

## Linear Programming Method for Investigating the Disposal Histories and Locations of Pollutant Sources in an Aquifer

S. Y. Chang\* and F. R. Kashani

*Civil & Environmental Engineering, North Carolina A&T State Univ., Greensboro, NC 27411, USA*

Received 29 May 2008; revised 30 August 2008; accepted 28 October 2008; published online 12 March 2009

**ABSTRACT.** A linear programming based methodology has been used to identify the unknown sources of groundwater pollution. A source identification model is demonstrated and evaluated using hypothetical groundwater systems for a steady state case and a transient case. The concept of linear super-position is employed to find waste disposal rates at disposal sites from concentrations at measurement wells. In the steady state case, the identification model finds the unknown location and magnitude of leaks in a pipe among the probable leak locations. Contamination data are obtained from sparsely located wells. Subsequently, the transient case source identification problem is tested for unsteady case of flow and transport with several disposal periods in the aquifer. The model results show that in the steady state case problem with perturbed data, the pipe leak is identified within 1% of the true flux value when the maximum error tolerance is set to 15%. In the transient case, a large number of measurements of concentration data spread over time and space are necessary to satisfactorily identify possible unknown sources of groundwater pollution. The injection rate values were found with a relative error of 13 to 32% and a normalized error of 17% when the disposal locations were known. The results remain unchanged after entering a dummy source to the unknown potential sources. Ignoring a measurement well location changes the identification results to a normalized error value of 21%. The linear programming model can be very useful in source identification with available measurement and aquifer parameters.

*Keywords:* ground-water pollution, random error, pollution sources, source identification, steady state, transient state, measurement well

### 1. Introduction

About 50% of the population of the United States depends on groundwater as the sole source of drinking water (Konieczki and Heilman, 2004). Since contamination of groundwater is detrimental to human health, the ecosystem, and the environment, the law requires EPA to identify contaminated sites and long-closed disposal sites. The sites that are deemed to represent the greatest potential threat are placed on a national priority list referred to as "Superfund" sites. After detection of the contamination in a groundwater system, the first step towards the effective remediation is the accurate identification of pollutant sources. Once source locations and release histories at a site are identified, the extent of the contaminant plumes in the subsurface can be analyzed, and the remediation system designed. A risk assessment study, which is directly related to this information, may then be completed. Therefore, the contaminant source identification problem is an important element of groundwater contaminant isolation, cleanup, and health risk assessment studies.

Source identification of groundwater pollution is an ill-posed inverse problem because of the stochastic nature of groundwater fate and transport. A problem can be categorized as a well-posed problem if it satisfies the following criteria: the solution exists, the solution is unique, and the solution is stable. Problems that do not satisfy these criteria are called ill-posed. Groundwater contaminant transport is a dispersive and therefore irreversible process thus modeling it with reverse time is an ill-posed problem (Atmadja and Bagtzoglou, 2001; Gaganis and Smith, 2006).

Numerous researchers have conducted studies focusing the groundwater source identification using optimization techniques or other mathematical methods. In the study by Gorelick et al. (1983) the problem of identifying the sources of groundwater contamination has been formulated as an optimization model. The optimization model uses a simulation model of groundwater solute transport incorporated with observation data as a series of constraints. Datta and Peralta (1986) introduced an 'expert system' to identify the location and magnitude of finite number of groundwater pollutant sources.

Similarly, Aral and Guan (1998) and Aral et al. (2001) formulated the source identification problem for groundwater contamination as a non-linear optimization model where location and release histories of contamination are unknown. The model objective is to minimize the residuals between simulated

---

\* Corresponding author. Tel.: +1 336 3347737; fax: +1 336 3347126.  
E-mail address: chang@ncat.edu (S. Y. Chang).

and measured contamination in observation sites similar to Gorelick et al. (1983). Aral and Guan (1998) proposed a genetic algorithm incorporated with simulation model as an approach to solve this non-linear optimization model. Guan and Aral (1999) conduct a complete discussion of this approach which they named as the Progressive Genetic Algorithm.

Additionally, Atmadja and Bagtzolgou (2001) conducted a state-of-art review of different mathematical methods that have been introduced during past years in recovering the time-release histories and identifying the contaminant source locations. They classified and summarized different methods such as heat transport inversion methods and contaminant transport inversion methods. These authors also introduced the related references and the limitations of each inversion method. Furthermore, Mahar and Datta (1997) proposed a non-linear optimization model incorporating the monitoring network design model with the identification. Another study by Mahar and Datta (2001) incorporates the parameter estimation with source identification of groundwater; the algorithm is an optimization-based methodology with non-linear programming techniques.

Likewise, Sun and Yeh (1990a and 1990b), Wilson and Liu (1996), Neupauer and Wilson (2001), Snodgrass and Kitaniadis (1997), and Skaggs and Kabala (1994) used mathematical inverse methods to identify the contamination sources. Neupauer et al. (2000) compared the relative effectiveness of the methods from studies of Skaggs and Kabala (1994) and Woodbury and Urych (1996). Sidauruk et al. (1998) presented an inverse method based on analytical solutions of contaminant transport problems. These authors use the logarithm of concentration to yield a linear relation between concentration and a combination of groundwater parameters. Skaliar and Ramirez (1998) proposed single point wise source identification method for general convection-diffusion transport process, involving distributed parameter systems with intricate geometry and uniformly distributed model parameters.

In this study, a linear optimization model is used to identify the pollution sources. Illustrations of the model are presented for the steady state and unsteady state cases of flow and transport. In the steady state case, the source identification model is used to identify the location of an instant leak of a conservative solute. The problem objective for the steady state case is to identify the true leak sources amongst the group of possible leak sources. The possible leak locations can be found by means of investigations such as maps, interviews and etc. The model is used to find the sources from simulation results. These results represent the closest match with the local groundwater concentration data. The concentration results at measurement well locations, due to a unit injection at each source, are arranged as the linear transform unit-concentration data. The linear transform unit-concentration data represents the reaction of the aquifer water quality at wells to the source activities. The steady state case identification model's sensitivity to the measurement data errors and necessary improvements is revised during this study. In the transient case the disposal frequency must be used as the input for the model. An abundant number of data needs to be obtained for transient case to locate the true sources amongst the possible sources.

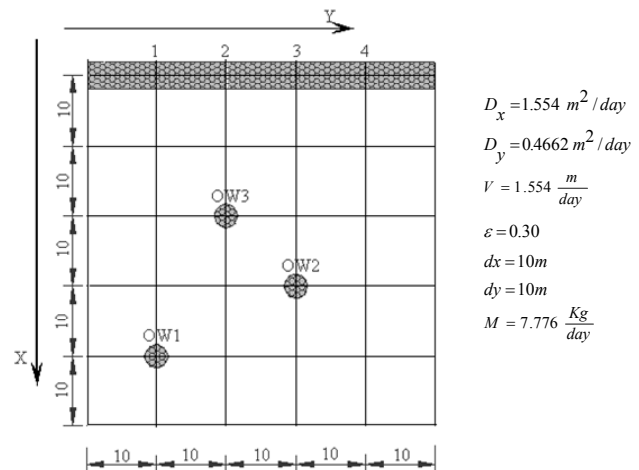
The model result is evaluated to illustrate different situations such as: entering perturbed concentration data, missing observations, and omitting one observation well.

## 2. Methodology

Two illustrative examples are used to demonstrate the use of the linear optimization and numerical simulation to a source identification problem. In the first example problem a steady state case of solute transport is demonstrated to evaluate the model performance in locating unknown sources of the pollution. In this case, concentration data is collected at a few spatially distributed locations. In the second example, identification of the source schedule is required for an unsteady state case of flow and transport in a two dimensional system in which several sources are responsible for the observed pollution.

### 2.1. Steady State Case

The steady state condition is when the solute concentration at any point in an aquifer remains constant with change of time. The source identification problem is presented for a confined aquifer containing a pipe system which is located above the water table in the aquifer. Necessary information about the pipe location and the aquifer parameters are presented in Figure 1.



**Notes:** Suspicious leak locations are numbered as 1, 2, 3, and 4. Actual leak is from location 2; OW1, OW2, and OW3 are the observation locations. Grid cells are 10 m by 10 m.

**Figure 1.** The hypothetical system, tested for steady-state source identification problem.

It is assumed that the groundwater flow pattern remains steady. The pipe has a leak with flux magnitude of 7.776 kg/day which is left undetected to the point that reaches the steady state conditions in the aquifer. Also the groundwater flows along the X axis and its velocity is 1.554 m/day. Three observation wells located downstream of the pipe are measure the pollutant concentration. It is also assumed that the pipe continues to leak until the concentration of the conservative solute

in the aquifer reached the steady state conditions and thus the observed concentration data at each site would not change with respect to time.

### 2.1.1. Steady State Simulation Model

The simulation model for the steady state problem is a two-dimensional steady-state diffusion dispersion analytical model in which pollutant M is injected at the origin instantaneously (Hunt, 1978):

$$C_c(x, y, \infty) = \frac{M \cdot e^{-\frac{xy}{2D_x D_y}}}{2\pi \varepsilon \sqrt{D_x D_y}} K_0 \left( \frac{VR}{2D_x} \right) \quad (1)$$

where

$$R = \sqrt{x^2 + y^2} D_x / D_y \quad (2)$$

$C_c$  is steady state concentration at coordinates  $x$  and  $y$  from injection origin;  $D_x$ ,  $D_y$  is constant principal values of dispersion tensors;  $\varepsilon$  is porosity of the aquifer;  $V$  is steady state velocity of water in aquifer along  $x$ ;  $D_x$  is unit cell dimension along  $x$ -axis;  $D_y$  is unit cell dimension along  $y$ -axis;  $M$  is Pollutant mass that is injected at the origin instantaneously;  $K_0$  is zero-order, modified Bessel function of the second kind.

### 2.1.2. Perturbation of Simulated Concentration

The field observations often contain errors that need to be considered in examining the proposed model performance. A computer model is used to generate a new set of data with random errors that also yields to a sensible reproduction of the observed contamination plume.

The observed concentrations are obtained from the simulation data perturbed with a random error. The perturbation of the simulation data is to represent the effect of measurement errors. In this study, the perturbed simulation concentrations  $\{C_{pert}\}$  is being used as a replacement for the concentration values at observation well locations. Computation of  $\{C_{pert}\}$  is as follows:

$$\{C_{pert}\} = \{C_s\} + \{\varepsilon r\} \quad (3)$$

where  $\{C_s\}$  represents the simulation concentration column matrix including the simulation data resulted at measurement well location, and  $\varepsilon r$  represents the random error column matrix where the number of rows is equal to the number of simulation concentrations. The random error term can be defined as:

$$\varepsilon r_i = Randn(a \times C_{s,i}) + \mu \quad (4)$$

where the random variable  $\varepsilon r_i$  is assumed to follow a nor-

mal distribution with mean value equal to  $\mu$  and standard deviation equal to  $a \times C_{s,i}$  where  $0 \leq a \leq 1.0$ .

The *Randn* function generates an array of random numbers whose elements are normally distributed with mean 0, and standard deviation equal to 1.

Such a scheme of randomly perturbing the simulated concentrations may be considered comparable to collecting and then testing multiple samples of contaminated groundwater at each observation well location and at each time period. A fraction value of 0.10 is used for  $(a)$  in the calculations. To carry out the computations, a MATLAB routine is used. Based on information regarding different measurement cases, other appropriate error terms can also be incorporated.

### 2.1.3. Steady State Case Model I

The source identification objective is to minimize the sum of absolute values of normalized residuals at all the observation sites. The normalized residual is defined as the difference between the simulated and observed concentration divided by the observed concentration. The objective function will be as follows:

$$Minimize \sum_{i=1}^m |r_{ni}| \quad (5)$$

where  $m$  is the number of collected observation data.

Since the absolute value is unrestricted in sign and can take a positive or negative value, for linear programming the value of normalized residuals are replaced with two positive values  $u_i$  and  $v_i$ . Variable  $r_{ni}$  will be replaced with following expression in the constraint set:

$$r_{ni} = u_i - v_i \quad (6)$$

where  $u_i$  is positive value representing the positive normalized residuals at measurement location  $i$ ;  $v_i$  is positive value representing the negative normalized residuals at measurement location  $i$ .

By using the variables  $u_i$  and  $v_i$  in the objective function as  $u_i + v_i$  it will be enforced that if  $r_{ni} > 0$  then the value of  $u_i$  has to be positive and  $v_i$  equal to zero and vice versa. Therefore the formulation will be as follows:

$$Min z = \sum_{i=1}^m (u_i + v_i) \quad (7)$$

Subject to:

$$\sum_{j=1}^n T_{ij} q_j - u_i + v_i = 1 \quad \forall i \in [1, m] \quad (8)$$

$$q_i \geq 0 \quad \forall i \in [1, m] \quad (9)$$

$$u_i \geq 0 \quad \forall i \in [1, m] \quad (10)$$

$$v_i \geq 0 \quad \forall i \in [1, m] \quad (11)$$

where  $T_{ij}$  is the linear transform unit-concentration data representing the simulated concentration at measurement site  $i$  from possible leak location  $j$  due to a unit leak flux (for example 1 kg/day) and at the end each is divided by observed concentration value related to the same measurement site;  $m$  is the number of observation sites and  $n$  is the number of possible sources (number of unknowns);  $q_i$  is the source flux magnitude which is entered to possible source location  $i$ . This is the product of the solute concentration of the source and volumetric disposal rate.

The simulated concentration minus the related observed concentration data is equal to the residual. This is defined by Equation 8 in the first constraint set. The equation is divided by the observed concentration to satisfy the normalized residual definition. Constraints which are defined by Equation 9 through Equation 11 enforce the non-negativity of the variables. To examine the capability of the optimization model in solving the source identification problem, the measurement data is initially produced from simulation run assuming there is no observation error.

#### 2.1.4. Steady State Case Model II

In the steady state case, a limited number of observation data is provided considering each well observes a constant value of concentration data. In the real situation, the measurement data contains errors. The model introduced by Equations 7 to 11 is sensitive to the data errors since the model is depending on such small amounts of measurement data. Therefore, improvements need to be made to the identification model in order to overcome the vulnerability of the identification model to such errors. By restricting the number of unknowns, the model can be less sensitive to the observation errors. The approach that is suggested by Gorelick et al. (1983) is to insert additional constraints into the model that limit the tolerance of the identified source flux to specified bracketed values. Furthermore, the objective function has to be improved to detect the smallest number of sources which explicate the measured concentrations within the specified tolerance. The objective function after the modification is the sum of two functions called primary objective and secondary objective. The primary objective is a binary decision variable which decides whether the source enters the model or not. Minimizing the absolute normalized residuals enters as the secondary objective. To apply both the primary and secondary objective with the related degree of importance, a large weighting factor multiplier is used for the primary objective (value of 10) and a small weighting factor (value of 1) is used for the secondary objective. The modified model becomes:

$$Min z = M \sum_{i=1}^n p_i + S \sum_{i=1}^m (u_i + v_i) \quad (12)$$

Subject to:

$$\sum_{j=1}^n T_{ij} q_j - u_i + v_i = 1 \quad \forall i \in [1, m] \quad (13)$$

$$q_i \leq M p_i \quad \forall i \in (1, 2, \dots, n) \quad (14)$$

$$u_i \leq u^* \quad \forall i \in (1, 2, \dots, m) \quad (15)$$

$$v_i \leq v^* \quad \forall i \in (1, 2, \dots, m) \quad (16)$$

$$p_i \in [0, 1] \quad \forall i \in (1, 2, \dots, n) \quad (17)$$

$$q_i \geq 0 \quad \forall i \in (1, 2, \dots, m) \quad (18)$$

$$u_i \geq 0 \quad \forall i \in (1, 2, \dots, m) \quad (19)$$

$$v_i \geq 0 \quad \forall i \in (1, 2, \dots, m) \quad (20)$$

where  $M$  is a large multiplier scalar coupled with the primary objective, a value of 10.0 is used in the example problem;  $S$  is a small multiplier scalar coupled with the secondary objective, a value of 1.0 is used in the example problem;  $p_i$  is a binary decision variable,  $p_i = 1$  if the potential source ( $i$ ) enters to the solution,  $p_i = 0$  otherwise, for  $n$  potential sources;  $u^*$  is error tolerance related to positive residuals;  $v^*$  is error tolerance related to negative residuals.

The above improved optimization model is a mixed integer program with the format of a fixed charge problem. The primary objective minimizes the number of sources and the secondary objective minimizes the sum of absolute normalized residuals. The first line in the constraint set defined by Equation 13 is the same as in Equation 8. The constraint in Equation 14 enforces that for each leak that enters to the basis, a large penalty will be charged to the objective function value. In Equation 15 and Equation 16 the residuals are enforced by the constraints to be bracketed within the specified tolerances.

## 2.2. Unsteady State Case

The source identification model is also applicable in the case of transient flow and transport in subsurface media. The optimization simulation model can be employed to identify the contaminant sources and release histories. The sources and release history are identified from observed concentration data collected over time. In the transient case the contaminant input may occur in various space and time. The identification model has to identify the time period and location of input for each disposal event. In comparing the steady state case, the model for the unsteady state case would not be as sensitive to the measurement data errors because the measurements are collected over time in various locations. In the unsteady state case, the potential source locations and their time periods which allow pollutants to enter the groundwater must be recogni-

zed. Each disposal event may have numerous consequent measurement data. Therefore the number of constraints is much greater than number of unknowns. The optimization model for the unsteady state will be as follows:

$$\text{Min } z = \sum_{i=1}^m (u_i + v_i) \quad (21)$$

Subject to:

$$\sum_{j=1}^n T_{ij} q_j - u_i + v_i = C_{obs,i} \quad \forall i \in [1, m] \quad (22)$$

$$q_i \geq 0 \quad \forall i \in (1, 2, \dots, m) \quad (23)$$

$$u_i \geq 0 \quad \forall i \in (1, 2, \dots, m) \quad (24)$$

$$v_i \geq 0 \quad \forall i \in (1, 2, \dots, m) \quad (25)$$

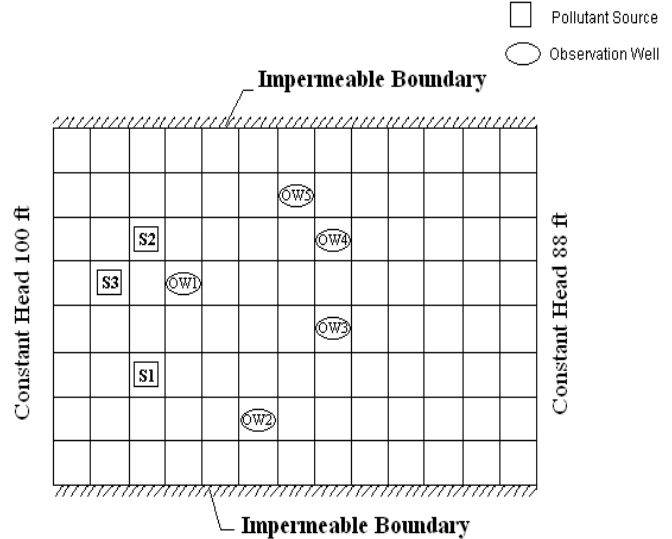
where  $C_{obs,i}$  is the observed concentration data at each time step and each measurement location which in the case study is obtained from perturbation of the simulation data;  $T_{ij}$  is the linear transform unit-concentration data representing the simulated concentration at measurement site  $i$  from possible leak location  $j$  due to a unit leak flux;  $m$  is total number of collected measurements equal to the number of measurement locations times total number of time steps in timeframe of the study.

The model is similar to the steady state case source identification model; however, the abundance of collected observation data will allow reducing the computation burden by using the ordinary residuals instead of normalized residuals. Therefore, in constraint set defined by Equation 22, the observation data is brought to the right hand side of the constraint and the definition of the  $T_{ij}$  is modified to suit the use of ordinary residuals in the model. The source identification objective however is to minimize the sum of absolute value of ordinary residuals at all the observation sites. The model employs linear superposition to find the waste disposal rates at disposal sites from concentration data at measurement wells. The disposal rate values are obtained by superposing the linear influences of injection rates into concentrations observed at measurement well locations.

This study applies the model to the case of transient flow, incorporated with the transient transport in the aquifer. Also, it uses the ordinary residuals in the objective function and constraints and perturbed concentration data is entered to the right side of the constraints. It is apparent that incorporating the abundant data (1000 measurement data) in this transient problem is effectively a replacement for the use of any integer constraints to overcome the data errors in the problem.

The model is applied to a hypothetical aquifer system with transient flow and transient transport of a conservative pollutant in an aquifer, shown in Figure 2. The aquifer system con-

tains two potential pollutant sources (sources S1 and S2, the source S3 enters as a dummy source for further analysis); there are also five measurement sites (OW1 ~ OW5) located in this aquifer. The hydraulic head varies linearly along both the constant head boundaries. The parameters of the example aquifer are represented in Table 1. The total study time of five years is divided into twenty equal time periods of three months each. Each injection period is three months, and for the first four time periods, the pollutant injection occurs in S1 and S2, the waste disposal rates at source wells are represented in Table 2.



**Notes:** The potential sources are marked S1, S2, and S3 where S1 and S2 are actual pollutant sources and S3 used as the dummy source. Five observation wells are placed at downstream of the potential sources; each grid cell has dimensions of 100 ft by 100 ft.

**Figure 2.** Schematic view of an aquifer system for unsteady state case study (customized from Mahar and Datta, 2000).

**Table 1.** Hydro-geometric Parameters of the Aquifer as Shown in Figure 2

Parameter	Value
$K_{xx}$ (ft/s)	0.002
$K_{yy}$ (ft/s)	0.002
Porosity ( $\mathcal{E}$ )	0.25
$\alpha_L$ (ft/s)	131.2
$\alpha_t / \alpha_L$	0.24
$D_L$ (distribution coefficient)	0.005
Bulk Density $P_b$	2.65
$\Delta X = \Delta Y$ (ft)	100
S (storage coefficient)	0.002

### 2.2.1. Two-dimensional Flow and Transport Equations

In the transient case study, the equation has been used to represent the two-dimensional flow and advection-dispersion solute transport through a non-homogenous anisotropic aquifer can be written in Cartesian tensor notation as (Konikow and Bredhoeft, 1978):

$$\frac{\partial}{\partial x_i} \left( T_{i,j} \frac{\partial h}{\partial x_j} \right) = S \frac{\partial h}{\partial t} + W \quad i, j = 1, 2 \quad (26)$$

where  $T_{i,j}$  is transmissivity tensor ( $L^2T^{-1}$ );  $h$  is the hydraulic head ( $L$ );  $W$  is volume flux per unit area (negative for inflow and positive sign for outflow,  $LT^{-1}$ );  $x_i, x_j$  is Cartesian coordinates ( $L$ ).

**Table 2.** Waste Disposal Rates and Periods for the Transient Case Study

1 <sup>st</sup> period of 3months of duration	Flow	Concentration
S1	0.166 cfs	10000 ppm
S2	0.106 cfs	10000 ppm
2 <sup>nd</sup> period of 3 months		
S1	0.053 cfs	10000 ppm
S2	0.208 cfs	10000 ppm
3 <sup>rd</sup> period of 3 month		
S1	0.166 cfs	10000 ppm
4 <sup>th</sup> period of 3 months		
S2	0.124 cfs	10000 ppm

If we only consider fluxes of direct withdrawal or recharge and steady leakage through a confining layer, streambed or lakebed, then  $W(x, y, t)$  can be expressed as:

$$W(x, y, t) = \underbrace{Q(x, y, t)}_{\text{Pumping}} + \underbrace{\frac{K_z}{m}(H_s - h)}_{\text{Leakage}} \quad (27)$$

where  $Q$  represents the rate of withdrawal (+) or recharge ( $-$ ), ( $LT^{-1}$ );  $K_z$  represents the vertical hydraulic conductivity of the confining layer, streambed or lakebed, ( $LT^{-1}$ ); and  $H_s$  represents the hydraulic head in the source bed, stream or lake ( $L$ ).

The seepage velocity can be written in Cartesian tensor notation as:

$$V_i = -\frac{K_{ij}}{\varepsilon} \frac{\partial h}{\partial x_j} \quad (i, j = 1, 2) \quad (28)$$

where  $V_i$  is the seepage velocity in the direction of  $x_i$  ( $LT^{-1}$ );  $K_{ij}$  is the hydraulic conductivity tensor ( $LT^{-1}$ );  $\varepsilon$  is the effective porosity of aquifer (dimensionless).

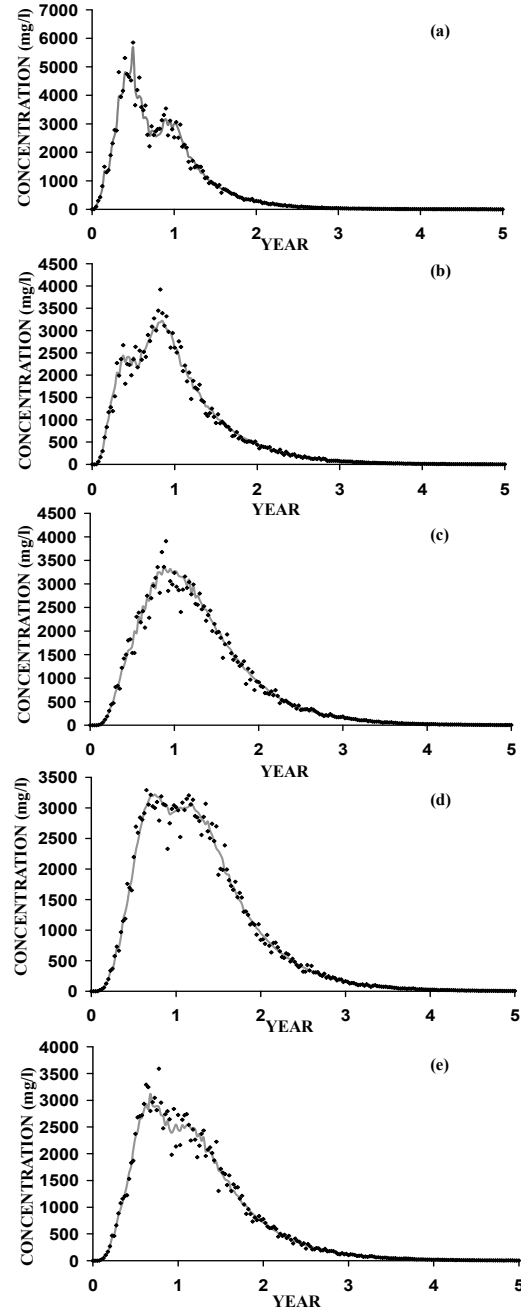
The equation used for two-dimensional transport and dispersion of a solute in groundwater flow may be written as:

$$\frac{\partial(Cb)}{\partial t} = \frac{\partial}{\partial x_i} \left( bD_{ij} \frac{\partial C}{\partial x_j} \right) - \frac{\partial}{\partial x_i} (bCV_i) - \frac{C'W}{\varepsilon} \quad (i, j = 1, 2) \quad (29)$$

where  $C$  is concentration of the dissolved chemical, ( $ML^{-3}$ );  $D_{ij}$  is hydrodynamic dispersion coefficient, ( $L^2T^{-1}$ );  $b$  is saturated thickness of the aquifer, ( $L$ ); and  $C'$  is concentra-

tion of the dissolved chemical in a source or sink fluid, ( $ML^{-3}$ ).

The transport simulation results are obtained based on the above equations, using USGS-MOC (USGS method of characteristics) computer code. Detailed derivations and discussions on the governing equations for flow and transport in groundwater systems are available in Konikow and Bredehoeft (1978).



**Notes:** The continuous line shows the simulation results at measurement wells, the perturbation results are shown as dots.

**Figure 3.** Pollutant concentration data for the transient problem in form of breakthrough curves (a), (b), (c), (d), and (e) for measurement wells 1, 2, 3, 4, and 5.

### 2.2.2. Measurement Simulation and Perturbation

To apply the source identification model to the example aquifer, a history of the observed concentrations at measurement sites is necessary. The concentration history data at each location is obtained from simulation results at that location. Concentration simulations are obtained by the USGS-Method of Characteristics (USGS-MOC) computer code (Konikow and Bredhoeft, 1978). The simulated concentration results are obtained in 10 time-steps during each time period at all well locations. The output data at each location is provided from the beginning of the first injection up to five years. The simulation results are provided in 10 time-steps for each of twenty time periods of three months. This means the total of 200 simulated concentration data for each observation site is provided. The simulated data is then perturbed to imitate the effects of measurement errors. Each perturbed datum was assumed to be sampled from a normal distribution and having the exact datum with a standard deviation equal to one-tenth of the magnitude of datum. Perturbation method is explained previously in the steady state case section.

A MATLAB subroutine has been employed to add the perturbation to the simulated data. The perturbed concentration data will be used as the measurement data in the identification problem. The simulation breakthrough curves along with scattered diagram of the perturbation results are presented in Figures 3a to 3e. The reason for inserting such concentration data is to create a set of realistic data for the identification problem inputs. Reviewing the breakthrough curves can give basic information about the sources' location but this information is not enough to gain complete knowledge about the release schedule. For instance in Figure 3a the pollutant reached the well No.1 almost immediately after entering the aquifer. The first plume peak reaches the well No.1 at 0.5 year and the last plume peak passes the measurement well at 1 year. Comparing to the well No.1, the well No.2 plume peak is extended longer with lower peak concentration value. This is because the well is not located immediately after sources. The perturbed data also contains random errors as high as 20% in peak points.

### 2.2.3. Evaluation Criterion for the Model Performance

In order to evaluate the performance of the model for the transient state a factor needs to be defined as the evaluation criterion. In this study, accuracy of the proposed source identification model is measured in terms of "normalized error" (NE) (Mahar and Datta, 2000). This factor is defined as:

$$NE = \frac{\sum_{i=1}^{in} \sum_{j=1}^{ns} |EF_{ij} - AF_{ij}|}{\sum_{i=1}^{in} \sum_{j=1}^{ns} AF_{ij}} \times 100 \quad (30)$$

where *in* is total number of potential time periods of source activity at potential source locations; *ns* is total number of potential source locations; *EF<sub>ij</sub>* is estimated source flux at source location *j*, during time period *i*; *AF<sub>ij</sub>* is actual source flux at

potential source location *j*, during time period *i*.

## 3. Results and Discussion

### 3.1. Steady State Case

#### 3.1.1. Results for Steady State Case Model I

The optimization model for the steady state case, Model I, is examined for the aquifer which is described in the Figure 1. The model for the steady state case is tested with the simulation data to assess the performance of the model. The simulation data is obtained from analytical solution that was previously discussed in Equations 1 and 2. After entering the aquifer parameters from Figure 1 into Equations 1 and 2, the simulation result at measurement wells OW1 to OW3 caused by a leak at location 2 on the pipe with magnitude of 90 mg/s (7.776 kg/day) is obtained and presented in Table 3.

**Table 3.** Simulation Results due to a Leak Magnitude of 7.776 kg/day at Location No. 2 (Used as Measurement Data)

Measurement station	X Coordinate	Y Coordinate	Cc kg/m <sup>3</sup>
OW1	40	-10	0.177
OW2	30	10	0.1114
OW3	20	0	1.8981

For purposes of evaluation of the source identification model, the simulation result is used in the constraint set. The results of such a test will show the accuracy of the linear programming model in identifying the magnitude and locations of the sources. The LINDO software, which is a linear programming model solver, has been employed for this example problem and the results are presented in Table 4. Results show that the model has clearly identified the source location and the leak magnitude. Total residuals are brought to 0.003 and the source flux level is 7.752 kg/day (99.7% accurate) in the linear programming model results. Also the LINDO results show another small source appearing in the output result which is because of round off errors. Introducing the simulation results as the concentration data to the identification authenticates the identification model's good performance with the error free data.

#### 3.1.2. Effect of Measurement Error to the Results of Model I

In the field situation the measurement data is subject to errors which should be considered in the source identification model. In this part of the study, the source identification Model I is applied to the steady state case with the perturbed measurement data. The perturbation method explained previously is applied to the simulation results from Table 3. The resulted values that are presented in Table 5 have been used as the measurement data for the source identification study.

The result of identification model with perturbed concentration values is presented in Table 6. The result has identified the source location S2, and shows another leak from the po-

tential source location S3. The total leak magnitude from sources S2 and S3 that is shown in the linear programming model results differ from the true leak magnitude. The difference between two values can be explained by the objective function value which is defined as total normalized residuals. The result confirms that the model is greatly responsive to the concentration data errors. Therefore, in order to have the model results resistant to the concentration data errors, the identification Model II is used for the perturbed concentration data in the steady state case.

**Table 4.** Results of Model I for Error Free Concentrations Compared to True Values

Potential source of leak	True leak magnitude (kg/day)	Linear Program results (kg/day)
1	0	0
2	7.776	7.752
3	0	0.0026
4	0	0

**Table 5.** Measurement Data due to a Leak Magnitude of 7.776 kg/day at Location No. 2 Obtained from Simulation Using Perturbation Method

Measurement station	Cs (kg/m <sup>3</sup> )	Cpert (kg/m <sup>3</sup> )
OW1	0.177	0.153349
OW2	0.1114	0.119358
OW3	1.8981	2.206276

**Table 6.** Results of Model II in Identifying the Leak Magnitude with Observation Errors Compared to True Values

Potential source of leak	True leak magnitude (kg/day)	Linear Program results (kg/day)
1	0	0
2	7.776	6.725215
3	0	0.1158773
4	0	0

### 3.1.3. Results of Steady State Model II

The example problem with perturbed data is used to test the improved model which is introduced in Equations 12 to 20. Two different runs were attempted using linear programming computer code LINGO. In the first one the bracketed tolerance is brought to 0.15 for both positive and negative residuals. The results are summarized in Table 7 (column a). For the second case, the positive and negative residuals were bracketed to 0.2 (Table 7, column b). Using binary decision variables  $p_i$  in objective function and the constraint set will make it impossible for imaginary sources to come into the result. Also the value of the error tolerance in Equations 15 and 16 will affect the accuracy of the resulted source flux. As shown in Table 7 the value of source flux resulting from maximum error tolerance of 0.15, is brought within 1% of the real source flux. In the

second case with maximum tolerance of 0.2, the result is overestimated by approximately 4% of source flux magnitude (Table 7, column b). In the case that the tolerance was limited to 0.05 or 0.1 the model became infeasible. The results of the modified model show that the model could be successful if the error bracket values are appropriately restricted.

**Table 7.** Linear Programming Results (in kg/day) Using Model II; (a) Perturbed Data Residuals Bracketed to 0.15, and (b) Perturbed Data Residuals Bracketed to 0.20

Potential source of leak	True leak magnitude (kg/day)	Linear Program results (a)	Linear Program results (b)
1	0	0	0
2	7.776	7.7347	8.0710
3	0	0	0
4	0	0	0

### 3.2. Unsteady State Case

The proposed model for the transient case is tested in different settings of the model parameters. The evaluation criterion of the model described previously was used to evaluate different scenarios of the model performance. In the first scenario, the source locations assumed to be known and the model is used to identify the rates and timing of the injection. The next scenario describes the situation that one dummy source location enters as an unknown. The third scenario studies the situation that one measurement well is omitted from the group of wells. The last scenario discusses the effect of missing measurement data on the source identification results.

#### 3.2.1. Evaluation of Source Flux Identification Model at Known Locations

In the transient case study the duration of each injection period and potential source locations are assumed to be known. Therefore the model is employed to identify the magnitude of the injection rate at each time period and source location.

The optimization model for the transient problem is solved using LINGO software. The result of the linear programming for the proposed example problem is presented in Table 8 (column a). The linear programming results in Table 8 (column a) properly identify the disposal events. The model results also determine the time and location of the injection identical to the true sources. For the sources with no pollutant injection the LP result gave a zero value of injection rate at the related time period. The results also indicate that using the additional integer constraints similar to that of steady state model II will not be necessary. Incorporation of numerous concentration data in the transient case is a good replacement for the integer constraints that were necessary for the perturbed data conditions in the steady state case. The injection rate values for each injection event from sources S1 and S2 are located within the range of 6 to 32% of the true injection rates. The normalized error value is 17.25% when the true sources S1 and S2 are entered as the only potential sources for the sour-



**Table 8.** LP Results of Source Fluxes Using Concentration Data with Random Error for (a) only Known Potential Sources S1 & S2 (b) Potential Sources S1, S2, and S3

Injection duration	source location	Actual injection rates	LP results w/only potential sources S1 & S2 (a)	Relative error (a) (%)	LP results including S3 (b)	Relative error (b) (%)
0 to 0.25 yr	S1	0.166	0.176	6	0.176	6
	S2	0.106	0.129	21	0.129	21
	S3	0	-	0	0	0
0.25 to 0.5 yr	S1	0.053	0.070	32	0.070	32
	S2	0.208	0.181	12.9	0.181	12.9
	S3	0	-	0	0	0
0.5 to 0.75 yr	S1	0.166	0.132	20	0.132	20
	S2	0	0	0	0	0
	S3	0	-	0	0	0
0.75 to 1 yr	S1	0	0	0	0	0
	S2	0.124	0.093	25	0.093	25
	S3	0	-	0	0	0
Normalized error				NE = 17.25%	NE = 17.25%	

**Table 9.** LP Results of source Fluxes at Known Locations Using Concentration Measurements with Random Error for (a) Assuming Measurement Well No.1 is not existent, (b) Assuming Lost Data

Injection duration	source location	Actual injection rates	LP results case (a)	Relative error: case (a) (%)	LP results case (b)	Relative error: case (b) (%)
0 to 0.25 yr	S1	0.166	0.168	1.2	0.004	97.6
	S2	0.106	0.143	34.9	0.207	95.2
	S3	0	0	0	0	0
0.25 to 0.5 yr	S1	0.053	0.077	45.3	0.205	287
	S2	0.208	0.170	18.3	0.129	38
	S3	0	0	0	0	0
0.5 to 0.75 yr	S1	0.166	0.128	22.9	0.109	34.3
	S2	0	0	0	0	0
	S3	0	0	0	0	0
0.75 to 1 yr	S1	0	0	0	0	0
	S2	0.124	0.103	16.9	0.096	22.6
	S3	0	0	0	0	0
Normalized error				NE = 21%	NE = 70.4%	

ce identification problem.

### 3.2.2. Performance Evaluation of Model for Multiple Potential Source Locations

The number of potential sources can be reduced if certain knowledge about location of sources is available. Information about the potential source locations is available through maps, interviews, previous records and other resources. In this case study, it is assumed that the numbers of potential source location are not well-known therefore one potential dummy source is added to the problem. Such a problem may be encountered when there are more locations that suspicious to the pollutant injections. Two sources of S1 and S2 are actual source locations with time varying magnitudes. Therefore it is assumed that a third potential source S3 at location (200 ft, 500 ft) is

entered to the case study. A new simulation model result is obtained for the unit injection rate at the dummy source location. The result of identification model for this case was obtained using LINGO software. The resulted values for the case without additional potential source and the case using additional potential source are presented in Table 8 (column b) for comparison.

Results show that adding an additional potential source will not seriously affect the final identification results. The result shows zero value of injection rate for the additional source location and the injection rate at actual source locations remains the same. Normalized error of the resulted injection rates for the case of entering one additional potential source location is 17.25%; this is exactly the same value as the case of using two actual sources.

### 3.2.3. Effect of Number Measurement Well Locations to the Source Identification Results

Identification of the pollutant timing and location depends on the collected measurement data through various time and places. Omitting one observation well location can impact the source identification accuracy. In order to investigate this matter, the observation well location No. 1 is disregarded. The model is used to identify the potential sources S1, S2, S3 using the measurement results obtained from observation well location OW2 through OW5. The results are presented in Table 9 (column a).

The value of normalized error of 21% states that the model identifies the sources with less accuracy than previous conditions. Well No. 1 is located immediately down-stream of potential sources; therefore, this change can have more impact on the accuracy of the identification results than disregarding any of the other four wells. Model results give a zero value for the sources with no injection rate and locate the injection events comparable to the true injections. The results also declare that disregarding well No. 1 does not have a serious impact on the identification results. This is probably because the other four measurement wells are located spatially in the face of groundwater flow direction in such an arrangement that they can collect concentration data effectively even when they are not immediately located downstream of potential sources. This will confirm that formation of the measurement wells in an aquifer can affect the source identification results. A general rule is to arrange the measurement well locations relative to the expected location of the sources such that the observed concentration data differs from one another and contain more information about source characteristics. More discussion about arrangement of measurement well locations can be found in a study conducted by Mahar and Datta (1996) which incorporates optimal monitoring network design methodology, with detection of groundwater contamination at the waste disposal facilities.

### 3.2.4. Effect of Missing Measurement Data to the Source Identification Results

In a real case, observed data may not be available from the beginning of the pollutant release to the aquifer. More frequent reading of concentration data may start after the contamination is detected. It is assumed that the measurement data does not exist for the first two time periods of the study. The result of the linear programming model for the case of missing data is presented in Table 9 (column b). Results indicate that the linear programming solution correctly located all the true source locations. The identification model calculates a small value for the source 1 during time period 1. Such results indicate that the interval of missing data includes those months when the effects of disposal at source 1 during period 1 are most prominent. The results also show a higher relative error in identifying the injection rate of the remaining sources. The value of the normalized error (NE = 70.4%) confirms that the accuracy of the identification model has been affected seriously under the missing data conditions.

## 4. Conclusions

The performance of a linear programming (LP) model for the groundwater source identification problem is evaluated for two illustrative examples of steady state and transient case. The LP model for the steady state example is used to identify the source locations along a pipe that leaks high concentrations of a conservative pollutant to an aquifer. Steady state Model I identified the source locations and gave correct flux level with 99.7 % accuracy when simulated concentration data were used. However, the same model fails to identify the correct sources and flux levels when perturbed concentration data is used. The steady state Model II is used when perturbed concentration data are used in the identification model. The results of steady state Model II with perturbed data identified the source location and recovered the flux level with 99% accuracy when the residuals were bracketed to 15%.

The LP model for the second illustrative example is used for condition of transient flow and transport. In this case, the pollutant was observed from different locations in an aquifer which is believed to be from waste disposal at several facilities. The purpose of identification model is to locate those disposal sites and the time period when the pollutant is released to the aquifer. The model results were analyzed for the various scenarios. For the first scenario that only the real sources are used as possible sources, the model results identify the disposal events and recover the disposal rates with a normalized error of 17.3%. Furthermore, adding another possible source location as a dummy source for the second scenario had results similar to the first scenario. Additionally, disregarding concentration data associated with measurement well No. 1 for the third scenario gave a normalized error of 21% of recovered disposal rates while the disposal events were well identified. Finally, the fourth scenario that the concentration data of the first two time periods are disregarded, the results were found with a normalized error of 70%. Incorporating abundant concentration data in the transient case is a replacement for the use of integer constraints that were necessary for the perturbed data conditions in the steady state case. The model can be used if the pollutants are continuously disposed to the aquifer in identified time frequencies.

The LP model presented in this paper can be useful in identifying pollutant sources for cleanup or remediation of contaminated aquifers. The perturbed data errors incorporated with the transient case in this study do not have a serious effect on identifying the source locations. However, these concentration data errors reduce the accuracy of the recovered release history. These perturbed data can have an impact on the linear effect unit-concentration response. For that reason the unit concentration response in conditions of the transient flow may not exactly describe the influence of the pollutant sources upon concentrations at observation wells. Generally, the observation site locations and their distance from source locations may affect the release histories recovered by the model as a result of dispersion phenomenon. In this study the aquifer hydrogeometric parameters are assumed to have no uncertainties but in a field situation aquifer parameters cannot be so well

defined without uncertainties.

**Acknowledgments.** This work was sponsored by the Department of Energy's Samuel Massie Chair of Excellence Program under grant number DE-FG01-94EW11425. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the funding agencies.

## References

- Aral, M.M., and Guan, J. (1998). *Identification of Groundwater Contaminant Sources and Release Histories using Genetic Algorithms*, Multimedia Environmental Simulations Laboratory, Technical Report MESL-01-98, Georgia Institute of Technology, Atlanta.
- Aral, M.M., Guan, J. and Maslia, M.L. (2001). Identification of contaminant source location and release history in aquifers, *J. Hydrol. Eng.*, 6(3), 225-234, doi:10.1061/(ASCE)1084-0699(2001)6:3(225).
- Atmadja, J., and Batzoglu, A.C. (2001). State of the Art Report on Mathematical Methods for Groundwater Pollution Source Identification, *Environ. Forensics*, 2, 205-214.
- Bassett, R.L., Buszka, P.M., Davidson, G.R., and Diaz, D.C. (1995). Identification of Groundwater Solute Sources Using Boron Isotopic Composition, *Environ. Sci. Technol.*, 29(12), 2915-2922.
- Datta, B. and Peralta, R.C. (1986). Expert pattern recognition for pollution source identification, Paper presented at the ASCE Water Forum '86, *World Water Issues in Evolution*, California.
- Dougherty, D.E., and Marryott, R.A. (1991). Optimal groundwater management: 1. Simulated annealing, *Water Resour. Res.*, 27(10), 2493-2508, doi:10.1029/91WR01468.
- Gaganis, P., and Smith, L. (2006). Evaluation of the uncertainty of groundwater model predictions associated with conceptual errors: A per-datum approach to model calibration, *Adv. Water Resour.*, 29(4), 503-514.
- Gorelick, S.M., and Remson, I. (1982). Optimal dynamic management of groundwater pollutant sources, *Water Resour. Res.*, 18(1), 71-76.
- Gorelick, S.M. (1982). A model for managing sources of groundwater pollution, *Water Resour. Res.*, 18(4), 773-781.
- Gorelick, S. M., Evans, B., and Remson, I. (1983). Identifying Sources of Groundwater Pollution: An Optimization Approach, *Water Resour. Res.*, 19(3), 779-790.
- Guan, J., and Aral, M.M. (1999). Progressive genetic algorithm for solution of optimization problems with nonlinear equality and inequality constraints, *Applied Mathematical Modeling*, 23, 329-343, doi:10.1016/S0307-904X(98)10082-3.
- Hunt, B. (1978). Dispersive Sources in Uniform Ground-Water Flow, *Journal of the Hydraulics Division, ASCE*, 104(1), 75-85.
- Jin, A. (1996). *An optimal estimation scheme for subsurface contaminant transport model using Kalman-Bucy filter*, Thesis, North Carolina A and T State University.
- Konieczki, A.D., and Heilman, J.A. (2004). *Water-use trends in the desert southwest-1950-2000*, Scientific Investigations Report 2004-5148, U.S. Department of the Interior, US. Geological Survey.
- LINDO SYSTEMS INC. (1999). *LINGO: the modeling language and optimizer*, Chicago, Illinois.
- Mahar, P.S., and Datta, B. (1997). Optimal monitoring network and ground-water pollution source identification, *J. Water Resour. Plann. Manage.*, 123(4), 199-207, doi:10.1061/(ASCE)0733-9496(1997)123:4(199).
- Mahar, P.S., and Datta, B. (2000). Identification of pollution sources in transient groundwater systems, *Water Resour. Manage.*, 14(3), 209-227, doi:10.1023/A:1026527901213.
- Mahar, P.S., and Datta, B. (2001). Optimal identification of groundwater pollution sources and parameter estimation, *J. Water Resour. Plann. Manage.*, 127(1), 20-29, doi:10.1061/(ASCE)0733-9496(2001)127:1(20).
- Meyer, P.D., Valocchi, A.J., and Eheart, J.W. (1994). Monitoring network design to provide initial detection of groundwater contamination, *Water Resour. Res.*, 30(9), 2647-2659.
- Neupauer, R.M., Borchers, B., and Wilson, J.L. (2000). Comparison of inverse methods for reconstructing the release history of a groundwater contamination source, *Water Resour. Res.*, 36(9), 2469-2475.
- Neupauer, R.M., and Wilson, J.L. (2001). Ad joint-derived location and travel time probabilities for a multidimensional groundwater system, *Water Resour. Res.*, 37(6), 1657-1668.
- Page, G.W. (1987). *Planning for groundwater protection*, (Orlando, FL: Academic Press Inc.).
- Skaggs, T.D., and Kabala, Z.J. (1994). Recovering the release history of a groundwater contaminant, *Water Resour. Res.*, 30(1), 71-79.
- Skliar, M. and Ramirez, W.F. (1998). Source Identification in Distributed Parameter Systems, *Applied Mathematics and Computer Science*, 8(4), 733-754.
- Snodgrass, M.F., and Kitanidis, P.K. (1997). A geostatistical approach to contaminant source identification, *Water Resour. Res.*, 33(4), 537-546.
- Sidauruk, P., Cheng, A.H.D., and Ouazar, D. (1998). Ground water contaminant source and transport parameter identification by correlation coefficient optimization, *Ground Water*, 36(2), 208-214.
- Sun, N.Z., and Yeh, W.W.G. (1990a). Coupled inverse problems in groundwater modeling: 1. Sensitivity analysis and parameter identification, *Water Resour. Res.*, 26(10), 2507-2525.
- Sun, N.Z., and Yeh, W.W.G. (1990b). Coupled inverse problems in groundwater modeling: 2. Identifiability and experimental design, *Water Resour. Res.*, 26(10), 2527-2540.
- Wilson, J.L., and Liu, J.L. (1996). Field Validation of the Backward-In-Time Advection Dispersion Theory, *Paper presented at the HSRC WERC Joint Conference on the Environment*, Albuquerque, New Mexico.
- Woodbury, A.D., and Ulrych, T.J. (1996). Minimum relative entropy and probabilistic inversion: Theory and application to recovering the release history of a groundwater contaminant, *Water Resour. Res.*, 32(9), 2671-2681.
- Zou, S., and Parr, A. (1995). Optimal Estimation of two-dimensional Contaminant Transport, *Ground Water*, 33, 319-325.