Converting Data to Decisions

ENVIRONMENTAL data aren't easy to obtain, and once obtained, they are often hard to interpret. For example, drilling into the earth to determine what kind of soil exists at any given spot in the substrate is not only expensive but also gives scientists just piecemeal information. Computer analysis with this information can be equally piecemeal. But earth scientists are learning that computer models can be made more meaningful when they are stochastic, meaning that they are based on a certain amount of probability. Now, with the capability of high-performance supercomputers in the National Nuclear Security Administration's Advanced Simulation and Computing (ASCI) program, Livermore scientists are exploring groundbreaking ideas in statistical theory that will help them use stochastic descriptions quantitatively and obtain a much more complete picture of soil composition.

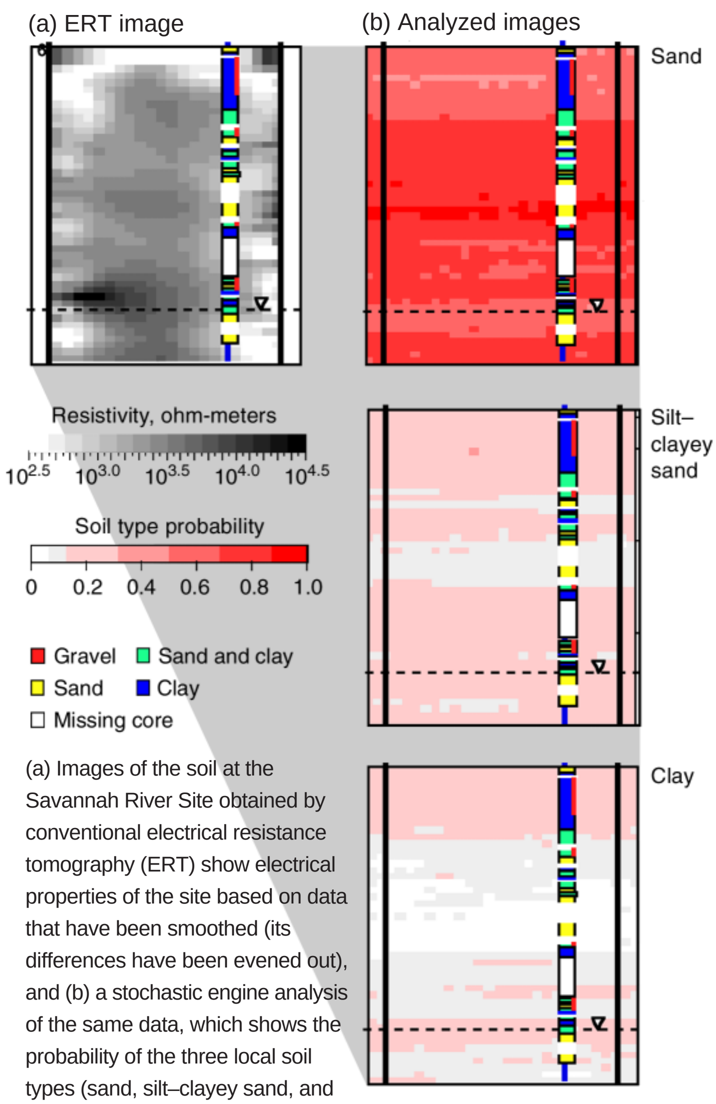
This new technology, called a stochastic engine, is a process that links predictive models, advanced statistical methods, and refined search methods. Using this technology, scientists can incorporate a proposed soil configuration into a computer model and produce a geophysical simulation. The simulated result is compared to actual data. If the result is consistent with observed data, then the simulation is boosted to the next phase of analysis.

The stochastic method is a powerful technique that is now in use. Livermore scientists are consulting on a project with the Westinghouse Savannah River Company in which the stochastic engine will assist in a major cleanup operation at the Savannah River Site in South Carolina. The method could also be applied to problems in stockpile stewardship, atmospheric dispersion, seismic velocities, and intelligence collection.

Cleanup Site Yields New Tool

The stochastic engine concept uses techniques developed at Livermore and was motivated by an innovative steam remediation cleanup being conducted by Southern California Edison at a Superfund site in Visalia, California, in which Laboratory scientists also participated. (See *S&TR*, January/February 1996, pp. 6–15.) During the project, more than 46 million pieces of data were obtained pertaining to the way steam, water, and contaminant flowed through the groundwater plumbing system. These data included temperatures, flow rates, pressures, and electrical resistance tomography (ERT) measurements. ERT, a technology developed at Livermore in

1993 and now available commercially, is similar to a computed tomography scan. It images soil resistivity, and that gives scientists information on soil properties such as temperature, soil type, and saturation. While the data collected from Visalia were rich and invaluable for Edison's operational decisions, the various data types could only be used independently. Observations and simulations could not be linked to provide the kind of cohesive understanding that would dramatically improve site operations and, most importantly, optimize the final outcome of the cleanup work.



(a) Images of the soil at the Savannah River Site obtained by conventional electrical resistance tomography (ERT) show electrical properties of the site based on data that have been smoothed (its differences have been evened out), and (b) a stochastic engine analysis of the same data, which shows the probability of the three local soil types (sand, silt–clayey sand, and clay) at each location.

The work at Visalia, while highly successful overall, is representative of a frustration that Livermore environmental scientists experience whenever they attempt to characterize soil compositions at cleanup sites: how to apply the powerful predictive capabilities of Livermore's supercomputers to complex, real situations. For the past year, Roger Aines and a multidisciplinary team have been discussing how to apply modern computational power and statistical search methods to extract maximum information from sparse initial data and then to improve the analysis on the fly as more data become available.

More Than One Right Answer

The power of the stochastic engine comes from its ability to refine a model by successively narrowing down the possible configurations of a hypothetical model. The refinement is done over progressive layers of data. In this process of model improvement through iteration, the stochastic engine uses an advanced statistical method called a hybrid Markov Chain Monte Carlo (MCMC)–Bayesian analysis.

In the MCMC analysis, a chain (or sequence) of configurations is considered. Each configuration undergoes a probability calculation that compares observed data to corresponding model predictions. If the predictions are acceptable (that is, probable for the configuration), the result of that calculation becomes the basis of the next configuration. This allows the process to rapidly search for good configurations in very complex situations. The Bayesian statistical method, based on the work of English mathematician Thomas Bayes, performs its part in the stochastic engine by comparing the probability calculations with real information to guide the statistical inference process.

Suppose a volume of soil is known to be composed of seven layers that could be either sand or silt, and an ERT measurement of that volume gives a value of 11. The stochastic approach calculates which configurations of silt and sand, and in which positions, give values close to 11. Each case with a value near 11 is passed on to the next stage of analysis. There, the model will continue to restrict possible configurations but base its decisions on other data types, such as water, temperature, or pressure.

For the simple case cited here, it is easy to calculate and compare all the possible configurations, but for a large area, such as the Visalia cleanup site, the possibilities are far too numerous. At Visalia, the MCMC–Bayesian method could help by performing an efficient intelligent search through the collection of possible soil configurations, rapidly identifying the configurations that most closely match all the data.

"It's not about trying to find the single best answer, but all of the good answers," says Aines. "In underground problems, there are usually multiple solutions that are consistent with the data."

The stochastic engine's ability to choose system configurations that are consistent with observed data allows much more tightly constrained (better restricted) answers than conventional methods. Only the ways the system can possibly exist are considered. Using the stochastic technique, for example, researchers can interpret ERT images to derive characteristic soil types for a site, rather than simply provide the electrical properties of the ground. The stochastic engine allows the available information to be used more effectively. It also allows the user to incorporate known constraints, such as the presence of a gravel layer observed in a well, to further guide the statistical inference.

It Doesn't Have to End with Dirt

The stochastic engine method has tremendous potential for use in disciplines that need to combine data and simulation. Currently, the team is working with a number of scientists from other Livermore directorates to identify unknown sources of toxic contaminants in the atmosphere, locate flaws in buildings, evaluate intelligence data, and expand tomography and x-ray imaging data.

The Savannah River Site project illustrates how the engine is being used in industrial partnerships. Livermore has been consulting with Westinghouse's Savannah River Company to clean up organic solvents from the soils and groundwater at the South Carolina site.

Since 1983, the company has been performing environmental cleanup of a site where, over time, solvents became a solvent plume that extended over 5 square kilometers. Now, Westinghouse is ready to present its cleanup results to regulators and assure the community that the remaining plume will not affect surface water bodies. The stochastic engine will be used to evaluate the effectiveness of source cleanup and to predict the ultimate effect of the remaining plume.

Challenges Ahead

Why hasn't the stochastic method been used before? For one thing, the complexity of the method has required robust computer power that simply has not been available until recently. For another, even with the power available on ASCI computers, some scientists are still skeptical of the method. Aines says that because underground problems are so complex, many people are displaying a "show me first" attitude toward the technology. "No one has done this before, so some believe it can't be done." The Savannah River Site project may prove that the engine is a feasible and valuable tool for environmental cleanup and more.

—Laurie Powers

Acknowledgments: The stochastic engine work described in this article was developed by a multidisciplinary team from Livermore's Energy and Environment, Engineering, and Computation directorates, including computational geoscientists John Nitao and Steve Carle; engineering statisticians Bill Hanley, Ron Glaser, and Sailes Sengupta; geophysicists Robin Newmark and Abe Ramirez; and computational scientist Kathy Dyer.

Key Words: Bayesian statistics, electrical resistance tomography, Markov chain, Monte Carlo method, Savannah River Site, stochastic engine, Superfund, Visalia cleanup.

For further information contact Roger Aines (925) 423-7184 (aines1@llnl.gov).

The sewer at Savannah River Site, South Carolina, that has been the source of solvents soaking into the soil. At the far right, geoscientist Steve Carle is shown at the sewer outfall. He was examining soil just below the outfall, looking at the silt layers that tend to control the migration of the solvent.