Calibration and Reliability in Groundwater Modelling (Proceedings of the ModelCARE 99 Conference held at Zürich, Switzerland, September 1999). IAHS Publ. no. 265, 2000.

Reliability-based evaluation of groundwater remediation strategies

H. W. REEVES, J. LEE, C. H. DOWDING
Department of Civil Engineering, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3109, USA
e-mail: h-reeves@nwu.edu

A. J. GRAETTINGER
Department of Civil and Environmental Engineering, Mineral Industries Building, Room 260, The University of Alabama, Tuscaloosa, Alabama 35487-0205, USA

Abstract Groundwater flow and contaminant transport modelling are combined with geological characterization data uncertainty in a computationally efficient method to estimate the reliability, or uncertainty, associated with various groundwater remediation designs. The reliability estimate provides information to the project regarding the probability of success of the design based on the current field data, the likelihood that new field data will improve the reliability of the performance estimate, and the location in the field where new data will most benefit the design. This approach recognises that reliability of design performance depends on both the uncertainty associated with the geological model and the impact of this uncertainty on the performance model.

INTRODUCTION

Characterization of a contaminated site to evaluate remediation alternatives requires answering two questions: "Are there enough field data to assess the performance and reliability of various remedial alternatives?" and "Where should the next data be collected to increase confidence in the remedial design?" These questions arise since groundwater flow and transport models typically are employed to simulate the behaviour of a contaminant plume under various remedial strategies, and because uncertainty in site hydrogeology leads to uncertainty in simulation results. The resolution of these questions requires consideration of both the inherent uncertainty associated with quantification of geological conditions and the impact of this uncertainty upon remediation system design.

To address these important questions, we build upon the ideas articulated by Freeze *et al.* (1990) regarding hydrogeological decision analysis and the notion of reliability-based design (Harr, 1987) where the uncertainty of meeting a performance objective for a project is quantified. A quantitative interrelationship between the site exploration process, the geological uncertainty model, the hydrological simulation model, and engineering reliability analysis is sought. The approach adopted to provide this quantification involves: a Bayesian statistical analysis to determine the best estimate for the spatially-correlated input parameters required by a numerical model, a Taylor Series approximation to relate the uncertainty in the dependent variables computed by the numerical model to the input parameters, and a determination of a β index or reliability index to facilitate the analysis of the problem.

APPROACH

Information regarding the geometry of geological units at a site may be classified as either *hard data* or *soft data* (Johnson, 1996). Hard data in the case of contaminant transport in groundwater are direct measurements of hydrogeological parameters at the site. Soft data are estimated values between direct measurement points. These data are obtained through geological interpretation, interpretation of geophysical tests, or experience with similar sites. Due to the typically small number of direct measurements at a site, soft data are essential in the modelling process. Hard and soft data must be combined to construct a geological model of the site. We adopt a Bayesian approach to perform this combination due to its ability to include soft data and to update the spatial covariance relationships for the site as new data are collected. We also note that the Bayesian approach with an uninformed prior produces estimation results that are very closely related to kriging (Handcock & Stein, 1993; Omre, 1987). Therefore, the results include traditional geostatistical approaches as a subset.

Ideally, we would like to obtain the probability distribution function describing the concentration in the subsurface for any time during the design life of the remediation scheme given a set of observations from the site. Formally, we may write a Bayes statement relating these parameters. However this type of Bayes approach is computationally too burdensome to be considered directly. We note that the vector of observations may lead to a large number of input parameters for a numerical model describing the behaviour of concentration with time. This yields a problem with too many dimensions to be effectively analysed. To make the problem more tractable, we will consider an estimate of the expected value and covariance of concentration based on vectors of two uncertain input parameters, transmissivity and initial concentration, to the model. Several methods may be used to estimate the uncertainty in the concentration values determined using transport models given uncertainties in the input parameters to the equation. We will apply a Taylor Series approach to propagate the uncertainty from the input parameters to the simulated concentrations. The Taylor Series approximation yields the following estimates for the expected value of concentration with time and the covariance of this computed concentration:

$$E[\mathbf{u}(t)] \approx \mathbf{g}(E[\mathbf{T}], E[\mathbf{u}_0]) \tag{1}$$

$$\begin{aligned} \mathrm{Cov}(\mathbf{u}(t_1), \mathbf{u}(t_2)) &= \mathbf{J}_{\mathbf{T}}(t_0, t_1)\mathrm{Cov}(\mathbf{T}, \mathbf{T})\mathbf{J}_{\mathbf{T}}^{\mathrm{T}}(t_0, t_2) + \\ \mathbf{J}_{\mathbf{u}_0}(t_0, t_1)&\mathrm{Cov}(\mathbf{u}_0, \mathbf{u}_0)\mathbf{J}_{\mathbf{u}_0}^{\mathrm{T}}(t_0, t_2) + \mathbf{J}_{\mathbf{T}}(t_0, t_1)\mathrm{Cov}(\mathbf{T}, \mathbf{u}_0)\mathbf{J}_{\mathbf{u}_0}^{\mathrm{T}}(t_0, t_2) \end{aligned} \tag{2}$$

Equation (1) states that the best estimate of concentration at every point in the discretized domain, $E[u_i]$, is the result given by the flow and transport model with the best estimates of the parameters, transmissivity, $\mathbf{T}$, and initial concentration, $\mathbf{u}_0$, as input. To evaluate the covariance equation (2) the Jacobian matrices relating changes in $\mathbf{T}$ and $\mathbf{u}_0$ to changes in $\mathbf{u}$ and the covariance matrices describing the spatial distribution of $\mathbf{T}$ and $\mathbf{u}_0$ must be determined.

Ultimately, the results of the analysis are presented in two ways: the estimated variance in the dependent variable (equation (2)), and a β or reliability index for the modelled remediation design. These two quantities are used to answer the driving questions posed in the introduction.

The β index is defined as the difference between the modelled response and the allowable performance at given points at a site divided by the standard deviation of the modelled behaviour. In terms of concentration this may be written as:

$$\beta = \frac{u_i - u_{allowed}}{\sigma_{u_i}} \tag{3}$$

where $u_{allowed}$ is the maximum concentration allowed at the compliance point and σ_{ui} is the estimated standard deviation of the modelled concentration at that point. By examining the β values for key locations at a site, the reliability or confidence in success of a modelled remedial design can be ascertained. If the β values are near zero, then the modelled remedial design implies that the concentrations will be near the maximum allowable concentration. If the β is large and negative, the modelled remedial system does not meet the desired performance, and if β is large and positive, the model predicts that the remedial scheme will be successful. The engineer may examine the computed β value and judge the reliability of the remedial scheme, if the system has a low reliability (β near zero or negative) the engineer can examine the u_i and σ_{ui} terms and determine if more sampling that may reduce the σ_{ui} can improve system reliability or if the remedial system must be redesigned.

If examination of the β values indicates that additional sampling may improve the confidence in the design, the site uncertainty and model sensitivity are combined to determine where the next data should be collected. The most important data will be data that reduces the estimated variance in the estimated performance variable. Thus, equation (2) is used and the location of the maximum value of Cov(**u**,**u**) is determined. To emphasize the use of this approximation technique to combine both field uncertainty and model response, the estimated covariance matrix has been called the importancy matrix by several authors (Tomasko *et al.*, 1988; Graettinger & Dowding, 1999).

METHODS AND SIMPLE EXAMPLE

Bayesian formulation and input parameters

To determine the expected value for transmissivity at every nodal point in the modelled domain and the covariance of transmissivity matrix that describes the spatial correlation of this parameter, we use a Bayesian approach that couples site data and a prior model describing the expected correlation. The prior model used for this example is based on developing a variogram and the use of an exponential autocorrelation function. This step is quite akin to kriging. If we assume that the parameters describing the autocorrelation are known, a simple updating scheme to determine the expected value for transmissivity and its covariance may be used (Gelman *et al.*, 1995):

$$E[\mathbf{T} \mid \mathbf{v}] = E[\mathbf{T}] + \mathrm{Cov}(\mathbf{v},\mathbf{T})\mathrm{Cov}(\mathbf{v})^{-1}(\mathbf{v} - E[\mathbf{v}]) \tag{4}$$

$$\mathrm{Cov}(\mathbf{T} \mid \mathbf{v}) = \mathrm{Cov}(\mathbf{T}) - \mathrm{Cov}(\mathbf{v},\mathbf{T})\mathrm{Cov}(\mathbf{v})^{-1}\mathrm{Cov}(\mathbf{T},\mathbf{v}) \tag{5}$$

Note that this analysis may be generalized to allow for the parameters describing the variogram to also be uncertain. Observations from the field are then used to update both the expected value of the input data and the values describing the variogram which are used in the autocorrelation function to give Cov(**T**,**v**). By choosing appropriate distributions for all the uncertain parameters, a procedure may be developed to update the expected value and covariance of transmissivity akin to that shown in equations (4) and (5). This generalized procedure, however, cannot be expressed in such a concise fashion.

Taylor Series implementation for flow equation

The key to this research project is the generation of the required Jacobian matrices described in equation (2) in an efficient manner. The derivatives were computed by applying ADIFOR (Bischof *et al.*, 1995), a program that interprets FORTRAN code and generates new code that evaluates specified derivatives, and by some hand modification of the code, generates analytical derivatives. This algorithm to compute the derivatives is much faster than typical numerical perturbation methods and a great deal faster than Monte Carlo analysis of these sensitivity matrices (Graettinger & Dowding, 1999). For large problems with many computational points, efficient determination of the Jacobian matrices is essential in the development of a practical reliability model.

To demonstrate this approach, we have applied the Taylor Series expansion described above to a two-dimensional, steady state groundwater flow model. The equations are solved using a linear finite element approach, and the thickness of the unit is allowed to vary by changing the transmissivity at each computational node in the domain. We will explore the use of the reliability approach by focusing on the sensitivity of piezometric head results and we will illustrate the use of the estimated covariance matrix (equation (2)) to direct sampling. This simple example is the first important step in analysing the full contaminant transport problem, but is sufficient to illustrate the nature of the procedure. Since the only uncertain input to the model is transmissivity, equation (2) is simplified to consider only the sensitivity of head results to changes in transmissivity. Figure 1 shows the simple domain used for this problem. The domain is discretized into 800 triangular elements and has 441 nodes. A regional piezometric head gradient is imposed across the site as noted on the figure, and the transmissivity in this synthetic domain varies linearly from 50 m^2 day^{-1} along the left-hand side of the domain to 100 m^2 day^{-1} along the right-hand side. Observations used in equations (4) and (5) are generated from the four sampling points shown on the figure. The prior model used in these equations specifies an average transmissivity across the domain based on the four data points and a covariance matrix based on an exponential autocorrelation function. This yields a predicted transmissivity field equivalent to one produced using ordinary kriging on these data. A proposed extraction well with a pumping rate of 500 m^3 day^{-1} is specified, as shown in Fig. 1, to generate a capture zone for a contaminant plume. Figures 2(a) and 2(b) show the expected value for transmissivity input to the domain resulting from application of equation (4) and the estimated standard deviation of transmissivity based on equation (5). Note that transmissivity values in areas of the domain away from the four samples tend to the average value of this parameter, which is the kriging result. We have assumed no nugget effect in the autocorrelation function, and, therefore, the standard deviation in

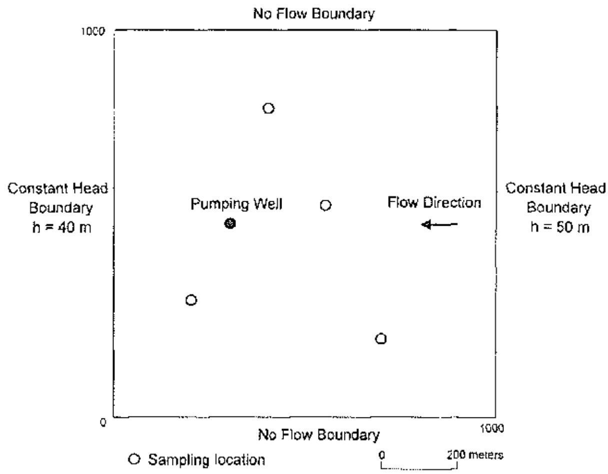

Fig. 1 Simple example domain with sample locations, proposed pumping well and imposed boundary conditions.

this parameter tends to zero at the measurement points and to its maximum value away from these points. Site exploration may be directed by examining these uncertainties. This type of analysis to direct exploration, however, does not consider the use of the site data for the design of the remediation system. To account for system design in the site exploration process, the behaviour of the design model must be considered. Figure 2(c)

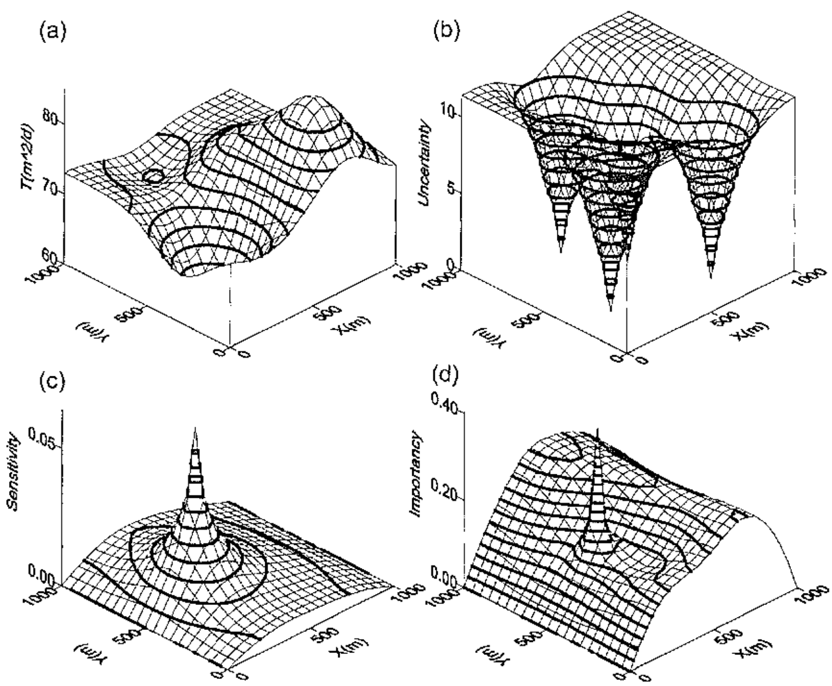

Fig. 2 (a) Transmissivity, (b) uncertainty in transmissivity, (c) sensitivity, and (d) estimated head covariance for simple directed sampling example.

presents the summed values for each node of the Jacobian matrix giving the sensitivity of head changes to changes in transmissivity. This figure indicates the total sensitivity of the piezometric head at the node to changes in transmissivity anywhere in the domain. Note that the highest sensitivity of the model results to changes in transmissivity occurring at the pumping well. We then expect that sampling near the pumping well will significantly impact the estimated uncertainty in piezometric head as computed using equation (2). Figure 2(d) shows the diagonal values of the estimated head covariance matrix using equation (2). These values are the appropriate indicators of the most important location to collect data as they are a combination of the values from Figs 2(b) and 2(c). Consideration of this importancy matrix directs data collection to reduce the maximum value of the entire estimated head covariance matrix and, thereby, to reduce the overall uncertainty in the computed result. For this case, the most important location to collect data is at the proposed pumping well. Data may be collected at this location and equations (4) and (5) used to update the input to the numerical model. Repeating this exercise yields the next most important location. This value occurs near $x = 450$ m and $y = 1000$ m. This next location is near where the uncertainty in estimated transmissivity as shown by Fig. 2(b) is large, but is still influenced by the pumping well. Reliability analysis using the computed β values may be used to determine when site exploration may cease.

Acknowledgements This work was funded in part by Grant R-827126-01-0 from the US Environmental Protection Agency. Dr Takeru Igusa has had a great impact on this research through his efforts on the Bayes formulation and analysis.

REFERENCES

Bischof, C., Carle, A., Khademi, P., Maure, A. & Hovland, P. (1995) *ADIFOR 2.0 User's Guide (Revision C)*. Report CRPC-95516-S, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, USA.

Freeze, R. A., Massmann, J. & Sperling, T. A. (1990) Hydrologic decision framework: Part 1. Motivation and background. *Ground Water* **27**(3), 567–575.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (1995) *Bayesian Data Analysis*. Chapman and Hall, London, UK.

Graettinger, A. J. & Dowding, C. H. (1999) Directing exploration with 3D FEM sensitivity and data uncertainty. *ASCE J. Geotech. Geoenv. Engng* **125**(11), 959–967.

Handcock, M. S. & Stein, M. L. (1993) A Bayesian analysis of kriging. *Technometrics* **35**(4), 403–410.

Harr, M. E. (1987) *Reliability-Based Design in Civil Engineering*. McGraw-Hill, New York.

Johnson, R. L. (1996) A Bayesian/geostatistical approach to the design of adaptive sampling programs. In: *Geostatistics for Environmental and Geotechnical Application* (ed. by R. M. Srivastava, S. Rouhani, M. V. Cromer & A. I. Johnson), ASTM STP 1283. American Society for Testing and Materials, Philadelphia, Pennsylvania, USA.

Omre, H. (1987) Bayesian kriging—merging observations and qualified guesses in kriging. *Mathematical Geology* **19**(1), 25–39.

Tomasko, D., Reeves, M. & Kelley, V. A. (1987) Parameter sensitivity and importance for radionuclide transport in double-porosity systems. In: *Geostatistical, Sensitivity, and Uncertainty Methods for Groundwater Flow and Radionuclide Transport Modeling* (ed. by B. E. Buxton) (Proc. DOE/AECL Conf., September 1987, San Francisco), 297–321. Battelle Press, Columbus, Ohio, USA.