of hardware for data maintenance, backup, and analysis. The application stated that itemized costs of $305,000 “will be provided by the Virgin Islands Water and Power Authority (WAPA) and the Office of Territorial and International Affairs.”

The project application is ambiguous about the nature of the hardware and operating systems that the GIS requires. A list of hardware for USGS use refers to a “microcomputer workstation.” Later references to “ESRI ARC/INFO workstation software,” however, indicate to sophisticated users that these applications require UNIX workstations and UNIX language operating systems. In general, UNIX is not used or supported in the Virgin Islands. Some Virgin Island officials associated with the GIS project feel they were not fully informed about the GIS operating environment and the long-term support costs to which they were committing.

A frustration for USGS in the project stems from the somewhat limited role the agency has had in providing high-quality data conversion services (digitizing). One USGS staff member explained that unfortunately the agency’s mandate only extends to providing data; the agency “can’t get involved in applications.”

The initial project proposal also was unclear about ownership of the digitized data. The proposal spoke in general terms: “All available digital data and attributes will be complete, accurate, and up-to-date at the end of the project, and will be available for use and transfer to the Department of Planning and Natural Resources (DPNR) (5).”

Given the conditions and environmental constraints discussed, a number of specific differences developed between the GIS implementation processes of the BVI and the USVI.

Initial System Planning Activities

British Virgin Islands

Although the BVI had no formal systems plan, consultants worked with the Town and Country Planning Department on four different occasions to provide insight into some aspect of GIS applications. Sometimes, the benefits from the consultations were neither the type nor the quality the department originally expected. In general, however, each provided some additional perspective on the possible benefits and perils of implementing a GIS.

U.S. Virgin Islands

The USVI apparently made few initial system planning efforts, although the U.S. Department of the Interior and USGS have wide experience in the territory. (An informed source claims a grant was made, possible by EPA, for a preceding $50,000 GIS project, but no mention of this has been found in the materials available for this article, either as a proposal, or in terms of specific products.) Possibly, this very familiarity led to a series of unexamined assumptions and diminished communications about the exact terms of the assistance and services that the USGS exchanged with several agencies of the Virgin Island government.

One indicator of the lack of system planning activities in the Virgin Islands is a proposal that the Virgin Islands Emergency Management Agency (VITEMA) circulated for a “Geographical Information Systems: Technical Operators Meeting.” This proposal, from an agency that has always been one of the most important participants in GIS activities, called for a “technical working group” to examine existing database management systems in the territory to develop a planning strategy for implementing GIS.[1] This proposal was dated September 28, 1992, 10 days before the USGS announced a demonstration of the completed products of Phase I of the “comprehensive geographic information system being developed for the USVI (6).”

[1] Ward, R.G. 1992. Geographical information systems: Technical operators meeting. Memorandum to Cyrille Singleton. VITEMA, St. Thomas, U.S. Virgin Islands (September).

Software Selected and Rationale

British Virgin Islands

In part because of the extended timeframe for the BVI planning and initiation of the GIS system, the Town and Country Planning Department never committed to a specific system configuration until the last stages of the planning process.

This process of "creative procrastination" had three synergistic results:

- PC power increased (and prices decreased) to the point where reasonably priced systems could perform many of the compute- and data-intensive operations demanded for graphics software mapping.
- ARC/INFO released the ARC/CAD version of its GIS software, which worked on PCs within the well-known AutoCAD drafting software. Architect and engineering offices in the BVI already used computer-aided design (CAD) software, so upgrading to include GIS functionality was relatively easy.
- The fourth GIS consultant to work in the BVI was experienced in implementing systems in the Caribbean and had special knowledge and access to early versions of both ARC/CAD and Version 1.0 of ARC/VIEW. These two tools, based on a last minute proposal by International Development Advisory Services (IDAS) of Miami, Florida, a private GIS support contractor, became the basis for the BVI GIS.

U.S. Virgin Islands

Workstation ARC/INFO 6.1 was selected as the basic software for the USVI GIS project because it is the USGS standard software. In the environment surrounding the USVI project, this seemed to be a sufficient explanation, although there may have been other reasons. Because of this decision, however, WAPA and the DPNR lack any means of updating map or attribute data files. Outside providers, such as the USGS's NMD, must perform that service. The USGS has noted that the only reason WAPA and DPNR lack these capabilities is because they (WAPA and DPNR) failed to provide the matching suite of hardware and software specified in the grant application.

The digitized water supply system offers an example of the extra costs that such a condition creates. The USGS built an ARC/INFO coverage by converting mapped data from AutoCAD source files, which they then linked to detailed attribute information about each element (e.g., pipe, valve, elbow) in the system. The USGS then used ARC/INFO network software, purchased with project funds, to build a network model that analyzes and displays the operation of the entire water distribution system—but no software in the Virgin Islands can run the network model.

Hardware Platforms

British Virgin Islands

The BVI GIS was originally installed on a Compaq 486-50, with a 21-inch screen. The Town and Country Planning Department soon learned, however, that data input would be more efficient if two or three smaller machines split the work, with the Compaq available for analysis and data quality checking. The department upgraded its existing office computers to handle the data entry. Users already feel the need for networked applications to share data more quickly. Plans for GIS expansion to other offices, such as the Electricity Corporation, increase the pressure for an extended local area network.

U.S. Virgin Islands

The system that the USGS used to build the USVI GIS database was a Data General UNIX workstation with one large digitizing tablet and one pen plotter. No matching or comparable hardware are installed anywhere in the USVI, as the original project proposal had foreseen. Observers tend to agree, however, that the failure to provide a specific hardware configuration is less significant than the lack of committed, senior, full-time technical staff. This staff is required to operate the level of GIS facility that the USGS envisioned.

Base Map Priorities and Layers Constructed

British Virgin Islands

Building a map database is proving to be a long process for the Town and Country Planning Department. This is complicated by the failure of a key digitizing contractor in Texas to provide property lines in a format conducive to constructing accurate property polygons. Operators in the Town and Country Planning Department have increased their data entry efficiencies, however, and most properties on the most densely inhabited islands have now been digitized.

Producing demonstration data displays accounts for a significant part of the cost of developing databases for the early phases of the BVI GIS implementation. These demonstrations aim to illustrate possible new application areas for other agencies and departments of the BVI that are interested in cooperating and sharing costs of additional system development. For example, the Electricity Corporation and the National Disaster Preparedness Agency need to map emergency services.

Converting the data (i.e., digitizing) in house in the BVI has produced costs and benefits. The costs revolve around the steep learning curve for data entry procedures and the constant distractions of responding quickly to “outsiders” who may be important long-term supporters of the GIS. The benefits include increasing staff skills and the ability to build constituencies for the program by promptly responding to real needs.

Coverage priorities for the BVI GIS include a national addressing system, completion of the territorial land use mapping, and accurate cadastral mapping (which has major environmental planning and management implications).

U.S. Virgin Islands

The USGS produced 44 coverages for the Virgin Islands from a variety of sources. Table 2 shows the major coverages and scales, by island.

Table 2. Major Coverages and Scales for the USVI

St. Croix	St. John	St. Thomas
STC water distribution 1:2,400	STJ water distribution 1:2,400	STT water distribution 1:2,400
STC roads 1:2,400	STJ roads 1:2,400	STT roads 1:2,400
STC building footprints 1:2,400	STJ building footprints 1:2,400	STT building footprints 1:2,400
STC shorelines 1:2,400	STJ shorelines 1:2,400	STT shorelines 1:2,400
STC DLG boundaries 1:24,000	STJ DLG boundaries 1:24,000	STT DLG boundaries 1:24,000
STC DLG roads 1:24,000	STC DLG roads 1:24,000	STT DLG roads 1:24,000
STC DLG hydrography 1:24,000	STJ DLG hydrography 1:24,000	STT DLG hydrography 1:24,000
STC DLG hypsography 1:24,000	STJ DLG hypsography 1:24,000	STT DLG hypsography 1:24,000

In addition, the following National Park Service data were converted and added to the data set but were not produced by the USGS:

- STJ NPS boundaries, 1:24,000
- STJ NPS roads, 1:24,000
- STJ NPS hydrography, 1:24,000
- STJ NPS hypsography, 1:24,000
- STJ NPS benthic communities, 1:24,000
- STJ NPS historical sites, 1:24,000
- STJ NPS vegetative cover, 1:24,000

Mapping for St. Thomas, St. John, and St. Croix identified a total of 10 coverages for each island. They are based on information from the USGS (“quad sheets” specifically demarking political boundaries, shorelines and streams, topography, and roads) and higher precision WAPA mapping, which derives from 1986 aerial photogrammetry, including left and right road boundaries, building footprints, shorelines, and water supply system data. The WAPA data are at 1:2,400 scale, an order of magnitude more precise than the USGS base map. St. John mapping consists of 10 added layers based on data that the Virgin Islands National Park (VINP) provided.

The original USVI project proposal referred to a two-phase process of database development, shown below (5):

Phase	Tasks
I. Base system development	Water distribution network Power distribution network Flood plain maps
II. Agency extensions (i.e., by USVI agencies)	Land use maps Transportation networks Emergency facility networks Tax parcel/land value

The USGS announced that Phase I was completed in October 1992 (6). Supposedly, the contents of these two phases were subsequently adjusted to reflect a different range of coverages, but the notion of a “Phase II” in which local agencies would assume more operating responsibilities was retained.

Ownership of, access to, and terms that govern the use of this digital data are confused. The USGS says it is unable to provide an authoritative catalog of the coverages because “one has not been produced.” WAPA says it has several diskettes of data in the safe but no equipment to manipulate them. The VINP has learned that it can use its own data as well as WAPA data converted and attributed by the USGS, but the park does not possess or use USGS digital line graph (DLG) data. To personnel in USVI agencies, USGS statements have clouded the question of access to the GIS information. For example, one such statement announced that the USGS Water Management Division cannot make the digitized data available to Virgin Island government agencies.

The DPNR apparently has no means of making direct use of the digital data. First, DPNR has no hardware or software that can use the data. Secondly, it has no operators who can build the GIS systems to actually apply the data to decision-making needs. The department is said to be preparing a new GIS proposal for training, hardware, and software for a new GIS system. According to unconfirmed rumors, this system will be based on MapGrafix, a Macintosh mapping system.

In operational terms, the data seem to belong to WAPA, which contributed major financial support and map resources to the project. WAPA has been helpful in providing copies of the digital data to other groups and agencies.

Environmental Problem-Solving in Local Decision-Making

British Virgin Islands

In the BVI, the first priority of the GIS facilities is to extend the National Addressing System and to improve the property ownership system. This will both improve postal services, as originally proposed, and provide better information for important revenue and financial analyses. Land use mapping and environmental impact assessments are important second priorities for GIS applications. Other features already developed for interim studies and analyses include mapping of significant coastal and wildlife features and environmentally sensitive areas from the Anegada Development Plan and mapping of important submarine habitats adjacent to Virgin Gorda.

The BVI have concentrated on developing GIS applications to address strategically important issues in the territory. Marine and coastal resources are vitally important to the BVI economy. They embody historical and cultural values, as well as maintain a high-quality environment to support charter yacht-based tourism, which is integral to the BVI economy. The Conservation and Fisheries Department is working with the Town and Country Planning Department to convert the country's coastal atlas to digital form (see section on Principal Users), as has been done on a demonstration basis for the Anegada and Virgin Gorda mapping.

U.S. Virgin Islands

In the USVI, environmental decision-making generally follows an adversarial, rather than a problem-solving format. A combination of historical and cultural factors have created the general assumption that the development process creates winners and losers. In this environment, information becomes an important tactical weapon, making it difficult to gather support for activities or programs that aim to make information more widely accessible. Technology is more acceptable, and more likely to receive leadership support, if it is justified on technical, less "political" terms.

The USGS team made a presentation on the Virgin Islands GIS in October 1992, after just completing the digital coverages for Phase I of the GIS project. The presentation emphasized that the GIS is intended to provide decision-makers with easily accessible spatial information (6).

Originally, the GIS was expected to benefit primarily the territory's three coastal zone commissions in their assessment of environmental effects of major development proposals. The ground-water protection program of the Division of Environmental Protection of the DPNR is using GIS analyses produced by the Water Resources Division of the USGS, employing the USVI GIS coverages with added data (e.g., wells) that the Water Resources Division is digitizing.

GIS Support Factors

According to the GIS support contractor for the BVI, a successful GIS requires three key support elements:

- GIS policy leadership
- GIS technical leadership
- Competent outside expert assistance

The following summarizes the comparative experience of the two programs for these three key implementation support factors:

Support Factors	BVI	USVI
GIS policy	Town and Country Planning Department director led project from chief minister's office	No GIS manager in government
GIS technician	BVI technician trained in the United States	No GIS specialist in government (draftsman at WAPA)
Outside support	UNDP and IDAS	USGS technical support and U.S. Department of the Interior financial support

Principal Users: Planned and Actual

British Virgin Islands

The BVI Department of Finance is the first user of data products from the GIS, based on the initial funding for the addressing system. This system is based on detailed parcel maps of the BVI so that the effective base map resolution of the BVI system is 1:2,500. This is considerably finer than the 1:24,000 scale of the USVI maps.

The Town and Country Planning Department, however, is working to recruit other users to the system, including:

- Public Works.
- Water and Sewerage Department for a systemwide map (which may eventually spin off as a separate system, given this department's long-term interest in engineering-quality facilities management information).
- Conservation and Fisheries.
- British Overseas Development Administration, which funded a coastal atlas for the BVI (7). At the suggestion of the Town and Country Planning Department, this mapping was developed in ARC/INFO. A proposal has been made to convert the coastal atlas to digital form for natural resource management applications, with a demonstration already developed showing the distribution of sensitive marine communities around Virgin Gorda.

In addition to these uses, the GIS group is starting to experiment with the use of remote sensing products in GIS production, which would encourage the use of GIS for natural resource change detection.

The Town and Country Planning Department's GIS specialist, Mikey Farara, is being reassigned to provide networking support (including GIS distribution over the network) for several government agencies. Meanwhile, the GIS operation is adding a cartographer to assist in tailoring GIS products to users' needs.

As enthusiasm for the GIS has blossomed in the BVI, managing for realistic expectations and stressing the investment costs that participating agencies can expect have become problems for the Town and Country Planning Department.

U.S. Virgin Islands

Complex evaluation issues face the USVI's three coastal zone commissions (one on each island). Therefore, land use planning in general and coastal zone permitting specifically were assumed to be important first users of the GIS. The DNPR, however, had no process to prepare the Division of Comprehensive and Coastal Zone Planning to implement this system. In addition, the scale of permitting decisions may be too fine for the GIS base map. (See discussion of scale below.)

WAPA is not using the GIS data. One senior manager characterized their experience with the GIS project as "paying a lot of money for a diskette of data that we keep locked in the safe."

The VINP (part of the U.S. National Park Service) and Biosphere Reserve have purchased a PC-based GIS system and employed an analyst to implement it for the park and adjacent areas on St. John and the surrounding seas. With this system, they plan to enter the USGS-developed data into the database. In addition, the Virgin Islands Resource Management Cooperative (a collaboration of research and resource management organizations) makes the VINP GIS data and analytical capabilities available to members, including government members.

At this time, the only major Virgin Islands government user of the GIS is the ground-water protection group of the DPNR's Division of Environmental Protection. Because they lack equipment or software to manipulate the GIS data already available, they use the Water Resources Division of USGS as a GIS contractor. This arrangement has two problems:

- *High costs:* Although the USGS "owns" the existing digital data, and processing private contracts would be complex, DPNR believes it could get similar services at cheaper prices from other vendors.
- *Inappropriate scale:* Environmental management processes in the Virgin Islands (and in most other small island states) require knowledge of property ownership, implying maximum map scales of 1:5,000 to 1:10,000. The USGS quad sheet scale of 1:24,000 is too coarse for many management purposes. Costs of remapping areas of concern at the higher resolution are high, and the problems of maintaining multiple map resolutions and sources are not trivial.

What GIS Can Do

Joseph Berry has proposed seven basic categories of "What GIS Can Do for You" (8). These applications can be related to the GIS products and proposals for the BVI and USVI, with special attention given to natural resource and environmental issues. Table 3 shows what coverages that have been or are being developed for the two systems can do.

Table 3 illustrates two contrasting issues separating the two jurisdictions. The USVI have the data available to perform a number of relatively complex analytical processes, especially in St. John. They have no capability to actually execute any such studies, however. The BVI, on the other hand, have proposed and often developed pilot or demonstration applications for several GIS uses but still need to develop the data resources to support these on a territorywide basis.

Lessons Learned

The comparative experience of these two very distinct GIS programs reinforces three basic lessons of information system design and implementation:

- *Plan, don't assume:* The prolonged, sometimes repetitious, planning process that evolved in the BVI

Table 3. Coverage Capabilities

Questions:	Analytical Function	USVI Application	USVI Status	BVI Application	BVI Status
Can you map it?	Mapping	USGS and WAPA-based coverages	Done	Land use cadastral	Done and proposed
Where is what?	Natural resource management	DLG hydrography	Done	Coastal atlas Sensitive areas	Proposed and partial
		DLG hypsography	Done	Significant features Population data	Major islands done
		Well inventory	Done	Land use	
		STJ national park coverages	Done	Sensitive areas	
Where has it changed?	Temporal	DLG and WAPA	Done	Land use updates population	Proposed
		Boundaries and roads 1982 to 1989			Proposed
What relationships exist?	Spatial			Land use Population data Coastal atlas	Partial Proposed Partial
Where is it best?	Suitability	STJ national park coverages (limited application)	Done	Land use Coastal atlas Sensitive areas Significant features	Partial
What affects what?	System	STJ national park coverages (limited application)	Done	Land use Coastal atlas Sensitive areas Significant features Population data	Partial and proposed
What if...?	Simulation	None discussed		Speculation, but no plans to implement yet	

involved multiple consultants providing often conflicting advice. This process served to educate policy-makers and managers in a much broader range of possibilities and avoidable problems than were available to the USVI. A corollary to the need for careful planning is the need to avoid making decisions or commitments to specific systems before such decisions are absolutely necessary. Especially in systems involving high technology, premature decisions often mean early obsolescence.

- *Implement in phases with early demonstration products:* Some issues, such as cadastral mapping and scale, are so subtle to inexperienced users that they need practice in real-life situations. If the USGS had spotted the scale problems at an early stage in the data conversion process, the USGS may have been able to provide a better solution. Some USVI critics claim the “1:24,000—one size fits all” attitude characterizes the federal approach.

- *Identify critical success factors for each situation:* In some environments (e.g., USVI), GIS is most attractive for its ability to provide enhanced powers of analysis. In others, such as the BVI, it is seen as a data integration tool and as a way to better inform political leadership and the public. To ensure success, major GIS implementations also need to meet the three major support requirements:
  - A political/senior management “chief”
  - A technical “chief”
  - Competent outside technical assistance

Finally, implementers should recognize that they have a stake in open information sharing. They should seek ways to redefine the decision-making process as a nonzero sum game: more information should benefit all parties. Of course, such changed attitudes require fundamental value shifts that take a long time to achieve and may have high short-term costs.

References

1. McElroy, J.L. 1991. The stages of tourist development in small Caribbean and Pacific islands. In: Proceedings of the International Symposium on the Island Economies: Policy Models for Planning Development, Lesbos, Greece (November).
2. Bureau of Economic Research, Virgin Island Department of Commerce. 1994. Economic indicators. St. Thomas, U.S. Virgin Islands (July).

3. Potter, L. 1984. Program notes from the Caribbean Conference of Planners, Kingston, Jamaica.

4. Potter, B., K. Green, and M. Goodwin. 1988. Management of natural resource information for the Virgin Islands National Park and Biosphere Reserve: Special biosphere reserve report. St. Thomas, U.S. Virgin Islands: Island Resources Foundation.

5. Government of the U.S. Virgin Islands. 1991. Application for technical assistance funds: Virgin Islands GIS system. Office of Territorial and International Affairs, U.S. Department of the Interior (April).

6. Parks, J.E. 1992. Invitation to a demonstration of the Virgin Islands GIS. U.S. Geological Survey, St. Thomas, U.S. Virgin Islands (October).

7. University of Manchester, British Overseas Development Administration. 1993. British Virgin Islands coastal atlas. Project was collaboration between graduate students at Manchester University, the British government, and the government of the British Virgin Islands.

8. Berry, J.K. 1994. What GIS can do for you. GIS World. Boulder, CO (May).

Application of GIS for Environmental Impact Analysis in a Traffic Relief Study

Bruce Stauffer
Advanced Technology Solutions, Inc., Lancaster, Pennsylvania

Xinhao Wang
University of Cincinnati, Cincinnati, Ohio

Abstract

This paper presents an application of a geographic information system (GIS) in a traffic relief study. Traffic congestion has severely affected the environmental quality and the quality of life for residents in the study area. A team of planners, environmental specialists, historians, landscape architects, traffic engineers, and GIS professionals organized to solve the problem. The team has evaluated the environmental and socioeconomic impacts of highway alignments from the very first step through every major decision for the duration of the project.

The GIS professionals have played a crucial role in maintaining constant and active interactions among members of the project team, federal and state agencies, and the public. GIS has helped to develop a natural and cultural resource inventory, identify contamination sources, assess environmental constraints, and evaluate proposed highway alignment alternatives. GIS provides an ideal atmosphere for professionals to analyze data, apply models, and make the best decisions. The high-quality map products that GIS creates enhance the quality of public presentations and reports. The authors feel that, as this project has progressed, more people have realized the benefit of using GIS.

Introduction

A traffic relief study, as one type of transportation project, aims to resolve traffic congestion problems through a combined strategy of upgrading existing infrastructure, building new infrastructure, and controlling traffic demand using congestion management strategies (CMS). This type of study proceeds through at least the following steps:

- Problem identification
- Data collection
- Preliminary design
- Environmental impact analysis
- Final design
- Construction

The process heavily involves federal, state, and local government agencies, as well as the public. The goal of the project is to develop an environmentally sound solution to the traffic congestion, which also happens to promote economic development and improve quality of life for people in the local area and the region. Environmental, social, and economic issues must be equally addressed from the very first step through the final design. Federal and state regulations generally require an environmental impact statement (EIS) when constructing new infrastructure or upgrading existing road systems. Preparing an EIS is a requirement for such a project and demands a significant commitment of time, money, staff, and technical resources.

A geographic information system (GIS) has the ability to process spatially referenced data for particular purposes. Along with the development of computer hardware and software, GIS has progressed from pure geoprocessing, to management of geographic information, to decision support (1). This paper presents the application of GIS in an ongoing traffic relief study in Marshalls Creek, Pennsylvania. The GIS function in this study has had various purposes:

- Inventory data compilation
- Spatial data analysis
- Map production
- Traffic modeling support
- Public presentations

The study shows that GIS can play an important and innovative role in transportation studies.

Project Description

The study area is in the Pocono region, located in northeastern Pennsylvania (see Figure 1). The Pocono Mountains and Delaware Water Gap National Recreation Area possess a wealth of natural and cultural resources. The area is famous for providing year-round vacation activities. Attractions include fishing, canoeing, and whitewater rafting in the summer and downhill and cross-country skiing, snowmobiling, snowboarding, and ice fishing in the winter. The area includes quiet woodland trails past a rushing waterfall and scenic settings for camping. In the fall, the area is ablaze with the brilliant colors of foliage. Various scenic sights, recreational sites, and national historic sites make the area ideal for attracting people to come for a day, a weekend, or a longer vacation.

Although tourism brings people to the Pocono area and promotes economic growth, it also brings a traffic congestion problem to the community. In addition, the influx of new home owners from New Jersey and New York adds the problem of commuter traffic to the area. The most troublesome section is in the vicinity of Marshalls Creek where U.S. Route 209 intersects with Pennsylvania (PA) Route 402. Two intersections are only about 500 feet apart. The traffic tieups can extend up to 3.5 miles on northbound Route 209, all the way back to Interstate Highway I-80. Emergency response times on U.S. Route 209 can be up to 20 minutes during peak traffic. The heavy traffic volume results in traffic accidents exceeding state averages on secondary roads as motorists seek alternative routes to avoid congestion. Through traffic traveling to and from New England using U.S. Route 209 as a connector between I-80 and I-84 makes the problem even worse. The year-round outdoor activities perpetuate the constant traffic problems that have severely affected the quality of life for residents in and around the Marshalls Creek area.

In response to the problems, the Pennsylvania Department of Transportation (PennDOT) selected a project team in February 1993 to conduct a traffic relief study in the Marshalls Creek area. The project team consists of individuals from seven firms and represents a wealth of experience in the variety of disciplines necessary to successfully complete this project. The team members include land use and traffic planners, biologists, historians, traffic and environmental engineers, surveyors, and GIS and global positioning system (GPS) professionals.

In addition to PennDOT, the funding agency, several federal and state regulatory agencies periodically review the development of the EIS to ensure that it meets regulations. These agencies are the Federal Highway Administration (FHWA), the U.S. Environmental Protection Agency (EPA), the U.S. Army Corps of Engineers, the Pennsylvania Department of Environmental Resources (DER), the Pennsylvania Historic Museum Commission, and the Pennsylvania Fish and Game Commission. Local planning commissions and citizen representatives also actively participate in advisory capacities. A series of agency coordination meetings, public meetings, and public information newsletters coordinates the activities of all participants over the course of the project.

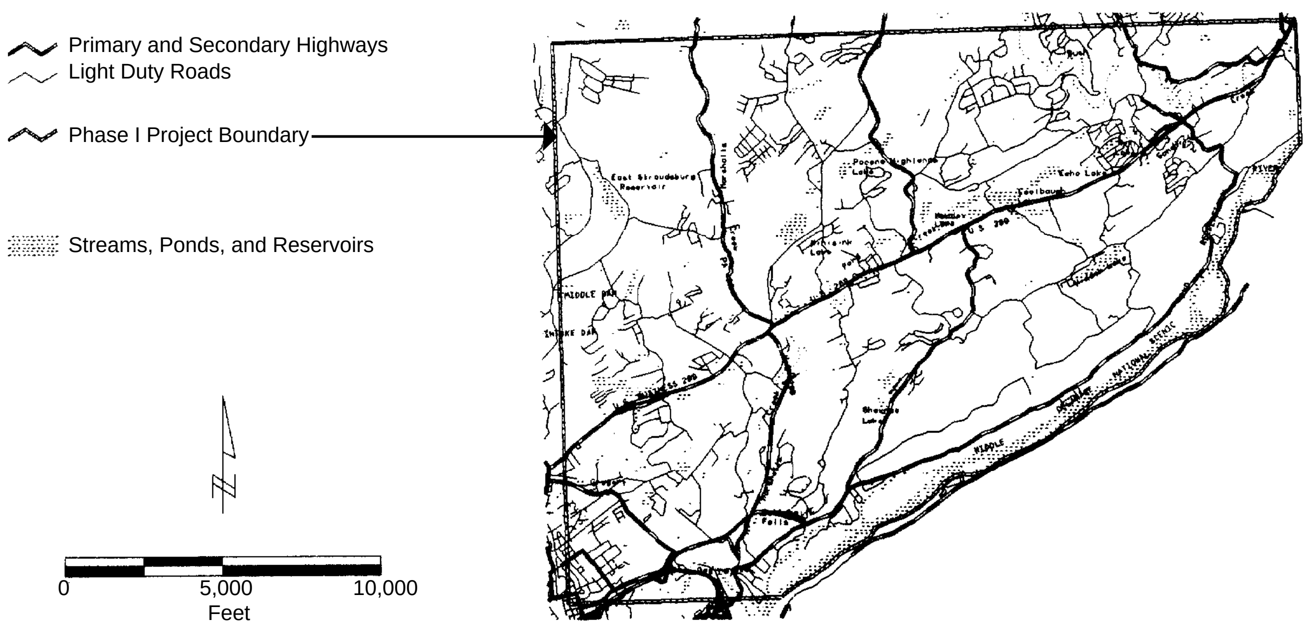


Figure 1. Phase I project area.

After collecting traffic data and performing traffic demand modeling, the team realized that adopting strategies to control traffic demand and upgrading existing roads through widening and intersection improvements would not suffice to meet demand projections for the design year 2015. Consequently, the team has determined a new road is needed to alleviate congestion in Marshalls Creek.

The study aims to identify alternatives to relieve traffic congestion along U.S. Route 209, PA Route 402, and Creek Road, and eliminate backups onto I-80 from U.S. Route 209. The alternatives should also improve air quality by reducing fuel consumption and vehicle emissions and facilitate travel through Marshalls Creek for local and through traffic. The improvements must comply with federal and state regulations. The study team must consider county and local government goals and objectives so that the traffic capacity improvements will be compatible with planned local development.

The project is being conducted in two phases. Phase I, which is complete, was an investigation broad in scope. It used inventory of secondary data to describe the environmental characteristics of the area. A traffic demand model identified the area for detailed study after a preliminary analysis of a wide range of baseline data. In Phase II, the team analyzes both primary and secondary data and delineates alternative alignments or transportation upgrading options that meet the need and minimize impacts. Analysis of environmental and engineering factors assists both in the determination of the most practical alternative and in preparation of the final EIS.

The nature of the study requires the analysis of a variety of data at different scales by different professionals. Through field investigation, the project team also constantly updates and adds new data to the existing database. The new data may be attribute data about some geographic features or may be locational data. The project team has found GIS to be an appropriate tool to meet the challenge of better conducting the study.

The team has used GIS extensively in both Phase I and Phase II studies. The two phases vary in data requirements, scales, and purposes of spatial analysis. With the support of GIS, the team has been able to quickly assemble data at adequate scales and present data in formats that are familiar to different professionals. GIS's data manipulation power distinguishes the different requirements of the two phases and at the same time, clearly depicts the linkage between the two phases. The following sections describe GIS applications that have helped facilitate the study and coordinate project team members, public agencies, and citizens.

GIS Application

GIS contains powerful tools to process spatially referenced data. These processes and their results are meaningless, however, without a clearly defined objective. Many professionals point out the importance of focusing GIS on practical problems. Using GIS is not an end; it is a means to represent the real world in both spatial and temporal dimensions. The benefits of using GIS can be summarized in three aspects:

- GIS helps to portray characteristics of the earth and monitor changes of the environment in space and time (2).
- GIS helps us to more deeply understand the meaning of spatial information and how that information can more faithfully reflect the true nature of spatially distributed processes (3).
- GIS helps us to model alternatives of actions and processes operating in the environment (2), to anticipate possible results of planning decisions (4), and to make better decisions.

This project demonstrates the advantages of applying GIS to solve practical problems from the above three aspects. An EIS requires extensive data about natural resources, land uses, infrastructure, and distribution of many interrelated socioeconomic factors. The accuracy and availability of required data depend on the scope of a study and the size of the study area. Our study shows that GIS, with its data retrieval, analysis, and reporting abilities, significantly improves the analysis. GIS helps to collect data at various scales, store data, and present data in forms that allow the project team to carry out the study in an innovative way.

Phase I Study

Phase I of the traffic relief study was completed in 1993. The goal of Phase I was to acquire understanding of the general features of the area and to use a traffic demand model for delineating an area for detailed study. The study area is approximately 52 square miles. To provide data for the preliminary analysis and the traffic demand modeling, the team developed baseline data inventory with GIS. Data were primarily secondary data that came from several different sources in different formats. For example, the U.S. Census Bureau 1990 population data were in TIGER format, the U.S. Fish and Wildlife Service National Wildlife Inventory (NWI) files in digital line graph (DLG) format, the U.S. Soil Conservation Service Monroe County Soils in DLG format, and the U.S. EPA Monroe County Natural Areas in ARC/INFO format. The majority of data sources were at scales between 1:15,000 and 1:24,000. With GIS tools, the team integrated these baseline data into a common presentation scale and projection. This process ensured an effective and comprehensive spatial analysis in the study area.

The team arranged and stored data in layers according to themes. Examples of the data layers included:

- Road center lines.
- Twenty-foot elevation contours.
- Utility lines.
- Water features.
- Subdivision boundaries.
- 1990 Census population by Census Tracts and Blocks.
- Flood plains.
- Geological formations.
- Public facilities, including schools, churches, and cemeteries.
- Political boundaries.
- Hazardous waste locations.
- Potential archaeological areas.

With these data layers, the team generated a series of 17 thematic maps to describe the features of the study areas. All maps were plotted on E-sized papers (48 inches x 36 inches) with the same map layout. The general reference map served as a base map for the other themes. It included several data layers to provide geographic references to the study area.

In addition to the base map, each individual theme map showed only one theme at a time, such as soils, subdivisions, and wetlands. Some theme maps showed derived data from the original data layers. For example, in developing a slope theme map, the team first built a three-dimensional surface from the 20-foot elevation contours, then calculated slope in degrees and aggregated areas based on a 10-degree interval. The slope theme map showed the result from the data processing. In addition, the project team created summary statistics tables to help team members gain knowledge about the study area. Table 1 is an example of the summary statistics for land use categories.

Table 1. Phase I Statistical Summaries for Land Use Categories

Land Use	Acres
Urban	10,628
Agricultural	1,183
Rangeland	12
Woodland	20,957
Water	1,394
Wetland	1,428
Transitional	422

During preparation of the Phase I inventory data and summary statistics, traffic planners performed traffic demand modeling to determine new road connections that would provide a minimum acceptable level of service in the year 2015. The modeling result was loaded into the GIS and converted into the same format and projection as other inventory data. Figure 2 displays the boundary that the traffic demand model delineated and the actual Phase II boundary. The two boundaries were not the

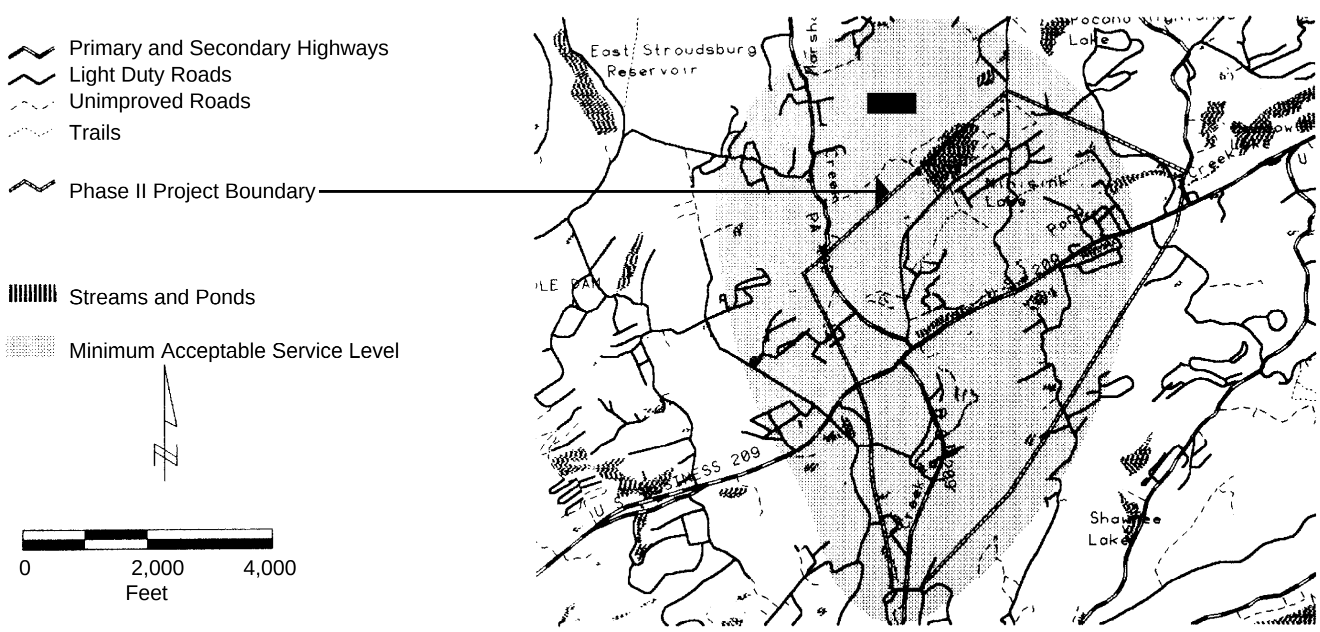


Figure 2. Phase II project area.

same. The thick line enclosed a Phase II study area that was delineated based on the traffic demand modeling and the team's understanding of environmental and other factors in the project area.

Phase II Study

The Phase II study is still ongoing. The area for the Phase II study is much smaller than that for Phase I. It is about 3.2 miles by 2.5 miles, or approximately 4 square miles.

The objective of the Phase II study is to conduct a detailed analysis for delineating a full range of feasible highway alignment alternatives. The alternatives must meet the needs of relieving traffic congestion and minimizing its impact on environmental and cultural resources.

Because the accuracy of the Phase I data was not sufficient for the Phase II study, the project team has collected data using different approaches to develop a similar set of baseline data at a finer scale. The major data source has been the photogrammetry data provided by PennDOT at a 1:2,400 scale. The data include:

- Road cartways
- Five-foot elevation contours
- Utility lines
- Water features
- Buildings footprints
- Bridges

The team directly digitized tax parcel boundaries from Monroe County Tax Assessor's maps that range in scales from 1:1,200 to 1:4,800. After digitizing each map sheet separately, the team merged them together to create a continuous parcel layer. The data layer has been adjusted to fit with the PennDOT photogrammetry data although the two data sets do not seem to match exactly. In addition, digital orthophotographs at 5-foot pixel resolution were also obtained for the project.

From these baseline data, the team has constructed Phase II data layers in four different ways.

The first approach digitizes from compilations on project base maps. For example, the team creates a land use data layer from the digital orthophotographs and infrared photography. The GIS group first plots the digital orthophotographs on a set of 1:2,400-scale map sheets. Road cartway and water features are plotted on top of the orthophotographs. Then land use specialists delineate land use boundaries with fine color markers and code land uses on maps according to the Anderson land use classification. In the end, the GIS group digitizes the land use boundaries from the compilations to create a land use data layer. Similarly, the 100-year flood plain data layer is delineated from compilations on project base maps with 5-foot contours and digitized.

The second approach derives new data layers from existing data. Buildings and structures are plotted at a 1:2,400 scale. Both historians and environmental engineers use the plots in their field investigations. After historians identify historic-eligible buildings on the plots, the GIS group develops an attribute data file that links the historic inventory data to the building geometry. Similarly, field investigations identify buildings and structures associated with contamination sites. The system stores types of contamination as building attributes. By overlaying the building data layer with the tax parcel data layer, the team can identify properties on which historic buildings or contamination sites are located.

The third approach constructs data layers by referencing Phase I data. For example, Phase II subdivision boundaries are derived from the digitized tax parcel boundaries by referring to the Phase I subdivision boundaries. Phase I subdivisions were manually compiled at 1:24,000 using approximate location, which did not align very well with the more accurate tax parcel boundaries. Using a 1:7,200-scale plot that shows both Phase I subdivision boundaries and the Phase II tax parcel boundaries, planners can verify and indicate properties associated with each subdivision. These properties are dissolved to create new boundaries for subdivisions that precisely fit with tax parcels. The same approach is used to refine public parks and private recreation areas.

The fourth approach obtains spatial data with GPS. The GPS surveyors collect accurate locational data about key features, such as boundaries of wetlands, site locations for hazardous waste, and locations of archaeological field samples. GPS data also supplement existing data, such as delineating footprints of new buildings to update the PennDOT baseline data. The integration of GIS and GPS provides the project team with accurate and up-to-date data.

Phase II data layers have provided much richer information for a detailed study of environmental features. They are merged in many different combinations to show the spatial distributions of different factors from different perspectives. Table 2 lists some of the map themes created for the Phase II study. All maps are plotted at a 1:7,200 scale.

In addition to using GIS as a data library and map production tool, we use GIS to support decision-making in two ways. First, the creation of a composite data layer has revealed the impact of alternative alignments on several composite constraints. The composite data layer is an overlay of several inventory layers and shows the various factors coincident at any location, and the relative importance of these factors.

Table 2. Selected Phase II Map Themes

Theme	Description
General reference	Road networks, hydrographic features, churches, municipal boundaries, and utilities
Parcels	Tax parcel boundaries
Community facilities	Public parks and private recreation, cemeteries, and public buildings
Flood plains	100-year flood plains
Land use	Current land uses
Subdivision	Approved subdivisions
Slope	Areas delineated with 5-degree slope intervals
Wetlands	Wetlands
Historic resources	Historic buildings and properties
Hazardous wastes	Hazardous waste sites and related buildings

These constraints have been identified by citizens, government agencies, and the project team. The major composite constraints include wetlands, historic properties, steep slopes (slope greater than 15 degrees), public parks and private recreation areas, 100-year flood plains, potential archaeological areas, prime agricultural land, subdivisions, and existing buildings and structures.

Two maps have been created from the composite constraints layer. One map shows the number of coincident constraint layers that occurs in any one location (see Figure 3). The other map shows the composite relative importance of coincident features. Both maps present a "sensitivity surface" view of the project area.

Secondly, the team uses GIS to perform interactive summary statistics for each alternative alignment. The project team analyzes the impact of the alignment alternatives on each individual constraint layer.

Two approaches define the impact areas for comparison. The first set impact areas are the areas enclosed by the footprint that traffic engineers delineated for each alternative. The second set impact areas are 300-foot buffers on both sides of the alignment delineated by the traffic engineers. The boundaries of the impact areas overlay on constraint data layers. Figure 4 displays wetlands crossed by alignment alternative ROW1B.

A set of summary statistics are calculated for each alignment. In the end, we compare the statistics for each alignment in a matrix (see Table 3). The matrix arranges constraints as rows and alignment alternatives as columns. The statistics include acres of selected features within each impact area, such as wetlands or high-quality watersheds, and total counts of features, such as historic-eligible buildings. The summary statistics also include listings of building names for businesses or public facilities within the impact areas.

The team has repeated the summary statistics several times as alignments shift. This procedure ensures that the final selected highway alignment minimizes environmental impacts, best meets project needs, and is the most cost-effective alignment to construct. The statistical matrix of impacts versus alignments is one of the critical evaluation criteria for comparing alignments and ultimately for selecting the final alignment.

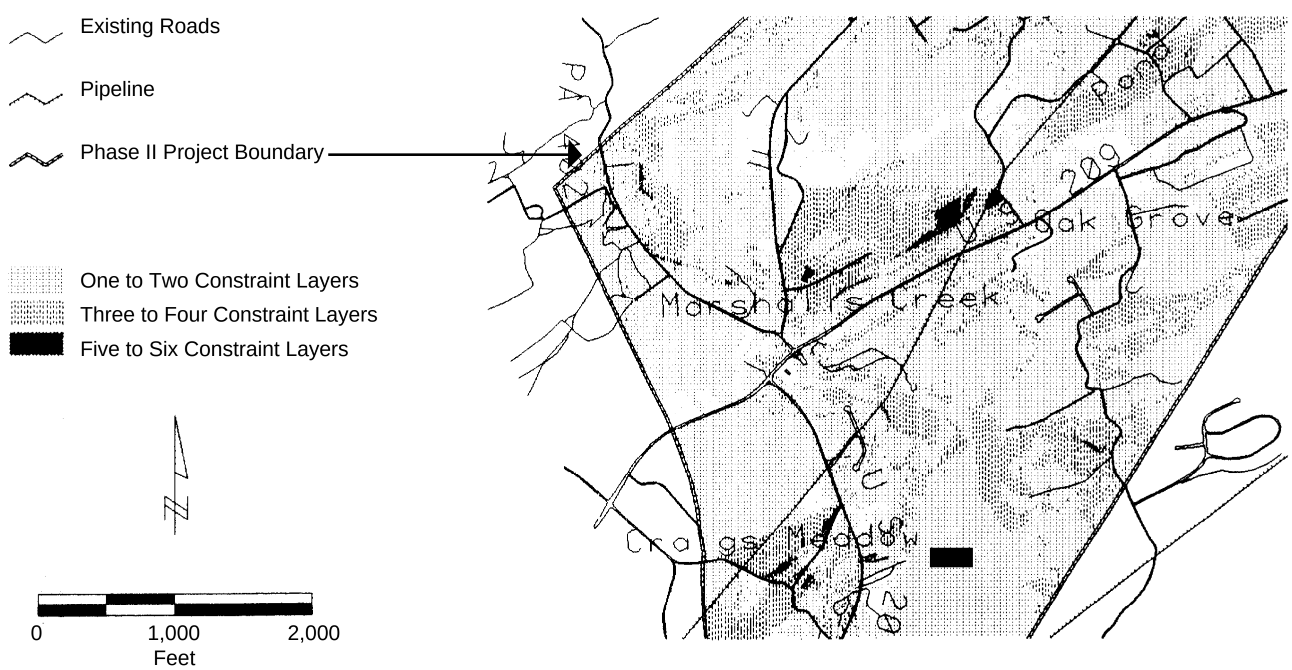


Figure 3. Composite constraints.

Table 3. Phase II Summary Statistics by Alternative

Constraints	Alignment Alternatives				
	ROW1A	ROW1B	ROW2A	ROW2B	ROW3
Wetlands					
Acres	2.76	3.27	3.25	1.8	13.64
Count	13	14	17	15	17
High-Quality Watersheds					
Acres	66.18	85.62	71.92	94.58	82.90
Hazardous Waste Parcels					
Acres	99.50	116.77	98.25	120.68	215.38
Count	8	5	11	10	6
Parcels With Historic-Eligible Buildings					
Acres	14.20	0.10	0.14	0.14	3.46
Count	3	1	2	2	2
Historic-Eligible Buildings					
Count	1	0	1	1	1

Benefit of Using GIS

In recent years, many federal, state, and local agencies have been actively acquiring and automating digital data (5). These databases provide various types of information at scales that are appropriate for a preliminary study covering a large area.

A more detailed study, which usually covers a smaller area, often requires more accurate data to describe the spatial distribution of relevant factors. GIS is flexible, allowing use of data at the scale and accuracy appropriate to the study purpose. The team has found that this feature helps improve the efficiency of the project without sacrificing the accuracy. This study has required two sets of scales ranging from 1:24,000 scale for Phase I, which required projectwide socioeconomic and environmental assessments, to 1:2,400 scale for Phase II, which requires detailed analysis for design of alternative alignments.

GIS has served as a digital database manager to assemble environmental, traffic, geographic, socioeconomic, and other data into a centralized project database. Data analyzed in this study originate from a variety of sources, such as PennDOT, U.S. Geological Survey (USGS), U.S. Census Bureau, Monroe County, and field survey. They are in different formats, including digital data in ARC/INFO, INTERGRAPH, and AutoCAD formats, GPS data, digital images, paper maps, and tabular data. Many of the public agencies and private organizations involved in this project already have digital data that GIS could easily use for this specific project. This has helped to reduce the overall costs of data collection and conversion.

This study demonstrates that GIS can support the information needs of many disciplines within a common framework and provide powerful, new tools for spatial analysis. Aside from technological considerations, GIS development initiates a higher-order systematization of geographic thinking (3), which is crucial to the success of a transportation project. In this project, GIS helps to determine the total impacts of alternative alignments on identified constraints. The team accomplishes this by superimposing the alternative alignments on constraint

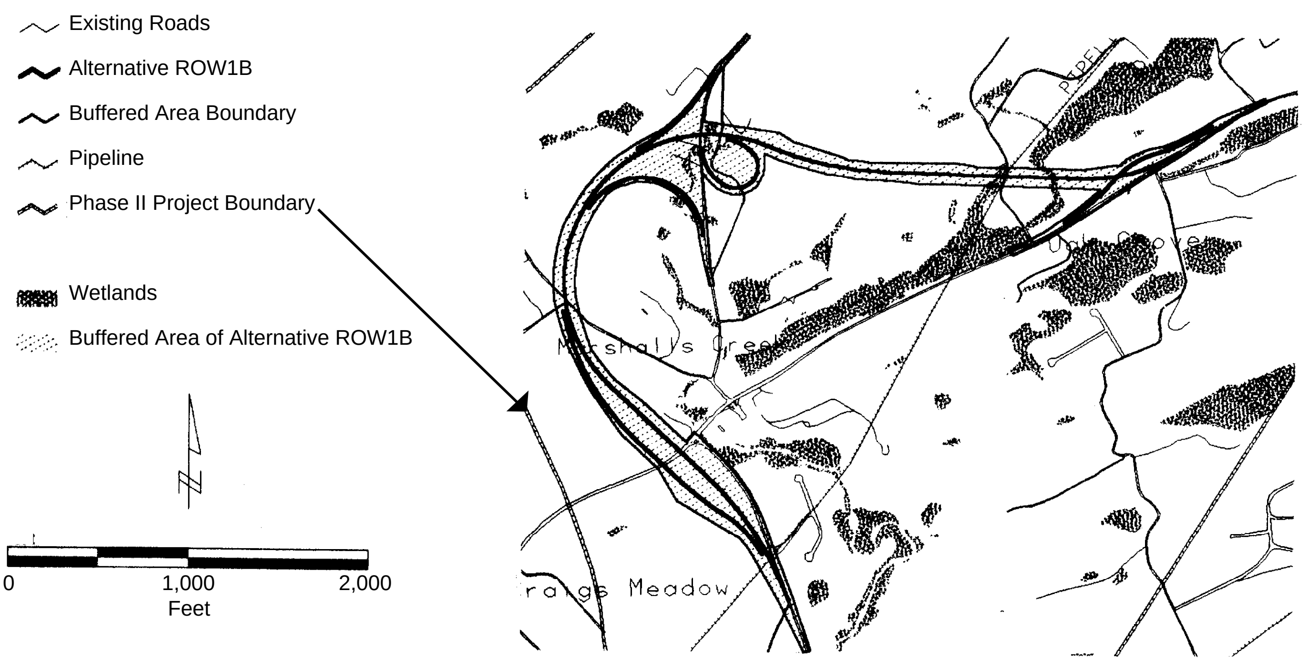


Figure 4. Wetlands crossed by an alternative alignment.

data layers to determine the total amount of each constraint layer that each alignment encounters. This approach supports the analysis of multiple alignments across the constraint surfaces for a variety of alternative scenarios. The spatial analysis tools and statistical function embedded in GIS prove to be very useful in such study.

Digital data that the GIS stores are used to summarize environmental, social, and economic data in many different ways. These functions include summing total acreage, listing entities of special interest, and counting numbers to provide useful baseline statistics for various alignments. Within a day, GIS accomplished what would have required several months of staff labor; GIS summarized impacts of 11 alternative alignments on all constraint layers. In addition, GIS creates composite constraints from individual data layers. A composite constraint data layer is created through a series of overlays to illustrate geographic coincidence of inventory themes.

Conventionally, engineers in a project such as this first delineate alternatives for alignments from the engineering perspective; they often consider factors such as steep slopes and costs. Then, other professionals, such as environmental specialists, historians, and planners, evaluate the alternatives from each point of view. GIS makes possible an early integration of environmental and engineering activities, ongoing communication with funding agencies and the public, and continual integration of a multidisciplinary team.

GIS helps to maintain high-quality data for the project. It allows for error checking and quality control of multiple data layers that would not be possible with conventional mapping. The team always compares a new data layer with other data to check for conflicts. Making check plots allows for quick identification of errors and missing data.

For instance, in the process of assigning building-use attributes to existing buildings, the GIS team first plotted buildings on a map and created a table with building identifiers. The field team then used the unique identifiers shown on the plots when noting building names and building uses in data collection tables. After relating the data table with the building data layer, the team found some buildings that did not have building use data. Moreover, some buildings were assigned uses that were out of range or seemed out of place, such as a residential building surrounded by several commercial buildings. The team highlighted the data for those buildings and sent them to engineers for verification. Through this data cleaning process, the team was able to obtain complete building use information.

Data quality directly affects project quality. Without GIS, this type of study often involves using and comparing maps at different scales, which frequently introduces serious errors. For each phase of this study, project team members have used a fixed-base map scale for all compilations. During data entry and data transformation, the team has kept accurate registration between data layers, ensuring data of the same resolution. In addition, the team has used FREQUENCY, one of the tools that GIS provides, to look for data values that are out of range as well as missing data. GIS tools have also helped to derive new relationships for features. For example, dissolving parcels has helped to create subdivision outlines, or overlaying historic buildings with parcels has helped to find the parcels on which they are located.

Planners, environmental specialists, historians, and landscape architects on the project team are responsible for field data collection, verification, and if necessary, compilation of field data into the standard project database. Wherever possible, the team has used GPS to eliminate the task of manual compilation and to improve accuracy of locating data. A fundamental requirement in applying the technology appropriately is to understand its capabilities, requirements, and limitations. Because several members manage inventory attributes, they need to know how to maintain unique identifiers for features so they can link up to the geometry. Because AutoCAD data transfers occur routinely with engineers, it is necessary to structure how the AutoCAD drawing files can be organized, as well as how certain attributes can be transferred by line color, layer name, or line width. The GIS group coordinates closely with other team specialists to identify quick, accurate, and cost-effective methods of data collection, data analysis, and presentation. In conjunction with the progress of the project, specialists from different fields have become familiar with the concept, requirements, and use of GIS. They now feel comfortable discussing alternatives while looking at results displayed on a computer screen.

GIS has created high-quality map products for public presentations and reports. GIS has also been used in several agency coordination meetings to display data and alternative alignments. GIS has allowed for different data layers to be displayed on the screen with specific combinations of features at various scales. Public agencies and citizens have been impressed by the clear and friendly graphic response to their questions. They have expressed interest in using the technology in future projects.

Once the study is complete, digital data assembled in this project will be an excellent resource for future projects in the study area. For these reasons, GIS provides more cost-effective project support for gathering, managing, and using data than that provided by paper and mylar maps.

Summary

For this project, the importance of innovation based on a solid scientific foundation cannot be overstated. In the current economic and regulatory climate, sound GIS methods are emerging as the only convincing and cost-effective means for locating, designing, and gaining approval for major public and private infrastructure projects.

The technology offers new and exciting tools for transportation planning.

The methodology that this project team has used can be successfully applied to other projects that require environmental assessment. The team has found GIS to be an extremely useful tool as users continue to learn its capabilities and the multiple tools that it offers. The regulatory agencies have repeatedly made favorable comments on how GIS can offer interactive viewing in a show-and-tell environment. This project is one of the first EIS projects to use GIS in a PennDOT-funded project. PennDOT appears convinced that GIS is an important component for conducting EIS for highway studies. More EIS projects probably will demand GIS services as part of the project approach, in part, because many federal and state regulatory agencies are increasingly using GIS.

References

1. Fischer, M.M. 1994. From conventional to knowledge-based geographic information systems. Computer, Environment, and Urban Systems 18(4):233-242.
2. Star, J., and J. Estes. 1990. Geographic information systems: An introduction. Englewood Cliffs, NJ: Prentice Hall, Inc.
3. Bracken, I., and C. Webster. 1990. Information technology in geography and planning, including principles of GIS. London, England: Routledge.
4. Burrough, P.A. 1986. Principles of geographic information system for land resources assessment. Oxford, England: Clarendon Press.
5. Hendrix, W.G., and D.J.A. Buckley. 1992. Use of a geographic information system for selection of sites for land application of sewage waste. J. Soil and Water Conserv. 47(3):271-275.