Applications of Diverse Data Combinations in Subsurface Characterization using D-optimality Based Pilot Point Methods (DBM)

Yong Jung* · G. Mahinthakumar

1Water Resources Research Division, Korea Institute of Construction Technology
2Department of Civil Engineering, North Carolina State University

ABSTRACT

Many cases of strategically designed groundwater remediation have lack of information of hydraulic conductivity or permeability, which can render remediation methods inefficient. Many studies have been carried out to minimize this shortcoming by determining detailed hydraulic information either through direct or indirect measurements. One popular method for hydraulic characterization is the pilot point method (PPM), where the hydraulic property is estimated at a small number of strategically selected points using secondary measurements such as hydraulic head or tracer concentration. This paper adopted a D-optimality based pilot point method (DBM) developed previously for hydraulic head measurements and extended it to include both hydraulic head and tracer measurements. Based on different combinations of trials, our analysis showed that DBM performs well when hydraulic head is used for pilot point selection and both hydraulic head and tracer measurements are used for determining the conductivity values.

Key words : Subsurface characteristics, Pilot point methods, D-optimality, Diverse data sets

1. Introduction

Pilot point method (PPM) is a well-developed inverse method designed to estimate hydrogeologic heterogeneity under either deterministic or stochastic conditions while honoring the existing hydraulic conductivity measurements. De Marsily et al. (1984) introduced this method to increase the ability of heterogeneity recognition with minimizing uncertainty. While this method decreased some of the drawbacks of other inverse methods such as overparameterization considerable degree of non-uniqueness and uncertainty still remained. Researchers have attempted to reduce non-uniqueness and uncertainty in various aspects of PPM. For instance, plausibility terms defined as boundaries of search spaces or solution smoothness, have been used in PPM (LaVene et al., 1995; RamaRao et al., 1995; Gomez-Hernandez et al., 1997; Kowalsky et al., 2012).

The quality and quantity of observations can significantly reduce the model error and parameter uncertainty. 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. LaVenue and Pickens (1992) first adopted this sensitivity terms in pilot point location selection and it is continually used by other researchers (LaVenue et al., 1995; RamaRao et al., 1995). Jung et al. (2011) adopted D-optimality criteria to search a highly sensitive collection of pilot points with respect to the quality improvement of measured data. 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. In the context of groundwater inverse problems, the term coupled inverse models is typically referred to problems where both flow and transport parameters are estimated simultaneously or both flow and transport measurements are involved in estimating either flow or transport parameters.
(Sun and Yeh, 1990). The latter definition is considered in this study. In this context, coupled inverse models increase the measured information beside hydraulic head values, such as steady state temperatures (Woodbury et al., 1987; Woodbury and Smith, 1988), mass transport (Strecker and Chu, 1986), solute arrival time quantile (Harvey and Gorelick, 1995). While coupled inverse problems has been investigated extensively in groundwater modeling (Strecker and Chu, 1986; Mishra and Parker, 1989; Median et al., 1990; Wen et al., 2002; Franssen et al., 2003; Liang et al., 2010), we are not aware of any work in the context of pilot point methods.

The objectives of this study are to (1) extend a previously developed D-optimality based PPM method to utilize tracer measurements in addition to hydraulic head measurements, 2) evaluate its performance for a simple 2-dimensional test problem to estimate hydraulic conductivity.

2. Forward Simulation Model

The pilot point inverse modeling procedure adopted in this paper requires repeatedly executing a forward groundwater model to obtain hydraulic heads and tracer concentrations. It is assumed that the groundwater flow is steady and the tracer is conservative. The domain is assumed to be two-dimensional. The governing equations describing steady state saturated groundwater flow and conservative transport are given by equations (1) and (2):

\[
\frac{\partial}{\partial x} (K b \frac{\partial h}{\partial x}) + \frac{\partial}{\partial y} (K b \frac{\partial h}{\partial y}) = 0 \quad (1)
\]

\[
\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} (D_x \frac{\partial C}{\partial x}) + \frac{\partial}{\partial y} (D_y \frac{\partial C}{\partial y}) - \frac{\partial}{\partial x} (V_x C) - \frac{\partial}{\partial y} (V_y C) \quad (2)
\]

Where \( K \) is the hydraulic conductivity with \( b \), thickness of confined aquifer, and \( h \) is the hydraulic head. \( V_x \) and \( V_y \) are flow velocities in the \( x \) and \( y \) directions that are computed using Darcy’s law from the hydraulic heads \( h \) obtained from equation (1). \( D_x \) and \( D_y \) are dispersion coefficients in the longitudinal and transverse horizontal directions that are dependent on \( V_x \) and \( V_y \). Equations (1) and (2) are discretized using central finite differences and solved in a MATLAB environment.

3. Calibration: D-optimality Based Method (DBM)

The D-optimality based method (DBM) for hydraulic conductivity characterization using hydraulic head measurements has been described in greater detail in Jung et al., 2011. Here we provide a brief overview. DBM involves two major steps: (1) search for a set of pilot point locations that are both sensitive and uncorrelated by maximizing the D-optimality metric (to be described later), (2) search for optimal hydraulic conductivity values at these locations so as to minimize the difference between the observed and calculated measurements. For these two procedures, optimization problems are formulated as a maximization of the first objective (D-optimality - equation (3)) and minimization of the second objective (sum of squared errors - equation (6)). For optimizing these objectives, genetic algorithm is adopted as a global search method.

The DBM approach used in this study follows the basic PPM approach outlined by De Marsily et al. (1984). First, an initial hydraulic conductivity field is obtained from existing hydraulic conductivity measurements using kriging; second, select pilot point locations where hydraulic conductivity values will be estimated; and, third, iteratively run forward models (flow and/or transport) to find optimal hydraulic conductivity values at the pilot point locations such that the resulting kriged field minimizes the discrepancy between calculated and observed hydraulic heads and/or transport concentrations. A finalized hydraulic conductivity distribution will be obtained when the objective function meets the stopping criteria in automated calibration. The stopping criteria of DBM 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.

3.1. Objective Functions

In DBM, two sequentially separated procedures are performed based on two different optimization criteria as shown in the flow chart (Fig. 1). First objective is maximization of determinant of squared sensitivity matrix while determining pilot point locations, and the second objective
Applications of Diverse Data Combinations in Subsurface Characterization using D-optimality based Pilot Point Methods (DBM)


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. Explanations in detail of objective functions are as shown below.

As the first objective function, D-optimality consists of a sensitivity matrix and determinant function as presented in equations (3) and (4). The maximum of the determinant of the Fisher information ($X^T X$), which is proportional to the inverse of covariance matrix, is equivalent to minimizing the norm of covariance matrix (Knopman et al., 1987). By maximizing D-optimality, we are achieving the following two desired goals: (1) selected pilot points are highly sensitive to the secondary information (thus less susceptible to measurement errors), and (2) the selected pilot points are less correlated to each other (less redundancy). In other words, small perturbations of the parameter (i.e., hydraulic conductivity) at selected pilot points will be sufficient to match calculated values to the measurements of secondary information (i.e., hydraulic heads and tracer concentrations). In addition, perturbations in hydraulic conductivity at any two points from this set will not produce similar changes in secondary information. In this study hydraulic heads, tracer concentrations, and arrival time quantiles of cumulated concentrations (Harvery and Gorelick, 1995) are used as measurement quantities to search for the pilot point locations. The concentration data and quantiles generally have more data points than steady state hydraulic head measurements, since they 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

$$\text{MAXIMIZE } obj_1 = |X^T X|$$

$$X_{n \times m} = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \\ \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & \cdots & c_m \end{bmatrix} \quad X_{nq \times m} = \begin{bmatrix} a_{1q} & a_{2q} & \cdots & a_{mq} \\ b_{1q} & b_{2q} & \cdots & b_{mq} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1q} & c_{2q} & \cdots & c_{mq} \end{bmatrix}$$

where

- $n$: number of observation wells
- $m$: number of pilot point locations
- $\tau_{nj}^q$: $q^{th}$ arrival time quantile at observation well $n$

Fig. 1. Flow chart of coupled inverse problems using DBM.
\[ Q(\tau) = \int_0^\infty C(t) dt \int_0^\tau C(t) dt \]  

where

- \( C \): concentration
- \( t \): time
- \( \tau_q \): \( q \)th arrival time quantile when \( Q(\tau_q) = q \)
- \( Q(\tau) \): normalized cumulative concentration

For the second objective of hydraulic conductivity estimation, the optimal combination of different information sets in terms of minimal residual between measured and calculated data is applied using binary coefficient (i.e. 1 or 0). Relative comparisons are adopted to combine various data sets together.

**MINIMIZE** 

\[ \text{obj.2} = W_a \sum_i \left( \frac{h_i^{obs} - h_i^{cal}}{h_i^{obs}} \right)^2 + W_b \sum_i \left( \frac{C_i^{obs} - C_i^{cal}}{C_i^{obs}} \right)^2 + W_c \sum_q \left( \frac{\tau_q^{obs} - \tau_q^{cal}}{\tau_q^{obs}} \right)^2 \]

where,

- \( W_{a,b,c} \): Weighting coefficient
- \( h_i^{obs}, C_i^{obs}, \tau_i^{obs} \): Observed hydraulic head, tracer concentration, and tracer arrival quantile
- \( h_i^{cal}, C_i^{cal}, \tau_i^{cal} \): Calculated hydraulic head, tracer concentration, and tracer arrival quantile

### 4. Numerical Examples of DBM

#### 4.1. Hypothetical Models

For synthetic model (i.e. true values), 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 \( \gamma(h) = c_0 + c_1 \cdot (1 - e^{-h/a}) \); 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 Fig. 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

---

Fig. 2. Hydraulic conductivity distribution and groundwater flow for the true K-field.

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 Fig. 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 true K-field.
4.2. Different Scenarios

Based on the true values of the hydraulic conductivity distribution, we generate two different hydraulic conductivity distributions, scenario 1 and 2. Four different locations are selected from true values 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 true values than scenario 2, since scenario 2 does not include the highly significant point (6, 6) as a hydraulic conductivity measurement. Fig. 4 shows the locations of hydraulic conductivity measurements and observations, and initial hydraulic conductivity distributions of scenario 1 and 2.

4.3. Numerical Comparison

The purpose of parameter estimation in this study is to obtain the optimal hydraulic conductivity distribution close to the true values based on several measurements. Therefore, if we want to see the performance of different combinations of various information sets in DBM, comparison between calculated and observed (in this study true K-field) hydraulic conductivities is a good indicator. As for numerical indicators, two different terms: average of square root hydraulic conductivity difference ($K_{diff}$) and maximum hydraulic conductivity differences ($MaxeK$) for whole grid points are shown in equations (7) and (8).

\[ K_{diff} = \frac{1}{N_t} \sum_{i=1}^{N_t} (K_{ci} - K_{ri})^2 \]  
(7)

\[ MaxeK = \text{Max} [|K_{ci} - K_{ri}|], \forall i \]  
(8)

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

Fig. 4. Hydraulic conductivity measurement locations and initial hydraulic conductivity distributions based on kriged measurements for scenario 1 and 2.
5. Results and Discussions

Before we tried to see the performance of coupled inverse problems, we carried out ten different trials of DBM with individual information sets (i.e. hydraulic head, tracer concentrations, and tracer arrival time quantiles). Fig. 5 gives the results of average of average square root hydraulic conductivity difference ($K_{\text{diff}}$) and average of maximum differences ($\text{MaxeK}$) between calculated and true hydraulic conductivities. Averages of $K_{\text{diff}}$ and $\text{MaxeK}$ 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 $K_{\text{diff}}$ was higher than when other information is used. Alternatively, for scenario 2 when only head information is used, lower averages of $K_{\text{diff}}$ and $\text{MaxeK}$ 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 true $K$-field and gives less variance of hydraulic conductivity searches. 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 DBM, using only one information set is not recommended for hydraulic conductivity search, when we have different kinds of information available.

For the comparison of coupled inverse problems, three

**Table 1.** Coupled inverse problems in different conditions ($W_a$, $W_b$, $W_c$: Weights for hydraulic heads, tracer concentration, and tracer arrival quantile, respectively)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>PPL Selection</th>
<th>$W_a$</th>
<th>$W_b$</th>
<th>$W_c$</th>
<th>$K_{\text{diff}}$</th>
<th>$\text{MaxeK}$</th>
<th>$K_{\text{diff}}$</th>
<th>$\text{MaxeK}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_1</td>
<td>Head Only</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5.46435</td>
<td>182.168</td>
<td>0.664628</td>
<td>23.09931</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.618474</td>
<td>31.4524</td>
<td>0.025131</td>
<td>1.259951</td>
</tr>
<tr>
<td></td>
<td>Head &amp; Con</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.944652</td>
<td>75.4416</td>
<td>1.725637</td>
<td>69.17361</td>
</tr>
<tr>
<td></td>
<td>Head &amp; Quan</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.33119</td>
<td>54.2901</td>
<td>0.423429</td>
<td>21.11354</td>
</tr>
<tr>
<td>S_2</td>
<td>Head Only</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.579468</td>
<td>23.0113</td>
<td>0.084991</td>
<td>0.82505</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.604876</td>
<td>24.1362</td>
<td>0.183736</td>
<td>5.112352</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.589397</td>
<td>53.8188</td>
<td>0.948019</td>
<td>36.9898</td>
</tr>
<tr>
<td></td>
<td>Head &amp; Quan</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>11.0705</td>
<td>355.5363</td>
<td>7.971674</td>
<td>241.2799</td>
</tr>
</tbody>
</table>
different pilot point location sets based on the applications of diverse information sets (i.e. head only, head and concentration, and head and quantile) are considered. For scenario 2, two different pilot point location sets were tested because optimized pilot point locations using hydraulic head or concentration are in very similar positions in the domain. 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. Table 1 presents the final results of averages of $K_{diff}$ and $MaxeK$ with diverse combinations. When hydraulic head and tracer arrival time quantile are used for head only pilot point locations, the averages of $K_{diff}$ 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, $K_{diff}$ 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 conductivities close to the true K-field or not. Clearly, in scenario 1, averages of $K_{diff}$ 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 for GA search (10000 shown in Table 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.

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 D-optimality based pilot point method (DBM), this coupled inverse problem is applied and compared with diverse conditions. 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) are individually 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 true K-field or not. In this study, plausibility terms are not applied to constrain the hydraulic conductivity search space to make it more similar to the true K-field, 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.

References

the worth of concentration data, *J. Hydrol.*, **281**, 281-295.


