United States
Environmental Protection
Agency

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

EPA/600/R-00/037
February 2000

Environmental Technology Verification Report

Environmental Decision Support Software

Decision*FX*, Inc.

Groundwater*FX*

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, VISUALIZATION, SAMPLE OPTIMIZATION, AND COST-BENEFIT ANALYSIS OF ENVIRONMENTAL DATA SETS**
TECHNOLOGY NAME:	**Groundwater*FX***
COMPANY:	**Decision*FX*, Inc.** **310 Country Lane** **Bosque Farms, NM 87068**
PHONE:	**(505) 869-0057**
WEBSITE:	**www.decisionFX.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 (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. This verification statement provides a summary of the test results of a demonstration of Decision*FX*'s Groundwater*FX* 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. The technical team performed a baseline analysis for each problem to be used as a basis of comparison.

Decision*FX* staff chose to use Groundwater*FX* to perform all three endpoints using data from the Site B and Site S sample optimization and cost-benefit problems. For both problems, Groundwater*FX* was used to define sample locations to characterize the extent of groundwater contamination above specified contaminant threshold concentrations. The software generated two-dimensional (2-D) base maps containing site features that were overlain with maps of concentrations or of probability of exceeding contamination threshold levels. Groundwater*FX* was also used to estimate the volume of water contaminated above the specified threshold concentrations and to provide exposure concentrations at specified locations for use in human health risk calculations. The estimates for volume and concentrations were done using probabilistic simulation. This permitted the analyst to provide statistical estimates of the confidence in the software's volume and concentration estimates. Details of the demonstration, including an evaluation of the software's performance, may be found in the report entitled *Environmental Technology Verification Report: Environmental Decision Support Software—DecisionFX, Inc., GroundwaterFX,* EPA/600/R-00/037.

TECHNOLOGY DESCRIPTION

Groundwater*FX* is a decision support system intended to provide decision makers and analysts a means of evaluating environmental information related to the nature and extent of contamination in groundwater. Key attributes of the product include the ability to delineate, provide visual feedback, and quantify uncertainties in the nature and extent of groundwater contamination (e.g., concentration distribution, probability distribution of exceeding a groundwater cleanup guideline); to provide objective recommendations on the number and location of sampling points; and to provide statistical information about the contamination (e.g., average volume of contamination, standard deviation, etc.). Groundwater*FX* runs on Windows 95 and 98 or NT platforms and on the Power Macintosh operating system.

VERIFICATION OF PERFORMANCE

The following performance characteristics of Groundwater*FX* were observed:

Decision Support: Groundwater*FX* is a probabilistic-based software designed to address 2-D and three-dimensional (3-D) groundwater contamination problems, including optimization of new sample locations and generation of cost-benefit information (e.g., evaluation of the probability of exceeding threshold concentrations). The software generated 2-D maps of the contamination and of the probability of exceeding a specified threshold concentration. Cost-benefit curves of the cost (volume) of remediation vs. the probability of exceeding a threshold concentration were generated in Excel using Groundwater*FX* output files. The software provided estimates of current and future exposure concentrations for use in human health risk calculations. The interpretations of statistical data permit the decision maker to evaluate future actions, such

as determining sampling locations or developing cleanup guidance, on the basis of the level of confidence placed in the analysis.

Documentation of the GroundwaterFX Analysis: Decision*FX* staff generated a report that provided an adequate explanation of the process and parameters used to analyze each problem. Documentation of data transfer, manipulations of the data, and analyses were included. The criteria used to select models for the simulation and the parameters for conducting the probabilistic assessment were provided in standard ASCII text files that are exportable to a number of software programs. Output files from the simulations were also provided for review.

Comparison with Baseline Analysis and Data: Decision*FX* used Groundwater*FX* to perform the visualization, sample optimization, and cost-benefit aspects of problems from Sites B and S. The analysis performed by Groundwater*FX* did not provide an adequate match to the data on either test problem. For Site B, the locations of wells in some simulations were incorrectly plotted on the site map. The maps of contaminant concentrations were generally consistent with the data near the source of contamination. However, the software did not represent the leading edge of the plume accurately. The maps showing the probability of exceeding a contaminant threshold were inconsistent with the baseline data, and the estimate of the volume of the plume was three to five times smaller than that obtained in the baseline analyses. The estimates of exposure concentrations for risk calculations were too low by a factor of 2 to 3 as compared to the baseline analyses. For Site S, Groundwater*FX*'s estimates of contaminant concentrations were an extremely poor match to the data and baseline analysis. As a result, estimates of the volume of contaminated groundwater and of exposure concentrations for risk calculations were substantially different from those suggested by the data and baseline analysis. In addition, the Groundwater*FX* estimates of exposure concentrations supplied for risk calculations were inconsistent with the contaminant concentration maps generated by the software.

Multiple Lines of Reasoning: The foundation of the Groundwater*FX* approach is a Monte Carlo simulator that produces multiple simulations of the distribution of contamination that are consistent with the known data. From these simulations, concentration and probability maps were produced to assist in evaluating the extent of contamination. This permits the decision-maker to evaluate future actions, such as determining sampling locations or developing cleanup guidance, on the basis of the level of confidence placed in the analysis.

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

Ease of Use: Groundwater*FX* is a sophisticated flow and transport code that incorporates Monte Carlo simulation in a 3-D framework. A high level of skill and experience is required to use it effectively.

Several features of Groundwater*FX* make the software package cumbersome to use. These include the need for a formatted data file for importing location and concentration data, the need to have all units of measurement in meters (USGS and state plane coordinates systems are typically measured in feet), the need to have all graphic files imported as a single bitmap (which prohibits the use of multiple layers in visualizations and requires coordinates of the bitmap to be provided when the bitmap is used as a base map for visualization), the inability to edit graphic bitmap files, and the absence of on-line help. Visualization output is limited to bitmaps of screen captures that can be imported into other software for processing. Overcoming these limitations to perform an analysis requires more work on the part of the software operator.

Groundwater*FX* exports text and graphics to standard word processing software directly. Graphic outputs are generated as bitmaps, which can be imported into other software to generate .jpg, and .cdr graphic files. Groundwater*FX* generates data files from statistical analysis and concentration estimates in ASCII format, which can be read by most software.

Efficiency and Range of Applicability: Two problems were completed and documented with 12 person-days of effort. However, the technical team concluded that the analyses were, at best, a first pass through the

problem; the procedure would need to be repeated several times to improve the accuracy of the analysis. The incomplete analysis was due primarily to the combination of the sophisticated approach of the software—e.g., Monte Carlo simulation of 3-D flow and transport—and the time constraints of the demonstration. Substantially more time would be required to properly analyze the problem. Groundwater*FX* provides the flexibility to address problems tailored to site-specific conditions.

Operator Skill Base: To use Groundwater*FX* efficiently, the operator should be knowledgeable in probabilistic modeling of groundwater flow and contaminant transport. Knowledge pertaining to conducting sample optimization analysis and performing cost-benefit problems would be beneficial.

Training and Technical Support: An analyst with the prerequisite skill base can be using Groundwater*FX* after three days of training. A users' manual is available to assist in operation of the software. Technical support is available through e-mail and over the phone.

Cost: Decision*FX* plans to sell Groundwater*FX* for $1000 for a single license. It will be supplied at no cost to state and federal regulators.

Overall Evaluation: The main strength of Groundwater*FX* is its technical approach using Monte Carlo simulation of flow and transport processes to address variability and uncertainty in groundwater contamination problems. The use of groundwater simulation models should be a better approach to sample optimization designs than the use of purely statistical or geostatistical simulation models. However, the analysis performed by Groundwater*FX* did not provide an adequate match to the data on either test problem. Thus, it was not possible to determine whether Groundwater*FX* can accurately estimate the extent of groundwater contamination. The technical team also concluded that the many ease-of-use issues identified above make the software cumbersome to use. In particular, visualization capabilities are limited, and the ability to import graphic files only in bitmap format can lead to problems in the analysis.

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 software can address these components of a credible analysis. However, other components, such as proper conceptualization and use of code, depend on the analyst's skills. 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 an evaluation of technology performance under specific, predetermined criteria and the appropriate quality assurance procedures. EPA makes no expressed or implied warranties as to the performance of the technology and does not certify that a technology will always, under circumstances other than those tested, operate at the levels verified. The end user is solely responsible for complying with any and all applicable federal, state, and local requirements.

EPA/600/R-00/037
February 2000

Environmental Technology Verification Report

Environmental Decision Support Software

Decision*FX*, Inc.

Groundwater*FX*

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 (DOE's) Environmental Management Program through the National Analytical Management Program (NAMP), funded and managed, through Interagency Agreement No. DW89937854 with Oak Ridge National Laboratory (ORNL), the verification effort described herein. This report has been peer-reviewed 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 Groundwater*FX* 2

2 GROUNDWATER*FX* CAPABILITIES 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 10
Ease of Use 11
Efficiency and Range of Applicability 11

4 GROUNDWATER*FX* EVALUATION 12
Groundwater*FX* Technical Approach 12
Description of Test Problems 12
Site B Sample Optimization and Cost-Benefit Problem 12
Site S Sample Optimization and Cost-Benefit Problem 14
Evaluation of Groundwater*FX* 14
Decision Support 14
Documentation of the Groundwater*FX* Analysis and Evaluation of the Technical Approach 15
Comparison of Groundwater*FX* Results with the Baseline Analysis and Data 16
Site B Sample Optimization and Cost-Benefit Problem 16
Site S Sample Optimization and Cost-Benefit Problem 23
Comment on Groundwater*FX* Site B and S Analyses 32
Multiple Lines of Reasoning 32
Secondary Evaluation Criteria 33
Ease of Use 33

Efficiency and Range of Applicability 33
Training and Technical Support..... 33
Additional Information about the Groundwater*FX* Software 33
Summary of Performance..... 34

5 GROUNDWATER*FX* UPDATE AND REPRESENTATIVE APPLICATIONS..... 36
Objective 36
Groundwater*FX* Update 36
Representative Applications 36

6 REFERENCES..... 40

Appendix A—Summary of Test Problems 41
Appendix B—Description of Interpolation Methods 47

List of Figures

1 Groundwater*FX*-generated map for Site B with sample locations color-coded to match TCE concentration 17
2 Groundwater*FX*-generated map of average TCE concentration at Site B at the time of the data collection 18
3 Baseline analysis of TCE concentration contours at 50 µg/L (green) and 500 µg/L (red) based on kriging interpolation with Surfer 19
4 Baseline analysis of TCE concentration (µg/L) contours based on kriging using GSLIB 20
5 Groundwater*FX*-generated map of the probability of the TCE concentration exceeding 50 µg/L 20
6 Baseline map of the probability of the TCE concentration exceeding 50 µg/L generated with GSLIB 21
7 Groundwater*FX*-simulated average CTC concentrations in the four layers based on original data plus three additional samples 25
8 Baseline analysis of CTC concentrations at 5-µg/L (blue) and 500-µg/L (red) contours based on Decision*FX* data set 27
9 Baseline analysis using the analytical solution to provide data points to generate contours at 5- and 500-µg/L CTC thresholds 28
10 Groundwater*FX* map of probability of exceeding 5 µg/L in layer 1 based on initial data 29
11 Average uranium concentrations in 2027 37
12 Average uranium concentrations in 2027 37
13 Probability map that uranium exceeds MCLs in 2027 38
14 Predicted uranium concentrations over time at well 413 with uncertainty error bars 39

List of Tables

1 Summary of test problems 6
2 Data supplied for the test problems 6
3 Site B groundwater contamination problem threshold levels 13
4 Groundwater*FX* and baseline analysis volume estimates at the 50% probability level for the Site B TCE contamination problem 22
5 Groundwater*FX* and GSLIB volume estimates at the 10% and 90% probability levels for the Site B TCE contamination problem 22
6 Groundwater*FX* volume estimates of CTC-contaminated groundwater for the Site S sample optimization problem. 30
7 Baseline volume estimates of CTC-contaminated groundwater for the Site S sample optimization problem. 30
8 Groundwater*FX* and baseline estimates for current CTC exposure concentrations (µg/L) for the Site S residential risk evaluation. 31
9 Groundwater*FX* and analytical estimates over time for CTC exposure concentrations (µg/L) for the Site S residential risk evaluation. 31
10 Groundwater*FX* performance summary 35

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 Steve Gardner (EPA NERL) and Budhendra Bhaduri (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 (ORO)]; for computer and network support, Leslie Bloom (ORNL); and for technical guidance and project management of the demonstration, David Carden and Regina Chung (ORO), 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 Bob Knowlton of Decision*FX*, Inc., 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
Characterization and Research Division
National Exposure Research Laboratory
P.O. Box 93478
Las Vegas, Nevada 89193-3478
(702) 798-2432

For more information on the Decsion*FX*, Inc., Groundwater*FX* product, contact

Bob Knowlton
Decision*FX*, Inc.
310 Country Lane
Bosque Farms, New Mexico 87068
(505) 869-0057

Abbreviations and Acronyms

ACL	alternate concentration limit
As	arsenic
ASCII	American Standard Code for Information Interchange (file format)
.bmp	bitmap file
BNL	Brookhaven National Laboratory
C_{95}	95th percentile concentration
Cd	cadmium
CD-ROM	compact disk—read only memory
Cr	chromium
CTC	carbon tetrachloride
DBCP	dibromochloroproprane
.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 Program (DOE)
EPA	U.S. Environmental Protection Agency
ESRI	Environmental Systems Research Institute
ETV	Environmental Technology Verification Program
FTP	file transfer protocol
Geo-EAS	Geostatistical Environmental Assessment Software
GSLIB	Geostatistical Software Library (software)
GUI	graphical user interface
IDW	inverse distance weighting
LHS	Latin hypercube sampling
MB	megabyte
MCL	maximum contaminant level
MHz	megahertz
MSL	mean sea level
NAMP	National Analytical Management Program (DOE)
NERL	National Exposure Research Laboratory (EPA)
NMERI	New Mexico Engineering Research Institute
NRC	Nuclear Regulatory Commission
ORD	Office of Research and Development (EPA)
ORNL	Oak Ridge National Laboratory
ORO	Oak Ridge Operations Office (DOE)
PCE	perchloroethene or tetrachloroethene
pdf	probability density function
ppm	parts per million
QA	quality assurance
QC	quality control
RAM	random access memory
RMSE	root mean square error
SADA	Spatial Analysis and Decision Assistance (software)
SCMT	Site Characterization and Monitoring Technology

TCA	trichloroethane
TCE	trichloroethene
Tc-99	technetium-99
UMTRA	Uranium Mill Tailings Remedial Action program (DOE)
UTRC	University of Tennessee Research Corporation
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 on the basis of testing and QA protocols developed with input from all major stakeholder and 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 levels) 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 have 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 these reports 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 Groundwater*FX*.

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 with 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 Groundwater*FX*

Groundwater*FX* is a decision support system intended to provide decision makers and analysts a means of evaluating environmental information relating to the nature and extent of contamination in groundwater contamination problems. Key attributes

of the tool include the ability to quantify uncertainties in the nature and extent of groundwater contamination; provide objective recommendations on the number and location of sampling points to delineate the contamination; provide visual feedback to a user on the nature and extent of the contamination (e.g., concentration distribution, probability distribution of exceeding a concentration threshold); and provide statistical information about the plume (e.g., average volume of contamination, standard deviation).

Decision*FX* staff chose to use Groundwater*FX* to perform all three endpoints using data from the Site B and Site S sample optimization and cost-benefit problems. For both problems, Groundwater*FX* was used to define sample locations to characterize the extent of groundwater contamination above specified contaminant threshold concentrations. The software generated two-dimensional (2-D) base maps containing site features that were overlain with maps of concentrations or of probability of exceeding contamination threshold levels. Groundwater*FX* was also used to estimate the volume of water contaminated above the specified threshold concentrations and to provide exposure concentrations at specified locations for use in human health risk calculations. The estimates for volume and exposure concentrations were done using probabilistic simulation. This approach permitted the analyst to provide statistical estimates of the confidence in the software's volume and concentration estimates.

The Site B problem was a 2-D groundwater contamination problem. Decision*FX* used Groundwater*FX* to perform probabilistic simulations of groundwater flow and transport. This analysis was used to identify and request four additional sample locations to further define the extent of the plume. On the basis of the final data set, the analyst used Groundwater*FX* to generate maps of the concentration distribution and probability distribution of exceeding the two threshold concentrations for trichloroethene (TCE), vinyl chloride (VC), and technetium-99 (Tc-99). The data were also used to generate a cost-benefit analysis of the volume contaminated vs. the cleanup threshold. Finally, Groundwater*FX* was used to estimate the exposure concentrations at two well locations 1 year and 5 years in the future as a basis for human health risk calculations.

The Site S sample optimization problem is a three-dimensional (3-D) groundwater contamination problem for a single contaminant, carbon tetrachloride (CTC). To address the 3-D nature of the problem, the Decision*FX* analyst divided the subsurface into four layers. The hydraulic parameters and data were used to perform probabilistic simulations of groundwater flow and transport. Groundwater*FX* was used to identify and request three additional sample locations to further define the plume. On the basis of the final data set, Groundwater*FX* was used to generate 2-D maps of the concentration distribution and probability distribution of exceeding the two threshold concentrations for CTC in the four layers. The data were also used to generate a cost-benefit analysis of the contaminated volume of groundwater which exceeded threshold concentrations. Finally, Groundwater*FX* was used to estimate exposure concentrations at two well locations under current conditions and at 1, 5, and 10 years in the future as a basis for human health risk calculations.

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

Section 2—Groundwater*FX* Capabilities

This section provides a general overview of the capabilities of Groundwater*FX*, a Decision*FX*, Inc., software product. Decision*FX*, Inc., supplied this information.

Groundwater*FX* is a decision support system intended to provide decision makers and analysts a means of evaluating environmental information relating to the nature and extent of contamination in groundwater contamination problems. Key attributes of the tool include its ability to

- quantify uncertainties in the nature and extent of soil contamination;
- provide objective recommendations on the number and location of sampling points to delineate the contamination;
- provide visual feedback to a user on the nature and extent of the contamination (e.g., concentration distribution, probability distribution of exceeding a soil guideline); and
- provide statistical information about the plume (e.g., average volume of contamination, standard deviation).

Groundwater*FX* relies mainly on flow and transport process model algorithms to assess the potential for contaminant migration and on operations research methods to provide guidance on key decision analysis needs (e.g., recommended location of monitor wells). The Groundwater*FX* methodology is an improvement over conventional groundwater modeling analysis approaches because it integrates the following features into a single software product:

1. it allows the user to simulate fate and transport for the source term, the vadose zone, and the saturated zone (a 3-D finite-difference model for flow and advective-dispersive solute transport);
2. it quantifies uncertainties through the use of Latin hypercube sampling (LHS) and Monte Carlo stochastic simulation techniques;
3. it honors hydraulic conductivity information and explicitly accounts for spatial variability through the use of geostatistical routines;
4. it honors observed water quality data, thereby providing a type of built-in calibration method;
5. it provides objective guidance on the placement of monitor wells based on an operations research algorithm (rather than by using expert judgment); and
6. it has visual display capabilities that allow a user to assess the uncertainties.

The Groundwater*FX* code is designed to provide decision analysis information on single analytes associated with contamination in groundwater. For multiple analytes of concern, multiple model runs must be performed. Though some investigators have used geostatistical approaches to analyze groundwater plume data, Decision*FX* recommends the use of mass-conservative process modeling methods to address these issues. Thus, Groundwater*FX* simulates the physics of flow and transport processes, providing a better understanding of the nature and extent of contamination, and quite often with fewer data points than a statistical or geostatistical approach would require.

Currently, Groundwater*FX* has versions that run on Windows 95, Windows NT, and Macintosh platforms. The software is written mainly in two languages: Fortran for the mathematical operations and C++ for the graphical user interface (GUI) functions. Development software was chosen for ease of use in porting to different platforms. The recommended computer configuration for running the Groundwater*FX* software on PC platforms is approximately 50 MB of hard-disk space for the program, about 100 MB of storage space for model runs, about 64 MB of RAM, and a reasonably fast Pentium processor (>100 MHz).

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 United States that 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 Decision*FX* are discussed in Section 4 as part of the evaluation of Groundwater*FX*'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 2-D and 3-D 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.

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

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 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 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 Groundwater*FX* 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 4000 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 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. On the basis of 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—Groundwater*FX* Evaluation

Groundwater*FX* Technical Approach

Groundwater*FX* is a probabilistic flow and transport model used to address groundwater contamination problems. The analyst takes the information provided from site characterization and develops a conceptual model of the source term, vadose zone flow, saturated zone flow, and contaminant transport in three dimensions. From the conceptual model and the site characterization data, the analyst chooses the model parameters necessary for Groundwater*FX* to perform the flow and transport simulation. Many parameters are assigned as a distribution of potential values. Groundwater*FX* randomly selects the model parameters from the distribution of potential values supplied by the software user and then performs a simulation of the problem. The process is repeated several times to obtain a distribution of potential outcomes.

In the initial stages of the analysis, there is often a wide spread in the distribution parameters. Therefore, 10 to 20 simulations are performed to determine the reasonableness of the distributions of the input parameters. The analyst uses his or her judgment to refine the parameter distributions. Then, the process is repeated until the results are generally consistent with the measured data. At this point, 100 to 150 simulations are performed. For each simulation, predicted concentrations are compared to the measured values. If the root mean square error (RMSE), the square root of the sum of the squares of the differences between measured and predicted values, is less than the analyst's defined limit, the simulation is viewed as representing the measured data.

The results from all simulations that pass the RMSE criteria are used to generate maps of the average predicted concentration from the multiple simulations and maps of the probability of exceeding specified contamination threshold levels. Because selection of the value to use for the RMSE limit is up to the analyst, an experienced analyst is required to choose this number correctly. If the RMSE is too large, there will be a poor match with the measured data. If it is too small, many simulations will be needed to find a large enough set of simulations that pass the RMSE conditioning criteria to provide meaningful statistics for generating probability maps. The average concentration maps and the probability maps are used to represent the nature and extent of the contamination visually and to perform estimates of the volume of contamination as a function of contaminant threshold and probability of exceeding the threshold. The probability maps are also used to guide decisions on future well placement in sample optimization problems.

Description of Test Problems

Groundwater*FX* was used on two test problems, Site B sample optimization and Site S sample optimization. During the demonstration, the Decision*FX* staff commented that the time to perform such an analysis was extremely limited, citing examples from their own experience in which each analysis easily required a person-month of effort. Decision*FX* therefore requested to be allowed to extend the sample optimization problems to include cost-benefit analysis and thereby remove the need to perform the analysis on a different data set. The technical team agreed at the time of the demonstration that this was a reasonable approach to demonstrating Groundwater*FX*'s capabilities. Therefore, Decision*FX* used Groundwater*FX* to provide cost-benefit estimates of the volume of contamination above certain problem- and contaminant-specific concentrations. Decision*FX* also computed the exposure concentrations at receptor locations at future times as part of a human health risk assessment. As part of the demonstration, more than 20 visualization outputs were generated. A few examples that display the range of Groundwater*FX*'s capabilities and features are included in this review. A general description of each test problem and the analysis performed using Groundwater*FX* follows. Detailed descriptions of all test problems are provided in Appendix A.

Site B Sample Optimization and Cost-Benefit Problem

The Site B problem was a 2-D groundwater contamination problem. The data supplied for analysis of Site B included surface maps of buildings, roads, and water bodies; hydraulic head data; and concentration data for three contaminants—TCE, VC, and Tc-99—in groundwater wells at over 25 different locations

during a year of sampling. Initial sampling attempted to define the central region of the plume, which extends over one mile and approaches a nearby river. The objective of the sample optimization problem was to develop a sampling strategy to define the region in which the groundwater contamination exceeds specified threshold concentrations (Table 3) with probability levels of 10, 50, and 90%. The 10% probability region is the region in which there is at least a 10% chance that the contamination will exceed the threshold level. Therefore, the 10% probability region predicts the maximum volume of contamination and the 90% probability region predicts the minimum. Two threshold concentrations were specified for each contaminant (Table 3).

The probability of exceeding a threshold concentration is used in a cost-benefit analysis of cleanup goals vs. cost of remediation. The analyst was also asked to calculate health risks associated with drinking 2 L/day of contaminated groundwater at two exposure points, on the basis of current conditions and conditions 5 years in the future. 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.

Decision*FX* staff chose to demonstrate the visualization, sample optimization, and cost-benefit analysis capabilities of Groundwater*FX*. For sample optimization, Groundwater*FX* simulates the flow and transport of the contaminants using a probabilistic approach. For the Site B problem, 44 input parameters were required to define the source term, the unsaturated zone, and the saturated zone. Of these, 17 parameters were assigned statistical distributions to quantify uncertainties. The analyst makes an initial estimate of the model parameters and their statistical distributions and performs a number of simulations. Next, the analyst evaluates the predicted concentrations from the simulations against the measured data and refines the choice of input parameters. The process is repeated until the analyst is satisfied with the choice of input parameters. At this point, typically 100 to 150 simulations are made. The output is compared to the known data; if the output is not consistent with the measured data, it is not used in constructing average concentration or probability maps. Consistency is judged through statistical criteria (RMSE) defined by the analyst. Typically, 40 of the 100 to 150 simulations pass the consistency test.

Table 3. Site B groundwater contamination problem threshold concentrations

Contaminant	**Threshold concentrations**
TCE	50, 500 (µg/L)
VC	50, 250 (µg/L)
Tc-99	10,000, 40,000 (pCi/L)

Using the data from the simulations that pass the RMSE statistical conditioning test, the analyst used the software to generate plots of the probability of exceeding concentration thresholds to assist in visual evaluation of the areas of largest uncertainty. Groundwater*FX* uses an operations research algorithm to quantitatively select optimal well locations on the basis of probability of exceedence. Initially, three additional well locations were selected to refine the plume estimate. The model simulations were then repeated. An additional location, bringing the total of new sample locations to 4, was requested to further define the extent of contamination.

With the final data set, the analyst used Groundwater*FX* to generate the average concentration distributions and the probability distribution of exceeding the two threshold concentrations for all three contaminants (TCE, VC, and Tc-99). These distributions were posted on a bitmap of the site to provide a visual frame of reference for the plume location. The statistical data on the nature and extent of contamination were exported to Excel and used to generate a cost-benefit analysis of the volume contaminated vs. cleanup threshold. Groundwater*FX* was also used to estimate exposure concentrations at two receptor locations at the time the data were collected and 5 years after that time. These estimates were imported into Microsoft Excel and used for evaluating human health risks. Since the risk calculations were performed independently of the Groundwater*FX* software and depended entirely on the skill of the analyst and not the software, the risk calculations were not evaluated. An evaluation was performed of the exposure concentrations used for the risk calculation.

Site S Sample Optimization and Cost-Benefit Problem

The Site S sample optimization and cost-benefit problem focuses on a 3-D groundwater contamination problem for a single contaminant, CTC. The data supplied for analysis of this problem included geologic cross-section data, hydraulic head data, hydrologic and transport parameters, and contaminant concentration data from 24 monitoring wells. Of these, data were collected at 5-ft vertical intervals for 19 wells, while data for the other 5 wells were collected at 40-ft vertical intervals. A total of 434 contaminant sample locations and values were provided to the analyst. The objectives of this problem were to develop a sampling strategy to define the 3-D region of the plume at threshold concentrations of 5 and 500 µg/L at confidence levels of 10, 50, and 90%; to estimate the volume of contaminated groundwater at the defined thresholds; and to calculate human health risks to support cost-benefit decisions. 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 developed 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.

Decision*FX* staff chose to demonstrate the visualization, sample optimization, and cost-benefit analysis capabilities of Groundwater*FX*. To address the 3-D nature of the problem, the Decision*FX* analyst divided the subsurface into four layers. The thickness of these layers was prescribed, going from the top to the bottom of aquifer, as 10, 20, 31, and 65 ft. For wells with a 5-ft vertical spacing, there were often multiple data points within each layer. When this occurred, contaminant concentration data within these regions were averaged over the layer.

For sample optimization, Groundwater*FX* simulates the flow and transport of the contaminants using a probabilistic approach. For the Site S problem, 73 input parameters were required to model the source term, the unsaturated zone, and the saturated zone. Of these, 29 parameters were assigned statistical distributions to quantify uncertainties. The procedure of data evaluation follows the same steps as discussed for the Site B.

Using the data from the simulations that pass the statistical conditioning tests, the analyst generated plots of the probability of exceeding threshold concentrations to visually evaluate the areas of largest uncertainty. Groundwater*FX* uses an operations research algorithm to quantitatively select optimal well locations on the basis of the probability of exceeding a threshold concentration. Three additional well locations were selected to refine the plume estimate. The data from these locations were used to refine the definition of plume locations.

With the final data set, the analyst used Groundwater*FX* to generate the average concentration distribution and the probability distribution of exceeding the two threshold concentrations for CTC. The statistical data on the nature and extent of contamination were exported to Microsoft Excel and used to generate a cost-benefit analysis of the volume contaminated vs. the cleanup threshold. Groundwater*FX* was also used to estimate concentrations at two receptor locations at the time the data were collected and for 1, 5, and 10 years after that time. These estimates were imported into Excel and used for evaluating human health risks. Since the risk calculations were performed independently of the Groundwater*FX* software and depended entirely on the skill of the analyst and not the software, the risk calculations were not evaluated. An evaluation was performed of the exposure concentrations supplied for the risk calculation.

Evaluation of Groundwater*FX* *Decision Support*

As noted earlier, Groundwater*FX* was designed as a decision support tool to evaluate environmental information relative to the nature and extent of contamination in groundwater. The software quantifies uncertainties and provides objective recommendations on sample location, statistical information about the contamination, and visual feedback on the extent of contamination. In the demonstration, Decision*FX* used Groundwater*FX* to import data on contaminant concentrations from ASCII text files and on surface structures (e.g., roads, lakes, and buildings) from bitmap graphical image files. Groundwater*FX* demonstrated the ability to integrate this information on a single

platform and place the information in a visual context. Groundwater*FX* generated 2-D maps of concentration contours and the probability of exceeding threshold values that support data interpretation. The software was used in the demonstration to generate the data necessary for producing cost-benefit curves. The cost-benefit curves were produced in an auxiliary software (Microsoft Excel). Groundwater*FX* was also used to provide suggestions for new sample locations on the basis of probabilistic analysis performed using the existing data. In addition, estimates of exposure concentrations were calculated for use in human health risk analysis. The translation of exposure to human health risk estimates was also produced in Microsoft Excel. The accuracy of the analyses is discussed below in the section comparing Groundwater*FX* results with baseline data and analysis.

Documentation of the Groundwater*FX* Analysis and Evaluation of the Technical Approach

For each analysis, Decision*FX* provided a detailed description of the manipulations necessary to take the data provided, import it into Groundwater*FX*, and perform the desired analysis. The steps proceeded logically and in a straightforward manner. Manipulations to format the data within the Groundwater*FX* format were relatively simple. Files containing data were supplied to the analyst using a .dbf format. Prior to using these files in Groundwater*FX*, the analyst had to import these files into another program (e.g., Microsoft Excel), reformat them to make the columns of data fit the Groundwater*FX* format, and save them in ASCII text file format. Units of measurement were converted from feet to meters. Decision*FX* provided information to support the choice of the different model parameters and their statistical distributions used in performing the sample optimization problem. In addition, information on model selection and the parameters for contouring were provided in the output files and the problem documentation.

To estimate the probability levels as to whether a contaminant exceeds a threshold concentration, Groundwater*FX* used an approach that was slightly different from the approach used in the baseline analysis. Groundwater*FX* mathematically divides the problem domain into a number of rectangular regions. It then performs multiple simulations with the data to estimate the range of possible distributions of contaminants in each region consistent with the measured data. For each simulation, the analyst computes the volume (or area in two dimensions) that exceeds the threshold concentration. This distribution of volumes is used to calculate the statistical nature of the distribution in estimated volumes.

In contrast, the baseline geostatistical analysis used an approach consistent with the EPA Data Quality Objective guidance (EPA 1994). The site was mathematically divided into a number of rectangular regions. Within each region, an analysis was made to determine a single estimate of the concentration. Using the statistical properties of the data, the analyst calculated the confidence that the contamination concentration does not exceed the threshold concentration in each region. This approach places the confidence question in each region of the analysis. There is more uncertainty as to the concentration within each region as compared to the total over the entire site. Therefore, the spread in estimated contaminated volume should be slightly larger for the baseline approach than for the Groundwater*FX* approach.

This does not imply that the Groundwater*FX* approach to estimating the volume that contains contaminants above the threshold concentration is technically incorrect. The approach supplies different information. In fact, the multiple simulation approach can be a more robust approach than that used in the baseline analysis. In effect, the baseline approach provides one simulation of the data that is used for decision purposes. The Groundwater*FX* approach can provide multiple (50–100) simulations of the data. Groundwater*FX* could have used the information from each simulation to develop a distribution of contamination values in each region and then could have directly estimated the 90% confidence level. If done correctly, this approach can provide a more technically defensible estimate than that of the baseline approach.

In performing the risk calculation, the Decision*FX* analyst was asked to estimate the risk at two residential receptor locations for each problem. Decision*FX* estimated the exposure concentration at the two requested locations, assumed that the wells were part of a distribution system, and calculated the average of the two wells. This is a nonstandard practice for evaluation of human health risk. Typically, it is assumed that a single well supplies the water needs for a single residence. The averaging used by Decision*FX* causes a lowering of the peak

risk estimate. To arrive at the average value, Decision*FX* used results from the suite of Monte Carlo simulations to calculate the mean, the standard deviation, and the 95% confidence limit concentration at each receptor location. Output files provided by Decision*FX* contained this information, and the technical evaluation was based on this information.

Comparison of Groundwater*FX* Results with the Baseline Analysis and Data

Site B Sample Optimization and Cost-Benefit Problem

The data supplied for analysis of Site B included surface maps of buildings, roads, and water bodies; hydraulic head data; and concentration data for three contaminants (TCE, VC, and Tc-99) taken at 25 groundwater wells during one year of sampling. Wells in which high concentrations of contamination were detected were sampled on a monthly basis, while others were sampled less frequently. Initial sampling attempted to define the central region of the plume, which extends more than one mile and approaches a nearby river. The objective of this problem was to develop a sampling strategy to define the region in which the groundwater contamination exceeds specified threshold concentrations (Table 3) with probability levels of 10, 50, and 90%. Decision*FX* staff requested four additional samples in two rounds of sampling to complete their analysis using Groundwater*FX*. The small number of additional samples reflects the technical strength of using groundwater flow and transport simulation to determine sample locations.

The concentration maps generated by Groundwater*FX* were compared to the baseline analysis concentration map. The technical team, in a few cases, took the data set compiled by Decision*FX* after sample optimization was completed and generated concentration contour maps to gain a better understanding of the differences between the baseline and Groundwater*FX* approaches. The baseline analyses consisted of data evaluation using several contouring algorithims available in Surfer and GSLIB (e.g., IDW, ordinary kriging, and indicator kriging). Multiple lines of reasoning were used during the baseline data analyses, generating hundreds of output files and maps. The Surfer data analysis focused on the use of IDW and ordinary kriging algorithms to contour contaminant concentrations. The Surfer kriging estimates were obtained with an anisotropy ratio of 0.5 and a direction of –40° (the direction of groundwater flow). Similarily, the GSLIB analyses used indicator kriging with the additional refinement of specifying spatial correlation lengths for a series of contaminant concentrations. The best match to the baseline data for evaluation of the Groundwater*FX* results was selected by comparing and contrasting the multiple outputs. Each of these baseline analyses used the data set provided to Decision*FX* after completion of the sample optimization and should correspond closely to the Groundwater*FX* estimates at the 10, 50, and 90% probability levels.

This report presents the results for TCE contamination. Similar types of output were generated for VC and TC-99. The TCE contamination was chosen as the basis of the evaluation because the Decision*FX* analyst noted that the volume estimates generated for VC and Tc-99 were believed to be incorrect. Problems encountered with the analyst's choice of the RMSE conditioning criteria during the demonstration required a reanalysis of the data, and there was not enough time to repeat all three analyses. Therefore, Decision*FX* decided to repeat only the TCE analysis to demonstrate Groundwater*FX*'s capabilities. The reanalysis did not have a major impact on the average concentration map. However, it did alter the estimates of the volume of contamination, particularly at the 10 and 90% probability levels. The problems with setting the RMSE conditioning criteria reflect a lack of adequate time during the demonstration to perform the analysis using this software.

Figure 1 shows the Groundwater*FX* sample locations (marked by triangles) on a site map with major water bodies, buildings, and railroad lines. The sample location triangles are color-coded to represent the measured TCE concentrations. This map includes the original sample locations plus the four additional samples selected by Decision*FX*. All of the wells are labeled, although the labeling is difficult to see in the visualization reproduced in this report. The technical team imported this file into Microsoft PowerPoint and used the zoom feature to magnify the image and examine the visualization. This examination verified that wells were in the correct location and that the color coding represented the measured values correctly. The technical team added larger labels on two wells, MW-141 and MW-152, to illustrate a problem found in the Decision*FX* analysis. MW-141 is near the bend of a stream in the east-central part of the map;

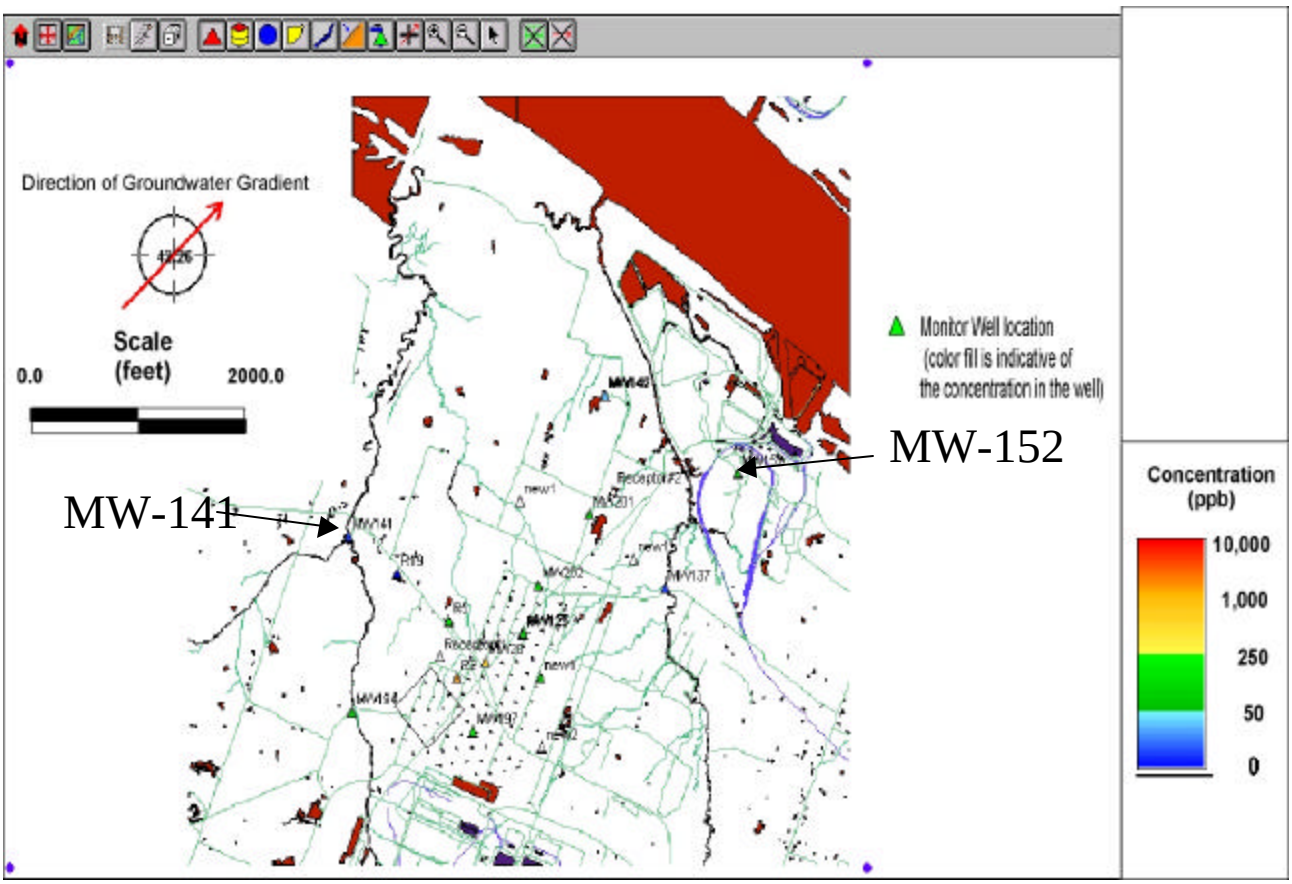

Figure 1. Groundwater*FX*-generated map for Site B with sample locations color-coded to match TCE concentration.

MW-152, located to the northeast of the large stream that drains into the river, is inside the blue loop that represents a railroad line.

Figure 2 is the site base map overlain with the average TCE concentrations as estimated by Groundwater*FX*. The threshold concentrations in the problem were designated as 50 and 500 µg/L. In Figure 2, concentrations estimated between 50 and 500 µg/L are green, and concentrations greater than 500 µg/L are orange, yellow, or red. Well locations are marked with triangles on the map and are color-coded. (This is difficult to see without enlargement.) The technical team noticed that the well locations were not plotted correctly on this site map. For example, it can be seen through comparison of Figures 1 and 2 that the locations of wells MW-141 and MW-152 have been moved by several hundred feet to the east and south. The cause for this inconsistency was determined to be operator error when combining the well locations with the background bitmap. The result moved the depiction of the contamination plume to the east and south, thus making direct comparison with the baseline analysis more difficult.

The technical team investigated the correlation between the plume map and the baseline data by importing Figure 2 into PowerPoint and enlarging the image. This review indicated that there was a poor match. At MW-152, data was collected monthly during the 1-year sampling period; the 12 measured values ranged from 201 to 245 µg/L. In Figure 2, the triangle representing MW-152 is color-coded green, consistent with the measured data (green represents 50–500 µg/L on the map). Even though the concentrations represented at the well locations are correct, the colored contour plume map in Figure 2 has this well located on the edge of the plume in the dark blue region (with dark blue representing 0 to 10 µg/L). Similar reviews of the data and the plume map were performed at MW-201 and MW-202. At MW-202, the 12 measured TCE concentrations ranged from 813 to 840 µg/L, and at MW-201 the TCE concentration ranged from 525 to 789 µg/L. The triangles representing these wells are

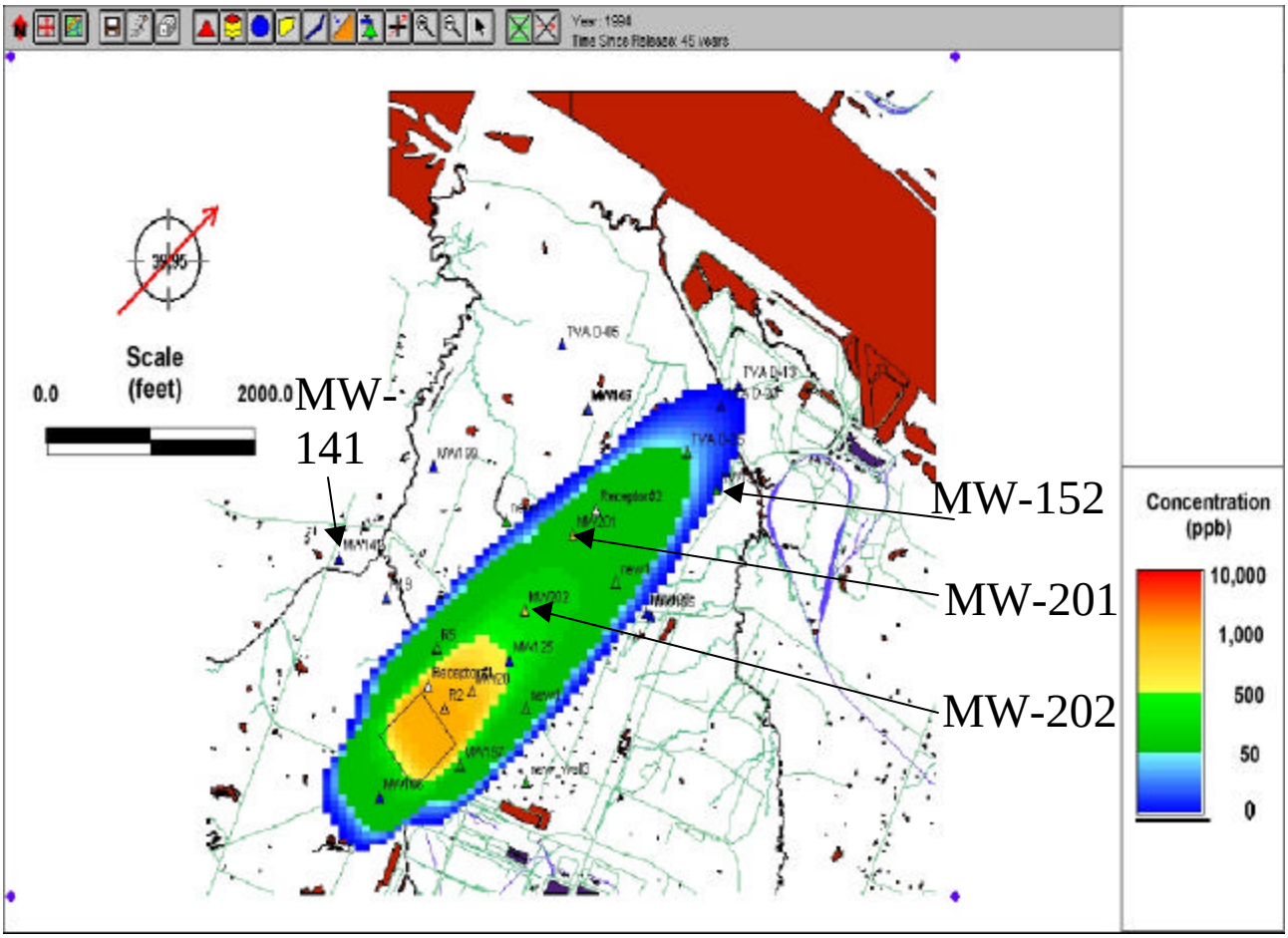

Figure 2. Groundwater*FX*-generated map of average TCE concentration at Site B at the time of the data collection.

both yellow, which represents a concentration greater than 500 µg/L (Figure 1). Again, this is consistent with the data. However, both of these wells are in the 50- to 500-µg/L zone (represented by green) of the plume map (Figure 2). The Groundwater*FX*-generated plume map also covers a much smaller area than would be expected, given the data.

Figure 3 represents the baseline analysis of the data set presented to Decision*FX* (original data plus data from the four locations determined through sample optimization) generated using the ordinary kriging interpolation in Surfer. TCE concentration contours at 50 and 500 µg/L are outlined in the figure. Well and receptor locations are marked. Figure 4 shows the baseline analysis produced with indicator kriging in GSLIB. In this figure, TCE concentrations between 5 and 500 µg/L are designated by blue; all other colors indicate concentrations exceeding 500 µg/L. In both baseline representations of the data, when more than one value was collected at a well location, the maximum value was used for interpolation. There are substantial differences between the baseline kriging interpretations of the data shown in Figures 3 and 4 and the Groundwater*FX* interpretation of the data shown in Figure 2. In both of the baseline analyses, the 500-µg/L contour extends much further to the north and east. Likewise, the 50-µg/L contour in the baseline analyses bends towards the east to include wells TVAD-25 and MW-152. The Groundwater*FX* analysis does not predict this shift to the north and east and consequently provides a poor match to the baseline data at these locations. The baseline interpolations are much more consistent with the data than is the Groundwater*FX* analysis.

In addition, both baseline analyses indicate that the 50-µg/L contour of the plume is not bounded to the north and east. This is consistent with the data because there are no sample locations down-gradient from MW-152, which has measured values between 201 and 245 µg/L. This implies that the sample

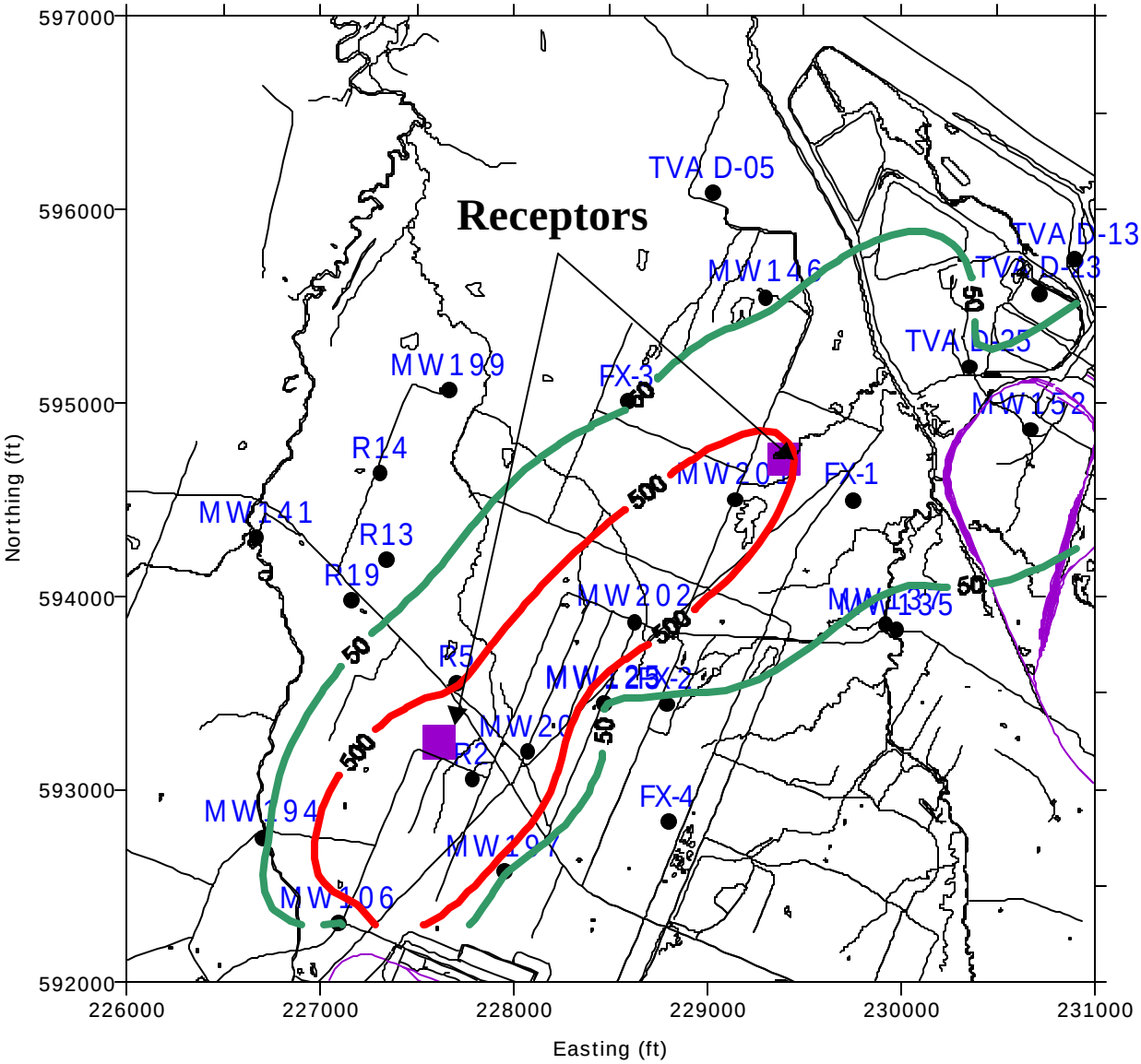

Figure 3. Baseline analysis of TCE concentration contours at 50 µg/L (green) and 500 µg/L (red) based on kriging interpolation with Surfer.

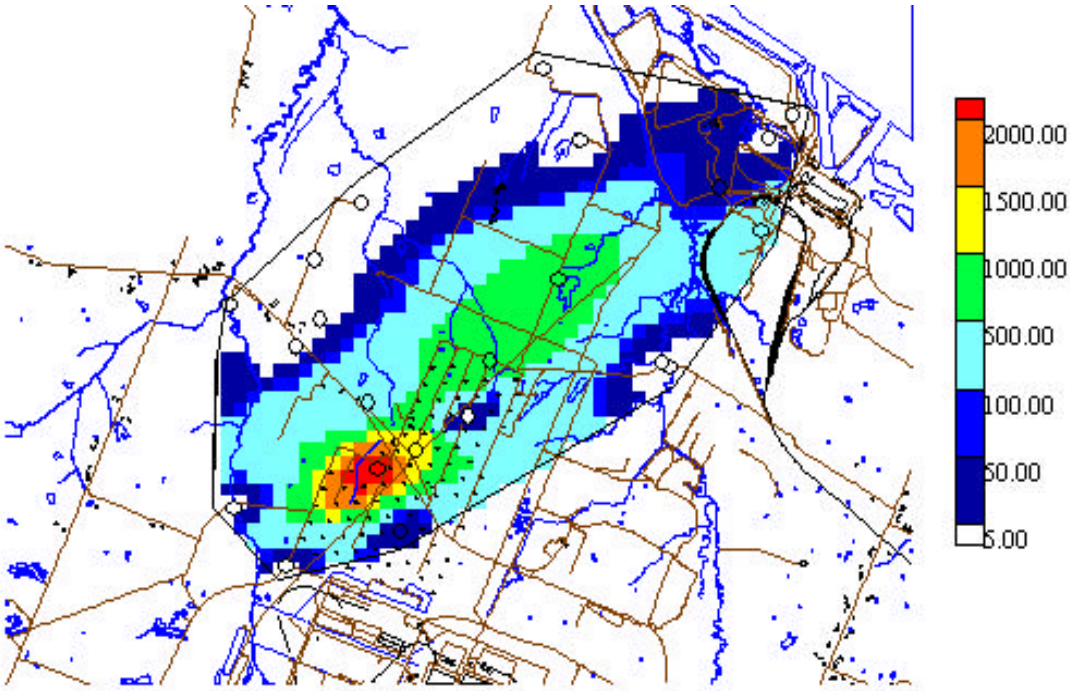

Figure 4. Baseline analysis of TCE concentration (µg/L) contours based on kriging using GSLIB.

optimization procedure in Groundwater*FX* may not have adequately characterized the plume.

Groundwater*FX* was also used to generate maps of the probability of exceeding the threshold concentrations for each of the three contaminants at each threshold concentration in Table 3. Figure 5 is the Groundwater*FX* map showing the probability that TCE exceeds the 50-µg/L threshold. The map contains a site map overlain by the probability map. In the probability map, regions in green have a 10 to 50% probability of exceeding the threshold, those in

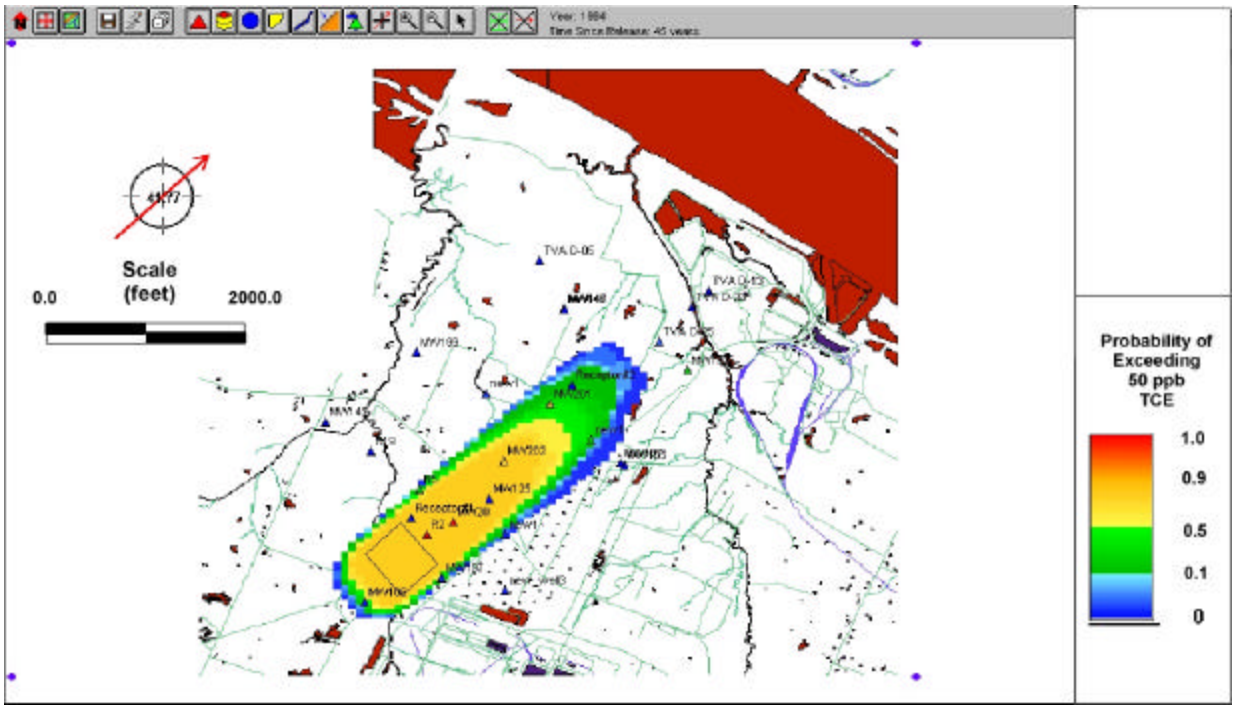

Figure 5. Groundwater*FX*-generated map of the probability of the TCE concentration exceeding 50 µg/L.

yellow have a probability of between 50 and 90%, and those in orange and red have a greater than 90% probability. The correlation between this map and the average concentration map generated by Groundwater*FX* is not clear. The average concentration map (Figure 2) shows a much larger area above the 50-µg/L concentration than does the probability map (Figure 5). Moreover, one would expect that the region of the plume with a concentration greater than 500 µg/L (depicted in yellow in Figure 2) would have a greater than 90% chance of exceeding 50 µg/L and be red in Figure 5. This is not the case.

For direct comparison with Figure 5, the technical team used indicator kriging in GSLIB to generate a map of the probability of exceeding the TCE threshold concentration of 50 µg/L (Figure 6) using the same data set as that used by Groundwater*FX*. Note there are large areas of red in Figure 6, indicating that there is a high probability that the 50-µg/L threshold has been exceeded; by comparison, no red areas appear in the Groundwater*FX*-generated probability map. In addition, the baseline analysis, as represented by Figure 6, indicates regions of high probability much further to the north and east as compared to the Groundwater*FX* analysis.

In its report documenting the analyses performed for the demonstration, Decision*FX* stated that for each Monte Carlo simulation that passed the RMSE conditioning criteria, the analyst calculated the volume of TCE-contaminated groundwater above the threshold concentration. This distribution of predicted volumes is used to define the volume estimate at the different probability levels. The 10% probability level volume estimate represents the volume for which only 10% of the estimated volumes are greater. The volume estimates were compared to the baseline analyses, which were derived through ordinary kriging using Surfer and indicator kriging using GSLIB.

Table 4 shows the estimates of the volume of contaminated groundwater at the 50% probability level generated by Groundwater*FX* and by the Surfer and GSLIB baseline analyses. The Groundwater*FX* estimates were approximately 70% lower than the Surfer baseline analyses at the 50-µg/L threshold and 50% lower at the 500-µg/L threshold. Likewise, the Groundwater*FX* estimates at the 50 µg/L and 500-µg/L thresholds were much lower than the estimates obtained using GSLIB. That the Groundwater*FX* volume estimates were consistently and substantially lower than the two baseline analyses at the 50% probability level indicates a poor match to the baseline analyses.

Table 5 presents the estimates of the volume of contaminated groundwater at the 10 and 90% probability levels generated by Groundwater*FX* and by the baseline GSLIB geostatistical analysis. At the 10% probability level, the Groundwater*FX* volume

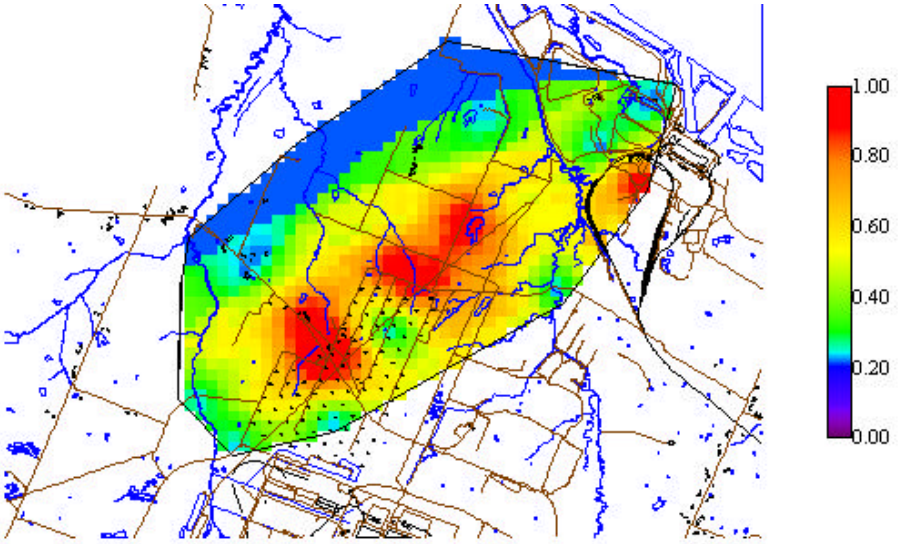

Figure 6. Baseline map of the probability of the TCE concentration exceeding 50 µg/L generated with GSLIB.

Table 4. Groundwater*FX* and baseline analysis volume estimates at the 50% probability level for the Site B TCE contamination problem

TCE threshold concentration	Groundwater*FX* estimate (ft^3)	Baseline estimates (ft^3)	
		Surfer analysis, ordinary kriging	GSLIB analysis, indicator kriging
50 µg/L	4.94E+07	1.74E+08	1.58E+08
500 µg/L	2.32E+07	5.40E+07	4.77E+07

Table 5. Groundwater*FX* and GSLIB volume estimates at 10% and 90% probability levels for the Site B TCE contamination problem

TCE threshold concentration	Estimate at 10% probability level (ft^3)		Estimate at 90% probability level (ft^3)	
	Groundwater*FX*	GSLIB	Groundwater*FX*	GSLIB
50 µg/L	6.25E+07	2.60E+08	3.42E+07	9.87E+07
500 µg/L	3.08E+07	1.03E+08	7.08E+06	4.25E+06

estimates were 76% lower than the baseline analysis for the 50-µg/L threshold and 66% lower for the 500-µg/L threshold, once again exhibiting the trend of Groundwater*FX* toward underestimating the volume of contaminated groundwater. At the 90% probability level, the Groundwater*FX* volume estimates were 65% lower than the baseline analysis for the 50-µg/L threshold but 66% higher for the 500-µg/L threshold.

The difference between the volume estimates at the maximum volume (10% probability level) and at the minimum volume (90% probability level) is much smaller for Groundwater*FX* than it is for the GSLIB baseline. This is particularly evident at the 500-µg/L threshold, where Groundwater*FX* volume estimates range from 7×10^6 to 3×10^7 (a difference of a factor of 4), while the baseline analysis volume estimates range from 4×10^6 to 1×10^8 (a factor of 25 difference). The cause for this difference is the technical approach used to estimate volumes. Groundwater*FX* performs multiple simulations and calculates the volume above the threshold for each simulation. This information is then used to calculate the probability of obtaining a certain volume. This method places the analysis on a global scale, as the entire problem domain is involved in the analysis. The baseline analysis estimates the concentration at each block of the modeled domain. Then estimates the probability that the concentration could exceed the threshold in each block. This places the analysis on a local (computational block) scale because it analyzes each block independently.

This difference in estimating volumes may partially explain the differences between the baseline and Groundwater*FX* analysis. However, the technical team still concluded that the Groundwater*FX* volume estimates are too low. This conclusion is based on the poor match between the data and the probability and concentration maps generated by Groundwater*FX* and on the observation that, at the 50-µg/L contour, the Groundwater*FX* volume estimate at the 10% probability level (6.3×10^7 ft^3 maximum volume) is still 50% lower than the baseline volume estimate at the 90% probability level (9.8×10^7 ft^3 minimum volume).

The technical team also noted the lack of consistency among the Groundwater*FX*-generated

estimates of contaminated volume as a function of probability levels and the probability maps. The Groundwater*FX* estimate of the volume of contaminated groundwater at the 90% probability level is consistent with the concentration map (Figure 2) but not with the probability map. The probability map (Figure 5) for the 50-µg/L threshold is not consistent with the measured data: it indicates that there is no area in which there is 90% probability that the concentration exceeds that threshold, but 8 of the 27 measured data values exceed the 50-µg/L threshold. Likewise, the probability map provided for the 500-µg/L threshold does not depict any region that is above the 90% probability level, yet 4 of the 27 measured values exceed the 500-µg/L level and the maximum measured value is 4648 µg/L. In contrast to the probability maps, volume estimates at the 90% probability level are nonzero, an indication the threshold has been exceeded.

Decision*FX* also used Groundwater*FX* to estimate exposure concentrations for assessment of human health risk at the two receptor locations. For the residential exposure scenario, the estimated groundwater concentrations for each constituent were used to estimate the 95th percentile upper confidence limit using Equation (1):

$$C_{95} = C_{\text{mean}} + Z_{95}(s/n^{1/2}) \qquad \text{(Eq. 1)}$$

where C_{95} is the 95th percentile concentration, Z_{95} is the standard normal variable for the 95th percentile, s is the standard deviation, and n is the number of samples. Decision*FX* decided to average the concentrations from the two receptor locations. From a technical perspective, this underestimates the maximum risk.

The Groundwater*FX* estimate for the 95th percentile TCE concentration was 506 µg/L. The technical team estimated the average concentration at two receptor locations (labeled on Figure 3) using kriging interpolation. For the first receptor, located near the highest TCE concentrations in the plume, the team estimated an average concentration of 1927 µg/L; for the second receptor, located near the edge of the 500-µg/L contour, the team determined an average concentration of 540 µg/L. Thus, the baseline average for these two locations is 1233 µg/L. It is clear that the estimate generated by Groundwater*FX* is low and inconsistent with the data and baseline analysis. The difference between the technical team's estimate and the Groundwater*FX* estimate would have been even larger had the technical team estimated the 95th percentile concentration. Given that the Groundwater*FX* 95th percentile TCE concentration was lower than the baseline estimates of the average concentration by at least a factor of 2, the technical team concluded that the Groundwater*FX* estimates are low and will lead to an underestimation of risk.

Groundwater*FX* was used to obtain estimates of the concentration 5 years into the future on the assumptions that the contaminant source was not removed and that groundwater flow remained unchanged. The predicted C_{95} estimate obtained as the average of the two well locations increased; however, the increase was only slight, to 605 µg/L. This is still lower than the technical team's estimates of the average concentration based on the initial conditions. The technical team did not attempt to produce a comparative analysis because of the difficulties in estimating an identical source term and flow rate consistent with those used by Decision*FX* and because the predicted future concentrations are clearly too low when compared to the baseline data. A comparative analysis of future predictions was performed for the Site S problem and is discussed later in this section.

A risk assessment was performed by using the exposure concentrations obtained by the Decision*FX* analyst. However, the analyst had to select the risk parameters and perform the risk calculations in Excel. Since risk assessment features are not part of the Groundwater*FX* software, these risk calculations are not evaluated.

A review of the Groundwater*FX* analyses for the two other contaminants, VC and Tc-99, led to similar conclusions about the performance of the software. For both VC and Tc-99, the Groundwater*FX* analysis tended to underestimate the spread of contamination as compared to the baseline data and analyses. The well locations were marked incorrectly (and in the same location as in the TCE analysis) for the Tc-99 analysis. However, the well locations were mapped correctly in the VC analysis.

Site S Sample Optimization and Cost-Benefit Problem

The data supplied for analysis of Site S included geologic cross-section data, hydraulic head measurements, and CTC concentration data for groundwater wells at 24 different locations during

one sampling period. Of the 24 wells, 5 were screened at three depths separated by 40 ft. The other 19 were screened at 5-ft intervals from the water table down to depths where further contamination was not detected. A total of 434 data points were provided to begin the analysis. The objective of this problem was to develop a sampling strategy to define the region in which the groundwater contamination exceeds 5 and 500 µg/L at confidence levels of 10, 50, and 90%.

The Decision*FX* analyst divided the subsurface into four layers. The thickness of the layers was prescribed, going from the top to the bottom of the aquifer, as 10, 20, 31, and 65 ft. For wells with 5-ft vertical spacing, there were often multiple data points in each layer. When this occurred, contaminant concentration data within these regions were averaged over the region. Using four vertical layers compressed the number of data points used in the analysis from 434 to 96. Decision*FX* requested 3 additional sample locations to complete the Groundwater*FX* analysis. The small number of additional samples reflects the technical strength of using groundwater flow and transport simulation to determine sample locations.

Using the data set that included the data from the three additional sample locations, the Groundwater*FX* analyst generated 2-D contour maps showing contaminant concentrations in each of the four layers and maps of the probability of exceeding the threshold concentrations of 5 and 500 µg/L for each layer. The concentration maps generated by Groundwater*FX* were compared to the baseline analysis concentration map. The original baseline analysis was performed at 10-ft vertical intervals that were substantially different from those chosen by Decision*FX*. The coarser vertical discretization used by Decision*FX* produced slightly different results than obtained in the original baseline analysis. To remove any differences between the baseline and the Decision*FX* analysis of the Site S problem, the baseline analysis was repeated using the four layers used in the Groundwater*FX* analysis, and the data set obtained by Decision*FX* after sample optimization was completed. In a few cases, the technical team used a more complete data set (based on an analytical solution to the flow and transport problem) than that supplied to Decision*FX* to generate concentration contour maps. This permitted a better understanding of the differences between the analytical solution (based on a more complete data set), the repeated baseline analysis using the Decision*FX* data set, and the Groundwater*FX* analysis.

Figure 7 is a composite of four bitmaps of screen captures of the Groundwater*FX*-generated maps for the CTC concentration in the four layers: layer 1 located 30–40 ft above mean sea level (MSL), layer 2 at 10–30 ft above MSL, layer 3 at 21 ft below to 10 ft above MSL, and layer 4 at 21–86 ft below MSL. The top of the water table is at 40 ft above MSL. Concentrations are color-coded as indicated in the color key provided at the bottom of the figure. Red, orange, and yellow indicate regions above 500 µg/L; green indicates regions between 5 and 500 µg/L; and blue indicates regions below 5 µg/L. The labeled monitoring well and receptor locations in Figure 7, though difficult to read, provide some frame of reference for the location of the concentration contours. The two receptor locations are marked with a triangle on each map. One receptor is located along the western edge of the current plume south of the plume midpoint. Although Figure 7 does not provide a scale of reference, Figure 8 indicates that the receptor location is near northing 251500 and easting 1296900. The second receptor is to the south of the current plume near the center of the plume in the east-west direction (northing 250000, easting 1297100). Groundwater flow is towards the south and in time, the second receptor will be exposed to contamination. The rectangular area on each map is the modeled source region because the highest Groundwater*FX*-predicted concentrations (layer 1) are in this area. Figure 7 appears to indicate that the bulk of the predicted contamination is in layer 1, with progressively less contamination in the deeper layers. Layer 1 is the only region with predicted concentrations in excess of the 500-µg/L threshold concentration (the yellow region in layer 1). All layers have predicted contamination between 5 and 500 µg/L (green region). Figure 7 appears to show that some of the predicted contamination has migrated north (opposite to the groundwater flow direction) of the source region (rectangle with highest concentrations). This is most likely a numerical artifact. Although a scale was not provided with the maps, it can be determined from the well locations that the Groundwater*FX* prediction indicates contamination has migrated 400 to 500 ft north (upstream) from the source region. This may be due to the modeling of dispersion processes, however, the spread upstream appears to

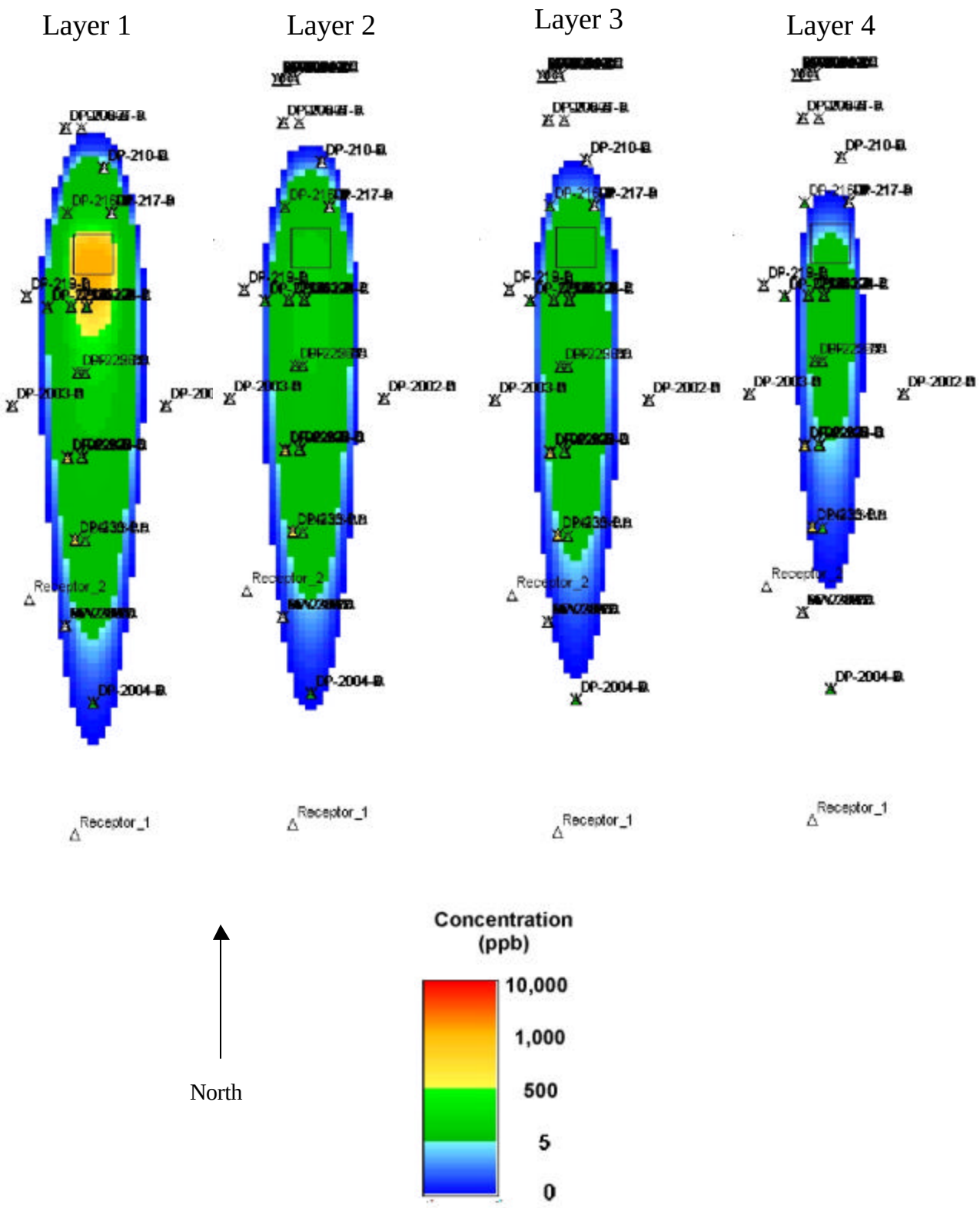

Figure 7. Groundwater*FX*-simulated average CTC concentrations in the four layers based on original data plus three additional samples.

be excessive compared to the technical team's observations on these types of problems.

The technical team also noted by comparing Figures 7 and 8 that the Groundwater*FX* analyst located the source region downstream from the measured peak CTC concentrations. This is clearly incorrect. In addition, the analyst did not account for the vertical component of groundwater flow that was evident in the data and described in the test problem. The analyst's choices of improper location of the source and omission of vertical flow component adversely impacted the Groundwater*FX* predictions and, as will be discussed, led to an inaccurate analysis.

The baseline analysis performed by kriging interpolation of the data supplied to Decision*FX* using Surfer is presented in Figure 8. The four layers correspond to those used by Groundwater*FX*. Small circles in the figure are well locations; some are labeled to provide a frame of reference. Receptor locations are marked with a diamond. A comparison of Figures 7 and 8 shows large differences. The baseline analysis for layer 1 (Figure 8) shows a small, narrow plume extending approximately 600 ft for the 500-µg/L contour (red zone) and 1,000 ft for the 5-µg/L contour (blue zone). By contrast, the Groundwater*FX* analysis shows the 5-µg/L contour extending approximately 4,000 ft. In the first three layers, the baseline analysis shows contamination much further to the north than is shown in the Groundwater*FX* analysis. The highest measured contamination occurred at wells DP-201 and DP-202 at a northing of approximately 255,000 ft. This baseline map is consistent with the data. The Groundwater*FX* peak concentration occurs at a northing of 253,800 ft, which is 1200 ft south of the peak values. The cause for this discrepancy is believed to be the source location chosen by the Decision*FX* analyst. Although the precise location of the source was not identified in the test problem, it could be located by the peak contaminant concentrations given to the analyst. Location of the source downgradient of the peak concentrations is incorrect and is indicative of operator error. Even had the Decision*FX* analyst located the source correctly, the length of the predicted plume is much longer than shown in the baseline analysis. Comparison of the other layers also shows major differences. In the baseline analysis, the 5-µg/L contour becomes successively longer, and the center of mass moves further south in each successive layer (i.e., as depth increases). This is consistent with the data and is indicative of a plume that is moving deeper as it travels to the south. In contrast, the Groundwater*FX* data shows the plume length getting smaller with depth. The baseline data and analysis also show each layer to have a region that exceeds the 500-µg/L threshold concentration. Groundwater*FX* did not indicate any contamination above 500 µg/L in layers 2 through 4.

Figure 9 supplies the technical team's concentration contours at 5 and 500 µg/L in the four layers used by Decision*FX* based on the analytical solution. The plume as derived from the analytical solution (Figure 9) is symmetric and is narrower and better-defined than the plume derived from the baseline analysis (Figure 8). These differences can be attributed to the increased information (greater number of data points) available for depicting the plume in the analytical solution. Comparison of the concentration maps (Figures 8 and 9) with the Groundwater*FX* average concentration maps (Figure 7) indicated that the Groundwater*FX* concentration maps were not consistent with the data. At many locations with high measured CTC concentrations, Groundwater*FX* predicted low concentrations. In order to gain a better understanding of the discrepancy, the technical team reviewed the input files prepared by Decision*FX*. The Decision*FX* analyst imported the initial data files into Excel and processed the data to obtain the average concentration in each layer. The review indicated that processing of the data was performed correctly. Thus, Groundwater*FX* started with the same data as used in the baseline analysis; however, it did not generate accurate maps with the data.

As part of the analysis, Groundwater*FX* was used to calculate the probability of exceeding the 5- and 500-µg/L CTC thresholds throughout the problem domain. Groundwater*FX* used this probability information in optimizing the selection of new sample locations. Figure 10 is a screen capture from Groundwater*FX* that presents the probability of exceeding 5 µg/L in layer 1 (the top 10 ft of the aquifer) at the current time, based on the final data set. Similar screen captures were provided for all layers and for both threshold concentrations at four times (the initial time and 1, 5, and 10 years into the future). In Figure 10, well identifiers and receptor locations are marked to provide a frame of reference. However, coordinate locations are not provided. A color key is provided, with the areas of highest probability in red and areas with the lowest

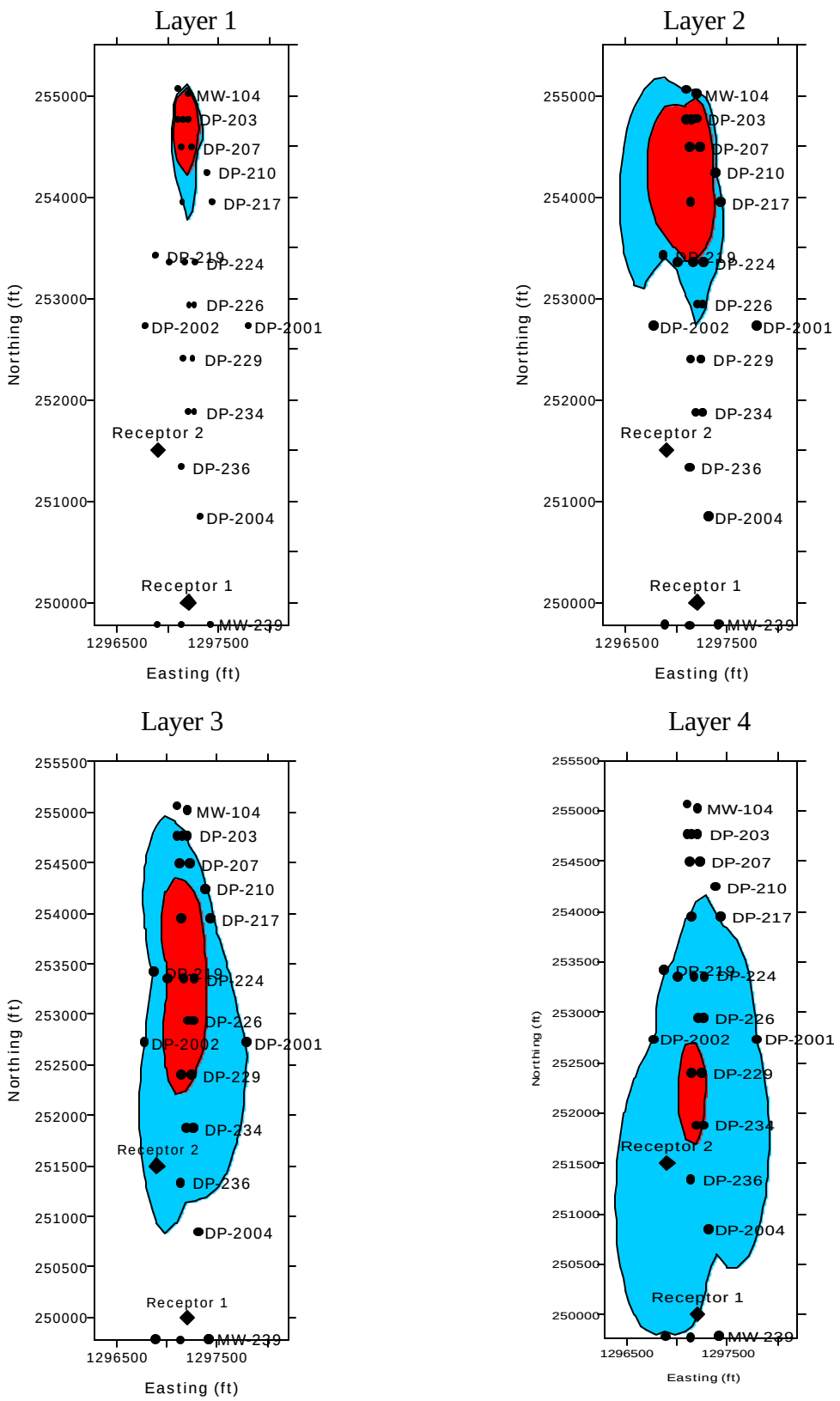

Figure 8. Baseline analysis of CTC concentrations at 5-µg/L (blue) and 500-µg/L (red) contours based on Decision*FX* data set.

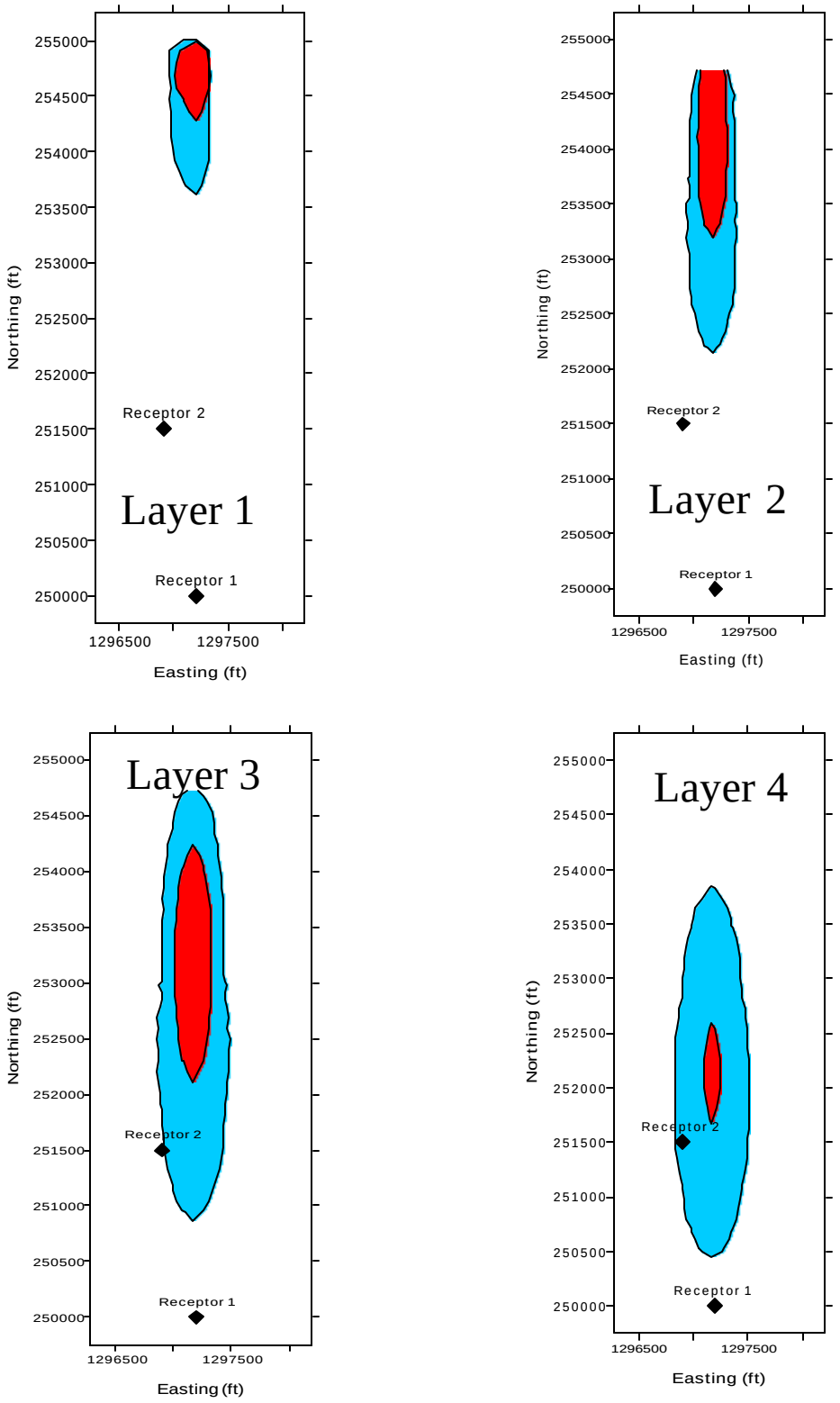

Figure 9. Baseline analysis using the analytical solution to provide data points to generate contours at 5- and 500-µg/L CTC thresholds.

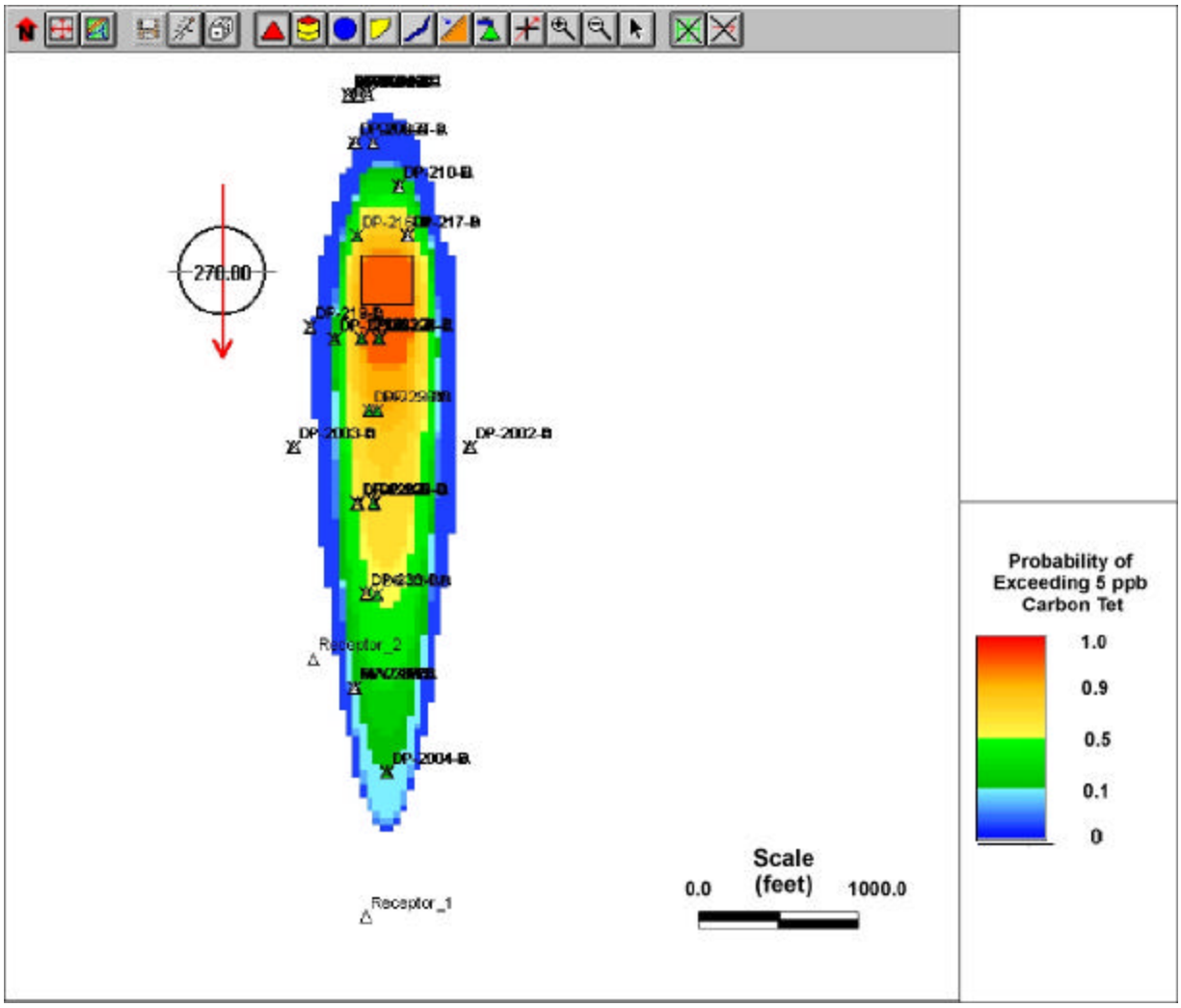

Figure 10. Groundwater*FX* map of the probability of exceeding 5 µg/L in layer 1 based on initial data.

probability in blue. The transition between yellow and green marks the 50% probability level. A comparison of Figure 10 with Figure 7, the Groundwater*FX*-generated map of average CTC concentrations, shows general agreement. Regions depicted as having an average concentration greater than 5 µg/L (green and yellow regions in Figure 7) have a greater than 50% probability of exceeding the threshold (yellow and red regions in Figure 10).

The CTC concentration and probability maps generated by Groundwater*FX* (Figures 7 and 10) were inconsistent with the data, the baseline analysis obtained using the same data as Groundwater*FX* (Figure 8), and the analytical solution (Figure 9). A review of the original data set supplied to Decision*FX* showed that 102 of the 434 data points had CTC concentrations greater than 500 µg/L, with the peak concentration exceeding 24,000 µg/L. In averaging the data into four vertical layers, between 3 to 7 data points (from a total of 27) in each layer exceeded the 500-µg/L threshold. For all layers, a total of 22 of the 108 data points were above this threshold. However, the Groundwater*FX* concentration maps did not show contamination above 500 µg/L in the three lowest layers. Groundwater*FX* was used to estimate, as a function of probability, the volume of contaminated groundwater above the two threshold values of 5 and 500 µg/L (Table 6). The technical team performed a baseline analysis using the same data provided to Decision*FX* after completion of the sample optimization. Baseline estimates were generated using kriging interpolation models in Surfer and are provided for each layer and for the entire site. As can be seen in Table 6, the Groundwater*FX* estimates at the 50% probability level were an order of magnitude lower than the technical team's

Table 6. Groundwater*FX* volume estimates of CTC-contaminated groundwater for the Site S sample optimization problem

CTC threshold concentration	Volume of contamination (ft^3)		
	10% probability level	50% probability level	90% probability level
5 µg/L	9.62E+7	4.39E+7	5.07E+6
500 µg/L	6.56E+6	8.87E+5	0

Table 7. Baseline volume estimates of CTC-contaminated groundwater for the Site S sample optimization problem

Layer	Volume (ft^3) > 5 mg/L	Volume (ft^3) > 500 mg/L
1. Surface (30 to 40 ft above MSL)	2.7E+6	1.59E+6
2. 10 to 30 ft above MSL	3.56E+7	1.67E+7
3. 20 ft below MSL to 10 ft above MSL	9.18E+7	2.28E+7
4. 85 to 20 ft below MSL	2.97E+8	1.40E+7
All layers	4.27E+8	5.51E+7

estimate at the 5-µg/L level and more than a factor of 50 lower at the 500-µg/L threshold level. In addition to using the Decision*FX* data set for estimating volumes, the analytical solution provided another basis for comparison. Comparison of the kriging baseline volume estimates to estimates obtained from the analytical solution indicated that the analytical solution estimates were 30 to 50% lower, resulting from better definition of the plume, as demonstrated in Figures 8 and 9 and discussed above. The agreement to within 50% is reasonable and consistent with the differing amounts of data used in the two analyses.

The technical team concluded that the Groundwater*FX* estimates were a poor match to the baseline volume estimates. Figures 8 and 9 along with Table 7 indicate that there are substantial volumes of contaminated groundwater in the lower layers. This is inconsistent with the concentration maps produced by Groundwater*FX*. The poor match between the data and the Groundwater*FX* concentration maps is believed to be the cause for the poor volume estimates. For example, the thickest vertical layer, layer 4, is the deepest; and the baseline analysis indicates that almost 70% of the contaminated volume above the 5-µg/L concentration threshold is in this layer. By contrast, Groundwater*FX* predicted that layer 4 had the smallest area of contamination as compared to all of the layers (see Figure 7).

Because of the poor match between the Groundwater*FX* analysis at the 50% probability level and the baseline analysis, the technical team concluded that it would not be meaningful to perform a comparison based on a geostatistical analysis of the data. However, even without the geostatistical analysis it is clear that the Groundwater*FX* 10% and 90% probability levels will not correspond to the data. For example, Groundwater*FX* indicates that at the 90% probability level, there is zero volume contaminated above 500 µg/L. However, approximately 20% of the data supplied to Groundwater*FX* exceeded the 500-µg/L threshold.

Decision*FX* also used Groundwater*FX* to estimate the exposure concentrations for a human health risk assessment at the two receptor locations. Decision*FX* followed the same approach as for Site B. For the residential exposure scenario, the estimated concentrations of CTC in groundwater were used to estimate the 95th percentile upper confidence limit using Equation 1. Decision*FX* combined the predicted concentrations at the two receptor locations to get an average concentration for risk at the site. Averaging underestimates the maximum human health risk.

Table 8 presents the Groundwater*FX* estimates for the mean and the 95th percentile CTC concentrations and the technical team's estimates of the average concentration at the two receptor locations (labeled on Figure 9) obtained using the same data as supplied to Decison*FX* after sample optimization. Groundwater*FX* predicts that both receptors would be exposed to concentrations greater than 5 µg/L. However, this is not consistent with the average concentration maps presented in Figure 7, which indicate that neither receptor would be exposed. The reason for this discrepancy could not be determined. Decision*FX* supplied the average exposure concentration at the two receptor locations for each of the Monte Carlo simulations that passed the RMSE conditioning criteria. However, the receptor locations were supplied on a local coordinate system (i.e., a coordinate system used by the Groundwater*FX* model). The technical team could not match the local coordinate system with the global system used to supply the data. Therefore, the exact location at which these concentrations were predicted to occur could not be determined.

As Table 8 indicates, the baseline average value is much lower than the Groundwater*FX* value for receptor 1 and much higher for receptor 2. The baseline analysis indicates that the contaminant has not reached receptor 1 at the initial time. This is consistent with the data. It is fortuitous that the maximum concentration of the two receptors for the baseline and the Groundwater*FX* analyses are almost identical. However, receptor 2 receives the highest exposure in the baseline analysis, while receptor 1 receives the highest exposure in the Groundwater*FX* analysis.

Groundwater*FX* was used to estimate the exposure concentrations at the two receptor locations for up to 10 years into the future if the source of contamination remained in place. Table 9 presents

Table 8. Groundwater*FX* and baseline estimates for current CTC exposure concentrations (µg/L) for the Site S residential risk evaluation

Receptor location	Baseline average	*FX* Average	*FX* C_{95}
1	0	258	397
2	240	24	38

Table 9. Groundwater*FX* and analytical estimates over time for CTC exposure concentrations (µg/L) for the Site S residential risk evaluation

Year	Receptor 1 location		Receptor 2 location	
	Analytical concentration	*FX* mean	Analytical concentration	*FX* mean
Current	0.2	258	18	24
1	92	331	34	30
5	239	896	65	73
10	404	2600	65	192

the Groundwater*FX* results and the analytical (known) concentrations for the test problem. From the concentration values for the analytical solution, it can be seen that the contamination does not reach the receptor 1 location in high concentrations until a year into the future. The concentration then continues to increase steadily over the next 9 years.

The concentrations predicted by Groundwater*FX* at the receptor 1 location are always much higher than the values given by the analytical solution and appear to be increasing more rapidly than the analytical solution values. For the receptor 2 location, the Groundwater*FX* values match the analytical solution reasonably well except around the 10-year time frame. The analytical solution for receptor 2 indicates a leveling off in CTC concentration after 5 years that is not shown in the Groundwater*FX* analysis. For the current conditions, the analytical solution indicates that receptor 2 receives higher exposure than receptor 1. By contrast, the Groundwater*FX* solution indicates receptor 1 always receives the highest exposure. Overall, Groundwater*FX* predicts much higher exposure concentrations than does the analytical solution. This is due to the overprediction of concentrations at receptor 1.

The accuracy of the Groundwater*FX* analysis as compared to the analytical solution is difficult to judge because of the problem in determining if the local coordinates used by Decision*FX* correspond to the same global coordinates as used for the receptors in the test problem and analytical solution. Assuming the coordinate systems are the same, the concentrations predicted by Groundwater*FX* at receptor 2 accurately matched the analytical solution for the first 5 years. The match at receptor 1 was poor, particularly at the current time and 10 years into the future.

The analytical solution indicated that the plume thickness was much less than the thickness of layer 4 (65 ft). The thickness of the plume could have been determined from the data supplied to the developer. Using the larger thickness caused a dilution effect and lowered the exposure concentrations. In addition, the analytical solution showed substantial contamination beneath the depth of layer 4 at the receptor 1 location. Both facts suggest that the Groundwater*FX* analysis should have been repeated with a finer vertical resolution. However, there was not time for the Decision*FX* analyst to repeat the analysis during the demonstration.

A risk assessment was performed by the Decision*FX* analyst using the exposure concentrations obtained by Groundwater*FX* in Microsoft Excel. However, the analyst had to make all of the decisions pertaining to selection of parameters and calculation of risk in Excel. Because the risk assessment feature is not part of Groundwater*FX*, the risk calculations were not evaluated.

Comment on the GroundwaterFX Site B and S Analyses

In both Groundwater*FX* analyses, there was a poor match between the output of Groundwater*FX* and the data and baseline analyses. The technical team could not determine any single reason for this, although a number of possible reasons were noticed. In particular, the analyst's choice of source location and neglect of the vertical component of flow on Site S basically precluded the model from matching the data. The Groundwater*FX* conceptual approach using Monte Carlo simulations is robust and should be able to perform a defensible analysis that matches the data. Following a review of the Groundwater*FX* results, the technical team concluded that the analyses were essentially a preliminary examination of the data and that the process would need to be repeated to refine parameter choices before either analysis could be considered to be representative of the baseline data and complete. Decision*FX* stated in its report that analysis of similar contamination problems could require two person-months of effort. In the demonstration, only 12 days were spent on the two problems, including the preparation of the documentation. In its report, DecisionFX also stated that "in the time allowed for the demonstration we were not able to get the quality of results normally sought in this type of analysis." In any event, although the technical approach appears promising in principle, it was not possible to determine if Groundwater*FX* can accurately estimate the extent of groundwater contamination.

Multiple Lines of Reasoning

Decision*FX* used Groundwater*FX* to provide a number of different approaches to examine the data. The foundation of the Groundwater*FX* approach is a Monte Carlo simulator that produces multiple simulations of the extent of contamination that are consistent with the known data. From these simulations, contaminant concentration and probability maps were produced to assist in data evaluation. The interpretations of statistical data permit the decision maker to evaluate future actions, such as determining sampling locations or

developing cleanup guidance, on the basis of the level of confidence placed in the analysis.

Secondary Evaluation Criteria

Ease of Use

Groundwater*FX* is a sophisticated flow and transport software that incorporates Monte Carlo simulation in a 3-D framework. A high level of skill and experience is required to use it effectively. All members of the technical review team who received training on this software noted that this product was complex and involved a high level of technical detail.

Several features of Groundwater*FX* make the software package cumbersome to use. These include the need for a formatted data file for importing location and concentration data, the need to have all units of measurement in meters (USGS and state plane coordinates systems are typically measured in feet), the need to have all graphic files imported as a single bitmap (which prohibits the use of multiple layers in visualizations and requires coordinates of the bitmap to be provided when the bitmap is used as a base map for visualization), the inability to edit graphic bitmap files, and the absence of on-line help. Visualization output is limited to bitmaps of screen captures that can be imported into other software for processing. Overcoming these limitations to perform an analysis requires more work on the part of the software operator—e.g., reformatting data files in an Excel spreadsheet and changing coordinates expressed in feet to meters to match the needs of Groundwater*FX*.

Groundwater*FX* exports text and graphics to standard word processing software directly. Graphic outputs are generated as bitmaps which can be imported into CorelDraw to generate .bmp, .jpg, and .cdr graphic files. Groundwater*FX* generated data files from statistical analysis and concentration estimates in ASCII format, which can be read by most software.

Efficiency and Range of Applicability

Groundwater*FX* was used to perform two sample optimization/cost-benefit problems with 12 person-days of effort. This included 2 days for post-processing of the bitmap graphic files, 1.5 days for post-processing of cost-benefit data on volumes of contamination, 1 day preparing a catalog of all files generated during the demonstration, and 4 days preparing the report documenting model assumptions, model outputs, and conclusions. The technical team concluded that the analyses were, at best, a first pass through the problem; the procedure would need to be repeated several times to improve the accuracy of the analysis. The incomplete analysis was due primarily to the combination of the sophisticated approach of the software—e.g., Monte Carlo simulation of 3-D flow and transport —and the time constraints of the demonstration. However, other ease-of-use issues, such as the need to process much of the input and output in software other than Groundwater*FX*, have a negative impact on efficiency.

Groundwater*FX* provides the flexibility to tailor the analysis to most groundwater contamination problems. It provides models for the source, vadose zone, and aquifer. The user has control over the choice of the many input parameters used to represent the flow and transport problem and the statistical distribution of these parameters.

Training and Technical Support

Decision*FX* provides a users' manual that discusses input parameters and contains screen captures of the pull-down menus used in the code. Technical support is supplied through e-mail. A 3-day training course is planned.

*Additional Information about the Groundwater*FX *Software*

Groundwater*FX* is a sophisticated software product and requires a skilled operator. To use Groundwater*FX* efficiently, the operator should be knowledgeable in probabilistic modeling of groundwater flow and contaminant transport. Knowledge pertaining to managing database files, contouring environmental data sets, conducting sample optimization analysis, and performing cost-benefit problems is also beneficial.

During the demonstration, Groundwater*FX* operated on a Windows 95 system. Two PCs were used for the demonstration. The first machine was a Micron 200-MHz Pentium with 64 MB of RAM, an 8.1-GB hard drive, a ZIP drive, an HP Model 8100 CD-Writer; and an external JAZ drive. The writing capabilities of the CD were used to provide output files containing data and visualizations for review. The JAZ drive was used to import data for the test problems. The second machine was a laptop SONY model PCG-719 with a 233-MHz Pentium MMX CPU, 32 MB of RAM, and a 2.1-GB hard drive. In

addition, training demonstrations were performed on a Macintosh machine to demonstrate that the software works on this platform, but the Macintosh was not used explicitly for the demonstration test problems.

DecisionFX plans to sell GroundwaterFX for $1000 for a single license. It will be supplied at no cost to State and Federal regulators.

Summary of Performance

A summary of the performance of Groundwater*FX* is presented in Table 10. The technical team observed that the main strength of Groundwater*FX* is its technical approach using Monte Carlo simulations of flow and transport processes to address variability and uncertainty in groundwater contamination problems. The use of groundwater simulation models should be beneficial in sample optimization designs as compared to purely statistical or geostatistical simulation models. However, the analyses performed by Groundwater*FX* did not provide an adequate match to the data and baseline analyses for either test problem. For Site B, monitoring well locations on some simulations were incorrectly plotted on the site map. The contaminant concentration maps were generally consistent with the data near the source of contamination. However, the leading edge of the plume was not represented accurately by Groundwater*FX*. The maps of the probability of exceeding a contaminant threshold were inconsistent with the data, and the Groundwater*FX* estimate of the volume of the plume was three to five times smaller than that obtained in the baseline analyses. In the Site B problem, estimates of exposure concentrations for risk calculations were too low by a factor of 2 to 3 as compared to the baseline analysis. For Site S, the contaminant concentration estimates were an extremely poor match to the data and baseline analysis. This caused estimates of the volume of contaminated groundwater and of exposure concentrations for risk calculations to be substantially different from the data and baseline analysis. In addition, the Groundwater*FX* estimates for exposure concentrations supplied for risk calculations were inconsistent with the Groundwater*FX* contaminant concentration maps. The technical team also concluded that the many ease-of-use issues identified earlier made the software cumbersome to use. In particular, visualization capabilities were limited, and the ability to only import graphic files in bitmap format can lead to problems in the analysis.

Table 10. Groundwater*FX* performance summary

Feature/parameter	Performance summary
Decision support	Groundwater*FX* is a probabilistic-based software product designed to address 3-D groundwater contamination problems, including optimization of new sample locations and generation of cost-benefit information (e.g., evaluation of the probability of exceeding threshold concentrations). The software generated 2-D maps of the contamination concentration and of the probability of exceeding a specified contamination concentration. Cost-benefit curves of the cost (volume) of remediation Vs. the probability of exceeding a threshold concentration were generated in Excel using Groundwater*FX* output files. Estimates of exposure concentrations in the present and in the future were prepared for use in human health risk calculations. The interpretations of statistical data permit the decision maker to evaluate future actions such as sample location or cleanup guidance on the basis of the level of confidence placed in the analysis.
Documentation of analysis	A detailed report documented the process, assumptions, and parameters used in the analysis. Output data files were provided to supplement the documentation.
Comparison with baseline analysis and data	The analysis performed by Groundwater*FX* did not provide an adequate match to the baseline data on either test problem. For Site B, well locations on some simulations were incorrectly plotted on the site map. The contaminant concentration maps were generally consistent with the data. However, the probability of exceedence maps were inconsistent with the baseline data, and the size of the plume was three to five times smaller than that obtained in the baseline analyses. Site B estimates of exposure concentrations for risk calculations were too low by a factor of 2 to 3. For Site S, the contaminant concentration estimates were an extremely poor match to the data and baseline analysis. This caused estimates of the volume of contaminated groundwater and exposure concentrations for risk calculations to be substantially different from the baseline data and analysis.
Multiple lines of reasoning	Groundwater*FX* provides concentration maps, probability maps and statistical evaluation of the model predictions that assist in multiple evaluations of the problem.
Ease of use	In general, the software is difficult to use for the following reasons: • Visualization output is limited to bitmaps of screen captures. • The software can only import bitmaps for use in visualization. • Maps cannot be annotated and modified (e.g., add scales); this must be performed in auxiliary software. • Data from statistical simulations cannot be processed; this task must be performed in auxiliary software. • Concentration data must follow a fixed format, and units of measurement must be in meters. • On-line help is not available.
Efficiency	Two problems completed and documented with 12 person-days of effort. However, the review team felt that the analysis would have been improved if more time had been available to complete the analysis.
Range of applicability	Groundwater*FX* provides the flexibility to tailor the analysis to most groundwater contamination problems.
Training and technical support	Users' manual One 3-day training course planned Technical support provided through e-mail Tutorial examples not provided with the software
Operator skill base	To efficiently use Groundwater*FX*, the operator should be knowledgeable in probabilistic modeling of groundwater flow and contaminant transport. Knowledge of sample optimization analysis and performing cost-benefit problems would be beneficial.
Platform	Demonstrated on a PC with Windows 95; can also operate on a Macintosh
Cost	$1000 for a single license; free to state and federal regulators

Section 5—Groundwater*FX* Update and Representative Applications

Objective

The purpose of this section is to allow the developer to provide information regarding new developments with its technology 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.

Groundwater*FX* Update

Since the EPA's Environmental Technology Verification (ETV) demonstration of DSSs took place in the fall of 1998, the Groundwater*FX* code has been updated with some new features that add greater flexibility and defensibility to the capabilities of the software. The modifications to the code include the following:

- A new user-interface option allows for much greater control in constructing a finite-difference grid for a groundwater problem, as well as greater specificity in inputting spatial information into the finite-difference grid. The new interface features are not unlike those offered in other high-end groundwater modeling interfaces such as Visual MODFLOW and GW-Vistas.
- Another very important addition to the code is the ability to condition/honor hydraulic head data. This option is similar to the one already employed in the code for conditioning water quality data, utilizing a statistical approach to matching simulated and observed data. The result is an even better potential for matching site conditions.
- The source term option has been given greater flexibility. Multiple source terms may now be simulated. Each source term can be input as a polygon, instead of just as a rectangle as in the previous version. In addition, the user may forgo the source term and vadose zone flow and transport and simply specify a flux to the water table. These options greatly enhance the usability of the code.
- The stream-aquifer interaction module has been enhanced to accommodate a wider range of possible configurations.
- Additional statistical reports have been added to the code for analysis of output data.

Representative Applications

As an example of the use of Groundwater*FX* in evaluating groundwater contamination problems, an analysis of the potential for natural attenuation is presented. A natural attenuation strategy requires that, within a reasonable time period, concentrations of the contaminants of concern be reduced below regulatory limits, or maximum contaminant levels (MCLs), by natural processes. Several potential natural attenuation processes can be considered:

- hydrodynamic dispersion of the contaminants (e.g., mass spreading and concentration reduction);
- degradation and/or decay (e.g., mass reduction);
- dilution from recharge or infiltration (e.g., areal recharge, stream/irrigation leakage); and/or
- flushing (e.g., discharge to a gaining stream).

Natural attenuation is applicable for organic contaminants (e.g., petroleum compounds) and inorganic constituents (e.g., metals). The main difference in processes between organic and inorganic constituents is the potential for degradation. For inorganics, the degradation of contaminants of concern probably has a minimal attenuation effect because biological processes are not very effective in reducing concentrations. Dilution, dispersion, and flushing are the main processes of interest for inorganics. For organic constituents, natural biodegradation processes may be present.

An example of this type of approach is found in the results of a natural attenuation analysis for a uranium mill tailings facility under the DOE's Uranium Mill Tailings Remedial Action (UMTRA) program. Figure 11 depicts the average contaminant plume distribution for uranium in 1997. The plume is discharging to the nearby stream, and dilution/ flushing is the dominant natural attenuation mechanism. The concentrations in the stream are well within acceptable limits for both human health

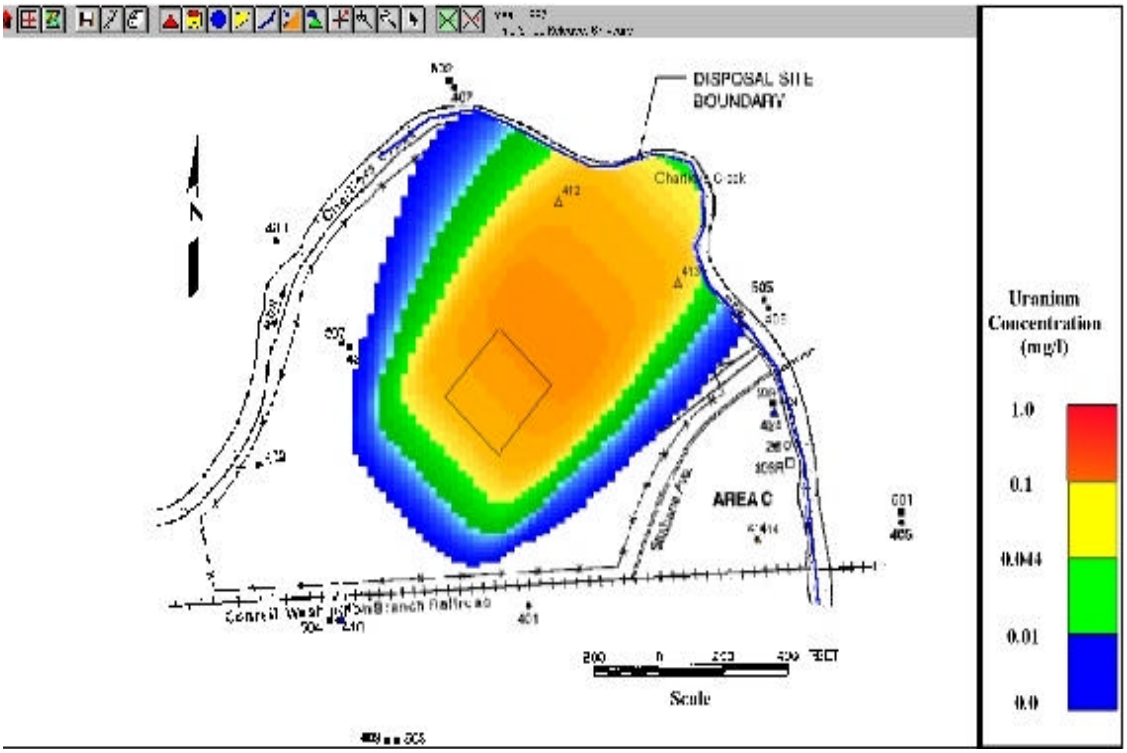

Figure 11. Average uranium concentrations in 1997.

and ecological concerns. The color contours on the plume are such that the green-to-yellow transition represents the concentration of the MCL. Therefore, the area of yellow-to-orange color is above acceptable limits.

Figure 12 shows the average contaminant plume concentrations 30 years after the previous plot. Over time the contaminants have attenuated to the point that, on average, the concentrations are less than the

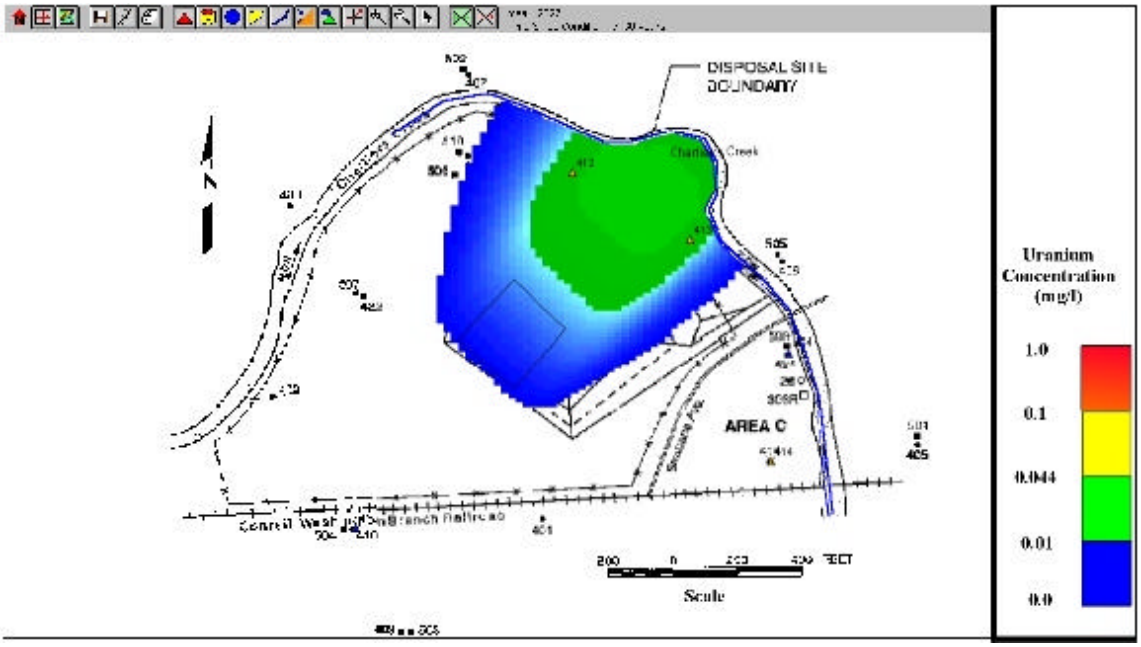

Figure 12. Average uranium concentrations in 2027.

MCL. However, the likelihood that the site is considered clean is not 100%.

Figure 13 shows the probability distribution for the same time frame as the previous plot—30 years after the baseline. The green regions of the plot indicate that there is a 5 to 10% probability that the concentrations may be above the MCLs at this time. In other words, on average we would expect the site to be cleaned up in 30 years, but there is still a 5 to 10% chance that it will not be within acceptable limits. Achieving essentially 100% likelihood of attaining compliance would take approximately 5 more years beyond this time. This uncertainty analysis allows the decision maker to plan for contingencies in monitoring duration and costs.

.

In addition to the visual depiction of the contaminant plumes just presented, the uncertainty analysis yields a statistical representation of likely concentrations in the monitoring wells through time (Figure 14). The power of this type of analysis is that the future monitoring of the site can be compared to the statistical distributions in this plot. As long as observed concentrations are less than the maximums shown in the upper error bars, the site is on track for natural attenuation. If, however, the concentrations monitored go above the uncertainty estimates, a reevaluation is in order.

If the uncertainties were addressed appropriately in the analysis, this situation should not occur, and the future monitoring should be within the predicted limits. From a regulatory perspective, this is advantageous. In a typical deterministic modeling scenario a calibrated model is used to predict concentrations at the compliance wells, yielding a single value for any given time frame of interest. If the monitored concentrations at a well are slightly above the predicted value at some time in the future, it is not clear whether the site is still on track for natural attenuation. With the uncertainty analysis, the analyst is provided likelihood estimates and a "comfort range" (the statistical spread on the predicted concentrations) for evaluating the performance of the remedy.

In addition to analyzing the potential for natural attenuation at this site, Groundwater*FX* was used to evaluate a potential pump-and-treat remedy. This type of active remedy would take an estimated 20 years to complete, at a cost of about $4.5M. From a cost-benefit standpoint, the monitored natural attenuation option is more favorable.

Groundwater*FX* analysis of the uranium mill tailings site in Riverton, Wyoming, resulted in the first natural attenuation remedy approved for a DOE UMTRA site, with concurrence by the Nuclear Regulatory Commission (NRC) following EPA guidelines and rules for compliance. Groundwater*FX* has also been used to demonstrate compliance for an alternate concentration limit (ACL) remedy at the Canonsburg, Pennsylvania, UMTRA site. NRC approval is pending

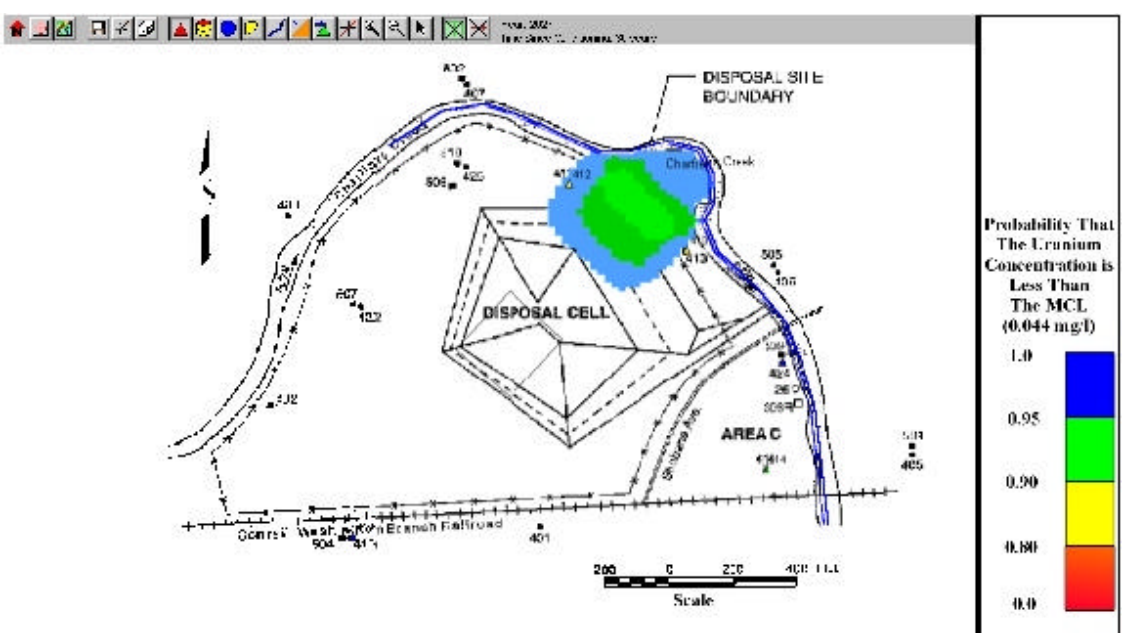

Figure 13. Map showing probability that uranium exceeds MCLs in 2027.

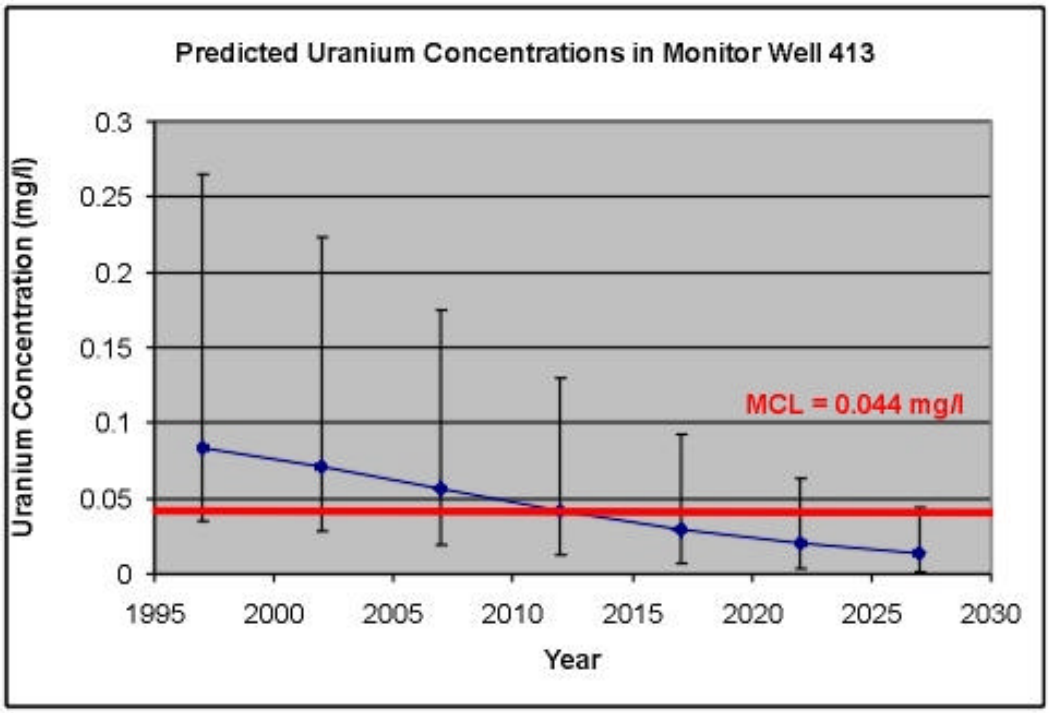

Figure 14. Predicted uranium concentrations over time at well 413 with uncertainty error bars.

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.

EPA (U.S. Environmental Protection Agency). 1994. *Guidance for the Data Quality Objective Process, QA/G-4.* EPA/600/R-96/055. U.S. Environmental Protection Agency, Washington, D.C.

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).

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.

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

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.