THE ENVIRONMENTAL TECHNOLOGY VERIFICATION PROGRAM

ETV

EPA
U.S. Environmental Protection Agency

ornl
Oak Ridge National Laboratory

ETV Joint Verification Statement

TECHNOLOGY TYPE:	**ENVIRONMENTAL DECISION SUPPORT SOFTWARE**
APPLICATION:	**INTEGRATION, VISUALIZATION, SAMPLE OPTIMIZATION, COST-BENEFIT, AND RISK ANALYSIS OF ENVIRONMENTAL DATA SETS**
TECHNOLOGY NAME:	**Spatial Analysis and Decision Assistance (SADA)**
COMPANY:	**University of Tennessee Research Corporation (UTRC) 1534 White Avenue, Suite 403 Knoxville, TN 37996-1527**
PHONE:	**(865) 241-5741**
WEBSITE:	**www.sis.utk.edu/cis/SADA**

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 the University of Tennessee Research Corporation's (UTRC's) Spatial Analysis and Decision Assistance (SADA)™ 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 remediation 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 the 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 University of Tennessee Research Corporation (UTRC) demonstrated Spatial Analysis and Decision Assistance (SADA) by performing visualization, sample optimization, and cost-benefit analysis for Sites N and S. Site N had two separate problems, and both were evaluated using SADA. The Site N problems were two-dimensional (2-D) soil contamination problems for three heavy metals (arsenic, chromium, and cadmium). In the Site N sample optimization 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 contained 524 data points on a 14-acre region of the site and required the analyst to perform a cost-benefit analysis of the remediation costs vs cleanup goal for each of the three contaminants. In addition, the analyst was asked to estimate the human health risks based on current conditions. The Site S test problem was a three-dimensional (3-D) groundwater contamination cost-benefit problem for a single contaminant (chlordane). The analyst was provided with a series of wells containing chlordane concentrations as a function of depth. The analyst was asked to define the region, mass, and volume of the plume at contaminant threshold concentrations of 5 and 500 μg/L. Based on this information and groundwater flow rates, estimates of current and future human health risks were requested.

SADA was used to integrate large quantities of data into a visual framework for assistance in understanding a site's contamination problem. For the Site N sample optimization problem, the data were used to develop a sampling scheme to characterize the site. Upon completion of the data-collection phase of the problem, maps with the probability of exceeding threshold concentrations were provided. For the Site N cost-benefit problem, SADA was used to estimate the cost of cleanup versus the cleanup threshold. Human health risks were evaluated on the basis of current conditions. For the Site S problem, SADA was used to estimate the volume of contamination above threshold levels and human health risks based on current conditions.

Details of the demonstration, including an evaluation of the software's performance, may be found in the report entitled *Environmental Technology Verification Report: University of Tennessee Research Corporation, Spatial Analysis and Decision Assistance (SADA)*, EPA/600/R-00/036.

TECHNOLOGY DESCRIPTION

SADA is an environmental software product that incorporates tools from various fields — including visualization, geospatial analysis, statistical analysis, human health risk assessment, cost-benefit analysis, sampling design, and decision analysis — into a dynamic and interactive environment. Each of these modules can be used independently or in an integrated fashion to address site-specific concerns in the characterization and remedial action design. SADA was designed to simplify and streamline several of the processes in environmental characterization, risk assessment, and cost-benefit analysis to bring the information together in a way that can help users make decisions about their particular site in a quick and cost-effective manner. SADA is designed to assist environmental professionals who need to examine the data within a spatial context. SADA runs on Windows 95, 98, and NT platforms.

VERIFICATION OF PERFORMANCE

The following performance characteristics of SADA were observed:

Decision Support: SADA was designed as a decision support tool and directly addresses environmental questions such as (1) the location and size of the area of contamination, (2) the size of the cleanup zone at a specified contaminant threshold concentration or risk level, (3) the confidence in predicting the area of contamination or cleanup zone, (4) the costs for remediating the cleanup zone, (5) the human health risks, and (6) the optimal location for the next set of samples to best define the extent of contamination. In the demonstration, UTRC was able to use SADA to quickly import data on contaminant concentrations, overlay site maps, and integrate this information on a single platform. SADA demonstrated the ability to place the information in a visual context and produced 2-D and 3-D maps that support data interpretation and decision making. SADA was used in the demonstration to automatically generate maps showing contaminant concentration, recommended cleanup zones, cost-benefit curves, and human health risk. These maps can be based on the probability of exceeding specified contaminant threshold concentrations or risk levels and at specified probability levels. SADA was also used to predict new sample locations based on statistical and/or geostatistical analyses of the existing data.

Documentation of the SADA Analysis: UTRC staff used SADA to generate reports that provided an adequate explanation of the process and parameters used to analyze each problem. Documentation of data transfer, manipulations of the data (e.g., how to treat contamination data as a function of depth in a well), and analyses were included. Model selection and parameters for statistical analysis and contouring were also provided in the exportable documentation.

Comparison with Baseline Analysis and Data: SADA was able to generate 2-D and 3-D maps of contaminant concentrations, human health risk, probability of exceeding contaminant threshold concentrations as a function of degree of probability, and remedial zone maps for specified contaminant thresholds and probability levels. The maps included posting of data at the sample location, color coding of sample points to represent a parameter (concentration or risk), contaminant concentration contours, human health point risks, and human health risk contours. SADA also generated cost-benefit curves for the cost of remediation vs the cleanup threshold. These curves could be calculated for varying degrees of probability in the data. For the Site N sample optimization problem, the SADA analysis generated an acceptable match to the data and the baseline analysis. When compared with the baseline geostatistical analysis that used the entire data set, SADA identified approximately 75% 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 the threshold concentrations matched, to within 21%, the baseline interpolation and geostatistical analyses at the 50% probability levels. Likewise, the area estimates at the 90% probability level were within 21% of the baseline analyses and geostatistical analysis. The slight differences between SADA and the baseline analysis were due to the different parameters used for interpolation. For the Site S cost-benefit test problem, at the 50% probability level there is good agreement between SADA, the baseline analysis using Surfer™, and the baseline geostatistical analysis. In fact, all three area estimates are within 13% of each other, indicating agreement. For the 10% probability level, the SADA area estimates are 6% less at the 5-µg/L threshold and 9% less at 500-µg/L than the baseline area estimates. The difference between the SADA results and the baseline analysis is due to the slightly different selection of boundaries of contamination and kriging parameters selected for the analyses. Overall, there is close agreement among the area estimates produced by SADA and the baseline geostatistical analysis. Both the noncarcinogenic and carcinogenic risks calculated by SADA for Site N and Site S were accurate and consistent with the baseline analysis and EPA's risk assessment guidance for Superfund for all of the test problems.

Multiple Lines of Reasoning: UTRC staff conducted multiple data explorations and evaluations that were supported by the statistical and geostatistical functions in SADA. This information provided a quantitative measure of the confidence that could be placed in the decision. Several data interpolation routines were considered on a problem-specific basis before UTRC staff selected the best one for data analysis. Several sample optimization schemes are available for use. Selection of a particular scheme depends on the objectives of the analysis and the amount of data.

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

Ease of Use: The demonstration showed that SADA was easy to use. The SADA graphical user interface has a logical structure to facilitate use of the options in the software package. SADA accepts database files in comma-delimited format; however, database files were supplied in .dbf format. The analyst imported the .dbf files into another software program (Microsoft Excel) and converted them into comma-delimited files. Drawing and map files could be read in .dxf format. Other common image file formats such as .jpg and .bmp were not supported. Visualization results can be output to any other Windows application that supports the use of the Clipboard, including commonly available software (e.g., Microsoft PowerPoint, Word, and WordPerfect).

Efficiency and Range of Applicability: SADA relies on a flexible database format with user-defined inputs. This provides a flexible platform that addresses problems efficiently and is tailored to the problem under study. The database permits filtering on the contaminant identifier and location. SADA has an auxiliary database that contains contaminants identified by name and Chemical Abstract Service (CAS) number. This feature facilitates data checking. SADA also has databases containing toxicological and exposure scenario parameters. These databases facilitate human health risk assessment. The software provides analysis on spatially correlated data and can simulate a wide range of environmental media and conditions (e.g., contaminant in groundwater, soil, sediment, or surface water; multiple contaminants on a single site) to be evaluated.

Training and Technical Support: SADA requires training for efficient and proper use. An analyst with a background in environmental problems and a basic knowledge of database operations, human health risk assessment, and statistics/geostatistics can be using SADA after one or two days of training. A detailed on-line help system is supplied with the software package. The on-line help provides examples of how to conduct analysis and gives recommendations on approaches to statistical/geostatistical modeling. Examples of software applications are provided as part of the software packages. A two-day training course is available. Technical support is available through e-mail.

Operator Skill Base: Effective use of all of the features of SADA requires that the operator possess a thorough understanding of the use of geospatial modeling in analyzing environmental problems and human health risk assessment. This includes an understanding of interpolation algorithms and geostatistics along with a fundamental knowledge of database manipulations, sample optimization, and cost-benefit analysis.

Platform: During the demonstration, SADA Beta Version 3.0 was operated on a Windows 95 operating system using a laptop with a 266-MHz Pentium processor, 128 MB of RAM, and 4 MB of video memory.

Cost: SADA will be distributed free over the Internet.

Overall Evaluation: The technical team concluded that the main strength of SADA is its technical approach to assist environmental decision-makers by defining areas of concern based on user-defined contaminant concentrations or human health risks. SADA's use of a geostatistical approach provides an estimate of the degree of uncertainty in the prediction that provides key information to assist in the selection of future sample locations and in determining cost-risk tradeoffs. The incorporation of databases of risk parameters, coupled with the pull-down menus in SADA, make risk calculations easy to perform. The integration of geostatistical analysis, human health risk assessment, cost-benefit analysis, sampling design, and decision analysis into a single software product makes SADA a powerful tool for analyzing spatially correlated data. SADA demonstrated the capability to accurately perform sample optimization analysis, estimate areas and volumes of contamination for cost-benefit analysis, and estimate the probability of exceeding threshold levels in concentration or risk.

The technical team did not notice any major limitations in SADA. Several minor limitations were noted. The 3-D visualizations provided only a qualitative depiction of the plume because a frame of reference (axis scale or surface maps) was not provided. Maps and drawings could be imported only as .dxf files; the capability to import other graphic formats would be beneficial. Finally, data files could be imported only in comma-delimited format, which requires reformatting in another software product.

A credible computer analysis of environmental problems requires good data, reliable and appropriate software, adequate conceptualization of the site, and a technically defensible problem analysis. The results of the demonstration show that the SADA software can be used to generate reliable and useful analyses for evaluating environmental contamination problems. This is the only component of a credible analysis that can be addressed by the software. The results of a SADA analysis can support decision making. Although SADA has been demonstrated to have the capability to produce reliable and useful analyses, improper use of the software can cause the results of the analysis to be misleading or inconsistent with the data. As with any complex environmental DSS product, the quality of the output is directly dependent on the skill of the operator.

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

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

David E. Reichle, Ph.D.
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.