Efficient Groundwater Remediation System Design Subject to Uncertainty Using Robust Optimization

Karen L. Ricciardi; George F. Pinder; and George P. Karatzas

Abstract: Many groundwater remediation designs for contaminant plume containment are developed using mathematically based groundwater flow models. These mathematical models are most effective as predictive tools when the parameters that govern groundwater flow are known with a high degree of certainty. The hydraulic conductivity of an aquifer, however, is uncertain, and so remediation designs developed using models employing one realization of the hydraulic conductivity field have an associated risk of failure of plume containment. To account for model uncertainty attributable to hydraulic conductivity in determining an optimal groundwater remediation design for plume containment, a method of optimization called robust optimization is utilized. This method of optimization is a multi-scenario approach whereby multiple hydraulic conductivity fields are examined simultaneously. By examining these fields simultaneously, the variability of the uncertainty is included in the model. To increase the efficiency of the robust optimization approach, a sampling technique is developed that allows the modeler to determine the minimum number of field realizations necessary to achieve a reliable remediation design.

DOI: 10.1061/(ASCE)0733-9496(2007)133:3(253)

CE Database subject headings: Ground-water management; Operation costs; Optimization; Optimization models; Algorithms; Hydrology; Hydrologic models; Remedial action; Uncertainty principles.

Introduction

Many active pump and treat groundwater-remediation designs, whether generated through the process of trial and error or by automated methods, are based upon groundwater models that use specified aquifer parameters (Gorelick 1983; Ahlfeld et al. 1988; Karatzas and Pinder 1993; Guan and Aral 1999). Efforts are often made to quantify the reliability of these models through sensitivity and uncertainty analysis (Konikow and Mercer 1988). Through uncertainty analysis, research has shown that it is possible to reduce the uncertainty in a groundwater flow model through strategic data collection (Wang and McTernan 2002). The hydraulic conductivity of a single hydrostratigraphic unit is spatially variable (Gelhar 1993) and because of this fact, the reduction of uncertainty in groundwater models is limited.

The hydraulic conductivity of an aquifer is the primary parameter that affects the results of a groundwater flow model. Even when adequate characterization of the spatial variability of hydraulic conductivity has been obtained through data collection, modelers are still limited in their ability to represent a spatially variable hydrostratigraphic unit due to the uncertainty associated with those areas where data have not been collected.

Reliability studies examine the relationship between an uncertain input parameter, and the effect that this uncertainty has on the model, by comparing the probability density functions (PDFs) of the input parameter and the model results. A number of reliability studies related to hydrologic models have been conducted (Konikow and Mercer 1988; Wagner 1999; Yenigul et al. 2006). Sensitivity studies conducted on hydrologic models address how small variations in the input parameters affect the model results (Ahlfeld et al. 1988; Hamed et al. 1996). The results of sensitivity studies can be used to evaluate the significance of various input parameters. Through sensitivity and reliability studies, it has been shown that remediation design models are very sensitive to changes in hydraulic conductivity and that there is a strong relationship between the uncertainty of hydraulic conductivity and the uncertainty in model results (Gorelick 1989; Gelhar 1993; Hamed et al. 1996).

It is possible to include the uncertainty of hydraulic conductivity into groundwater flow analysis by using a multi-scenario method such as a Monte Carlo approach (Hamed et al. 1996; Yenigul et al. 2006). In the Monte Carlo approach, multiple realizations of hydraulic conductivity fields are randomly generated within the bounds of the uncertainty parameters provided. The groundwater model is simulated for each realization, and the results from these models are examined collectively. The Monte Carlo approach is most often used to model groundwater flow and contaminant transport and in reliability studies.

Stochastic methods have been applied in a variety of manners to incorporate the uncertainty in the aquifer's properties into groundwater management problems. A type of stochastic programming called chance-constrained modeling was used in early efforts to include uncertainty directly into optimal remediation designs (Tung 1986). These models utilize a technique whereby
the domain of the uncertain parameter value, in other words all possible values of this parameter, is determined through analysis of the associated PDF. A particular percentile of the distribution, whichever value produces an appropriate conservative result, is then used in a deterministic model. This minimum or maximum value equates with a particular percentile of the data. Wagner and Gorelick (1987), under the hypothesis of normality, split water quality constraints into two parts: a deterministic, expected value component, and a stochastic one related to a specified reliability level. By doing such, they were able to account for the variability in the uncertain parameters in a chance-constrained model. Optimization models developed through stochastic programming often result in conservative systems.

Genetic algorithms (GA) have also been used as a solution technique in the development of a robust approach to optimal groundwater remediation design that takes into account the uncertainty of hydraulic conductivity values (Hilton and Culver 2005). Within each generation of the robust GA, all designs are evaluated using the same realization of the heterogeneous hydraulic conductivity field, but the realizations vary between GA iterations. Ongoing performance of the designs is measured and is used in the GA evolution process resulting in a conservative remediation design that takes into account the uncertainty in hydraulic conductivity.

Wagner et al. have simultaneously considered the operation of a pumping design for groundwater containment as well as the uncertainty in the aquifer properties through the application of a stochastic program with simple recourse (Wagner et al. 1992). This approach is a multiobjective stochastic programming approach, which may be considered a precursor to the robust optimization approach utilized in this work. This model minimized the expected total costs of remediation over a number of realizations of outcomes of the random parameter, namely the hydraulic conductivity of the aquifer. This approach considers the uncertainty in an explicit manner, resulting in a single remediation design that includes the data from the multiple realizations representative of the uncertainty.

Robust optimization (RO) is a multiobjective approach that can be used to develop a least-cost pump-and-treat remediation design or redesign for containment of a contaminated aquifer subject to parameter uncertainty (Mulvey et al. 1995; Karatzas and Pinder 1997; Watkins and McKinney 1995). The RO approach analyzes, simultaneously, simulations that use a number of different aquifer parameter values, thereby including the variance of the uncertain parameter in the model. Mulvey presented his vision of the RO algorithm, not by defining a set of rules, but rather by presenting the ideas for including uncertainty in the optimization problem. Mulvey states that there are multiple ways in which the RO approach can be implemented. This paper presents one possible implementation of the RO algorithm and how this approach can be efficiently applied to the contaminant containment problem.

Robust optimization was first explored as a method for developing a risk-based groundwater remediation system by Watkins and McKinney in 1995 (Watkins and McKinney 1995, 1997). In this new application of Mulvey’s approach, RO was used to quantify the risk associated with pump and treat groundwater remediation designs subject to uncertainty in hydraulic conductivity. The objective in their design was plume containment. The source of uncertainty in their models was the spatial variability of hydraulic conductivity that is known to exist in single hydrostratigraphic units. Pumping rates and well locations were determined in their model so that the resulting remediation design was both cost effective and risk averse.

Since Watkins and McKinney’s work, RO has not been explored further as a viable method for including uncertainty into the remediation design. Because RO is a multiobjective approach, solving RO problems can be numerically intensive. The work presented in this paper explores robust optimization in a manner whereby the number of scenarios necessary to obtain a representative solution is placed under scrutiny. Whereas this work is related to the work of Watkins and McKinney, the focus of this study is quite different. The focus of this work is to increase the efficacy of RO in groundwater remediation design and as such the research presented in this paper answers new questions related to the application of RO.

In addition to the main focus of this paper being different from that of Watkins and McKinney, there are other factors that differentiate this work from previous analysis related to RO. The work presented in this paper utilizes a different implementation of Mulvey’s RO algorithm than Watkins and McKinney’s. The approach taken in this paper is one where a series of nested, deterministic problems are solved and the uncertainty is considered in an implicit manner through postanalysis of the deterministic optimization solutions. This work assumes fixed well locations and as such is a linear optimization problem whereby pumping rates from each of the wells is determined for a least cost remediation design. And finally, in this work, the variability is represented through the generation of perfectly homogeneous fields differing in hydraulic conductivity values. The values of the perfectly homogeneous fields are determined through sampling the probability density function that characterizes the uncertainty.

The uncertainty in the hydraulic conductivity is described herein by the use of a lognormal distribution, as has been historically done (Law 1944; Bennion and Griffiths 1965), and also by a beta distribution. The beta distribution is used to avoid complications that arise when sampling a distribution that has an unbounded domain of values, all of which have positive probability of occurrence, such as the lognormal distribution.

Sampling of the PDF will be done using equal-area sampling. This technique allows for the simplification of the RO formulation so that unnecessary calculations are avoided. As the number of samples analyzed increases, the optimal solutions converge to a reliable design. Using equal area sampling, the number of samples required in this multiobjective approach is minimized.

Because this is a contaminant containment problem, the objective function is convex. Thus, a quasi-Newton steepest descent algorithm is utilized to solve this RO problem in an efficient manner.

Methodology

Design Approach

The formulation of the optimization problem that results in a minimum operation and maintenance cost pump-and-treat groundwater-remediation design, subject to hydraulic-gradients constraints that will assure containment of a contaminated plume, is
Objective: \[
\min \sum_{k=1}^{I} \alpha_k q_k
\]  \hspace{1cm} (1)

Subject to: \[
g - g_i \leq 0, \hspace{0.5cm} i = 1, \ldots, m
\]  \hspace{1cm} (2)

\[
0 \leq q_k \leq \max(q)
\]  \hspace{1cm} (3)

where \(I\) = total number of possible wells in the design; \(\alpha_k\) = cost per unit pumping at well \(k\); \(q_k\) = pumping rate at well \(k\); \(g\) = prespecified required gradient necessary for containment; \(g_i\) = gradient response to pumping \((q_1, q_2, \ldots, q_I)\) at location \(i\) in the model; \(m\) = total number of locations where the gradient constraint must be satisfied; and \(\max(q)\) = maximum allowable pumping from any well \(k\).

A stochastic analysis is conducted to evaluate the effects of the uncertainty on the outcomes of the optimization model. In the stochastic analysis of the optimal remediation design subject to uncertainty the uncertain parameter is sampled multiple times, and an optimal remediation design is determined for each sampled value (Ghanem 1991). The resultant mean and standard deviation of the optimal remediation designs are then used to determine a final remediation design and to quantify the uncertainty that the resultant remediation design will be successful.

RO is a method of optimization that has been developed specifically for problems that include uncertain parameter values (Mulvey et al. 1995). Each scenario represents one possible realization of the uncertain parameter field. In this problem it is assumed that all of the parameter statistics are known. When the statistics of an input parameter are unknown, a different approach is assumed that all of the parameter statistics are known. When the statistics of an input parameter are unknown, a different approach is introduced.

The values of hydraulic conductivity that differentiate each scenario (one from the other) are determined through sampling the PDF that characterizes the uncertainty. Because our objective at this point is to represent the model uncertainty attributable to hydraulic conductivity as a regional variable, rather than spatial variability of the hydraulic conductivity, each scenario is represented by a perfectly homogeneous hydraulic conductivity field. Later analyses introduce spatial variability into the model. Once the possible realizations have been determined, they are used to calculate an optimal remediation design.

A number of different interpretations of the RO technique are possible. In the most general sense RO can be expressed as the following:

Objective: \[
\min_{S \in \Omega} \left\{ \min_{q_k} \sum_{k} \alpha_k q_k + \omega \frac{1}{|N_S|} \sum_{S_1 \in \Omega, S_1 \neq S} \max(0, \xi_{S_1}) \right\}
\]  \hspace{1cm} (6)

Subject to: \[
g - g_i \leq 0 \hspace{0.5cm} \text{for} \hspace{0.5cm} S^*
\]  \hspace{1cm} (7)

\[
0 \leq q_k \leq \max(q) \hspace{0.5cm} \forall \hspace{0.5cm} S \text{ and } S^*
\]  \hspace{1cm} (8)

where \(\Omega = \text{set of all possible scenarios}; \ S^* = \text{selected scenario}; \ S = \text{scenarios that are not considered to be the selected scenario}; \ \omega = \text{total weight for the penalty term}; \ |N_S| = \text{probability that scenario } S \text{ exists}; \ 1/|N_S| = \text{individual weight of scenario } S; \ \text{and } \xi_{S_1} = \text{sum of the constraint violations for the scenario } S_1 \neq S \text{ which is only positive when a constraint is not satisfied, i.e., } (g - g_i) > 0.

The penalty costs in the nested RO problems have three components (Eq. (6)). The first is the constraint violation, \(\xi_{S_1}\), for each scenario not considered to be the selected scenario, \(S\) (i.e., \(S \neq S^*\)). The second is the total weight, \(\omega\), which can be thought of as a risk-aversion term. The third is the weight associated with the frequency of occurrence of each scenario, \(1/|N_S|\). For example, if a scenario not equal to \(S^*\) has low probability of occurrence and the constraint is violated for that particular scenario subject to a specified pumping schedule, the penalty cost will be lower. This third component of the penalty term prevents the optimal design from being biased toward improbable scenarios.

Assuming each scenario to be the selected scenario, \(S^*\), the optimal pumping combination, \(q^*\), is determined for that scenario. (Note that \(q = (q_1, q_2, \ldots, q_I)\) where \(I\) = total number of wells.) The final remediation design is then selected as the minimum over all of the pumping combinations determined for each scenario.

In the RO formulation, there is a cost associated with meeting the constraints for the selected scenario and a cost associated with not meeting the constraints for all other scenarios. If the penalty
weight is very large, then the penalty costs will dominate the value of the total RO objective function. In the set of given scenarios, there is one scenario whose optimal solution satisfies the constraints of all the given scenarios because it is the most restrictive. The RO objective function evaluated at this solution has no penalty costs and hence it is the most conservative solution given a set of scenarios. The pumping cost associated with the most conservative solution is the maximum optimal pumping cost required of any of the scenarios to meet the given constraints. If the penalty weight is very large in the RO formulation, then this maximal cost and most conservative solution will be the minimum RO solution over the set of scenarios. Similarly, if the penalty weight is very small, then the least conservative solution will be selected as the minimum RO solution over the set of scenarios. Thus, in determining a solution using the approach presented, it is essential that the value of the penalty weight be such that there is a balance between the cost associated with meeting the constraints for a given scenario and the cost associated with violating the constraints for all other scenarios being considered.

Considerable computational savings can be realized through the application of this technique. It is not necessary to examine cases where all of the scenarios are assumed to be the selected scenario if the order in which the selected scenarios are chosen is done in the following logical manner. The first scenario, assumed to be the true scenario, or the selected scenario, has the smallest hydraulic conductivity value. The following scenario assumed to be the selected scenario is that associated with the next highest hydraulic conductivity value of all the remaining scenarios. The analysis continues to examine the selected scenarios with hydraulic conductivity values of increasing values. The optimal pumping associated with the scenarios associated with increasing hydraulic conductivity values are examined sequentially until an increase in total RO cost (pumping and penalty) is observed for successive scenarios (Fig. 1). The total cost for selected scenarios associated with higher hydraulic conductivity values than those examined thus far very probably will be higher. There is no need to conduct further computations, as a minimum robust solution has been determined.

Calculations of the objective function are computationally intensive. For each combination of pumping, the value of the objective function requires the calculation of the hydraulic-gradient values for each scenario, not considered to be the selected scenario, in order to determine the sum of the constraint violations, \(\xi_\text{S}\). The feasible region of the RO problem, however, is defined by the constraints applied to the selected scenario.

The hydraulic-head response for a confined aquifer is linear with respect to pumping, resulting in a convex objective function with linear constraints. A quasi-Newton steepest descent method is used to solve this problem (Press et al. 1992).

### Determining a Set of Scenarios

In an effort to minimize the number of samples required to gain a representative set of hydraulic-conductivity fields, we have applied a new method of sampling called equal-area sampling. Equal-area sampling is derived from the Latin-hypercube method of sampling (Press et al. 1992). The PDF that describes the uncertainty in hydraulic conductivity is separated into equal areas. Rather than employ random sampling from each equal-area interval defined in Latin-hypercube sampling, the individual samples, \(x_i\), are selected as those that divide the PDF into equal areas. This ensures equal probability for each of the scenarios and allows the use of a single individual penalty weighting term for each scenario in the RO problem. This weight is equal to one over the total number of scenarios analyzed, so if nine samples are desired, then the PDF will be separated into ten equal areas (Fig. 2), and the individual weighting term for each scenario is equal to one divided by nine, or one ninth, that is \(1/N_s = 1/9\) for all \(S\) where \(|\Omega|\) = cardinality of the set \(\Omega\).

The equal area sample values for a given distribution curve, \(x_i\), where \(i=1,\ldots,n\) and \(n=\text{number of equal area samples}\), can be determined by solving the following equation for \(x_i\):
Table 1. Maximum Hydraulic Conductivity Value, m/hr, Obtained Using Equal Area Sampling

<table>
<thead>
<tr>
<th>Number of scenarios</th>
<th>Lognormal</th>
<th>β distribution</th>
<th>65% of lognormal</th>
<th>95% of lognormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.0134</td>
<td>0.0129</td>
<td>0.0103</td>
<td>0.0128</td>
</tr>
<tr>
<td>10</td>
<td>0.0149</td>
<td>0.0142</td>
<td>0.0107</td>
<td>0.0139</td>
</tr>
<tr>
<td>30</td>
<td>0.0174</td>
<td>0.0162</td>
<td>0.0110</td>
<td>0.0152</td>
</tr>
<tr>
<td>50</td>
<td>0.0186</td>
<td>0.0171</td>
<td>0.0111</td>
<td>0.0156</td>
</tr>
<tr>
<td>100</td>
<td>0.0201</td>
<td>0.0171</td>
<td>0.0112</td>
<td>0.0160</td>
</tr>
</tbody>
</table>

\[
\frac{1}{n+1} = \int_{0}^{x_{i}} P(x)dx \quad (9)
\]

where \( P(x) = \text{given PDF} \).

Equal-area sampling also allows the modeler to observe convergence of the solution to a reliable solution. We define a reliable solution to be one where the value of the solution does not change a significant amount as the number of scenarios examined increases. When using the described application of RO to solve the plume-containment problem, as the number of scenarios examined grows the solution increases monotonically.

To understand why this is the case, recall that the RO solution is the sum of the pumping cost and the cost associated with the violations of the constraints. As the number of scenarios increases, the largest hydraulic-conductivity value sampled from the representative PDF increases in value (Table 1). When the highest hydraulic-conductivity value is not the selected scenario in the RO approach, the violations of the constraints associated with the high hydraulic-conductivity value will be large, as will be the penalty term in the objective function. When the highest conductivity value is identified with the selected scenario, a large amount of pumping is necessary to satisfy the constraints.

Historically a lognormal-distribution curve has been used to describe the uncertainty in the hydraulic conductivity (Law 1944; Csallany and Walton 1963; Gelhar 1993). When examining a histogram of hydraulic-conductivity data, the lognormal distribution does conform to the data. The domain of the lognormal distribution, however, is not bounded from above and therefore has positive probability values associated with all positive hydraulic-conductivity values. Because all positive hydraulic-conductivity values have a positive probability of occurrence, when equal-area sampling is applied to the lognormal distribution, as one increases the number of scenarios examined the highest hydraulic-conductivity value will increase without bound (Table 1). When the highest hydraulic-conductivity value increases without bound, the solution will not converge as the number of scenarios increases.

The solutions of Eq. (9) for the case of \( n \) equal-area samples for the lognormal-distribution function are given by

\[
x_i = \exp \left[ \sqrt{2} \sigma \text{erf}^{-1} \left( \frac{n - i}{n} \right) + \mu \right] \quad (10)
\]

where \( \mu = \text{mean of ln(x)} \) and \( \sigma = \text{standard deviation of ln(x)} \). Notice that as \( i \) increases and as \( n \) increases in Eq. (10), the value of \( x_i \) approaches infinity. This is due to the limit property of the inverse error function, \( \text{erf}^{-1}(y) \), that is as \( y \) approaches the value 1 from values less than 1, then \( \text{erf}^{-1}(y) \) approaches infinity.

Because problems arise in solving the RO problem when using the lognormal distribution to describe the uncertainty in the hydraulic conductivity, one may truncate the lognormal-distribution function to obtain a limited range of positive hydraulic-conductivity values. Using this strategy, as the number of scenarios, \( n \), goes to infinity in Eq. (9), the largest value of hydraulic conductivity is bounded above.

When using a truncated lognormal distribution to describe the uncertainty, the degree to which the distribution is truncated is determined by the modeler. Because the amount of truncation is subjective, we examine an alternative to truncation of the lognormal-distribution function based upon a PDF that can have a form similar to the lognormal distribution, but has positive probabilities over a bounded domain. One such distribution is the beta-distribution function.

A skewed version of the unimodal form of the beta distribution can mimic a lognormal distribution (Johnson et al. 1994) (Fig. 3). When ample data are available, the domain supported by the beta distribution is determined by the minimum and maximum value of the hydraulic conductivity value measured, and the optimal beta-distribution-shape parameters will be such that the beta-distribution function fits the histogram of hydraulic-conductivity data. The functional form of the beta distribution is as follows:

\[
P(y) = \frac{1}{\Gamma(p+1, q+1)} \left( \frac{y-a}{b-a} \right)^p \left( \frac{b-y}{b-a} \right)^q \quad (11)
\]

The function \( \Gamma(p+1, q+1) = \beta(a, b) \) function. The values \( a \) and \( b \) define the lower and upper bounds of the range of possible hydraulic-conductivity values. The parameters \( p \) and \( q \) describe the shape of the beta-distribution function (Johnson et al. 1994).

A beta-distribution curve can be fit to the hypothetical-distribution data of the hydraulic conductivity, just as a lognormal distribution was fit to these data (Riccioardi et al. 2005). In this analysis the beta distribution that best approximates the hypothetical lognormal-distribution function is used. The domain of hydraulic-conductivity values is set equal to \([0, 1]\). The modal value of the beta distribution is required to be the same as the modal value of the representative lognormal distribution, imposing a relationship between the shape parameters \( p \) and \( q \). (In a nontheoretical problem, the true data would determine the range of values chosen and the modal value of the distribution.) The method of least squares is used to determine the shape parameters, \( p \) and \( q \), of the unimodal form of the beta distribution that best fits the representative histogram of the hydraulic-conductivity distribution.

Fig. 3. Both a lognormal and a skewed unimodal beta distribution curve can be used to represent the uncertainty in the hydraulic conductivity.
The integral of the beta-distribution function is not expressible in closed form, so an analytic expression for the equal-area sample points cannot be derived for this distribution. Using a modified version of Lentz's method to obtain a numerical approximation of the integral of the incomplete beta function (Press et al. 1992), a linear search is performed for each of the equal-area samples of the beta distribution.

When a beta distribution is used to describe the uncertainty in the hydraulic conductivity, convergence of the remediation design with an increasing number of scenarios examined in the problem is observed. The design to which this analysis converges is the solution to this problem. Because the data determine the range of values covered by the beta distribution, this optimal solution is not dependent upon the discretion of the modeler. This solution is objective.

A number of sampling approaches are examined and compared to the average solution determined stochastically. In each analysis, the number of samples examined increases and convergence of the solution is observed. By examination of the solutions determined in this way, a conservative remediation system is established. It should be noted that as the number of scenarios increases, the reliability of the solution as one that is representative of the uncertainty increases. This reliability should not be confused with the concept that a greater number of samples will decrease the variance of the uncertainty distribution, thereby increasing the reliability of the solution. The uncertainty distribution in this model is fixed and so as more samples' values are collected, the mean and the standard deviation of the population of sampled values will approach the mean and standard deviation of the PDF from which the samples were determined. The sampling approaches considered include random sampling applied to a lognormal distribution, equal-area sampling applied to a lognormal distribution, equal-area sampling applied to a truncated lognormal distribution, and equal-area sampling applied to a beta distribution.

The final analysis includes spatial variability of hydraulic conductivity superimposed upon the design uncertainty in the model. After determining the hydraulic conductivity values that represent the design uncertainty, multiple spatially variable fields are generated for each of the sampled values. Solving the problem applied to these scenarios is numerically intensive due to the increase in the number of scenarios considered.

**Sample Problem**

For this investigation a finite-element-groundwater-flow model is used. It has 40 nodes by 25 nodes, equally spaced, and represents an area that is 1,170 m by 720 m (Fig. 4).

This is a confined aquifer with boundary conditions such that there is no flow out of the northern and southern boundaries of this model and there are constant head conditions of 50 and 30 m on the western and eastern boundaries, respectively. These conditions create a uniform hydraulic gradient across the model. The model has one layer of uniform thickness of 70 m.

Two problems are presented using this model. The first problem involves two possible extraction wells located in the interior of the desired capture zone. The two wells are located at Points 1 and 2 in Fig. 4. The second problem involves six possible well locations, also located in the interior of the desired capture zone. These six wells are located at Points 1, 2, 3, 4, 5, and 6 in Fig. 4. The remediation design is implemented for 3 years, at the end of which the steady-state flow field is analyzed. The cost per unit of pumping from each well for three years is set equal to the constant $43,920. The capture zone is realized by placing gradient constraints on pairs of nodes placed along a line surrounding the desired capture zone boundary. The gradient constraints are such that the flow of groundwater must be toward the wells after 3 years of pumping.

In the stochastic analysis, 100 random scenarios are generated by sampling the lognormal distribution with median of exp(μ) equal to 0.01 m/h, i.e., μ = ln(0.01), and standard deviation of the related normal distribution, σ, equal to 0.3. For each scenario an optimal cost pumping design is determined for containment.
Fig. 6. Six well total cost solutions (pumping cost + penalty cost) as a function of the number of scenarios utilized to characterize the uncertainty. The sample values that characterize the scenarios used to describe the uncertainty were determined through randomly sampling the lognormal distribution or through equal area sampling of the beta distribution, the lognormal distribution, 95% of the lognormal distribution, or 65% of the lognormal distribution.

These results are compared with the results from the new application of RO (Figs. 5 and 6, and Tables 2 and 3).

In the new application of the RO model, the hydraulic conductivity is determined for each scenario in a given study using one of these strategies, that is a hypothetical-lognormal distribution, a truncated-lognormal distribution, or a beta distribution. The lognormal distribution used in this study is the same lognormal distribution used in the stochastic analysis (Fig. 3).

To examine how the variance of the uncertain parameter influences the solution to the remediation problem, the results of the two-well problem attained using different standard deviations of the related normal distribution are compared. The standard deviation values of the related normal distribution to the lognormal distribution with median of 0.01 m/hr examined are 0.3, 0.1, 0.03, 0.003, and 0.0.

The penalty weight considered initially is $8.0 \times 10^6$. This value was chosen because using this value the solution to the remediation problem is achieved when the selected scenario is higher than the mean hydraulic conductivity value, but lower than the highest sampled hydraulic conductivity value. This indicates that the solution is a conservative remediation system and that there is a balance between the operational costs of remediation and the penalty costs associated with failure of the system due to risk associated with the uncertainty of the hydraulic conductivity. To examine the effect of the penalty weight on the remediation solution, the results of the two-well problem using penalty weights of $8.0 \times 10^5$ and $8.0 \times 10^7$ are compared to the results attained using $8.0 \times 10^6$.

Analyses are conducted on the same lognormal-distribution function truncated at the 95th percentile and at the 65th percentile. Truncation at the 95th percentile captures almost all of the
distribution and eliminates from consideration only those high values associated with the tail of the lognormal distribution. Truncation at the 65th percentile maintains the modal values of the distribution, while eliminating a much larger proportion of the tail of the distribution.

The beta distribution that best fits the representative hydraulic conductivity distribution employs the beta distribution as described by the aforementioned lognormal distribution parameters given by $a=0$, $b=1$, $p=10.094$, and $q=1094.34$ (Fig. 3).

Once the equal-area sample values have been calculated from the PDF, the scenarios are defined assuming perfectly homogeneous aquifers, each with the different sampled hydraulic conductivity values. Spatial variability is introduced into the two-well problem by superimposing 15 spatially correlated randomly distributed two-dimensional matrices upon each of the scenario values. These matrices were generated using the subroutine SGSIM that is part of the GSLIB package developed by the Stanford Center for Reservoir Forecasting (Deutsch and Journel 1998). The matrices have a correlation length of 120 m, which equates with a correlation over a maximum of four nodes in the model. Simple kriging is applied to obtain a field that follows a Gaussian distribution with mean of zero and standard deviation of 0.03. These spatially distributed fields are superimposed upon uniform hydraulic conductivity fields determined through equal area sampling of the lognormal distribution.

The RO solutions for scenarios with sequentially larger mean values are examined in the order of increasing mean hydraulic conductivity values. In this analysis, hydraulic conductivity fields with the same mean are examined simultaneously. These scenarios are differentiated by different randomly distributed spatially variable fields. Because the mean values are the same, and because the spatial variability is randomly distributed, the RO solutions of scenarios with the same mean are determined in the order in which the scenarios were generated.

### Results

In this section we report on the effects of the variance of the uncertain parameter, as well as the effects of variable penalty weights on the solution to the two-well problem. We also examine the results of the two-well and the six-well problems as a function of the PDF used to describe the uncertainty and the number of scenarios used to represent the uncertainty in the optimization problem.

The results of the stochastic analysis on the two-well problem indicate that the optimal remediation cost for 100 individual randomly generated hydraulic conductivity fields has an average solution that indicates that Well 1 should pump 6.154 m³/hr and Well 2 should pump 18.388 m³/hr for 3 years with a total pumping cost of $1,077,868. The standard deviation on pumping from Well 1 is 0.705, Well 2 is 2.070 and pumping cost is $112,614. This solution reflects an average solution and hence is not considered to be a conservative system. When applying the nested optimization problem, provided the weight is appropriately chosen, the solution will be more risk averse, and more conservative than the average stochastic result. The results of applying the nested optimization problem with 50 scenarios sampled from the same distribution as that used in the stochastic analysis results in a pumping design where Well 1 pumps 8.98 m³/hr, Well 2 pumps 20.89 m³/hr, and the total cost of pumping is $1,311,890. This result is greater than the result determined stochastically, and hence is a conservative remediation system.

The applied RO problem is solved for the two-well problem with incrementally larger populations of scenarios consisting of samples determined through equal-area sampling. Initially no spatial distribution is considered. Each distribution is examined independently for convergence to an optimal solution as the population of scenarios increases. Random sampling of the lognormal distribution is also examined. Finally, spatial variability of hydraulic conductivity is added to the scenarios examined in the two-well problem to determine its effect on the optimal solution.

The solutions to the two-well problem where the variance in

### Table 4. Two Well RO Solutions Where 50 Scenarios are Considered. These Scenarios Are Determined through Equal-Area Sampling of Lognormal Distributions Defined by Different Variance Parameter Values. The Penalty Weight in the RO Problem Is the Same for All Cases: $8.0 \times 10^6$.

<table>
<thead>
<tr>
<th>Standard deviation</th>
<th>RO solution ($)</th>
<th>Pumping cost ($)</th>
<th>Penalty</th>
<th>Pumping Well 1 (m³/h)</th>
<th>Pumping Well 2 (m³/h)</th>
<th>Selected scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.30</td>
<td>1,449,664</td>
<td>1,311,890</td>
<td>137,774</td>
<td>8.98</td>
<td>20.89</td>
<td>41</td>
</tr>
<tr>
<td>0.10</td>
<td>1,069,553</td>
<td>1,029,485</td>
<td>40,068</td>
<td>6.59</td>
<td>16.85</td>
<td>36</td>
</tr>
<tr>
<td>0.03</td>
<td>953,638</td>
<td>943,402</td>
<td>10,236</td>
<td>5.62</td>
<td>15.86</td>
<td>35</td>
</tr>
<tr>
<td>0.003</td>
<td>910,928</td>
<td>910,022</td>
<td>905</td>
<td>5.31</td>
<td>15.41</td>
<td>46</td>
</tr>
<tr>
<td>0.00</td>
<td>907,120</td>
<td>907,120</td>
<td>0.00</td>
<td>5.25</td>
<td>15.41</td>
<td>All same</td>
</tr>
</tbody>
</table>

### Table 5. Two Well RO Solutions Where 50 Scenarios Are Considered. These Scenarios Are Determined through Equal-Area Sampling of the Same Lognormal Distribution. The Penalty Weight Used in Each of These RO Problems Is Variable.

<table>
<thead>
<tr>
<th>Penalty weight</th>
<th>RO solution ($)</th>
<th>Pumping cost ($)</th>
<th>Penalty</th>
<th>Pumping Well 1 (m³/h)</th>
<th>Pumping Well 2 (m³/h)</th>
<th>Selected scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>$8 \times 10^5$</td>
<td>1,075,501</td>
<td>892,894</td>
<td>182,607</td>
<td>6.84</td>
<td>13.49</td>
<td>4</td>
</tr>
<tr>
<td>$8 \times 10^6$</td>
<td>1,449,664</td>
<td>1,311,890</td>
<td>137,774</td>
<td>8.98</td>
<td>20.89</td>
<td>4</td>
</tr>
<tr>
<td>$8 \times 10^7$</td>
<td>1,683,278</td>
<td>1,678,622</td>
<td>4,656</td>
<td>9.92</td>
<td>28.30</td>
<td>48</td>
</tr>
</tbody>
</table>
the uncertain parameter is different are summarized in Table 4. In this analysis 50 scenarios are considered. The solution determined when the variance is zero is equal to the optimal pumping design when the hydraulic conductivity value is equal to the mean hydraulic conductivity value for all other distributions considered. The selected scenario for all sample sets determined from distributions with nonzero standard deviations is associated with a hydraulic conductivity value that is greater than the mean of the distributions, resulting in a conservative remediation system. The pumping rates for systems where the variance is greater are higher. As the variance decreases, the cost of the optimal solution decreases and approaches the solution to the optimization problem where the hydraulic conductivity is assumed to be a known value, that is the variance is zero. The relationship between variance and total RO cost appears to be almost linear for this problem.

The optimal solutions to the two-well problem where the penalty weights are different are summarized in Table 5. As the penalty weight increases, the effect of the violations, and hence the variability, affects the solution more significantly. High penalty weights are associated with a design with very little risk. Violations of the constraints for scenarios with high hydraulic conductivity values are high, thereby increasing the penalty costs. This drives the solutions to the RO problem to be high.

The number of scenarios examined is plotted versus the costs of the solutions to the two-well problem in Fig. 5 and the six-well problem in Fig. 6. When random sampling is employed, the solutions to the RO problem do not appear to converge to a robust optimum as quickly as the solutions determined when equal-area sampling is employed. When equal-area sampling is used, the values of the solutions to the RO problem monotonically increase as the number of scenarios examined increases. As the number of scenarios increases, convergence to a solution is more easily observed than when random sampling is employed.

The solutions to the problem using equal-area sampling on the lognormal-distribution curve seem to increase without bound as the number of samples increase in both the two-well and the six-well problems (Figs. 5 and 6). This is due to the influence of the highest hydraulic-conductivity value on the RO problem. As noted earlier, as the number of samples drawn from a lognormal-hydraulic-conductivity distribution increases in equal-area sampling, the highest hydraulic conductivity value increases without bound.

When the truncated lognormal-distribution curve is used to represent the uncertainty in the hydraulic conductivity, convergence of the optimal solution does occur as the number of samples examined increases. The value to which the RO problem converges is dependent upon the chosen degree of truncation (Figs. 5 and 6). Truncation at 65% greatly reduces the RO solution because the high hydraulic conductivity sample values are excluded from consideration in the optimization problem. Despite the fact that the 65% truncated distribution maintains the unimodal character of the lognormal distribution, the effects of severe truncation of the high values limit the significance of the optimal solution.

Convergence is also observed using equal-area sampling applied to the beta-distribution function. The advantage to using the beta-distribution function over the truncated-lognormal-distribution function is that the range of hydraulic-conductivity values spanned by the beta-distribution function is determined by the variation in the hydraulic conductivity.

In both the two-well and the six-well problems where equal-area sampling is conducted on the beta-distribution to determine the representative scenarios, the change in the RO solution as a function of the number of scenarios considered appears to be minimal after 30 scenarios are considered (Figs. 5 and 6). The results of the final pumping design for the two-well and the six-well problem when 30 scenarios are considered are recorded in Tables 4 and 5. Because the change in the RO solution is monotonically increasing and convergence can be determined using this optimization approach, it is possible to determine scenario set size that is representative of the uncertainty. This is a major advantage of this method of optimization over other approaches because a minimum number of scenarios can be determined, thereby minimizing the computational effort required to determine a representative solution.

In the final analysis conducted on the two-well problem the hydraulic conductivity fields contain correlated spatial variability. In this model the value of the optimal solution increases over the value without the spatial variability. The results of the two-well problem with and without spatial variability are summarized in Tables 6 and 7. The value of the new optimal solution is slightly higher than the result without spatial variability because the spatial variability increases the hydraulic conductivity in some locations, thereby increasing the required pumping necessary to satisfy the gradient constraints. This increase is reflected in a solution with a slightly higher cost.

### Conclusions

Robust optimization is a method of optimization that addresses the uncertainty in the groundwater-flow model while taking into account the spatial variability. The results of this optimization approach are summarized in Tables 6 and 7. The value of the new optimal solution is slightly higher than the result without spatial variability because the spatial variability increases the hydraulic conductivity in some locations, thereby increasing the required pumping necessary to satisfy the gradient constraints. This increase is reflected in a solution with a slightly higher cost.

<table>
<thead>
<tr>
<th>Number of scenarios</th>
<th>RO solution ($)</th>
<th>Pumping cost ($)</th>
<th>Penalty cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1,109,643</td>
<td>1,109,639</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1,182,860</td>
<td>1,156,515</td>
<td>29,345</td>
</tr>
<tr>
<td>5</td>
<td>1,256,097</td>
<td>1,150,221</td>
<td>105,876</td>
</tr>
<tr>
<td>7</td>
<td>1,295,109</td>
<td>1,189,134</td>
<td>105,975</td>
</tr>
<tr>
<td>10</td>
<td>1,333,844</td>
<td>1,229,053</td>
<td>104,787</td>
</tr>
<tr>
<td>15</td>
<td>1,374,214</td>
<td>1,291,424</td>
<td>82,790</td>
</tr>
<tr>
<td>20</td>
<td>1,390,539</td>
<td>1,284,572</td>
<td>105,967</td>
</tr>
<tr>
<td>30</td>
<td>1,418,626</td>
<td>1,286,636</td>
<td>131,990</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of scenarios</th>
<th>RO solution ($)</th>
<th>Pumping cost ($)</th>
<th>Penalty cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1,167,160</td>
<td>1,165,856</td>
<td>10,304</td>
</tr>
<tr>
<td>4</td>
<td>1,229,503</td>
<td>1,206,482</td>
<td>23,020</td>
</tr>
<tr>
<td>5</td>
<td>1,275,247</td>
<td>1,257,034</td>
<td>18,313</td>
</tr>
<tr>
<td>7</td>
<td>1,327,327</td>
<td>1,303,633</td>
<td>23,693</td>
</tr>
<tr>
<td>10</td>
<td>1,378,176</td>
<td>1,295,069</td>
<td>83,107</td>
</tr>
<tr>
<td>15</td>
<td>1,406,016</td>
<td>1,295,376</td>
<td>110,639</td>
</tr>
<tr>
<td>20</td>
<td>1,430,678</td>
<td>1,303,326</td>
<td>127,352</td>
</tr>
<tr>
<td>30</td>
<td>1,456,029</td>
<td>1,318,874</td>
<td>131,155</td>
</tr>
</tbody>
</table>
account the standard deviation of the uncertain parameter, in our case hydraulic conductivity. Robust optimization does this by utilizing a multiscenario approach to uncertainty whereby each scenario has a weight that is associated with its frequency of occurrence. By using equal-area sampling, the RO problem is greatly simplified because, through equal-area sampling, fewer samples are required to represent the uncertainty distribution than if random sampling were employed and with equal area sampling each scenario has equal weight in the RO problem.

The application RO in this work has shown that RO can be used to develop a design that accounts for variable degrees of risk. The RO problem has been applied in a nested optimization problem, whereby it has been shown, through comparison with stochastic results, that it is possible to determine a conservative result that takes into account the uncertainty. The degree to which the resultant system is conservative is dependent upon the penalty weight assigned in the RO problem. Recall that the degree of risk is modeled by the penalty weight.

Because robust optimization is a multiscenario approach, and hence a computationally intensive process, it is important to keep to a minimum the number of scenarios necessary to represent the uncertainty in the hydraulic conductivity. Convergence of the solutions determined through the application of RO presented in this paper is easily observed when applying equal-area sampling to the truncated-lognormal distribution and to the beta distribution. Convergence in our test problems is observed after approximately 30 scenarios.

The introduction of spatial variability to this model increases the resulting solution cost value. Efforts have been made to include the uncertainty of aquifer parameters into optimal groundwater remediation designs through the use of optimization methods that represent uncertainty with a multiscenario approach. Because of the nature of sampling hydraulic conductivity in the field, most field based models are developed with some degree of uncertainty and hence remediation designs based upon these models have risk associated with this uncertainty. In an effort to simplify the interpretation of methods that utilize a multiscenario approach to including uncertainty in the optimization model, this work illustrates a new method by which a multiscenario approach is utilized to determine a deterministic approach that takes into account the risks associated with the uncertainty. The limitation of this method is that it is reliant upon a multiscenario approach, and hence can be computationally intensive when the number of scenarios necessary to represent the uncertainty is large. It has been shown, however, that the number of representative scenarios in this method is smaller than the typical number of scenarios used to represent uncertainty in a stochastic approach. The resultant deterministic model determined through application of this nested optimization problem is determined to be one of the scenarios investigated in the multiscenario representation of the uncertain parameter. This deterministic model is dependent upon a prespecified penalty weight of the risk associated with the uncertainty. This penalty weight is prespecified by the modeler. Through careful selection of the penalty weight, it is possible to determine a conservative remediation system for an aquifer where the physical parameters are uncertain.

The optimization approach developed in this research has been applied to a hypothetical single layer homogeneous aquifer. A pump-and-treat groundwater remediation design has been developed for contaminant containment whereby groundwater flow constraints have been placed upon this system. This optimization model can also accommodate the examination of multiple hydrosratigraphic units in the groundwater model.

Acknowledgments

This work was made possible through funding by the U.S. Department of Energy, Contract No. DE-FG07-97ER62525.

Notation

The following symbols are used in this paper:

- \( \mathbf{A} \) = matrix operator acting on the control vector \( \mathbf{x} \);
- \( a, b \) = lower and upper bounds of the support of the beta function;
- \( \mathbf{b} \) = required constraint vector;
- \( f(x) \) = cost as a function of arbitrary control vector \( x \);
- \( g \) = maximum allowable hydraulic gradient;
- \( g_i \) = hydraulic gradient measured at location \( i \);
- \( I_j \) = incomplete beta function;
- \( K \) = hydraulic conductivity;
- \( m \) = total number of gradient constraint locations;
- \( \max(q) \) = maximum allowable pumping from each well;
- \( n \) = number of equal area samples;
- \( P(x) \) = given probability density function;
- \( p, q \) = shape parameters of the beta function;
- \( q_k \) = pumping combination from all wells;
- \( S \) = scenario;
- \( S^* \) = scenario obtained using the selected hydraulic-conductivity field;
- \( t \) = amount of truncation of the lognormal distribution expressed as a decimal;
- \( x_k \) = \( k \)th equal area sample value;
- \( \alpha_k \) = cost per unit pumping from well \( k \);
- \( \beta \) = beta function;
- \( \mu \) = mean of the lognormal distribution of the hydraulic conductivity data;
- \( \xi_k \) = sum of the positive constraint violations;
- \( \rho(z) \) = violation as a function of arbitrary vector \( z \);
- \( \sigma \) = standard deviation of the lognormal distribution of the hydraulic conductivity data;
- \( \Omega \) = the set of all possible scenarios;
- \( \omega \) = total penalty weight; and
- \( 1/N_3 \) = individual weight of Scenario \( S \).

References


