ABSTRACT


Contaminated groundwater sites requiring remediation treatment exist all over the world. However, many of these sites have inadequate hydraulic information leading to inefficient cleanup strategies. Thus, many studies investigate the best way to obtain accurate hydraulic information (i.e., hydraulic conductivity or permeability) of the subsurface from secondary measurements such as hydraulic head and tracer/contaminant concentrations. Due to the limited amount of information and their associated uncertainty, these problems are often ill posed and non-unique. To minimize these shortcomings, different methods have been applied. An increasingly popular method to determine hydraulic conductivities in a given site is the pilot point method (PPM). The physical meaning of pilot points is that they are not direct measurement points; rather these selected points are added to the parameter search procedure to reduce the ill-posedness.

Two different evolutionary search based PPM approaches are developed in thesis as follows: 1) D-optimality sensitivity based method (SBM) that uses D-optimality criterion to search for pilot points and then a subsequent search for hydraulic conductivity values at these points, 2) A simultaneous search-based method (SSBM), where pilot points and hydraulic conductivities are searched simultaneously. These methods are first tested with hydraulic head measurements and then with both hydraulic head and concentration measurements using several synthetic problem scenarios.

Results show that the selected pilot points using SBM lead to a more accurate hydraulic conductivity characterization than a random set of pilot points and a sequential PPM developed previously. SSBM provide comparable hydraulic conductivity distributions to those from SBM, however it has more variance in searching hydraulic conductivities when performed over several trials. When tracer concentration measurements are added to the head measurements, both methods result in improved hydraulic conductivity characterization.
Evolutionary Algorithm based Pilot Point Methods for Subsurface Characterization

by
Yong Jung

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Civil Engineering

Raleigh, North Carolina
2008

APPROVED BY:

Chair of Advisory Committee
(Dr. G. (Kumar) Mahinthakumar)

Co-chair of Advisory Committee
(Dr. Ranji S. Ranjithan)

(Dr. Sankar Arumugam)

(Dr. Sujit Ghosh)
DEDICATION

This work is dedicated to my wife, Dr. Sunyoung Bae
who feareth the Lord, Jesus Christ.
Yong Jung was born in Kwangju, South Korea where he was raised until his college year. He got bachelor degree from Chosun University in the major of civil engineering. After his college, he worked for the division of civil engineering of the 91st Airbase Construction Group in the Korean Air force as an officer for four years. In addition he worked for airport design in Daewoo Engineering Company. In 2001, he started his Master of Science degree at North Carolina State University under direction of Prof. Robert Borden investigating the transport of emulsified edible oil for bioremediation in subsurface. After his master degree, he worked for environmental consulting firm, Solutions-IES, as an environmental engineer. After one year, he enrolled Ph.D. program in Civil Engineering at the same University of his master degree in 2004. His major study was improving the subsurface characterization using evolutionary computations.
ACKNOWLEDGEMENTS

I would like to appreciate Dr. G. Mahinthakumar for all his endeavors to guide me and mentor me with wonderful insight. I would not forget his caring of my family during the study. I would like to thank Dr Ranji Ranjithan for his profound academic advice and continuous encouragements, and specially thank to Dr. Sankar Arumugam, and Dr. Sujit Ghosh for their time and effort as committees.

I would like to thank to the officemates, Matthew Clayton, Emily Zechman, Li Liu, Baha Mirghani, Michael Tryby, Kyutae Lee, Jitendra Kumar, and Xin Jin for their collaborations and encouragements.

I am truly indebted to my lovely parents (Dangchae Jung, and Yanglim Choi), parent-in-laws (Shinhan Bae, and Hwaja Tae), brother’s family (Kwon Jung, Hyunyoung Park, Gyum, and Sol) and sister’s family (Hairang Jung, Youngju Park, Yubin, Yudam, and Yuha), sister-in-laws (Kichung Bae and Juyoung Bae) for their true support and prayers.

Finally I truly present my love to my adorable God’s gift, MIN JUNG.

(This work was partially supported by National Science Foundation (NSF) under Grant No. BES-0238623. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.)
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... vii
LIST OF FIGURES ........................................................................................................ viii

Chapter 1: Introduction ................................................................................................. 1

Chapter 2: Development and application of a Sensitivity based Pilot Point Method (SBM) for subsurface characterization ................................................................. 3
  2.1 Introduction ............................................................................................................. 3
     2.1.1 Pilot points ....................................................................................................... 5
     2.1.2 Hydraulic conductivity estimation .................................................................. 7
  2.2 Background ........................................................................................................... 8
     2.2.1 Genetic Algorithm (GA) ................................................................................ 8
     2.2.2 Kriging .......................................................................................................... 9
     2.2.3 Forward Flow Model .................................................................................... 10
  2.3 Pilot Point Methodology (PPM) .......................................................................... 11
     2.3.1 Determination of pilot points ....................................................................... 12
     2.3.2 Determination of hydraulic conductivity values .......................................... 15
  2.4 Test Problem and Results .................................................................................... 17
     2.4.1 Synthetic Model ............................................................................................ 17
     2.4.2 Pilot point locations ..................................................................................... 21
     2.4.3 Values at pilot point locations ...................................................................... 25
     2.4.4 Comparison with sequential search method ............................................... 28
  2.5 Conclusion ........................................................................................................... 30
  2.6 References .......................................................................................................... 32

Chapter 3: Development of a simultaneous search based pilot point method (SSBM) and its comparison with the sensitivity based pilot point method (SBM) .......... 37
  3.1 Introduction ........................................................................................................... 37
  3.2 Methodology ........................................................................................................ 41
     3.2.1 Simulation-Optimization Model .................................................................. 41
     3.2.2 Simultaneous Search-Based Method (SSBM) ............................................. 43
  3.3 Numerical Experiments ...................................................................................... 44
     3.3.1 Scenarios ...................................................................................................... 45
     3.3.2 Objective and Penalty .................................................................................. 47
  3.4 Results .................................................................................................................. 48
     3.4.1 SSBM and RANDOM .................................................................................. 48
     3.4.2 SSBM and SBM .......................................................................................... 53
  3.5 Summary and Final Remarks .............................................................................. 56
  3.6 References .......................................................................................................... 57

Chapter 4: Extension of SBM and SSBM to use both hydraulic head and tracer measurements ........................................................................................................... 61
  4.1 Introduction .......................................................................................................... 61
  4.2 Simulation Model ................................................................................................. 65
     4.2.1 Forward Model .............................................................................................. 65
4.2.2 Max/Minimization Algorithm ................................................................. 66
4.3 Calibration: Sensitivity Based Method (SBM) ............................................ 67
   4.3.1 Overview of SBM ............................................................................. 67
   4.3.2 Objective Functions ......................................................................... 68
4.4 Numerical Examples of SBM ..................................................................... 71
   4.4.1 Hypothetical Models ......................................................................... 71
   4.4.2 Different Scenarios .......................................................................... 74
   4.4.3 Numerical Comparison ..................................................................... 75
4.5 Results and Discussions .......................................................................... 75
4.6 Conclusion ............................................................................................... 85
4.7 References .............................................................................................. 87
Chapter 5: Conclusions .................................................................................. 90
LIST OF TABLES

<Table 2.1> Chosen genetic algorithm options in MATLAB................................. 21
<Table 2.2> D-optimality values of final selected pilot point locations for SBM....... 22
<Table 2.3> Head error values with sequentially selected points.......................... 29
<Table 4.1> Coupled inverse problems in different conditions............................. 79
LIST OF FIGURES

<Figure 2.1> Synthetic kriging results for this study (Left: semi-variogram graph with sill and range (x-axis: distance, y-axis: semi-variogram), Right: 3-dimensional figure for kriged hydraulic conductivity values (x- and y-axes: grid points, z-axis: hydraulic conductivity values), orange color indicates higher hydraulic conductivity) ........................................ 10

<Figure 2.2> Flowchart of the sensitivity based pilot point method .................................. 12

<Figure 2.3> Synthetic groundwater field (unknown reality field; square = known hydraulic conductivity locations, triangle = hydraulic head measurement locations) and unknown reality of K-field ...................................................... 18

<Figure 2.4> Mean uniform (Left) and mean non-uniform (Right) flow contours for homogeneous hydraulic conductivity condition with constant heads at upstream (20 m) and downstream (15 m) respectively. ....................... 19

<Figure 2.5> Initial K-field distributions for four scenarios ........................................... 20

<Figure 2.6> Final selections of pilot points in four scenarios .................................... 22

<Figure 2.7> Comparisons of D-optimality values between random pilot point sets and GA optimized pilot point sets ........................................................................................................ 23

<Figure 2.8> Sum of correlations (left axis) of pilot point locations and their sensitivity metric (right axis) from random and SBM ................................................................. 24

<Figure 2.8> Sum of correlations (left axis) of pilot point locations and their sensitivity metric (right axis) from random and SBM. (Continued) ......................... 25

<Figure 2.9> Head error values from pilot point locations from D-optimality (Det-PPL) and sensitivities of randomly selected locations (Random) ...................... 26

<Figure 2.10> Contour maps of scenario_2 (Top left: unknown reality, Top right: initial contour, Bottom left: without penalty, Bottom right: with penalty) ....... 28

<Figure 2.11> Sequentially searched pilot point locations with the selection order and final hydraulic conductivity distribution (Using four selection) ............ 29

<Figure 2.11> Sequentially searched pilot point locations with the selection order and final hydraulic conductivity distribution (Using four selection) (Continued) .................................................................................. 30
<Figure 3.1>  Simultaneous search based pilot point method .................................................. 44

<Figure 3.2>  Hydraulic conductivity distribution of unknown reality ................................. 45

<Figure 3.3>  Initial hydraulic conductivity distributions (From left top to right bottom: Scenario_1, Scenario_2, Scenario_3, and Scenario_4) ........................................... 46

<Figure 3.4>  Comparison of minimum fitness comparison from random locations and SSBM .............................................................................................................. 50

<Figure 3.5>  Final pilot point locations using SSBM with measurement points of hydraulic conductivities and hydraulic heads for four different scenarios 51

<Figure 3.6>  Final hydraulic conductivity contour maps (From top left to right: Scenario_1, Scenario_2, Scenario_3, and Scenario_4) ........................................... 52

<Figure 3.7>  Average minimum fitness from thirty trials and average hydraulic conductivity difference ........................................................................................................... 54

<Figure 3.8>  Contour maps of unknown reality, initial distribution of scenario 2 (Top left and Top right) and final distribution from SBM and SSBM respectively (Bottom left and bottom right) ............................................................... 55

<Figure 4.1>  Flow chart of coupled inverse problems using SBM ........................................ 68

<Figure 4.2>  Hydraulic conductivity distribution and groundwater flow for the unknown reality ......................................................................................................................... 72

<Figure 4.3>  Initial tracer injection line with the transport depending on time in the domain (triangles: observation points, and dark line: Tracer injection points in initial injection domain) ........................................................................... 73

<Figure 4.4>  Hydraulic conductivity measurement locations and initial hydraulic conductivity distributions based on kriged measurements for scenario 1 and 2 .................................................................................................................. 74

<Figure 4.5>  Comparison between various information based pilot point method using average of relative hydraulic conductivities and average of maximum hydraulic conductivity differences .............................................................. 76

<Figure 4.6>  Final pilot point locations from each information set for scenario 1 and 2 (squares: hydraulic conductivity measurements, triangles:
head/concentration observation locations, and circles: optimal pilot point locations) .......................................................... 77

<Figure 4.7> Combination of pilot point locations from head and concentration or quantile separately (In scenario 1, the combinations of head and concentration or head and quantile are located in the same points) ........ 79

<Figure 4.8> Comparison between SBM and SSBM .................................................. 81

<Figure 4.9> Final pilot point locations using SSBM .................................................. 81

<Figure 4.10> Hydraulic conductivity distributions from SBM and SSBM for scenario 1 and 2 .............................................................................................................. 82

<Figure 4.10> Hydraulic conductivity distributions from SBM and SSBM for scenario 1 and 2 (Continued) .............................................................................................................. 83

<Figure 4.10> Hydraulic conductivity distributions from SBM and SSBM for scenario 1 and 2 (Continued) .............................................................................................................. 84
Chapter 1: Introduction

For effective groundwater resource management, accurate knowledge regarding the hydraulic conductivity distribution (K-field) of the aquifer is extremely important. Due to the cost and difficulty of obtaining K measurements, most sites have very few sparsely distributed K-measurements. At many sites, however, secondary measurements such as head and concentration measurements may be available. These measurements can be used to augment K measurements via parameter estimation techniques to improve K-field characterization. Since the number of parameters involved in characterizing the entire K-field can be very large compared to the number of measurements available, the parameter space should often need to be reduced in the estimation process to avoid problems such as non-uniqueness. Some approaches to reduce parameter space in K-field characterization include statistical representation, zonal representation, and more recently, pilot point techniques. The pilot Point Method (PPM) was first suggested by De Marsily et al. (1984) for reducing the parameter space for K-field characterization. Pilot point methods reduce uncertainty in the K-field estimation process by identifying a few key locations in the aquifer where the K-values could be adjusted to honor secondary measurements. Once the K-values are obtained at these locations, interpolation (or Kriging) could be used to predict the K-values in the remaining space. The main advantage of PPM over other methods is its increased flexibility in choosing strategic locations for parameter estimation. Even though pilot point methods are very effective in reducing the parameter space, the strategies for the choice of pilot points still remains an open question. The overall goal of this dissertation is to develop two new pilot point methods and test their effectiveness in solving a synthetic hydraulic conductivity estimation problem. Also, most pilot methods investigated so far have used only hydraulic heads as the only secondary measurement. The effect of adding other secondary measurements such as concentration data needs to be studied. Therefore, a second goal of this dissertation is to study how additional types of secondary data can improve the effectiveness of the PPMs developed through this research.
This dissertation is organized into three major chapters, each constituting a journal manuscript. The first paper, chapter 2, develops a sensitivity based pilot point method (SBM) and tests its effectiveness on synthetic problem scenarios using hydraulic head data as the only secondary measurement. The second paper, chapter 3, develops simultaneous search based method (SSBM) and tests and compares its effectiveness on the same scenarios used earlier. The third paper, chapter 4, extends both SBM and SSBM to use concentration measurements in addition to hydraulic head measurements.
Chapter 2: Development and application of a Sensitivity based Pilot Point Method (SBM) for subsurface characterization

2.1 Introduction

Inadequate representation of the hydraulic conductivity or permeability of the subsurface is one of the greatest sources of uncertainty in developing good groundwater flow and transport models for field applications. Unfortunately, very few direct hydraulic conductivity measurements are available for most sites. Given this reality, considerable research has been expended during the last few decades in developing methods that can improve hydraulic conductivity estimates from secondary measurements such as hydraulic head or tracer concentrations. These methods fall under the category of inverse modeling methods as hydraulic conductivity field is input to a groundwater flow and transport model while hydraulic heads and tracer concentrations are output. Given the potentially large parameter space and lack of adequate secondary measurements, combined with the uncertainties of prior information and simulation models, these inverse problems are generally ill-posed suffering from non-uniqueness, instability, and non-existence. A number of texts and publications including Yeh (1986), Carrera and Neuman (1986), Dietrich and Newsam (1990), and McLaughlin (1996) provide excellent reviews on this topic.

In order to reduce the parameter space for stabilization of inverse problem, Stallman (1956) introduced zonal representation of hydraulic conductivity field, which has been applied to deterministic or stochastic condition of hydraulic conductivity (Coats et al. 1970, Emsellem and de Marsily 1971, Neuman and Yakowitz 1979, Clifton and Neuman 1982, Neuman 1982, and Carrera and Neuman 1986a, 1986b, Medina and Carrera 1996). As an alternative, others used a geostatistical representation (Kitanidis and Vomvoris 1983, de Marsily et al. 1984, Certes and de Marsily 1991, and Sun et al. 1995). In
addition, regularization is employed to the objective function to further reduce non-uniqueness (Emsellem and de Marsily 1971, and Carrera and Neuman 1986a, 1986b). Regardless of the representation of the hydraulic conductivity field, both optimization based (i.e., nonlinear least squares) and geostatistical inversion based techniques can be used. Geostatistical inversion methods are based on a linearized relationship between the geostatistical models of the hydraulic conductivity with secondary measurements such as hydraulic heads (Gutjahr et al. 1978, Keidser and Rosbjerg 1991, and Sun and Yeh 1992). However, the linearization assumption used in these methods is often valid for small variances in the hydraulic conductivity field, which may not represent reality (Certes and de Marsily 1991, RamaRao et al. 1995, LaVenue et al. 1995, and Hendricks-Franssen 2000). Optimization based methods adjust the hydraulic conductivity field in a systematic fashion to match the secondary measurements by simulating these measurements using a forward groundwater flow or transport model. While these methods have a more general applicability, unless the parameter space is reduced by a proper representation of the hydraulic conductivity field, they can suffer from issues such as non-uniqueness and instability. Pilot point methods (PPM) offer a promising alternative to zonal models and improvement over traditional geostatistical models in reducing the parameter space in an optimization based inversion approach. This method was first proposed by De Marsily et al. (1984) where a set of calibration points called “pilot points” are selected from the model domain at which the hydraulic conductivity is adjusted to match the head measurements. A number of researchers have modified the PPM and applied it to different situations in terms of reducing instability that results from over-parameterization and huge fluctuations of estimated parameter values (Fasanino et al. 1986, Certes et al. 1991, LaVenue et al. 1992, 1995, 2001, RamaRao et al. 1995, Oliver et al. 1996, Cooley 2000, Vesselinov et al. 2001a, 2001b, Medina et al. 2003, Hernandez et al. 2003, Doherty 2003, Kowalsky et al. 2004, Alcolea et al. 2005).

The basic approach of PPM proposed by De Marsily et al. (1984) is as follows; 1) initialize hydraulic conductivity field using a geostatistical method (ordinary/universal...
kriging) based on measured hydraulic conductivity values, 2) select pilot point locations where no values of measured hydraulic conductivity are available and use only kriged values, 3) iterate to find optimal values of hydraulic conductivity at selected pilot point locations by minimizing an objective function (generally, the sum of squared differences between observed hydraulic heads and calculated hydraulic heads obtained using a forward model) and kriged hydraulic conductivity field, 4) a final hydraulic conductivity field will be obtained by previous iterations and kriging. Comparative studies for the PPM with other nonlinear inverse approaches can be found in Keidser and Rosbjerg 1991, McLaughlin et al. 1996, Zimmerman et al. 1998, Floris et al. 2001, Carrera et al. 2005. Following the given basic approach, we focus on the optimal number and location of pilot points to prevent over-parameterization and reduce instability.

2.1.1 Pilot points

Location and number of pilot points are major factors of stability in parameter identification using pilot point method. Pilot point locations can be selected based on a number of criteria including empirical and sensitivity considerations. Several authors including De Marsily et al. (1984), Fasanino et al. (1986), Certes and de Marsily 1991, Doherty (2003), and Hernandez et al. (2003) used empirical considerations such as, 1) density of uniformly distributed points, and 2) geological indication of large amount of heterogeneity (e.g. high gradient of measured hydraulic heads). An efficient sensitivity based approach to select pilot point locations was introduced by LaVenue and Pickens (1992). They employed adjoint sensitivity techniques that calculate sensitivities of hydraulic conductivity to measured hydraulic heads at each potential pilot point location. These sensitivities are ranked and the locations with the highest sensitivities are successively added to the pool of pilot points until no further improvement is possible. In addition, several applications with the same manner of above using adjoint sensitivity analysis were used to find optimal pilot point locations at different situations (RamRao et al. 1995, LaVenue et al. 1995, Hendricks-Franssen 2000).
In general, the number of pilot points should be less than or equal to the number of secondary measurements available to prevent over parameterization and non-uniqueness. Beyond this constraint, there is no fixed rule on the optimal number of pilot points for a given scenario. One may fix the number and location of pilot points during whole procedure (de Marsily et al. 1984, Fasanino et al. 1986, Certes and de Marsily 1991, Romero et al. 2000, and Doherty 2003) or iteratively add the pilot points while searching for their locations (LaVenue and Pickens 1992, RamRao et al. 1995, LaVenue et al. 1995, and LaVenue and de Marsily 2001). Both approaches have been widely used. RamRao et al (1995) suggested that sequential addition is superior to simultaneous selection. For the number of pilot points, RamRao et al. (1995) and LaVenue et al. (1995) suggested that fewer number of pilot points has better performance, since it is less likely to give large fluctuations in the hydraulic conductivity field. Furthermore, De Marsily (1978) suggested that number of pilot points should be less than the number of measured hydraulic conductivity values (LaVenue and De Marsily 2001). However, Capilla et al. (1997) and Doherty (2003) pointed out that more pilot points might yield a more accurate or realistic hydraulic conductivity field. Thus, as pointed out by Chavent et al (1998), there is usually an optimal number of pilot points as too few pilot points may dishonor the secondary measurements and too large a number may cause over-parameterization. In our opinion, the optimal number of pilot points would be the minimal number required to honor the secondary measurements such as head and concentrations. This would minimize the chance of having too many correlated points (i.e., points that can invoke similar changes in secondary measurements thus leading to potential non-uniqueness by redundancy). Having a minimal number of highly sensitive points is also less likely to destroy the original covariance structure of the measured hydraulic conductivities. Preserving the covariance structure might be an important consideration in cases where there are a significant number of measured hydraulic conductivity values.
Instead of using pilot point locations, Sahuquillo et al. (1992), Gomez-Hernanez et al. (1997), Capilla et al (1998), Hendricks Franssen et al. (1999), Wen et al. (1998a, 1998b, 1999, 2002) used master point locations to reduce the parameter space. The resulting method is known as the Sequential Self-Calibration (SSC) method. SSC is conceptually similar to PPM except the use of master points, which are randomly selected imaginary grid points and used as hydraulic conductivity perturbation locations. Master point locations, however, should fulfill the basic need of one-third variogram correlation range of hydraulic conductivity measurements. For this reason, only one or two master location(s) should be in the correlation range as the rule of thumb. For instance, Wen et al. (1998b) used the range of the hydraulic conductivity variogram and number of observation wells as criteria to decide the number of master locations and assign fewer master locations for fewer wells or wide-ranging correlation lengths. As in most previous studies, this study also uses the number of observation wells as a basic criterion for the determination of number of pilot point locations. However, their locations will be based on criteria that include both sensitivity and parameter correlation. This is different from sensitivity based methods used by LaVenue and Picken(1992), RamaRao et al (1995), and LaVenue et al (1995). The criteria will be discussed in detail in the methodology section (2.3).

2.1.2 Hydraulic conductivity estimation

Once the potential pilot points are found, hydraulic conductivity needs to be estimated at these points by adjusting them to match the secondary measurements. In estimating the hydraulic conductivity, many applications of PPM and SSC have used gradient-based search methods (de Marsily 1984, Fasanino et al. 1986, RamaRao et al. 1995, LaVenue et al. 1995, and Wen et al. 1998, 1999). A potential shortcoming of gradient-based search methods is that they are local search methods; this might cause it to converge to a local minimum if several local minima exist. For overcoming this deficiency, a global search method such as genetic algorithm (GA) can be employed. One can find many
applications of GA for hydraulic conductivity or permeability estimation in petroleum reservoir and groundwater domains (Sen et al. 1995, Bush et al. 1996, Guerreiro et al. 1998, Romero et al. 2000, Karpouzos et al. 2001, Romero et al. 2002). In this study we propose a GA based search approach for pilot point location selection as well as hydraulic conductivity estimation at the selected pilot points.

The remainder of this paper is organized as follows. The next section, Background, provides a description of the computational components, section 2.3 describes the pilot point method developed in this paper, including descriptions of genetic algorithm, geostatistics, groundwater flow model, and numerical experiment. Section 2.4 presents the test problem with results and discussion of application of the pilot point method, followed by Section 2.5, Conclusions.

2.2 Background

The pilot point method developed in this paper uses three main computational components: (a) a genetic algorithm based search method for locating pilot points and determining hydraulic conductivity values, (b) simple kriging for interpolation of K-values, and, (c) a groundwater flow model for determining hydraulic head values from K-field. For simplicity, these computational components are implemented in a MATLAB environment.

2.2.1 Genetic Algorithm (GA)

One of the popular global heuristic optimization methods is the genetic algorithm (GA) suggested by Holland (1975). The basic idea came from the process of natural evolution. As selected species in nature survive in a certain environment, numerically selected values that have the best fit in a particular condition can be found as a result of GA operation. A GA consists of different operational functions such as representation,
selection, crossover, mutation, and elitism. In GA the decision variables are usually encoded as real or binary strings. In this paper, the decision variables are either the pilot point locations (integer node indices in the grid) or hydraulic conductivity values. For simplicity, real representation is used for these decision variables. The built-in genetic algorithm function in the MATLAB GA toolbox is used in this paper.

2.2.2 Kriging

Kriging/co-kriging is an interpolation tool to obtain unbiased predicted values of hydraulic conductivities over a set of grid points based on observed correlated values at certain points. Moreover, this has the advantage of reproducing smooth surface of optimal hydraulic conductivities from hydraulic conductivity perturbations at pilot point locations. In most applications, kriging is based on the assumption that the observed values are generated from an intrinsic stationary stochastic process defined over a space or region. Intrinsic stationarity implies that the mean of the process is constant and the variance of the difference of two response values depends only on the distance between the locations. More specifically, if \( Z(s) \) represents hydraulic conductivity at location \( s \), then \( \{Z(s), s \in D\} \) is said to be intrinsic stationary, if (i) \( E[Z(s)] = \mu, \forall s \in D \) and (ii) \( \text{Var}[Z(s_1) - Z(s_2)] = 2\gamma(s_1 - s_2), \forall s_1, s_2 \in D \).

For this study, the exponential model \( (\gamma(h) = c_0 + c_1 \cdot (1 - e^{-3|h|/a})) \) for semi-variogram is applied where the sill, \( c_1 \),represents the maximum of semi-variogram of measured hydraulic conductivity; range, \( a \) denotes the correlation distance of observations, beyond this distance observations are not correlated; and nugget, \( c_0 \) represents micro scale variance of observations observed at the same point. We assume that the observed hydraulic conductivities are log-normally distributed. Initial hydraulic conductivity field is generated by kriging the initially observed data. Then, the hydraulic conductivities are perturbed at the pilot point locations to honor the observed hydraulic head values.
Kriging is reapplied using the perturbed and measured conductivity values to produce the final estimate of hydraulic conductivity values. Figure 2.1 shows the semi-variogram and hydraulic conductivity field from a given synthetic example for this study.

![Figure 2.1](image)

**Figure 2.1** Synthetic kriging results for this study (Left: semi-variogram graph with sill and range (x-axis: distance, y-axis: semi-variogram), Right: 3-dimensional figure for kriged hydraulic conductivity values (x- and y-axes: grid points, z-axis: hydraulic conductivity values), orange color indicates higher hydraulic conductivity)

### 2.2.3 Forward Flow Model

The forward groundwater flow model is based on a central difference approximation of the steady state groundwater flow equation:

\[
\frac{\partial}{\partial x} \left( T(x, y) \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( T(x, y) \frac{\partial h}{\partial y} \right) = 0
\]

**eq 2.1**

where

\[ T(x, y) \text{ = Isotropic, heterogeneous, transmissivity field(} k(x, y) \times \text{Aquifer depth)} \]

\[ h(x, y) \text{ = Head field} \]
This forward model is used to obtain the head field $h(x,y)$ for a given hydraulic conductivity field $k(x,y)$.

### 2.3 Pilot Point Methodology (PPM)

The PPM developed in this paper involves two steps: one, selecting the pilot point locations (section 3.1) and two, determining hydraulic conductivities at the pilot point locations (section 3.2). Prior to performing these two steps, a synthetic hydraulic conductivity field (‘unknown reality’) and an ‘initial hydraulic conductivity field’ must be generated for testing purposes. In this study the unknown reality is generated by kriging arbitrarily assumed measured hydraulic conductivity values at selected points. The initial hydraulic conductivity field is generated by excluding a subset of the original hydraulic conductivity measurements and then kriging these values. The hydraulic head measurements at the observation locations are generated by performing a groundwater flow simulation using the unknown reality hydraulic field. In step 1 of the PPM algorithm, starting with the initial hydraulic conductivity field, genetic algorithm is used to search for the set of pilot points that give the maximum D-optimality value (equation 2.2). In step 2, the hydraulic conductivities at the selected pilot point locations are perturbed to minimize the difference between the observed and calculated hydraulic head values (equation 2.4). The observed head values are generated from the unknown reality. The calculated head values are based on the kriged hydraulic conductivity field that combines the original hydraulic conductivity measurements with the pilot point perturbed hydraulic conductivity values. A schematic of this procedure with step 1 on the left side and step 2 on the right side is illustrated in Figure 2.2.
As shown in figure 2.2, the two separate procedures are sequentially carried out.

2.3.1 Determination of pilot points

The most sensitive locations with least correlation ought to be selected as pilot points, as this will maximize the data worth of the hydraulic head measurements. In other words, these would be the most unique set of points that require minimal perturbations of hydraulic conductivities to match hydraulic head observations. A criterion that satisfies these conditions is the D-optimality criterion mathematically expressed as (J. Kiefer and W. Keifer, 1959, Fedorov et al., 1968)

\[
\text{MAXIMIZE } obj_1 = \det[X^TX]
\]

\text{eq 2.2)
\[
\begin{bmatrix}
\frac{\partial h_1}{\partial k_1} & \frac{\partial h_1}{\partial k_2} & \cdots & \frac{\partial h_1}{\partial k_m} \\
\frac{\partial h_2}{\partial k_1} & \frac{\partial h_2}{\partial k_2} & \cdots & \frac{\partial h_2}{\partial k_m} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial h_n}{\partial k_1} & \frac{\partial h_n}{\partial k_2} & \cdots & \frac{\partial h_n}{\partial k_m}
\end{bmatrix}
\]

where

\(\text{obj}_1\): objective function in first procedure

\(X\): sensitivity matrix

\(h\): hydraulic head observation (n-number of hydraulic head observations)

\(k\): hydraulic conductivity values at potential pilot point locations

\(m\): number of selected pilot points

A key quantity in this procedure is the sensitivity matrix \((X)\) shown in equation 2.3 that measures the sensitivity of hydraulic heads with respect to hydraulic conductivities at potential pilot point locations. As equation 2 shows, we are maximizing the determinant of the Fisher information matrix \(X^T X\), which is known as D-optimality. If \(A = X^T X\) and variance-covariance matrix \(B = A^{-1}\), the correlation matrix \(C\) is simply the covariance B matrix scaled both column wise and row wise by the square root of the respective diagonal entries (Hill 2003). Thus the Fisher information matrix \(A\) is proportional to the inverse of covariance matrix \(B\). Also, maximization of determinant of the Fisher information matrix is mathematically equivalent to the problem of minimizing the norm of the covariance matrix (Knopman et al., 1987). The set of pilot points chosen in this manner satisfy the following criteria: these are the least correlated set of points in terms of hydraulic head response at the head observation locations to hydraulic conductivity
changes. In other words, changes in hydraulic conductivity at any two points from this set will invoke dissimilar changes in head values at the head observation points. In an optimization sense, these set of points are less prone to non-uniqueness as we are more likely to find a unique set of hydraulic conductivity values at these points to match the head observations. As opposed to minimizing the determinant or norm of the covariance matrix, the D-optimality criterion (i.e., maximizing the determinant of the Fisher information matrix) has two additional advantages. First, it does not require calculation of the correlation matrix thus resulting in significant computational savings and smaller impact of round off errors. Second, D-optimality also ensures that the pilot points selected are highly sensitive to the head observations; a larger determinant value will ensure larger diagonal values (higher sensitivity) and smaller off diagonal values (lower correlation). Higher sensitivity means that smaller perturbations are required at the pilot points to match the head observations thus nearly preserving the original correlation structure of the measured hydraulic conductivities.

Similar to the D-optimality criterion, the sequential search method (SSM) employed by LaVenue et al. (1992) using adjoint sensitivity also ensures that the points selected are maximally sensitive. Since the points are selected one at a time, in terms of the scale of sensitivity, norm function (maximum of the matrix singular value decomposition) to get magnitude of the matrix elements is sufficient. However, unlike D-optimality, SSM does not ensure that the combined set of pilot points is least correlated as described earlier. Due to these advantages, D-optimality has been used for a number of applications including parameter estimation in groundwater (Knopman et al. 1987) and characterization of water distribution systems (Bush et al. 1998). To our knowledge, it has not been used in pilot point selection.

In order to evaluate the D-optimality criterion, the number of pilot points (\(m\) in equation 3), need to be known apriori. In general, the number of pilot points should be less than or equal to the number of head measurements. Using a larger number would lead to
redundant pilot points and the matrix $X$ will be rank deficient (number of unknowns $m >$ number of rows $n$). This would likely make the fisher information matrix $(X^T X)$ to be linearly dependent forcing its determinant (D-optimality) to be essentially zero. A number $m$ that is too small will not make efficient use of the observed head values. Based on these observations and preliminary tests conducted with different number of pilot points, in this study, we chose the number of pilot points to be one less than the total number of head measurements. Since the number of head measurements is always four in this study we fix the number of pilot points to three (see 4.2).

Heuristic unconstrained search optimization approaches such as GA’s frequently enforce constraints by the use of penalty functions. These are simply penalty values added to the objective function to encourage search towards a feasible space. Since the version of the MATLAB GA toolbox used in this paper did not support constraints, penalties were used to enforce constraints. In the pilot point location search procedure, bound constraints limiting the pilot points to nodes in the computational domain are enforced by adding a penalty whenever this is violated. The measurement locations were avoided in the search by encoding the decision variables (pilot point locations) by excluding the measurement locations. The D-optimality criterion automatically ensures that no two pilot points are the same since the value of D-optimality in this case would be zero.

2.3.2 Determination of hydraulic conductivity values

Once the pilot point locations are determined, the hydraulic conductivity values need to be determined at these locations from the available hydraulic head measurements. This inverse problem can be formulated as an optimization problem. The objective is to minimize the difference between the observed and the simulated hydraulic head values by perturbing the hydraulic conductivities at the pilot point locations. This difference can be quantified using a sum of squared residual as expressed below by equation (2.4):
MINIMIZEx \( \text{obj}_2 = \sum (h_{\text{observed}} - h_{\text{calculated}})^2 \) \hspace{1cm} \text{eq 2.4}

where

\( \text{obj}_2 \): objective function

\( h_{\text{observed}} \): observed hydraulic head values

\( h_{\text{calculated}} \): calculated hydraulic head values

Since hydraulic heads are primarily sensitive to spatial variations of hydraulic conductivity than their absolute magnitudes, it is customary to constrain the magnitude of hydraulic conductivity based on prior information to reduce the search space (e.g., LaVenue et al. 1992). In this paper, we use the geometric mean of measured hydraulic conductivity values to narrow down the hydraulic conductivity search space. To accomplish this, we add a penalty to the objective in equation 2.5 if the geometric mean of the hydraulic conductivity values at the pilot points (gp) deviates from the geometric mean of the measured hydraulic conductivity values (gm) by more than 75%.

\[
\text{if } \left| \frac{\text{gm} - \text{gp}}{\text{gm}} \right| > 0.75
\]

\[ \text{obj}_2 = \text{obj}_2 + \frac{\left| \text{gm} - \text{gp} \right|}{\text{gm}} \] \hspace{1cm} \text{eq 2.5}

end

...
2.4 Test Problem and Results

2.4.1 Synthetic Model

A 2-dimensional synthetic site (700 m by 1000 m) with regular shape of grid blocks (100 m by 100 m) shown in figure 2.3 is used for numerical tests. This synthetic groundwater flow field is based on a confined, isotropic, and heterogeneous aquifer exhibiting mean or non-mean uniform flow presented in figure 2.4. The flow field is obtained by applying constant head boundary conditions at the left (upstream) and right (downstream) boundaries (scenario 1 and 2) or left (upstream) and bottom (downstream) boundaries (scenario 3 and 4) and no-flow boundary conditions at the other sides. The upstream and downstream boundaries have 20 m and 15 m constant hydraulic heads respectively.

The synthetic hydraulic conductivity field (unknown reality) is generated by interpolating (kriging) nine randomly generated hydraulic conductivity measurements at random points. The exponential semi-variogram model used for kriging and the resulting K-field is the same as shown earlier in Figure 2.1. Based on these hydraulic conductivity measurements, the hydraulic head measurements are generated at four arbitrary observation locations using the groundwater flow mode. The nine hydraulic conductivity measurement points (squares) and the four hydraulic head observation points (triangles) are shown in Figure 2.3.
<Figure 2.3> Synthetic groundwater field (unknown reality field; square = known hydraulic conductivity locations, triangle = hydraulic head measurement locations) and unknown reality of K-field
To generate different initial K-field scenarios for the numerical experiments, four hydraulic conductivity measurements are selected from the original nine measurements by excluding five measurements. Two different synthetic K-fields (Figure 2.5) with two different flow conditions (Figure 2.4) constituting a total of four scenarios are used in the experimental study. Scenarios 1 and 2 correspond to the case with mean uniform flow and scenarios 3 and 4 correspond to the case with mean non-uniform flow. Of these four, the initial K-field shared by Scenarios 2 and 4 is significantly different from the unknown reality when compared to the one used by Scenario 1 and 3. The relative hydraulic conductivity error metric as defined by equation 2.6 is used to evaluate the efficiency of each method in matching the unknown reality.
Field 1 (Scenario 1 & 3)  Field 2 (Scenario 2 & 4)

<Figure 2.5> Initial K-field distributions for four scenarios

\[
\text{Re } K \text{ Error} = \frac{1}{N_t} \sqrt{\sum_{i=1}^{N_t} (K_{ci} - K_{ri})^2}
\]

\text{eq 2.6)

where

- \(N_t\): Total number of known hydraulic conductivity
- \(K_{ci}\): Calculated hydraulic conductivity
- \(K_{ri}\): True hydraulic conductivity

The ‘ga’ function in the MATLAB GA toolbox is used to search for the pilot point locations (first procedure) and hydraulic conductivity values at the selected pilot point locations (second procedure). Real encoding is used for representing the populations in these experiments. Population size and maximum number of generations are fixed at 1000 and 100 respectively with chosen function types in table 2.1. Calculations are carried out on a PC with a Pentium 4 CPU, 3.0 GHz clock speed, and 1.0 GB RAM.
2.4.2 Pilot point locations

Based on the considerations discussed earlier, the number of pilot points was fixed at three (one less than the total number of head measurements) throughout the search process. Fixing this number significantly simplifies the search process for the pilot point locations. The locations of these three pilot points are then searched using GA based on the D-optimality criterion (equation 2.2). The sensitivity matrix X (equation 2.3) in the D-optimality criterion is calculated using central finite difference using a fixed perturbation of $\Delta K = 1$. The measurement points are excluded from the search as described in section 4.1. The final selected pilot point locations (displayed as circles) are presented in figure 2.6 for the four scenarios. The final pilot point locations in figure 2.6 are well distributed in the given domains of four scenarios. This is probably due to the fact that the D-optimality criterion automatically ensures distribution as pilot points that are close together would likely be correlated thus resulting in small D-optimality values.
The final D-optimality values for the pilot point sets from the GA search are presented in Table 2.2 for each scenario. One might notice that these values are extremely small. This is plausible as hydraulic heads are known to have very low sensitivities to the magnitude of hydraulic conductivities that are perturbed at the pilot points. However, hydraulic heads are sensitive to the variability of hydraulic conductivity that is exploited when the pilot points are selected in a collective fashion.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>D-optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario_1</td>
<td>18E-10</td>
</tr>
<tr>
<td>Scenario_2</td>
<td>9.81E-15</td>
</tr>
<tr>
<td>Scenario_3</td>
<td>1.17E-14</td>
</tr>
<tr>
<td>Scenario_4</td>
<td>1.26E-14</td>
</tr>
</tbody>
</table>
In order to evaluate the effectiveness of GA in searching pilot point locations with high D-optimality values, we calculated D-optimality values of thirty random pilot point sets (3 points in each set) in each scenario and compared it to the values obtained by the GA search. These are presented in figure 2.7. In the figure 2.7, squares represent the D-optimality value from GA search and the diamonds represent values for randomly selected pilot point locations. Clearly, for all four scenarios, the D-optimality values for the pilot points searched by GA are one or two orders of magnitude greater than those for a random set of pilot points. This indicates that GA is effective in searching for pilot points with high D-optimality values.

\[ \text{Figure 2.7} \quad \text{Comparisons of D-optimality values between random pilot point sets and GA optimized pilot point sets} \]
To evaluate the effectiveness of the D-optimality metric in SBM, two additional metrics are calculated: (1) Sum_corr = Sum of correlations between individual pilot point locations (sum of the absolute values of the off-diagonal entries of the correlation matrix described in section 3.1), (2) Norm_sens = Norm of sensitivity matrix. The Sum_corr metric measures the correlation among the selected set of pilot points. A low sum_corr value is better for estimation as this will minimize non-uniqueness. The norm_sens metric measures the sensitivity of the pilot points to the measured hydraulic heads. A high norm_sens metric is better for estimation as this will reduce the amount of perturbation required to match the measured head values. In Figure 2.8 we compare these metrics for SBM pilot points and random pilot points for the 4 scenarios for 30 random trials. In most cases, the pilot points from SBM are less correlated with significantly larger sensitivities. Occasionally randomly selected pilot point locations provide less correlated locations but with insignificant sensitivity values. In other cases, the random pilot point locations are highly sensitive to the hydraulic head measurements, but these locations are strongly correlated each other. In other words, the SBM based pilot points provide a good compromise between these two metrics.

<Figure 2.8>  Sum of correlations (left axis) of pilot point locations and their sensitivity metric (right axis) from random and SBM.
2.4.3 Values at pilot point locations

Once the pilot point locations are identified, the values of hydraulic conductivity at these points need to be estimated. As discussed section 4.2, hydraulic conductivities at the selected pilot points are perturbed to minimize the differences of observed and calculated hydraulic heads at observation points. Observed heads correspond to the unknown reality and the calculated heads correspond to the iteratively estimated hydraulic conductivity distribution in the GA search. In the absence of non-uniqueness, the estimated conductivity is close to the unknown reality when the differences between the head values are small. In Figure 2.9 we compare the objective function values (equation 2.4) and relative hydraulic conductivity differences (equation 2.6) for the random pilot points and D-optimality based pilot points for all four scenarios for 30 random trials. For fair comparison of the same GA parameters are used for the hydraulic conductivity search at the D-optimality based pilot points and random pilot points. Overall, the D-optimality based pilot points yield several orders of magnitude better fitness than random pilot points. However, occasionally, randomly selected pilot point locations give better fitness value. For example, in scenario 3, approximately one third the thirty trials result in slightly better performance in terms of objective value for the random case. However in relative K error comparison SBM shows better result at the same locations. When initial
hydraulic conductivity distribution is significantly different from the unknown reality (scenario 2 and 4), SBM provide more accurate hydraulic conductivity estimation with minimum fluctuations. When comparing the cases with different flow conditions, mean non-uniform flow cases (scenarios 3 and 4) result in slightly larger fluctuations in hydraulic conductivity estimates and larger hydraulic head errors for both SBM and Random. Other than that, different flow conditions do not significantly influence hydraulic conductivity estimation for these problem sets. From these results, we can conclude that, in general, the D-optimality based pilot point selection is superior to random selection and is a viable and promising alternative to other methods of pilot point selection for various conditions.

<Figure 2.9>  Head error values from pilot point locations from D-optimality (Det-PPL) and sensitivities of randomly selected locations (Random)
The hydraulic conductivity contour maps of unknown reality, initial K-field, and the solutions obtained with and without the magnitude constraint for K (equation 2.5) are presented in figure 2.10 for scenario 2. For this illustration, scenario 2 is chosen over other scenarios since its initial K-field is very different from the unknown reality thus presenting a more difficult problem. Visually examining these figures, we can see that both the constrained and unconstrained searches show a dramatic improvement over the initial K-field in predicting the unknown reality. Also, both methods identify the pattern of the K-variation extremely well particularly on the left side. The head error values of both these cases are three orders of magnitude better (5.09E-4) than the value corresponding to the initial K-field (1.50E-1). However, average error in the actual K-values is higher for the unconstrained case (0.7345) than the constrained case (0.4605). Also, the variance of K-field between different search trials is also significantly reduced for the constrained case (data not shown). This shows that using prior information to constrain the K-values will improve the estimate of K-field especially in the absence of tracer data (i.e., concentration measurements). In real problems, geometric mean of K can be obtained if some hydraulic head measurements and average flow velocity are available. In highly heterogeneous aquifers, however, using this constraint can lead to erroneous estimates of K-field. In these cases, every effort should be made to collect tracer measurements in addition to head measurements that can be used in the K-field estimation process.
<Figure 2.10> Contour maps of scenario_2 (Top left: unknown reality, Top right: initial contour, Bottom left: without penalty, Bottom right: with penalty)

2.4.4 Comparison with sequential search method

Here we compare the effectiveness of our D-optimality based PPM method with the sequential search method developed by Rama Rao et al. (1995). This method sequentially repeats the process of searching for points with the highest sensitivity and then adjusting the conductivity value at these locations until no further improvement is possible. The results are presented for scenario 2 in table 2.3 and figure 2.9. From Table 3, we see that while the fitness is improving with increasing pilot points, that even after 4 pilot points, the fitness value is still three orders of magnitude larger than the D-Optimality based method.
### Table 2.3: Head error values with sequentially selected points

<table>
<thead>
<tr>
<th>No.</th>
<th>X</th>
<th>Y</th>
<th>Hydraulic Conductivity (K)</th>
<th>Head error</th>
<th>Average error of actual Ks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>4</td>
<td>49.245</td>
<td>0.112</td>
<td>1.881</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>2</td>
<td>61.857</td>
<td>0.112</td>
<td>1.896</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>84.744</td>
<td>0.104</td>
<td>2.002</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>4</td>
<td>71.586</td>
<td>0.092</td>
<td>2.010</td>
</tr>
</tbody>
</table>

*Figure 2.11* Sequentially searched pilot point locations with the selection order and final hydraulic conductivity distribution (Using four selection)
A D-optimality based pilot point method is developed in this study for hydraulic conductivity estimation from existing K and head measurements. D-optimality based sensitivity, which includes covariance terms between potentially selected pilot points, satisfies the requirements of maximizing sensitivity with minimum correlation between pilot points. Since maximally sensitive points require minimal perturbation to match head measurements, the covariance structure of original measured hydraulic conductivities is preserved well. Also, since the points selected using this criterion are minimally correlated the estimated conductivities are closer to the unknown reality. Four synthetic scenarios with different initial K-field distributions and different flow conditions are examined. In all four scenarios, D-optimality based PPM provides better results than a random set of pilot points and different flow conditions have minimal influences on
hydraulic conductivity search process. Furthermore, SBM is shown to perform much better than a previously developed sequential search method.

There are several limitations of this study that could be explored in the future:

- The impact of the accuracy of the sensitivity matrix $X$ (equation 2.3) has not been examined. A major consideration in the accuracy is the size of denominator ($\Delta K$) used in calculating the sensitivity matrix $X$. If the head response is nonlinear to changes in $K$ values, the fixed $\Delta K$ value of 1 is used in this study might not be appropriate if the perturbation of hydraulic conductivity values at the pilot points in the second step are of different magnitude.

- Comparison with another method such as simultaneously searching for pilot point locations and their hydraulic conductivity values to match the head measurements has not been pursued. This will be investigated in a subsequent paper.

- The effect of incorporating tracer concentration measurements in the pilot point selection and hydraulic conductivity estimation has not been studied. This will be investigated in a subsequent paper.

- Effectiveness of this method for different test problems, especially those with larger mesh sizes need to be investigated. This could be pursued as a future study. Computational improvement using parallel computing and/or surrogate modeling should be considered to reduce the computational burden of the large number of intensive forward model calculations involved in this method.
2.6 References


Fedorov, V.V. and A. Pazman, Design of physical experiments, Fortschritte der Physik 16, 325-355, 1968.


Oliver, Dean S., Nanqun He, and Albert C. Reynolds, Conditioning permeability fields to pressure data, 5th European Conference on the Mathematics of Oil Recovery, Leoben, Austria, 3-6, Sept. 1996.


Vesselinov, Velimir V., Shlomo P. Neuman, and Walter A. Illman, Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff equivalent parameters, high-resolution stochastic imaging and scale effects, WRR, Vol. 37, No. 12, 3019-3041, Dec. 2001b.


Chapter 3: Development of a simultaneous search based pilot point method (SSBM) and its comparison with the sensitivity based pilot point method (SBM)

3.1 Introduction

Hydraulic conductivity is a significant system parameter required to represent groundwater flow and contaminant transport. For most sites, it is difficult to obtain an adequate amount of system parameter information (i.e. hydraulic conductivity and porosity) needed to design more accurate strategies of remediation. On the other hand, state variables or dependent variables such as hydraulic head and contaminant concentration measurements are readily available. Many studies to characterize the subsurface have been performed to overcome deficient information about the unknown system parameters using known state variables in the manner of inverse problem. To solve inverse problems, researchers have developed various techniques (hereafter known as “inverse techniques”) over several decades. Inverse techniques for subsurface characterization can be categorized by objective criteria or parameterization. Chavent (1979) classified inverse techniques using objective criteria: equation error criterion (direct method) or output error criterion (indirect method). Since the indirect method is convenient to apply to general optimization techniques, most of studies have been focused on using output error criterion based on the difference between observed and calculated information. Additionally, this indirect method is applicable to conditions where observations are limited. However, most minimization problems for hydrogeologic parameter identification are usually nonlinear and non-convex (Yeh 1986), and need appropriate minimization tools. Based on parameterization (McLaughlin and Twonley 1996), there are three different methods: Zonal methods, Geostatistical representation, and hybrid methods. Zonal approaches primarily used to stabilize inverse problem (Stallman, 1956) divide the site into a number of discretized blocks corresponding to the parameters required to represent a site. Each divided zone is assumed to have homogeneous hydrogeologic properties. Geostatistical representation generates a
stationary random parameter field or a deterministic hydraulic conductivity field using kriging application based on measurements in the domain. Typically a zonal approach has better performance including measurement errors although it has limitations to describe a fairly complex aquifer. Contrarily, geostatistical representation performs well in large-scale heterogeneous conditions since it is more efficient when expressing spatial variables in the domain. Pilot point method (PPM) is one of the popular non-linear indirect methods using geostatistical representation. De Marsily et al. (1984) introduced this method in order to minimize uncertainty with enhanced heterogeneity recognition. Pilot points in which hydraulic conductivities will be perturbed are selected at where hydraulic conductivity measurements are not occurred. The perturbations at pilot points to find optimal hydraulic values to minimize the difference of calculated and observed hydraulic heads in the domain. Hydraulic conductivity values at points excluded from pilot points and measurement points are obtained from kriging function as an interpolation method. Based on this procedure, calculated hydraulic conductivity distribution is generated much closer to the unknown reality. Hydraulic conductivity measurements with the final perturbed values at the pilot points can effectively represent the heterogeneity of the unknown reality.

As previously mentioned, most minimizations of groundwater inverse problems are nonlinear and non-convex. To overcome these conditions, various optimization algorithms have been applied and modified. The basic optimization concept of an indirect inverse technique using different algorithms starts from initial parameter measurements; minimization algorithms iteratively improve these initial parameters to honor observed secondary information. After several automated calibrations for hydraulic conductivity estimation (Emsellem and De Marsily 1971, Bard 1974, and Yeh 1975) have been performed beyond analytical methods, researchers have continually applied different optimization techniques such as gradient-based methods, stochastic methods, and global search methods (e.g. genetic algorithm, and simulate annealing) to minimize output error to find optimum parameters. Different optimization techniques: Quasi-Newton (Chavent
and Bissell 1998, and Certes and de Marsily 1991), Gauss-Marquardt-Lavenberg (Doherty 2003, Alcolea et al. 2006), and Steepest decent (LaVenue and de Marsily 2001) of gradient-based methods have been used in hydraulic conductivity (or permeability) search problems as optimization tools. In gradient-based method usage, generally, hydraulic conductivities are updated with the information of search directions and specified step sizes. Especially, the gradient-based method was used from the invention of the pilot point methods (De Marsily et al. 1984) and as well as other researchers (Fasanino et al. 1986, RamaRao et al. 1995, LaVenue and de Marsily 2001, Hernandez et al. 2003, and Doherty 2003). Stochastic methods, particularly, are adopted widely where measurement and model errors involved. Maximum likelihood (Carrera and Neuman 1986a,b; Samper and Neuman 1989), maximum a posteriori (Kowalsky et al. 2004), and stochastic partial differential equation (Sun and Yeh 1992) have all also been used in hydraulic conductivity identification. In addition some researchers (e.g. Jang and Choe, 2002) applied mixture of gradient based and stochastic methods to improve searching ability. Recently, global or near global search methods (e.g. Genetic Algorithm (GA) and Simulate Annealing (SA)) are increasingly popular because gradient-based optimization approaches may have difficulties in search within discontinuous decision space and the increased presence of multiple local optima. In addition, global optimizations are flexible to incorporate information from various resources and readily adaptable to massive parallel computing. Mayer and Huang (1999) compared GA with a local search method in parameter identification (i.e. hydraulic conductivity) for zonal and stochastic distributions, and addressed that GA gave a more robust solution in a three dimensional inverse problem. As an example of GA applications in groundwater, we can easily find for groundwater monitoring and remediation design (Ritzel et al. 1994, Cieniawski et al. 1995, Rogers et al. 1995, Aly and Peralta 1999, Chan Hilton and Culver 2000, and Yoon and Shoemaker 2001), and groundwater management (Morshed and Kaluarachchi, 2000, and Rao et al. 2003). In subsurface parameter identification, additionally, a number of researchers have applied genetic algorithms (Bush and Carter 1996, Guerreiro et al. 1998, Mayer and Huang 1999, Romero et al. 2000, Samuel and Jha 2003, Tsai et al. 2005),
simulated annealing (Sen et al. 1995), and hybrid methods (Tsai et al. 2003). However, pilot point methods are rarely used global optimization tools (Karpouzos et al. 2001). The objective of this paper is to introduce another extension of global optimization usage in pilot point methods: simultaneous search of pilot point locations and parameter values.

In the previous work (Chapter 2), sensitivity-based pilot point methods (SBM) produced satisfactory results on searching for pilot point locations and hydraulic conductivity distributions. The sensitivity-based method consists of two separate procedures: the first procedure is the pilot point location search based on D-optimality sensitivity followed by the second procedure, hydraulic conductivity values search by minimizing the sum of squared differences between observed and calculated hydraulic heads. Because the perturbations of hydraulic conductivities at the most sensitive pilot point locations could give substantial reduction of difference with minimum influences on the covariance structure of hydraulic conductivity measurements. However, there is an interesting point we want to investigate for hydraulic conductivity searching procedure. In terms of D-optimality, randomly selected pilot point locations did not achieve good results in the first procedure. However, some random pilot point locations gave better results in the second procedure. Presumably, in the first procedure sensitivity is based on the fixed $\partial K$ value (in the previous work, $\partial K$ was fixed to 1), which is subtracted from the initial hydraulic conductivity value in the denominators of the sensitivity matrix. However, in the second procedure, the changes of hydraulic conductivity values are within several orders magnitude by using GA. The differences in size between $\partial K$ in the first and size of searched hydraulic conductivity in the second procedures undermine the ability of sensitivity analysis in the first procedure. Therefore in this paper, a simultaneous search of pilot point locations and hydraulic conductivities, without considering the size of $\partial K$ for sensitivity analysis, is developed and compared to the sensitivity-based method.
3.2 Methodology

3.2.1 Simulation-Optimization Model

The major components of the simulation-optimization model for this study are a numerical groundwater flow model, a measurements (i.e. hydraulic conductivity) interpolation method as an input parameter generator, and a genetic algorithm optimization tool. For the numerical groundwater flow, the finite difference method with central approximations is applied to the partial differential equations. Gaussian elimination is used to obtain calculated head values at whole grid points from given boundary conditions. The synthetic groundwater flow field for this study is confined, heterogeneous, and isotropic. As an interpolation method for hydraulic conductivity distributions, an ordinary kriging is adopted to obtain unbiased hydraulic conductivities for all grid points from several observed correlated values at certain points. For this study, an exponential model (equation 3.1) as a semi-variogram is selected to verify the correlation structure of measurements or perturbed values at pilot points with measurements, and applied to the generation of hydraulic conductivity distribution. For this interpolation procedure, hydraulic conductivity measurements are assumed log-normally distributed. Initial hydraulic conductivity distribution generated from measurements is iteratively modified using perturbed hydraulic conductivities at pilot points to search for feasible values under perturbation criteria (or an objective function of optimization). The sum of squared residual of hydraulic heads between observed and calculated is used as the objective function. The final hydraulic conductivity distribution is kriged based on measurements with the final perturbed values. As an optimization tool for the perturbation, genetic algorithms are applied to minimize the perturbation criterion.

\[ \gamma(h) = \begin{cases} 
0 & h = 0 \\
 c_0 + c_1(1 - e^{-3|h|/a}) & h \neq 0 
\end{cases} \]

\text{eq 3.1)
where,

\[ a \]: Range (correlation distance of observations)
\[ c_0 \]: Nugget (micro scale variance of observations)
\[ c_1 \]: Sill (maximum of semi-variogram of measurements)
\[ h \]: Distance

A genetic algorithm suggested by Holland (1975) is one of the major stochastic global search methods. The principle of genetic algorithm is based on the natural evolution of the wild environment. As selected species in nature survive for a certain environment, numerically selected values that have the best fit in a particular condition can be found as a result of genetic algorithm operation. The constituents of GA are representation/encoding, operators (selection, mutation, crossover, and elitism), and objective function. The general procedures of GA operation are as follows: 1. Randomly generate initial population using encoded individuals, 2. Select some individuals based on the evaluation of the objective function, 3. Regenerate different populations based on selected individuals using GA operators, 4. Iteratively continue 2 and 3 until objective function meets certain criteria. Representation defines the data type of individual input for objective function. Typical types of representation are binary coding and real coding. In this study real coding is selected to indicate node numbers and hydraulic conductivities. As operators we used selection, mutation, and crossover functions. Selection is performed on the previous generation to choose individuals for the next generation. To generate different populations, mutation gives minor random changes of individuals in the population and crossover combines two individuals. The major benefit of using GA is that a solution from GA is out of a population of candidate solutions instead of one solution from direct search, which gives more exploration in the search procedure. For this study an embedded genetic algorithm in MATLAB is used to select pilot point locations and hydraulic conductivity values simultaneously.
3.2.2 Simultaneous Search-Based Method (SSBM)

Simultaneous Search-Based Method (SSBM) is one of the different searching types in pilot point methods using global optimization approaches. SSBM searches simultaneously pilot point locations and hydraulic conductivities at pilot points. SSBM is only dependent on the sum of squared residual of hydraulic heads between observed and calculated without considering sensitivity values of pilot point locations. Accordingly the size of $\partial K$ in denominator of sensitivity matrix is not considered in SSBM procedure. However, in the GA search, pilot point locations are selected presumably based on the ability to minimize the objective function, even though it is not numerically expressed in the search procedure. This shorted procedure (without consideration of sensitivity) reduces computation time and procedure, since in the sensitivity-based method (SBM) the groundwater flow model should be run twice as much as in the SSBM in order to get the sensitivity matrix of the pilot point locations and hydraulic conductivity values at those selected locations. Figure 3.1 shows a flowchart of the simultaneous search procedure. Initial hydraulic conductivity distribution is kriged from measurements, and pilot point locations and values are added through the GA searching procedure to minimize objective value. Thus in GA search, pilot point locations and hydraulic conductivity values for selected locations are iteratively modified to meet the GA termination criteria. In this experiment we assume that three pilot point locations are optimum number for searching the best hydraulic conductivity field.
3.3 Numerical Experiments

For numerical tests, a 2-dimensional synthetic site (700 m by 1000 m) is generated and divided into 8 by 11 grid points. As boundary conditions for this numerical site, Dirichlet (left and right sides of boundary) and Neuman (top and bottom sides of boundary) boundaries are applied. Specially leftmost and rightmost sides for Dirichlet boundaries have 20 m and 15 m constant hydraulic heads respectively. We also assume the conditions of this synthetic groundwater flow field to be a confined, isotropic, and heterogeneous aquifer.
3.3.1 Scenarios

Based on nine different hydraulic conductivity values at well-distributed locations (data not shown) in a given site, ordinary kriging as an interpolation method using an exponential model (equation 3.1) is employed to generate an unknown reality with the certain parameter values such as range (3.4026) and sill (0.8619). The hydraulic conductivity contour map of the generated unknown reality is presented in Figure 3.2. Hydraulic conductivities of the unknown reality are distributed between 5 and 85 mm/second, which can be recognized as a confined sandy aquifer. Figure 3.2 shows dark blue spot as lower hydraulic conductivity region in right upper side, and left lower side of contour map presents higher hydraulic conductivity region.

![Figure 3.2](image)

<Figure 3.2> Hydraulic conductivity distribution of unknown reality

From this unknown reality, we generate four different scenarios in order to verify the SSBM procedure. To generate four different scenarios, we randomly select four different hydraulic conductivity values out of the unknown reality and assume that those four selected points and values are measurement points and true values of hydraulic
conductivity in the field. Then, the initial hydraulic conductivity distributions presented in Figure 3.3 of four different scenarios are generated using the ordinary kriging based on the measurements. Scenario_2 and scenario_3 have totally different pattern of hydraulic conductivity distributions, because these scenarios exclude point (6,6) that has significantly lower hydraulic conductivity values to compare to other locations in the unknown reality. But scenario_1 and scenario_4 are somewhat similar to the unknown reality.

<Figure 3.3> Initial hydraulic conductivity distributions (From left top to right bottom: Scenario_1, Scenario_2, Scenario_3, and Scenario_4)

For more verifications of the performance of SSBM in terms of minimum fitness values, the objective functions from SSBM are compared to those from random pilot point locations. In random pilot points selection, three pilot point locations are selected for
thirty trials without any sensitivity criteria. Random pilot point location selection followed the given steps: 1. Select pilot point locations randomly from selection pool consisting of node number of grid points excluding hydraulic conductivity and hydraulic head measurement locations. 2. Search hydraulic conductivity values using GA for previously selected pilot point locations. 3. Give average minimum fitness value for thirty trials.

3.3.2 Objective and Penalty

We are using the same objective function of the sensitivity-based method in searching for the hydraulic conductivity values at selected pilot point locations. The sum of difference between observed and calculated hydraulic heads is the major objective function for SSBM without any sensitivity calculations. As previously mentioned, pilot point locations searches are totally dependant on the efficiency in the minimization of the given objective function, therefore in SSBM, less restrictions are applied on the pilot point location search and more freedom to generate better hydraulic conductivity are expected. Equation 3.2 presents the objective function for this study. In future study, we can add more secondary information such as tracer or contaminant distribution to produce a hydraulic conductivity field much closer to the unknown reality.

\[
\text{MINIMIZE } obj = \sum (h_{\text{observed}} - h_{\text{calculated}})^2 \quad \text{eq } 3.2
\]

Where

- \( h_{\text{observed}} \): Observed hydraulic heads
- \( h_{\text{calculated}} \): Calculated hydraulic heads

We penalized the GA search for two major purposes. Firstly, the GA search for pilot point locations should be within the parameterized grid points. When selected pilot points are outside the node boundaries, we assign a penalty to guide the GA search within the
node boundaries of the modeled site. In addition, the hydraulic conductivity search is penalized based on the measurement of the geometric mean of hydraulic conductivities. Geometric mean of hydraulic conductivity is applied to describe the average groundwater flow. Simply, this geometric mean value is obtained from hydraulic conductivity measurements at different locations using the Darcy’s equation. The application of measured geometric mean values for the criteria of penalization of searching hydraulic conductivities honors hydraulic conductivity measurements from the field. On the other hand, this restriction can possibly lessen the searching ability, though for this study the geometric mean penalty function works better to generate a hydraulic conductivity field close to the unknown reality. The range of searching hydraulic conductivities is 75 percent of the measured geometric mean for this study. Equation 3.3 shows the geometric penalty function.

\[ \text{if } \frac{|\text{gm} - \text{gp}|}{\text{gm}} > 0.75 \] 

\[ \text{obj} = \text{obj} + \frac{|\text{gm} - \text{gp}|}{\text{gm}} \]

end

... 

where

- \( \text{gm} \): geometric mean of hydraulic conductivity measurements
- \( \text{gp} \): geometric mean of hydraulic conductivities at pilot points

### 3.4 Results

#### 3.4.1 SSBM and RANDOM
Firstly, we compared minimum fitness values from randomly selected pilot point locations to the SSBM solutions. In Figure 3.4, the minimum fitness values of random pilot point locations for thirty trials are presented as diamond shapes, and empty triangles represent the average minimum fitness values. Additionally, the results of SSBM are presented in square dots with average minimum fitness values in an empty circle. From these results clearly we can see that randomly selected pilot point locations, which have been used widely in the past (De Marsily et al. 1984, Fasanino et al. 1986, Doherty 2003, and Hernandez et al. 2003), are less able to give more precise information for four different scenarios. The differences of the average minimum fitness between SSBM and random locations are approximately one and a half or two orders of magnitude. Even though scenario_2 and scenario_3 shows some better results with random locations, these occasions are very rare and insignificant from the point of view of this experiment. Empirical selection of pilot point locations is possible to overcome this shortcoming. However, it may increase potentials to give more uncertainties in the hydraulic conductivity distribution. Thus several researchers used sensitivity analysis (LaVenue and Pickens 1992, RamRao et al. 1995a, LaVenue et al. 1995b, and Hendricks-Franssen 2000) for pilot point location selection. However, from previous experiment (Chapter 2) sensitivity analysis also has a limitation in searching for optimal hydraulic conductivities. As briefly explained in the introduction, a fixed size (changes of hydraulic conductivity values in sensitivity matrix) of denominator in the sensitivity measurement has been used and this can minimize the ability to search better hydraulic conductivity distributions, because in the second procedure hydraulic conductivity values are explored without searching ranges. In other words, different sizes of the denominator in sensitivity matrix can lead to other pilot point locations, which means the functional expectation of the first selection of sensitive pilot points can be weaken in the second procedure. Therefore instead of using sensitivity measurement, SSBM searches simultaneously for pilot point locations and hydraulic conductivity values without sensitivity consideration and will be compared to SBM in next section.
Final pilot point locations using SSBM in different scenarios are presented in Figure 3.5. As a result of scenarios 1 and 2, the three selected pilot point locations keep distance to the hydraulic conductivity measurement locations that may reduce the direct effect on hydraulic conductivity measurements during optimal hydraulic conductivities search. In general, to reduce direct effect (i.e. introducing more uncertainties on prediction) on hydraulic conductivity measurements, a certain distance between the pilot point and hydraulic conductivity measurement locations has to be kept as scenario 1 and 2 shown in Figure 3.5. However, scenarios 3 and 4 have gathered pilot points at one specific location. In terms of computation, these points are redundant, since these can be replace to one pilot point. In addition, these points may not give enough recognition of
heterogeneity in domain. In the unknown reality the average hydraulic conductivity value is 57.058 and at the point (6,6) the hydraulic conductivity changes dramatically to a value of 7.255 shown in Figure 3.2. Because the point (6,6), which has substantial influence on the secondary information (hydraulic head), is not included in the hydraulic conductivity measurement points for scenario_2 and scenario_3, the search process locates a pilot point exactly (6,6) or very nearby to that significant location. However, in the case of scenario_4, selected pilot point locations are gathered together, which can be reduced to one pilot point locations instead of two points at same location. This matter can be considered for further study.

<Figure 3.5> Final pilot point locations using SSBM with measurement points of hydraulic conductivities and hydraulic heads for four different scenarios
Figure 3.6 shows the final distribution of hydraulic conductivities for different scenarios found using SSBM. This final contours using SSBM are based on the best result of thirty trials for each scenario in terms of minimum fitness values. Every contour map has a similar pattern to the unknown reality except scenarios 3 and 4. However, especially scenario_4 has a very small minimum fitness (1.55E-05). Thus in terms of minimization of fitness, SSBM works very efficiently. From this result we can say that the SSBM is properly searching for hydraulic conductivity distribution while only depending on hydraulic head measurement information. To have a better representation of hydraulic conductivities, we probably need to add more information such as contaminant/tracer concentration as additional secondary information.

<Figure 3.6> Final hydraulic conductivity contour maps (From top left to right: Scenario_1, Scenario_2, Scenaro_3, and Scenario_4)
3.4.2 SSBM and SBM

To compare the SSBM to the SBM we used two different criteria: average minimum fitness value and average hydraulic conductivity difference from the unknown reality for each scenario.

As previously described, four different scenarios are generated from four selected measured hydraulic conductivity values out of known values. Each scenario we compared hydraulic conductivity values at whole grid points for average hydraulic conductivity difference. Based on average hydraulic conductivity difference, SBM and SSBM are compared. Equation 3.4 is the average hydraulic conductivity difference.

\[
\text{Average } K \text{ diff} = \frac{\sqrt{\sum_{i=1}^{N_{av}} (UV_i - CK_i)^2}}{N_{av}} \quad \text{eq 3.4)}
\]

where

- \( UV \): Known hydraulic conductivity at whole grid points
- \( CK \): Searched hydraulic conductivity
- \( N_{av} \): Number of total known hydraulic conductivity

In terms of the average minimum fitness value, the SSBM produces better results for all scenarios except scenario_2, although it has more variance. Simply, the SSBM can find the best solution, but has an increased chance to find worse solutions. The SBM, however, is able to minimize the fitness function as much as the SSBM with less variance, since the result from the SBM dose not show big fluctuations in minimum fitness values. The minimum fitness value for scenario_3 seems outrageously high compared to the other scenarios. Our overall goal is to find a hydraulic conductivity field close to the unknown reality for a field application. The lower graph of Figure 3.7 shows the average difference of hydraulic conductivities, which contrasts with the result of
minimum fitness value. SSBM produces similar values of average hydraulic conductivity difference to what the SBM generates, though SSBM has larger variances of hydraulic conductivities. Even though SBM found worse results in terms of minimum fitness values, this method produces hydraulic conductivity distributions with less variance, which means that SBM is more stable in searching hydraulic conductivities for a real-world field.

![Average Minimum Fitness](image1)

![Average K Difference](image2)

*Figure 3.7* Average minimum fitness from thirty trials and average hydraulic conductivity difference
In scenario_3 we can clearly see that SSBM needs additional information or more constraint to decrease the variance and to find a better way to search for hydraulic conductivity. In addition, Figure 3.8 shows the final hydraulic conductivity distribution of scenario_2. The contour maps from SBM and SSBM follow the pattern of the unknown reality. Visually, SBM generates a much closer hydraulic conductivity distribution. The minimum fitness value from SBM ($2.94 \times 10^{-4}$), however, is one order of magnitude larger than the minimum fitness from SSBM ($2.38 \times 10^{-5}$). This result shows that SSBM has an increased possibility to produce non-unique solutions. In addition, the fixed denominator size for sensitivity matrix in SBM is insignificant to the final hydraulic conductivity distribution search.

<Figure 3.8> Contour maps of unknown reality, initial distribution of scenario 2 (Top left and Top right) and final distribution from SBM and SSBM respectively (Bottom left and bottom right)
3.5 Summary and Final Remarks

For reliable and efficient groundwater management, accurate subsurface information is necessary. Specifically, we have focused on finding hydraulic conductivities, which are highly correlated to the groundwater flow and contaminant transport. As a subsurface characterization tool for a heterogeneous environment, the pilot point method was applied and improved using genetic algorithm as a more efficient search technique. In our study we adopted simultaneously searched pilot point locations and hydraulic conductivity values at those selected locations, and named it the simultaneous search-based method (SSBM). The results of the SSBM show that the application of random locations as pilot points is not as capable at producing an accurate characterization of the subsurface with same conditions and SSBM gives better minimum fitness values. Additionally SSBM was compared to the previously developed sensitivity-based method (SBM) (Chapter 2) in terms of average minimum fitness value and average hydraulic conductivity difference for four different scenarios. When considering only average minimum fitness values, SSBM has more variance, even though generally SSBM shows better performance. However when a different criterion, average hydraulic conductivity difference, was applied, SSBM has larger differences for the given scenarios. Therefore, even though SSBM performs well enough to be an alternative to SBM, it requires more information or more constraints in order to provide a hydraulic conductivity distribution much closer to the unknown reality. For the further studies, non-uniqueness of results from both SSBM and SBM can be studied using the Evolutionary Algorithm to Generate Alternative (EAGA, Zechman and Ranjithan 2004), because non-uniqueness is major concern of inverse problems and if we have more information, such as alternative hydraulic conductivity distributions, we can predict or apply remediation designs more effectively. In addition, the efficiency of application of additional information such as contaminant/tracer concentration and dispersivity can reduce uncertainty can be further explored to develop more field applicable methods.
3.6 References


Dakuo, He and Wang Fuli, Feedback penalty function based on genetic algorithm, Proceedings of the 4th World Congress on Intelligent Control and Automation, Shanghai, P.R. China, June 10-14, 2002.


Chapter 4: Extension of SBM and SSBM to use both hydraulic head and tracer measurements

4.1 Introduction

Pilot point method (PPM) is a well-developed inverse method designed to present hydrogeologic heterogeneity under either deterministic or stochastic conditions while honoring the hydraulic conductivity measurements. De Marsily et al. (1984) introduced this method to increase the ability of heterogeneity recognition while minimizing uncertainty, rather than using zonal method or stochastic inversion, which may not represent heterogeneity over the real field due to the possible large homogenous grids, linearization between hydraulic conductivities and heads, and small variance assumption of hydraulic conductivity distribution (Certes and de Marsily 1991, Rama Rao et al. 1995, LaVenue et al. 1995, and Hendricks-Franssen 2000). Even though this method decreased some of the drawbacks of other methods, a considerable degree of non-uniqueness and uncertainty still remained in many practices.

Thus researchers have attempted to reduce non-uniqueness and uncertainty in various aspects of the problems. For instance, plausibility terms defined as boundaries of search spaces, or solution smoothness, have been used in PPM (Rama Rao et al. 1995, LaVenue et al. 1995, and Gomez-Hernandez et al. 1997) and also in other parameterization applications (Carrera and Neuman 1986 a, b). Even though the estimated hydraulic conductivity distribution efficiently minimizes the objective function (e.g. difference of the observed and calculated secondary information), occasionally the result may not be realistic. In addition, unbounded fluctuation of hydraulic conductivity has more chances to produce more uncertainty. According to the Rama Rao et al. (1995), plausibility terms are embedded in pilot point method because the generation of hydraulic conductivity fields are honoring the direct measurement of hydraulic conductivity values; the covariance shape of observation of hydraulic conductivity is maintained; and predefined
searching spaces limit the searching range close to reality. Also the number of unknown parameters is significantly related to reduction of uncertainty inherent in the system modeling error and the parameter error. If we increase the number of unknown parameters, such as the number of pilot points, generally it will reduce the system modeling error however the parameter uncertainty will be increased. Therefore the number of unknown parameter should be decided depending on the efficiency of reducing both the system modeling error and parameter error (Yeh 1986). In addition, the quality and quantity of observations can significantly reduce the model error and parameter uncertainty. More and better observations allow us to define more unknown parameters in the search procedure and directly improve the objective criteria. Furthermore, the amount of available data is significantly related to the non-uniqueness in a model. Therefore many researchers have applied effort to have better observations.

For the quality improvement of measured data, sensitivity techniques have been used to minimize the effects of the observed secondary information errors, because small changes of hydraulic conductivities at sensitive pilot point locations highly affect the secondary information changes and vice versa. In other words, even though measurements of secondary information contains some degree of errors, the optimized hydraulic conductivities will not be highly influenced from it. Additionally, sensitive techniques minimize the fluctuation of the covariance structure of measured hydraulic conductivity because of minimal changes of final hydraulic conductivity values from the measured. LaVenue and Pickens (1992) first adopted this sensitivity terms in pilot point location selection and it is continually used by other researchers (RamaRao et al. 1995, and LaVenue et al. 1995). In their method, a sequentially selected pilot point is successively added into the pool of pilot point locations followed by hydraulic conductivity search procedure for this pilot point. After finding an optimal value of hydraulic conductivity at the additional pilot point, they assumed the added point as one of the measurements and keep searching for more pilot points until optimal objective values are fulfilled. Also, Jung and his colleagues (Submitted) demonstrated that one collection of pilot point
locations was found to produce hydraulic conductivity distribution close to the unknown reality using D-optimality criterion for searching highly sensitive pilot point locations. They showed that D-optimality provided a robust method in sensitivity based pilot point technique under different conditions. Detail of D-optimality based pilot point methods will be presented in the Calibration section.

As a way to improve the quantity of observation data to reduce parameter errors and model non-uniqueness, coupled inverse models have garnered more attention. Previously most efforts on inverse problems have been concentrated on flow equations because the hydraulic head distributions are directly associated with the hydraulic conductivities. In addition, several issues are generated in the application of coupled inverse problems, for example: various dimensions, scales, and accuracies of observed data; cross effect between observed state variables and system parameters; many possible selections of performance criterion; and many options of experimental design (Sun and Yeh 1990, and Medina et al. 1990). Beside hydraulic head values, researchers have applied additional information (i.e. steady state temperatures (Woodbury et al. 1987, and Woodbury and Smith 1988), mass transport (Strecker and Chu 1986, and Van Rooy et al. 1989), solute arrival time quantile (Harvey and Gorelick 1995), etc.) for coupled inverse problems.

Most coupled inverse problems were flow-transport inversions. Before combining two different simulations, Murty and Scott (1977) and Umari et al. (1979) carried out inversion of mass transport to estimate the dispersivities of a hypothetical 2-dimensional aquifer based on well-known flow field. After these trials, Strecker and Chu (1986) combined groundwater flow and contaminant transport code and used two distinguished stages to identify flow and transport parameters. Firstly, transmissivity was identified from only head observations and the final result from this first step was used in the second stage for dispersivity coefficient identification from concentration observations. The reasons they separate these two steps are the slow convergence of the searching procedure and large amount of time consumed running the transport code for every
Following the earlier work, Van Rooy et al. (1989) also applied two stages for flow and transport parameter estimation sequentially except that different kinds of parameters; source concentration and dispersivity, were estimated in the second stage. However, we can say that the two methods above did not fully use the available information, because the contaminant transport is also highly related to the condition of transmissivity distributions. Also we can see several trials that show the possibility of transport simulation usage for flow parameter identification: Sun and Yeh (1990) demonstrated that flow parameters could be estimated from hydraulic head and/or contaminant concentrations; Mishra and Parker (1989) compared sequential run of flow and transport model to the coupled process from transient unsaturated flow simulations and their results showed that simultaneous estimation has smaller estimation error in the identification of hydraulic and transport parameters; Median et al. (1990) presented additional concentration data along with the head observation and improved the ability of model parameter identification in terms of estimated parameter error reduction to the unknown reality; and Wen et al. (2002) added transient tracer data and found the combined information improved transport prediction as well. As suggested, Wagner and Gorelick (1987), and Medina and Carrera (1996) also utilize all possible information simultaneously in groundwater flow and contaminant transport to estimate model parameters and their uncertainty. Xiang et al. (1992) used both observation data for the parameter estimation using robust $L_1$ parameter estimator for heterogeneous inverse aquifer. Furthermore, Wagner (1992) searched contaminant location and release history using the identified model parameters from the coupled inverse model.

In a different sequence of coupled inverse problems, Keidser and Rosbjerg (1991) used the two-stage feedback optimization to find out final log transmissivity. In the first step they used the information of both hydraulic heads and concentration to estimate hydraulic conductivity distribution that was then applied to identify transport parameters such as dispersivities and source concentration in the second step. Based on estimated transport parameters in the second step, final hydraulic conductivity distribution was re-estimated.
in the first step using various parameterizations (e.g. pilot point approach, zonal parameterization associated with kriging approach, kriging approach using prior information of the log transmissivity, and pure zonal approach without incorporation of transmissivity observation). In their observation, pilot point method was the best at large local heterogeneity recognition.

The objectives of this study are to improve quality and quantities of observations together using sensitivity based pilot point method with transport information, and propose the optimal technique of hydraulic conductivity identification in coupled inverse method.

4.2 Simulation Model

4.2.1 Forward Model

A finite difference approximation of the differential equation for flow and transport estimation is applied for the numerical simulation. To get hydraulic heads and concentrations at all grid points Gaussian elimination is used. Predefined boundary conditions in the ground water flow simulation are also employed in transport code. In the process of combining these two different codes to make coupled inverse problems, flow velocity data for whole grid points from the flow simulation are transferred to the transport code and applied to the advection terms. Steady state hydraulic head measurements and transient tracer concentrations will be used for coupled inverse method in this study. The storage of concentration can be divided into dissolved and sorbed concentration. For this study, retardation factor is one, which means sorbed concentration is not considered. In addition we are assuming that tracer is used in this numerical study, therefore the first order decay is not considered either. Groundwater governing equation of flow and transport used in this study is shown in equations 4.1 and 4.2.
\[ \frac{\partial}{\partial x} (T_x \frac{\partial \phi}{\partial x}) + \frac{\partial}{\partial y} (T_y \frac{\partial \phi}{\partial y}) = 0 \quad \text{eq 4.1) } \]

\[ D_L \frac{\partial^2 C}{\partial x^2} + D_{TH} \frac{\partial^2 C}{\partial y^2} - \nu \frac{\partial C}{\partial x} = \frac{\partial C}{\partial t} \quad \text{eq 4.2) } \]

where \( T \) is the transmissivity obtained the product of hydraulic conductivity and thickness of confined aquifer, and \( \phi \) is hydraulic heads in equation 4.1. \( D_L \) and \( D_{TH} \) are dispersion coefficient of longitudinal and transverse horizontal direction, and \( \nu \) is flow velocity in equation 4.2.

### 4.2.2 Max/Minimization Algorithm

Based on the numerical formulations of groundwater flow and transport, hydraulic heads, and tracer concentration data at each time step can be obtained and can be the basic criteria for searching for the highly sensitive pilot point locations and the optimal hydraulic conductivity distribution for this study. SBM provides two different procedures (i.e. sensitive pilot point location search, and hydraulic conductivity distribution search). Detail of SBM procedure will be in calibration part. For these two procedures, the indirect inverse problems are formulated as a maximization of the first objective and minimization of the second objective. For optimizing two objectives, a genetic algorithm (Holland 1975) is adopted as a global search method. Most current studies have focused on using local search methods, such as Quasi-Newton (Chavent and Bissell 1998, and Certes and de Marsily 1991), Gauss-Marquardt-Lavenberg (Doherty 2003, Alcolea et al. 2006), and Steepest decent (LaVenue and de Marsily 2001), all gradient-based methods. In the presence of multiple local maxima and minima for pilot points and residual respectively, local search methods may produce non-unique solutions and may not be suitable for this given condition.
4.3 Calibration: Sensitivity Based Method (SBM)

4.3.1 Overview of SBM

The sensitivity based method follows the basic approaches De Marsily et al. (1984) developed as given; first, the initial hydraulic conductivity field is obtained from hydraulic conductivity measurements using ordinary or co-kriging; second, select pilot point locations where hydraulic conductivity values will be perturbed. These selected pilot points are not physically measured points. The significant difference of SBM to the basic approaches is the usage of sensitivity concept and selection of the least correlated set of pilot point locations using D-optimality. The beneficial effects of D-optimality usage will be shown in the explanation of objective functions. The third step is to iteratively run forward models (flow and/or transport) to find the optimal hydraulic conductivities using kriged hydraulic conductivity data with measurements and the newly optimized hydraulic conductivity values at pilot point locations. The criterion of this procedure is the discrepancy between calculated and observed hydraulic heads and/or transport concentrations, which will be minimized. Finally, a finalized hydraulic conductivity distribution will be obtained when the objective function meets the stopping criteria in automated calibration. The stopping criteria of SBM in a Genetic Algorithm application are the consecutive number of generations without objective function improvement, the installed stop time with no changes of objective, and/or total time limit for whole generations.
4.3.2 Objective Functions

In SBM, two sequentially separated procedures are performed based on two different optimization criteria as shown in the flow chart (Figure 4.1). First objective is maximization of determinant of squared sensitivity matrix while determining pilot point locations, and the second objective for the determination of hydraulic conductivities at pilot point locations is the sum of squared residuals of observed and calculated secondary information such as hydraulic heads, tracer concentration, and/or tracer arrival quantile.

As a first objective function, maximum of D-optimality have been used. D-optimality consists of a sensitivity matrix and determinant function as presented in equation 4.3. The maximum of the determinant of the Fisher information \( (X^TX) \), which is proportional to the inverse of covariance matrix, is equivalent to minimizing the norm of covariance matrix (Knopman et al. 1987). Therefore the final selection of pilot points has less correlated pilot points using this objective function. Thus following this technique we are
achieving both criteria as given: selected pilot points that are highly sensitive to the secondary information and a set of pilot points that are less correlated each other. In other words, small perturbations at selected pilot points give enough changes to match calculated values to the measurements of secondary information. Perturbations in hydraulic conductivity at any two points from this set will not produce similar changes in secondary information. In addition, the advantages of D-optimality as a criterion are the computational savings because the calculation of covariance matrix is not required, even though we are considering the covariance in our process, and guarantee of highly sensitive pilot points selection.

For the sensitivity matrix, we are measuring the degree of changes of secondary information by perturbing hydraulic conductivities at the selected potential pilot point locations. In this study hydraulic heads, tracer concentrations, and quantiles of cumulated concentrations are applied to search for the pilot point locations. For hydraulic heads we assume that this groundwater flow is steady state. However tracer concentrations are transient, thus sensitivity matrix of tracer concentration is dependent on measurement time. Furthermore we are able to get arrival time quantiles of concentration (Harvey and Gorelick 1995) as one dependent variable of the sensitivity matrix. These concentration data and quantiles give more information than steady state hydraulic head measurements, since their data sets are dependent on time (e.g. every 10 days measurements) and selected quantile times (e.g. 0.25, 0.5, and 0.75). The formulations of first objective function, D-optimality including various data sensitivity matrixes, and tracer arrival time quantiles are presented as

\[
\text{MAXIMIZE } obj \_1 = \det[X^T X] \tag{eq 4.3}
\]
For the second objective of hydraulic conductivity estimation, a commonly used weighting method is applied to find out the optimal combination of different information sets in terms of minimal residual between measured and calculated data. Relative comparisons are applied to combine various data sets, which means different unit products can be evaluated together. \( W \) is the binary weighting coefficient indicating whether this component is considered in the procedure (i.e. 1 or 0). As previously used to find pilot point locations, there are three different measured data sets to figure out the optimal hydraulic conductivity distribution.
\[
\text{MINIMIZE } \ obj_{-2} = W_a \sum_i \left( \frac{h_i^{\text{obs}} - h_i^{\text{cal}}}{h_i^{\text{obs}}} \right)^2 + W_b \sum_i \left( \frac{C_i^{\text{obs}} - C_i^{\text{cal}}}{C_i^{\text{obs}}} \right)^2 + W_c \sum_q \left( \frac{\tau_q^{\text{obs}} - \tau_q^{\text{cal}}}{\tau_q^{\text{obs}}} \right)^2 \quad \text{eq 4.6}
\]

where,

\[ W_{a,h,or\,c} : \text{Weighting coefficient} \]

\[ h^{\text{obs}}, C^{\text{obs}}, \tau^{\text{obs}} : \text{Observed hydraulic head, tracer concentration, and tracer arrival quantile} \]

\[ h^{\text{cal}}, C^{\text{cal}}, \tau^{\text{cal}} : \text{Calculated hydraulic head, tracer concentration, and tracer arrival quantile} \]

### 4.4 Numerical Examples of SBM

#### 4.4.1 Hypothetical Models

For synthetic model (i.e. unknown reality), we generate a heterogeneous, confined, and isotropic site (1000 m by 700 m) using the kriging method as a tool of interpolation. In the kriging process, the exponential model \( \gamma(h) = c_0 + c_1 \cdot (1 - e^{-3h/a}) \); where \( c_1 \) is sill, \( a \) is range, \( c_0 \) is nugget, and \( h \) is distance) is applied to generate heterogeneous hydraulic conductivity distribution. In this study sill \( (c_1) \) and range \( (a) \) values were 0.8619 and 3.4026 respectively without a nugget effect. The synthetic hydraulic conductivity was distributed between 5 and 85 mm/second. The flow field based on the synthetically generated hydraulic conductivity distribution has two Dirichlet boundaries and Neuman boundaries at leftmost and rightmost sides, and top and bottom sides of the synthetic site respectively. For Dirichlet boundaries 20 m (leftmost side) and 15 m (rightmost side) are constantly applied without changes. In the flow figure, lower hydraulic conductivity location has significant flow bump, other than that mostly groundwater flow does not have vital changes in the given domain based on flow direction.
As previously mentioned, boundary conditions are imported from the flow code and pass the groundwater velocity profile to the advection terms in the contaminant transport code. In this synthetic transport code, we are applying instant a small line source as a tracer injection at the line shown in the Figure 4.3 as a thick black line. We are assuming that we already know other information such as longitudinal and transverse dispersivities (100, and 20), molecular dispersion coefficient (0.01), retardation factor (1), and concentration of injected tracer (200 PPM) and injection period (2 days). The time period for concentration measurements at selected observation locations; the same locations as hydraulic head observations, is every 10 days period from 400 days generation and Figure 4.3 shows the 20, 100, 200 and 300 day concentration contours (from left upper corner to right bottom corner) after injection using the hydraulic conductivity distribution of the unknown reality.
Figure 4.3: Initial tracer injection line with the transport depending on time in the domain (triangles: observation points, and dark line: Tracer injection points in initial injection domain)
4.4.2 Different Scenarios

Based on the unknown reality of the hydraulic conductivity distribution, we generate two different hydraulic conductivity distributions, scenario 1 and 2. Four different locations are selected from unknown reality as hydraulic conductivity measurements, and based on them initial conditions are generated by ordinary kriging function. Each scenario has four hydraulic head and/or concentration observation locations and does not change locations with time. Scenario 1 has closer pattern of hydraulic conductivity distribution to the unknown reality than scenario 2, since scenario 2 does not include the highly significant point (6, 6) as a hydraulic conductivity measurement. Figure 4.4 shows the locations of hydraulic conductivity measurements and observations, and initial hydraulic conductivity distributions of scenario 1 and 2.

![Figure 4.4](image)

Hydraulic conductivity measurement locations and initial hydraulic conductivity distributions based on kriged measurements for scenario 1 and 2
4.4.3 Numerical Comparison

The purpose of parameter estimation in this study is to obtain the optimal hydraulic conductivity distribution close to the unknown reality based on several measurements. Therefore, if we want to see the performance of different combinations of various information sets in SBM, comparison between calculated and observed (in this study unknown reality) hydraulic conductivities is a good indicator. As for numerical indicators, two different terms: average relative hydraulic conductivity difference ($ReK$) and maximum hydraulic conductivity differences ($MaxeK$) for whole grid points are shown in equations 4.7 and 4.8.

\[
ReK = \frac{1}{N_t} \sqrt{\sum_{i=1}^{N_t} (K_{ci} - K_{ri})^2} \quad \text{eq 4.7}
\]

\[
MaxeK = \text{Max} \left[ |K_{ci} - K_{ri}| \right], \quad \forall i \quad \text{eq 4.8}
\]

where $K_{ci}$ and $K_{ri}$ are calculated and unknown real hydraulic conductivity at grid point $i$, and $N_t$ is total number of grid points.

4.5 Results and Discussions

Before we tried to see the performance of coupled inverse problems, we carried out ten different trials of SBM with individual information sets (i.e. hydraulic head, tracer concentrations, and tracer arrival time quantiles). Figure 4.5 gives the results of average of average relative hydraulic conductivity differences ($ReK$) and average of maximum differences ($MaxeK$) between calculated and unknown real hydraulic conductivities. Averages of $ReK$ and $MaxeK$ show the similar pattern in different data applications for both scenario 1 and 2. For scenario 1 when only head information is used for optimizing pilot point locations and hydraulic conductivities at those selected locations, average of
ReK was higher than when other information is used.  Alternatively, for scenario 2 when only head information is used, lower averages of ReK and MaxeK are found. Using only concentration data gives significant reduction in variance of hydraulic conductivity solutions for scenario 1, however in scenario 2, the variance increased with only concentration data application. Furthermore, tracer arrival time quantile produces even higher fluctuations in the hydraulic conductivity search. In scenario 2 when only head information is used, the significant pilot point location (6, 6), which has lowest hydraulic conductivity value, was selected in the pilot point location search. Therefore the final hydraulic conductivity values much closer to the unknown reality and gives less variance of hydraulic conductivity searches. In Figure 4.6, the pattern of final pilot point locations of only head or concentration used is quite similar to each other for both scenarios. When we use coupled inverse problems for hydraulic conductivity estimation, pilot point locations from either head only or concentration only can be used to find optimal hydraulic conductivity distribution. For the second procedure of SBM, using only one information set is not recommended for hydraulic conductivity search, when we have different kinds of information available.

<Figure 4.5>  Comparison between various information based pilot point method using average of relative hydraulic conductivities and average of maximum hydraulic conductivity differences
<Figure 4.6> Final pilot point locations from each information set for scenario 1 and 2 (squares: hydraulic conductivity measurements, triangles: head/concentration observation locations, and circles: optimal pilot point locations)
For the comparison of coupled inverse problems, three different pilot point location sets based on the application of information (i.e. head only, head and concentration, and head and quantile) are considered. Pilot point locations using hydraulic head or concentration are being optimized in very similar positions in the domain for scenario 2. Therefore, except scenario 2 we have three different pilot point location sets. One set of pilot point locations is from only hydraulic head, and the other sets are the combination of final sets from head and concentration individually or head only with tracer arrival time quantile only. The combination of pilot point locations is shown in Figure 4.7. Using the selected pilot point locations, different combinations of information are used to get hydraulic conductivity values. As a conclusive procedure of this study, firstly we only used hydraulic head data and then combined hydraulic head data with tracer concentrations for hydraulic conductivity search. Finally hydraulic head data with tracer arrival time quantile was used for both pilot point locations from only head and the combination from head and quantiles for both pilot point locations and hydraulic conductivity searches. As shown in Table 4.1, when hydraulic head and tracer arrival quantiles are used for head only pilot point locations, the averages of ReK and MaxeK are increased and also the variances of both numerical criteria follow a similar pattern. Furthermore for combined locations based on heads and quantiles, ReK and MaxeK are getting significantly worse with increased variances for both scenarios. The reason for this is possibly that the combined pilot point locations introduce the different covariance structure and gives more uncertainties in the hydraulic conductivity search procedure. However, the coupled inverse problem of hydraulic head and tracer concentration for hydraulic conductivity search is promising the robustness of hydraulic conductivity estimation either initial hydraulic conductivity close to the unknown reality or not. Clearly, in scenario 1, averages of ReK and MaxeK are significantly decreased with minimized variances for both, and in scenario 2, both numerical criteria are close to the best result from head only. When both information of head and concentration is used for both procedure with bigger population size (10000 shown in Table 4.1), the accuracy of searching hydraulic conductivity is increased in scenario 1. Based on these results, selection of pilot point
locations based only on hydraulic heads and fitting of hydraulic conductivities from hydraulic heads and tracer concentrations are recommended in groundwater flow parameter identification.

<Table 4.1>  Coupled inverse problems in different conditions

<table>
<thead>
<tr>
<th>Scenario</th>
<th>PPL Selection</th>
<th>Weight</th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wa</td>
<td>Wb</td>
<td>Wc</td>
</tr>
<tr>
<td>S_1</td>
<td>Head Only</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Head &amp; Con</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Head &amp; Quan</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S_2</td>
<td>Head Only</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Head &amp; Quan</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Furthermore we tried to compare SBM with simultaneous search based method (SSBM) in coupled inverse problem. SSBM is based on only one objective function instead of using two objective functions as in SBM. The objective function in SSBM is the residual of measured and calculated secondary information. The decision variables in SSBM are

<Figure 4.7>  Combination of pilot point locations from head and concentration or quantile separately (In scenario 1, the combinations of head and concentration or head and quantile are located in the same points)
the pilot point locations and hydraulic conductivity values for those selected pilot point locations, and are optimized simultaneously without sensitivity consideration. Detailed explanation is in the Chapter 3. In this study, increased population size was applied to give more objective function evaluations for the comparison to the SBM, since SBM has two separate procedures and therefore has more calculations. When we compared relative hydraulic conductivities between SBM and SSBM using combined inverse problems (Figure 4.8), for scenario 1, SBM gives more constant relative hydraulic conductivity values and also maximum errors are smaller. As a previous chapter has shown, SSBM sometimes gives a good solution in scenario 1 in terms of maximum error; however, the pilot point locations are consolidating into the upper right corner of the given synthetic site. Figure 4.9 shows the final pilot point locations with the best result based on maximum errors for scenario 1 and 2 from SSBM using the coupled (hydraulic head and tracer concentration) inverse method. The selected pilot point locations for scenario 2 are around the point (6,6), which is a significant location for representing heterogeneity close to the unknown reality as previously mentioned. However these points may destruct the measured covariance structure of hydraulic conductivities and may be redundant when searching for optimal hydraulic conductivities. In other words, we can select one pilot point location near to (6,6) or at the exact point and may produce better hydraulic conductivity prediction.
<Figure 4.8>  Comparison between SBM and SSBM

<Figure 4.9>  Final pilot point locations using SSBM
Unknown Reality

Initial Condition of Scenario 1

Final Distribution of Scenario 1 using SBM (H_HC)

<Figure 4.10> Hydraulic conductivity distributions from SBM and SSBM for scenario 1 and 2
Final Distribution of Scenario 1 using SSBM (HC)

Initial Condition of Scenario 2

Final Distribution of Scenario 2 using SBM (H_HC)

<Figure 4.10> Hydraulic conductivity distributions from SBM and SSBM for scenario 1 and 2 (Continued)
Figure 4.10 shows the final hydraulic conductivity distributions using SBM and SSBM of hydraulic conductivity estimation in combination of hydraulic heads and tracer concentrations. Beside final hydraulic conductivity distributions, semi-variograms based on measured and optimize final hydraulic conductivities at pilot points are presented. These results are the preeminent results from the ten trials for both of SBM and SSBM in terms of ReK and MaxeK. If we see hydraulic conductivity distribution, the unknown reality has two separated section left and right sides. SBM shows the total domain pattern as segregated pattern between left and right sides of domain. However, SSBM has better distribution in the left side of domain and also catches lower hydraulic conductivity distributions around (6, 6) well for both scenarios. Based on the best values of ReK and MaxeK, SSBM produces slightly better results except MaxeK in scenario2 (Data not shown). Furthermore, we need to consider about variance of hydraulic conductivity searches and covariance structure of final hydraulic conductivity distribution. The variances of hydraulic conductivity searches are shown in Figure 4.9. As we found, SSBM has larger fluctuations in hydraulic conductivity search. For this bigger variance, we need to consider to providing the optimization tool more chances or more time to
explore the results before supplying the premature results, which may reduce bigger variance. In terms of covariance structure, SBM follows well to the measured values and produces closer covariance structure to the unknown reality. Covariance structure in SSBM for scenario 2 has significant change in sill value.

4.6 Conclusion

Several techniques have been investigated to improve various aspects of subsurface parameter identification over the recent decades. One method to significantly improve the techniques is the application of coupled inverse problems. In sensitivity based pilot point method (SBM), this coupled inverse problem is applied and compared with other applications. For the first step of the pilot point locations search, three different information sets (e.g. hydraulic heads, tracer concentrations, and tracer arrival time quantiles) individually are applied and compared to each other. In this comparison, pilot point locations based on hydraulic heads has similar trends to the pilot point locations from tracer concentrations. Thus hydraulic head based pilot point locations and the combination of each head and quantiles respectively based pilot point locations were selected and applied to the second procedure comparisons. In the second step, the procedure of hydraulic conductivity search was based on the combination of various information sets in order to give an idea of the optimal set of information based on given pilot point locations. In the results, a set of measurements of hydraulic heads and tracer concentrations was a promising tool for hydraulic conductivity search with the pilot point locations based on heads for either initial hydraulic conductivities close to the unknown reality or not. In this study, plausibility terms are not applied to constrain the hydraulic conductivity search space to make it more similar to the unknown reality, because tracer concentration information inherently provide the magnitude of hydraulic conductivity while hydraulic heads gives the trends of hydraulic conductivity distributions. Tracer arrival time quantiles with hydraulic heads for pilot point locations and hydraulic conductivity search introduced more uncertainties in our case. In addition, SSBM with
additional information (i.e. tracer concentration) performed comparably well to the coupled inverse problems in SBM, and it reduced computational times. However, it still produced more variance of the hydraulic conductivity search process.

In this study applying SBM and SSBM in coupled inverse problems addresses deterministic condition based hydraulic conductivity distribution. As some researchers have mentioned in their studies, the assumption of deterministic condition of hydraulic conductivity distributions may mislead in the design of remediation procedures or risk assessments. Thus, the degree of non-uniqueness needs to be handled in SBM and SSBM. The application of Evolutionary Algorithm to Generate Alternative (EAGA; Zechman and Ranjithan 2004) for searching the degree of non-uniqueness and generating a set of possible alternatives with or without noisy measurements might be a good area for more study. Also regardless of uniqueness problems, parallel computational methods should be considered since global search methods can be too computationally intensive in a large problem.
4.7 References


Chapter 5: Conclusions

Two variations of the pilot point method was developed for hydraulic conductivity characterization and applied to several synthetic scenarios.

As the first method, sensitivity based on D-optimality (SBM) is applied to find the optimal pilot point locations. This criterion attempts to minimize correlation among pilot points while maximizing D-optimality. Final pilot points from this criterion are well distributed in domains. In addition, since the selected pilot points are minimally correlated, they efficiently utilize the secondary measurements and thereby improve the uniqueness properties of final hydraulic conductivities. Since selected pilot points are highly sensitive to the secondary information, only small perturbations are needed to minimize the objective and these small changes of hydraulic conductivities at the pilot points does not substantially influence the covariance structure of the measured hydraulic conductivity values.

The second method, simultaneous search based method (SSBM), is developed to find pilot point locations and hydraulic conductivity values concurrently. Simultaneous search can reduce the time of computation and restriction of search space, since this method does not consider sensitivity of pilot points. However, minimum restriction of pilot point location search may produce the potential of non-unique results. In the numerical experiments, SSBM shows better performance in terms of minimum fitness of secondary information. However, it is found that the comparison between calculated and observed hydraulic conductivities shows considerable differences for given scenarios. Therefore, even though SSBM shows the capability to be an alternative of SBM, it requires more assistance from other information to reduce the variances of the hydraulic conductivity search.

Finally, both SBM and SSBM are extended to use tracer measurements in addition to hydraulic head measurements. In the previously described numerical experiments
hydraulic heads were the only secondary information used to find hydraulic conductivities in the domain. However, hydraulic conductivity distribution based only on hydraulic heads without constraints on its magnitude will lead to unrealistic and non-unique solutions. When tracer concentration data is introduced in the searching procedure, these data can provide information on the magnitude of hydraulic conductivity distribution. Therefore, when tracer concentration measurements are incorporated in SBM and SSBM, magnitude constraints become unnecessary for the hydraulic conductivity search. The combination of hydraulic heads and concentrations show improved robustness of hydraulic conductivity estimation for multiple scenarios.

For further studies in this area, computational improvement can be considered for future studies by using parallel computing and/or surrogate modeling, because pilot point method requires a significant amount of computation for forward calculations and kriging estimations. In addition, the degree of non-uniqueness of SBM and SSBM can be explored by using alternative generation techniques such as Evolutionary Algorithm to Generate Alternative (EAGA; Zechman and Ranjithan 2004).