THE ENVIRONMENTAL TECHNOLOGY VERIFICATION PROGRAM

U.S. Environmental Protection Agency

Oak Ridge National Laboratory

ETV Joint Verification Statement

TECHNOLOGY TYPE:	**ENVIRONMENTAL DECISION SUPPORT SOFTWARE**
APPLICATION:	**INTEGRATION, VISUALIZATION, SAMPLE OPTIMIZATION, AND COST-BENEFIT ANALYSIS OF ENVIRONMENTAL DATA SETS**
TECHNOLOGY NAME:	**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.