Conf 9308107--3

Comments on
"Use of Conditional Simulation in Nuclear Waste Site Performance Assessment"
by Carol Gotway

Darryl J. Downing
Mathematical Sciences Section
P. O. Box 2008, Building 6012
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6367

ABSTRACT

This paper discusses Carol Gotway's paper, "The Use of Conditional Simulation in Nuclear Waste Site Performance Assessment." The paper centers on the use of conditional simulation and the use of geostatistical methods to simulate an entire field of values for subsequent use in a complex computer model. The issues of sampling designs for geostatistics, semivariogram estimation and anisotropy, turning bands method for random field generation, and estimation of the cumulative distribution function are brought out.

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

SEP 29 1993
OSTI

Research sponsored by the Applied Mathematical Sciences Program, Office of Energy Research, U. S. Department of Energy, under contract DE-AC05-84OR21400 with the Martin Marietta Energy Systems, Inc.

MASTER
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Comments on
"Use of Conditional Simulation in Nuclear Waste Site Performance Assessment"
by Carol Gotway.

I would like to begin by thanking Carol Gotway for an interesting and informative paper on conditional simulation. I found this paper engaging for many reasons, but the main one is the use of spatial statistics in doing simulations. Before going on to discuss some of the issues in Gotway's paper, I would like to add that I felt the paper was clearly written, had a nice theme and explained the material well enough to keep me interested and informed. This type of introductory paper on a methodology is something I personally find very absorbing and would like to see more of in the statistics journals. Technometrics should be applauded for publishing it.

Gotway's paper begins by explaining the need for doing assessments. The Environmental Protection Agency (EPA) has set up guidelines for doing assessments and regulations that must be met before nuclear waste can be stored. These are very real concerns, and the storage facilities need to establish that they will meet the requirements. For example, EPA 40 CFR 191.16 states in part that the disposal systems should be "... designed to provide reasonable expectation that for 1,000 years after disposal, undisturbed performance of the disposal system shall not cause radionuclide concentrations to exceed specified limits." For most of us 1,000 years is more than a lifetime, and experiments that last this long will probably never be published.

The question is how can we, with the tools we have today, establish with some degree of confidence that a requirement like that above will be achieved? Over the last several years this has been demonstrated through the use of simulation analysis of the system being studied. I am familiar with the study of nuclear power plants where interest is on the determination of the core temperature and the associated probability that it would exceed some critical temperature. These analyses were performed by doing Monte Carlo simulations, running computer codes several times, and examining the distribution of the output of interest. The major difference between the work that Gotway presents and some of the work that was done in the past is that

in examining groundwater travel time (GWTT) a whole field of transmissivity measurements are being collected to use in the analysis, including measured values (this is the conditional part). In doing this one wants to maintain the spatial correlation and uncertainty of the values of the known measurements in the simulation so that this will be incorporated through the model to the output variable of interest (GWTT). Iman and Conover (1983) used what is termed Latin Hypercube Sampling (first mentioned by McKay, Conover, and Beckman (1979)) along with a technique to induce correlation between variables. I mention this to point out that correlation between variables was considered to be important then, as well as now; but, to my knowledge, there was very little concern for spatial relationships. The reliance on computer codes is similar. In Gotway's example one cannot run an experiment for a thousand years, and in the nuclear power plant one is not going to run it into a critical region to see the effect.

Gotway's examination of the transmissivity measurements taken around the Waste Isolation Pilot Plant (WIPP) are cogent and a necessary adjunct to the geostatistical analysis that followed. It is not clear to me what one would do if the test for normality of the log transmissivity values had been rejected. Comments on this would be welcome.

The geostatistical analysis is well presented, and the level of detail is sufficient to understand the essentials of the process without being overwhelmed. The estimation of the empirical directional semivariograms is clearly stated, and the subsequent fitting of the parametric models is also. There are several parametric models that one can choose from (see Cressie(1991)), and the proper selection is partly an art that comes with experience. Sensitivity of the model output to the choice of model for the variogram is an obvious area of research. I found the discussion on anisotropy and the way that it was handled in the paper to be illuminating. Combining the east-west and north-south semivariograms based on geometric anisotropy appears to be a reasonable way to handle the problem. A procedure to test for geometric anisotropy does not seem to be presented, and I wonder if much work has been done regarding this problem. The degree of anisotropy could greatly influence the design for future sampling of a site as mentioned in Journel and Huijbregts (1978). Morris (1991) proposed an index that can be used to compare spatial designs for the purpose of estimating the

semivariogram. His analysis is for the isotropic case, but the point he makes is valid for the anisotropic case also. Anisotropy is a very important problem in spatial statistics and is a topic of much research.

The section on conditional simulation and turning bands was very interesting. The question of how one should go about simulating a stochastic process with a given variance-covariance structure is well known in multivariate analysis. The assertion has been made that, when the covariance matrix takes on a size greater than 1000X1000, problems start to occur in factoring the covariance matrix using traditional numerical methods, for example, when using the Cholesky decomposition. The turning bands method was developed as an alternative to the Cholesky decomposition method that allows for generation of random field realizations using a univariate random number generator with a specific one-dimensional covariance structure. This covariance structure depends on the covariance structure of the random field to be simulated. Most of the literature discusses the case where the covariance is isotropic, but Gotway points out that geometric anisotropy can be accounted for also. The manner in which this is done is probably numerical rather than analytical, and the quality of the approximation would be important in assessing the goodness of the simulation of the 2-dimensional random field. The method of turning bands was developed by Matheron (1973) due to the problems of factorizing large matrices. Today new efficient routines allow the Cholesky decomposition of very large matrices (as large as 40000x40000, see Ng (1993)) with minimal computational time. Thus, although the method of turning bands may have been necessary in the past, it may not be today, in some cases. The idea of conditional simulation, to construct a new realization that passes through the data (or as Gotway puts it - "honors the data"), has a definite appeal. The simple kriged surface using the data alone is much too smooth, and the conditional simulation preserves what is known from the data but adds the uncertainty at other points.

The final portion of the paper discusses the groundwater flow equations and transport of particles through the aquifer. These relationships and particle tracking algorithms allow the computation of GWTT. The important questions are about the distribution of GWTT and whether the regulations will be satisfied with some specified confidence. The problem of fitting

a distribution to the output has been looked at by several authors (e.g., Iman and Conover (1980), Bowman, Serbin, and Shenton (1980), and Downing, Gardner, and Hoffman (1983)). The accuracy and precision of the resulting cumulative distribution function for predicting the true system response is still being investigated. It is a function of the goodness of the input data and how well it portrays reality, the number of computer runs that can be made, the quality of the model (its accuracy and precision), and a list of other factors that affect the output. The ability to separate these issues and determine their effect is important to the analysis of simulation models and their output. Current papers that discuss related issues are Sacks, Welch, Mitchell, and Wynn (1989) and Currin, Mitchell, Morris, and Ylvisaker (1991).

In summary, I would like to mention that work on computer simulation models is extremely important. These are the primary tools we have for investigating many complex problems, for example ground water contamination, molecular dynamics, climate modeling, and nuclear reactor accidents. Many of these models are being investigated at the national laboratories, but unfortunately often without any help from statisticians. The paper by Gotway shows another example of how statistics can be applied to a complicated process to yield practical results. I congratulate her on a well written and informative piece of work.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ADDITIONAL REFERENCES

Bowman, K. O., Serbin, C. A., Shenton, L. R. (1980), "Explicit/Approximate Solutions for S_B," *Communications in Statistics - Simulation and Computation*, **BP(2)**, 127-32.

Cressie, N. (1991), *Statistics for Spatial Data* , Wiley, New York.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991), "Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments," *Journal of the American Statistical Association*, **86**, 953-963.

Downing, D. J., Gardner, R. H., and Hoffman, F. O. (1985), "An Examination of Response-Surface Methodologies for Uncertainty Analysis in Assessment Models," *Technometrics*, **27**, 151-163.

Iman, R. L. and Conover, W. J., (1980), "Small Sample Sensitivity Analysis Techniques for Computer Models with an Application in Risk Assessment," *Communications in Statistics - Theory Methods.*, **A9(17)**, 1749-1842.

Iman, R. L. and Conover, W. J. (1982), "A Distribution Free Approach to Inducing Rank Correlation Among Input Variables," *Communications in Statistics - Simulation and Computation*, **11(3)**, 311-360.

Jounel, A. G. and Huijbregts, C. J. (1978), *Mining Geostatistics*, Academic Press, London.

Matheron, G. (1973), "The Intrinsic Random Functions and Their Applications," *Advances in Applied Probability*, **5**, 439-468.

McKay, M. D., Conover, W. J., and Beckman, R. J. (May 1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, **2(2)**, 239-245.

Morris, M. D. (1991), "On Counting the Number of Data Pairs for Semivariogram Estimation," *Mathematical Geology*, **23**, 929-943.

Ng, E. G. and Peyton, B. W. (1993), "Block Sparse Cholesky Algorithms on Advanced Uniprocessor Computers," *SIAM-Journal of Scientific Computing*, 1034-1056.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments," *Statistical Science*, **4**, 409-423.

END

DATE FILMED

11/10/93