United States
Environmental Protection
Agency

Office of Research and
Development
Washington, D.C. 20460

EPA/600/R-99/093
October 1999

Environmental Technology Verification Report

Environmental Decision Support Software

Environmental Software SitePro™ Version 3.0

THE ENVIRONMENTAL TECHNOLOGY VERIFICATION PROGRAM

U.S. Environmental Protection Agency

Oak Ridge National Laboratory

ETV Joint Verification Statement

TECHNOLOGY TYPE:	**ENVIRONMENTAL DECISION SUPPORT SOFTWARE**	
APPLICATION:	**INTEGRATION AND VISUALIZATION OF ENVIRONMENTAL DATA SETS**	
TECHNOLOGY NAME:	**SitePro™ Version 3.0**	
COMPANY:	**Environmental Software**	
ADDRESS:	**17011 Beach Blvd., Suite 900** **Huntington Beach, CA. 92647**	**PHONE: (714) 379-7000** **FAX: (714) 379-7001**
WEBSITE: **E-MAIL:**	**www.envsoft.com** **info@envsoft.com**	

The U.S. Environmental Protection Agency (EPA) has created the Environmental Technology Verification Program (ETV) to facilitate the deployment of innovative or improved environmental technologies through performance verification and dissemination of information. The goal of the ETV Program is to further environmental protection by substantially accelerating the acceptance and use of improved and cost-effective technologies. ETV seeks to achieve this goal by providing high-quality, peer-reviewed data on technology performance to those involved in the design, distribution, financing, permitting, purchase, and use of environmental technologies.

ETV works in partnership with recognized standards and testing organizations and stakeholder groups consisting of regulators, buyers, and vendor organizations, with the full participation of individual technology developers. The program evaluates the performance of innovative technologies by developing test plans that are responsive to the needs of stakeholders, conducting field or laboratory tests (as appropriate), collecting and analyzing data, and preparing peer-reviewed reports. All evaluations are conducted in accordance with rigorous quality assurance protocols to ensure that data of known and adequate quality are generated and that the results are defensible.

The Site Characterization and Monitoring Technologies Pilot (SCMT), one of 12 technology areas under ETV, is administered by EPA's National Exposure Research Laboratory (NERL). With the support of the U.S. Department of Energy's (DOE's) Environmental Management program, NERL selected a team from Brookhaven National Laboratory (BNL) and Oak Ridge National Laboratory (ORNL) to perform the verification of environmental decision support software. This verification statement provides a summary of the test results of a demonstration of Environmental Software's SitePro™ environmental decision support software product.

DEMONSTRATION DESCRIPTION

In September 1998, the performance of five decision support software (DSS) products were evaluated at the New Mexico Engineering Research Institute, located in Albuquerque, New Mexico. In October 1998, a sixth DSS product was tested at BNL in Upton, New York. Each technology was independently

evaluated by comparing its analysis results with measured field data and, in some cases, known analytical solutions to the problem.

Depending on the software, each was assessed for its ability to evaluate one or more of the following endpoints of environmental contamination problems: visualization, sample optimization, and cost-benefit analysis. The capabilities of the DSS were evaluated in the following areas: (1) the effectiveness of integrating data and models to produce information that supports the decision, and (2) the information and approach used to support the analysis. Secondary evaluation objectives were to examine DSS for its reliability, resource requirements, range of applicability, and ease of operation. The verification study focused on the developers' analysis of multiple test problems with different levels of complexity. Each developer analyzed a minimum of three test problems. These test problems, generated mostly from actual environmental data from six real remediation sites, were identified as Sites A, B, D, N, S, and T. The use of real data challenged the software systems because of the variability in natural systems.

Environmental Software staff used SitePro Version 3.0 to perform the visualization endpoint using data from Sites D, S, and T. Sites D and S have groundwater contamination, and Site T has soil contamination. The intent of the SitePro analyses was to demonstrate the software's capability to integrate large quantities of data into a visual framework for assistance in understanding a site's contamination problem. Because SitePro was not developed to address sample optimization or cost-benefit problems, Environmental Software staff did not attempt to perform these aspects of the test problems.

During the demonstration SitePro was used to import data from many different sources (drawing and data files) and integrate these into the SitePro platform. Database manipulations (sort and query), GIS operations (multiple layers on maps, hot-linking of the data to the maps), data analysis (creating contours of water level and contaminant concentration, geologic boring maps, and geologic cross-section maps) and visualization (two-dimensional maps containing site features, contour levels, sample locations, and measured values) were demonstrated. Details of the demonstration, including an evaluation of the software's performance, may be found in the report entitled *Environmental Technology Verification Report: Environmental Software, SitePro™ Version 3.0,*EPA/600/R-99/093.

TECHNOLOGY DESCRIPTION

SitePro is a software application designed to help environmental professionals quickly and comprehensively characterize and manage information relevant to understanding environmental contamination problems. SitePro integrates a database, a geographic information system (GIS), computer-aided design (CAD), mapping, contouring, boring logs, cross-sections, graphing, imaging and reporting inside one application. This integration provides support for decisions pertaining to monitoring and remediation. SitePro can be used to manage various types of environmental data including data on contaminated soil and water, air emissions, wastewater, and health and safety parameters. The software allows environmental professionals to manage and share their site data using a single file. SitePro runs on Windows 95 and 98 and NT platforms.

VERIFICATION OF PERFORMANCE

The following performance characteristics of SitePro Version 3.0 were observed:

Decision Support: SitePro was able to quickly import electronic data on contaminant concentrations, geologic structure, and surface structure from a variety of sources and integrate this information on a single platform. SitePro was able to display the information in a visual context to support data interpretation.

Documentation of the SitePro Analysis: Environmental Software staff used SitePro to generate reports that provided an adequate explanation of the process and parameters used to analyze each problem. Documentation of data transfer, manipulations of the data (e.g., how to treat contamination data as a

function of depth in a well), and analyses were included. Model selection and parameters for contouring were also provided in the exportable documentation.

Comparison with Baseline Analysis and Data: SitePro was able to generate geologic boring maps and cross sections that accurately matched the data. The software was also able to generate hydraulic head measurement and contaminant concentration maps. The maps ranged from posting of data at the sample location to contours generated through inverse distance weighting (IDW) interpolation routines. In general, SitePro-generated contour maps were consistent with the measured data and baseline analysis. In a few cases, however, the SitePro predicted contours did not completely match the data. The cause for the poor agreement was the choice of contouring parameters used by the analyst.

Multiple Lines of Reasoning: Environmental Software chose not to use SitePro to provide multiple interpretations of the data with different modeling parameters. SitePro has several contouring algorithms, but only one contouring algorithm (IDW) was used in the demonstration. In addition, different parameters could have been used in the IDW algorithm to explore the data. Performing multiple interpolations of the data using different interpolation routines and parameters would have provided multiple views of the data that generally assist in data interpretation.

In addition to performance criteria, the following secondary criteria were evaluated:

Ease of Use: The demonstration showed that SitePro was extremely easy to use. The SitePro platform has a logical structure to permit use of the options in the software package. SitePro was demonstrated to import and export data in a wide range of formats. During the demonstration, one of four .dxf files provided by Environmental Software could not be read by other .dxf readers.

Efficiency and Range of Applicability: SitePro has a flexible database structure that supports multiple data input formats. This provided a flexible platform which addressed problems efficiently because the platform could be tailored to the problem under study. The database permits queries on a wide range of fields (e.g., chemical name, date, concentration, and well identifiers) and also permits filtering (e.g., include only the maximum concentration at a location over a range of sample dates). The software allows evaluation of a wide range of environmental conditions (e.g., contaminants in different media). Completion of three problems required one person-week of effort.

Operator Skill Base: To use SitePro efficiently, the operator should have a basic understanding of the use of computer software in analyzing environmental problems. This includes fundamental knowledge about GIS, CAD, and database files. In addition, skills in contouring environmental data is also key to achieving satisfactory results.

Training and Technical Support: SitePro requires minimal training for efficient use. An analyst with the prerequisite skill base can be using the software within a day. SitePro offers a wide range of options for training and technical support. A detailed on-line help system is supplied with the software package. In addition, a user's manual is available to assist in operation of the software. A step-by-step tutorial provided with the software package covers the major features. Two one-day training courses (introduction and advanced) are available if desired. Technical support is available for a yearly maintenance fee.

Cost: At the time of the demonstration, SitePro was priced at $2295 for a single license. Educational and multiple license discounts are available. In addition, new clients are required to subscribe to one year of technical support at $275 per year.

Overall Evaluation: The main strength of SitePro was its ability to easily integrate and manage information to allow analysis and spatial visualization of the data. SitePro was capable of managing data files from a wide range of sources, querying the data files to examine specific issues, and generating

boring logs, geologic cross-sections, and contour maps for hydraulic head and contamination. SitePro's capabilities and ease of use make it suitable for assisting with complex environmental contamination problems including multiple sources and contaminants. For an operator with the proper qualifications, the software is easy to use, so it is a good choice for users who do not operate the software on a regular basis. The main limitation of SitePro observed in the demonstration was that, in some instances, the contours generated by SitePro showed poor agreement with the actual data. However, the poor agreement was due to the analyst's choice of contouring parameters. Two minor limitations of SitePro that were noted in the demonstration were the poor legibility of the geologic cross-section and boring maps and the inability to read one SitePro file (in .dxf format) using other software programs.

The credibility of a computer analysis of environmental problems requires good data, reliable and appropriate software, adequate conceptualization of the site, and a technically defensible problem analysis. The results of the demonstration showed that the SitePro software can be used to generate reliable and useful analyses for evaluating environmental contamination problems. This is the component of a credible analysis that can be addressed by the software; other components such as proper conceptualization and use of the code depend on the analyst's skills. The results of a SitePro analysis can support decision-making. SitePro has been employed in a variety of environmental applications. Although the SitePro has been demonstrated to have the capability to produce reliable and useful analyses, improper use of the software can cause the results of the analysis to be misleading or inconsistent with the data. As with any complex environmental DSS product, the quality of the output is directly dependent on the skill of the operator.

As with any technology selection, the user must determine if this technology is appropriate for the application and the project data quality objectives. For more information on this and other verified technologies visit the ETV web site at http://www.epa.gov/etv.

Gary J. Foley, Ph.D
Director
National Exposure Research Laboratory
Office of Research and Development

David E. Reichle
ORNL Associate Laboratory Director
Life Sciences and Environmental Technologies

NOTICE: EPA verifications are based on evaluations of technology performance under specific, predetermined criteria and appropriate quality assurance procedures. EPA, ORNL, and BNL make no expressed or implied warranties as to the performance of the technology and do not certify that a technology will always operate as verified. The end user is solely responsible for complying with any and all applicable federal, state, and local requirements. Mention of commercial product names does not imply endorsement.

EPA/600/R-99/093
October 1999

Environmental Technology Verification Report

Environmental Decision Support Software

Environmental Software SitePro™ Version 3.0

By

Terry Sullivan
Brookhaven National Laboratory
Upton, New York 11983

Anthony Q. Armstrong
Amy B. Dindal
Roger A. Jenkins
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831

Jeff Osleeb
Hunter College
New York, New York 10021

Eric N. Koglin
U.S. Environmental Protection Agency
Environmental Sciences Division
National Exposure Research Laboratory
Las Vegas, Nevada 89193-3478

ornl

Notice

The U.S. Environmental Protection Agency (EPA), through its Office of Research and Development (ORD), and the U.S. Department of Energy's Environmental Management Program through the National Analytical Management Program (NAMP), funded and managed, through Interagency Agreement No. DW89937854 with Oak Ridge National Laboratory, the verification effort described herein. This report has been peer and administratively reviewed and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use of a specific product.

Table of Contents

List of Figures v
List of Tables vii
Foreword ix
Acknowledgments xi
Abbreviations and Acronyms xiii

1 INTRODUCTION 1
Background 1
Demonstration Overview 2
Summary of Analysis Performed by SitePro Version 3.0 2

2 SITEPRO VERSION 3.0 TECHNOLOGY DESCRIPTION 4

3 DEMONSTRATION PROCESS AND DESIGN 5
Introduction 5
Development of Test Problems 5
Test Problem Definition 5
Summary of Test Problems 5
Analysis of Test Problems 6
Preparation of Demonstration Plan 8
Summary of Demonstration Activities 8
Evaluation Criteria 9
Criteria for Assessing Decision Support 9
Documentation of the Analysis and Evaluation of the Technical Approach 10
Comparison of Projected Results with the Data and Baseline Analysis 10
Use of Multiple Lines of Reasoning 10
Secondary Evaluation Criteria 10
Documentation of Software 10
Training and Technical Support 11
Ease of Use 11
Efficiency and Range of Applicability 11

4 SITEPRO VERSION 3.0 EVALUATION 12
Description of Test Problems 12
Site D 12
Site S 12
Site T 12
Evaluation of SitePro Version 3.0 13
Decision Support 13
Documentation of the SitePro Analysis and Evaluation of the Technical Approach 13
Comparison of SitePro Results with the Baseline Analysis 13
Site D 13
Site S 15
Site T 24
Multiple Lines of Reasoning 26

Secondary Evaluation Criteria 26
Ease of Use 26
Efficiency and Range of Applicability 26
Training and Technical Support 26
Additional Information about the SitePro Software 26
Summary of Performance 27

5 SITEPRO VERSION 3.0 UPDATE AND REPRESENTATIVE APPLICATIONS 29
Objective 29
Technology Update 29
Representative Applications 30

6 REFERENCES 31

Appendix A — Summary of Test Problems 32
Appendix B — Description of Interpolation Methods 37

List of Figures

1 Fourth-quarter Site D TCE contours generated by SitePro 14
2 Comparison of SitePro analysis, baseline analysis, and the baseline data for Site D fourth-quarter concentration contours 16
3 SitePro-generated contours for hydraulic head levels at Site D 17
4 Comparison of SitePro analysis, baseline analysis, and the baseline data for Site S water levels for the cost-benefit problem 18
5 Site S chlordane 5 and 500-µg/L contours as generated by SitePro 20
6 Site S chlordane 5 and 500-µg/L contours for the analytical solution and the baseline contour obtained using kriging with an anisotropy ratio of 0.3 21
7 Site S borehole log generated by Site Pro for monitoring well MW-245a 22
8 Geologic cross-section map for Site S generated by SitePro 23
9 Site T ethylene dibromide contours above threshold concentration of 21 µg/L 24
10 Site T contours for all four chemicals above their threshold concentrations 25
11 Site T area sampling coverage map with proposed new sampling locations 25

List of Tables

1 Summary of test problems 6
2 Data supplied for the test problems 7
3 Site T soil contamination threshold concentrations 13
4 SitePro Version 3.0 performance summary 28

Foreword

The U.S. Environmental Protection Agency (EPA) is charged by Congress with protecting the nation's natural resources. The National Exposure Research Laboratory (NERL) is EPA's center for the investigation of technical and management approaches for identifying and quantifying risks to human health and the environment. NERL's research goals are to (1) develop and evaluate technologies for the characterization and monitoring of air, soil, and water; (2) support regulatory and policy decisions; and (3) provide the science support needed to ensure effective implementation of environmental regulations and strategies.

EPA created the Environmental Technology Verification (ETV) Program to facilitate the deployment of innovative technologies through performance verification and information dissemination. The goal of the ETV Program is to further environmental protection by substantially accelerating the acceptance and use of improved and cost-effective technologies. The ETV Program is intended to assist and inform those involved in the design, distribution, permitting, and purchase of environmental technologies. This program is administered by NERL's Environmental Sciences Division in Las Vegas, Nevada.

The U.S. Department of Energy's (DOE's) Environmental Management (EM) program has entered into active partnership with EPA, providing cooperative technical management and funding support. DOE EM realizes that its goals for rapid and cost-effective cleanup hinge on the deployment of innovative environmental characterization and monitoring technologies. To this end, DOE EM shares the goals and objectives of the ETV.

Candidate technologies for these programs originate from the private sector and must be commercially ready. Through the ETV Program, developers are given the opportunity to conduct rigorous demonstrations of their technologies under realistic field conditions. By completing the evaluation and distributing the results, EPA establishes a baseline for acceptance and use of these technologies.

Gary J. Foley, Ph.D.
Director
National Exposure Research Laboratory
Office of Research and Development

Acknowledgments

The authors wish to acknowledge the support of all those who helped plan and conduct the demonstration, analyze the data, and prepare this report. In particular, we recognize the technical expertise of Randy Breeden and Mike Gansecki (EPA Region 8) and Budhendra Bhudari (ORNL) who were peer reviewers of this report. For internal peer review, we thank Marlon Mezquita (EPA Region 9); for technical and logistical support during the demonstration, Dennis Morrison (NMERI); for evaluation of training during the demonstration, Marlon Mezquita and Gary Hartman (DOE's Oak Ridge Operations Office); for computer and network support, Leslie Bloom (ORNL); and for technical guidance and project management of the demonstration, David Carden, and Regina Chung (DOE Oak Ridge Operations Office), David Bottrell (DOE Headquarters), Stan Morton (DOE Idaho Operations Office), Deana Crumbling (EPA's Technology Innovation Office), and Stephen Billets (EPA NERL). The authors also acknowledge the participation of Bern Baumgartner and Dave Low of Environmental Software, who performed the analyses during the demonstration.

For more information on the Decision Support Software Technology Demonstration, contact

Eric N. Koglin
Project Technical Leader
Environmental Protection Agency
Environmental Sciences Division
National Exposure Research Laboratory
P. O. Box 93478
Las Vegas, Nevada 89193-3478
(702) 798-2432

For more information on the Environmental Software SitePro product, contact

Dave Low
Environmental Software
17011 Beach Blvd., Suite 900
Huntington Beach, CA 92647
(714) 379-7000
dlow@envsoft.com
www.envsoft.com

Abbreviations and Acronyms

As	arsenic
.bmp	bitmap (file format)
BNL	Brookhaven National Laboratory
CAD	computer-aided design
CAS	Chemical Abstract Service
CTC	carbon tetrachloride
Cd	cadmium
CD	compact disk
Cr	chromium
d	day
DBCP	dibromochloropropane
dbf	database file
DCA	dichloroethane
DCE	dichloroethene
DCP	dichloropropane
DOE	U.S. Department of Energy
DSS	decision support software
dxf	data exchange format (file)
EDB	ethylene dibromide
EM	Environmental Management
EMIS	environmental management information system
EPA	U.S. Environmental Protection Agency
ES&H	environmental, safety, and health
ESRI	Environmental Systems Research Institute
ETV	Environmental Technology Verification Program
FTP	file transfer protocol
GB	gigabyte
GEO-AS	Geostatistical Environmental Assessment Software
GIS	geographical information system
GSLIB	Geostatistical Software Library Version 2.0
GUI	graphical user interface
ICTF	Intermodal Container Transfer Facility
IDW	inverse distance weighting
.jpg	JPEG file interchange format
L	liter
MB	megabyte
MCSP	Microsoft Certified Service Provider
MHz	megahertz (used to define the clock speed on computer processors)
NAMP	National Analytical Management Program (DOE)
NERL	National Exposure Research Laboratory (EPA)
NMERI	New Mexico Engineering Research Institute
ORD	Office of Research and Development
ORNL	Oak Ridge National Laboratory
PCE	perchloroethene or tetrachloroethene
ppb	parts per billion
ppm	parts per million

QA	quality assurance
QC	quality control
RAM	random access memory
ROM	read only memory
SADA	Spatial Analysis and Decision Assistance
SCMT	Site Characterization and Monitoring Technology
.shp	Shape file
SQL	structured query language
TCA	trichloroethane
TCE	trichloroethene
Tc-99	technetium-99
VC	vinyl chloride
VOC	volatile organic compound
2-D	two-dimensional
3-D	three-dimensional

Section 1 — Introduction

Background

The U.S. Environmental Protection Agency (EPA) has created the Environmental Technology Verification Program (ETV) to facilitate the deployment of innovative or improved environmental technologies through performance verification and dissemination of information. The goal of the ETV Program is to further environmental protection by substantially accelerating the acceptance and use of improved and cost-effective technologies. ETV seeks to achieve this goal by providing high-quality, peer-reviewed data on technology performance to those involved in the design, distribution, financing, permitting, purchase, and use of environmental technologies.

ETV works in partnership with recognized standards and testing organizations and stakeholder groups consisting of regulators, buyers, and vendor organizations, with the full participation of individual technology developers. The program evaluates the performance of innovative technologies by developing test plans that are responsive to the needs of stakeholders, conducting field or laboratory tests (as appropriate), collecting and analyzing data, and preparing peer-reviewed reports. All evaluations are conducted in accordance with rigorous quality assurance (QA) protocols to ensure that data of known and adequate quality are generated and that the results are defensible.

ETV is a voluntary program that seeks to provide objective performance information to all of the actors in the environmental marketplace and to assist them in making informed technology decisions. ETV does not rank technologies or compare their performance, label or list technologies as acceptable or unacceptable, seek to determine "best available technology," nor approve or disapprove technologies. The program does not evaluate technologies at the bench or pilot scale and does not conduct or support research.

The program now operates 12 pilots covering a broad range of environmental areas. ETV has begun with a 5-year pilot phase (1995–2000) to test a wide range of partner and procedural alternatives in various pilot areas, as well as the true market demand for and response to such a program. In these pilots, EPA utilizes the expertise of partner "verification organizations" to design efficient processes for conducting performance tests of innovative technologies. These expert partners are both public and private organizations, including federal laboratories, states, industry consortia, and private sector facilities. Verification organizations oversee and report verification activities based on testing and QA protocols developed with input from all major stakeholder/customer groups associated with the technology area. The demonstration described in this report was administered by the Site Characterization and Monitoring Technology (SCMT) Pilot. (To learn more about ETV, visit ETV's Web site at http://www.epa.gov/etv.)

The SCMT pilot is administered by EPA's National Exposure Research Laboratory (NERL). With the support of the U.S. Department of Energy's (DOE's) Environmental Management (EM) program, NERL selected a team from Brookhaven National Laboratory (BNL) and Oak Ridge National Laboratory (ORNL) to perform the verification of environmental decision support software. Decision support software (DSS) is designed to integrate measured or modeled data (such as soil or groundwater contamination concentrations) into a framework that can be used for decision-making purposes. There are many potential ways to use such software, including visualization of the nature and extent of contamination, locating optimum future samples, assessing costs of cleanup versus benefits obtained, or estimating human health or ecological risks. The primary objective of this demonstration was to conduct an independent evaluation of each software's capability to evaluate three common endpoints of environmental remediation problems: visualization, sample optimization, and cost-benefit analysis. These endpoints were defined as follows.

- *Visualization* — using the software to organize and display site and contamination data in ways that promote understanding of current conditions, problems, potential solutions, and eventual cleanup choices;

- *Sample optimization* — selecting the minimum number of samples needed to define a contaminated area within a predetermined statistical confidence;

- *Cost-benefit analysis* —assessment of either the size of the zone to be remediated according to cleanup goals, or estimation of human health risks due to the contaminants. These can be related to costs of cleanup.

The developers were permitted to select the endpoints that they wished to demonstrate because each piece of software had unique features and focused on different aspects of the three endpoints. Some focused entirely on visualization and did not attempt sample optimization or cost-benefit analysis, while others focused on the technical aspects of generating cost-benefit or sample-optimization analysis, with a minor emphasis on visualization. The evaluation of the DSS focused only on the analyses conducted during the demonstration. No penalty was assessed for performing only part of the problem (e.g., performing only visualization).

Evaluation of a software package that is used for complex environmental problems is by necessity primarily qualitative in nature. It is not meaningful to quantitatively evaluate how well predictions match at locations where data has not been collected. (This is discussed in more detail in Appendix B.) In addition, the selection of a software product for a particular application relies heavily on the user's background, personal preferences (for instance, some people prefer Microsoft Word, while others prefer Corel WordPerfect for word processing), and the intended use of the software (for example, spreadsheets can be used for managing data; however, programs specifically designed for database management would be a better choice for this type of application). The objective of this report is to provide sufficient information to judge whether the DSS product has the analysis capabilities and features that will be useful for the types of problems typically encountered by the reader.

Demonstration Overview

In September 1998, a demonstration was conducted to verify the performance of five environmental software programs: Environmental Visualizations System (C Tech Development Corp.), ArcView and associated software extenders [Environmental Systems Research Institute (ESRI)], Groundwater*FX* (Decision*FX* Corp.), Sampling*FX* (Decision*FX* Corp.), and SitePro (Environmental Software Corp.). In October, a sixth software package from the University of Tennessee Research Corporation, Spatial Analysis and Decision Assistance (SADA), was tested. This report contains the evaluation for SitePro version 3.0.

Each developer was asked to use its own software to address a minimum of three test problems. In preparation for the demonstration, ten sites were identified as having data sets that might provide useful test cases for the demonstration. All of this data received a quality control review to screen out sites that did not have adequate data sets. After the review, ten test problems were developed from field data at six different sites. Each site was given a unique identifier (Sites A, B, D, N, S, and T). Each test problem focused on different aspects of environmental remediation problems. From the complete data sets, test problems that were subsets of the entire data set were prepared. The demonstration technical team performed an independent analysis of each of the ten test problems to ensure that the data sets were complete.

All developers were required to choose either Site S or Site N as one of their three problems because these sites had the most data available for developing a quantitative evaluation of DSS performance.

Each DSS was evaluated on its own merits based on the evaluation criteria presented in Section 3. Because of the inherent variability in soil and subsurface contamination, most of the evaluation criteria are qualitative. Even when a direct comparison is made between the developer's analysis and the baseline analysis, different numerical algorithms and assumptions used to interpolate data between measured values at known locations make it almost impossible to make a quantitative judgement as to which technical approach is superior. The comparisons, however, do permit an evaluation of whether the analysis is consistent with the data supplied for the analysis and therefore useful in supporting remediation decisions.

Summary of Analysis Performed by SitePro Version 3.0

SitePro provides environmental decision support through its integration of data from multiple sources (spreadsheet, drawing, and database files) into a platform that supports query operations, data manipulation, and visualization. SitePro places the

information into a visual context through two-dimensional (2-D) mapping of data and surface features. The software integrates the following fundamental tools into one application: data and database management, computer-aided design (CAD), geographic information systems (GIS) mapping, contouring, boring logs generation, geologic cross-section visualization, graphing, imaging, and reporting. SitePro allows analysts to manage and share their site data using a single file.

Environmental Software staff chose to use SitePro Version 3.0 to perform the visualization endpoint using data from Sites D, S, and T. The intent of the SitePro analyses was to demonstrate the capability to integrate large quantities of data into a visual framework for assistance in understanding a site's contamination problem. SitePro was not developed to address sample optimization or cost-benefit problems and did not attempt to perform these aspects of the test problems.

During the demonstration SitePro was used to import data from many different sources (drawing and data files) and integrate this into the SitePro platform. Database manipulations (sort and query), GIS operations (multiple layers on maps, hot-linking of the data to the maps), data analysis (creating contours of water level and contaminant concentration, geologic boring maps, and geologic cross-section maps), and visualization (2-D maps containing site features, contour levels, sample locations, and measured values) were demonstrated.

Section 2 contains a brief description of the capabilities of SitePro. Section 3 outlines the process followed in conducting the demonstration. This includes the approach used to develop the test problems, a summary description of the ten test problems, the approach used to perform the baseline analyses used for comparison with the developer's analyses, and the evaluation criteria. Section 4 presents a technical review of the analyses performed by of SitePro. This includes a detailed discussion of the problems attempted, comparisons of the SitePro analyses and the baseline results, and an evaluation of SitePro against the criteria established in Section 3. Section 5 presents an update on the SitePro technology and provides examples of representative applications of SitePro in environmental problem-solving.

Section 2 — SitePro Version 3.0 Technology Description

This section provides a general overview of the capabilities of Environmental Software's SitePro software product. The information was supplied by Environmental Software.

SitePro is a software application designed to support environmental professionals in comprehensively characterizing and managing environmental sites. SitePro integrates several fundamental tools that are used by environmental professionals into one application: data and database management, CAD, GIS, mapping, contouring, boring logs, cross sections, graphing, imaging, and reporting. This integration provides support for decisions pertaining to monitoring and remediation activities. SitePro is used to manage data on various environmental issues and media, including contaminated soil and water, air emissions, wastewater, and health and safety parameters. SitePro allows environmental professionals to manage and share their site data using a single file.

SitePro's specific features are categorized as mapping, site assessment and characterization, and data presentation and reporting. SitePro maps site and facility features using CAD and GIS features. Enhancement of an existing site map or site feature is accomplished with SitePro's map drawing and generation tools. SitePro supports field data entry and analysis to make field decisions. Users can generate contour maps during soil and groundwater investigations and soil-gas surveys. Hydrographs and chemical concentration curves can be generated with SitePro's graphing tool. Report-ready boring logs, well construction logs, and geologic cross-sections can be generated during site characterization activities to improve site understanding.

SitePro's database and report queries allow users to display and compare field data with historic data and maps. SitePro's customizable lookup tables allow users to easily modify the database and querying features. For instance, users can customize SitePro with specific chemicals, soil types, and monitoring parameters. Contours of parameters (e.g., contaminant concentration, hydraulic head) are generated and displayed on the map. Double-clicking on a map object (e.g., a monitoring well) hot-links the user to all the site data for the selected object. Map layers, objects, and labels can be turned on or off to focus presentation information. SitePro includes numerous features that automate data output material, such as tabular reports, graphs, maps, boring and well construction logs, and geologic cross-sections. SitePro's output may be printed or saved in an external file format.

SitePro's open, flexible, and scalable architecture includes data exchange tools that allow data import and export to more than 25 industry-standard map and data formats. The software integrates with Microsoft Office and Microsoft Exchange. Environmental Software has earned the status of a Microsoft Certified Solution Provider (MCSP).

The Microsoft Windows 95 and 98 and NT operating systems support SitePro software. System requirements include a 90-MHz Pentium processor with a minimum of 32 MB of RAM (48 MB RAM for NT workstations), 50 MB of available hard disk space, and a 1.44-MB floppy disk drive (CD-ROM drive preferred).

SitePro is part of a modular and integrated e3™ product line developed by Environmental Software. SitePro sells for $2,295 (U.S.) and is available directly from Environmental Software.[1] The company also offers an upgrade, maintenance, and service package for $275 per year that is required for new users.

Technical assistance is contained in SitePro's on-line help and user's guide. In addition, up-to-date information about current releases can be obtained from Environmental Software's web support (www.envsoft.com). Environmental Software offers a suite of training options, including two 1-day SitePro certification training classes (Introductory and Advanced). Environmental Software also offers consulting services at competitive rates for data management, data loading, customization, and software integration.

[1] Price listed is for the desktop version. Volume discounts are available, and server products are priced separately.

Section 3 — Demonstration Process and Design

Introduction

The objective of this demonstration was to conduct an independent evaluation of the capabilities of several DSSs in the following areas: (1) effectiveness in integrating data and models to produce information that supports decisions pertaining to environmental contamination problems, and (2) the information and approach used to support the analysis. Specifically, three endpoints were evaluated:

- *Visualization* — Visualization software was evaluated in terms of its ability to integrate site and contamination data in a coherent and accurate fashion that aids in understanding the contamination problem. Tools used in visualization can range from data display in graphical or contour form to integrating site maps and aerial photos into the results.

- *Sample optimization* — Sample optimization was evaluated for soil and groundwater contamination problems in terms of the software's ability to select the minimum number of samples needed to define a contaminated region with a specified level of confidence.

- *Cost-benefit analysis* — Cost-benefit analysis involved either defining the size of remediation zone as a function of the cleanup goal or evaluating the potential human health risk. For problems that defined the contamination zone, the cost could be evaluated in terms of the size of the zone, and cost-benefit analysis could be performed for different cleanup levels or different statistical confidence levels. For problems that calculated human health risk, the cost-benefit calculation would require computing the cost to remediate the contamination as a function of reduction in health risk.

Secondary evaluation objectives for this demonstration were to examine the reliability, resource requirements, range of applicability, and ease of operation of the DSS. The developers participated in this demonstration in order to highlight the range and utility of their software in addressing the three endpoints discussed above. Actual users might achieve results that are less reliable, as reliable, or more reliable than those achieved in this demonstration, depending on their expertise in using a given software to solve environmental problems.

Development of Test Problems

Test Problem Definition

A problem development team was formed to collect, prepare, and conduct the baseline analysis of the data. A large effort was initiated to collect data sets from actual sites with an extensive data collection history. Literature review and contact with different government agencies (EPA field offices, DOE, the U.S. Department of Defense, and the United States Geological Survey) identified ten different sites throughout the U.S which had the potential for developing test problems for the demonstration. The data from these ten sites were screened for completeness of data, range of environmental conditions covered, and potential for developing challenging and defensible test problems for the three endpoints of the demonstration. The objective of the screening was to obtain a set of problems that covered a wide range of contaminants (metals, organics, and radionuclides), site conditions, and source conditions (spills, continual slow release, and multiple releases over time). On the basis of this screening, six sites were selected for development of test problems. Of these six sites, four had sufficient information to provide multiple test problems. This provided a total of ten test problems for use in the demonstration.

Summary of Test Problems

A detailed description of the ten test problems was supplied to the developers as part of the demonstration (Sullivan, Armstrong, and Osleeb 1998). A general description of each of the problems can be found in Appendix A. This description includes the operating history of the site, the contaminants of concern, and the objectives of the test problem (e.g., define the volume over which the contaminant concentration exceeds 100 µg/L). The test problems analyzed by Environmental Software are discussed in Section 4 as part of the evaluation of SitePro's performance.

Table 1 summarizes the ten problems by site identifier, location of contamination (soil or groundwater), problem endpoints, and contaminants of concern. The visualization endpoint could be performed on all ten problems. In addition, there were four sample optimization problems, four cost-benefit problems, and two problems that combined sample optimization and cost-benefit issues. The range of contaminants considered included metals, volatile organic compounds (VOCs), and radionuclides. The range of environmental conditions included two- and three-dimensional soil and groundwater contamination problems over varying geologic, hydrologic, and environmental settings. Table 2 provides a summary of the types of data supplied with each problem.

Analysis of Test Problems

Prior to the demonstration, the demonstration technical team performed a quality control examination of all data sets and test problems. This involved reviewing database files for improper data (e.g., negative concentrations), removing information that was not necessary for the demonstration (e.g., site descriptors), and limiting the data to the contaminants, the region of the site, and the time frame covered by the test problems (e.g., only data from one year for three contaminants). For sample optimization problems, a limited data set was prepared for the developers as a starting point for the analysis. The remainder of the data were reserved to provide input concentrations to developers for their sample optimization analysis.

For cost-benefit problems, the analysts were provided with an extensive data set for each test problem with a few data points reserved for checking the DSS analysis. The data quality review also involved importing all graphics files (e.g., .dxf and .bmp) that contained information on surface structures such as buildings, roads, and water bodies to ensure that they were readable and useful for problem development. Many of the drawing files were prepared as ESRI shape files compatible with ArcView™. ArcView was also used to examine the graphics files.

Once the quality control evaluation was completed, the test problems were developed. The test problems were designed to be manageable within the time frame of the demonstration and were often a subset of the total data set. For example, in some cases, test problems were developed for a selected region of the site. In other cases, the database could have contained information for tens of contaminants, while the test problems themselves were limited to the three or four principal contaminants. At some sites, data were available over time periods exceeding 10 years. For the DSS test problems, the analysts were typically supplied chemical and hydrologic data for a few sampling periods.

Once the test problems were developed, the demonstration technical team conducted a complete analysis of each test problem. These analyses served as the baseline for evaluating results from the developers. Each analysis consisted of taking the

Table 1. Summary of test problems

Site identifier	Media	Problem endpoints	Contaminants
A	Groundwater	Visualization, sample optimization	Dichloroethene, trichloroethene
A	Groundwater	Visualization, cost-benefit	Perchloroethene, trichloroethane
B	Groundwater	Visualization, sample optimization, cost-benefit	Trichloroethene, vinyl-chloride, technetium-99
D	Groundwater	Visualization, sample optimization, cost-benefit	Dichloroethene, dichloroethane, trichloroethene, perchloroethene
N	Soil	Visualization, sample optimization	Arsenic, cadmium, chromium
N	Soil	Visualization, cost-benefit	Arsenic, cadmium, chromium
S	Groundwater	Visualization, sample optimization	Carbon tetrachloride
S	Groundwater	Visualization, cost-benefit	Chlordane
T	Soil	Visualization, sample optimization	Ethylene dibromide, dibromochloropropane, dichloropropane, carbon tetrachloride
T	Groundwater	Visualization, cost-benefit	Ethylene dibromide, dibromochloropropane, dichloropropane, carbon tetrachloride

Table 2. Data supplied for the test problems

Site history	Industrial operations, environmental settings, site descriptions
Surface structure	Road and building locations, topography, aerial photos
Sample locations	x, y, z coordinates for soil surface samples soil borings groundwater wells
Contaminants	Concentration data as a function of time and location (x, y, and z) for metals, inorganics, organics, radioactive contaminants
Geology	Soil boring profiles, bedrock stratigraphy
Hydrogeology	Hydraulic conductivities in each stratigraphic unit; hydraulic head measurements and locations
Transport parameters	Sorption coefficient (K_d), biodegradation rates, dispersion coefficients, porosity, bulk density
Human health risk	Exposure pathways and parameters, receptor location

entire data set and obtaining an estimate of the plume boundaries for the specified threshold contaminant concentrations and estimating the area of contamination above the specified thresholds for each contaminant.

The independent data analysis was performed using Surfer™ (Golden Software 1996). Surfer was selected for the task because it is a widely used, commercially available software package with the functionality necessary to examine the data. This functionality includes the ability to import drawing files to use as layers in the map, and the ability to interpolate data in two dimensions. Surfer has eight different interpolation methods, each of which can be customized by changing model parameters, to generate contours. These different contouring options were used to generate multiple views of the interpolated regions of contamination and hydrologic information. The best fit to the data was used as the baseline analysis. For three-dimensional (3-D) problems, the data were grouped by elevation to provide a series of 2-D slices of the problem. The distance between slices ranged between 5 and 10 ft depending on the availability of data. Compilation of vertical slices generated 3-D depictions of the data sets. Comparisons of the baseline analysis to the SitePro results are presented in Section 4.

In addition to Surfer, two other software packages were used to provide an independent analysis of the data and to provide an alternative representation for comparison with the Surfer results. The Geostatistical Software Library Version 2.0 (GSLIB) and Geostatistical Environmental Assessment Software Version 1.1 (Geo-EAS) were selected because both provide enhanced geostatistical routines that assist in data exploration and selection of modeling parameters to provide extensive evaluations of the data from a spatial context (Deutsch and Journel 1992; Englund and Sparks 1991). These three analyses provide multiple lines of reasoning, particularly for the test problems that involved geostatistics. The results from Surfer, GSLIB, and Geo-EAS were compared and contrasted to determine the best fit of the data, thus providing a more robust baseline analysis for comparison to the developers' results.

Under actual site conditions, uncertainties and natural variability make it impossible to define plume boundaries exactly. In these case studies, the baseline analyses serve as a guideline for evaluating the accuracy of the analyses prepared by the developers. Reasonable agreement should be obtained between the baseline and the developer's results. A discussion of the technical approaches and limitations to estimating physical properties at locations that are between data collection points is provided in Appendix B.

To minimize problems in evaluating the software associated with uncertainties in the data, the developers were required to perform an analysis of one problem from either Site N or Site S. For Site N, with over 5000 soil contamination data points, the baseline analysis reflected the actual site conditions closely; and if the developers performed an accurate analysis, the correlation between the two should be high. For Site S, the test problems used actual contamination data as the basis for developing a problem with a known solution. In both Site S problems, the data were modified to simulate a constant source term to the aquifer in which the

movement of the contaminant can be described by the classic advective-dispersive transport equation. Transport parameters were based on the actual data. These assumptions permitted release to the aquifer and subsequent transport to be represented by a partial differential equation that was solved analytically. This analytical solution could be used to determine the concentration at any point in the aquifer at any time. Therefore, the developer's results can be compared against calculated concentrations with known accuracy.

After completion of the development of the ten test problems, a predemonstration test was conducted. In the predemonstration, the developers were supplied with a problem taken from Site D that was similar to test problems for the demonstration. The objective of the predemonstration was to provide the developers with a sample problem with the level of complexity envisioned for the demonstration. In addition, the predemonstration allowed the developers to process data from a typical problem in advance of the demonstration and allowed the demonstration technical team to determine if any problems occurred during data transfer or because of problem definition. The results of the predemonstration were used to refine the problems used in the demonstration.

Preparation of Demonstration Plan

In conjunction with the development of the test problems, a demonstration plan (Sullivan and Armstrong 1998) was prepared to ensure that all aspects of the demonstration were documented and scientifically sound and that operational procedures were conducted within quality assurance (QA)/quality control (QC) specifications. The demonstration plan covered

- the roles and responsibilities of demonstration participants;
- the procedures governing demonstration activities such as data collection to define test problems and data preparation, analysis, and interpretation;
- the experimental design of the demonstration;
- the evaluation criteria against which the DSS would be judged; and
- QA and QC procedures for conducting the demonstration and for assessing the quality of the information generated from the demonstration.

All parties involved with implementation of the plan approved and signed the demonstration plan prior to the start of the demonstration.

Summary of Demonstration Activities

On September 14–25, 1998, the Site Characterization and Monitoring Technology Pilot, in cooperation with DOE's National Analytical Management Program, conducted a demonstration to verify the performance of five environmental DSS packages. The demonstration was conducted at the New Mexico Engineering Research Institute, Albuquerque, New Mexico. An additional software package was tested on October 26–29, 1998, at Brookhaven National Laboratory, Upton, New York.

The first morning of the demonstration was devoted to a brief presentation of the ten test problems, a discussion of the output requirements to be provided from the developers for evaluation, and transferring the data to the developers. The data from all ten test problems—along with a narrative that provided a description of the each site, the problems to be solved, the names of data files, structure of the data files, and a list of output requirements—were given to the developers. The developers were asked to address a minimum of three test problems for each software product.

Upon completion of the review of the ten test problems and the discussion of the outputs required from the developers, the developers received data sets for the problems by file transfer protocol (FTP) from a remote server or on a high-capacity removable disk. Developers downloaded the data sets to their own personal computers, which they had supplied for the demonstration. Once the data transfers of the test problems were complete and the technical team had verified that each developer had received the data sets intact, the developers were allowed to proceed with the analysis at their own pace. During the demonstration, the technical team observed the developers, answered questions, and provided data as requested by the developers for the sample optimization test problems. The developers were given 2 weeks to complete the analysis for the test problems that they selected.

The third day of the demonstration was visitors' day, an open house during which people interested in DSS could learn about the various products being tested. During the morning of visitors' day,

presenters from EPA, DOE, and the demonstration technical team outlined the format and content of the demonstration. This was followed by a presentation from the developers on the capabilities of their respective software products. In the afternoon, attendees were free to meet with the developers for a demonstration of the software products and further discussion.

Prior to leaving the test facility, the developers were required to provide the demonstration technical team with the final output files generated by their software. These output files were transferred by FTP to an anonymous server or copied to a zip drive or compact disk–read only memory (CD-ROM). The technical team verified that all files generated by the developers during the demonstration were provided and intact. The developers were given a 10-day period after the demonstration to provide a written narrative of the work that was performed and a discussion of their results.

Evaluation Criteria

One important objective of DSS is to integrate data and models to produce information that supports an environmental decision. Therefore, the overriding performance goal in this demonstration was to provide a credible analysis. The credibility of a software and computer analysis is built on four components:

- good data,
- adequate and reliable software,
- adequate conceptualization of the site, and
- well-executed problem analysis (van der Heijde and Kanzer 1997).

In this demonstration, substantial efforts were taken to evaluate the data and remove data of poor quality prior to presenting it to the developers. Therefore, the developers were directed to assume that the data were of good quality. The technical team provided the developers with detailed site maps and test problem instructions on the requested analysis and assisted in site conceptualization. Thus, the demonstration was primarily to test the adequacy of the software and the skills of the analyst. The developers operated their own software on their own computers throughout the demonstration.

Attempting to define and measure credibility makes this demonstration far different from most demonstrations in the ETV program in which measurement devices are evaluated. In the typical ETV demonstrations, quality can be measured in a quantitative and statistical manner. This is not true for DSS. While there are some quantitative measures, there are also many qualitative measures. The criteria for evaluating the DSS's ability to support a credible analysis are discussed below. In addition a number of secondary objectives, also discussed below, were used to evaluate the software. These included documentation of software, training and technical support, ease of use of the software, efficiency, and range of applicability.

Criteria for Assessing Decision Support

The developers were asked to use their software to answer questions pertaining to environmental contamination problems. For visualization tools, integration of geologic data, contaminant data, and site maps to define the contamination region at specified concentration levels was requested. For software tools that address sample optimization questions, the developers were asked to suggest optimum sampling locations, subject to constraints on the number of samples or on the confidence with which contamination concentrations were known. For software tools that address cost-benefit problems, the developers were asked either to define the volume (or area) of contamination and, if possible, supply the statistical confidence with which the estimate was made, or to estimate human health risks resulting from exposure to the contamination.

The criterion for evaluation was the credibility of the analyses to support the decision. This evaluation was based on several points, including

- documentation of the use of the models, input parameters, and assumptions;
- presentation of the results in a clear and consistent manner;
- comparison of model results with the data and baseline analyses;
- evaluation of the use of the models; and
- use of multiple lines of reasoning to support the decision.

The following sections provide more detail on each of these topics.

Documentation of the Analysis and Evaluation of the Technical Approach

The developers were requested to supply a concise description of the objectives of the analysis, the procedures used in the analysis, the conclusions of the analysis with technical justification of the conclusions, and a graphical display of the results of the analysis. Documentation of key input parameters and modeling assumptions was also requested. Guidance was provided on the quantity and type of information requested to perform the evaluation.

Based on observations obtained during the demonstration and the documentation supplied by the developers, the use of the models was evaluated and compared to standard practices. Issues in proper use of the models include selection of appropriate contouring parameters, spatial and temporal discretization, solution techniques, and parameter selection.

This evaluation was performed as a QA check to determine if standard practices were followed. This evaluation was useful in determining whether the cause of discrepancies between model projections and the data resulted from operator actions or from the model itself and was instrumental in understanding the role of the operator in obtaining quality results.

Comparison of Projected Results with the Data and Baseline Analysis

Quantitative comparisons between DSS-generated predictions and the data or baseline analyses were performed and evaluated. In addition, DSS-generated estimates of the mass and volume of contamination were compared to the baseline analyses to evaluate the ability of the software to determine the extent of contamination. For visualization and cost-benefit problems, developers were given a detailed data set for the test problem with only a few data points held back for checking the consistency of the analysis. For sample optimization problems, the developers were provided with a limited data set to begin the problem. In this case, the data not supplied to the developers were used for checking the accuracy of the sample optimization analysis. However, because of the inherent variability in environmental systems and the choice of different models and parameters by the analysts, quantitative measures of the accuracy of the analysis are difficult to obtain and defend. Therefore, qualitative evaluations of how well the model projections reproduced the trends in the data were also performed.

A major component of the analysis of environmental data sets involves predicting physical or chemical properties (contaminant concentrations, hydraulic head, thickness of a geologic layer, etc.) at locations between measured data. This process, called interpolation, is often critical in developing an understanding of the nature and extent of the environmental problem. The premise of interpolation is that the estimated value of a parameter is a weighted average of measured values around it. Different interpolation routines use different criteria to select the weights. Due to the importance of obtaining estimates of data between measured data points in many fields of science, a wide number of interpolation routines exist. Three classes of interpolation routines commonly used in environmental analysis are nearest neighbor, inverse distance, and kriging. These three classes of interpolation, and their strengths and limitations, are discussed in detail in Appendix B.

Use of Multiple Lines of Reasoning

Environmental decisions are often made with uncertainties because of an incomplete understanding of the problem and lack of information, time, and/or resources. Therefore, multiple lines of reasoning are valuable in obtaining a credible analysis. Multiple lines of reasoning may incorporate statistical analyses, which in addition to providing an answer, provide an estimate of the probability that the answer is correct. Multiple lines of reasoning may also incorporate alternative conceptual models or multiple simulations with different parameter sets. The DSS packages were evaluated on their capabilities to provide multiple lines of reasoning.

Secondary Evaluation Criteria

Documentation of Software

The software was evaluated in terms of its documentation. Complete documentation includes detailed instructions on how to use the software package, examples of verification tests performed with the software package, a discussion of all output files generated by the software package, a discussion of how the output files may be used by other programs (e.g., ability to be directly imported into an Excel spreadsheet), and an explanation of the theory behind the technical approach used in the software package.

Training and Technical Support

The developers were asked to list the necessary background knowledge necessary to successfully operate the software package (i.e., basic understanding of hydrology, geology, geostatistics, etc.) and the auxiliary software used by the software package (e.g., Excel). In addition, the operating systems (e.g., Unix, Windows NT) under which the DSS can be used was requested. A discussion of training, software documentation, and technical support provided by the developers was also required.

Ease of Use

Ease of use is one of the most important factors to users of computer software. Ease of use was evaluated by an examination of the software package's operation and on the basis of adequate on-line help, the availability of technical support, the flexibility to change input parameters and databases used by the software package, and the time required for an experienced user to set up the model and prepare the analysis (that is, input preparation time, time required to run the simulation, and time required to prepare graphical output).

The demonstration technical team observed the operation of each software product during the demonstration to assist in determining the ease of use. These observations documented operation and the technical skills required for operation. In addition, several members of the technical team were given a 4-hour tutorial by each developer on their respective software to gain an understanding of the training level required for software operation as well as the functionalities of each software.

Efficiency and Range of Applicability

Efficiency was evaluated on the basis of the resource requirements used to evaluate the test problems. This was assessed through the number of problems completed as a function of time required for the analysis and computing capabilities.

Range of applicability is defined as a measure of the software's ability to represent a wide range of environmental conditions and was evaluated through the range of conditions over which the software was tested and the number of problems analyzed.

Section 4 — SitePro Version 3.0 Evaluation

Description of Test Problems

Environmental Software's SitePro is a data integration and visualization tool. Environmental Software used SitePro to perform the visualization aspects on problems for Sites D, S, and T. SitePro was used to integrate site maps, surface structures, and hydrologic and contamination data to create a representation of the current status at the site. As part of the demonstration, several dozen visualization outputs were generated. A few examples that display the range of SitePro's capabilities and features are included in this report. A general description of each test problem and the analysis performed using SitePro follows. Detailed descriptions of all test problems are provided in Appendix A and in Sullivan, Armstrong, and Osleeb (1998).

Site D

The Site D problem was a 3-D groundwater contamination problem. The data supplied for the analysis of Site D included surface maps of buildings, roads, and water bodies; concentration data on four contaminants [perchloroethene (PCE), dichlorethene (DCE), trichloroethene (TCE), and trichloroethane (TCA)] in groundwater wells at different depths and locations for four quarters of one year; hydraulic head data; and geologic boring data. Environmental Software chose to evaluate only one contaminant (TCE), at only one threshold concentration (50 µg/L). SitePro was used to generate contours of water level and of the maximum measured TCE concentrations in four data sets representing sampling during the four quarters of a year. A site map with buildings, roads, and other surface features was included on the water level and TCE contour maps to provide a reference for the contaminant location.

Site S

This test problem was a 3-D groundwater contamination cost-benefit problem for a single contaminant (chlordane). The data consisted of measurements collected from a series of wells. In each well, chlordane concentrations were measured as a function of depth. Site S contained the most extensive and reliable data set for evaluating accuracy. To focus on accuracy, the problem was simplified by removing information regarding surface structures (e.g., buildings and roads) and by limiting the test problem to only one contaminant. The analyst was asked to define the region, mass, and volume of the plume at contamination concentrations of 5 and 500 µg/L. The analysis could be extended to include definition of the plume volumes as a function of three confidence levels, 10, 50, and 90%. SitePro estimated the areal extent of the plume by using the maximum concentration at each well location. This effectively reduced the problem to a 2-D problem. SitePro was used to generate the following output for this problem:

- hydraulic head contour map with well locations to provide reference;
- borehole log for one well;
- geologic cross-section map for a series of wells; and
- chlordane concentration contour profiles at 5 and 500 µg/L.

Site T

The Site T problem was a 2-D soil contamination sample optimization problem. The data supplied for analysis of this problem included surface drawings of buildings and roads and soil contamination data for four organic contaminants [ethylene dibromide (EDB), dibromochloroproprane (DBCP), dichloropropane (DCP), and carbon tetrachloride (CTC)]. This test problem was designed as a method for assessing the accuracy with which the software can be used to predict sample locations to define the extent of surface and subsurface soil contamination. The design objective was to generate a 3-D rendering of the soil contamination in two stages. In the first stage, the analysts were asked to develop a sampling strategy to define surface areas on the site in which the soil contamination exceeded the threshold concentrations given in Table 3 with confidence levels of 10, 50 and 90% on a 50 by 50 ft grid. In the second stage, after defining the region of surface contamination, the analysts were asked to define subsurface contamination in the regions found to be above the threshold at the 90% confidence limit. The problem definition required subsurface sampling locations on a 10-ft vertical scale to fully characterize the soil contamination at depths from 0 to 30 ft below ground surface (the approximate location of the aquifer). SitePro was used to generate contour maps of the surface areas on Site T that exceeded specified threshold concentrations for each

Table 3. Site T soil contamination threshold concentrations

Contaminant	Threshold concentration (: g/kg)
Ethylene dibromide (EDB)	21
Dichloropropane (DCP)	500
Dibromochloropropane (DBCP)	50
Carbon tetrachloride (CTC)	5

contaminant (see Appendix A). Buildings and surface features were included on the map to provide a frame of reference. SitePro also generated a single map containing all contaminants above the threshold concentrations. SitePro was used to estimate the surface area of soil contamination for each of the four contaminants by contouring the data provided. As noted earlier, SitePro cannot perform sample optimization. However, SitePro does have the ability to draw a fixed radius around each sampling point to provide a coverage map. Based on the coverage and contoured data, recommended sampling locations were provided. SitePro was used to generate the following outputs for this problem:

- contour maps of the areas exceeding the specified threshold concentrations for each contaminant;
- combination of all contaminants on a single map; and
- sample coverage map and proposed new sample locations.

Evaluation of SitePro Version 3.0

Decision Support

During the demonstration, it was observed that SitePro was able to quickly import and integrate data on contaminant concentrations, geologic structure, and surface structures from a variety of sources with different formats. SitePro was able to place this information in a visual context to support data interpretation.

Documentation of the SitePro Analysis and Evaluation of the Technical Approach

For each analysis, Environmental Software provided a step-by-step description of the manipulations necessary to import the data provided and to perform the required analyses. The steps proceeded logically and in a straightforward manner. Manipulations to format the data within the SitePro architecture were relatively simple. For example, the Site T data file (dbf, containing sample locations and measured contaminant concentrations) and a drawing file (dxf, containing site maps) were imported directly into SitePro's data management system. The SitePro database provided an integrated structure that was coupled with SitePro's analysis tools (e.g., contouring/mapping, cross sections, boring/well logs, graphing, and reporting). In addition, Site T data was hot-linked to the Site T map generated from the drawing files. These hot-links enabled the analyst to view the site map, click on the sample location, and obtain the desired database information. Another useful feature of the software was direct export of the output to standard, commercially available word processing software. Documentation of data transfer, manipulations (for example, how to treat contamination data as a function of depth in a well), and analyses were included. Model selection and parameters for contouring were also provided in the test problem documentation. Overall, the technical approach used by Environmental Software followed standard practices.

Comparison of SitePro Results with the Baseline Analysis

Site D

Environmental Software used SitePro to generate TCE contours for four different sampling periods based on the maximum observed concentration in each well during the sampling period. This problem was designed as a sample optimization problem with a limited initial data set. In addition, during any given sampling period, only a fraction of the wells were sampled. This limited data set did not lend itself to generating contours at the levels requested in the problem description. Therefore, Environmental Software elected to generate contours consistent with the available data. SitePro used a nearest neighbor search method in conjunction with an inverse distance weighting (IDW) method to create the contours. In each case, the exponent of 2 or 4 was used, depending on the sampling period (i.e., the weight used for interpolation was $1/D^2$ or $1/D^4$, where D is the distance from the interpolation point to the measured data). Figure 1 displays the SitePro TCE contours for the fourth sampling period

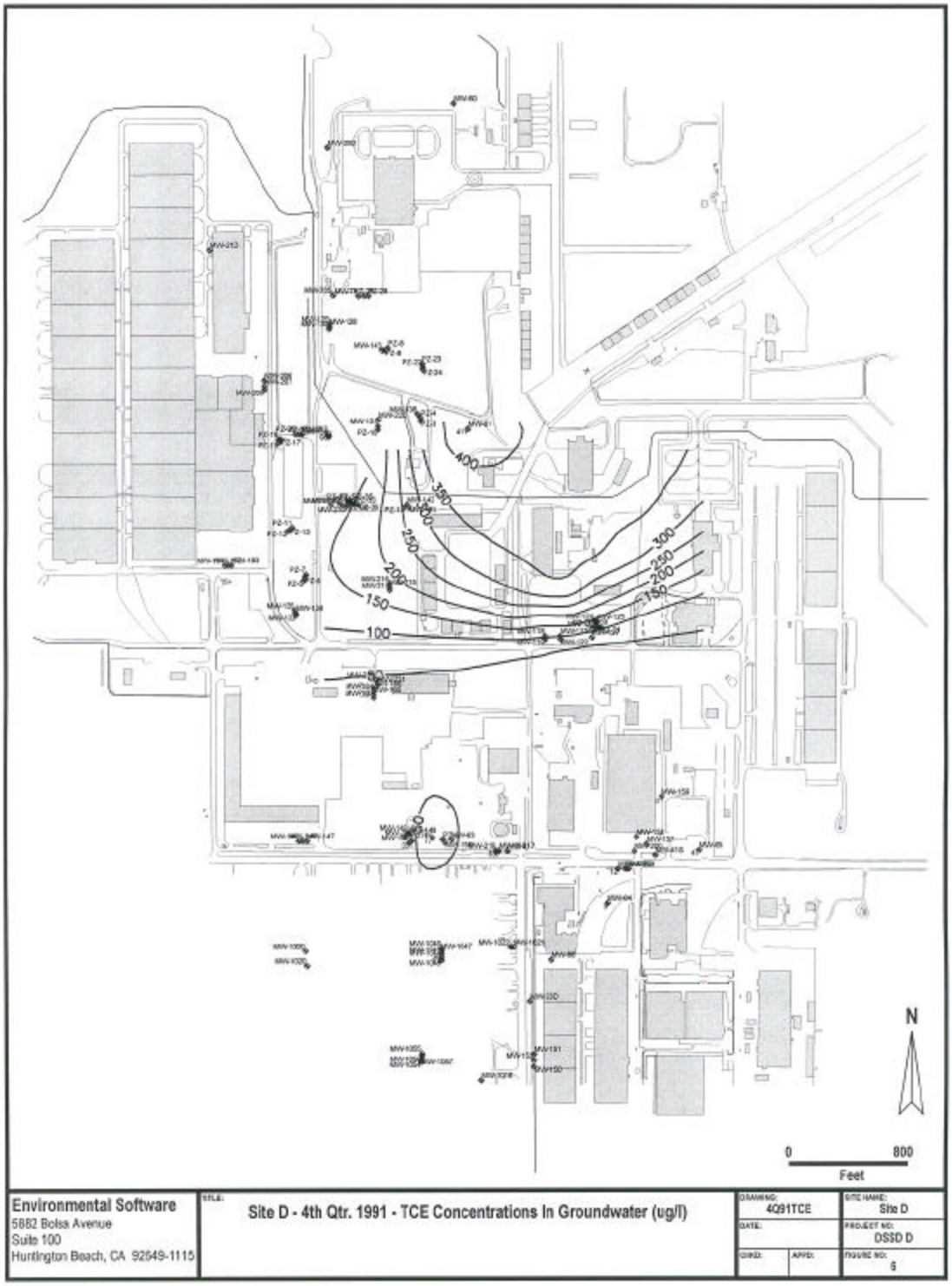


Figure 1. Fourth-quarter Site D TCE contours generated by SitePro.

on a base map including buildings and roads. Identifiers for every well and the measured concentrations for the ten wells used for generating the contours are posted on the map. Although these are not easy to see due to the small scale of the figure, Environmental Software provided large-format output (22 by 36 in.) of this map in which these details were clearly visible.

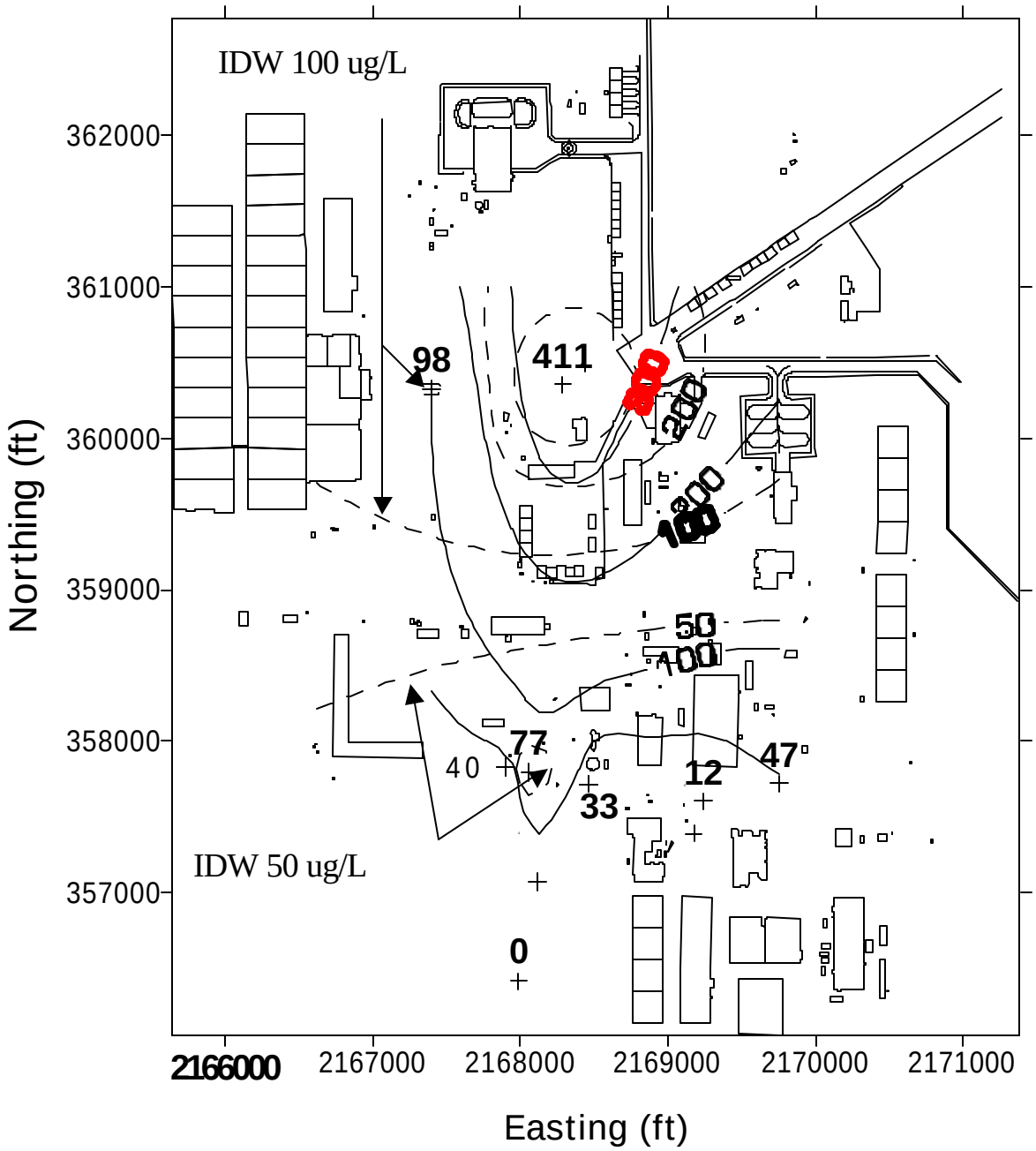
The SitePro interpolations were found to be generally consistent with the data for Site D. However, there were differences between the baseline and the SitePro analyses (see Figure 2). To create Figure 2, Surfer was used to interpolate, through kriging (see Appendix B), the same data set given to Environmental Software to provide the baseline contour (solid line in Figure 2). Then the SitePro analysis was recreated using Surfer, and the same contour parameters (IDW) used by SitePro (dashed line). In addition, the measured TCE values (in units of µg/L) from the baseline data were plotted on the figure. From Figure 2, it can be seen that the SitePro contours are generally consistent with the baseline data. A small circle representing the 50-µg/L contour surrounds the well with the measured value of 77 µg/L. A larger circle, representing the 300-µg/L contour (the maximum contour shown in the figure), surrounds the well with a measured value of 411 µg/L. The well with the measured value of 98 µg/L appears surrounded by the small region of the 100-µg/L contour. This 100-µg/L contour excludes the measured value of 98 µg/L, consistent with the data. However, the main branch of the 100-µg/L contour indicates contamination far to the south of this point.

While the SitePro analysis is generally consistent with the baseline data, the baseline kriging analysis performed by the demonstration technical team, shown in Figure 2, provided a better match to the baseline data. The baseline kriging analysis had predicted higher concentrations between the peak value of 411 µg/L and the value of 77 µg/L approximately 3000 ft farther south. This is consistent with measured data provided for other sampling periods between these two points. Also, the baseline analysis indicates the 50-µg/L contour is near a series of wells that are in the same area (i.e., wells near a northing of 358000) that had measured TCE values of 40, 77, 40, 33, 12, and 47 µg/L. The SitePro analysis had the 50-µg/L contour farther to the north. The cause for the discrepancy between the baseline and SitePro TCE concentration contours was the choice of contouring algorithm. Environmental Software used an IDW algorithm with the distance weighted to the second power. The poor match between the Site D baseline contours and the SitePro-generated contours indicates the limitations in the SitePro contouring algorithm used in the analysis. The impact of this discrepancy on decision making depends on the exact application and the decision that needs to be made. For example, if the objective of the analysis is to determine the general trends in the data and visualize the location of the contamination relative to site features, the SitePro analysis may be adequate. If the objective of the analysis is to determine the location of the 50-µg/L contour, the SitePro analysis may not be adequate.

Environmental Software also used SitePro to generate a contour map of water levels measured at the site (Figure 3). It was found that the SitePro evaluation was generally consistent with the data and with the baseline analysis performed using kriging.

Site S

SitePro was used to generate a contour map of the Site S water levels. There were some discrepancies between SitePro results and the baseline data. Figure 4 shows the water-level contours generated by SitePro using IDW with a weight of 2, the baseline analysis generated by the Surfer software package using kriging, and the locations of the wells with measured water levels. As Figure 4 indicates, both methods produce similar results that are consistent with the data when the distance between measurement points (wells) is short. This occurs in the upper half of the map (water-level contours of 34, 35, and 36 ft). When the distance between wells increases, the contours generated by SitePro do not match the baseline data or analysis. This can be seen at the water-level contour of 33 ft (Figure 4). SitePro also obtained a poor fit to the data at the upper part of the domain (water level 36.5 ft). However, when the demonstration technical team generated the water-level contours using another software program, Surfer, with the same contouring algorithms and parameters used by SitePro, the results were consistent with those produced by the analyst of Environmental Software using SitePro. SitePro used IDW algorithms for generating contour maps of spatial data. SitePro has other contouring algorithms (nearest neighbor, dip projection where the slope of the data is weighted using inverse distance weights, and linear interpolation) that could



SitePro contour (IDW) ---------
Baseline contour (kriging) ———
TCE measured value (µg/L) + 40
Contour levels are at 300, 200, 100, and 50 : g/L

Figure 2. Comparison of SitePro analysis (dashed line), baseline analysis (red solid line), and the baseline data (posted near crosses) for Site D TCE fourth-quarter concentration contours.

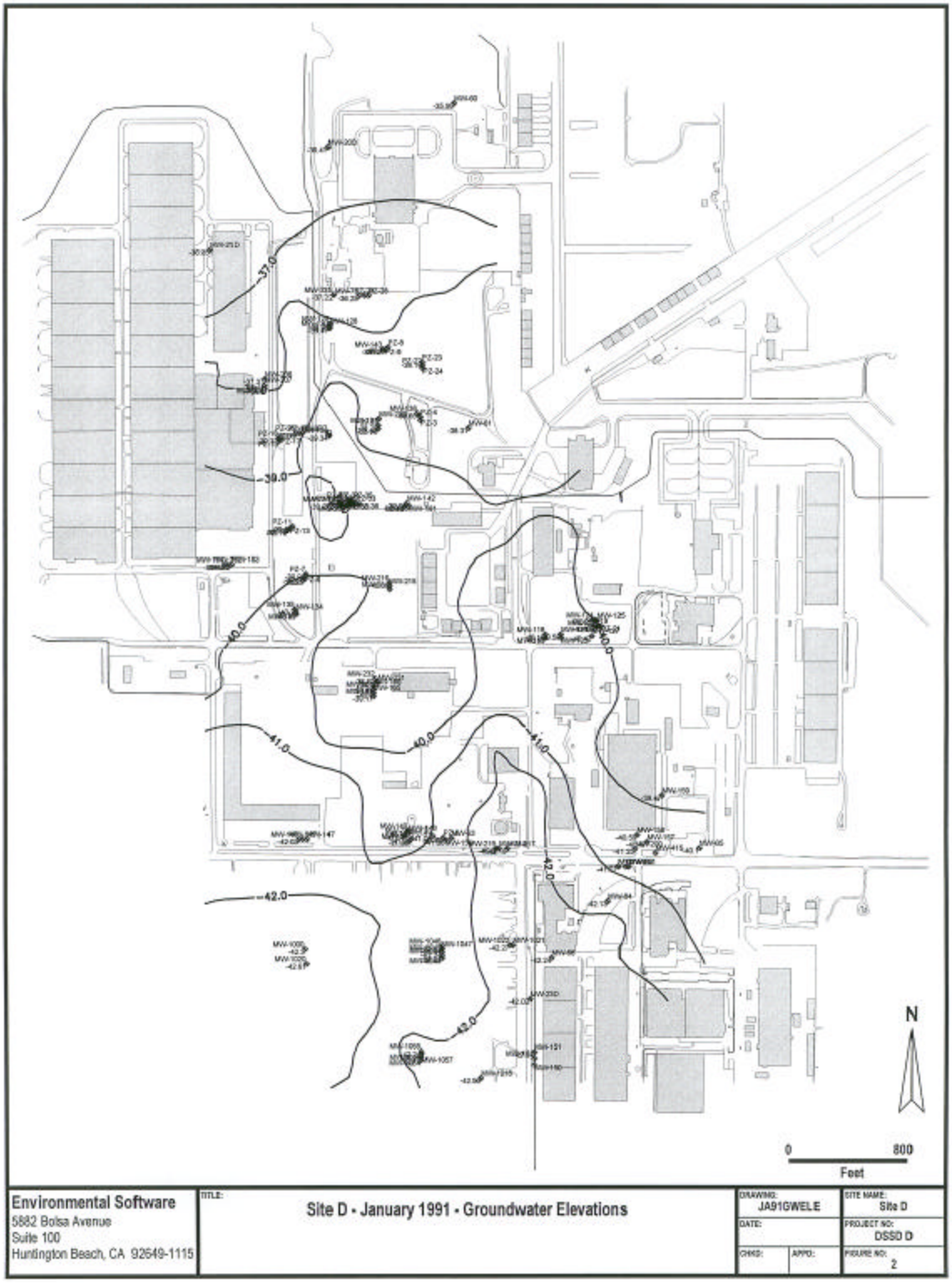


Figure 3. SitePro-generated contours for hydraulic head levels at Site D.

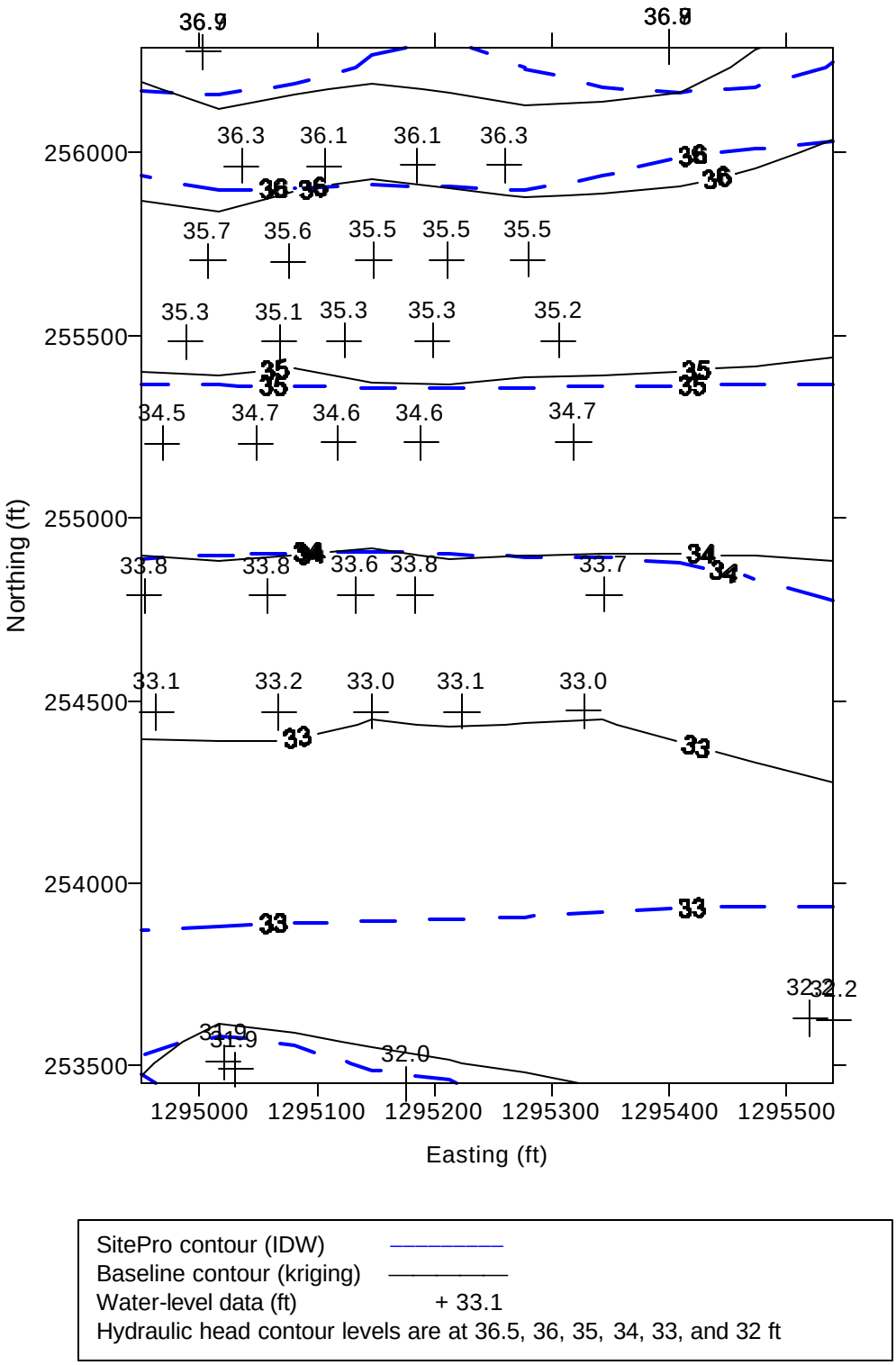


Figure 4. Comparison of SitePro analysis (dashed line), baseline analysis (solid line), and baseline data (posted above the crosses) for Site S water levels for the cost-benefit problems

have been used for data interpretation. SitePro does not contain kriging-contouring algorithms that are often used to interpolate environmental data sets. The results from the Site S problem indicate that the SitePro contouring algorithm choices are limited and may not always produce results that are consistent with the data.

The Site S contamination data set consisted of chlordane concentration values provided on a 5-ft vertical spacing at a series of wells. This provided a 3-D description of contamination. Environmental Software used SitePro to query the database to select the maximum concentration in each well for use in generating a contour map of the areal extent of chlordane contamination at the 5- and 500-µg/L level (Figure 5). On this map, the well identifier and the maximum value in each well was plotted; the resulting concentration contours were consistent with the baseline data. Figure 5 was provided as a printed document that was read by an optical scanner and placed into this report. The electronic version of this file was provided in .dxf and could not be read by other software products designed to read this type of file (Surfer, CorelDraw, and MapInfo).

The contamination data for Site S were generated using a simulation model with a constant source of contaminants supplied to the aquifer. Hydrologic and chlordane transport parameters were obtained from field data and were taken to be constant over the problem domain. These assumptions permitted release and transport through the aquifer to be represented by a partial differential equation that was solved analytically. The analytical solution provides an exact reference for chlordane concentrations to evaluate the SitePro results. The results of the analytical solution are plotted on Figure 6. For 500-µg/L contour, the analytical solution has a length of approximately 1000 ft and a maximum width of 100 ft. The analytical 500-µg/L contour reaches approximately 100 ft farther south than the transect defined by wells DP-121 through DP-125 (Figures 5 and 6). This is approximately 300 ft farther south than the contour projected by SitePro. The cause for this discrepancy is the lack of precise knowledge of the data at every location in space. For the transect defined by wells DP-121 through DP-125, the analytical solution gives a peak concentration of 558 µg/L on the centerline. In the data supplied to the analysts, DP-123 is 21 ft off of the centerline and has a peak concentration of 484 µg/L. Therefore, as shown in Figure 5, the SitePro 500-µg/L contour does not extend this far. In general, the SitePro analysis was consistent with the data provided but not an exact match to the known analytical solution. This was a result of incomplete knowledge about the system and not a limitation of SitePro.

The demonstration technical team also performed a baseline analysis with the same data set supplied to Environmental Software; however, kriging with an anisotropy ratio of 0.3 was used (Figure 6). The baseline analysis provided essentially the same results as the SitePro analysis depicted in Figure 5. This was expected because the baseline data bounds the 5- and 500-µg/L contours on each side of the plume. The minor differences between the SitePro and baseline analyses at the leading edge of the 5- and 500-µg/L contours, with the baseline analysis predicting a slightly larger contour area, were due to the choice of interpolation parameters used to generate the baseline contours.

SitePro was able to accurately generate standard boring logs from data for Site S (e.g., Figure 7). At MW-245a there was a surface soil layer 5 ft thick underlain by a 6-ft-thick cobble layer. The surface layer was underlain by a 185-ft-thick layer of poorly graded sands. A confining clay layer 23 ft thick separated the poorly graded sand from a silty sand 145 ft thick. The depth of the well was 355.8 ft, as shown in Figure 7. In this instance, accuracy was judged through comparison of the location of the soil layers as depicted by SitePro and the data. SitePro provided an exact match with baseline data. In this example the soil classification designators are sometimes difficult to read and a legend, which would have been useful in interpreting the figure, was not provided.

SitePro was able to accurately generate geologic cross-section maps for Site S (Figure 8). Accuracy was judged by comparison of the SitePro representation in the cross-section map with the data. SitePro provided an exact match with baseline data. Site S exhibited a few thin surficial layers underlain by a thick layer of poorly graded sands. A thin clay layer ranging in thickness from 10 to 25 ft underlies the upper sand layer. A thick silty sand layer was underneath the clay layer. This is clearly depicted in the geologic cross-section generated by SitePro (Figure 8). The quality of the text on this figure could have been clearer to improve readability.

Figure 5. Site S chlordane 5 (solid line) and 500-µg/L (shaded) contours as generated by Site Pro.

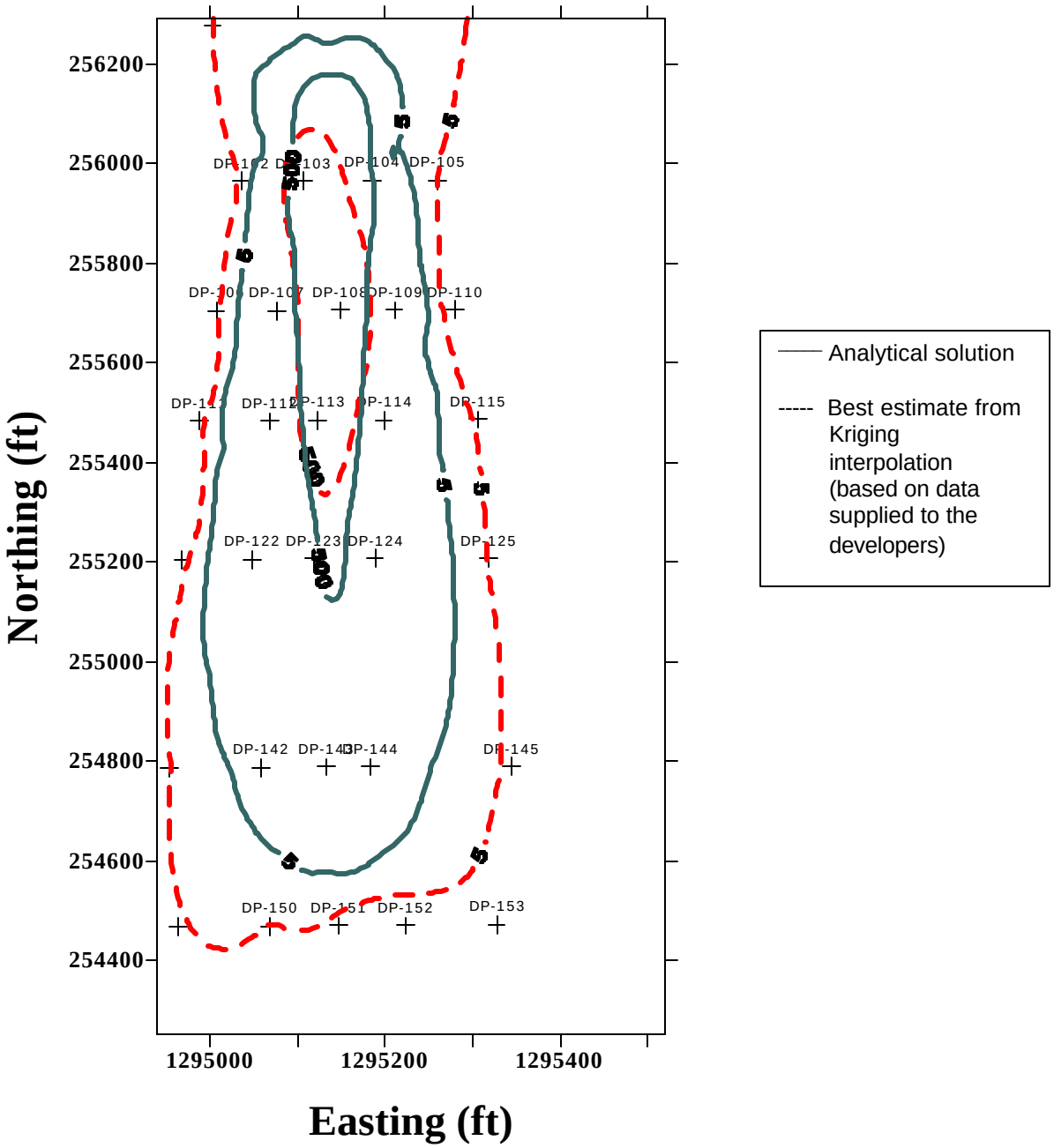


Figure 6. Site S chlordane 5 and 500-µg/L contours for the analytical solution (solid lines) and the baseline contour (dashed line) obtained using kriging with an anisotropy ratio of 0.3. Data locations are posted (+) and labeled on the map.

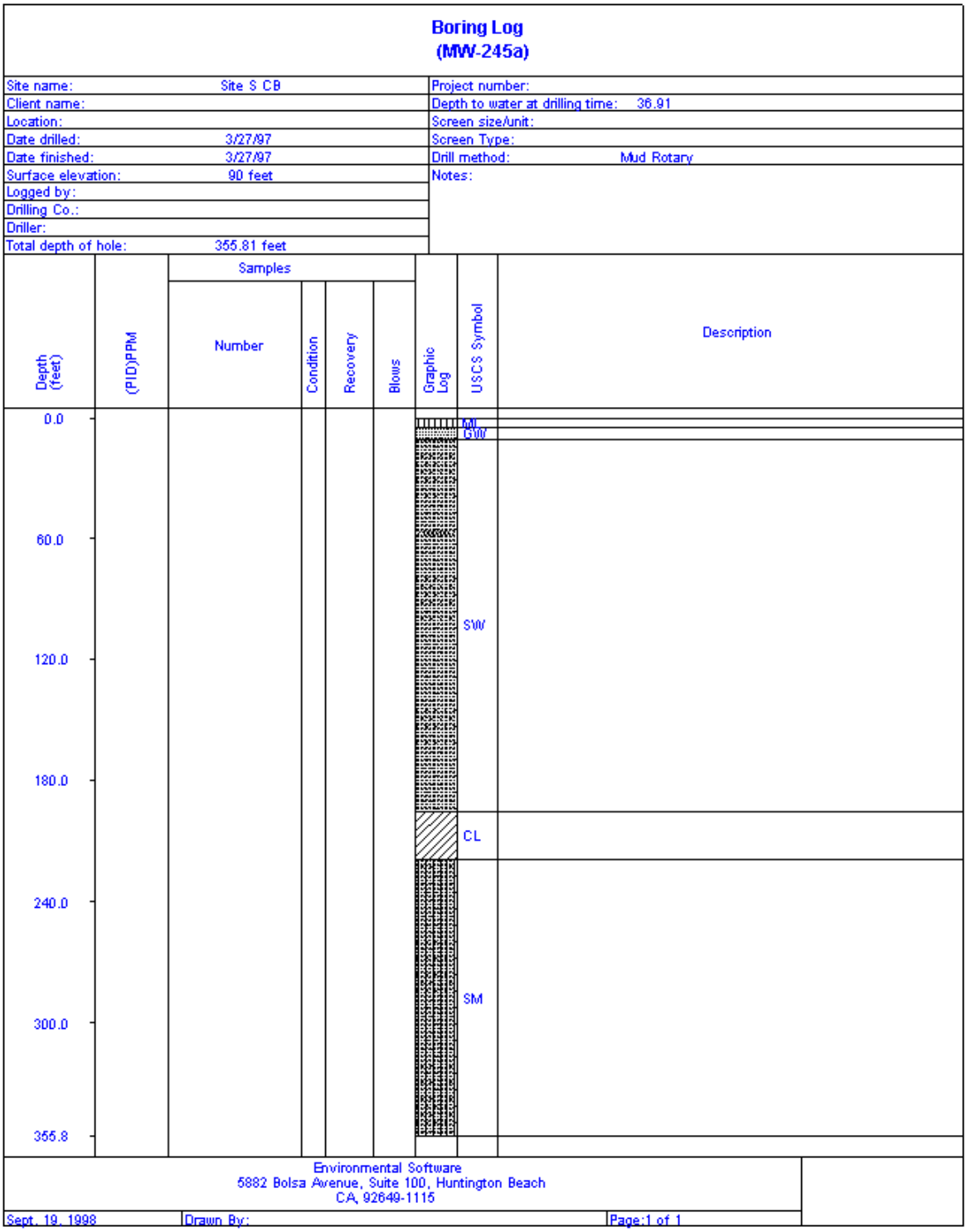


Figure 7. Site S borehole log generated by SitePro for monitoring well MW-245a.

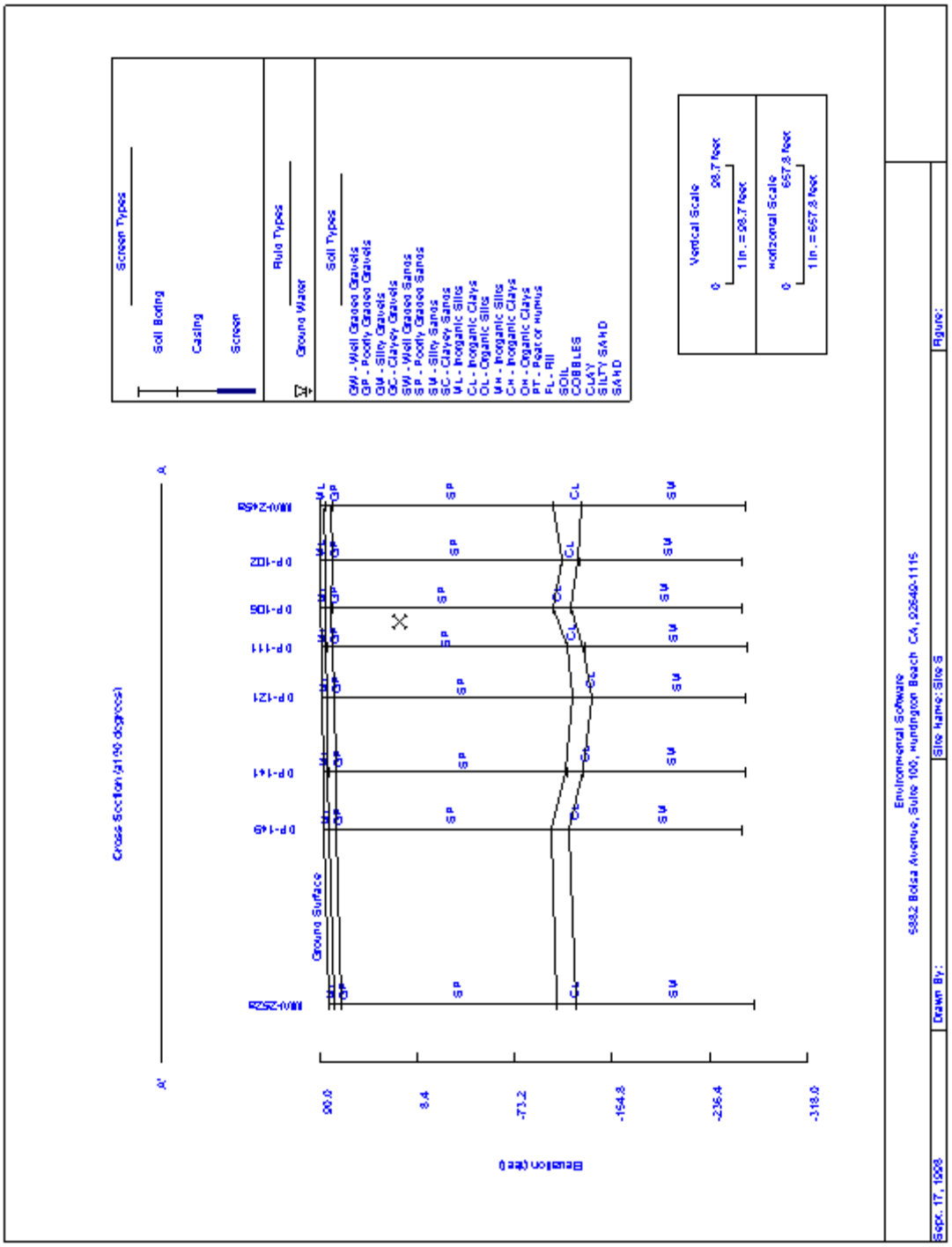


Figure 8. Geologic cross-section map for Site S generated by SitePro.

Site T

SitePro was used to place boundaries on the regions of surface soil contamination for each of the four contaminants by contouring the data. Estimates of the total area of contamination were not provided; however, the contours generated by SitePro were compared with the baseline analysis and were consistent with the baseline data. Figure 9 is a screen capture from SitePro that shows the site map, including buildings, sample locations, sample identifiers, measured EDB concentration, and surface soil EDB concentration contours. The display is confusing because of the amount of information supplied. However, the operator of SitePro can easily turn off (or on) what will be displayed in the figure. The figure shows the contamination localized to the northwest corner of the site (upper left-hand corner) near a building in which the chemicals were handled. This depiction was consistent with the baseline data. Figure 10 shows the site map, sample locations, and contours above the Table 3 thresholds for all four contaminants. The contaminants are color-coded on the map to permit differentiation between them. The map permits a rapid comparison of whether all contaminants originate from the same source area. Three contaminants—EDB, DCP, and DBCP—have a similar profile and appear to emanate from the same source. The fourth contaminant, CTC, has a high concentration at one sample location in the north-central part of the site (upper central section of the figure). The depictions of the four contaminants were consistent with baseline data and the baseline analysis of the data. The CTC contours extend off-site because of high CTC concentrations measured at one location, the contouring algorithm, and the lack of bounding sampling locations in that area.

The soil contamination problem for Site T was designed as a sample optimization problem. Typically, sample optimization methods use statistical analysis of the existing data to predict the optimum location for collecting additional data while minimizing uncertainties in the analysis. SitePro does not have statistical routines for performing sample optimization. However, the software was used to draw a fixed radius around every sampling point to provide a coverage map. Based on the coverage and the contoured data, the analyst used his judgment to recommend five sampling locations, as shown in Figure 11. Existing sample locations are circled in red (25-ft radius) and black (50-ft radius). Proposed sampling locations in Figure 11 are circled in blue and designated as prop-01 through prop-05. The proposed locations can be found surrounding the area of high contamination near the edge of the building in the northwest corner of the site and near the high CTC concentration area

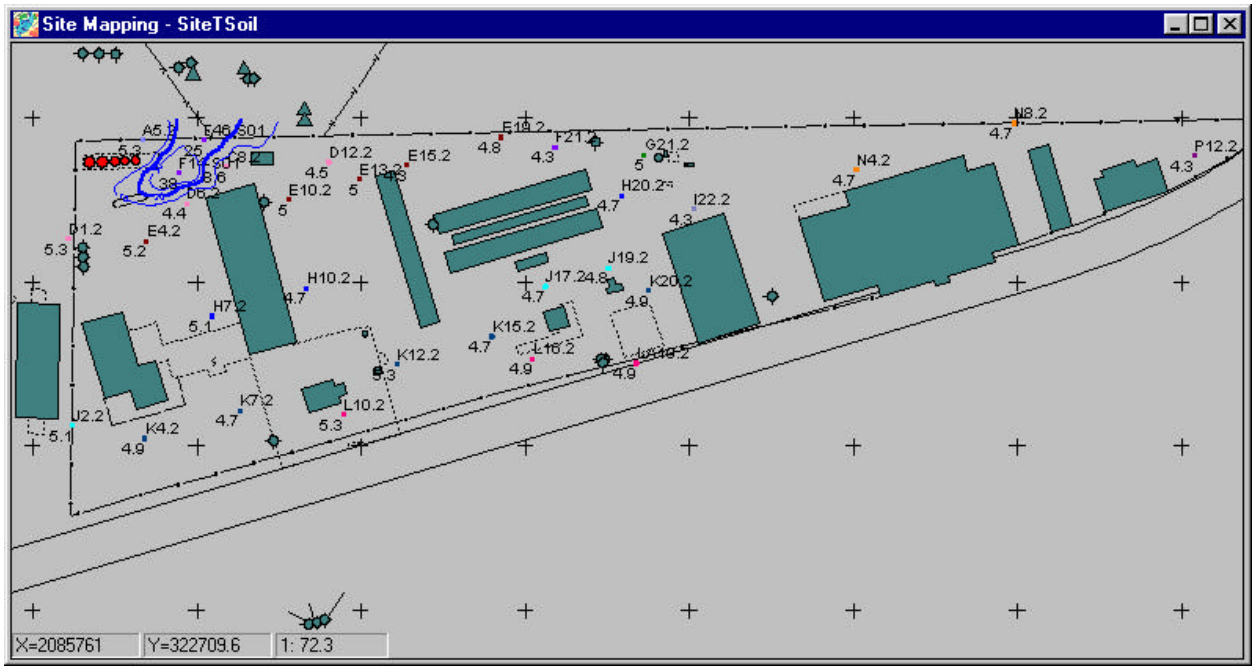


Figure 9. Site T ethylene dibromide contours above threshold concentration of 21 µg/L.

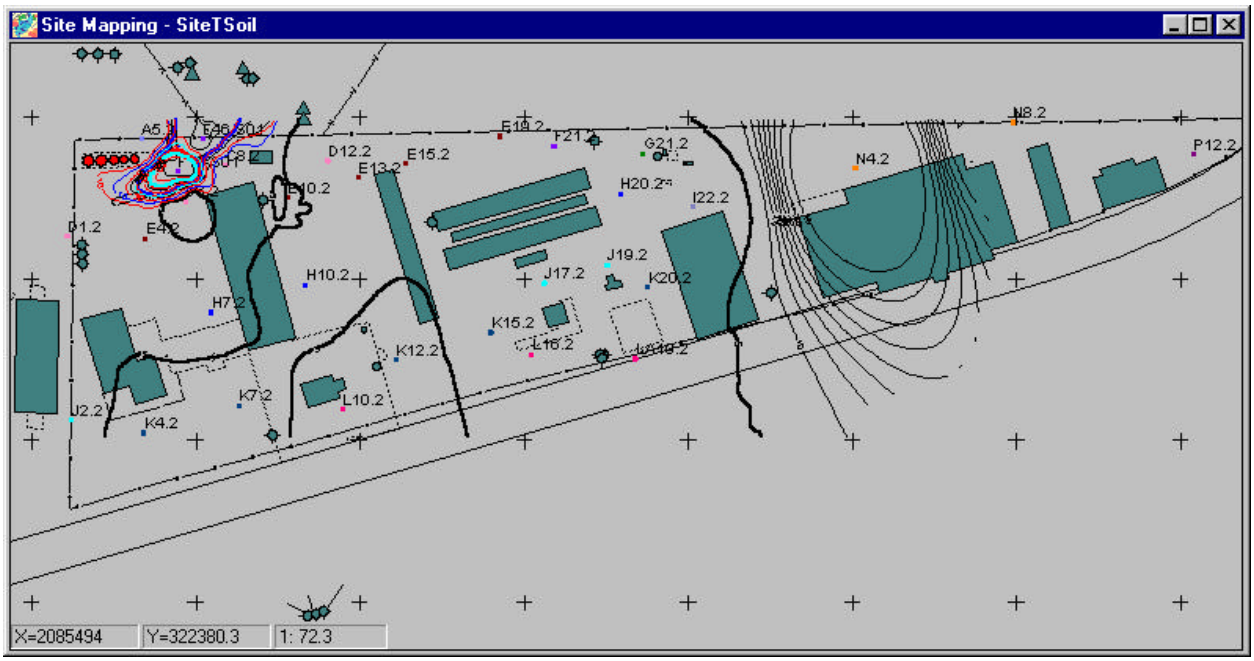


Figure 10. Site T contours for all four chemicals (CTC — black, EDB — blue, DCP — red, and DBCP — light blue) above their threshold concentrations.

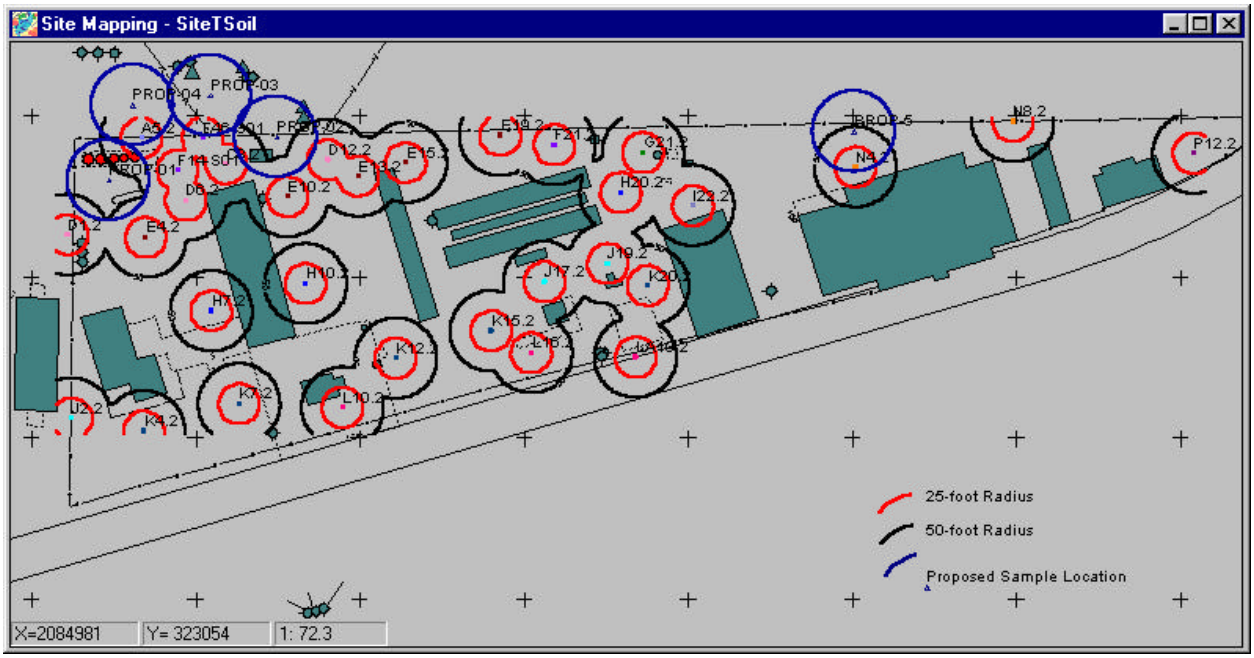


Figure 11. Site T area sampling coverage map with proposed new sampling locations.

in the north-central portion of the site. The legibility of the figure could have been improved by more clearly distinguishing the proposed sample locations. Environmental Software did not request data at these locations to further complete the sample optimization portion of the problem. Therefore, it is not possible to evaluate the technical accuracy of the predicted sample locations. It is possible to state that the SitePro contouring was consistent with the baseline data and recommended sample locations are reasonable.

Multiple Lines of Reasoning

Environmental Software chose not to use SitePro to provide multiple interpretations of the data with different modeling parameters. SitePro has several contouring algorithms, but only one contouring algorithm (IDW) was used in the demonstration. In addition, different parameters could have been used in the IDW algorithm to explore the data. Performing multiple interpolations of the data using different interpolation routines and parameters would have provided multiple views of the data that generally assist in data interpretation.

Secondary Evaluation Criteria

Ease of Use

The analysis team found that SitePro was easy to use. It has a graphical user interface (GUI) with a logical structure to permit use of the options in the software. It has a flexible database structure that supports multiple data input formats. During the demonstration SitePro was able to incorporate drawing exchange format (.dxf) files containing surface features and .dbf files containing data on contamination, hydrology, and geologic structure. The GUI provides a platform to address problems efficiently and to tailor data formatting to the problem under study. The GIS features in SitePro permit use of multiple layers in generating maps (e.g., multiple contaminants on a single map) and permit rapid viewing of the data at the sample locations (e.g., the database is hot-linked to the maps presented in the GUI). The software contains an extensive chemical identification database, which permits comparisons to ensure matching between the chemical name and chemical identifiers (e.g., CAS number). The database structure permits queries on a wide range of fields (e.g., chemical name, date, concentration, well identifiers) and also allows filtering (e.g., include only data between certain dates and maximum concentration at a location over a range of dates; use only data with appropriate data qualifiers).

Reviewers from the technical team received 4 hours of training. The reviewers agreed that SitePro appeared to be easy to use and that minimal training would be necessary to use the basic features of SitePro.

SitePro can export text and graphics to many different formats, including CAD formats (.dxf), GIS formats [Shape files (.shp)], and standard word processing software (.jpg and ASCII text). It generates .bmp and .dxf visualization files that can be read by a large number of software products. It can be used to generate large-scale maps (2 by 3 ft) to enhance visual clarity of the output. However, one .dxf file created by SitePro (Figure 5) was unable to be read by a number of other software programs. The three other .dxf files generated by SitePro were readable.

Efficiency and Range of Applicability

During the demonstration, Environmental Software provided one staff member for one week and an additional staff person for two days. During this time, both spent one full day on demonstrating the software to visitors. Including completion of the report documenting their efforts, Environmental Software used a total of approximately five staff-days to evaluate the three problems. The demonstration showed that the software was capable of rapidly importing data to perform an evaluation.

Training and Technical Support

Environmental Software provides a number of options for SitePro training and technical support, as follows:

- Two one-day training courses: an introductory course aimed at teaching the most commonly used features of SitePro and a more advanced course that teaches more sophisticated database query operations and the generation of boring logs and cross-section maps. These courses cost $900 for up to six people. An additional charge of $65 per person covers course supplies and materials.
- Technical support at $275 per year
- An extensive on-line help manual
- Tutorial case studies provided with SitePro

The users' manual provided with SitePro provides detailed instructions on how to operate the software. The manual is well-organized, with ten chapters that cover the features of SitePro, including importing and exporting data and creating maps, graphs, and reports. There are also illustrations of the pull-down menus that the operator sees when using SitePro.

Additional Information about the SitePro Software

To use SitePro efficiently the operator should have a basic understanding of the use of computer software in analyzing environmental problems. This includes a fundamental knowledge of GIS, CAD, and database files. In addition, knowledge about contouring environmental data sets is beneficial.

During the demonstration, SitePro Version 3.0 was operated on a Windows 95 operating system. Two machines were used during the demonstration. One machine was a laptop with a 120-MHz Pentium processor, 40 MB of RAM, and a 1.2-GB hard drive. Problems with the mouse on the laptop, unrelated to the use of SitePro, required the rental of a second machine. The second machine had a 266-MHz Pentium processor, 64 MB of RAM, a 2.0-GB hard drive, and 1 MB of video memory for a 19-inch monitor with an 800 H 600 resolution. The minimum requirements to operate SitePro are a 90-MHz Pentium processor, 50 MB of storage memory, and 32 MB of RAM.

SitePro costs $2295 for a single license. Educational and multiple license discounts are available. In addition, new clients are required to purchase one year of technical support at a cost of $275.

Summary of Performance

A summary of SitePro's performance is presented in Table 4. Overall, the technical team observed that the main strength of SitePro was its ability to easily integrate data and maps in a single platform, allowing spatial visualization of the data. The GUI platform appeared to be easy to use and had logical pull-down menus and on-line help. SitePro supported a wide range of formats for importing and exporting data, including CAD files (.dxf), GIS files (.shp), graphics files (.bmp and .jpg), and data files (.dbf, ASCII text). The main limitation of SitePro was that contour maps used to represent and visualize environmental parameters (concentration, hydraulic head, etc.) did not always provide an acceptable match to the measured data. Other minor limitations of SitePro that were noted in the demonstration had to do with the legibility of the geologic cross-section and boring maps and the transferability of one .dxf format file to other software programs. These might have been correctable through more operator intervention (e.g., changing font size, etc.), but this capability was not demonstrated.

Table 4. SitePro Version 3.0 performance summary

Feature/parameter	Performance summary
Decision support	SitePro integrated data, site maps, and surface features into a 2-D spatial representation. Query and sort capabilities permitted investigation of the data against threshold concentrations.
Documentation of analysis	Documentation of the process and parameters were provided and assumptions explained. Model parameters, queries, and maps were exported to word processing files to document the analysis. Graphical output was prepared in bitmap (.bmp) and drawing exchange format (.dxf).
Comparison with baseline analysis and data	Borehole logs accurately matched data. Geologic cross section accurately matched data. Concentration and hydrologic head contours were consistent with the IDW model assumptions. However, the IDW contouring algorithm was shown to produce a poor fit to the data under certain conditions. Accurately mapped wells, buildings, and site features. Accurately posted data to sample locations. Generated sample coverage map based on sample locations. Hot-linked data to well locations.
Multiple lines of reasoning	Not demonstrated
Ease of use	User-friendly, logical layout of menus, and query capabilities. Imported and exported files in many different formats (CAD, GIS, graphics, database, ASCII). One of four .dxf files produced by SitePro could not be read by other dxf readers.
Efficiency	Three problems completed with one man-week effort.
Range of applicability	SitePro contains an extensive database on chemicals and their properties and visualizes, in two dimensions, soil and groundwater contamination problems. Query capabilities permit flexibility in the analysis to handle a wide range of conditions. The flexible database structure permits tailoring the analysis to site-specific problems.
Training and technical support	User's manual Documentation of contouring routines On-line help Tutorials (test cases) for training Two one-day training courses (introductory and advanced) One-day training course on network administration of SitePro Upgrade, maintenance, and technical support for $275/year
Operator skill base	Basic knowledge about environmental data and geographical information systems (GIS), computer-aided design (CAD), database files and contouring
Operating system	Windows 95
Cost	$2295

Section 5 — SitePro Version 3.0 Update and Representative Applications

Objective

The purpose of this section is to allow Environmental Software to provide information regarding new developments with SitePro since the demonstration activities. In addition, the developer has provided a list of representative applications in which its technology has been or is currently being used.

Technology Update

Environmental Software released SitePro 4.0 and EMIServer 1.0 as part of its e3 product line in June 1999. The e3 product suite is a complete environmental, safety, and health (ES&H) solution that brings a client-server, intranet ready environmental management information system (EMIS) to environmental professionals. The component-based architecture of the e3 suite delivers domain ES&H solutions with information technology features essential for successful enterprise ES&H implementation. SitePro 4.0 is the first of many client applications that will work in conjunction with EMIServer.

SitePro 4.0 is a client application that facilitates the assessment, remediation, and monitoring of contaminated soil and water. Environmental engineers, technical managers, consultants, and government regulators can use SitePro's integrated database, GIS, CAD, mapping, and imaging tools for spatial and temporal visualization of contaminants in water, soil, and groundwater. SitePro can output maps, drawings, charts, and reports that summarize environmental history and status at a site. EMIServer 1.0 is the relational database that stores data for SitePro 4.0 and other e3 client applications. New features of the SitePro 4.0 release include the following:

- New database: Microsoft Structured Query Language (SQL) Server 7.0-based EMIServer
- Fully client/server
- New Microsoft Outlook style user interface
- New GIS features, including image registration
- New CAD engine (TurboCAD)
- New boring logs
- Customizable
- Flexible log templates
- Fully automated
- New cross sections
- Fences or cross sections
- Many optional annotations: dense nonaqueous phase liquid, light nonaqueous phase liquid, concentrations, etc.
- Full automation
- New generalized database query tool for all reports, maps, and graphs
- Provides access to all data
- Allows users to add computations to data queries
- Allows users to save/reuse queries
- New graphing tool that handles dates, multiple axes
- New power data manager provides spreadsheet editing capability
- New import feature that provides pre-loading data check and status reports
- New reporting tool that provides cleaner, more user-customizable reports
- New site summary
- Intranet-enabled access to EMIServer
- Browser-enabled via Microsoft Internet Explorer
- Dynamically customized tree-structured data views
- Complements sibling clients, including SiteBrowser (Q1/99) and SiteAnalyst (Q2/99)

The basic machine system requirements for SitePro Version 4.0 are slightly greater than for Version 3.0. The minimum requirements for the server are a 266-MHz Pentium with 64 MB of RAM and 2 GB of storage space. The client requirements are a 266-MHz Pentium with 64 MB of RAM and 50 MB of storage space. Both the server and the client must have the appropriate software to use Microsoft SQL. The exact hardware and software requirements can be obtained from Environmental Software.

Environmental Software is offering three training courses for SitePro Version 4.0. The first is a one-day introductory course that teaches clients to use the most important features of SitePro. A one-day advanced training course that builds upon the introductory training course is also offered. The advanced course is geared towards advanced

applications of the CAD, GIS, and database tools in SitePro. The student learns to make custom log and cross-section templates, create basemaps, perform sophisticated database queries, and generate query templates in the advanced class. The introductory and advanced classes are often offered on consecutive days. The cost for each course is $900 for up to six people plus an additional $65 per person for supplies and materials. The third training class is for EMIServer administration. The course focuses on network installation, database security, data quality control strategies, data access, and database maintenance for SitePro Version 4.0. Completion of the Microsoft SQL Server training course is strongly recommended as a prerequisite for the EMIServer administration course. This course is $575 per person and lasts one day. All prices quoted are for training at an Environmental Software location. Training at other sites can be made available for an additional fee.

Representative Applications

SitePro continues to expand its customer base. New clients include major corporations such as Westinghouse, Lockheed Martin, and Boeing. SitePro is also being used to manage and visualize environmental data sets by several environmental consulting firms. Examples include use of SitePro on the Beede Waste Oil Superfund site in New Hampshire and the Port of Los Angeles West Basin Intermodal Container Transfer Facility (ICTF) project. Detailed descriptions of these experiences can be found on the Environmental Software web page (www.envsoft.com).

Section 6 — References

Deutsch, C. V., and A. Journel. 1992. Geostatistical Software Library Version 2.0 and User's Guide for GSLIB 2.0. Oxford Press.

Englund, E. J., and A. R. Sparks. 1991. Geo-EAS (Geostatistical Environmental Assessment Software) and User's Guide, Version 1.1. EPA 600/4-88/033.

Golden Software. 1996. Surfer Version 6.04, June 24. Golden Software Inc., Golden Colorado.

Sullivan, T. M., and A. Q. Armstrong. 1998. "Decision Support Software Technology Demonstration Plan." Environmental & Waste Technology Center, Brookhaven National Laboratory, Upton, N.Y., September.

Sullivan, T. M., A. Q. Armstrong, and J. P. Osleeb. 1998. "Problem Descriptions for the Decision Support Software Demonstration." Environmental & Waste Technology Center, Brookhaven National Laboratory, Upton, N.Y., September.

van der Heijde, P. K. M., and D. A. Kanzer. 1997. *Ground-Water Model Testing: Systematic Evaluation and Testing of Code Functionality and Performance*. EPA/600/R-97/007. National Risk Management, Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH.

Appendix A — Summary of Test Problems

Site A: Sample Optimization Problem

Site A has been in operation since the late 1940s as an industrial machine plant that used solvents and degreasing agents. It overlies an important aquifer that supplies more than 2.7 million gal of water per day for industrial, commercial, and residential use. Site characterization and monitoring activities were initiated in the early 1980s, and it was determined that agricultural and industrial activities were sources of contamination. The industrial plant was shut down in 1985. The primary concern is volatile organic compounds (VOCs) in the aquifer and their potential migration to public water supplies. Source control is considered an important remediation objective to prevent further spreading of contamination.

The objective of this Site A problem was to challenge the software's capabilities as a sample optimization tool. The Site A test problem presents a three-dimensional (3-D) groundwater contamination scenario where two VOCs, dichloroethene (DCE) and trichloroethene (TCE), are present. The data that were supplied to the analysts included information on hydraulic head, subsurface geologic structure, and chemical concentrations from seven wells that covered an approximately 1000-ft square. Chemical analysis data were collected at 5-ft intervals from each well.

The design objective of this test problem was for the analyst to predict the optimum sample locations to define the depth and location of the plume at contamination levels exceeding the threshold concentration (either 10 or 100 µg/L). Because of the limited data set provided to the analysts and the variability found in natural systems, the analysts were asked to estimate the plume size and shape as well as the confidence in their prediction. A high level of confidence indicates that there is a high probability that the contaminant exceeds the threshold at that location. For example, at the 10-µg/L threshold, the 90% confidence level plume is defined as the region in which there is greater than a 90% chance that the contaminant concentration exceeds 10 µg/L. The analysts were asked to define the plume for three confidence levels—10% (maximum plume, low certainty, and larger region), 50% (nominal plume), and 90% (minimum plume, high certainty, and smaller region). The initial data set provided to the analyst was a subset of the available baseline data and intended to be insufficient for fully defining the extent of contamination in any dimension. The analyst used the initial data set to make a preliminary estimate of the dimensions of the plume and the level of confidence in the prediction. In order to improve the confidence and better define the plume boundaries, the analyst needed to determine where the next sample should be collected. The analyst conveyed this information to the demonstration technical team, which then provided the analyst with the contamination data from the specified location or locations. This iterative process continued until the analyst reached the test problem design objective.

Site A: Cost-Benefit Problem

The objectives of the Site A cost-benefit problem were (1) to determine the accuracy with which the software predicts plume boundaries to define the extent of a 3-D groundwater contamination problem on a large scale (the problem domain is approximately 1 square mile) and (2) to evaluate human health risk estimates resulting from exposure to contaminated groundwater. The VOC contaminants of concern for the cost-benefit problem were perchloroethene (PCE) and trichloroethane (TCA).

In this test problem analysts were to define the location and depth of the PCE plume at concentrations of 100 and 500 µg/L and TCA concentrations of 5 and 50 µg/L at confidence levels of 10 (maximum plume), 50 (nominal plume), and 90% (minimum plume). This information could be used in a cost-benefit analysis of remediation goals versus cost of remediation. The analysts were provided with geological information, borehole logs, hydraulic data, and an extensive chemical analysis data set consisting of more than 80 wells. Chemical analysis data were collected at 5-ft intervals from each well. Data from a few wells were withheld from the analysts to provide a reference to check interpolation routines. Once the analysts defined the PCE and TCA plumes, they were asked to calculate the human health risks associated with drinking 2 L/d of

contaminated groundwater at two defined exposure points over the next 5 years. One exposure point was in the central region of the plume and one was at the outer edge. This information could be used in a cost-benefit analysis of reduction of human health risk as a function of remediation.

Site B: Sample Optimization and Cost-Benefit Problem

Site B is located in a sparsely populated area of the southern United States on a 1350-acre site about 3 miles south of a large river. The site is typical of many metal fabrication or industrial facilities because it has numerous potential sources of contamination (e.g., material storage areas, process activity areas, service facilities, and waste management areas). As with many large manufacturing facilities, accidental releases from laboratory activities and cleaning operations introduced solvents and other organic chemicals into the environment, contaminating soil, groundwater, and surface waters.

The objective of the Site B test problem was to challenge the software's capabilities as a sample optimization and cost-benefit tool. The test problem presents a two-dimensional (2-D) groundwater contamination scenario with three contaminants—vinyl chloride (VC), TCE, and technetium-99 (Tc-99). Chemical analysis data were collected at a series of groundwater monitoring wells on quarterly basis for more than 10 years along the direction of flow near the centerline of the plume. The analysts were supplied with data from one sampling period.

There were two design objectives for this test problem. First, the analyst was to predict the optimum sample location to define the depth and location of the plume at specified contaminant threshold concentrations with confidence levels of 50, 75, and 90%. The initial data set provided to the analyst was a subset of the available baseline data and was intended to be insufficient for fully defining the extent of contamination in two dimensions. The analyst used the initial data set to make a preliminary estimate of the dimensions of the plume and the level of confidence in the prediction. In order to improve the confidence in defining the plume boundaries, the analyst needed to determine the location for collecting the next sample. The analyst conveyed this information to the demonstration technical team, who then provided the analyst with the contamination data from the specified location or locations. This iterative process continued until the analyst reached the design objective.

Once the location and depth of the plume was defined, the second design objective was addressed. The second design objective was to estimate the volume of contamination at the specified threshold concentrations at confidence levels of 50, 75, and 90%. This information could be used in a cost-benefit analysis of remediation goals versus cost of remediation. Also, if possible, the analyst was asked to calculate health risks associated with drinking 2 L/d of contaminated groundwater from two exposure points in the plume. One exposure point was near the centerline of the plume, while the other was on the edge of the plume. This information could be used in a cost-benefit analysis of reduction of human health risk as a function of remediation.

Site D: Sample Optimization and Cost-Benefit Problem

Site D is located in the western United States and consists of about 3000 acres of land bounded by municipal areas on the west and southwest and unincorporated areas on northwest and east. The site has been an active industrial facility since it began operation in 1936. Operations have included maintenance and repair of aircraft and, recently, the maintenance and repair of communications equipment and electronics. The aquifer beneath the site is several hundred feet thick and consists of three or four different layers of sand or silty sand. The primary concern is VOC contamination of soil and groundwater as well as contamination of soil with metals.

The objective of the Site D problem was to test the software's capability as a tool for sample optimization and cost-benefit problems. This test problem was a 3-D groundwater sample optimization problem for four VOC contaminants—PCE, DCE, TCE, and trichloroethane (TCA). The test problem required the developer to predict the optimum sample locations to define the region of the contamination that exceeded threshold concentrations for each contaminant. Contaminant data were supplied for a series of wells screened at different depths for four quarters in a 1-year time frame. This initial data set was insufficient to fully define the extent of contamination. The analyst used the initial data set to make a preliminary estimate of the

dimensions of the plume and the level of confidence in the prediction. In order to improve the confidence in the prediction of the plume boundaries, the analyst needed to determine the location for collecting the next sample. The analyst conveyed this information to the demonstration technical team, who then provided the analyst with the contamination data from the specified location or locations. This iterative process was continued until the analyst determined that the data could support definition of the location and depth of the plume exceeding the threshold concentrations with confidence levels of 10, 50, and 90% for each contaminant.

After the analyst was satisfied that the sample optimization problem was complete and the plume was defined, he or she was given the option to continue and perform a cost-benefit analysis. At Site D, the cost-benefit problem required estimation of the volume of contamination at specified threshold concentrations with confidence levels of 10, 50, and 90%. This information could then be used in a cost-benefit analysis of remediation goals versus cost of remediation.

Site N: Sample Optimization Problem

Site N is located in a sparsely populated area of the southern United States and is typical of many metal fabrication or industrial facilities in that it has numerous potential sources of contamination (e.g., material storage areas, process activity areas, service facilities, and waste management areas). Industrial operations include feed and withdrawal of material from the primary process; recovery of heavy metals from various waste materials and treatment of industrial wastes. The primary concern is contamination of the surface soils by heavy metals.

The objective of the Site N sample optimization problem was to challenge the software's capability as a sample optimization tool to define the areal extent of contamination. The Site N data set contains the most extensive and reliable data for evaluating the accuracy of the analysis for a soil contamination problem. To focus only on the accuracy of the soil sample optimization analysis, the problem was simplified by removing information regarding groundwater contamination at this site, and it was limited to three contaminants. The Site N test problem involves surface soil contamination (a 2-D problem) for three contaminants—arsenic (As), cadmium (Cd), and chromium (Cr). Initial sampling indicated a small contaminated region on the site; however, the initial sampling was limited to only a small area (less than 5% of the site area).

The design objective of this test problem was for the analyst to develop a sampling plan that defines the extent of contamination on the 150-acre site based on exceedence of the specified threshold concentrations with confidence levels of 10, 50% and 90%. Budgetary constraints limited the total expenditure for sampling to $96,000. Sample costs were $1200 per sample, which included collecting and analyzing the surface soil sample for all three contaminants. Therefore, the number of additional samples had to be less than 80. The analyst used the initial data to define the areas of contamination and predict the location of additional samples. The analyst was then provided with additional data at these locations and could perform the sample optimization process again until the areal extent of contamination was defined or the maximum number of samples (80) was attained. If the analyst determined that 80 samples was insufficient to adequately characterize the entire 150-acre site, the analyst was asked to use the software to select the regions with the highest probability of containing contaminated soil.

Site N: Cost-Benefit Problem

The objective of the Site N cost-benefit problem was to challenge the software's ability to perform cost-benefit analysis as defined in terms of area of contaminated soil above threshold concentrations and/or estimates of human health risk from exposure to contaminated soil. This test problem considers surface soil contamination (2-D) for three contaminants—As, Cd, and Cr. The analysts were given an extensive data set for a small region of the site and asked to conduct a cost-benefit analysis to evaluate the cost for remediation to achieve specified threshold concentrations. If possible, an estimate of the confidence in the projected remediation areas was provided at the 50 and 90% confidence limits. For human health risk analysis, two scenarios were considered. The first was the case of an on-site worker who was assumed to have consumed 500 mg/d of soil for one year during excavation activities. The worker would have worked in all areas of the site during the excavation process. The second scenario considered a resident who was assumed to live on a

200- by 100-ft area at a specified location on the site and to have consumed 100 mg/d of soil for 30 years. This information could be used in a cost-benefit (i.e., reduction of human health risk) analysis as a function of remediation.

Site S: Sample Optimization Problem

Site S has been in operation since 1966. It was an industrial fertilizer plant producing pesticides and fertilizer and used industrial solvents such as carbon tetrachloride (CTC) to clean equipment. Recently, it was determined that routine process operations were causing a release of CTC onto the ground; the CTC was then leaching into the subsurface. Measurements of the CTC concentration in groundwater have been as high as 80 ppm a few hundred feet down-gradient from the source area. The site boundary is approximately 5000 ft from the facility where the release occurred. Sentinel wells at the boundary are not contaminated with CTC.

The objective of the Site S sample optimization problem was to challenge the software's capability as a sample optimization tool. The test problem involved a 3-D groundwater contamination scenario for a single contaminant, CTC. To focus only on the accuracy of the analysis, the problem was simplified. Information regarding surface structures (e.g., buildings and roads) was not supplied to the analysts. In addition, the data set was modified such that the contaminant concentrations were known exactly at each point (i.e., release and transport parameters were specified, and concentrations could be determined from an analytical solution). This analytical solution permitted a reliable benchmark for evaluating the accuracy of the software's predictions.

The design objective of this test problem was for the analyst to define the location and depth of the plume at CTC concentrations exceeding 5 and 500 µg/L with confidence levels of 10, 50, and 90%. The initial data set provided to the analysts was insufficient to define the plume accurately. The analyst used the initial data to make a preliminary estimate of the dimensions of the plume and the level of confidence in the prediction. In order to improve the confidence in the predicted plume boundaries, the analyst needed to determine where the next sample should be collected. The analyst conveyed this information to the demonstration technical team, who then provided the analyst with the contamination data from the specified location or locations. This iterative process continued until the analyst reached the design objective.

Site S: Cost-Benefit Problem

The objective of the Site S cost-benefit problem was to challenge the software's capability as a cost-benefit tool. The test problem involved a 3-D groundwater cost-benefit problem for a single contaminant, chlordane. Analysts were given an extensive data set consisting of data from 34 wells over an area that was 2000 ft long and 1000 ft wide. Vertical chlordane contamination concentrations were provided at 5-ft intervals from the water table to beneath the deepest observed contamination.

This test problem had three design objectives. The first was to define the region, mass, and volume of the plume at chlordane concentrations of 5 and 500 µg/L. The second objective was to extend the analysis to define the plume volumes as a function of three confidence levels—10, 50, and 90%. This information could be used in a cost-benefit analysis of remediation goals versus cost of remediation. The third objective was to evaluate the human health risk at three drinking-water wells near the site, assuming that a resident drinks 2 L/d of water from a well screened over a 10-ft interval across the maximum chlordane concentration in the plume. The analysts were asked to estimate the health risks at two locations at times of 1, 5 and 10 years in the future. For the health risk analysis, the analysts were told to assume source control preventing further release of chlordane to the aquifer. This information could be used in a cost-benefit analysis of reduction of human health risk as a function of remediation.

Site T: Sample Optimization Problem

Site T was developed in the 1950s as an area to store agricultural equipment as well as fertilizers, pesticides, herbicides, and insecticides. The site consists of 18 acres in an undeveloped area of the western United States, with the nearest residence being approximately 0.5 miles north of the site. Mixing operations (fertilizers and

pesticides or herbicides and insecticides) were discontinued or replaced in the 1980s when concentrations of pesticides and herbicides in soil and wastewater were determined to be of concern.

The objective of the Site T sample optimization problem was to challenge the software's capability as a sample optimization tool. The test problem presents a surface and subsurface soil contamination scenario for four VOCs: ethylene dibromide (EDB), dichloropropane (DCP), dibromochloropropane (DBCP), and CTC. This sample optimization problem had two stages. In the first stage, the analysts were asked to prepare a sampling strategy to define the areal extent of surface soil contamination that exceeded the threshold concentrations listed in Table A-1 with confidence levels of 10, 50 and 90% on a 50- by 50-ft grid. This was done in an iterative fashion in which the analysts would request data at additional locations and repeat the analysis until they could determine, with the aid of their software, that the plume was adequately defined.

The stage two design objective addressed subsurface contamination. After defining the region of surface contamination, the analysts were asked to define subsurface contamination in the regions found to have surface contamination above the 90% confidence limit. In stage two, the analysts were asked to suggest subsurface sampling locations on a 10-ft vertical scale to fully characterize the soil contamination at depths from 0 to 30 ft below ground surface (the approximate location of the aquifer).

Table A-1. Site T soil contamination threshold concentrations

Contaminant	Threshold concentration (μg/kg)
Ethylene dibromide (EDB)	21
Dichloropropane (DCP)	500
Dibromochloropropane (DBCP)	50
Carbon tetrachloride (CTC)	5

Site T: Cost-Benefit Problem

The objective of the Site T cost-benefit problem was to challenge the software's capability as a cost-benefit tool. The test problem involved a 3-D groundwater contamination scenario with four VOCs (EDB, DCB, DBCP, and CTC). The analysts were given an extensive data set and asked to estimate the volume, mass, and location of the plumes at specified threshold concentrations for each VOC. If possible, the analysts were asked to estimate the 50 and 90% confidence plumes at the specified concentrations. This information could be used in a cost-benefit analysis of various remediation goals versus the cost of remediation. For health risk cost-benefit analysis, the analysts were asked to evaluate the risks to a residential receptor (with location and well screen depth specified) and an on-site receptor over the next 10 years. For the residential receptor, consumption of 2 L/d of groundwater was the exposure pathway. For the on-site receptor, groundwater consumption of 1 L/d was the exposure pathway. For both human health risk estimates, the analysts were told to assume removal of any and all future sources that may impact the groundwater. This information could be used in a cost-benefit analysis of various remediation goals versus the cost of remediation.

Appendix B — Description of Interpolation Methods

A major component of the analysis of environmental data sets involves predicting physical or chemical properties (contaminant concentrations, hydraulic head, thickness of a geologic layer, etc.) at locations between measured data. This process, called interpolation, is often critical in developing an understanding of the nature and extent of the environmental problem. The premise of interpolation is that the estimated value of a parameter is a weighted average of measured values around it. Different interpolation routines use different criteria to select the weights. Because of the importance of obtaining estimates of parameters between measured data points in many fields of science, a wide number of interpolation routines exist.

Three classes of interpolation routines commonly used in environmental analysis are nearest neighbor, inverse distance, and kriging. These three classes cover the range found in the software used in the demonstration and use increasingly complex models to select their weighting functions.

Nearest neighbor is the simplest interpolation routine. In this approach, the estimated value of a parameter is set to the value of the spatially nearest neighbor. This routine is most useful when the analyst has a lot of data and is estimating parameters at only a few locations. Another simple interpolation scheme is averaging of nearby data points. This scheme is an extension of the nearest neighbor approach and interpolates parameter values as an average of the measured values within the neighborhood (specified distance). The weights for averaging interpolation are all equal to $1/n$, where n is the number of data points used in the average. The nearest neighbor and averaging interpolation routines do not use any information about the location of the data values.

Inverse distance weighting (IDW) interpolation is another simple interpolation routine that is widely used. It does account for the spatial distance between data values and the interpolation location. Estimates of the parameter are obtained from a weighted average of neighboring measured values. The weights of IDW interpolation are proportional to the inverse of these distances raised to a power. The assigned weights are fractions that are normalized such that the sum of all the weights is equal to 1.0. In environmental problems, contaminant concentrations typically vary by several orders of magnitude. For example, the concentration may be a few thousand micrograms per liter near the source and tens of micrograms per liter away from the source. With IDW, the extremely high concentrations tend to have influence over large distances, causing smearing of the estimated area of contamination. For example, for a location that is 100 m from a measured value of 5 μg/L and 1000 m from a measured value of 5000 μg/L, using a distance weighting factor of 1 in IDW yields a weight of 5000/1000 for the high-concentration data point and 5/100 for the low-concentration data point. Thus, the predicted value is much more heavily influenced by the large measured value that is physically farther from the location at which an estimate is desired. To minimize this problem, the inverted distance weight can be increased to further reduce the effect of data points located farther away. IDW does not directly account for spatial correlation that often exists in the data. The choice of the power used to obtain the interpolation weights is dependent on the skills of the analyst and is often obtained through trial and error.

The third class of interpolation schemes is kriging. Kriging attempts to develop an estimate of the spatial correlation in the data to assist in interpolation. Spatial correlation represents the correlation between two measurements as a function of the distance and direction between their locations. Ordinary kriging interpolation methods assume that the spatial correlation function is based on the assumption that the measured data points are normally distributed. This kriging method is often used in environmental contamination problems and was used by some DSS products in the demonstration and in the baseline analysis. If the data are neither lognormal nor normally distributed, interpolations can be handled with indicator kriging. Some of the DSS products in this demonstration used this approach. Indicator kriging differs from ordinary kriging in that it makes no assumption on the distribution of data and is essentially a nonparametric counterpart to ordinary kriging.

Both kriging approaches involve two steps. In the first step, the measured data are examined to determine the spatial correlation structure that exists in the data. The parameters that describe the correlation structure are calculated as a variogram. The variogram merely describes the spatial relationship between data points. Fitting a model to the variogram is the most important and technically challenging step. In the second step, the kriging process interpolates data values at unsampled locations by a moving-average technique that uses the results from the variogram to calculate the weighting factors. In kriging, the spatial correlation structure is quantitatively evaluated and used to calculate the interpolation weights.

Although geostatistical-based interpolation approaches are more mathematically rigorous than the simple interpolation approaches using nearest neighbor or IDW, they are not necessarily better representations of the data. Statistical and geostatistical approaches attempt to minimize a mathematical constraint, similar to a least squares minimization used in curve-fitting of data. While the solution provided is the "best" answer within the mathematical constraints applied to the problem, it is not necessarily the best fit of the data. There are two reasons for this.

First, in most environmental problems, the data are insufficient to determine the optimum model to use to assess the data. Typically, there are several different models that can provide a defensible assessment of the spatial correlation in the data. Each of these models has its own strengths and limitations, and the model choice is subjective. In principle, selection of a geostatistical model is equivalent to picking the functional form of the equation when curve-fitting. For example, given three pairs of data points, (1,1), (2,4) and (3,9), the analyst may choose to determine the best-fit line. Doing so gives the expression $y = 4x - 3.33$, where y is the dependent variable and x is the independent variable. This has a goodness of fit correlation of 0.97, which most would consider to be a good fit of the data. This equation is the "best" linear fit of the data constrained to minimization of the sum of the squares of the residuals (difference between measured value and predicted value at the locations of measured values). Other functional forms (e.g., exponential, trigonometric, and polynomial) could be used to assess the data. Each of these would give a different "best" estimate for interpolation of the data. In this example, the data match exactly with $y = x^2$, and this is the best match of this data. However, that this is the best match cannot be known with any high degree of confidence.

This conundrum leads to the second reason for the difficulty, if not impossibility, of finding the most appropriate model to use for interpolation—which is that unless the analyst is extremely fortunate, the measured data will not conform to the mathematical model used to represent the data. This difficulty is often attributed to the variability found in natural systems, but is in fact a measure of the difference between the model and the real-world data. To continue with the previous example, assume that another data point is collected at $x = 2.5$ and the value is $y = 6.67$. This latest value falls on the previous linear best-fit line, and the correlation coefficient increases to 0.98. Further, it does not fall on the curve $y = x^2$. The best-fit 2nd-order polynomial now changes from $y = x^2$ to become $y = 0.85x^2 + 0.67x - 0.55$. The one data point dramatically changed the "best"-fit parameters for the polynomial and therefore the estimated value at locations that do not have measured values.

Lack of any clear basis for choosing one mathematical model over another and the fact that the data are not distributed in a manner consistent with the simple mathematical functions in the model also apply to the statistical and geostatistical approaches, albeit in a more complicated manner. In natural systems, the complexity increases over the above example because of the multidimensional spatial characteristics of environmental problems. This example highlighted the difficulty in concluding that one data representation is better than another. At best, the interpolation can be reviewed to determine if it is consistent with the data. The example also highlights the need for multiple lines of reasoning when assessing environmental data sets. Examining the data through use of different contouring algorithms and model parameters often helps lead to a more consistent understanding of the data and helps eliminate poor choices for interpolation parameters.