Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 7, No. 3, pp. 503–529, September 2011

THE EFFECTS OF LARGE SCALE VARIATIONS IN A PHYSICAL MODEL ON THE RELEVANCE OF UNCERTAINTY IN AN OPTIMIZATION DESIGN APPLIED TO A GROUND WATER SYSTEM

Karen L. Ricciardi

Abstract: This work focuses on interpreting the importance of considering uncertainty in determining a reliable and cost effective design for remediation of a physical system, namely a ground water system. The utilization of optimization in this process is critical as it provides a mechanism for determining cost effective designs. In these optimization problems calculating values of the objective and constraint functions requires the application of mathematical simulations of the physical system. A genetic algorithm (GA) is utilized to solve the optimization problems. Uncertainty is considered for a variety of physical systems using a multi-scenario approach. The physical systems examined differ in their large scale features. The differences in the affects of the uncertainty on the solutions to the optimization problems applied to each of the systems are compared using statistical tools. Through correlation measures, the significance of uncertainty in each of the systems is determined. The results of this work indicate that large scale features of the physical system for ground water remediation design problems dictate the relevance of uncertainty in determining reliable cost effective designs.

Key words: *hydrology, uncertainty, ground, water management, correlation, optimization*

Mathematics Subject Classification: *86A05, 90C90, 62M20*

---

1 Introduction

Numerical models are used extensively as predictive tools for making ground water management decisions. The accuracy of predictive ground water flow models and subsequent reliability of management designs is dependent upon the ability to accurately represent material properties of a ground water system, also known as a ground water aquifer. Uncertainty in these parameter values can result in ground water management designs that are not reliable. The degree to which uncertainty affects the reliability is a function of the large scale differences in the geologic environments of the physical systems. Optimization, statistics and model simulations provide the tools for which the physics of ground water flow models are shown to be an important consideration in determining the relevance of uncertainty in determining cost effective reliable remediation designs.

The spread of contaminated ground water is often prevented by pumping ground water out of the ground, thereby influencing ground water flow. When contaminated ground water is extracted from the ground it is treated, or purified, then safely discarded. Remediation systems that operate in the manner are referred to as pump-and-treat designs. The installation and operation of pump-and-treat designs are costly, and so optimization methods are commonly used to determine cost effective designs.

A parameter value of a ground water system that greatly influences the dynamics of flow is the material property of the ground known as the hydraulic conductivity, K [14]. Hydraulic conductivity defines the rate at which water, under pressure, is able to flow through the ground. In one dimensional space this relationship is defined by Darcy's Law:

$$v = -K\frac{dh}{dx}, \tag{1.1}$$

where v is the specific discharge of water, or velocity of water flow $[LT^{-1}]$, K is the hydraulic conductivity of the medium $[LT^{-1}]$, h is the hydraulic head, or level of the surface of the ground water subject to atmospheric pressure $[L]$, and x is spatial location [L] [5].

It is difficult to measure hydraulic conductivity values in the field. While geologist can provide a modeler with a rough map of hydrologically distinct regions, the boundaries between these regions is often not precise. Further, it has been well documented that even in hydrologically similar regions there is spatial variability in the hydraulic conductivity [7]. The uncertainty of hydraulic conductivity has been the subject of a great deal of research related to ground water management [16,1,8,6,21]. Unlike in previous studies where the focus has been on the natural spatial variability of hydraulic conductivity, here we examine the effects of uncertainty in the boundaries of distinct hydrologic fields on the determination of reliable optimal cost pump-and-treat ground water remediation designs. Further, these effects are examined for regions with distinctively different geologic settings.

Regions with similar hydrologic properties characterized by similar hydraulic conductivity values are called hydrostratigraphic fields. In this work, a heterogeneous field is one where there is more than one hydrostratigraphic field. The locations and orientations of the boundaries of the hydrostratigraphic fields can have a significant impact on the reliability of management designs. These effects have been shown to exist in the determination of reliable pump-and-treat remediation designs for containment where each field is spatial variability but the boundaries of the fields are assumed to be fixed [16]. Here we examine the uncertainty in the location of the boundary itself for seven different models each representative of a distinctly different geologic environment.

Numerical ground water flow models for heterogeneous aquifers are built using a block-structured approach. Such an approach limits the possible geometries of the boundaries. Modeling complicated geologic environments, such as vertically sloping boundaries, alternating layers of sands and silt or fingering of sand and silt, multilayered numerical models requires fine discretization of the fields. The computational intensity of models with fine discretization is often prohibitive. In some geologic settings it is shown that the uncertainty associated with course block-structured models in heterogeneous fields does not significantly affect the reliability of an optimal pump-and-treat ground water remediation design [10]. While such a model ignores the finer details of the geometry of the boundary of distinct hydrologic fields, the overall geometry in the model with respect to the design constraints affects the degree to which a reliable ground water management design can be determined.

Point source data of hydraulic conductivity, such as the data gathered from water wells, provides sparse data from which hydraulic conductivity values are determined. In those areas where the hydraulic conductivity value is not known, geospatial statistical measures are used to interpolate K values. While spatial variability in K is a source of uncertainty in these models, it has been shown that the uncertainty in mean K values for heterogeneous fields is most significant when the contrast between the mean K values is high [21]. In this work only systems with highly contrasting mean K values for distinct hydrologic fields are considered.

The uncertainty in the locations of distinct hydrostratigraphic regions for seven different geologic environments is analyzed with respect to a pump-and-treat ground water remediation design. The determination of a set of scenarios that represent the uncertainty in these boundaries uses an approach that involves techniques in image processing to segment the hydraulic conductivity field [15, 3, 19]. Unlike support vector machines [2] and methods in cluster modeling [11] that are typically applied used to separate field of like data, the method presented assumes no prescribed geometries for the boundaries of the distinct K fields and all observations of K are regarded as true values. Purely geospatial statistical methods, such as Multiple Indicator Kriging [17] require that decisions be made about the statistical techniques that can influence the weight of specific data points in determining the location of the boundary. Geospatial statistical techniques are used to generate full field categorization data from the sparse data, then methods in image processing are use to determine the boundaries that define each of the realizations of different heterogeneous hydraulic conductivity fields in the set of multiple scenarios.

In examining the uncertainty effects in a ground water management design, one can take a purely mathematical approach whereby the quantification of the uncertainty is well defined and it is possible to calculate the uncertainty in the response [20]. These methods require that the uncertainty in the K is known. While such an approach is possible for fields where the hydrologic properties are uniform, in heterogeneous models where the boundary of the distinct hydrologic fields is uncertain, such an approach is not straight forward. In this work, a stochastic approach is taken in modeling the uncertainty in the location of the boundaries of distinct hydrologic fields for each of the geologic environments. Stochastic approaches are commonly used to interpret the effects of uncertainty on pump-and-treat management designs [18]. Here the boundary is modeled using a multi-scenario approach. The location of the boundary in each of the scenarios is determined using a method derived from techniques in image processing. The optimal pump-and-treat ground water remediation design is determined using a genetic algorithm (GA) for each of the scenarios. Statistical measures are used to draw conclusions about the optimal solutions with respect to the uncertainty considered in each model.

The specific objective of this work is to examine how uncertainty in a physical model is expressed in the reliability of a solution to an optimization problem. In particular, this work examines optimization problems where the objective function and the constraint functions are dependent upon the physical model. The application presented in this work is the determination of an optimal pump-and-treat ground water remediation design where there is uncertainty in the physical model that describes ground water flow.

Methods used to model ground water flow are presented in the context of identifying significant features of ground water flow systems that are relevant to pump-and-treat ground water remediation designs. An optimization problem that results in a least-cost pump-and-treat remediation design is formulated. Details of a GA are discussed to determine the solutions to the optimization problem. And the methods section concludes with a discussion of how uncertainty in the ground water flow model is represented in the optimization problem and how the solutions are analyzed. Following the methods section is a description of the different physical ground water flow models and the remediation design constraints analyzed in this study. The parameter values of the GA applied to these problems is also presented. The results of this work are divided into two sections, one where differences in the optimal solutions are observed for the different models where uncertainty is not considered, and another where uncertainty is considered. The main results are highlighted throughout the results section and the conclusions section summarized more broad key results of this work.

2 Methods

2.1 Ground Water Flow Model

A pump-and-treat ground water remediation design that contains the flow of contaminated ground water through manipulation of the hydraulic head values is the focus of this study. Determining a least-cost design that meets the imposed flow constraints for containment requires solving the partial differential equation that describes the dynamics of ground water flow in a fully saturated, three-dimensional, porous media:

$$\nabla(\mathbf{K} \cdot \nabla h) = S_s \cdot \frac{\partial h}{\partial t} + f, \tag{2.1}$$

where ∇ is the spatial differential operator, $\mathbf{K}$ is the hydraulic conductivity tensor [LT^{-1}], h is the hydraulic head [L], S_s is the specific elastic storage [L^{-1}], t is time [T] and f represents a source/sink (per unit volume) term [T^{-1}]. It is within the source/sink term, f, that the pumping rates are prescribed.

To solve for the hydraulic head in the flow equation (Equation (2.1)) defined by a given K field and pumping design in space and time, the ground water flow simulator MODFLOW-2000 is applied [9]. This simulator takes a finite difference approach to solve the flow equation. The problems examined in this work seek steady-state solutions, and the aquifers in the model are assumed to be confined. The implications of examining a steady-state solution are that $\frac{\partial h}{\partial t}$ is set equal to zero in Equation (2.1) and confined aquifers are those where the hydraulic conductivity values for each cell in the finite-difference mesh are constant. These are common assumptions to make for examining new methodologies applied to ground water management problems. Under these conditions the head response to pumping rates is linear allowing for a response matrix approach to be utilized [8]. This approach greatly reduces the number of calls to the numerical solver, MODFLOW-2000, for the flow equations, thereby reducing the computational intensity of the problem.

The response matrix approach to modeling the hydraulic head values for a management design is performed as follows: MODFLOW-2000 is utilized to calculate the hydraulic head values for each of the models under conditions where none of the wells are actively pumping. These values are called the ambient head values. The hydraulic head values are then calculated using MODFLOW-2000 for models where one well is assigned one pumping unit and all other wells are inactive. The difference between the ambient head value and the head value in response to one unit of pumping is the hydraulic head response associated with the active well. The head responses are calculated for each of the proposed wells in the management model. These response values are collectively called the response matrix. The total head response to a management design defined by pumping from multiple wells at variable rates is then calculated. These head values equal the sum of the individual head responses from each of the wells scaled by the pumping rate assigned to each of the wells. The total head response is then added to the ambient head values to obtain the steady-state head values for the model.

2.2 Optimal Cost Pump-and-Treat Design

The cost associated with the remediation design is the sum of the cost of installation of the active wells and the cost of remediating the volume of water extracted from each of the wells [4]. It is here assumed that the cost of remediating the extracted water does not change with time and is not dependent upon the concentration of the contaminant in the extracted water. Such an assumption is reasonable for short term projects where water is sufficiently remediated using methods that are low cost, not dependent upon the concentration of contaminant in the water, and where maintenance and operational costs of the wells is approximately linear with respect to the quantity of pumping. When it is necessary to implement a management design over a long period of time and when the technology necessary to sufficiently remediate the water is costly and is dependent upon the concentration of contaminant in the water, it is not reasonable to assume the cost per unit of pumping is constant.

A fixed number of wells are considered at specified locations. Flow constraints are placed upon the system so that the gradient of the hydraulic heads at specified locations is towards the wells. The optimization problem that minimizes the cost associated with this remediation plan is as follows:

$$\begin{aligned} &\text{Objective:} && \min \textstyle\sum_{i=1}^{n} (CA_i + Rq_i) \\ &\text{Subject to:} && g_j > \max g, j = 1, ..., m \\ & && 0 \le q_i \le \max q, i = 1, ..., n, \end{aligned} \tag{2.2}$$

where C is a real valued constant representative of a fixed installation cost for each active well, A_i is the activity function defined below, R is the daily cost of ground water remediation per unit volume of water per, q_i is the pumping rate at well i measured in units of volume per day, n is the total number of wells considered, $\max g$ is the maximum allowable hydraulic head gradient so that flow is towards the well at any of the m constraint locations, g_j is the gradient at the constraint location j given a pumping design of $q = (q_1, q_2, ..., q_n)$, and $\max q$ is the maximum amount of pumping from any of the wells. Here pumping at any specific well, q_i, must be positive. This condition implies that only extraction wells are considered (no injection wells are considered) in the resulting remediation plan. The activity function is given by the following:

$$A_i = \begin{cases} 0, & \text{if } q_i = 0 \\ 1, & \text{if } q_i > 0. \end{cases} \tag{2.3}$$

The value of the hydraulic head gradient, g_j, is defined in terms of the finite difference mesh utilized in the ground water simulator (Figure 1). At any of the constraint locations, j, the gradient, g_j, is modeled by the difference in hydraulic head values at adjacent nodes oriented such that flow from the outer node, k, to the inner node, $k-1$, would be towards the well. Then $g_j = (h_{j,k} - h_{j,k-1})/\Delta x$, where $h_{j,k}$ and $h_{j,k-1}$ represent the hydraulic head values at adjacent nodes of the mesh associated with constraint j, and Δx represent the spatial distance between the adjacent nodes.

2.3 Derivative-Free Optimization

A derivative-free method of optimization is used to solve this optimization problem since the objective function for this problem is discontinuous and hence not differentiable due to the activity function in Equation (2.3).

The maximum pumping constraint is taken into consideration in this problem by defining a discrete and finite set of permissible pumping rates for each well. The permissible rates are between 0 and $\max q$. By defining the set of pumping rates in this way, the decision space for this optimization problem consists of a large, but discrete set of variables. A genetic algorithm (GA) is employed to solve this modified optimization problem. Within the GA each member of each generation of the populations represents a possible remediation design plan, i.e. combination of pumping rates at the proposed well locations.

Here a minimum fitness value is sought. The fitness value is equal to the value of the objective function, plus a scalar multiple of the sum of the violations of the gradient constraints at the constraint locations should the given pumping design fail to meet the constraints (Equation (2.2)). This fitness value is thus defined by the following equation:

$$\sum_{i=1}^{n}(CA_i + Rq_i) + \omega \sum_{j=1}^{m} \max\{0, \max(g - g_j)\}, \tag{2.4}$$

where all variables are as stated previously, and ω is a penalty weight. By defining the fitness value in this manner, the gradient constraints are considered in this optimization problem. While it is possible that the gradient constraints may be violated using this formulation, by setting the penalty weight appropriately, violations of the constraints are unlikely.

In the GA, subsequent generations are determined by employing the rules of elitism, mating priority, crossover, mutation and random selection. To generate a new population, two of the members of the mating population are randomly selected as a match. Random crossover locations are chosen to create a member of the next generation. The members of the new generation may undergo a random mutation, however the rate of mutation is small. It should be noted that some proportion of the new generation is not determined through mating and is randomly generated in the manner in which the initial population was generated.

The parameter values that define the GA used to solve for optimal management designs are determined by examining the reliability of the solutions. The reliability of optimal designs determined through a GA for a particular model is found by evaluating the set of solutions to the GA implemented with multiple randomly generated initial populations. Reliable solutions are those that are not dependent upon the initial population. The details of the parameters that define the GAs used for different models is presented in Section 3: Implementation.

2.4 Representation of Uncertainty

The hydraulic conductivity of the ground through which water flows in a ground water aquifer is not uniform. Further, hydraulic conductivity measurements often only exist at sparse locations in a given region where water wells allow for the testing of the flow properties. In order to solve the ground water flow model using a finite difference approach, however, it is necessary to specify at every cell in the finite difference mesh a fixed hydraulic conductivity value, K. Assignment of K values where no data have been collected introduces uncertainty into the solutions to the flow model.

Heterogeneous aquifers considered in this work are those where the region being modeled is characterized by contrasting hydraulic conductivity fields called hydrostratigraphic units. An example of contrasting fields is a unit consisting of clean sand, where water flows with ease (high K), and a unit consisting of sand mixed with silt, where the fine grains of the silt block fluid flow (low K). Geologists understand that observed contrasting hydrologic properties measured at different ground water wells indicates that the wells are located in distinct hydrostratigraphic units separated by sharp boundaries. Mapping these boundaries based upon sparse data is typically not precise.

The uncertainty considered in this work is the uncertainty in the locations of the boundaries of distinct hydrostratigraphic units. This uncertainty is represented using a multi-scenario approach, whereby multiple realizations of the K field are generated using assumed known information that characterizes the uncertain parameter. The fields representative of the uncertainty in the boundaries of the hydrostratigraphic regions were generated using a data driven approach. First the data are classified into distinct fields and each of the data points is assigned an integer-indicator value associated with the field to which it belongs. The classification indicator values are then interpolated over the entire field and a contour technique is used to determine the boundary between the contrasting fields.

Segmentation methods refer to those methods in image processing whereby the colors assigned to each pixel in an image are categorized by particular groups. Each category or group is usually representative of some range of colors on a given spectrum, or continuum, of colors. A single color, called an index color, is then assigned to each category. Segmentation of the image occurs when index color for each category is assigned to each pixel of the image according the pixel's original membership in the said category [19].

To apply segmentation methods in this context, the number of hydrostratigraphic regions, n_r, is determined as follows. Each hydrostratigraphic unit is assigned an integer value ordered in accordance with the ordering of the mean K value for the unit, $I_1, I_2, ..., I_{n_r}$. These integer values are analogous to the index colors described above for segmentation of an image. For example, if there are 4 units with mean K values of 0.001, 0.01, 0.1 and 1.0 md^{-1}, the unit with mean K of 0.001 md^{-1} would be assigned the category value of $I_1 = 1$, the unit with mean of 0.01 md^{-1} the value of $I_2 = 2$, etc. If the hydrostratigraphic fields are representative of distinct geologic regions, such as a sand unit and a silt-sand unit, this process of categorizing the data is not difficult because the K values will differ by orders of magnitude and membership in a particular field, for example sand or silt-sand, will be easy to determine. If the hydrostratigraphic units are not distinct, then techniques for determining the modes of a multi-modal distribution must be employed and membership of the data to a given distribution must be based upon likelihood measures [12].

To determine the location of the boundaries, the integer-indicator values are interpolated for all nodes of the finite difference mesh where data does not exist using universal kriging. Universal kriging utilizes a regression model to obtain a surrogate for the K category field. Here a linear regression model is utilized with a Gaussian correlation function, $\mathcal{R}(\theta, d)$, i.e. a Gaussian correlation model is utilized. The Gaussian correlation function is of the following form:

$$\mathcal{R}(\theta, d) = \prod_{j=1}^{N} \exp(-\theta_j d_j^2), \tag{2.5}$$

where θ is the correlation factor, here defined to be 1 for for each cell in the finite difference mesh, $j = 1..N$, and d_j is the distance from a fixed location in the model and each cell in the finite different mesh.

The number of candidate well locations in the remediation design, n, is equal to the number of cells in the finite difference mesh, N, only when all well possible well locations are considered. Typically optimization problems for pump-and-treat ground water remediation designs do not consider all cells within a model as candidate well locations and n is less than N.

An upper bound on the correlation distance is set to be 5 nodes within the finite difference mesh. In the physical models, this distance equates to 100 m. The Matlab kriging toolbox, DACE, is used in this application [13]. Through kriging in this manner, each node of the finite difference mesh is assigned a numeric value. These values may not be integers themselves but they fall within the integer category values assigned to each data point.

Different contour curves of the category field are then used to generate multiple realizations of heterogeneous K fields that represent the uncertainty of the boundary locations of the distinct hydrostratigraphic units. The location of the contour associated with the indicator value that is midway between the initial integer-indicator values is said to represent the boundary location that is most likely. For example, if two indicator values are 1 and 2, then the most likely contour would be associated with the interpolated indicator value of 1.5 in the field. To generate a set of realizations representative of the uncertainty in the boundary locations, a truncated normal distribution with a mean of zero and standard deviation of one is randomly sampled. The contour level, c_{bnd}, that determines the boundary between the hydrostratigraphic fields associated with the two categories is given by the following equation:

$$c_{bnd} = \frac{r}{\varsigma}(c_\alpha - c_\beta) + \frac{(c_\alpha + c_\beta)}{2}, \tag{2.6}$$

where c_α and c_β are the integer-indicator values assigned to physically adjacent region, α and β in the space, r is the random value sampled from the normal distribution and ς is a parameter that scales the random variable in such a way that the resultant c_{bnd} is between c_α and c_β. The assignment of a value to ς is related to the level of uncertainty in the location of the boundary where higher values relate to less uncertainty. In this study ς is equal to the value 4.

2.5 Stochastic Analysis

For each of the models, A through G, 100 K fields are generated to represent the uncertainty in the boundary between the low and high K fields. The least-cost pump-and-treat ground water remediation design is determined by solving the optimization problem using a GA for each of the 100 realizations. As such, for each model 100 optimal pump-and-treat remediation designs are determined. These sets of remediation designs are analyzed and compared using statistical measures.

The sets of 100 remediation designs generated for each model are differentiated by three quantitative measures: the number of active wells, the locations of the active wells and the pumping rates assigned at each of the active wells in the remediation designs. These sets of remediation designs are grouped according to subsets of designs where the active wells within each subset are the same. For example, if the solution results for a model contain remediation designs where wells 1 and 3 are active in 30% of the designs and the remaining 70% of the solutions are designs characterized by active wells 1, 2 and 4, then the total solution set contains two subsets of designs. Each of the subsets within the solution set is assigned an integer value that is herein called the "design index" for the model. The reason it is necessary to make a distinction between designs with different active wells is that the distance between potential well locations in the models is such that statistical comparisons between pumping rates at different wells is meaningless.

The design index provides a label for each subset of the set of remediation designs determined for each model. Associated with each design index are the number of wells that are active in the design and the well number for each active well indicating the location of the well in the model (Figure 2). The well labels are the same for all models and are assigned based upon the well location. The mean pumping rates are determined for the individual active wells within each of the design subsets. The standard deviation of the pumping rates is also calculated for these wells and is used as a rough measure of the sensitivity of the remediation design to the uncertainty of the physical parameter of the model within the design subset.

The response of the pumping rates within each design to the uncertainty in the physical parameter in the model using statistical measures. In this analysis the correlation coefficients and P-values are used to compare the optimal pumping rates at each well in each design with the contour levels, c_{bnd}, (Equation (2.6)) that define the uncertainty in the K field in each model. When P-values are calculated to be less than 0.05, it is believed that the correlation between the compared sets, quantified by the correlation coefficient, is significant.

Different optimal remediation designs for each model are defined by active wells at different locations within the model. The frequency of each design in the full set of design solutions is determined. This frequency is measured in percent of the total number of solutions. Further, for each design the range of contour levels that define the K fields in the associated models are determined. When the range of contour levels for distinct designs is disjoint, it is possible to conclude that categorization of solutions by design is a function of the contour level. If the range of contour levels for different designs overlap, it is possible that multiple locally optimal solutions exist in the optimization problem.

The parameters for the GA used to solve the optimization problem were set so that the algorithm is 80% reliable for all of the models examined. Since the GA is not 100% reliable, no consideration is given to solutions that appear to be suboptimal such as those designs associated with anomalously high operational costs.

3 Implementation

3.1 Ground Water Flow Model

Seven hypothetical ground water flow models were generated for this study. These seven models contrast the physical features of the geology with the parameters for the flow model as well as the parameters of the remediation design. It is through these seven models that it is possible to draw associations and conclusions regarding the effects of the uncertainty in the boundary of the hydrostratigraphic units in different geologic settings.

The seven models are all single layer models representative of a region that is 1000 m by 1000 m in size and 100 m thick. No flow boundaries exist on the northern and southern boundaries of the aquifer, while constant head boundaries of -20 m and -25 m are set at the western and eastern boundaries, respectively. Aquifer vertical recharge is not modeled here. The models are all of confined aquifers and the flow equation is run to simulate a steady-state condition. A uniform finite difference mesh of 50 cells by 50 cells is defined over the given area so that each cell represents a 20 m by 20 m region in the aquifer (Figure 2). The seven models represent seven different geologic settings, each with two distinct hydrologic regions. These seven models can be grouped into two comparison groups. In the first group (Figure 3) there are three environments where the modeled regions consist of two hydrologic zones that are roughly the same size. The locations of low K and high K zones differ in these models, as does the orientation of the boundary between these zones with respect to the ambient flow fields. In the second comparison group (Figure 4) there are four different settings also with two distinct hydrologic regions. In this group the zones of low and high K are not equal in size and so the effects of a smaller zone of high or low K in a model that is primarily of a contrasting K value are examined. These models differ in the locations of the boundary between the contrasting K values, as well as the K values assigned to the defined regions in the model with these K values.

A hypothetical data set of K values is created for each of the seven models. This set consists of 13 randomly generated values, sampled from distribution fields representative of observed uncertainty in K values that have been measured in the field (Table 1) [7]. Two units with distinct hydrologic properties are defined for each model. The units are geologically representative of clean sand characterized by a mean K value of 1.0 md^{-1} and a silt-sand by a mean K value of 0.01 md^{-1}. The locations of the hypothetical measurement are the same for all the models (Figure 2), but the values of K measured differ. Utilizing the segmentation algorithm, the most likely location of the boundaries of the hydrostratigraphic units for each of these models are illustrated in Figures 3 and 4.

A set of 100 realizations of the K field are generated using the segmentation algorithm for each model. By performing the analysis on sets of increasing size, it was determined that the statistical characteristics of the solution sets do not change when more than 100 scenarios are considered. For this reason it is concluded that an ensemble size of 100 is sufficiently large to represent the uncertainty in the location of the boundaries of the units in each model. The integer value assigned to those observation locations within the region of low K is 1, while the integer representing the high K value is 2. When the randomly sampled contour level, c_{bnd}, determined is greater than 1.5, the boundary of the high K regions is closer to the observation points where high K is measured, thereby reducing the area where high K values are assigned in the model. When the contour level is below 1.5, the boundary is closer to the low K observation points, thereby reducing the area where low K values are assigned. Examples of different boundary locations for models B and E are illustrated in Figure 5. It is significant to note how the proximity of the boundary to potential well locations in the remediation design changes as c_{bnd} changes.

3.2 Remediation Design

A least-cost pump-and-treat remediation design is sought for each of the models that will reverse ambient ground water flow from the west towards the east along 14 gradient constraint locations. These constraint locations are positioned at every other node along a 550 m northerly to southerly transect that is located 280 m from the eastern boundary of the model (Figure 2). Ten possible well locations are positioned up-stream of the gradient constraints. The installation cost applied to a well, A_i, that is activated is 500 dollars. Pumping rates are bounded below by a rate of 0 m^3d^{-1} and above by a rate of 500 m^3d^{-1}. The daily cost for remediation per unit pumping, R, is equal to 1. The cost for remediation and the installation cost are artificial in this example. Their values are set such that the introduction of a new well will be considered in an optimal remediation design only when maximum pumping is not sufficient from a remediation design with fewer wells. Ground water flow reversal is assumed to be achieved when the difference in hydraulic head values at adjacent nodes along the constraint location transect in the direction of the wells is greater than 0.01 m. This implies that at the outer cell of the coupled gradient constraint pairs, the hydraulic head is 0.01 m greater than the inner cell, ensuring that ground water flow is in the direction necessary for flow towards the wells. The parameter values of the optimization problem in Equation (2.2) are summarized in Table 2.

3.3 Parameters of the genetic algorithm

There are ten potential well locations in the remediation design, and so there are ten decision variables in the optimization problem. The application of the genetic algorithm (GA) necessitates the construction of a discrete and finite set of possible pumping rates for each of the wells. For these problems the same set is utilized, namely q_i may assume the following set of values: 0, 5, 10, 15,. . . , 495, 500. Given these parameters of the optimization problem, there are 101^{10} possible pumping designs for this problem.

The penalty weight, ω, for violations of the constraints is 10^5 (Equation (2.4)). This value is one where the numeric value of the costs associated with the pumping and treating of the ground water are at the same scale as the violations of the ground water flow constraints. It is a value that is prohibitive of violations of the constraints, but is still such that variations in the pumping rates and hence operational costs, are optimized in this problem.

The members of the population carried to the next generation through elitism consist of the top 10% of best fit members of the given population. The mating population consists of the top 50% of the best fit members, including the elite. Mating is conducted by applying one cross-over event with subsequent mutation occurring in 0.05% of the resulting pumping rates. Mating accounts for 70% of the population of the next generation. The remaining 20% of the next generation of pumping rates are created by randomly sampling the set of possible pumping rates for each well. The GA creates new generations of populations of pumping combinations for the wells until after 25 generations there is no change in the determination of a best fit pumping design for all of the members of the population.

The reliability of the GA is sensitive to the initial population size. To determine an initial population size where the results of the GA return the same solution in 80% of the instances, implying that a stable locally optimal solution is obtained, different initial population sizes were considered. Using 20 different randomly generated initial populations, the GA is applied to the seven geologic models where uncertainty is not considered. In four of the models (Models B, C, D and E) an initial population size of 10,000 is such that the GA returns the same result in all 20 runs, i.e. is 100% reliable. In three of the models (Models A, F and G) an initial population size of 25,000 is needed to obtain 100% reliability. Only in Model F is an initial population size of 25,000 less than 100%, and in this case the reliability is found to be 80%. For all models with and without consideration of uncertainty, an initial population size of 25,000 is used in the GA.

4 Results

Solving optimization problems where the objective and constraint functions are dependent upon solving for a physical model using a numerical simulation are computationally expensive. When uncertainty in the model is taken into consideration using multi-scenario approaches, it is advantageous to understand the effects of uncertainty on the solutions to the optimization problems. The results of this work illustrate how the physical properties of a model, namely a ground water model, play an crucial role in interpreting and quantifying the effects of uncertainty on optimization results, used here for determining a reliable ground water remediation system.

4.1 Geologic Variations for Models with No Uncertainty

When uncertainty is not considered, optimal remediation design plans determined for the seven hydrologic models provide insight into the effects of large scale differences in the physical model. In these models the hydrostratigraphic fields of a heterogeneous aquifer are perfectly homogeneous with hydraulic conductivity values of 0.01 md^{-1} and 1.0 md^{-1} and the boundaries of these fields are fixed at the most likely locations as depicted in Figure 3 and 4. The least-cost pump-and-treat remediation solutions for these models are given in Figures 6 and 7.

Result 1: There is a direct relationship between the amount of pumping required at a well to meet the flow constraints and the K value where the wells and the gradient constraint are positioned.

Rationale for Result 1: It is clear that the geologic environments represented by models A-F require different pump-and-treat ground water remediation designs depending upon the locations of the gradient constraints and the proposed well locations with respect to the boundaries of the distinct hydrologic fields.

For example, the remediation design for Model F clearly illustrates how the ground water flow at the gradient constraint locations is affected by extraction of ground water in regions of low K versus high K. In Model F the south-eastern quadrant of the model is characterized by a low K region while the surrounding quadrants are characterized by high K values (Figure 4). The ambient groundwater flow fields in Model F at the gradient constraint locations are similar (Figure 9). To satisfy the desired flow constraints in the remediation design, pumping from wells 7, 9 and 10 with pumping rates of 90, 455 and 120 m^3d^{-1} are required respectively. Well 7 is located in the low K region where less pumping is necessary to satisfy the gradient constraints than at well 9, located in the high K region. Pumping at well 10 supplements the drawdown of the hydraulic head needed to meet the flow constraints in the high K region.

Result 2: There is an indirect effect of pumping at a well on the flow at the gradient constraint locations caused by changing the hydraulic head values at the *boundaries* of hydrostratigraphic regions where the ambient ground water flow gradient is small.

Rationale for Result 2: The constant head boundaries placed upon the western and eastern boundaries of the models directly affect the ambient ground water flow within the interior of the model (Figures 8 and 9). If the K fields are perfectly homogeneous along the ground water flow path, then the ambient flow would be uniform across the model. Such a field is depicted in Model A where the flow path is from west to east. In Model A, despite the fact that the region to the north is characterized by low K values and the region to the south is characterized by high K values, along any flow path that runs directly west to east the hydraulic conductivity is the same.

When the boundary between the regions of high K and low K are not co-linear with the flow path, the ambient flow fields in the interior of the model are not uniform. Ground water flows more quickly through high K fields than low K fields. For models where the boundaries of a connected high K field include two types, a single valued constant head boundary and a unspecified head value such as that defined by an adjacent low K field in the interior of the modeled region, the hydraulic head variations within the high K field are much less than the head variations in the adjacent low K field. Examples of such high K fields are illustrated in Models B, C, D, and E (Figures 3 and 4). The most efficient way to change the flow in these high K fields is to pump ground water out of wells in the surrounding low K field so that the head values in the cells interior to the model that bound the high K region are such that the gradient constraints are satisfied. Pumping within the low K regions directly effects the heads along the delineating boundaries of the low and high K regions, resulting in a secondary effect of changing the flow within the high K region so that the gradient constraints are satisfied.

It should be noted that it is generally not prudent to pump directly from low permeable regions of the aquifer because this inevitably increases the drawdown at the well. As drawdown increases, greater effort is needed to pump water from a greater depth out of the well. Assuming the cost of pumping is linear for these cases is not realistic and is a limitation of this model. Further, increasing the drawdown significantly at a well may result in lowering the hydraulic head to the point where it is below the top of the aquifer. If this happens, then the flow does not occur in a confined aquifer, as is assumed in these models. Making an incorrect assumption about confined versus unconfined flow can result in highly inaccurate predictions of the hydraulic head. While these affects were not considered in the analysis of the affects of uncertainty on remediation design in this work, future models should consider avoiding such problems by the addition of a minimum head constraint at the wells in the optimization problems.

In models C and D the gradient constraint are located in high K regions where the constant head boundary values assume a single value. In these models the optimal pumping designs are ones where the wells are located solely within the low K regions adjacent to the high K regions. In fact the only models where pumping wells are located in a high K field are those models where the gradient constraint locations are within high K fields bounded by constant head boundaries characterized by different constant head values (Models A, F and G).

Results summary: Optimal pump-and-treat ground water remediation designs are defined by the locations of active pumping wells and the rate at which water is pumped out of the ground. Both the well locations and pumping rates are functions of three factors. First, they are directly related to the manner in which the cost function in the optimization is constructed. Often, this is the only concern for those applying optimization techniques to ground water management problems. But two overlooked factors are those related to the physical model itself. There is a direct relationship between pumping and ground water flow at the gradient constraint locations. This direct relationship is due to the hydraulic conductivity value of the field where the active well and gradient constraints are located. There is also an indirect effect of pumping on the flow at the gradient constraint locations caused by changing the hydraulic head values at the boundaries of contrasting K fields in the model.

4.2 Uncertainty in the Boundary

There are three features that define an optimal pumping design for each of the models: the number of active wells, the locations of the wells and the rate at which the wells are pumping. In each of the models, changes in the location of the boundaries between low and high K fields results in changes in the resultant optimal pump-and-treat remediation design. The effects of the uncertainty in the boundaries of the regions for different geologic environments is assessed by examining multiple optimal solutions applied to multiple scenarios generated to represent the uncertainty. The following results indicate that the geologic environment plays a significant role in determining the effects of uncertainty.

The optimal remediation design results applied to the multiple scenarios generated to represent the uncertainty for each of the geologic models are summarized in Table 3.

4.2.1 Well Locations

Result 3: If all of the solutions determined in the stochastic analysis result in remediation designs with the same numbers and locations of active wells, then the uncertainty in the boundary between contrasting hydrostratigraphic fields is expressed in the variability of the pumping rates at the active wells only. From a design perspective these cases are ideal because there is no uncertainty capital costs associated with the installation of the wells because the number of active wells and locations of the wells can be determined with 100% certainty.

Rationale for Result 3: The results summarized in Table 3 indicate that there is only one design index associated with the solution sets for Models A, B, C and E. This result implies that uncertainty in the boundary between high and low K fields does not affect the number of active wells or the well locations in the optimal remediation designs for these models. Only the pumping rate at each of the active wells is affected by uncertainty.

Result 4: When the physical effects on the optimal pumping design *are* affected by uncertainty in the boundary between contrasting hydrostratigraphic fields, then uncertainty in the optimal designs is expressed in the pumping rates at the active wells as well as the number and locations of the active wells.

Rationale for Result 4: Multiple Design Indices are listed for the optimal remediation designs determined for the multiple scenarios generated to represent the uncertainty in Models D, F and G. For each of these models, changes in the boundary locations resulted in changes in the direct and indirect effects of pumping on the ground water flow at the gradient constraint locations.

In Model D, as the contour level, c_{bnd}, increases, the boundary moves closer to the region where high K values are observed, causing the area of the high K region to decrease. To achieve the desired head gradient within the high K region most efficiently, the active well located in the adjacent low K region, must be as close to the boundary as possible. And so as the boundary moves closer to the observation points of high K, the activation well moves from well 3 to either well 2 or well 8.

There are two possible pumping designs for Model G, both with two active wells pumping in the high K field. In both designs well 10 is aggressively pumping thereby forcing water to flow towards the wells. The no-flow condition along the northern boundary of the model invokes a uniform gradient of groundwater flow along the northernmost transect of the model. This boundary effect is extremely difficult to overcome and so pumping at well 10, closest to this boundary and closest to the gradient constraints is needed.

In the interior of the Model G, aggressive pumping from either well 3 or well 8 is used to meet the gradient constraints further south. Along the southern boundary there is a no flow condition, however, the low K region in the southwestern region impacts the gradient in the high K field. The groundwater flow along the southernmost flow path from west to east is not uniform. Along this path a much higher ambient head gradient exists in the low K region than in the high K region. Since the ambient gradient in the high K region is smaller in the south than in the north, pumping from a centrally located well, such as well 3 or well 8, will reach the desired gradient constraint condition. The decision to pump from well 3 or 8 is related to the location of the boundary since there is an indirect effect on the gradient at the constraint locations in the high K field in the southern region of this model.

In Model F, four different remediation designs, indicated by four design indices in Table 3, are determined for the scenarios associated with different contour levels, c_{bnd}. For most scenarios three wells are required to be active to meet the flow constraints. In cases where the boundary is closer to the observed low K regions, the gradient constraints in the high K region require substantial pumping to meet the flow constraints at wells 8 and 10. While wells 8 and 10 affect the gradient in the high K field, additional pumping is required to affect the gradient in the low K field. Recall that effects on the boundary within the interior of the model primarily occur from pumping in a low K region and here wells 8 and 10 are located in a high K region. Pumping from either well 6 or 7 will equally satisfy the gradient constraints in the low K field.

As the contour level, c_{bnd}, increases and the boundary location moves towards observations points with high K values in Model F, the active pumping well in the low K region is located centrally at well 7, while the wells in the high K region are positioned at well locations 9 and 10. When the contour level is very high, the design only requires that two wells are active, namely wells 8 and 10. At this point, well 8 is located within the low K

field (with lower contour level it is in the high K field). Despite the fact that low pumping is generally needed to influence ground water flow in low K fields, pumping from well 8 remains aggressive since the ambient head gradient in the low K region is greater in Model F than in it is in Models B and E. Aggressive pumping is required meet the desired flow constraints.

Results summary: The number of active pumping wells and the location of these wells in an optimal pump-and-treat ground water remediation design is not affected by uncertainty in the boundary between low and high K fields *if* the uncertainty does not change the direct and indirect effects of pumping on the hydraulic head gradient at the flow constraint locations. If the direct and indirect effects are impacted by changes in the boundary, then the number and location of the active wells in the optimal designs is impacted as well. Uncertainty in the boundary, in these cases, results in solutions where the number of active wells, the locations and the pumping rates are uncertain.

4.2.2 Pumping Rates and Reliability

The rate of pumping from each of the wells in the remediation design is almost always related to the K value at the well. Wells in low K regions generally pump less than wells in high K regions. Of course there are exceptions to this generality, as seen in pumping from well 8 in Model F when the contour level, c_{bnd}, is high. Statistical measures are used to examine how the uncertainty in the boundary between low and high K fields is related to the optimal pumping rates determined at the wells. The mean and standard deviation of the pumping rates for the different optimal designs for each model are reported in Table 3. The correlation coefficient and P-values between the pumping rates and the contour level values are also reported in Table 3 to determine the likelihood that the changes in pumping are related to the uncertainty.

Result 5: Low variance in pumping rates is not always an indication of a reliable design.

Rationale for Result 5: The pumping rates for optimal designs determined for Models A, C and D when uncertainty is considered are all characterized by low standard deviations that are less than 10% of the mean rate. Due to the discretization of the pumping rates used in the GA, drawing conclusions about the changes in pumping rate with the changes in the contour level values, c_{bnd}, as indicated by the correlation coefficient and the P-value are questionable. But the resultant designs for Models A and C have only one design index, indicating certainty in the well locations for the design. Since the standard deviation of the pumping rates are low for these models it can be concluded that these designs are reliable. No further measures are necessary to better define the boundary in these models.

The optimal designs determined for Model D, however, have three possible designs that differ in the location of the remediation well with respect to the gradient constraints. In these designs the location of the well is shown to be dependent upon the contour level, c_{bnd}, and so there is risk associated with determining the placement of the well. Since wells are expensive to set and because the location of the well is highly dependent upon the contour level, efforts should be made to better classify the location of the boundary.

Result 5: High variance in pumping rates may be attributable to the uncertainty in the boundary. The physics of the model is a factor in making this assessment.

Rationale for Result 5: In Models B and E the standard deviation of the pumping rates is greater than 35% of the mean. There is only one optimal design however for each of these models, i.e. the optimal location of the pumping wells is the same for all scenarios. In these models there is a very strong correlation between the pumping rates and the contour level, c_{bnd}, and this correlation is significant as indicated by the P-value being well below 0.05.

These statistical measures indicate that the pumping rates in these designs are a function of the location of the boundary between the low and high K fields. In these cases it would be strongly suggested that further geophysical data be collected to reduce the uncertainty in the boundary location so that reliable, cost effective pumping can be employed.

In Model G two remediation designs are determined for all values of the contour level, c_{bnd}. The standard deviation of the pumping rates determined at well 10 is low and the correlation with the contour level is not significant. These observations support the conclusion that pumping from well 10 not driven by the boundary location, but rather by the no flow condition placed along the northern boundary of the modeled region. The high correlation coefficients and low P-values for the pumping rates at wells 3 and 8, interior to the model, suggest that the location of the boundary has a significant effect on the remediation design. In this model 84% of the designs invoke pumping at wells 3 and 10 so it may be recommended in this case that this design be implemented and revised once an observation of K is determined at well 10.

Result 6: The geometry and uncertainty in the location of the boundary between contrasting K fields results in very different expressions of the uncertainty in the remediation design when the contrasting K fields are reversed.

Rationale for Result 6: Model D and Model F are characterized by an almost homogeneous K field with a contrasting region in the southeast corner of the model. In Model D the field is primarily one of low K with the contrasting field having high K values, while in Model F the field is primarily one of high K with a contrasting field of low K. The optimization results with the consideration of uncertainty for Model D resulted in a single well design with a pumping rate that was known with certainty. The results for Model F, however, produced the most variable results of all of the models examined in this study.

The optimization designs obtained when multiple scenarios were used to represent the uncertainty in Model F resulted in four possible designs. The standard deviations of the pumping rates are generally large in all of these designs (Table 3). By examining the statistical correlations between the pumping rates and the contour levels, c_{bnd}, used to obtain these results it was found that in some of these designs the pumping rates are dependent upon the contour levels. These dependencies were observed in the design associated with design indices 1 and 2. In designs associated with design indices 3 and 4 the statistics indicate that there is still ambiguity over the dependence of pumping upon the contour level.

The location and uncertainty in the boundary between the low and high K values in Model F are exactly the same as those observed in Model D, but in Model D there is a small region of high K in a predominantly low K field. While the standard deviations of pumping in Model D were lower than those observed in Model F, recall that in Model D three possible designs were determined. To determine a low risk remediation design for a geologic environment similar to that in Model F would require that further information about the location of the boundary between the low and high K be obtained.

Results summary: The uncertainty in the boundary between low and high K regions in a ground water model results in uncertainty in optimal cost pump-and-treat ground water remediation designs. The uncertainty in the remediation designs, however, is not always correlated with the uncertainty in the model. Unlike optimization problems where uncertainty a specific parameter of the problem arises in the mathematical formulation of the problem, the indirect uncertainty in the model as it relates to uncertainty in the optimization problem requires that the modeler understand the physics of the model to best interpret the uncertainty results.

5 Conclusions

The research presented herein examines the uncertainty in the boundary between distinct hydrologic regions in a ground water model characterized by contrasting hydraulic conductivity values for the development of optimal pump-and-treat remediation designs. Different geologic environments are presented in this study where the uncertainty in the boundary of the low and high K fields is considered. The uncertainty is represented using a multi-scenario approach whereby each scenario represent a K field in a given environment with a different boundary location as defined by a contour level. The technique used to determine the boundary location is based upon image segmentation method approaches coupled with contour methodologies. A genetic algorithm is used to solve the optimization problem whose objective is to determine a least cost pumping design. Installation costs are considered in this model. The results of these studies indicate that reliable ground water remediation designs are affected by the geometry of different geologic environments in highly variable ways.

The results of this research indicate that while the uncertainty in all of the models considered is due to the uncertainty in the boundary between the low and high K fields, the physical model plays a substantial role in how this uncertainty affects the uncertainty in the optimization results. In these ground water remediation design problems having a clear understanding of the physics of the model, it is possible to draw conclusions about how the optimal remediation designs are dependent upon the geologic environment. By classifying physical features of the model related to the decision variables and constraints in the optimization problem, observations that involve the ground water flow model provide insight into an interpretation of the optimization results that go well beyond a purely statistical interpretation.

This work involves the interpretation of seven distinct physical models and while it is possible to draw conclusions about how the geometries of the physical models affect the optimization results, further investigations are necessary to quantify mathematically these responses.

Statistical measures are useful tools for examining the correlation between the imposed uncertainty and the uncertainty in the optimal results. When physical models are considered and the uncertainty is such that uncertainty can change the dynamic of the modeled system dramatically with respect to changes in the decision variables, information provided by statistical measures may not be sufficient for making decision for reliable designs. This work has shown that such dramatic changes are possible in applying optimization methods to obtain optimal pump-and-treat ground water remediation designs.

Using mathematical models as prediction tools for physical systems is common practice in optimization problems. If there is uncertainty in the physical system, multi-scenario approaches are often used to generate multiple optimal solutions from which reliable results are obtained. This research suggests that before one takes this approach, they look beyond the multi-scenario results to determine whether or not the features of the physical model can be used to understand how uncertainty in a physical model is expressed in the results of an optimization problem that is dependent upon solving the physical system.

References

[1] D.A. Bau and A.S. Mayer, Optimal design of pump-and-treat systems under uncertain hydraulic and plume distribution, *J. Cont. Hydrol.* 100 (2008) 30–46.

[2] B.E. Boser, I.M. Guyon and V.N. Vapnik, A training algorithm for optimal margin classifiers, in *5th Annual ACM Workshop on COLT*, D. Haussler (ed.), ACM Press, New York, 1992, pp. 144–152.

[3] D. Comaniciu and P. Meer, Robust analysis of feature spaces: color image segmentation, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 750–755.

[4] U.S. Environmental Protection Agency (EPA), *Cost Analyses for Selected Groundwater Cleanup Projects: Pump and Treat Systems and Permeable Reactive Barriers*, EPA 542-R-00-013, ¡http://www.clu-in.org¿. February, 2001.

[5] R.A. Freeze and J.A. Cherry, *Groundwater*, Prentice Hall, Englewood Cliffs, 1979.

[6] R.A. Freeze and S. M. Gorelick, Convergence of Stochastic Optimization and Decision Analysis in the Engineering Design of Aquifer Remediation, *Ground Water* 37 (1999) 934–954.

[7] L.W. Gelhar, *Stochastic Subsurface Hydrology*, Prentice Hall, Englewood Cliffs, 1993.

[8] S.M. Gorelick, A review of distributed parameter groundwater management modeling methods, *Wat. Resour. Res.* 19 (1983) 305–319.

[9] A.W. Harbaugh, E.R. Banta, M.C. Hill and M.G. McDonald, *MODFLOW - 2000, the U.S. Geological Survey modular groundwater model - User guide to modularization concepts and the Ground-Water Flow Process*, U.S. Geological Survey Open-File Report 00-92 (2000) 121p.

[10] Hill, M. C., The Practical Use of Simplicity in Developing Ground Water Models, *Ground Water* 44 (2006) 775–781.

[11] L. Kaufman and P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, 1990.

[12] F.L. Gall, Determination of the modes of a Multinomial distribution, *Statistic & Probability Letters* 62 (2003) 325–333.

[13] S.N. Lophaven, H.B. Nielsen, J. Sondergaard, DACE: A Matlab Kriging Toolbox, Version 2.0, *IMM Informatics and Mathematical Modelling*, Technical University of Denmark, Technical Report IMM-TR-2002-12, 2002.

[14] B. Nilsson, A.L. Hojberg, J.C. Refsgaard and L. Troldborg, Uncertainty in geological and hydrogeological data, *Hydrology and Earth System Sciences* 11 (2007) 1551–1561.

[15] T. Pavlidas, *Structural Pattern Recognition*, Springer-Verlag, Berlin, 1977.

[16] K.L. Ricciardi, Remediation of heterogeneous aquifers subject to uncertainty, *Ground Water* 47 (2009) 675–685.

[17] J. Vann, D. Guibal, M. Harley, Multiple Indicator Kriging — Is it Suited to My Deposit?, in *4th International Mining Geology Conference*, Coolum, Qld, 2000, pp. 187–194.

[18] B.J. Wagner, Evaluating data worth for groundwater managment under uncertainty, *J. of Wat. Res. Pl. and Management* 125 (1999) 281–288.

[19] R. Wilson, *Image Segmentation and Uncertainty (Pattern Recognition & Image Processing)*, Research Studies Press, 1987.

[20] C.L. Winter and D.M. Tartakovsky, Groundwater flow in heterogeneous composite aquifers, *Wat. Res. Rsrch* 38 2002: Art. No. 1148.

[21] C.L. Winter, A. Guadagnini, D. Nychka and D.M. Tartakovsky, Multivariate sensitivity analysis of saturated flow through simulated highly heterogeneous groundwater aquifers, *J. Comp. Physics* 217 (2006) 166–175.

Manuscript received 3 October 2010
revised 14 January 2011
accepted for publication 22 March 2011

Karen L. Ricciardi
University of Massachusetts Boston, Department of Mathematics
E-mail address: Karen.Ricciardi@umb.edu

Table 1: Hypothetical values of hydraulic conductivity, $K(md^{-1})$, observed in the given numbered sample at the (x, y) coordinate, measured in m, for each of the modeled regions, K_{A-G}. The coordinate (0,0) is located in the south-west corner of the modeled region, and coordinates pairs where the x coordinate is positive are to the east of (0,0) while those where the y coordinate is positive are north of (0,0). These observation points are graphed in Figure 2, labeled by the Sample number.

Sample	x	y	K_A	K_B	K_C	K_D	K_E	K_F	K_G
1	100	800	0.0103	0.95	0.0095	0.0097	0.0100	0.958	0.934
2	140	120	0.96	0.92	0.0095	0.0095	1.05	0.975	0.0106
3	160	320	1.03	1.09	0.0098	0.0096	1.13	1.14	0.0112
4	240	660	0.0105	0.85	0.0102	0.0102	0.0098	1.08	0.954
5	260	880	0.0095	1.02	0.0099	0.0095	0.0111	1.04	1.014
6	360	220	0.95	1.01	0.0113	0.0106	0.95	1.07	0.0095
7	460	540	0.0076	1.08	0.0104	0.0104	0.0092	1.20	1.002
8	600	120	0.94	0.0076	1.07	0.98	0.0102	0.0094	0.948
9	620	820	0.0099	0.0089	0.992	0.0102	0.0103	1.03	1.14
10	760	340	1.12	0.0111	1.11	0.93	0.0094	0.0124	0.934
11	780	880	0.0116	0.0095	1.05	0.0106	0.0092	0.965	0.975
12	900	660	0.0092	0.0097	1.07	0.0103	0.0106	0.942	0.926
13	920	80	1.18	0.0113	0.967	0.99	0.0093	0.0121	1.014

Table 2: Values of parameters used to define the remediation design in Equations (2.2) and (2.4).

Description	Parameter	Value
Number of possible wells	n	10
Number of constraint location	m	14
Cost for well installation	C($)	500
Daily cost of ground water treatment	R ($)	1
Maximum pumping rate	max $q(m^3d^{-1})$	500
Maximum allowable gradient	max $g(mm^{-1})$	0.01
Penalty weight	ω	10^5

Table 3: Statistical results of the optimal pump-and-treat ground water remediation designs for each of the models. The correlation coefficients (Corr. Coef.) and P-values are between the pumping rates at individual wells and the contour level, c_{bnd}, used to generate the location of the boundary between the low and high K regions. The locations of the wells are graphed in Figure 2. The % of Solns is the percent of solutions in the solution set that have the associated statistical properties in the range of c_{bnd} indicated.

Model	Design Index	No. Wells	Well	Pump. Mean $m^3 d^{-1}$	STD	Corr. Coeff.	P-value	c_{bnd} range	% of Solns.
A	1	2	6	230	15	-0.1440	0.3184	all	100
			8	223	12	-0.4840	0.0004		
B	1	2	7	113	66	-0.8120	0.0000	all	98
			9	137	80	-0.8808	0.0000		
C	1	1	3	81	3	0.6463	0.0000	all	100
D	1	1	3	57	5	-0.2296	0.2398	< 1.53	56
	2	1	2	51	5	-0.7176	0.0195	> 1.53	20
	3	1	8	57	3	0.9622	0.0000	> 1.53	20
E	1	1	7	94	35	-0.9252	0.0000	all	100
F	1	2	8	349	76	-0.9798	0.0000	> 1.67	18
			10	335	55	0.9024	0.0009		
	2	3	7	77	69	-0.8849	0.0000	[1.5, 1.67]	28
			9	395	76	-0.4846	0.0791		
			10	145	34	0.7255	0.0033		
	3	3	7	83	44	-0.9134	0.0015	[1.31, 1.5]	16
			8	369	50	-0.5450	0.1625		
			10	286	26	0.0485	0.9092		
	4	3	6	22	9	-0.4455	0.1271	[1.31, 1.5]	26
			8	435	51	-0.8618	0.0002		
			10	258	25	-0.3757	0.2058		
G	1	2	3	344	60	-0.8856	0.0000	all	84
			10	207	29	0.1491	0.3460		
	2	2	8	327	53	-0.9810	0.0000	all	16
			10	213	14	0.7188	0.0445		

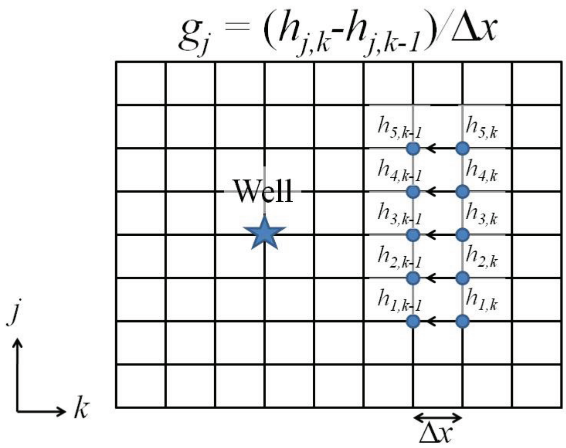

Figure 1: Calculation of the gradient is performed by taking the difference in the hydraulic head values, $h_{j,k}$, at adjacent cells in line with the desired flow path towards the well location(s), then dividing by the distance between these cells, Δx.

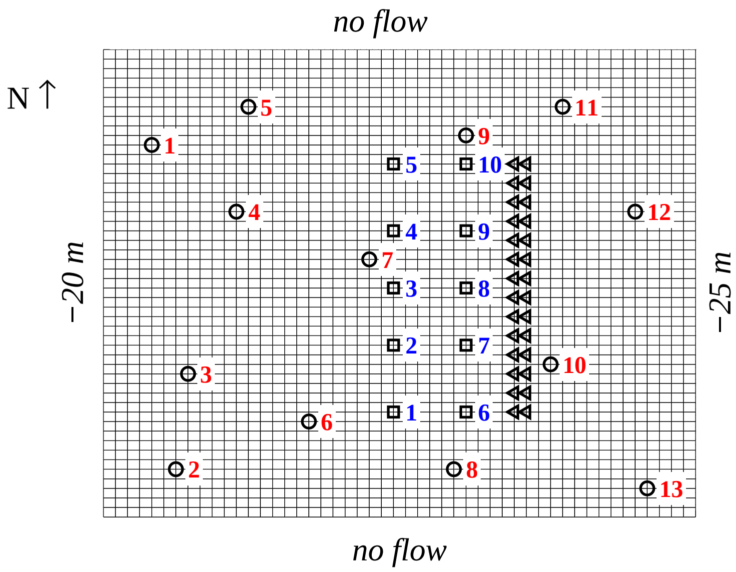

Figure 2: Hypothetical groundwater flow model. The squares represent the ten potential well locations for the remediation design and the associated well numbers; the triangles are located at the gradient constraint locations and are oriented so that they face in the direction of the desired flow; the circles are the locations of the thirteen hypothetical K measurements within the region (Table 1).

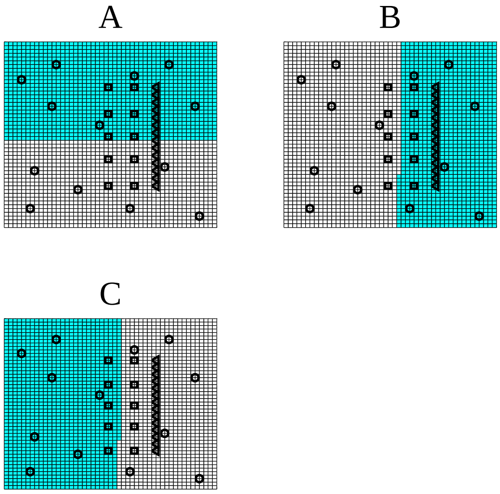

Figure 3: Models A, B and C are of heterogeneous fields with regions of low and high K that are roughly equal in size. In these models the boundary between the low and high K fields occurs at the location that is most likely, $c_{bnd} = 1.5$. The shaded region represents the area of low K with a value of 0.01 md^{-1}, while the white region represents the area of high K with a value of 1.0 md^{-1}. The circles represent the locations where the K values are observed (Table 1), the squares represent potential well locations in the remediation design and the triangles represent the locations of the gradient constraints.

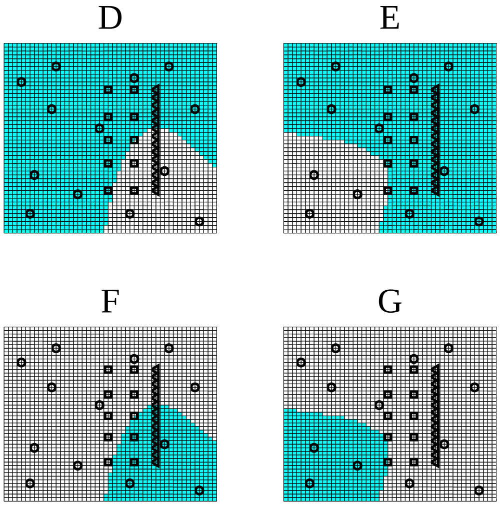

Figure 4: Models D, E, F and G are of heterogeneous fields with regions of low and high K that are not equal in size. In these models the boundary between the low and high K fields occurs at the location that is most likely. The shaded region represents the area of low K with value of 0.01 md^{-1}, while the white region represents the area of high K with value of 1.0 md^{-1}. The circles represent the locations where the K values have been observed (Table 1), the squares represent potential well locations in the remediation design and the triangles represent the locations of the gradient constraints.

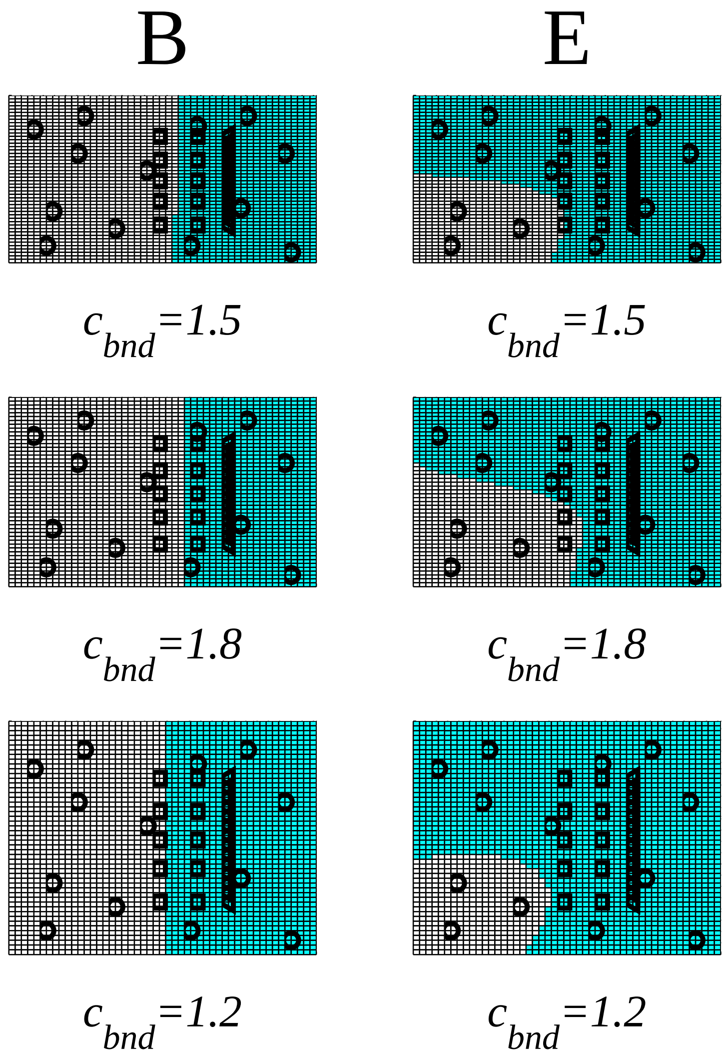

Figure 5: Three example K fields for models B and E generated so that the locations of the boundary between the contrasting K values is based upon randomly sampling the probability density function used to describe the uncertainty in the boundary location. The proximity of the boundary to the potential wells in the remediation design is different. The shaded region represents the area of low K with value of 0.01 md^{-1}, while the white region represents the area of high K with value of 1.0 md^{-1}.

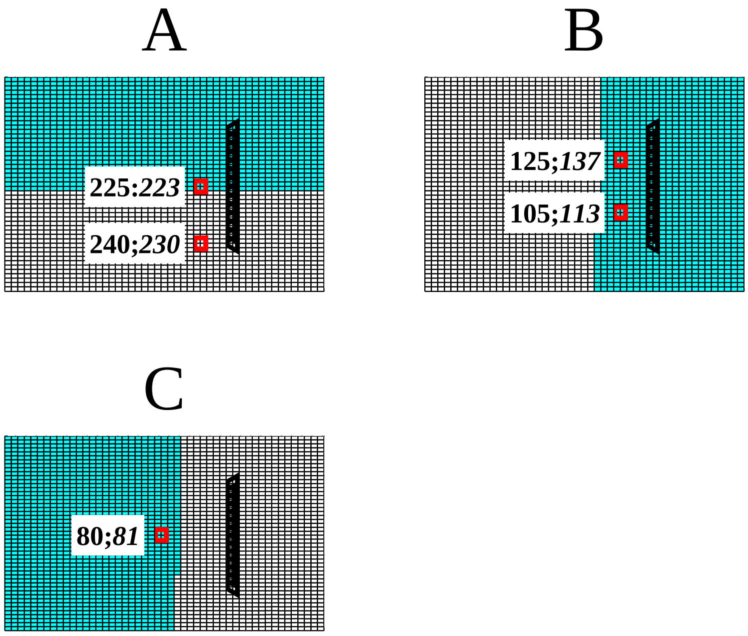

Figure 6: Optimal pumping designs for Models A through C, representing different geologic environments. The boundary of the distinct hydrologic fields in these models is at the 1.5 contour level, c_{bnd}, of the indicator value representing the most likely location of the boundary given the observation data. The pumping at the wells is measured in units of m^3d^{-1}. The non-italicized values represent the optimal pumping rates when the boundary between the K fields is located in the most likely location ($c_{bnd} = 1.5$). The italicized values are the mean values for designs when uncertainty is considered. More complete statistical results on the sets of solutions given different K fields are reported in Table 3.

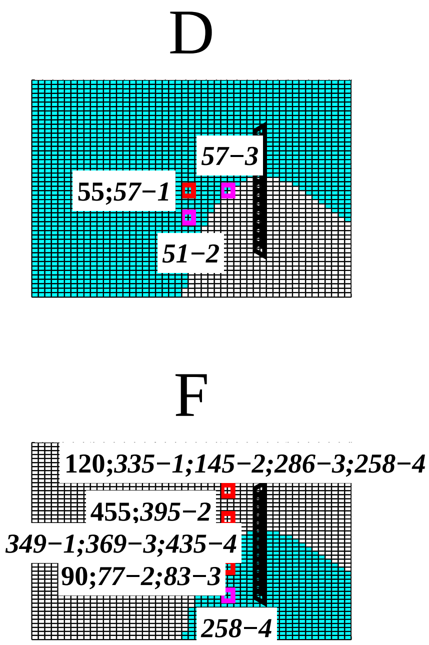

Figure 7: Optimal pumping designs for Models D through G, representing different geologic environments. The boundary of the distinct hydrologic fields in these models is at the 1.5 contour level, c_{bnd}, of the indicator value representing the most likely location of the boundary given the observation data. The pumping at the wells is measured in units of m^3d^{-1}. The non-italicized values represent the optimal pumping rates when the boundary between the K fields is located in the most likely location ($c_{bnd} = 1.5$). The italicized values are the mean values for designs when uncertainty is considered. If different remediation designs were determined for the set of K fields that represent the uncertainty, than the different pumping designs are indicated by the number that follows the dash after the pumping rate. More complete statistical results on the sets of solutions given different K fields are reported in Table 3.

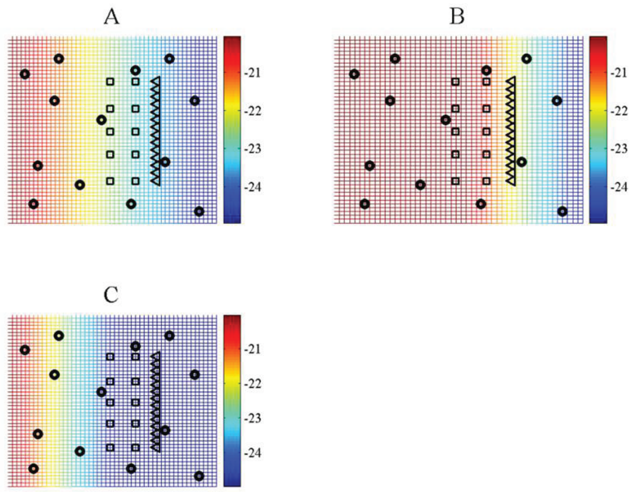

Figure 8: Ambient hydraulic heads for Models A,B and C. Red areas represent regions where the head values are close to -21 m while blue areas represent regions where the value is close to -25 m. Flow is from red to blue.

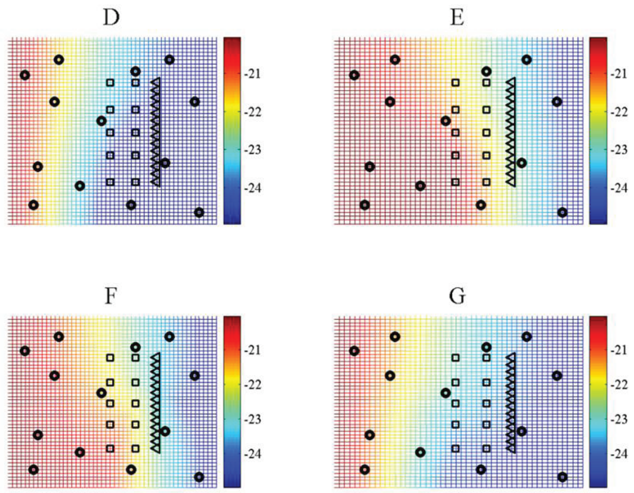

Figure 9: Ambient hydraulic heads for models D, E, F and G. Red areas represent regions where the head values are close to -21 m while blue areas represent regions where the value is close to -25 m. Flow is from red to blue.