United States Environmental Protection Agency
Office of Research and Development
Washington, D.C. 20460
EPA/600/R-00/038
March 2000

Environmental Technology Verification Report

Environmental Decision Support Software

Decision*FX,* Inc.

Sampling*FX*

THE ENVIRONMENTAL TECHNOLOGY VERIFICATION PROGRAM

ETV

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:	**Sampling*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 Sampling*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*, Inc., chose to use Sampling*FX* to perform the visualization, sample optimization, and cost-benefit endpoints for four problems from three sites (A, N, and T). Sampling*FX* was used to provide objective guidance on the selection of optimum locations for new samples, to quantify the nature and extent of contamination as a function of probability, to estimate exposure concentrations for human health risk analysis, and to estimate cleanup volumes as a function of cleanup level. The Site A sample optimization test problem was a three-dimensional (3-D) groundwater problem for two volatile organic compounds (VOCs), dichloroethene (DCE), and trichlorethene (TCE). The Site N sample optimization problem was a two-dimensional (2-D) soil contamination problem for three heavy metals (arsenic, cadmium, and chromium). In this problem, data were supplied over a limited area of the site, and the analyst was asked to develop a sampling strategy that characterized the remainder of the 125-acre site while taking only 80 additional samples. The Site N cost-benefit problem considered the same contaminants as the sample optimization problem and had 524 data points on a 14-acre site. The objective of this problem was to define the areas in which the contamination exceeded threshold concentrations. In addition, the analysts were asked to estimate the human health risks based on current conditions. The Site T test problem was a 2-D soil contamination problem. This problem included four contaminants: ethylene dibromide (EDB), dichloropropane (DCP), dibromochloropropane (DBCP), and carbon tetrachloride (CTC).

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., SamplingFX*, EPA/600-00/038.

TECHNOLOGY DESCRIPTION
Sampling*FX* is a geostatistics-based software program intended to provide decision makers and analysts a means of evaluating environmental information relative to the nature and extent of contamination in surface and subsurface soils. Key attributes of the product include the ability to delineate, provide visual feedback on, and quantify uncertainties in the nature and extent of soil contamination (e.g., concentration distribution, probability of exceeding a soil cleanup guideline); to provide objective recommendations on the number and location of sample locations; and to provide statistical information about the contamination (e.g., average volume of contamination, standard deviation, etc.). Sampling*FX* runs on Windows 95, 98 or NT platforms and on the Power Macintosh system.

VERIFICATION OF PERFORMANCE
The following performance characteristics of Sampling*FX* were observed:

Decision Support: In the demonstration, Decision*FX* used Sampling*FX* to import data on contaminant concentrations and surface structures from ASCII text files and bitmap graphical image files. Sampling*FX* demonstrated the ability to integrate this information on a single platform and place the information in a visual context. It generated 2-D maps of concentration contours, maps showing the probability of exceeding threshold concentrations, and variance maps that support data interpretation. The software was used in the demonstration to generate the data necessary for producing cost-benefit curves and estimating human health risk. The cost-benefit curves and risks were produced in auxiliary software, Microsoft Excel. The accuracy of the analyses is discussed in Section 4 of the report.

Documentation of the SamplingFX Analysis: The Decision*FX* analyst generated a report that provided an adequate explanation of the process and the parameters used to analyze each problem. Documentation of data transfer, manipulations of the data, and analyses were included. Model selection and parameters for conducting the probabilistic assessment were provided in standard ASCII text files that are exportable to a number of software programs.

Comparison with Baseline Analysis and Data: The concentration contours produced by Sampling*FX* during the demonstration were compared to the data and to baseline analyses performed using data interpolation and geostatistics. The visualizations produced by Sampling*FX* were often limited by a lack of a frame of reference or site map, and this lack made comparison more difficult. In the 3-D groundwater contamination sample optimization problem for Site A, the Sampling*FX* concentration contours and probability maps did not match the data or the baseline analysis. For the Site N sample optimization problem, the Sampling*FX* analysis generated an acceptable match to the data and the baseline analysis. When compared to the baseline geostatistical analysis with the entire data set, Sampling*FX* identified approximately 70% of the site that had arsenic contamination above 125 mg/kg with the constraint of an additional 80 samples to characterize the entire 125-acre site. For the Site N cost-benefit problem, contaminant contour and probability maps were consistent with the baseline interpolation and geostatistical analysis. Estimates of the area where the contamination exceeded threshold concentrations matched, to within 20%, the baseline interpolation and geostatistical analyses at the 50% probability levels. The area estimates at the 10% probability level (at least a 10% chance that contamination exceeds the threshold concentration) were considerably different from the baseline geostatistical analysis. This was due to different definitions of the probability level. Sampling*FX* performs multiple simulations of the concentration distribution at the site. It then calculates the area above the threshold for each simulation and uses these areas to estimate the probability of an area's exceeding the threshold. Consequently, the area probability estimates are for the site as a whole. By contrast, the baseline geostatistical analysis used an approach that is consistent with EPA data quality objective guidance and defined the area estimates based on the probability that a given location could exceed the threshold. There is much more uncertainty on a local scale, and therefore, the area estimates for the baseline geostatistics analysis show a wider variation than the Sampling*FX* analysis at the different probability levels. Sampling*FX* was used to estimate exposure concentrations for risk assessment calculations for two scenarios—exposures for on-site workers and residential exposures. The Sampling*FX* values for the worker scenario were consistent with the baseline analysis for two of the three contaminants but incorrect and too low for one contaminant, arsenic. The exposure concentrations generated by Sampling*FX* for the residential scenario were inconsistent with the data and considerably lower than the baseline estimates for all three contaminants (As, Cd, and Cr). For the Site T soil contamination problem, contaminant contour and probability maps were consistent with the data and the baseline analysis for each of the four contaminants (EDB, DCP, DBCP, and CTC). Estimates of the area where the contamination exceeded threshold concentrations did not match the baseline interpolation analysis and appeared to be inconsistent with the concentration and probability of exceedence maps generated by Sampling*FX*.

Multiple Lines of Reasoning: Decision*FX* used Sampling*FX* to provide a number of different analysis approaches to examine the data. The foundation of its approach is a Monte Carlo simulator that produces multiple simulations of the existing data that are consistent with the known data. From these simulations, concentration maps, variance maps, and probability maps were produced to assist in data evaluation. This permits the decision maker to evaluate future actions such as sample location or cleanup guidance based on the level of confidence placed in the analysis.

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

Ease of Use: During the demonstration it was observed that in general Sampling*FX* was not user-friendly. Sampling*FX* has (or lacks) several features that 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, and the need to have all graphic files imported as a single bitmap, as well as the absence of on-line help. Visualization output is limited to screen captures that can be imported into other software for processing. Visualization output was often supplied without a frame of reference (coordinate scale or site map), and this makes data interpretation more difficult. While each of these limitations can be overcome and the analysis performed, it requires more work on the part of the software operator.

Sampling*FX* exhibited the capability to export ASCII text and graphics to standard word processing software directly. Screen captures from Sampling*FX* were imported into CorelDraw to generate .jpg and .cdr graphic files that can be read by a large number of software products. Sampling*FX* generated data files from statistical analysis and concentration estimates in ASCII format, which can be read by most softwares.

Efficiency and Range of Applicability: Sampling*FX* was used to complete four sample optimization/cost-benefit problems with 12 person days of effort. This was slightly longer than the technical team would have anticipated and was due primarily to the extensive post-processing of maps and data required for the analysis. However, Sampling*FX* provides the flexibility to address problems tailored to site-specific conditions. The user has control over the choice of the parameters that control the geostatistical simulations, and the software allows a wide range of environmental conditions [e.g., contaminants in different media (groundwater or soil)] to be evaluated. Its applicability to 3-D groundwater contamination problems is not clear. Theoretically, one should be able to use the software for this type of problem. However, the results provided for the Site A 3-D test problem were not consistent with the data.

Operator Skill Base: To efficiently use Sampling*FX*, the operator should be knowledgeable in the use of statistics and geostatistics in analyzing data for environmental contamination problems. In addition, knowledge about managing database files, contouring environmental data sets, and conducting sample optimization and cost-benefit problems is beneficial for proper use of the software.

Training and Technical Support: An analyst with the prerequisite skill base can use Sampling*FX* after one or two 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* intends to sell Sampling*FX* for $500 for a single license. It will be supplied at no cost to state and federal regulators.

Overall Evaluation: The technical team's evaluation of Sampling*FX* was based on observation and training supplied during the demonstration, the documentation of the analyses performed during the demonstration, the Sampling*FX* users' guide, the visualization maps provided for the analyses, and the evaluation team's experience with software products that perform similar functions. The technical team concluded that the main strength of Sampling*FX* is its technical approach to solving the sample optimization problem. The use of the multiple simulations of the data to generate probability and concentration maps provides a technically robust framework for conducting sample optimization problems. The technical team concluded that there were several limitations in the application of Sampling*FX* to environmental contamination problems. Sampling*FX* was unable to produce an adequate match to the data for the Site A 3-D sample optimization problem; was unable to match exposure concentrations for risk calculations; and produced area estimates that were not consistent with its own probability and concentration maps (Sites N and T). In addition, the Decision*FX* analyst used a nonstandard approach for estimating the probabilities of a given area of contamination. The approach underestimates contamination areas at low probabilities. 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.

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. Sampling*FX* can be an appropriate choice for some environmental contamination problems, and the results of the analysis can support decision making. As with any software product, improper use of the software can cause the results of the analysis to be misleading or inconsistent with the data. In general, the quality of the output is directly dependent on the skill of the operator.

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

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

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

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

EPA600/R-00/038
March 2000

Environmental Technology Verification Report

Environmental Decision Support Software

Decision*FX*, Inc.

Sampling*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

List of Tables

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., Sampling*FX* product, contact

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

Abbreviations and Acronyms

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
MB	megabyte
MHz	megahertz
NAMP	National Analytical Management Program (DOE)
NERL	National Exposure Research Laboratory (EPA)
NMERI	New Mexico Engineering Research Institute
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
SADA	Spatial Analysis and Decision Assistance (software)
SCMT	Site Characterization and Monitoring Technology
TCA	trichloroethane
TCE	trichloroethene
Tc-99	technetium-99
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 based on testing and QA protocols developed with input from all major stakeholder/customer groups associated with the technology area. The demonstration described in this report was administered by the Site Characterization and Monitoring Technology (SCMT) Pilot. (To learn more about ETV, visit ETV's Web site at http://www.epa.gov/etv.)

The SCMT pilot is administered by EPA's National Exposure Research Laboratory (NERL). With the support of the U.S. Department of Energy's (DOE's) Environmental Management (EM) program, NERL selected a team from Brookhaven National Laboratory (BNL) and Oak Ridge National Laboratory (ORNL) to perform the verification of environmental decision support software. Decision support software (DSS) is designed to integrate measured or modeled data (such as soil or groundwater contamination 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*, Inc.), Sampling*FX* (Decision*FX*, Inc.), 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 Sampling*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 based on the evaluation criteria presented in Section 3. Because of the inherent variability in soil and subsurface contamination, most of the evaluation criteria are qualitative. Even when a direct comparison is made between the developer's analysis and the baseline analysis, different numerical algorithms and assumptions used to interpolate data between measured values at known locations make it almost impossible to make a quantitative judgement as to which technical approach is superior. The comparisons, however, do permit an evaluation of whether the analysis is consistent with the data supplied for the analysis and therefore useful in supporting remediation decisions.

Summary of Analysis Performed by Sampling*FX*

Sampling*FX* is a geostatistics-based software product designed to provide decision makers and analysts a means of evaluating environmental information relative to the nature and extent of contamination in surface and subsurface soils. Sampling*FX* quantifies uncertainties and provides additional sample location recommendations, statistical information about the contamination, and visual feedback on the extent of contamination.

In the demonstration, Decision*FX* used Sampling*FX* to import data on contaminant concentrations and surface structures from ASCII text files and bitmap graphical image files. Sampling*FX* demonstrated the ability to integrate this information on a single platform and place the information in a visual context. Sampling*FX* generated two-dimensional (2-D) maps of concentration contours, maps showing the probability of exceeding threshold concentrations, and variance maps that support data interpretation. Sampling*FX* was used in the demonstration to generate the data necessary for producing cost-benefit curves and estimates of concentrations at receptor locations for use in human health risk analysis. The cost-benefit curves and risk analysis were produced in an auxiliary software package (Microsoft Excel).

Decision*FX* staff chose to use Sampling*FX* to perform all three endpoints using data from the Site A sample optimization problem, the Site N sample optimization problem, the Site N cost-benefit problem, and the Site T sample optimization problem. During the demonstration, visualization results were presented for all four problems. For the Site A sample optimization problem, Sampling*FX* was used to define sample locations to characterize the three-dimensional (3-D) volume of groundwater contaminated above specified contamination threshold concentrations. For the Site N and Site T soil sample optimization problems, Sampling*FX* was used to specify surface soil sample locations for site characterization. For the Site N cost-benefit problem, Sampling*FX* was used to define the area of the site that had contamination above specified threshold concentrations as a function of probability and estimate the cost of remediation. In addition, exposure concentrations were estimated for use in calculating residential and on-site worker risk.

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

Section 2 — Sampling*FX* Capabilities

The following section provides a general overview of the capabilities of Sampling*FX*, a Decision*FX*, Inc., software product. Decision*FX* supplied this information.

Sampling*FX* is a DSS intended to provide decision makers and analysts a means of evaluating environmental information relative to the nature and extent of contamination in surface and subsurface soils. Key attributes of the tool include the 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).

Sampling*FX* relies mainly on geostatistical algorithms [from the Geostatistical Software Library (GSLIB)] to analyze spatial aspects of soil contamination data and operations research methods to provide guidance on key decision analysis needs (e.g., recommended location of samples).

Sampling*FX* is an improvement over conventional sampling and analysis approaches because it provides information on spatial variability (something that traditional statistical approaches ignore) and objective guidance on sampling placement (rather than using expert judgment).

Currently, Sampling*FX* has versions that operate 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. The development software was chosen for ease of use in transferring between different platforms. The recommended computer configuration for running the Sampling*FX* software on PC platforms is approximately 15 MB of hard-disk space for the program, about 10 MB of storage space for model runs, about 32 MB of RAM, and a Pentium processor with reasonable speed (>100 MHz).

The Sampling*FX* code is intended for use in providing decision analysis information on single analytes associated with contamination in surface or subsurface soils. The methodology is based on geostatistics and therefore is applicable to other parameters exhibiting spatial correlation (e.g., hydraulic conductivity distributions). For multiple analytes of concern, multiple model runs must be performed.

Section 3 — Demonstration Process and Design

Introduction

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

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

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

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

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

Development of Test Problems

Test Problem Definition

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

Summary of Test Problems

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

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

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

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

Table 1. Summary of test problems

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

Table 2. Data supplied for the test problems

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

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

The independent data analysis was performed using Surfer™ (Golden Software 1996). Surfer was selected for the task because it is a widely used, commercially available software package with the functionality necessary to examine the data. This functionality includes the ability to import drawing files to use as layers in the map, and the ability to interpolate data in two dimensions. Surfer has eight different interpolation methods, each of which can be customized by changing model parameters, to generate contours. These different contouring options were used to generate multiple views of the interpolated regions of contamination and hydrologic information. The best fit to the data was used as the baseline analysis. For 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 Sampling*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 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 an anonymous server or copied to a zip drive or compact disk–read only memory (CD-ROM). The technical team verified that all files generated by the developers during the demonstration were provided and intact. The developers were given a 10-day period after the demonstration to provide a written narrative of the work that was performed and a discussion of their results.

Evaluation Criteria

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

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

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

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

Criteria for Assessing Decision Support

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

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

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

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

Documentation of the Analysis and Evaluation of the Technical Approach

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

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

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

Comparison of Projected Results with the Data and Baseline Analysis

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

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

Use of Multiple Lines of Reasoning

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

Secondary Evaluation Criteria

Documentation of Software

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

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

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

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

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

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

Section 4 — Sampling*FX* Evaluation

Sampling*FX* Technical Approach

The technical approach applied in Sampling*FX* is based on geostatistics. Geostatistical methods are based on the premise that measured variables located close to each other will have similar values, while variables far apart will have little correlation between their corresponding values. A statistical measure for this interrelationship is summarized by the correlation between variables measured at different points in space. This measure or related measures, such as the variogram and covariance, form the central idea around which linear estimation methods in geostatistics operate. The use of correlation measures also separates this estimation method from other deterministic interpolation algorithms such as inverse distance, linear interpolation, splines, and quadrature methods. Using a statistical estimator allows the estimation error to be calculated along with the estimate. Thus, a geostatistical method provides both the most likely value and an estimate of the range of other possible values for a given location. This is important information because the spatial variability present in most parameters is such that error-free estimation is not possible. In fact, there are often many possible solutions to the estimation problem that agree with the measurements (Appendix B). Kriging is one of the more common geostatistical methods used to provide smoothed estimates of variables.

The geostatistical framework also allows an alternative to single estimates of the distribution of values, such as kriging. This alternative falls under various names depending on the method used to implement it: conditional simulation, sequential simulation, constrained simulation, stochastic interpolation, or fractal interpolation. In these approaches, multiple realizations of the data are performed. The intent of each simulation is to estimate values between the measured values while maintaining the same statistical characteristics as the measured data. The result is to produce estimated values that have the same statistical structure as the actual measured values and agree with the measured points.

The approach used by Sampling*FX* to address the variability is based on the sequential Gaussian simulator from GSLIB (Deutsch and Journel 1992).

The sequential simulation procedure creates multiple realizations from the data that honor the data points and reproduce two important statistical measures, the mean and the covariance function. The results are generated random fields that agree with the observed data and have the same amount of variability as the field from which the data were drawn. The inherent concept is that each generated field provides an equally likely realization of the value being simulated and that each field is consistent with the data. Each of the generated fields is used to estimate the variability in the value at the different estimation points. The geostatistical concepts associated with the Sampling*FX* approach take into account the spatial distribution of contaminant concentration data, including the autocorrelation that exists between samples that are taken close to each other. Thus, this tool is clearly distinguished from more conventional, statistically based sampling schemes that do not account for spatial correlation.

Sampling*FX* Implementation of Geostatistical Approach

Sampling*FX* imports measured data, defines a grid (i.e., divides the area of concern into a number of rectangular blocks), and predicts contaminant concentrations at unsampled locations using the sequential simulation procedure. This procedure generates multiple realizations of predicted concentrations, with each simulation honoring the existing data but reflecting the variability inherent in the data field. The suite of simulated concentrations is then used to estimate probability density functions (pdfs) and subsequently provides a probabilistic description of soil contaminant distribution. This description reflects the probability that the threshold concentration will be exceeded at a point within the site domain.

The Sampling*FX* user has the ability to step through each of the stochastic simulation results to observe the variability in predicted concentration distributions. Alternatively, the mean or average concentration distribution from all the stochastic runs may be displayed. The user can specify a desired concentration threshold of concern (e.g., a soil cleanup level) and view a 2-D map of the probability of exceeding that concentration. The user can also query the model to produce a statistical report on the area of contamination that exceeds a user-specified concentration threshold. This data may be important for evaluating the uncertainties associated with cleanup costs, especially if different land use scenarios are being considered.

The geostatistical routines in the code also allow the user to view the spatial distribution of either the standard deviation or variance of the predicted concentration distribution. The variance is a measure of variability or uncertainty in the predicted concentration distribution. Where the variance is high, there is less confidence in the predicted concentration information. When the user invokes the sampling optimization algorithm, the variance information is combined with a user-specified concentration threshold to predict the best location(s) for additional sampling. This sampling optimization algorithm is aimed at providing uncertainty reduction where it is most needed.

Sampling*FX* provides model output in a variety of forms. The code has the capability to determine the location of several probability contours for any user-specified threshold concentration. Because the results of each set of simulations are saved in an internal database, it is not necessary to rerun the simulations for each threshold concentration and each contour. By envisioning contours that connect points of equal probability, the Sampling*FX* user can examine measures of uncertainty in estimated concentrations at selected intervals along a specified contour. These uncertainties provide the quantitative means to locate either a single sampling point or multiple sampling locations objectively.

Description of Test Problems

Sampling*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 surface and subsurface soils. The software is designed to address and visualize the nature and extent of soil contamination, provide recommendations for sample locations, and provide statistical information on the nature and extent of contamination. Sampling*FX* was used on four problems: Site A sample optimization, Site N sample optimization, Site N cost-benefit, and Site T sample optimization. All three endpoints of the demonstration were addressed. As part of the demonstration, over 100 visualization outputs were generated. A few examples that display the range of Sampling*FX*'s capabilities and features are included in this report. A general description of each test problem and the analysis performed using Sampling*FX* follows. Detailed descriptions of all test problems are provided in Appendix A and in Sullivan, Armstrong, and Osleeb (1998).

Site A Sample Optimization Problem

The Site A problem was a 3-D groundwater contamination problem. The data supplied for the analysis of Site A included surface maps of roads, buildings, and water bodies; concentration data on two contaminants—trichloroethene (TCE) and dichloroethene (DCE)—in groundwater wells at different depths and locations; hydraulic head data; and geologic structure data. This test problem was designed as a method for assessing the accuracy with which the software can be used to predict sample locations to define the source of a groundwater plume and define the extent of contamination.

Sampling*FX* performs a 2-D analysis. To address the 3-D characteristics of this problem, Decision*FX* divided the contamination data into six vertical strata 10 ft thick to a depth of 50 ft below the water table. The decision to stop at a depth of 50 ft was based on two considerations: the fact that the confining layer began to appear at depths greater than 50 ft in some regions and a desire to minimize the time required for the analysis during the demonstration. Although a complete analysis could have been performed, the emphasis was on the process for completing an evaluation, and Decision*FX* determined that analyzing six layers was sufficient to demonstrate Sampling*FX*'s capabilities.

For sample optimization, Sampling*FX* works with one contaminant per simulation. To save computational time, TCE was arbitrarily chosen by the Decision*FX* analyst as the contaminant for performing the sample optimization analysis. Originally, only the data for 7 sample locations (e.g., groundwater wells) were supplied in the problem domain. Decision*FX* staff requested 26 additional sample locations to provide enough data to apply geostatistics. This information was used to generate the next set of 10 sample locations using expert judgment of the Decision*FX* analyst to fill in data gaps. Thus, Decision*FX* used a total of 43 groundwater well locations in the analysis.

Based on the final data set using the 43 groundwater wells, Sampling*FX* was used to generate the concentration distribution, the variance distribution, and the probability distribution of exceeding two threshold concentrations (10 and 100 μg/L) for both TCE and DCE. The variance is the square of the difference between the value at a location and the mean value. Variance distribution maps emphasize regions that are far from the mean, and the variance can be related to uncertainty. These analyses were performed for each of the six vertical strata defined by Decision*FX*. In all, 48 maps were generated during the analysis.

The statistical data on the nature and extent of contamination were exported to Excel and also used to generate a cost-benefit analysis of the area contaminated vs cleanup threshold. This was not requested as part of the problem definition but was performed by Decision*FX* to highlight the software's capability in this area.

Site N Sample Optimization Problem

The Site N sample optimization problem was a surface soil contamination problem for three contaminants (As, Cd, and Cr). The analysts were given an extensive data set for a small, highly contaminated region of the site (<10 acres) and asked to develop a sample optimization scheme to define the extent of contamination for the entire site (125 acres). Table 3 presents contaminant threshold concentrations for each contaminant. The test problem was designed to assess the accuracy with which the software can be used to predict sample locations to define the extent of surface soil contamination. Budgetary constraints limited the number of additional sample locations to 80. Because of the limited number of samples, the analyst was asked to supply estimates of the extent of contamination based on the confidence in the results.

Sampling*FX* was used to perform an iterative analysis where several suggested sample locations were requested. Although three contaminants are present, arsenic, which had the highest measured concentrations, was chosen by Decision*FX* as the reference contaminant to be used for sample optimization decisions. This is a legitimate approach because in practice a single sample location would be selected and measurements performed for all three contaminants as opposed to selecting three different sample locations and measuring each for a single contaminant. Initially, data were supplied only for a small area of the site. Therefore, a random sample generator was used to distribute 20 samples throughout the site. With this data set, variograms were constructed and a Monte Carlo geostatistical simulator was used to generate 50–100 realizations of the data. The geostatistical simulator matches the data at known sampling locations and estimates contaminant concentrations at other unsampled locations. This information is used to generate variance maps and maps of the probability of exceeding the threshold concentrations. An operations research algorithm selects the next sampling locations based on the variance and probability information. Additional sampling locations were selected based on expert judgment to fill in data gaps and reduce variance. In the second round, 14 additional sample locations were selected. This process was continued until data at the 80 additional sample locations were provided. Sampling*FX* used all of the data to produce the following information:

- concentration contour maps for the three contaminants (As, Cd, and Cr),
- variance contour maps for the three contaminants, and
- probability maps of exceeding threshold concentrations for each contaminant at the two threshold concentrations (Table 3).

Table 3. Site N soil contamination threshold concentrations for the sample optimization problem

Contaminant	Minimum threshold concentration (mg/kg)	Maximum threshold concentration (mg/kg)
Arsenic (As)	125	500
Cadmium (Cd)	70	700
Chromium (Cr)	370	3700

Site N Cost-Benefit Problem

The Site N cost-benefit problem was a surface soil contamination cost-benefit problem for the same three contaminants used in the Site N sample optimization problem—As, Cd, and Cr. The developers were given an extensive data set for a 14-acre region of the site and asked to conduct a cost-benefit analysis to evaluate the area and cost for remediation to achieve specified threshold concentrations (shown in Table 4).

Sampling*FX* was used to estimate the areal extent of the soil contamination by taking the supplied concentration data and using its geostatistical simulator to estimate concentrations, the variance in the predicted concentrations, and the probability of exceeding threshold concentrations. The software generated the following output for each contaminant for this problem:

- a site map with roads and water bodies overlain with concentration contours at the specified threshold concentrations,
- variance contour maps for the contaminants, and
- probability maps of exceeding threshold concentrations for each contaminant at the two threshold concentrations (Table 4).

In addition, Sampling*FX* was used to calculate the exposure concentrations for use in calculating human health risk. The Decision*FX* analyst was able to take these concentrations and import them into Excel and perform a risk calculation. However, the risk calculation was performed independent of Sampling*FX* software and depended entirely on the skill of the analyst and not the software. Therefore, it is not evaluated in this report. An evaluation was performed of the exposure concentrations supplied for the risk calculation.

Site T Sample Optimization Problem

The Site T problem was a 2-D soil contamination sample optimization problem. The data supplied for analysis of this problem included surface drawings of buildings and roads and soil contamination data for four organic contaminants [ethylene dibromide (EDB), dibromochloroproprane (DBCP), dichloropropane (DCP), and carbon tetrachloride (CTC)]. This test problem was designed as a method for assessing the accuracy with which the software can be used to predict sample locations to define the extent of surface and subsurface soil contamination. The design objective was to generate a 3-D rendering of the soil contamination in two stages. In the first stage, the analysts were asked to develop a sampling strategy to define surface areas on the site in which the soil contamination exceeded the threshold concentrations given in Table 5 with probability levels of 10, 50, and 90% on a 50 H 50 ft grid. In the second stage, after defining the region of surface contamination, the analysts were asked to define subsurface contamination in the regions found to be above the threshold at the 90% probability level. The problem definition required subsurface

Table 4. Site N soil contamination threshold concentrations for the cost-benefit problem

Contaminant	Minimum threshold concentration (mg/kg)	Maximum threshold concentration (mg/kg)
Arsenic (As)	75	500
Cadmium (Cd)	70	700
Chromium (Cr)	370	3700

Table 5. Site T soil contamination threshold concentrations

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

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).

Sampling*FX* was used to perform an iterative analysis in which several suggested sample locations were requested. Sampling*FX* used the original data set, which contained 32 sample locations with data for each of the four contaminants. Decision*FX* arbitrarily selected DBCP as the contaminant for developing the sampling network. With the DBCP data set, variograms were constructed and a Monte Carlo geostatistical simulator was used to generate 50 realizations of the data. The geostatistical simulator matches the data at known sampling locations and estimates contaminant concentrations at other, unsampled locations. This information was used to generate variance maps and maps of the probability of exceeding the threshold concentrations. An operations research algorithm was used to select the next sampling locations based on variance and the probability information. Additional sampling locations were selected based on expert judgment to fill in data gaps and reduce variance. In the second round, 16 additional sample locations were selected. This process was repeated for a third round (30 additional samples) and a fourth round (16 samples) until a total of 64 additional sample locations were provided. Sampling*FX* used data from the all of the 64 sample locations to produce the following:

- concentration contour maps for the four contaminants (CTC, DBCP, DCP, and EDB),
- variance contour maps for the four contaminants,
- probability maps of exceeding threshold concentrations for each contaminant at the threshold concentrations (Table 5), and
- a graph of the estimated area of contamination as a function of the number of samples collected.

Evaluation of Sampling*FX* *Decision Support*

In the demonstration, Decision*FX* used Sampling*FX* to import data on contaminant concentrations and surface structures from ASCII text files and bitmap graphical image files. The software demonstrated the ability to integrate this information on a single platform and place the information in a visual context. It generated 2-D maps of concentration contours, maps showing the probability of exceeding threshold concentrations, and variance maps that support data interpretation. Sampling*FX* was used in the demonstration to generate the data necessary for producing cost-benefit curves and was used to estimate human health risk. The cost-benefit curves and risk estimates were produced in auxiliary software (Microsoft Excel). The accuracy of the analyses is discussed in the section on comparison of Sampling*FX* results with baseline data and analysis.

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

For each problem, Decision*FX* provided a detailed description of the steps necessary to import the provided data into Sampling*FX* and perform the desired analysis. The steps proceeded logically, and manipulations to format the data into the Sampling*FX* format were relatively simple. Files containing data were supplied to the analyst using a .dbf format. Before these files were used in Sampling*FX*, they were imported into another program (e.g., Microsoft Excel) and saved in ASCII text file format. Decision*FX* also provided rationales for the choice of the different model approaches (geostatistical-based, variance-based, or based on expert judgment) used in performing the sample optimization problem. Model selection and parameters for contouring were provided in the output files and problem documentation.

In general, the probabilistic simulation approach used by Sampling*FX* provides a robust mathematical foundation for performing the analysis. However, in performing estimates of the regions in which a contaminant exceeds a threshold concentration as a function of probability, Sampling*FX* used an approach that was slightly different than the approach used in the baseline analysis. Specifically, Sampling*FX* divides the domain mathematically 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 software computes the volume (or area in two dimensions) that exceeds the threshold concentration. From the distribution of the multiple simulations, Sampling*FX* calculates the probability that a volume of soil will exceed the threshold concentration.

The baseline geostatistical analysis was performed with a slightly different approach, one that used the EPA data quality objective guidance (EPA 1994), which tends to maximize the volume estimates at low probabilities. In the baseline analysis, the site was mathematically divided into a number of rectangular regions (grids). Within each region, an analysis was made to determine a single estimate of the concentration. Using the statistical properties of the data, the probability that the contamination does not exceed the threshold concentration in each region is calculated. This approach places the probability question in each mathematical grid of the analysis. There is more uncertainty as to the concentration within each region as compared to the total over the entire site because of averaging over a larger area for the entire site. Therefore, the baseline approach will predict larger volumes of contamination for the high-probability estimate (e.g., <10% chance of exceeding a threshold concentration) as compared to the Sampling*FX* approach and lower volumes for the low-probability estimate (e.g., >10% chance of exceeding a threshold concentration). The Sampling*FX* and baseline estimates of contaminated volume at the mean value should be similar.

This does not imply that the Sampling*FX* approach is technically incorrect. The approach simply supplies different information. In fact, as described above, 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 Sampling*FX* approach provides multiple (50–100) simulations of the data. Sampling*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 area of contamination as a function of probability of exceeding a threshold concentration. If done correctly, this approach may provide a more defensible estimate than the baseline approach.

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

Site A Sample Optimization Problem

In the Site A sample optimization problem, data on contaminants in seven groundwater wells were supplied on a 5-ft vertical spacing. These groundwater data were taken near a suspected contaminant source. Site maps containing buildings and a river were provided in bitmap form to assist in the data evaluation. Decision*FX* used Sampling*FX* to develop a sample optimization scheme to define the 3-D extent of TCE and DCE contamination in the source region. Sampling*FX* operates in 2-D space; therefore, the analyst divided the vertical domain into 10-ft-thick slices and analyzed the data at 10-ft intervals down to 50 ft below the water table. This yielded six 2-D slices for the analysis. Decision*FX* requested an additional 36 sample locations (groundwater wells) in two rounds of sampling to complete its analysis using Sampling*FX*.

The results generated by Sampling*FX* were compared to a baseline analysis concentration map. The baseline analysis used the entire data set to define the zones of contamination above the threshold concentration. The baseline analysis also divided the subsurface into 10-ft vertical sections. However, the baseline analysis used the maximum concentration within the section, in contrast to the Decision*FX* analysis, which used the concentration at a 10-ft spacing. Therefore, the baseline analysis will tend to predict larger areas of contamination. Decision*FX* generated maps of concentration, variance, and the probability of exceeding the threshold at each threshold for each contaminant for each depth (six depths). A total of 48 maps were prepared as part of the analysis, and each was visually compared to the baseline analyses. The demonstration technical team, in a few cases, took the same data set supplied to Decision*FX* after sample optimization was completed and generated concentration contour maps. This permitted a better understanding of the differences between the baseline and the Sampling*FX* approaches. To illustrate the Sampling*FX* approach, this report presents the results for DCE contamination at 30 ft below the water table. Similar types of output were generated for the other elevations and contaminants.

Figure 1 was generated using Sampling*FX* to show the average DCE concentration at 30 ft below the water table. The “zone of interest” identified on Figure 1 delineates the area for which data were provided for analysis. This region contains less than half the area shown in the figure. (The analyst stated that region outside of the zone of interest does not contain information meaningful to the demonstration and should be disregarded) The reason the Decision*FX* analyst chose to conduct the analysis on this much broader region is not clear. Within the border, the outline of the zone of contamination can be determined with careful scrutiny. Color-coded circles (which are difficult to see in this figure) represent data-collection locations. As can be determined from the color key, the blue section of the map represents the areas where the predicted concentration is less than 10 : g/L, the green section

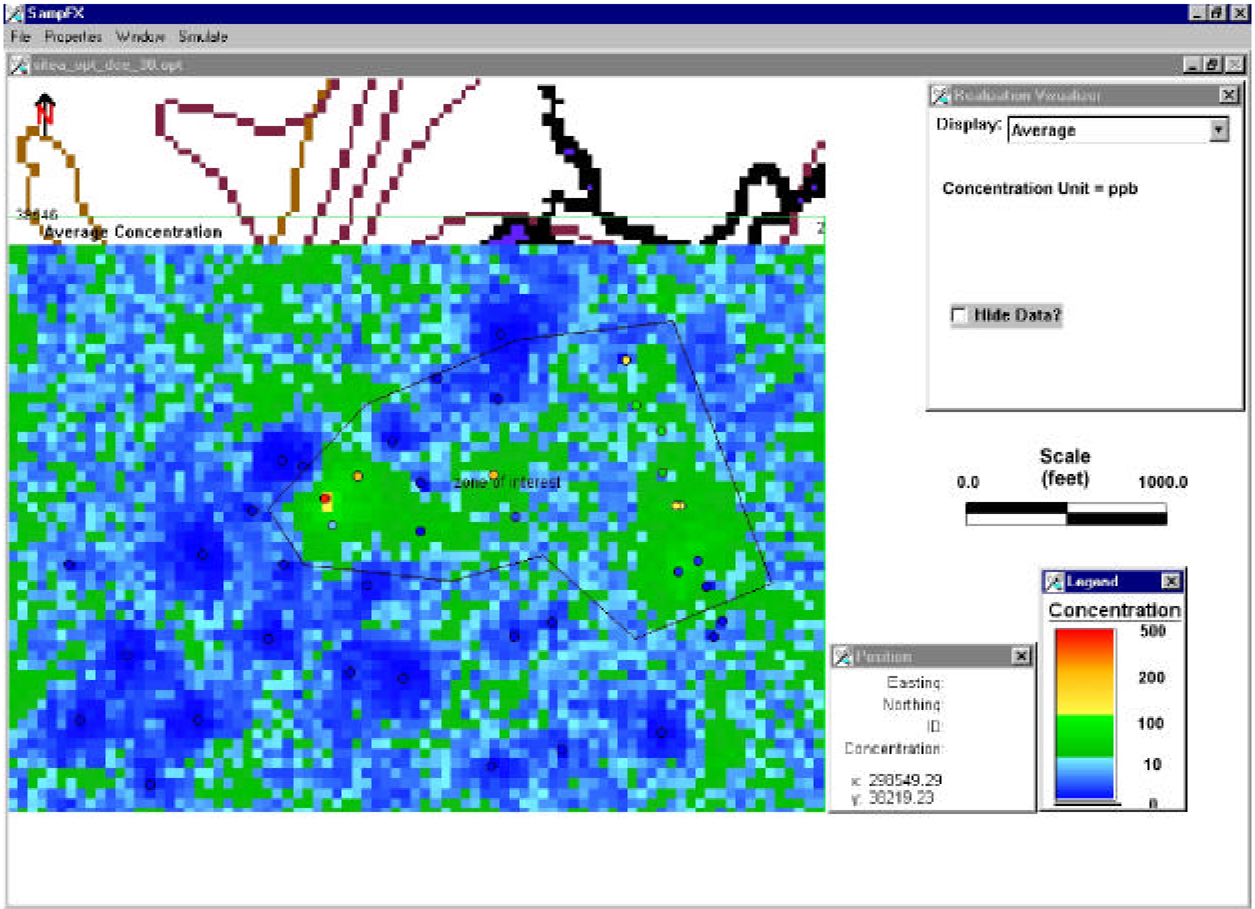

Figure 1. Sampling*FX* map for average DCE concentrations 30 ft below the water table at Site A.

represents the area with concentrations between 10 and 100 :g/L, and yellow and red regions have concentrations greater than 100 :g/L.

While the map contains all of the information needed to understand the contamination problem, interpretation of the map requires someone with experience in analyzing this type of information. The presentation of information is not very clear nor easily understood. Similar remarks apply to the variance and probability of exceeding a threshold concentration maps provided by Decision*FX*. The variance maps and the information they provide are discussed in the section evaluating the performance of Sampling*FX* in the Site N sample optimization problem and are not addressed in this section. The variance maps provide information on the areas of high uncertainty and therefore are useful in sample optimization.

Figure 1 exemplifies the problems the demonstration technical team had in performing an evaluation of the Sampling*FX* results. The concentration map overlies the base map and makes defining exact locations of plume boundaries impossible. The map does not contain coordinates to provide a reference for evaluation. The color scheme for the plot covers the entire range of concentration.(0 to 500:g/L). It is difficult to delineate the two threshold concentrations of 10 and 100 :g/L defined in the test problem for the purpose of comparison with the maps generated by the technical team. For these reasons, the demonstration technical team was unable to perform a detailed quantitative analysis of the Sampling*FX* output and only performed a visual comparison of the outputs. Decision*FX* did supply output files with the estimated concentration in each spatial location modeled by Sampling*FX* for the concentration maps. These files were reviewed to further understand the Sampling*FX* analysis results.

In Figure 1, the plume can be seen as the green area in the zone of interest. It is difficult to discern a pattern of contamination in the figure. Figure 2 presents a map of the probability of DCE exceeding a threshold concentration of 10 :g/L. In this map, the blue regions have less than a 25% probability of exceeding the threshold, green sections have

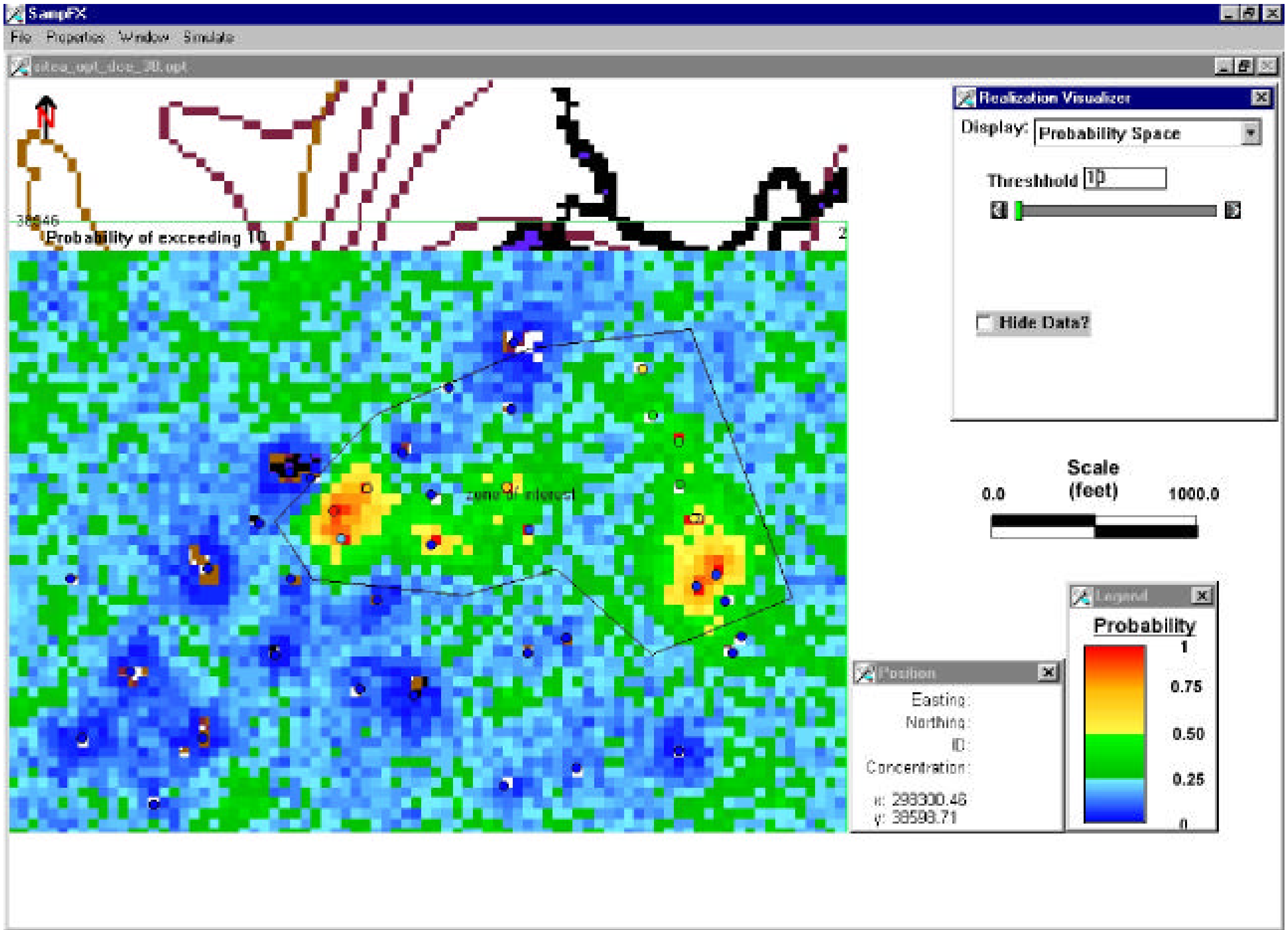

Figure 2. Sampling*FX*-generated probability map of regions exceeding the DCE 10-: g/L threshold at 30 ft below the water table for Site A.

between 25 and 50% probability, and yellow and red sections have greater than 50% probability. The zone of contamination can be seen more clearly in this figure than in Figure 1. Figure 2 indicates that there are several regions that have a greater than 50% probability of exceeding the 10-: g/L threshold concentration.

Figure 3 presents the baseline analysis of the complete data set. This map represents the maximum concentration in the region between 25 and 35 ft below ground surface. The water table ranged from 3 to 12 ft below ground surface, with an average of 5 ft. This baseline analysis map is most comparable to the map produced by Sampling*FX*. The baseline map was created using the kriging interpolation routine in the Surfer software package. After multiple sets of kriging parameters were evaluated, it was determined that an anisotropy ratio of 0.3 and a direction of ! 80E best represented the impacts of the direction of groundwater flow and transverse spread on the contamination data. In this figure, sample locations are marked with a circle. Buildings and the river are also included on the map. The map represents the region with DCE concentrations above 10 : g/L in blue and that above 100 : g/L in red. This map shows a long continuous plume originating from a building in the west of the map. The baseline analysis and results are different than those obtained by Decision*FX*.

Figure 3 is based on the complete data set and may therefore provide different results than found by Decision*FX*. Therefore, in an attempt to resolve the differences between the two approaches, the baseline approach was repeated using Surfer to examine only the data obtained by Decision*FX* through their sample optimization process. Figure 4 presents a contour map based on kriging of the data used by Decision*FX*. The map includes the measured concentrations posted to the right of the data locations. There are four areas of DCE contamination above 100 : g/L. These areas correspond to the data locations with measured values in excess of the 100-: g/L threshold. The 10-: g/L contour shows a continuous plume

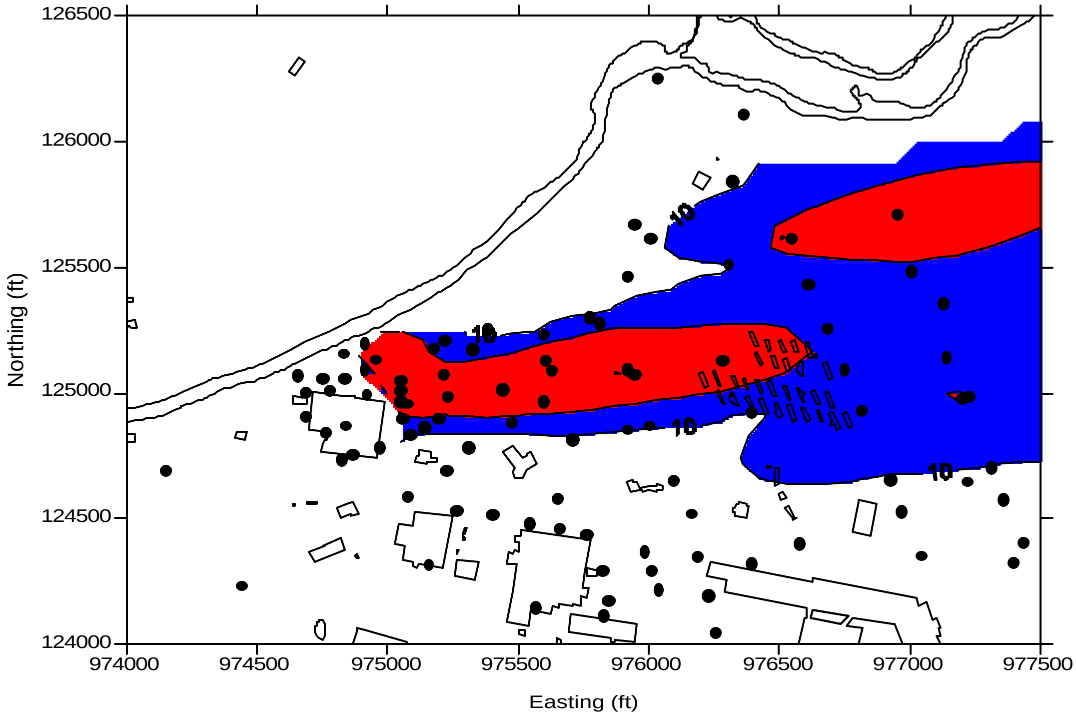

Figure 3. Baseline analysis of DCE concentration contours for the region between 25 and 35 ft below grade for Site A. The contours were generated using kriging. The blue region is the 10-: g/L contour, and the red region is the 100-: g/L contour.

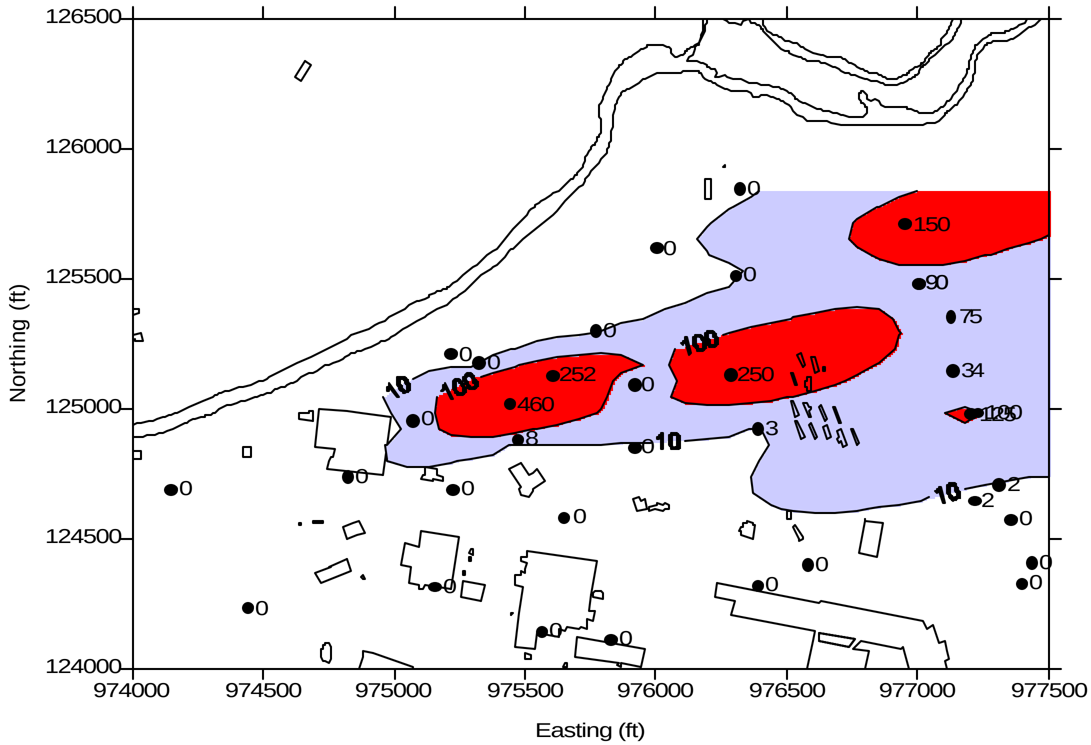

Figure 4. Site A DCE contours generated using the data set supplied to Decision*FX* and kriging for interpolation. The blue region is the 10-: g/L contour, and the red region is the 100-: g/L contour. Measured values are posted next to their sampling locations.

originating from the source region in the west and traveling to the eastern edge of the domain. The results obtained by the technical team using the data supplied to Decision*FX* are similar to the baseline analysis but different than the maps presented by Decision*FX* (Figures 1 and 2).

Figure 5 was generated by the technical team to be used to compare with Figure 1. Figure 5 shows the DCE concentrations above 10 : g/L in blue and above 100 : g/L in red. To create Figure 5, the technical team used Surfer to krig the data used by the Sampling*FX* analyst to create Figure 1. These data are averaged concentrations at each point rather than the actual concentrations provided by the technical team. (The actual concentrations are included in Figure 5 next to each sample location). Figures 1 and 5 are similar. In both cases, only an extremely small region is shown to be above the 100-: g/kg threshold concentration. The 10-: g/kg contours (green in Figure 1 and blue in Figure 5) are essentially the same. The consistency between Figures 1 and 5 indicates that the differences between maps (Figures 1 and 2 generated by Sampling*FX* as compared to Figures 3 and 4 generated by the technical team) are not due to differences in interpolation routines but result from the treatment of the data in the geostatistical simulation performed using Sampling*FX*. Comparison of the measured values on Figure 5 with the contours based on predicted values, indicates that there is a poor match between the actual data and the Sampling*FX* output.

The geostatistical simulation of Sampling*FX* appears to be less consistent with the measured data than the kriging interpolations produced by Surfer. This may be due to the application of Sampling*FX* to a

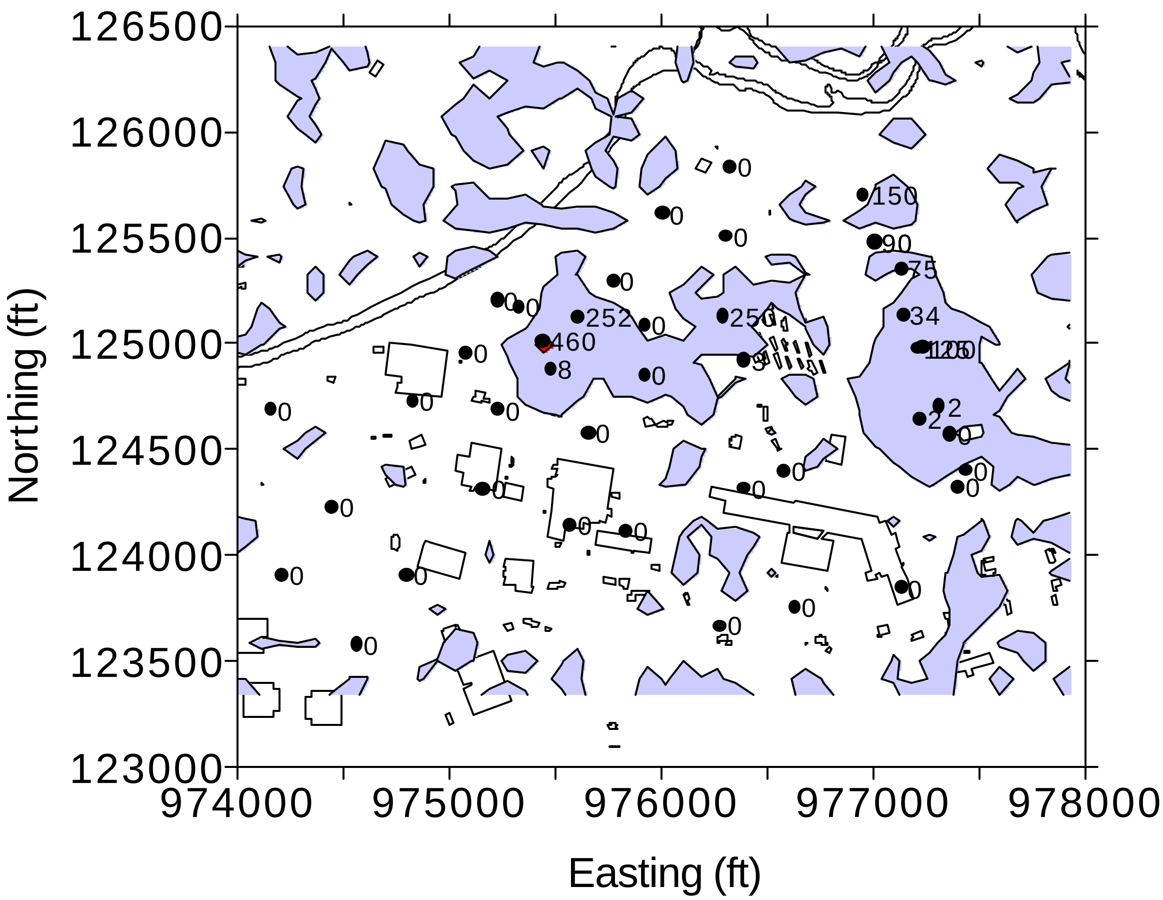

Figure 5. DCE concentration contours at 30 ft below the water table for Site A. These contours were generated using kriging interpolation based on the concentration values generated at each grid point in the Sampling*FX* analysis. The blue region is the 10-: g/kg contour, and the red region is the 100-: g/L contour. Measured values are posted next to their sampling locations.

groundwater flow contamination problem or to poor choice of model parameters. In the report of the demonstration results, the Decision*FX* analyst noted that this is not the typical application of the code. In addition, the Decision*FX* analyst stated in the Sampling*FX* report that more time would have been useful in improving the analysis. This was the last problem attempted by Decision*FX*, and sample optimization data was still being supplied on the last day of the 2-week demonstration.

Although Decision*FX* supplied estimates of the probability levels of the contamination using the Sampling*FX* geostatistics routines, the technical team decided not to evaluate that feature on this problem. The lack of consistency with the average concentration data indicated to the demonstration technical team that a meaningful evaluation could not be performed. For similar reasons, the area estimates for the plume were not reviewed extensively. It should be noted that the Sampling*FX* areas were much smaller than the areas obtained by kriging in the baseline analysis. These features of Sampling*FX* are reviewed on the Site N sample optimization and cost-benefit problems.

Site N Sample Optimization Problem

The Site N sample optimization problem was designed to evaluate the capability of the DSS to optimize sample locations for surface soil contaminated with heavy metals. The initial data set provided to the analyst contained data indicating contamination above the threshold concentrations in the southwest corner of the site. Figure 6 presents the site map generated by the technical test team with the initial sample locations marked with the symbol +. The map also shows arsenic-concentration contours at the two threshold concentrations. The region containing the initial data covers only a small fraction of the entire site. The map also contains locations of roads, surface water ponds, and creeks. Decision*FX* used Sampling*FX* to develop a sample optimization scheme to define the degree of contamination on the remaining portion of the site for As, Cd, and Cr. Decision*FX* generated maps of average concentration, variance in concentration, and probability of exceeding the threshold concentrations listed in Table 3. The entire set of maps generated by Decision*FX* was examined as part of the review process. In this report, arsenic contamination at the 125-mg/kg threshold concentration is presented. This set of maps and the findings of the review are similar to those found with the other contaminants and threshold concentrations.

Figure 7 is a screen capture presenting the final sample location and predicted arsenic concentrations with a base map of the site, including roads and water bodies. The process for selecting sample locations has been described in the problem description part of this section. Sample locations are denoted with a color-coded circle. The red circle in the southwest region of the site is an anomaly created forming the bitmap image required by Sampling*FX*. It is not a product of the software. The 80 additional sample locations are distributed throughout the site. A few samples were taken just outside the site boundary. These samples did not show any contamination above the threshold concentrations.

Figure 8 is a screen capture presenting the final estimate of base map concentrations and the sample locations selected using Sampling*FX* throughout Site N. Sample locations are marked with a circle. The sample locations are somewhat difficult to see in this figure and are presented more effectively in Figure 7. On the map, highest concentrations are denoted in red and lowest concentrations in blue. From the map, it can be seen that the entire site has been covered and that more samples have been taken in the regions of higher concentrations. The yellow area indicates concentrations above the 500-mg/kg threshold, and the green area indicates concentrations between the 125- and 500-mg/kg thresholds.

Figure 8 illustrates the problems the demonstration technical team had performing evaluations of the Sampling*FX* results. The concentration map covers the base map and makes defining exact locations of plume boundaries impossible. The map does not contain coordinates to provide a reference for evaluation. For these reasons, the demonstration technical team was unable to perform a detailed quantitative analysis of the Sampling*FX* output and performed visual inspection of the outputs. Decision*FX* did supply output files with the estimated concentration in each spatial location modeled by Sampling*FX* for the concentration maps. These files were used to further understand the output of Sampling*FX*.

Figure 9 presents a screen capture of the variance map for arsenic. This map can be used to identify

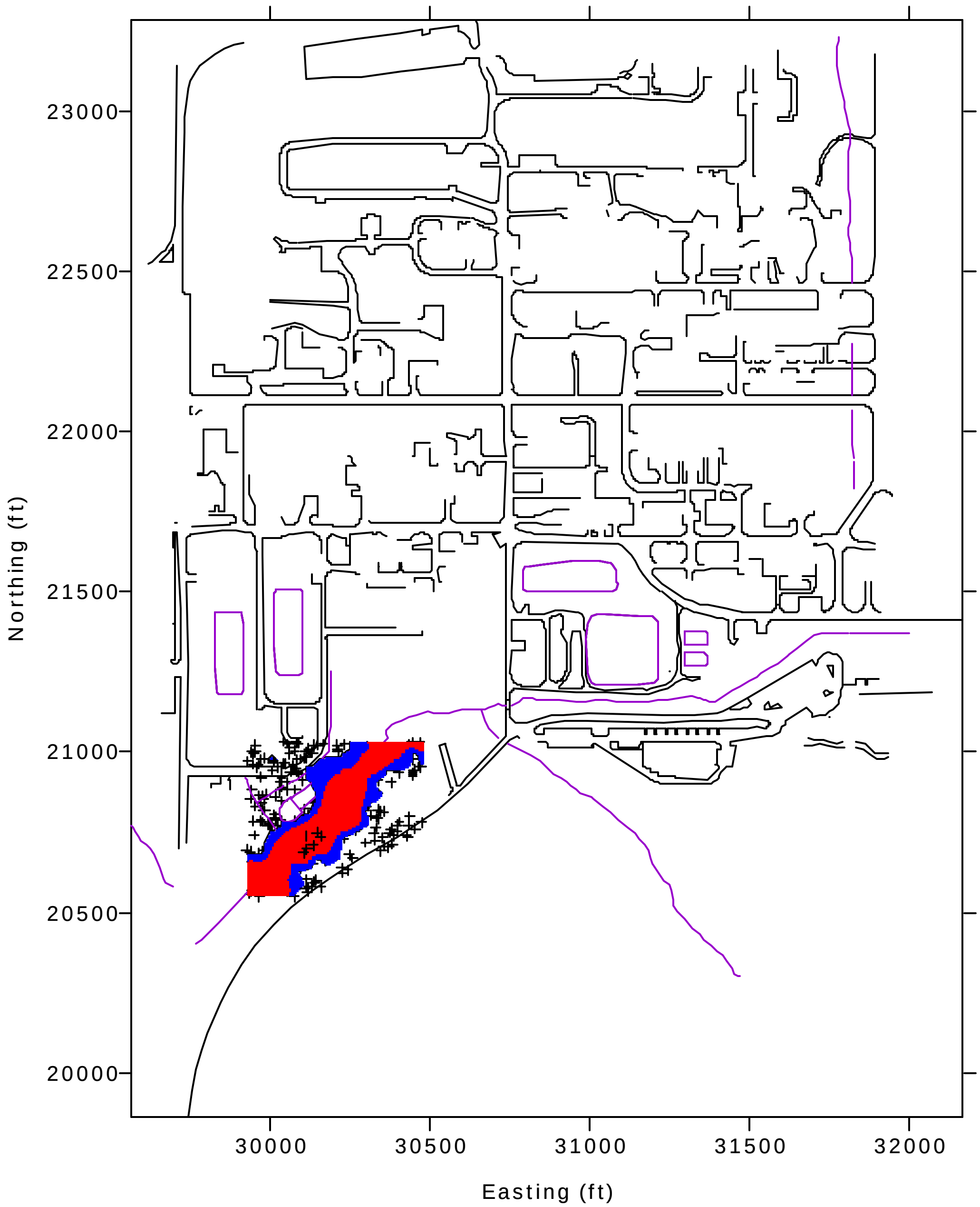

Figure 6. Initial data locations and arsenic contours at the two threshold concentrations, 125 mg/kg (blue) and 500 mg/kg (red), for Site N. Sample locations are marked by a +.

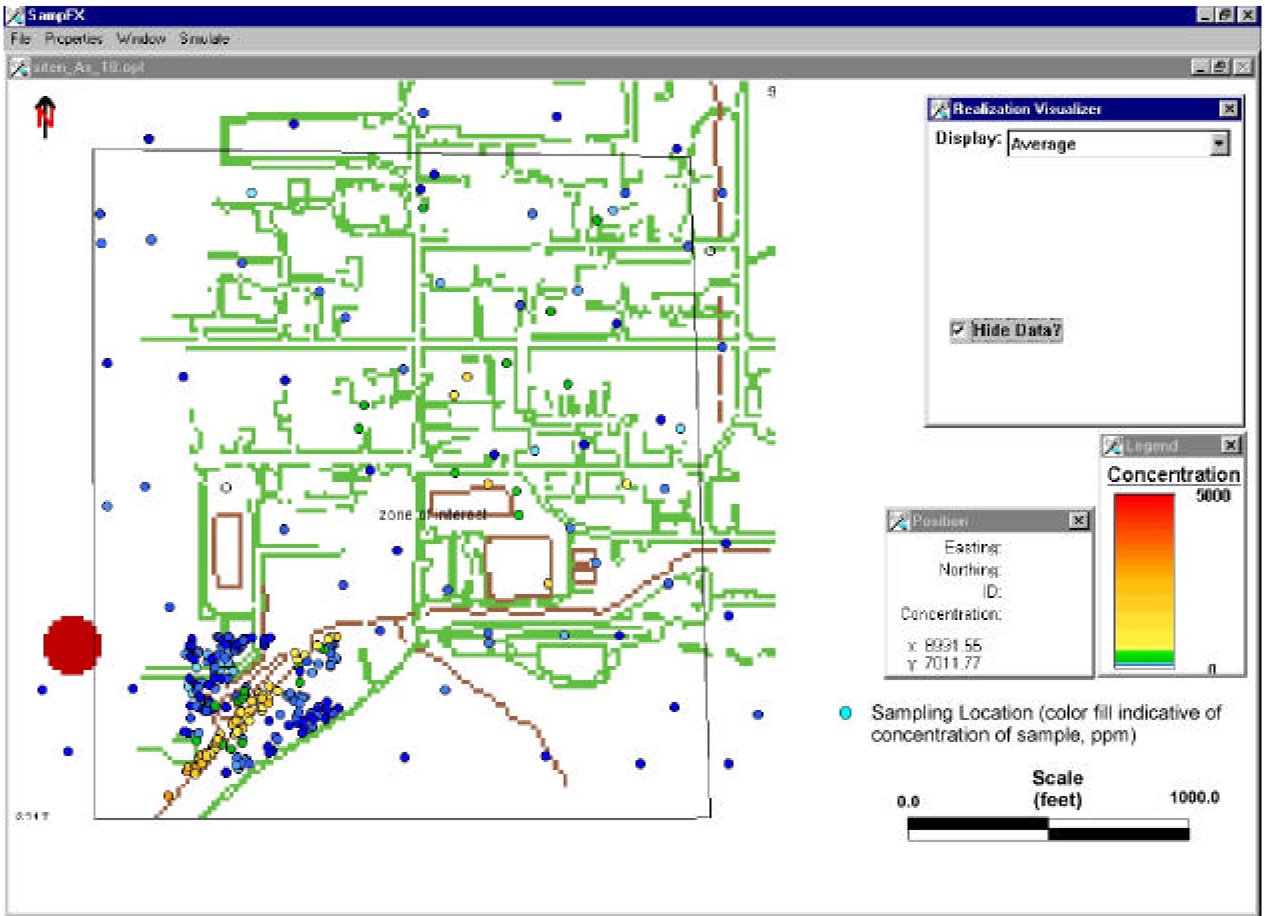

Figure 7. Final sample locations generated by Sampling*FX* for the Site N sample optimization problem.

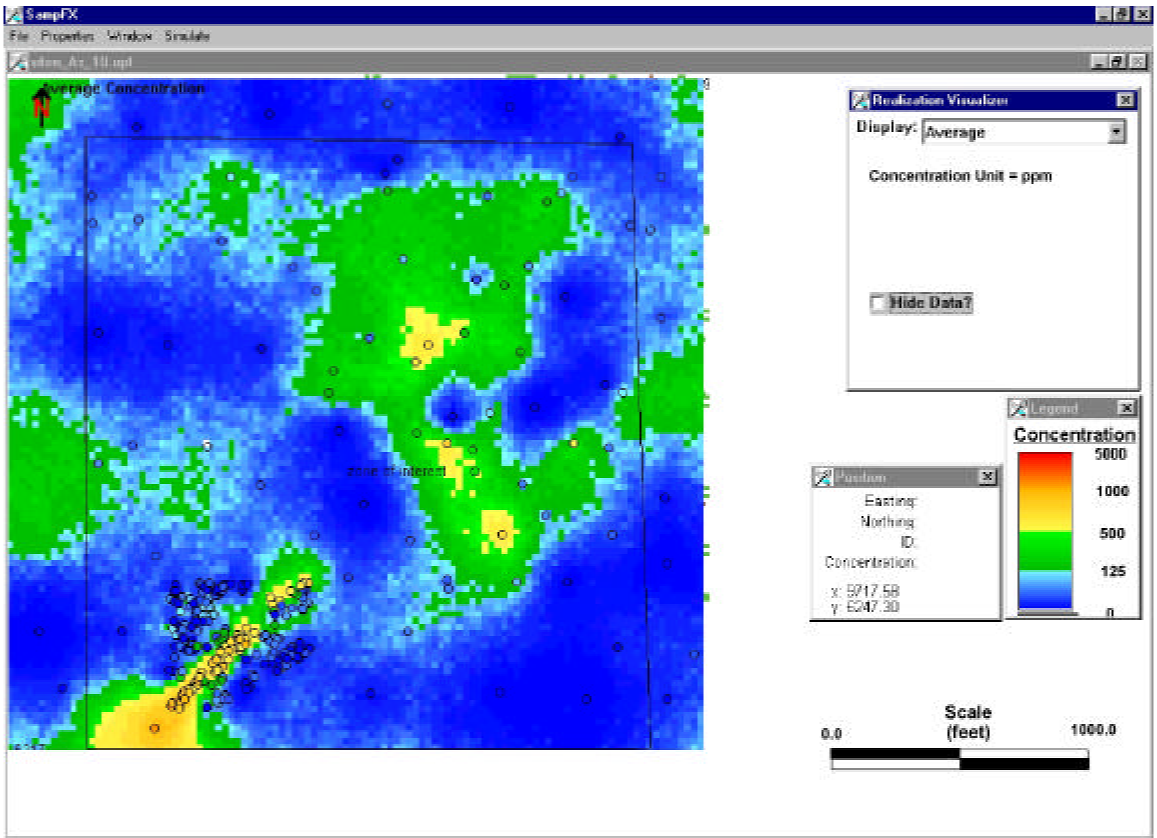

Figure 8. Sampling*FX*'s final contour map for average arsenic concentration for the Site N sample optimization problem.

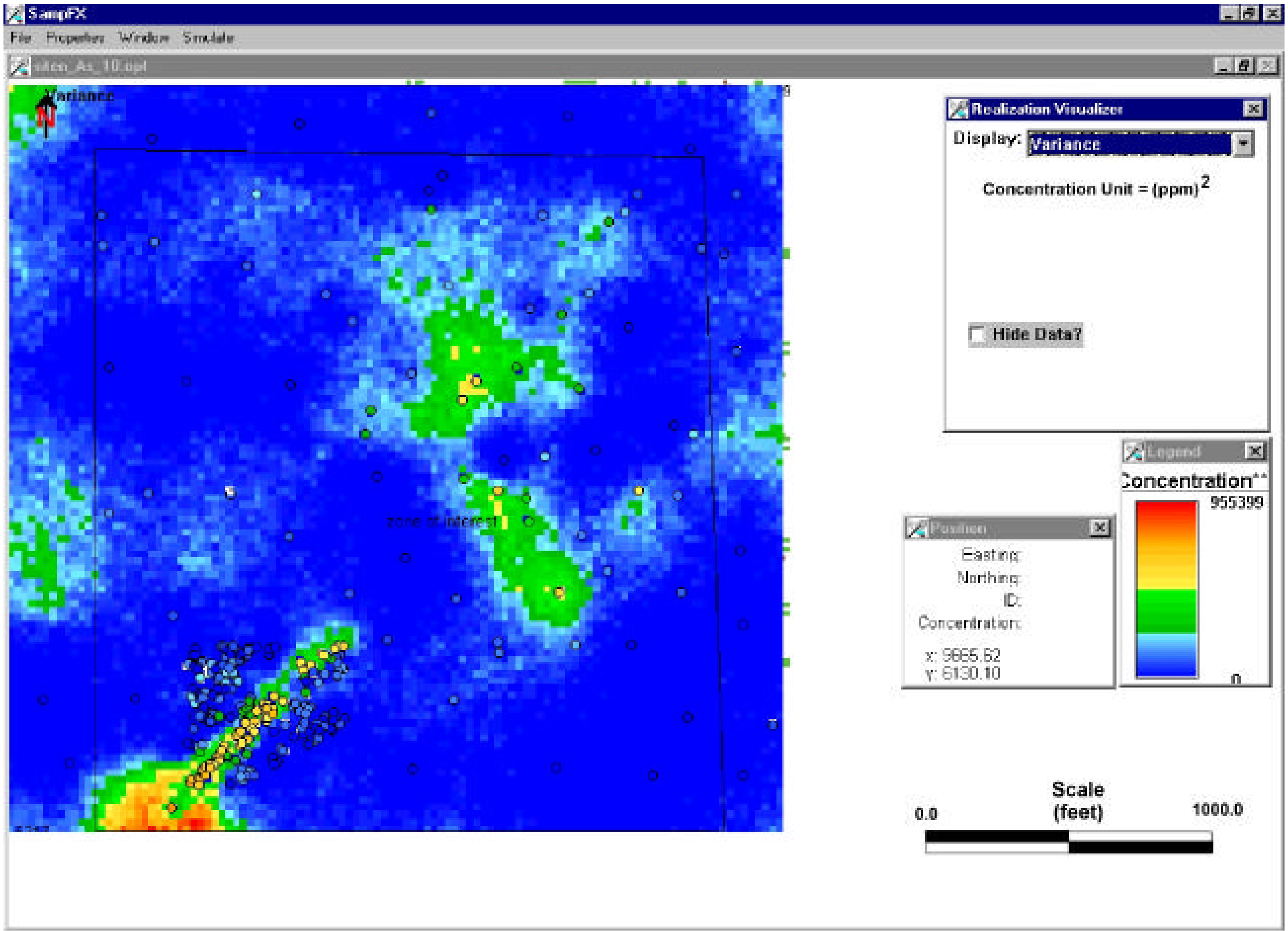

Figure 9. Sampling*FX* variance map for arsenic concentrations for the Site N sample optimization problem.

areas of high variance, which are related to uncertainty, and therefore provide guidance on sample locations. Variance is a measure of the difference between the predicted contamination level and its mean value at that location. Low variance shows the data or mean is consistent with the modeled values. High variance shows no consistency between the data or mean and the modeled values for the location. Therefore, areas of high variance are areas of high uncertainty in the model predictions. Comparing Figure 9 with Figure 8, it can be seen that the variance map does highlight the areas of high variance in the southeast corner and the central area of the site.

Figure 10 is a screen capture of a map that presents the probability of exceeding 125-mg/kg threshold for arsenic. The partially obscured (clearly visible in Figure 7) red circular region in Figure 10 to the north of the data supplied to the analyst is part of the Sampling*FX* visualization background map and does not indicate a high probability region. A comparison of Figures 8 and 10 shows some unexpected differences. In particular, the region with a probability greater than 50% of exceeding the 125-mg/kg threshold (the region in yellow, orange, or red in Figure 10) is much smaller than the 125-mg/kg concentration contour (the region in green, yellow, or orange in Figure 8). This difference results from the different interpretations of the data. Figure 8 presents the average concentration, which does not necessarily correspond directly to the 50% probability level of Figure 10. Furthermore, the technical team expected that regions of the site with an average arsenic concentration greater than 500 mg/kg (the yellow regions in Figure 8) should have a high probability (>90%) of exceeding the 125 mg/kg threshold. However, these regions are displayed in Figure 10 as having a probability of between 50 and 75% (the yellow regions of Figure 10) of exceeding the threshold. The technical team was unable to determine the causes for the apparent discrepancy between the average concentration map (Figure 8) and the probability map (Figure 10).

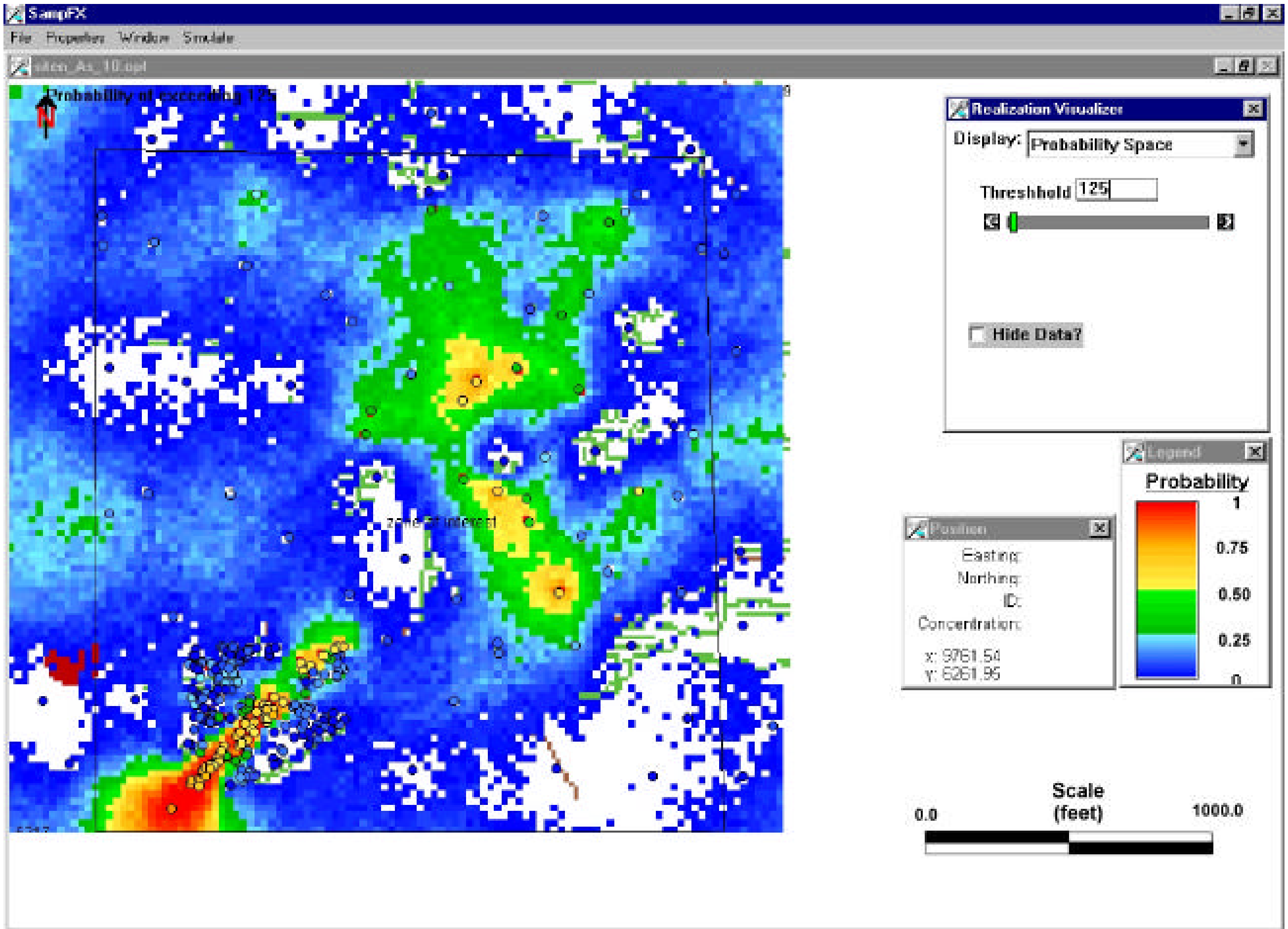

Figure 10. Sampling*FX* map of the probability of exceeding the 125-mg/kg threshold for arsenic, Site N sample optimization problem.

Figure 11 presents the baseline analysis of Site N conducted using the entire data set (4187 points). The results were generated using the Surfer software package and using kriging for data interpolation. A precise comparison of Figures 8 and 11 is difficult because of the lack of a base map with site features or coordinates on Figure 8. However, it can be seen that Sampling*FX* was able to locate most, though not all, of the regions contaminated above the arsenic threshold concentration of 500 mg/kg while being limited, by sampling costs, to using only 80 data points (2% of the complete data set).

The software was able to find the three major contamination areas in the central region of the site that had arsenic concentrations in excess of 500 mg/kg. Sampling*FX* did not locate small regions in the northeast corner and the north central part of the site that had several measured points above the 500-mg/kg threshold concentration. These regions were areas that would be covered by a circle with less than a 50-ft radius. Given the limited number of additional samples, it is not surprising that they were not found. Overall, Sampling*FX* did a very good job of defining the contamination regions with limited additional data.

Sampling*FX* was also able to find and define the regions containing arsenic contamination above 125 mg/kg. The Sampling*FX* analysis predicted a much broader area of contamination than that found in the baseline kriging analysis. However, the area where the probability of exceeding the arsenic 125-mg/kg threshold is greater than 50% (Figure 10) is slightly less than the baseline kriging analysis. This reflects the uncertainty resulting from the limited number of data points. Sampling*FX* predicted one large area of contamination above the 125-mg/kg threshold that was not as large as presented. This is in the southwest region of the site, due north of the high-concentration region supplied to the analyst (approximately an easting of 31,100

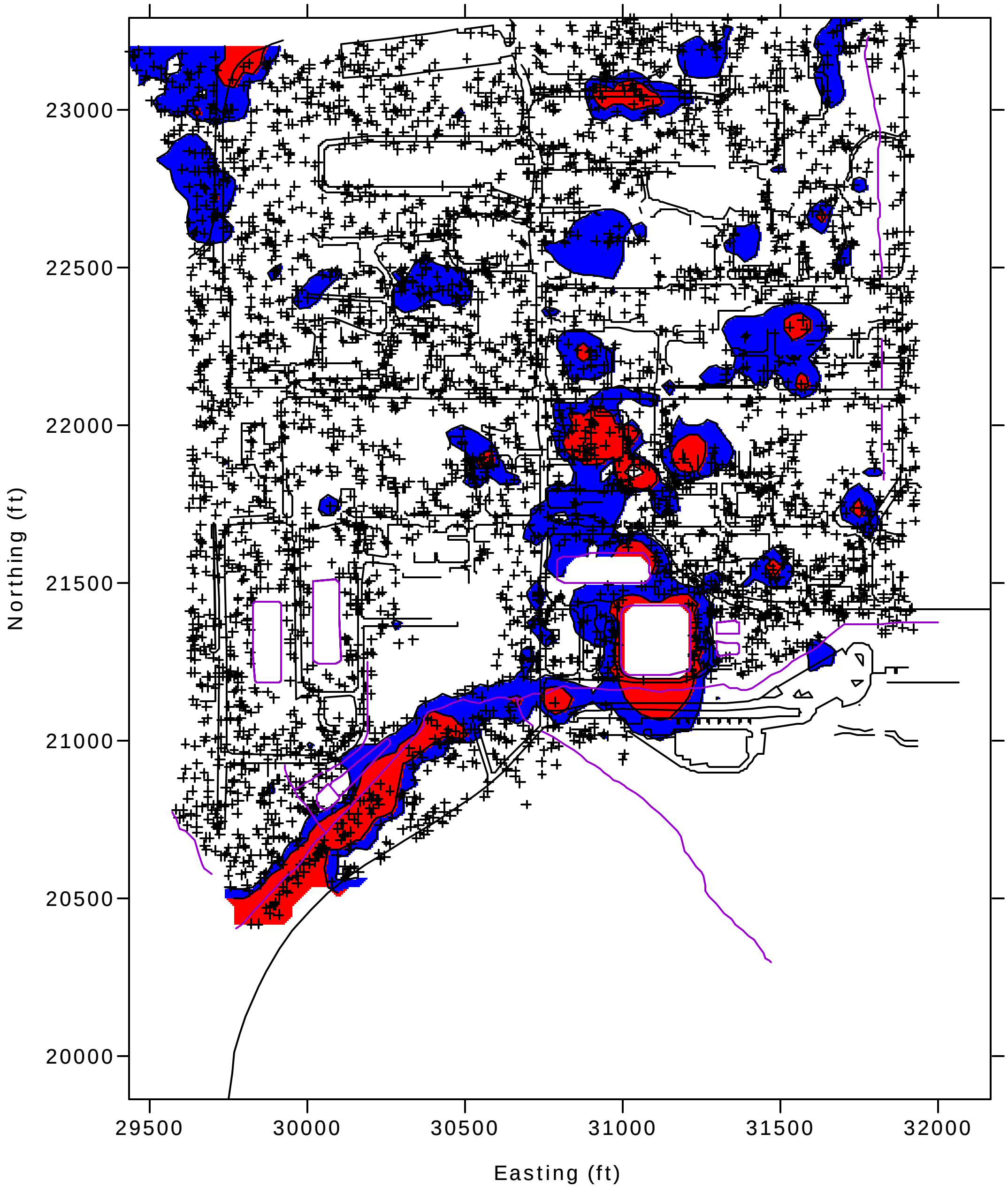

Figure 11. Baseline analysis (4187 data points) for arsenic for the Site N sample optimization problem. The blue region represents the 125-mg/kg contour level, and the red, the 500-mg/kg level.

and a northing of 21,750 in Figure 11). The cause for the overestimation can be determined by examining the complete data set. In the complete data set, there are two samples in this region with arsenic concentrations slightly above the threshold. It happens that the Decision*FX* analyst requested data at one of these samples that contained high concentration of arsenic. With the limited number of samples, the influence of this data point on the concentration contours is spread out over a larger region than for the complete data set. Examining the probability map for exceeding the threshold concentration of 125 mg/kg (Figure 10) it can be seen that other than at the exact location of the sample (the yellow mark), there is less than a 25% probability (light blue region) of exceeding the threshold. This example illustrates the advantages of performing the geostatistical analysis and the problems of having incomplete knowledge about contamination.

SamplingFX was used to perform a similar analysis for arsenic at the upper threshold and chromium and cadmium at both threshold concentrations. A total of nine maps similar to those shown in Figures 8–10 were prepared. Through comparison of all the SamplingFX results with the baseline geostatistical analysis and the baseline kriging analysis, it has been demonstrated that SamplingFX provided reasonable and accurate characterization of this site given the constraint of only 80 additional sample locations.

Figure 12 presents an analysis of the area of arsenic contamination vs the probability of exceeding the 125-mg/kg threshold for each round of sampling. DecisionFX generates this analysis by exporting statistical information produced by SamplingFX into Microsoft Excel and generating the graph. The graph illustrates the value of collecting additional data to refine the estimate of contaminated area. The area estimated by SamplingFX with a 50% probability of exceeding the arsenic concentration of 125 mg/kg is approximately 675,000 ft^2. The baseline geostatistical analysis, using the entire data set, estimated the area with a 50% probability of exceeding the arsenic concentration of 125 mg/kg to be 955,000 ft^2. As previously discussed, SamplingFX did not identify all areas with arsenic contamination during the sampling optimization exercise, with the result that it estimated a smaller area of arsenic contamination. When compared to the baseline geostatistical analysis with the entire data set, SamplingFX identified approximately 70% of the entire site that had arsenic contamination above 125 mg/kg. The technical team concluded that this is a reasonable match considering the constraint of 80 additional samples to characterize the entire site.

Site N Cost-Benefit Problem

SamplingFX was used to evaluate the surface soil contamination data for three contaminants—As, Cd, and Cr—at Site N. In this problem, 524 data points were supplied over a 14-acre region of the site. In addition, a bitmap containing the roads, creeks, and surface water bodies was supplied to assist in the interpretation of the data. Figure 13 presents a map generated by SamplingFX of the sample locations marked with a color-coded circle. The color key is to the right of the diagram. Unfortunately, the scale does not list the concentrations that correspond to the colors. However, red areas are the highest concentration and blue the lowest. This map forms the basis for further analysis. DecisionFX generated maps of average concentration, variance in concentration, and the probability of exceeding the

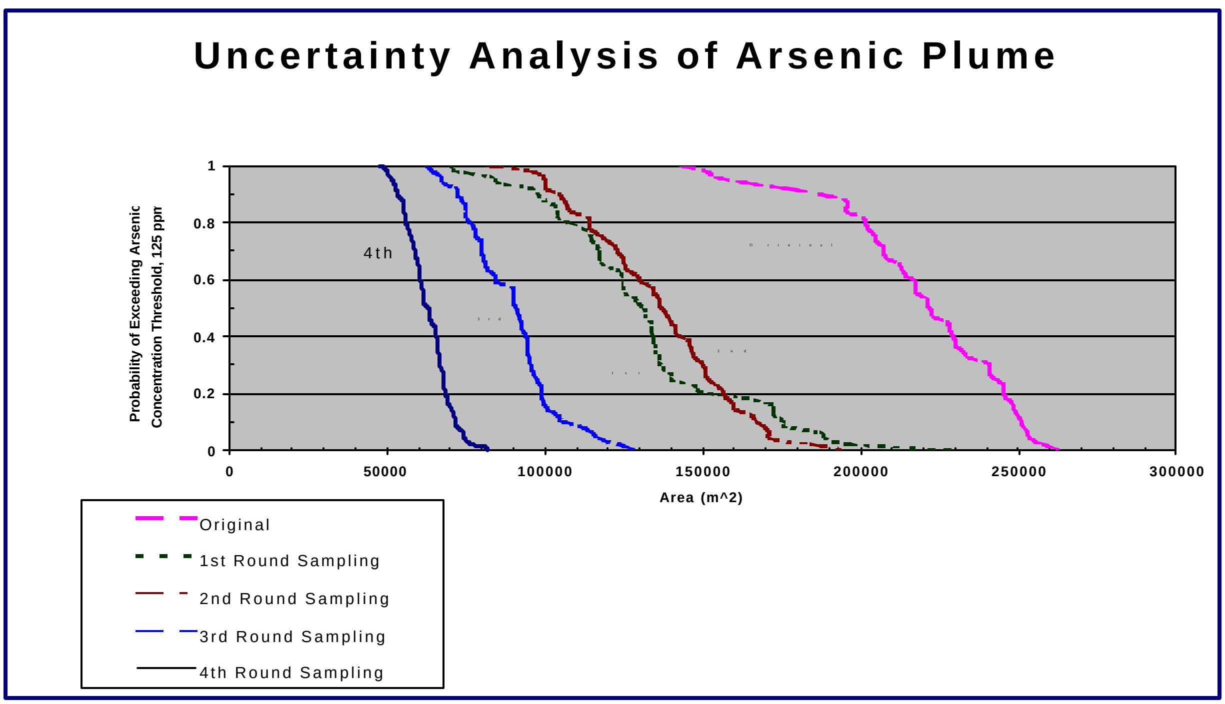

Figure 12. SamplingFX analysis for area of contamination as a function of the probability of exceeding the 125-mg/kg arsenic threshold. The area is calculated for each round of sampling during the sample optimization problem.

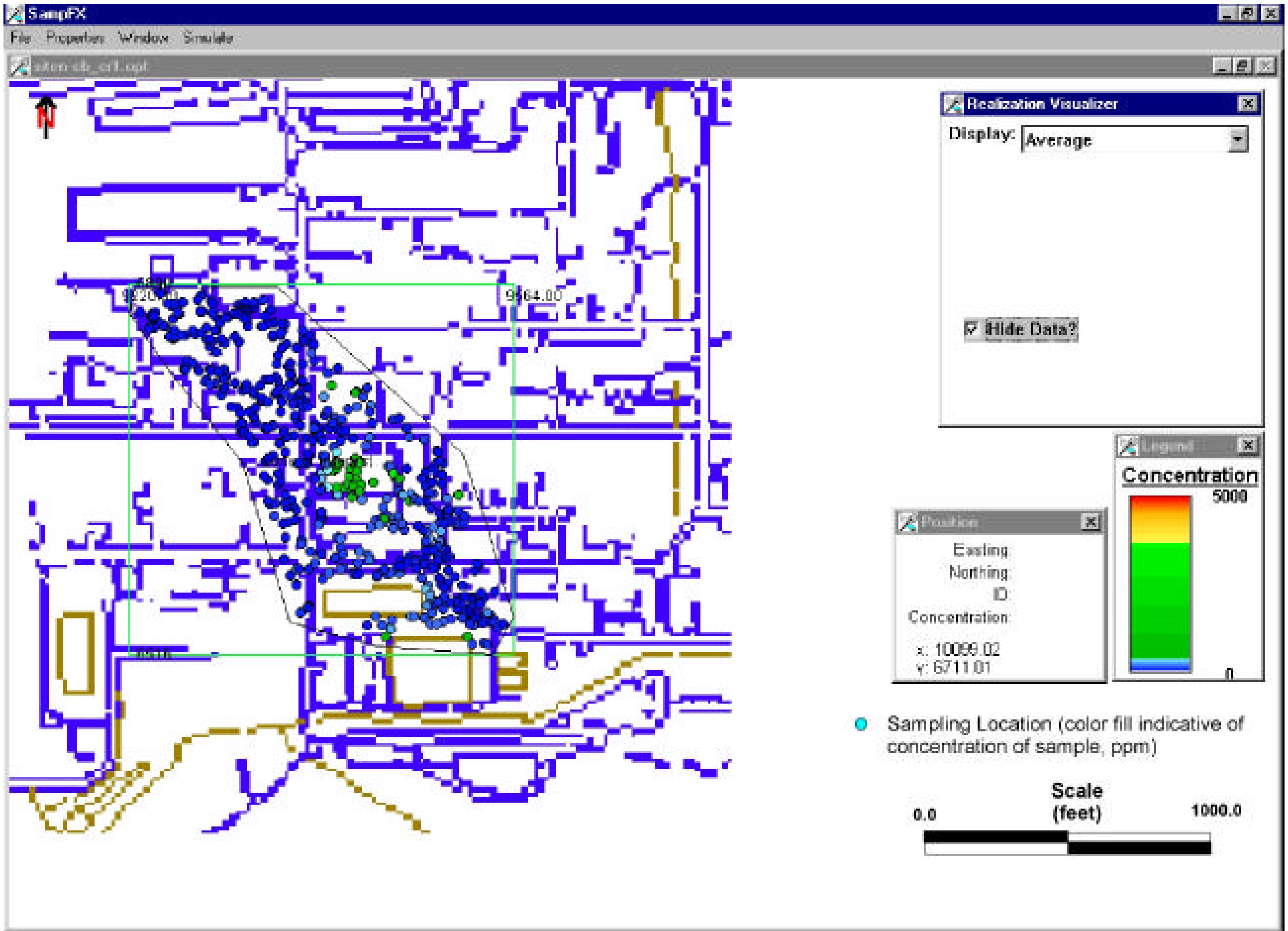

Figure 13. Sampling*FX* map for the Site N cost-benefit problem containing soil sample locations color-coded to match measured arsenic concentrations.

threshold concentrations that are shown in Table 4. The entire set of maps generated by Decision*FX* was examined as part of the evaluation process; however, only arsenic contamination information at the 75-mg/kg threshold is presented. Arsenic was chosen to represent the Decision*FX* analysis because this was the information presented in the Decision*FX* report from the demonstration activities.

Figure 14 is a screen capture presenting the final estimate of the average arsenic concentration contours based on the data supplied to Decision*FX*. Sample locations are marked with a color-coded circle. The color key is to the right of the diagram and concentrations are labeled on the key. The sample locations are somewhat difficult to see in this figure. On the map, highest concentrations are denoted in red and lowest concentrations in blue. In Figure 14, yellow areas are above the 500-mg/kg threshold for arsenic, and green areas indicate concentrations between 75 and 500 mg/kg. This figure also contains a polygon denoting the region of interest for the analysis. Regions outside of this polygon do not contain data and are model extrapolations that should be disregarded in the analysis. At distances that are far from the nearest measured data location, the model sets the projected concentration to the mean value. In this example, the mean value lies between 75 and 500 mg/kg. Therefore, most of the region outside the area of interest is green.

Figure 14 illustrates the problems the demonstration technical team had with the technical evaluation of the Sampling*FX* results. The concentration map covers the base map and makes defining exact locations of plume boundaries impossible. The map does not contain coordinates to provide a reference for evaluation. For these reasons, the demonstration technical team was unable to perform a detailed quantitative analysis of the Sampling*FX* output and performed visual inspection of the outputs.

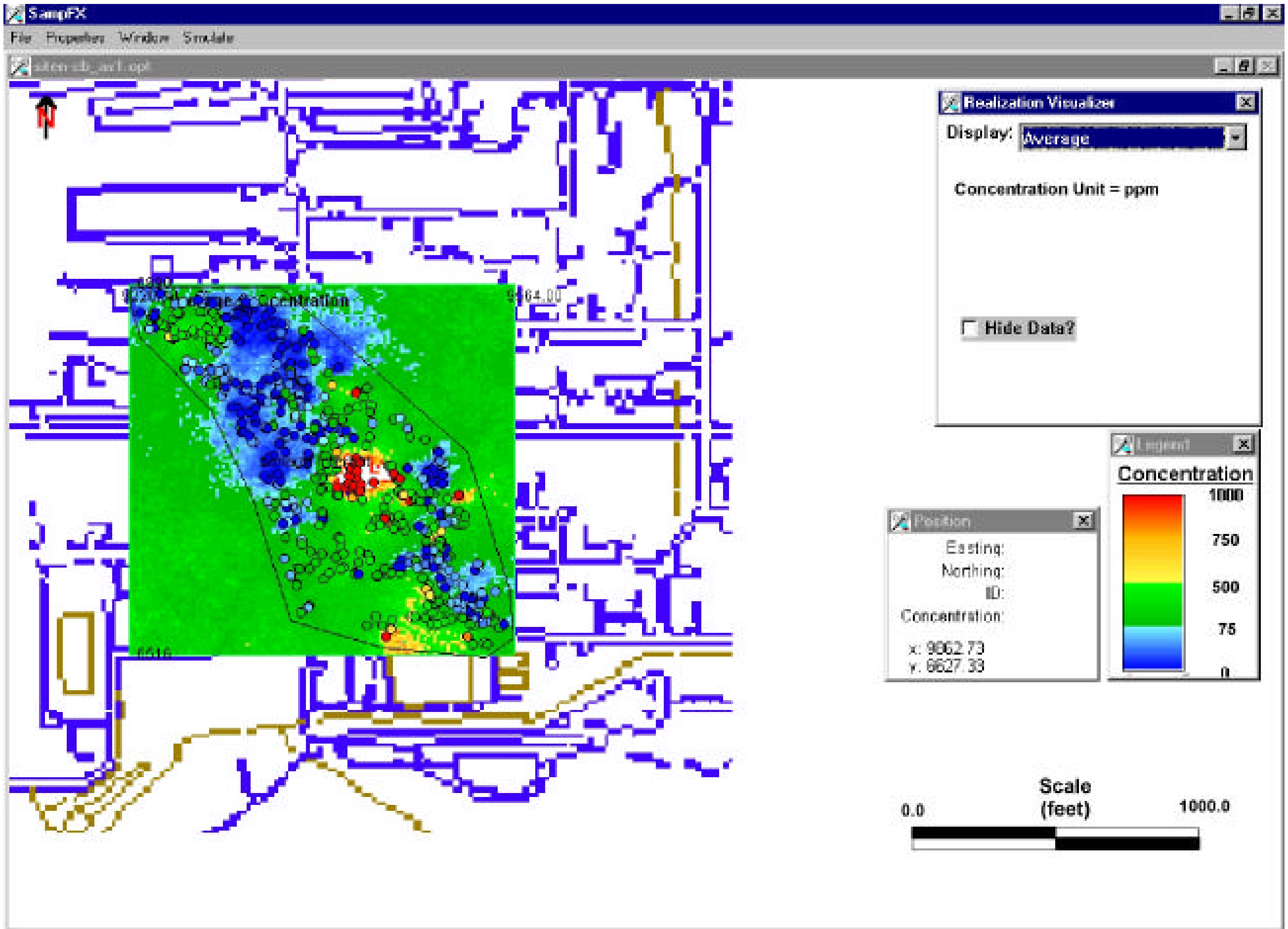

Figure 14. Sampling*FX* estimate of average arsenic concentration for the Site N cost-benefit problem.

Figure 15 presents a screen capture of a map that presents the probability of exceeding the threshold concentration of 75 mg/kg for arsenic. A comparison of Figures 14 and 15 shows that they have the same spatial characteristics. In particular, the region with a probability greater than 50% of exceeding the 75-mg/kg threshold (the region in yellow, orange, or red in Figure 15) is similar to the 75-mg/kg concentration contour (the region in green, yellow or orange in Figure 14). Also, there are substantial areas with a greater than 90% probability (red regions) of exceeding the 75-mg/kg threshold. In this problem, unlike the Site N sample optimization problem, there are enough data to make the average concentration correspond closely to the 50% probability level and to identify areas of high probability.

Figure 16 presents the analysis of the Site N cost-benefit problem conducted by demonstration technical team using the entire data set. These results were generated using the Surfer software and kriging for data interpolation. Precise comparison of Figures 14 and 16 is difficult because Figure 14 lacks a base map with site features or coordinates in the region of the analysis. However, it can be seen that the Sampling*FX* analysis and the baseline analysis correspond closely. Both indicate the region above the 500-mg/kg arsenic threshold concentration in the center (red on the baseline analysis; yellow, orange or red on the Sampling*FX* analysis), with several smaller high-concentration regions throughout the site. Both analyses also indicate the regions with contamination below the lower arsenic threshold concentration of 75 mg/kg (clear on baseline analysis map; blue on Sampling*FX* analysis map).

The problem definition requested that the analyst estimate the area of contamination at three probability levels—10, 50, and 90%—for each threshold concentration. The probability level corresponds to the amount of uncertainty in the decision. The 10% probability level is the level at which the analyst believes that there is at least a 10% probability that the contaminant concentration at a

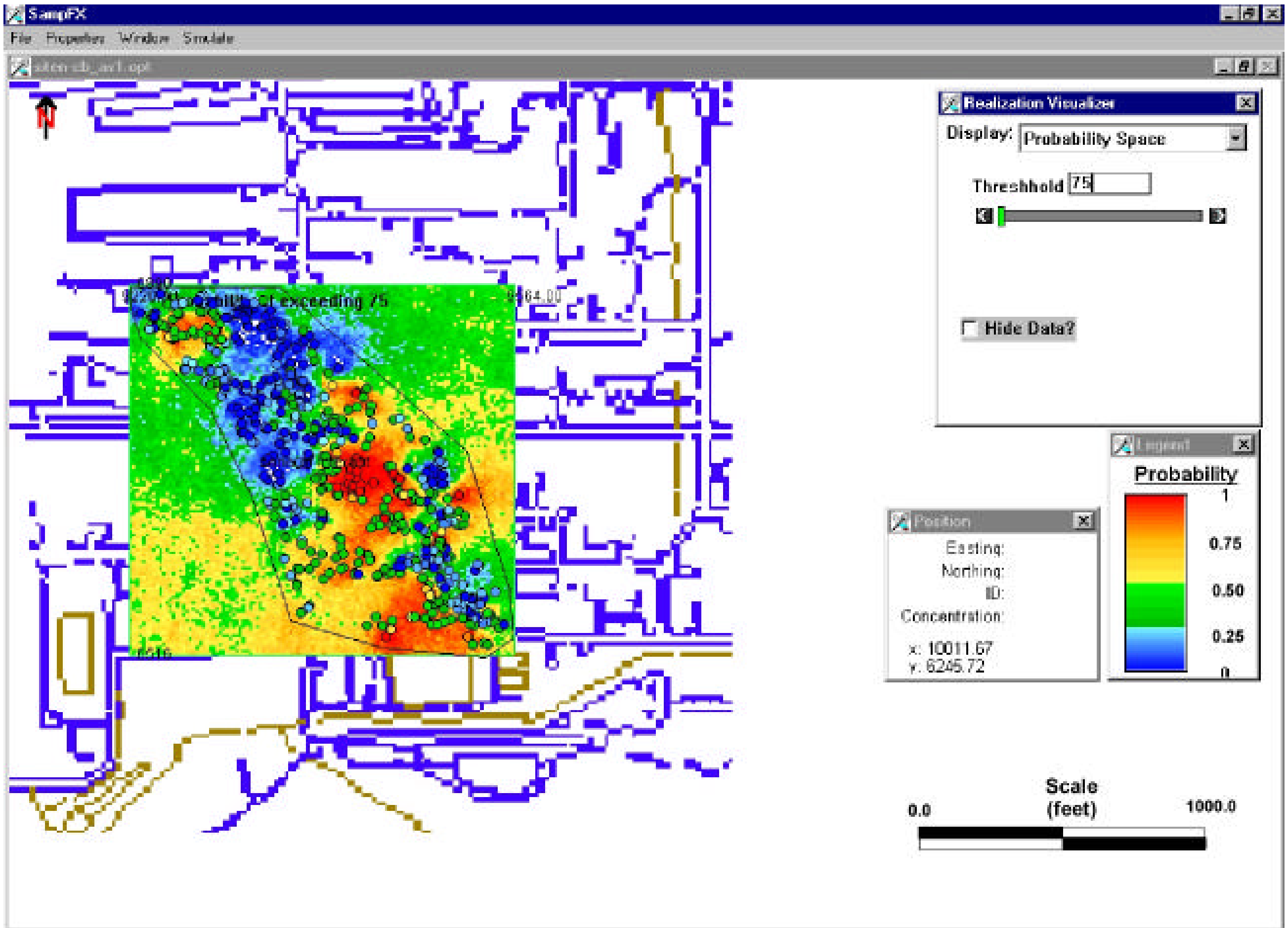

Figure 15. Sampling*FX*-generated map of the probability of exceeding the 75-mg/kg threshold for arsenic for the Site N cost-benefit problem.

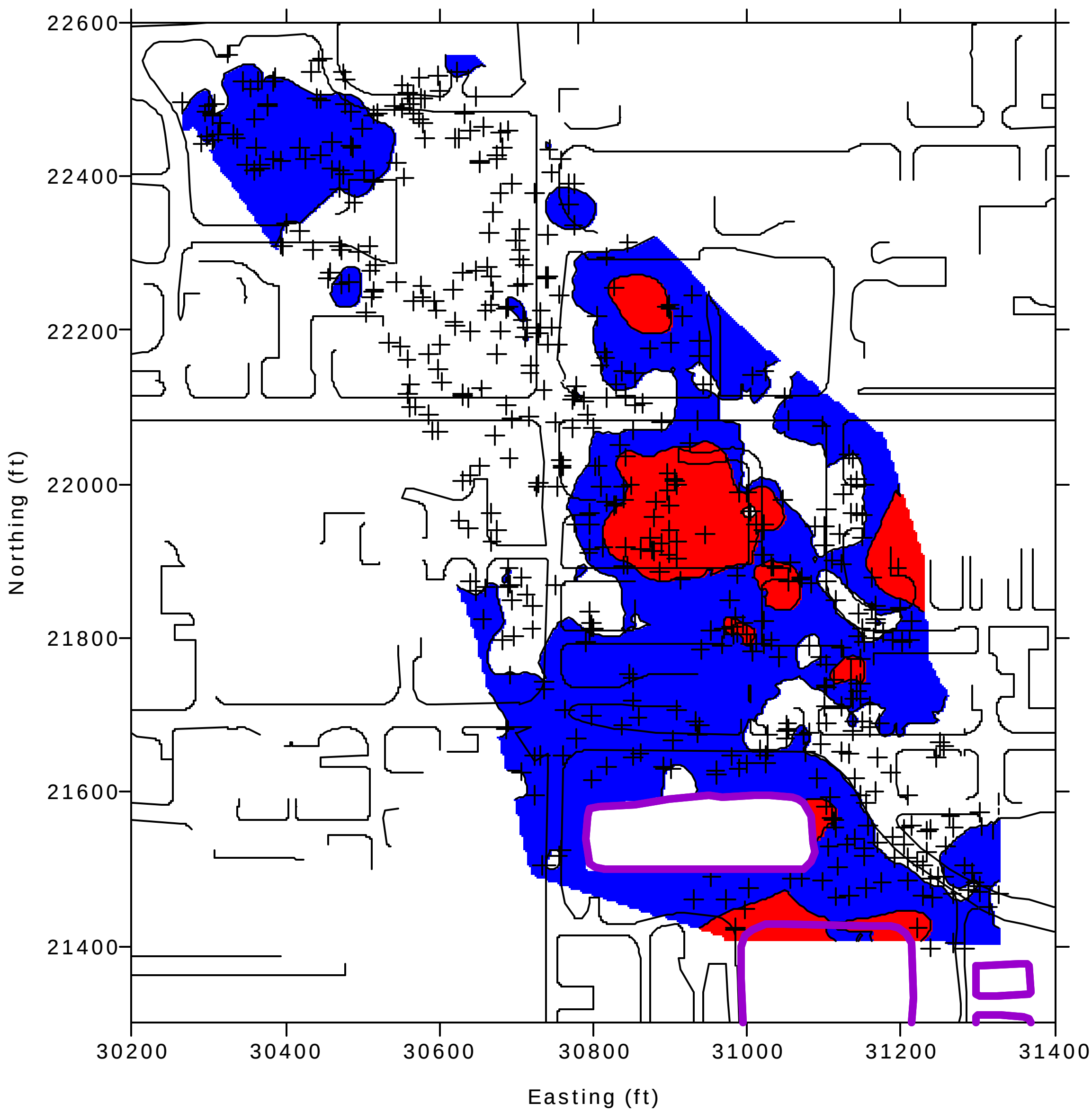

Figure 16. Baseline analysis of Site N cost-benefit arsenic concentration contours performed using Surfer with kriging interpolation of the data. Areas in blue correspond to regions above the 75-mg/kg threshold for arsenic. Areas in red correspond to regions above the 500-mg/kg threshold.

specified location exceeds the threshold concentration. This leads to a larger estimate of the area of contamination as compared to the 50% probability level. Similarly, the 90% probability level corresponds to level at which the analyst believes that there is a 90% probability that the contaminant concentration at a specified location exceeds the threshold concentration. For comparison, area estimates were generated using Surfer to interpolate the data using kriging and an independent, somewhat different geostatistical approach than that used in Sampling*FX*. The differences in approach have been discussed previously in this section (under "Sampling*FX* Technical Approach"). Table 6 presents the estimates of the area of contamination derived from the baseline kriging analysis, from the baseline geostatistical analysis at the 50% probability level, and from the Sampling*FX* analysis at the 50% probability level. As the table indicates, the three estimates show reasonable agreement at the 50% probability level. Estimates are generally within 20% of each other, and there is no clear pattern indicating that one method always over- or underestimates area as compared to the others. Considering that all three approaches used slightly different boundaries and slightly different parameters for kriging, the technical team concluded that the agreement is reasonable.

At the 50% probability level, Sampling*FX* predicts that the chromium concentration exceeds the threshold concentration of 3700 mg/kg in an area of 96 ft^2 (Table 6). This area represents one block of the simulation domain and is the minimum non-zero area estimate for the analysis. The other two baseline approaches predict zero area above the threshold. The maximum chromium concentration in the data set was 3366 mg/kg. The estimate of one simulation block exceeding the threshold arises from the multiple simulation statistical approach used in the Sampling*FX* analysis. The multiple simulation approach indicates that even though the maximum measured value is 9% less than the threshold, there still exists a 50% probability that a measured value could exceed the threshold. This reflects the fact that there is no guarantee that the maximum measured value corresponds to the actual maximum value.

Table 7 compares the estimates of the area of contamination at the 10% probability level (maximum area) generated by the baseline geostatistical analysis and by Sampling*FX*. The table also includes an area estimate based on the Sampling*FX* 90% probability level (minimum area) for comparison. Sampling*FX* supplied the area estimates in units of square meters because it requires all measurements to be in meters. The values were converted to square feet by the technical team for comparison with the baseline analyses.

As indicated in Table 7, the baseline geostatistical approach usually predicts a greater area exceeding the threshold at the 10% probability than does the Sampling*FX* estimate at the same probability level. In addition, Sampling*FX*'s 10, 50, and 90%

Table 6. Baseline and Sampling*FX* estimates of the area of contamination at the 50% probability level for the Site N cost-benefit problem

Contaminant	Threshold concentration (mg/kg)	Area of contamination (ft^2)		
		Baseline kriging with Surfer	Baseline kriging with geostatistical 50% probability level	Sampling*FX* 50% probability level
Arsenic	75	330,000	389,000	362,000
	500	57,000	44,000	52,200
Cadmium	70	285,000	325,000	263,000
	700	17,300	17,000	19,000
Chromium	370	37,100	30,500	44,400
	3700	0	0	96

Table 7. Baseline and Sampling*FX* estimates of the area of contamination at the 10% probability level for the Site N cost-benefit problem, with the Sampling*FX* 90% probability level added for comparison

Contaminant	Threshold concentration (mg/kg)	Area of contamination (ft^2)		
		Baseline kriging with geostatistical 10% probability level	Sampling*FX* 10% probability level	Sampling*FX* 90% probability level
Arsenic	75	461,000	374,000	350,000
	500	135,000	59,200	47,800
Cadmium	70	402,000	272,000	251,000
	700	22,100	22,600	16,200
Chromium	370	77,500	50,000	39,800
	3700	0	0	0

estimates show a much narrower range of values (Table 6) as compared to the baseline geostatistical approach. This is due to the differences in approach in estimating area. Sampling*FX* calculates the total area that will exceed the threshold for each of its multiple simulations of the data. Each simulation area estimate is then analyzed to determine the distribution of contaminated area. This distribution is used to determine the statistical probabilities of a given area exceeding the threshold. The baseline approach performs a single simulation and calculates the probability level at each computational point in the analysis. There tends to be more variability in the predicted concentration at any one location as compared to the mean concentration for the entire site. Therefore, the range of area estimates will be greater for the baseline approach.

The small variation in Sampling*FX*'s estimated areas is not consistent with the wide variation in area between the 10% probability level as shown on the software's arsenic probability map (any color other than dark blue in Figure 15) and the 90% probability level (red regions in Figure 15). Figure15 seems to show a substantial difference between the areas for arsenic at the 75-mg/kg threshold, whereas Table 7 indicates only a 10% difference. The larger spread in area seen on Figure 15 is more consistent with the baseline geostatistical analysis. The reason for the discrepancy between the probability map (Figure 15) and area estimates in Table 7, both of which were provided by Decision*FX*, is that the probability map represents the variability on a local scale while the area estimates pertain to variability over the total simulation region.

Sampling*FX* was used to perform a similar analysis for arsenic at the upper threshold concentration and for both threshold concentrations for chromium and cadmium. A total of eight maps similar to Figures 14 and 15 were prepared. In addition, a variance map was produced for each contaminant. Through comparison of all of the Sampling*FX* results with the baseline geostatistical analysis and with the baseline kriging analysis, it has been demonstrated that Sampling*FX* provided reasonable and accurate characterization of this site and provided accurate estimates of the area of contamination for this problem at the 50% probability level. Area estimates at the 10 and 90% probability levels cannot be accurately evaluated, primarily because of the different approach used by Sampling*FX* to estimate areas. Discrepancies between the probability maps and area estimates further complicated any attempts at comparison.

Decision*FX* also used Sampling*FX* to estimate exposure concentrations for human health risk assessment. For the industrial exposure scenario, all measured soil concentrations for each constituent were used to estimate the 95th percentile upper confidence limit using Eq. (1):

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

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. Equation (1) provides the variation in the mean of the entire data, and the C_{95} value obtained from Eq. (1) can be interpreted as the 95th percentile upper confidence limit that the mean will be less than C_{95}. The use of the variation in the mean may be appropriate for a worker who travels over the entire site.

The estimates for C_{95} generated by Sampling*FX* for the three contaminants are shown in Table 8. The table also contains the mean value, the standard deviation generated from the 524 samples, and the 95th percentile concentration as evaluated by the demonstration technical team. Comparing the two C_{95} estimates, it is clear that the values calculated by Sampling*FX* match for cadmium and chromium but are low for arsenic.

Decision*FX* used Sampling*FX* to provide an estimate of exposure concentrations for a residential scenario covering a 200 H 100 ft area at the location of the maximum contamination on the site. In this case, estimated concentrations were 500, 600, and 350 mg/kg for As, Cd, and Cr, respectively. Exact details of the computation, such as the location used to estimate the concentrations, were not supplied. However, the results are questionable. First, in the data supplied to Sampling*FX*, arsenic exists at much higher concentrations than cadmium, as shown by their respective means and standard deviations. In fact, arsenic concentrations are higher than cadmium at 494 out of the 524 sampling locations. Further-more, at the 30 sample locations where cadmium concentrations exceeded the arsenic value, only one had a value greater than 600. So these points were clearly not the same ones used to arrive at an average cadmium value of 600 mg/kg over the 200 H 100 ft area. Further, visual inspection of the contour map (Figure 16) shows a 200 H 100 ft region (easting 30,800–31,000, northing 21,900–22,000) where the concentration of arsenic is at least 500 mg/kg.

The technical team examined the data from the region of high arsenic concentration to determine the 95th upper confidence limit concentration (C_{95}) using Eq. (1). The 200 H 100 ft area selected was located at easting 30,822–31,022 and northing 21,914–22,014. In this region, there were 27 data points. Table 9 presents the technical team's estimates for the mean, the standard deviation, and C_{95}. These estimates are clearly much higher than those obtained by Decision*FX*.

The Decision*FX* analyst performed a risk assessment using the exposure concentrations obtained by Sampling*FX*. However, the analyst had to make all of the decisions pertaining to selection of parameters and calculation of risk. This feature is not part of Sampling*FX*; thus, the risk calculations are not evaluated.

Table 8. Comparison of Sampling*FX* and baseline estimates for the 95th percentile exposure concentrations (mg/kg) for the Site N worker risk evaluation

Contaminant	Sampling*FX* C_{95} estimate	Mean value	Standard deviation	Technical team C_{95} estimate
As	222.6	221.9	522.1	265.4
Cd	168.9	142.4	309.4	168.9
Cr	126.3	104.4	255.7	126.3

Table 9. Comparison of Sampling*FX* and baseline estimates for the 95th percentile exposure concentrations (mg/kg) for the Site N residential risk evaluation

Contaminant	Sampling*FX* C_{95} estimate	Mean value	Standard deviation	Technical team C_{95} estimate
As	500	1588	1547	2154
Cd	600	919	896	1247
Cr	350	820	692	1073

Site T Sample Optimization Problem

Originally, the analyst was provided soil data for four contaminants (CTC, EDB, DBCP, and DCP) at 32 sample locations. Site maps containing building and fence locations were provided to assist in the analysis. Decision*FX* used Sampling*FX* to develop a sample optimization scheme to define the extent of soil contamination throughout the site for four of the contaminants. The analyst proceeded through four rounds of sampling, requesting data at 64 additional sample locations. For each contaminant, the analyst used the 96 data points to generate maps of average concentration, the variance in concentration, and the probability of exceeding the threshold concentrations shown in Table 5. The technical team examined the entire set of maps generated by Decision*FX* as part of the evaluation process. This report presents the Sampling*FX* results for EDB contamination at the 21-µg/kg threshold concentration.

Figure 17 is a screen capture from Sampling*FX* showing the final 96 sample locations marked with circles and the average EDB concentration based on the data. In this figure, regions in blue are below the EDB threshold concentration of 21µg/kg; all other regions are above the threshold. The computational blocks, visible in the diagram, represent an area of approximately 50 ft^2. These blocks are the minimum area in the computational evaluation and are consistent with the test problem description, which requested that analysts locate the contamination within a resolution of a 50-ft square. Examination of Figure 17 indicates that there are seven blocks above the threshold concentration, with two or three others that are close to the threshold concentration.

Figure 17 exemplifies the problems the demonstration technical team had performing a technical evaluation of the Sampling*FX* results. The concentration map does not contain a site map or spatial coordinates. There is no visible frame of reference for evaluating the location of the contamination. For this reason, the technical team was unable to perform a detailed quantitative analysis of the Sampling*FX* output and performed only a visual inspection of the outputs.

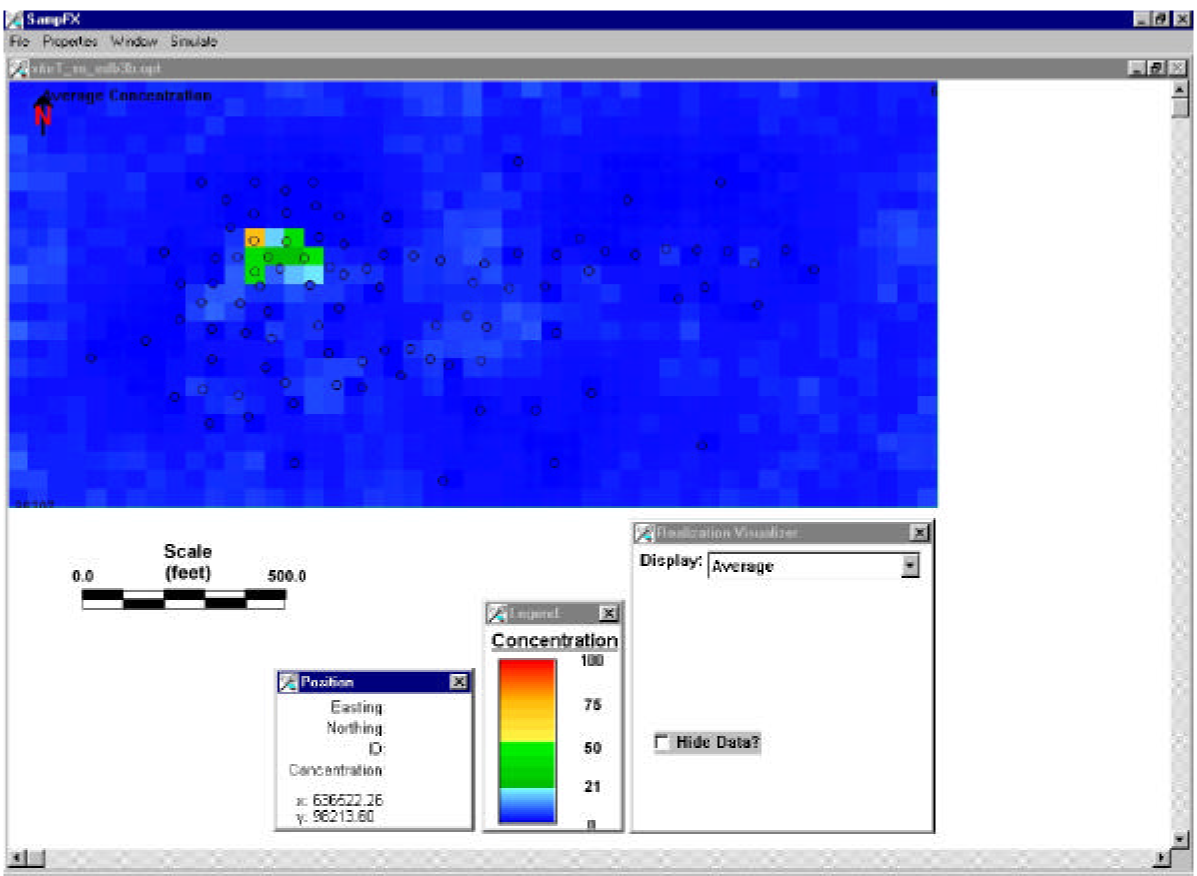

Figure 17. Sampling*FX* map of the average EDB concentration at Site T for the surface soil sample optimization problem.

Figure 18 is the Sampling*FX* map indicating the probability of exceeding the 21-µg/kg EDB threshold concentration. The base map of the site that is visible beneath the probability map provides some frame of reference for the contamination locations. Examination of Figure 18 in terms of the probability of exceeding the threshold indicates that there are five blocks (colored red) with >90% probability, one block (colored yellow) with >50% probability, three blocks (colored green) that have between 25 and 50% probability, and one block (colored light blue) that has a 10–25% probability. (Note that the white areas were not defined by the analyst.) This is consistent with the information provided in Figure 17.

Figure 19 presents the technical team's analysis of the Site T sample optimization problem using the entire data set. The results were generated using the Surfer software package and kriging for data interpolation. Figure 19 contains the site map, concentration contours for EDB at 21 µg/kg (blue) and 500 (red) µg/kg, and the sample locations (circles). Comparison with the Sampling*FX* results in Figures 17 and 18 indicates a reasonable match between the Sampling*FX* analysis with limited data (96 sample locations) and the baseline analysis with the complete data set (273 sample locations). Sampling*FX* accurately defined the contamination zone in the northeast corner of the site. However, the Sampling*FX* analysis missed the small zone of contamination approximately 300 ft to the west of the main area of contamination. This zone is smaller than the 50-ft spacing stated in the test problem, and therefore, it is reasonable to expect that the sample optimization process would miss this contamination.

As part of the test problem, the analyst was asked to calculate the soil surface area that had contamination levels greater than the threshold concentrations in Table 5. Table 10 presents the Sampling*FX* and the

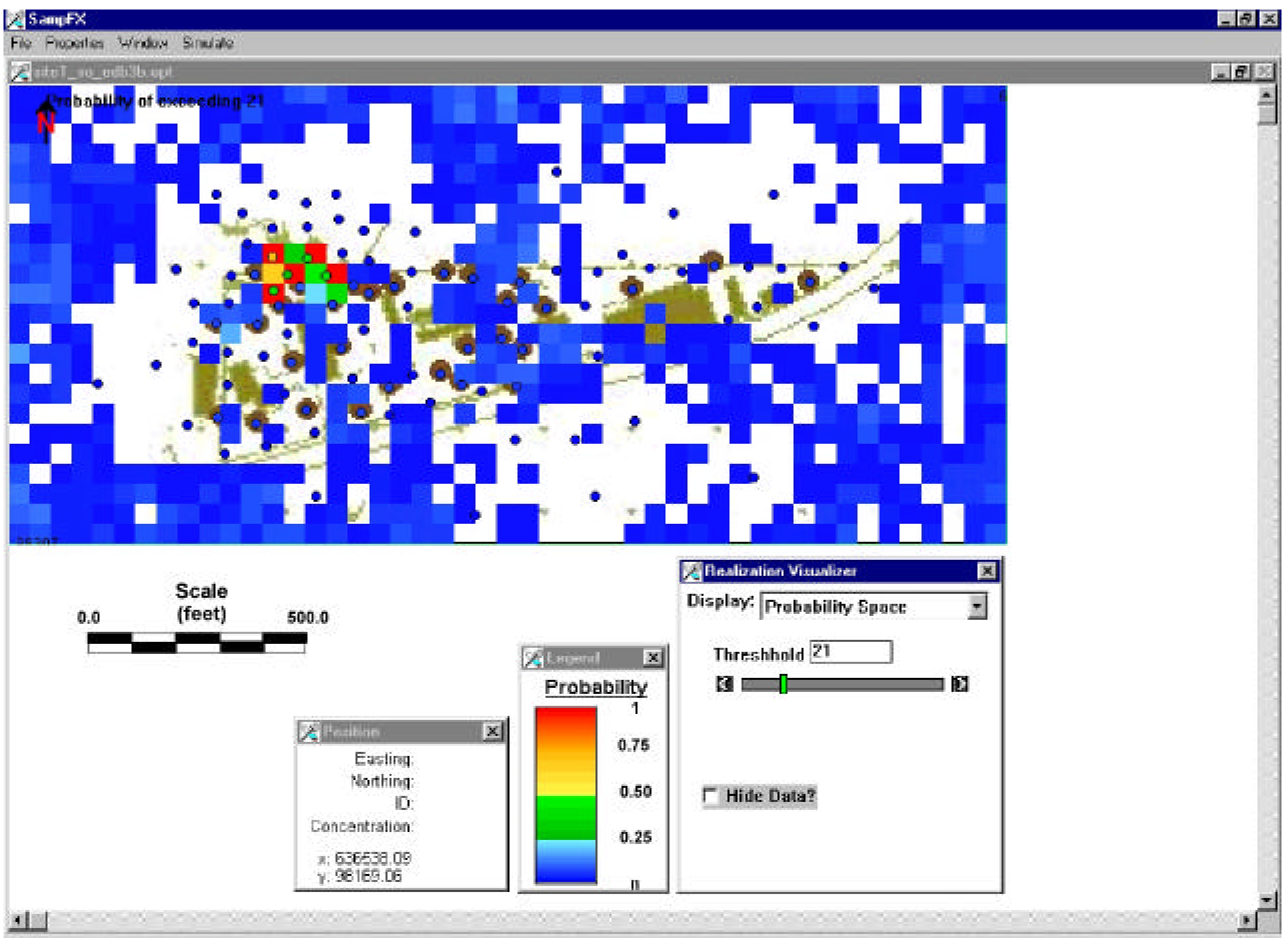

Figure 18. Sampling*FX* map of the probability for exceeding the EDB 21-: g/kg threshold at Site T for the surface soil sample optimization problem.

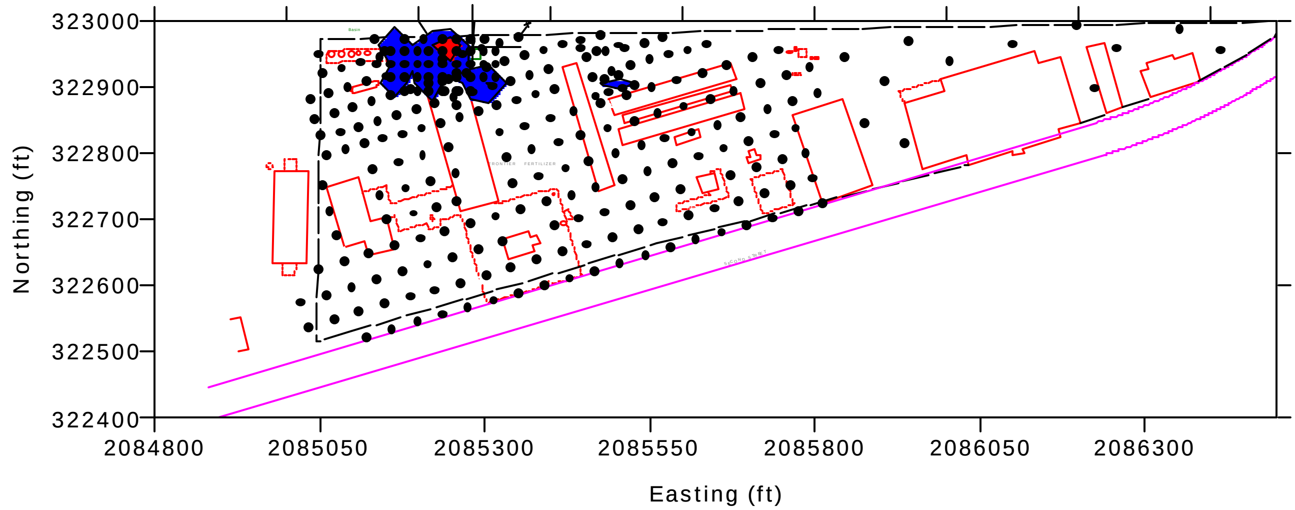

Figure 19. Baseline analysis concentration contour map of EDB contamination at 21 µg/kg (blue) and 500 µg/kg (red) for the Site T surface soil sample optimization problem. The analysis was performed using Surfer and the complete data set.

Table 10. Sampling*FX* estimates of the area of contamination at three probability levels and baseline area estimates for the Site T sample optimization problem

Constituent	Threshold concentration (mg/kg)	Area of contamination (ft^2)			
		90% probability level	50% probability level	10% probability level	Baseline kriging with Surfer
CTC	5	80,400	101,400	120,000	71,500
DCP	500	2,386	2,386	4,772	1,000
DBCP	50	2,386	4,765	7,157	8,950
EDB	21	19,100	27,400	38,200	14,200

baseline analysis area estimates. To obtain the area estimates and the probability levels, Sampling*FX* performs multiple simulations consistent with the data. In each simulation, the software calculates the area that exceeds the mean threshold concentration. Sampling*FX* uses the area estimates from each simulation to calculate the probability levels. The software supplied the resulting area estimates in units of square meters because it requires all measurements to be in meters. The values were converted to square feet by the authors of this report. The baseline analysis was conducted with the complete data set, using Surfer and kriging to interpolate the data. There are substantial differences between estimates, and it appears that the Sampling*FX* estimates are inconsistent with its own concentration and probability maps for EDB (Figures 17 and 18). For EDB, the technical team estimated an area of 14,200 ft^2; the Sampling*FX* estimate at the 50% probability level was 27,400 ft^2. The Sampling*FX* estimate is larger than the baseline analysis by a factor of 2 and corresponds to 11½ of the 50-ft^2 computational blocks used in the Sampling*FX* analysis. This estimate is also apparently inconsistent with the information supplied in the probability map (Figure 18). As noted above, Figure 18 has six blocks (red and yellow regions in the figure) above the 50% probability level. This corresponds to an area of 15,000 ft^2 and is a reasonable match with the baseline analysis. Similarly, the Figure 18 area above the 90% probability level (red zone) is 12,500 ft^2, and the area above the 10% probability level (red, yellow, green, and light blue regions) is 25,000 ft^2. These values are quite different from those reported by Decision*FX*. The cause for the discrepancies between the area estimates obtained from visual inspection of Figure 18 and those supplied by Decision*FX* are not known. The area estimates generated by visual inspection of Sampling*FX* probability plots are consistent with the baseline analysis.

The CTC analysis was similarly reviewed and the comparison between the baseline analysis (273 data points) and the SamplingFX analysis (96 data points) is reported in Table 10. Again, it was concluded that the CTC concentration contours and probability of exceedence maps are reasonably consistent with the baseline analysis and the data. However, the area estimates are larger than the baseline analysis and were inconsistent with the other Sampling*FX* output maps. The Sampling*FX* probability map appears to indicate that the 90% probability level area is 35,000 ft^2 (14 blocks). The 50% probability level estimated from the map is 47,500 ft^2 (19 blocks), as compared to the baseline estimate 71,500 ft^2 obtained using the complete data set. An accurate estimate of the 10% probability level could not be obtained from information supplied by Decision*FX*; however, it was greater than 100,000 ft^2. To ensure that the difference in area estimates was not due to the different data sets, the technical team used the same data set (96 samples) supplied to Decision*FX* after completion of their sample optimization. In this case, the technical team's estimate of area increased slightly to 75,800 ft^2. This corresponds closely with the estimate based on the complete data set (273 samples). In addition, visual comparison of the regions of contamination based on the Decision*FX* sample optimization data set and the complete data set matched closely.

The Sampling*FX* area estimate for DCP above the 500-:g/kg threshold at the 50% probability level was 2383 ft^2. This corresponds to one region (approximately 50 ft^2) and is the minimum nonzero area that can be produced by the analysis. The technical team's value of 1000 ft^2 is therefore consistent with the Sampling*FX* estimate for DCP. The Sampling*FX* area estimate for DBCP above the 50-:g/kg threshold at the 50% probability level was 4765 ft^2. This is approximately half of the technical team's baseline area estimate.

Multiple Lines of Reasoning

Decision*FX* used Sampling*FX* to provide a number of different approaches to examine the data. The foundation of the Decision*FX* approach is a Monte Carlo simulator that produces multiple simulations of the existing data that are consistent with the known data. From these simulations, concentration maps, variance maps, and probability maps were produced to assist in data evaluation. This permits the decision maker to evaluate future actions such as sample location or cleanup guidance based on the level of confidence placed in the analysis.

Secondary Evaluation Criteria

Ease of Use

During the demonstration it was observed that Sampling*FX* is not user-friendly. However, the graphical user interface (GUI) was easy to use. The GUI provided a platform to address problems efficiently and to tailor data formatting to the problem under study.

Sampling*FX* has (or lacks) several features that 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 coordinate systems are typically measured in feet), and the need to have all graphic files imported as a single bitmap. The graphic files limitation prohibits the use of multiple layers in visualizations and requires that coordinates of the bitmap be provided when it is used as a base map for contamination data. In addition, graphic bitmap files cannot be edited, and the software does not have an on-line help feature. Visualization output is limited to screen captures that can be imported into other software for processing. Visualization output was often supplied without a frame of reference (coordinate scale or site map), and this makes data interpretation more difficult. While each of these limitations can be overcome and the analysis performed, it requires more work on the part of the software operator (e.g., a data file could be reformatted in a spreadsheet and coordinates in feet changed to meters to match the needs of Sampling*FX*).

Sampling*FX* exhibited the capability to export text and graphics to standard word processing software directly. Screen captures from Sampling*FX* were imported into CorelDraw to generate .jpg and .cdr graphic files that can be read by a large number of software products. Sampling*FX* generated data files from statistical analysis and concentration estimates in ASCII format that can be read by many software products.

Efficiency and Range of Applicability
Sampling*FX* was used to complete four sample optimization/cost-benefit problems with 12 person-days of effort. This included 4 days for analysis, 5 days for postprocessing of the data to perform cost-benefit analysis and add legends and scales to the maps, and 3 days for preparing the report. This was slightly longer than the technical team would have anticipated and was due primarily to the extensive postprocessing of maps and data required for the analysis. In addition, a newly installed version of Windows 98 created hardware problems for the analyst.

Sampling*FX* provides the flexibility to address problems efficiently and can be tailored to the problem under study. The user has control over the choice of the parameters that control the geostatistical simulations performed by Sampling*FX*. In addition, the software allows evaluation of a wide range of environmental conditions (e.g., contaminants in different media: groundwater or soil). Sampling*FX* should be applicable to almost any soil contamination problem. Its usefulness in 3-D groundwater contamination problems is not clear. Theoretically, one should be able to use the model for this type of problem. However, the results provided on the Site A 3-D test problem were not consistent with the data.

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

Additional Information about the SamplingFX Software
To use Sampling*FX* efficiently, the operator should be knowledgeable in the use of statistics and geostatistics to analyze environmental contamination problems. In addition, knowledge about managing database files, contouring environmental data sets, and analyzing sample optimization and cost-benefit problems is beneficial.

During the demonstration, Sampling*FX* was run on a Windows 95 operating 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, a Macintosh machine was brought to demonstrate that the software worked on this platform. Training demonstrations were performed on the Macintosh machine, but it was not used explicitly for the demonstration problem sets.

Decision*FX* plans to sell Sampling*FX* for $500 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 Sampling*FX* is presented in Table 11. The technical team concluded that the main strength of Sampling*FX* is its technical approach to solving the sample optimization problem. The use of the multiple simulations of the data to generate probability and concentration maps provides a technically robust framework for conducting sample optimization problems. For the two soil contamination problems, Sites N and T, the sample optimization procedure defined the contaminated region with far fewer samples than collected during the original site characterization sampling activities.

The technical team found that there were several limitations in the application of Sampling*FX* to environmental contamination problems. Sampling*FX* was unable to produce an adequate match to the data for the Site A 3-D sample optimization problem; was unable to match exposure concentrations for risk calculations for the Site N cost-benefit residential scenario; used a nonstandard approach for estimating the probabilities of a given area of contamination; and produced area estimates that were not consistent with its own probability and concentration maps for Sites N and T. 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.

Table 11. Sampling*FX* performance summary

Feature/parameter	**Performance summary**
Decision support	Sampling*FX* integrated data and site maps into 2-D spatial representations. Sampling*FX* is a geostatistics-based software designed to address sample optimization problems by predicting sample locations. It is also designed to generate cost-benefit information (e.g., evaluation of the probability of exceeding threshold concentrations) that was exported into Excel to generate cost-benefit curves that were a function of probability of exceedence. Sampling*FX* can also estimate exposure concentrations at receptor locations for health risk analysis. Maps of the contamination and the probability of exceeding a specified contamination concentration were generated. The statistical data interpretations permit the decision-maker to evaluate future actions such as sample location or cleanup guidance based on probability.
Documentation of analysis	A detailed report documented the technical approach, assumptions, and parameters used in the analysis.
Comparison with baseline analysis and data	Sample optimization procedures for the Site A groundwater contamination problem selected a sampling network, but the contaminant concentration and probability maps were not consistent with the data. Sample optimization procedures for the Site N and T soil contamination problems were able to place sample locations accurately and estimate contamination contours and generate probability maps consistent with the data. Site N cost-benefit analysis of the area above threshold concentrations was consistent with the baseline and geostatistical analysis at the 50% probability level, but markedly different at other probability levels. This was due to the technical approach used in Sampling*FX,* which does not conform to EPA DQO guidance. Estimates of soil contamination exposure concentrations for the residential risk calculations were incorrect and too low as compared to the data and baseline analysis. Site T concentration contours and probability maps generated by Sampling*FX* were consistent with the baseline analysis and data, but the cost-benefit analysis of the area above the threshold concentration was inconsistent with the baseline analysis and with Sampling*FX*–generated probability maps.
Multiple lines of reasoning	Sampling*FX* provides a number of different approaches to examine the data as well as multiple simulations to assist in quantifying uncertainties. These include concentration maps, variance maps, and probability maps that were produced to assist in data evaluation.
Ease of Use	Sampling*FX* is not user friendly for the following reasons: • Visualization output is limited to screen captures. • The software can only import bitmaps for use in visualization. • Map cannot be annotated and modified (e.g., to add scales); this must be performed in auxiliary software. • Data from statistical simulations cannot be processed; this task must be handled in auxiliary software. • Concentration data must follow a fixed format; units of measurement must be in meters. • On-line help not available.
Efficiency	Four problems completed and documented with 12 person-days of effort.
Range of applicability	Sampling*FX* is designed to handle any form of spatially correlated data. Therefore, it can handle contamination in soils and groundwater. The applicability to 3-D contamination problems was attempted but not demonstrated.
Training and technical support	Users manual 1½-day training course planned Technical support through e-mail Tutorial examples are not provided with the software
Operator Skill Base	Detailed understanding of statistical and geostatistical analysis procedures for contamination problems. Knowledge of cost-benefit analysis procedures would be beneficial.
Platform	Windows 95 demonstrated; Macintosh product available
Cost	$500 for a single license; free to state and federal regulators

Section 5 — Sampling*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.

Sampling*FX* Update

Decision*FX* is in the process of upgrading the Sampling*FX* DSS from Version 1.0 to Version 2.0. Most of the improvements in the software are a result of lessons learned in the demonstration and comments supplied in the verification report.

Representative Applications

The analysis of a lead-contaminated site at Sandia National Laboratories is an example of the type of analysis that can be performed with Sampling*FX*. The site is a 5-acre firing and testing facility. A conventional EPA-style sampling approach was applied at the site using a star and grid pattern for the sampling network design. Two soil concentration thresholds were considered, representing residential and industrial land use exposure scenarios. Figure 20 shows the probability distribution of exceeding the two threshold limits for the two land use scenarios. Areas in red have a high probability of exceeding the threshold, while areas in blue have a low probability. The cleanup volumes are markedly different for each land use scenario, and the uncertainties are different as well.

A cost analysis for cleanup of the site resulted in the estimates shown in Figure 21. The uncertainty in the residential cleanup is about 20% of the total cost and is fairly significant (±$500K) in terms of budget. The range in costs was determined by selecting different confidence levels for cleanup.

In addition to quantifying the uncertainty in the cleanup volume, Decision*FX* used the operations research methods in Sampling*FX* to optimize the sampling network design. The logic here is that for

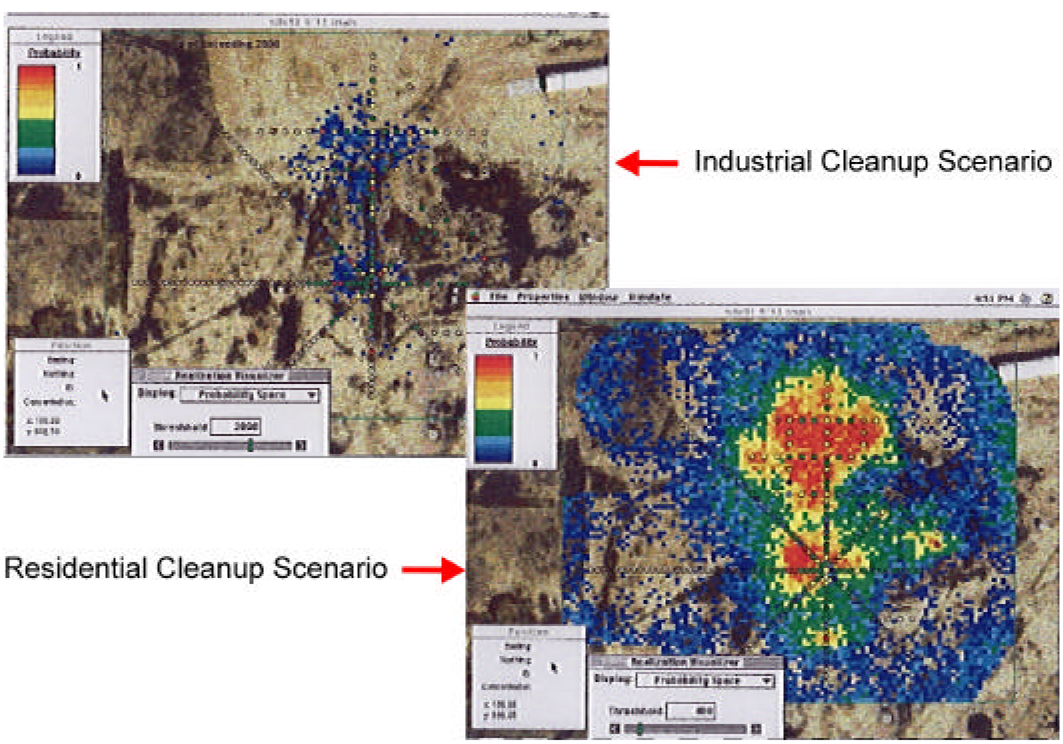

Figure 20. Site cleanup maps for industrial and residential standards.

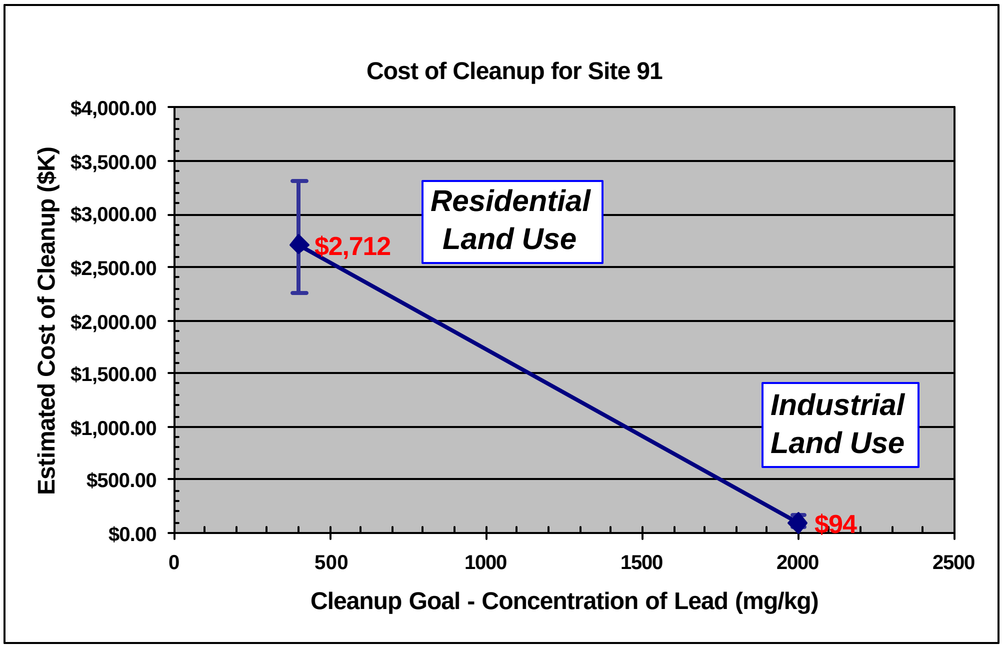

Figure 21. Cleanup costs as a function of threshold concentration. The range of costs reflects different confidence levels in meeting cleanup goals.

each sampling event an analysis is performed, uncertainties are estimated, and new sampling locations are chosen to efficiently reduce uncertainties. After a certain point additional samples do little to refine the definition of the nature and extent of contamination. In this case the plume statistics were stable after five rounds of sampling, with a total of 65 samples collected (as opposed to the 350 samples collected with the conventional EPA baseline approach). Figure 22 shows a probability plot representing the uncertainty in the cleanup area for the residential cleanup scenario for each round of sampling.

The baseline approach at this site initially used a star pattern for sample network design, followed by a grid-sampling pattern in an area of elevated concentrations. Another traditional EPA approach that can be contrasted with these methods is a straight grid sampling method. Use of the design criteria from EPA's Data Quality Objectives (DQO) guidance (EPA 1994) to estimate the number of samples in a uniform grid yields an estimate of about 650 samples to cover the site adequately. A geostatistical analysis of the data from a grid sampling approach, employing 650 samples throughout the site, yields area estimates that are significantly less than either the baseline approach or the optimal sampling approach using the Sampling*FX* operations research methodology. The cost estimate for a residential cleanup scenario using the EPA uniform grid analysis is on the order of $1.7M, ±$180K. This is less than the baseline and Sampling*FX* estimates of $2.7M, ±$500K, because of the suboptimal sampling strategy. If the uniform grid sampling method were used on this site, it is likely that the cleanup volume would be underestimated and that confirmatory sampling would have shown the deficiency. With the uniform grid sampling approach, the final cost of cleanup would probably be greater than the cost resulting from baseline method because of the need for a second round of mobilization for the cleanup work.

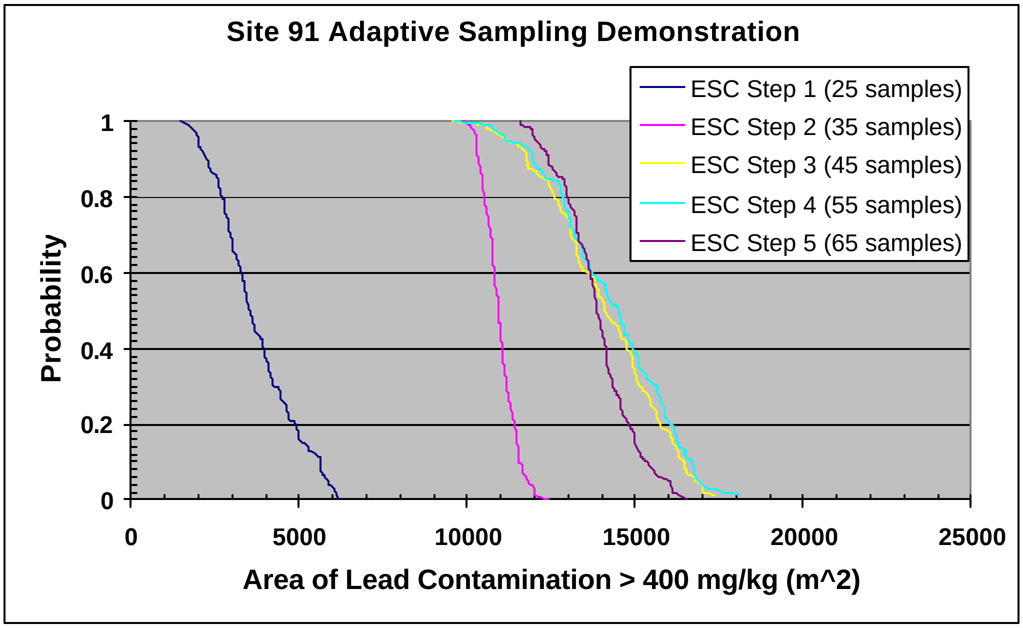

Figure 22. Cleanup costs as a function of number of samples collected. The range of estimated areas reflects different probability levels in meeting cleanup goals.

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).

Table A-1. Site T soil contamination threshold concentrations

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

Site T: Cost-Benefit Problem

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

Appendix B — Description of Interpolation Methods

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

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

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

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

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

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

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

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

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

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