

# Locate the Pollution Source

Shen Quan  
Yang Zhenyu  
He Xiaofei  
Zhejiang University  
Hangzhou, China

Advisor: Zhang Chong

## Summary

We develop a model for a strategy to detect new pollution. Three processes govern the movements of pollutants in groundwater: advection, dispersion, and retardation. Information from the wells is used to

- determine the rate and direction of groundwater movement,
- determine the horizontal and vertical extent of the pollutants, and
- analyze the underground structure and characteristics.

Regarding the diversity and complexity of the given data, we employ a two-step data selection to determine the pollutants most likely to cause new pollution during this period of time. We refine the data to choose those chemicals that best represent the variation during this period of time. Then, by using a grid-search algorithm, we write a computer program to simulate the movement process and identify the location and start time of the pollution source. The program is written in C and runs on a PC. Four kinds of new pollution sources are located. The graph resulting from our model is in a good agreement with the given data. Finally, we test parameter sensitivity.

## Assumptions

- All soil and aquifer properties are homogeneous and isotropic throughout both the saturated zone and the unsaturated zone.
- The aquifer consists of sand and gravel.

---

*The UMAP Journal* 20 (3) (1999) 355–368. ©Copyright 1999 by COMAP, Inc. All rights reserved. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice. Abstracting with credit is permitted, but copyrights for components of this work owned by others than COMAP must be honored. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior permission from COMAP.

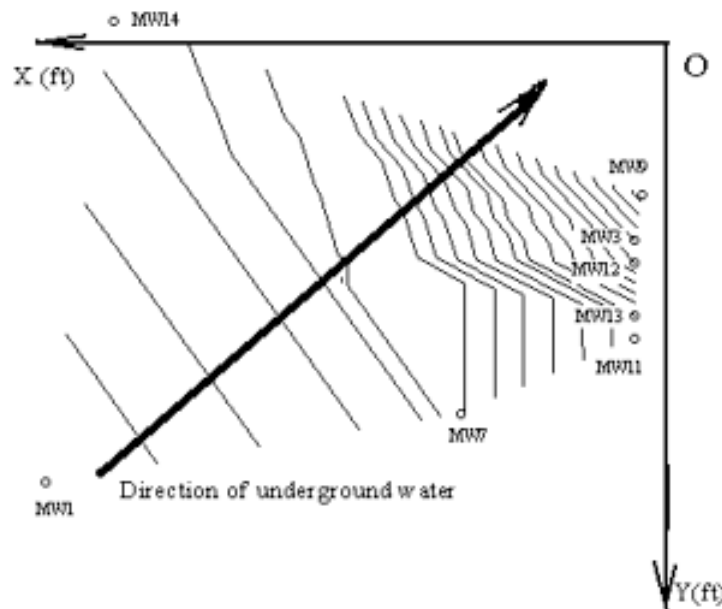
- Steady, uniform water flow occurs only in the vertical direction throughout the unsaturated zone, and only in the horizontal (longitudinal) plane in the saturated zone in the direction of groundwater velocity.
- Physical processes play the greatest role, while the chemical processes are negligible.
- All the parameters describing the characteristic of both zones are constant throughout the monitoring period.
- All the sources of the pollutants are point sources.

## Problem One

This problem is to estimate the location and start time of the source, so we consider the movement process of the pollution and the structure of the underground.

### Data Analysis and Processing

We assume that there is no interaction between pollutants so that we can process each pollutant separately. With the given data of the coordinate and water level of each well, we plot the water level map by using linear interpolation on the elevations of the monitoring wells, as in **Figure 1**. For simplicity of computation, we assume that all the underground water flows in the same direction.



**Figure 1.** Water-level map.

## Data Selection

Because we have thousands of data points about concentration of various pollutants, we must select data carefully. We do this in the following steps:

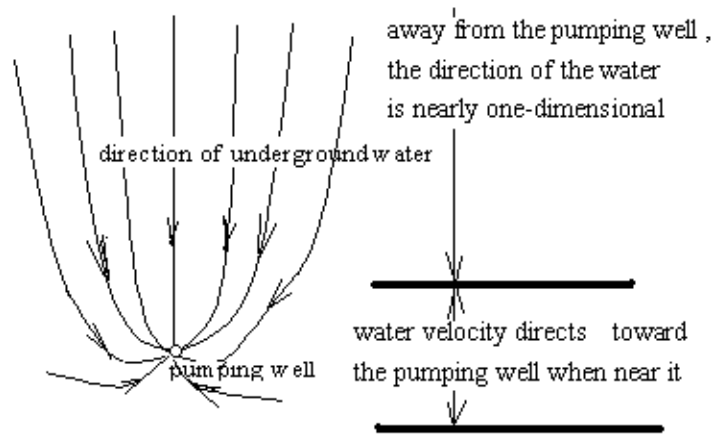
- Because pollutants are strongly influenced by layers of different permeability, measurements of critical parameters and pollutant concentrations need to be done at intervals over the depth of the aquifer. We need a method for sampling at different depths in an aquifer. By analyzing the data set, we find that almost every pollutant affects only one part of a well (top, middle, or bottom). Thus, for each pollutant we need to consider only the effect on one layer of the well. Furthermore, the data from the bottom of each well (if any) remain constant or nearly so, hence we can neglect such data.
- We delete the data for some pollutants, such as tetrachloroethane, acrolein, benzene, bromomethane, chlorobenzene, cobalt, and so on, because there are hardly any changes in concentrations of these pollutants in each well.
- We think that the pulse fluctuation about the pollutant concentration during a relatively stable period, such as for manganese, is caused by random factors. Thus, we eliminate these pollutants from the data set.
- There is a particular constituent, the CarbonTotalOrganic, whose concentration value decreases significantly, from more than 1000 to less than 1.5. Thus, we eliminate it.
- Now only four pollutants remain: calcium, chloride, magnesium, and TDS.

## Reselection

For each remaining pollutant, to accurately reflect the tendency for the concentration to change, we reselect its data as follows:

- For each well, we choose two concentration values for each year, one from the first half of the year and the other from the second half.
- Because we do not know the locations of MW-27 and MW-33 and, moreover, the concentration changes in these two wells are small, we do not consider their data.
- According to the groundwater flow direction, the average concentration value of MW-9 should not be higher than that of MW-3 and MW-12, which contradicts the given data for calcium, chloride, and so on. This is also true for barium. (In 1997, concentrations in MW3M and MW12M vary from 50 to 85, whereas they vary from 80 to 95 in MW9M.) Therefore, we think that MW-9 is a pumping well (**Figure 2**). Thus, we do not use the data from MW-9 in our analysis.

Finally, we list in **Table 1** the data for calcium that we use to calculate the source location.



**Figure 2.** Groundwater movement near a pumping well.

**Table 1.**  
Data for calcium used in the model.

Date	MW-3M	MW-7M	MW-11T	MW-12M
12/7/93	41	50	39	42
3/7/94	42	50	43	47
9/19/94	42	45	41	41
7/10/95	36.5	54.3	44.7	59.5
10/10/95	19.2	53	43.2	54.7
3/6/96	62.4	65.1	50.7	82.4
10/9/96	60.2	61.9	53.3	87.6
3/18/97	63.8	125	53.2	87.6
12/15/97	61.4	115	63.8	88.4

- According to the data, there is some pollutant detected in an early year such as 1990; we name it the *background concentration* ( $C_b$ ). We think that the later pollutants' concentrations consist of background concentration plus new injected concentration. According to **Figure 1**, MW-1 must be at the headwater level. Moreover, the data from its bottom hardly change during this period according to the data set. Thus, we estimate  $C_b$  using data from MW-1B as follows:

$$C_b = \text{arithmetic mean of the concentration value from MW-1B} \\ \text{during this period for a certain pollutant}$$

In **Table 2** we collect the symbols used in this paper and their definitions.

**Table 2.**  
Symbols used.

$\alpha_L$	horizontal dispersion coefficient (m)
$\alpha_T$	vertical dispersion coefficient (m)
$C$	pollutant concentration (mg/liter)
$C_b$	background concentration (described above)
$C_0$	concentration in the pollutant source (mg/liter)
$D$	pervasion coefficient (m <sup>2</sup> /s)
$H$	water level (ft)
$I$	hydraulic gradient
$K$	hydraulic conductivity (gal/day/ft <sup>2</sup> )
$L$	horizontal distance in the direction of water flow (ft)
$m$	discharge rate of the pollutant (mg/day)
$n$	effective porosity
$q$	discharge rate of the pollutant (liter/day)
$R_d$	retardation factor
$S$	compound parameter
$t_0$	start time of the pollution (yr)
$\theta$	angle between the direction of underground water and the $x$ -axis
$V_d$	groundwater velocity (ft/day)
$W$	hantush function
$(x_0, y_0)$	pollution source coordinate

## Model Design

### Model Formulation

The movement of pollutants consists of advection, dispersion, and retardation. Furthermore, regarding the large scale of the area, the vertical movement is negligible. Thus, movement of pollutant in the soil (saturated and unsaturated) can be described by the following two-dimensional equation:

$$R_d \frac{\partial C}{\partial t} = V_d \alpha_L \frac{\partial^2 C}{\partial x^2} + V_d \alpha_t \frac{\partial^2 C}{\partial y^2} - V_d \frac{\partial C}{\partial x}. \quad (1)$$

### Model Explanation

The model equation applies to steady uniform flow. An analytical solution to the equation can be developed for both continuous (step-function) and pulsed inputs of pollutants as boundary conditions. A step function implies the input of a constant concentration pollutant for an infinite amount of time, while a pulse load is a constant concentration input for a finite amount of time. The terms “infinite” and “finite” are relative to the time frame of the analysis.

We assume that the pollution source is applied as a step function (continuously) with the following boundary conditions:

$$\begin{aligned} C(x, y, 0) &= 0, & (x, y) &\neq (0, 0); \\ C(0, 0, t) &= C_0; \\ C(\pm\infty, y, t) &= C(x, \pm\infty, t) = 0, & t &\geq 0. \end{aligned}$$

## Model Solution

The function is a second-order partial differential equation. Equations of this form apply to a wide variety of problems, including mass transport, fluid dynamics, and heat transfer.

For an instantaneous point source at time  $t = 0$ , there is an analytical solution of the form

$$C(x, y, t) = S \exp\left(\frac{x}{2\alpha_L}\right) [W(0, b) - W(t, b)], \quad (2)$$

where

$$m = C_0 q, \quad S = \frac{m}{4\pi V_d (\alpha_L \alpha_T)^{1/2}},$$

and  $W(u, b)$  is the *hantush function*

$$W(u, b) = \int_u^\infty \frac{\exp\left[-y - \frac{b^2}{2y}\right]}{y} dy \quad \text{with} \quad b = \sqrt{\frac{x^2}{4\alpha_L^2} + \frac{y^2}{4\alpha_L \alpha_T}}.$$

Before computing, we classified the parameter used according to our assumptions above:

- During the data processing, the coordination and time of the pollution source are unknown, and so is the value of  $m$ . Thus,  $x_0$ ,  $y_0$ ,  $t_0$ , and  $S$  are variable.
- The parameters  $\alpha_L$ ,  $\alpha_T$ ,  $\theta$ , and  $V_d$  are constants.

The main task is to find the location and the start time of the pollutants. Hence, we develop a grid-search optimization routine to get an optimized solution:

- We estimate the location of the pollutant source and transform coordinates as follows:
  - Set the point of the pollutant source to be the new origin.
  - Set the new  $x$ -axis to be parallel to the direction of the underground water flow.
  - Set the new  $y$ -axis to be perpendicular to the new  $x$ -axis.
- We construct an equation to calculate the movement of the pollutant under the ground. We calculate the concentration changes in each well and compare with the changes according to the data set. We repeatedly adjust the location of the pollution source, the value  $S$ , and the value  $t_0$  (detailed in the following) until there is a satisfactory agreement. The criterion for convergence is the sum of the squares of the residuals between the data and the model predictions. The objective function to be minimized is

$$\sum_i [(C_i - C_b) - C'_i]^2,$$

where  $C_i$  is the pollutant concentration data value for well  $i$ ,  $C'_i$  is the model prediction, and  $C_b$  is the background pollution level.

## Parameter Estimation

We estimate the parameters for the saturated zone as following:

- **Hydraulic Conductivity  $K$ :** We consider hydraulic conductivity, measured in gallons per day per square foot, only in the horizontal direction. According to the literature,  $K = 265 \text{ gpd/ft}^2$  ( $1 \text{ gpd/ft}^2 = 4.72 \times 10^{-5} \text{ cm/sec}$ ).
- **Hydraulic Gradient:** According to **Figure 1**, made by interpolation, we assume that the direction of the underground water is one-dimensional.
- **Ground-Water (Interstitial Pore Water) Velocity  $V_d$ :** According to Darcy's Law,  $V_d$  is defined as

$$V_d = -KI/n,$$

where  $I$  is the hydraulic gradient,  $K$  is hydraulic conductivity, and  $n$  is effective porosity. We assume that the soil type of the saturated zone is sand with porosity 20%, so we estimate  $V_d = 1.5 \text{ ft/day}$ .

- **Dispersion Coefficient  $\alpha$ :** This coefficient incorporates two forms of dispersive process: dynamic dispersion and molecular diffusion. According to the literature, the horizontal dispersion coefficient and the vertical dispersion coefficient are approximately equal. Both have the estimated value 25 ft.
- **Retardation Factor:** Retardation is based on pollutant characteristics and aquifer composition. Since its effect is not very significant, we estimate  $R_d = 1$ .
- **Concentration in Pollution Source:** According to the literature, when the water table is usually sufficiently high so that the pollutant directly enters ground water, the  $C_0$  value is the estimate of the source concentration.

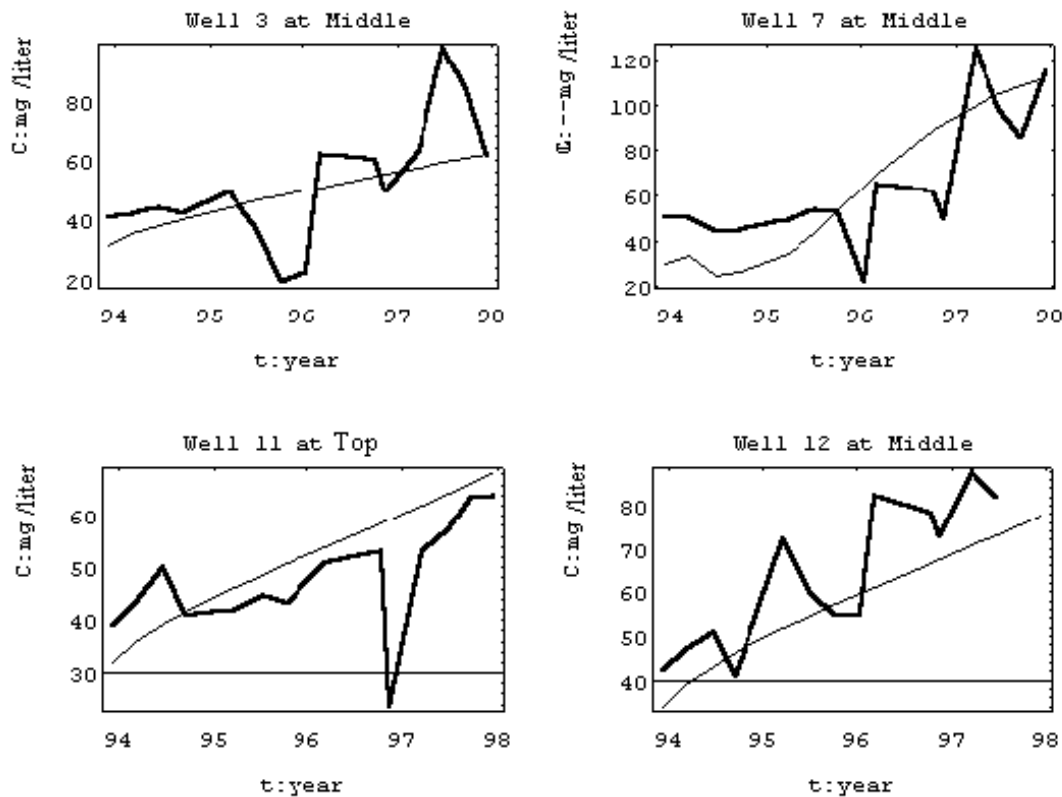
## Results

There are four new pollutants: calcium, chloride, magnesium, and TDS. The location and the start time for the pollution sources, as predicted by our model, are in **Table 3**.

**Table 3.**  
Source and start time for pollutants.

Pollutant	$x$ -coordinate (ft)	$y$ -coordinate (ft)	Start time (m/d/y)
TDS	7077	6538	8/12/91
Magnesium	6423	7461	1/1/94
Chloride	6931	5823	5/18/91
Calcium	7750	6040	9/1/93

Finally, we mimic the movement process of the pollutant in reverse and compare with the given data set (**Figure 3**). From the graphs, we conclude:



**Figure 3.** Calcium concentrations at four wells. The thick curves are data, the thin curves are model predictions.

- For near-ideal conditions, the model is suitable; for regular use, a more robust model is desired.
- Even though the two curves do not fit very well, they show a similar change tendency.

## Sensitivity Analysis

We conduct a rudimentary sensitivity analysis to explain the stability of our model. We separately vary the values of the constants  $\alpha_L$ ,  $\alpha_T$ ,  $\theta$ , and  $V_d$  by 10% and compute the corresponding changes in the values of the location and time of the pollution source (**Table 4**)

The model demonstrates good stability, but  $\theta$  has a relatively significant influence on the result of the model. Thus, it is reasonable to consider the parameter  $\theta$  as a variable and repeat our grid-searching algorithm in a five-dimensional space of  $\theta$ ,  $x_0$ ,  $y_0$ ,  $t_0$ , and  $S$ . For calcium, we get the comparative results shown in **Table 5**.

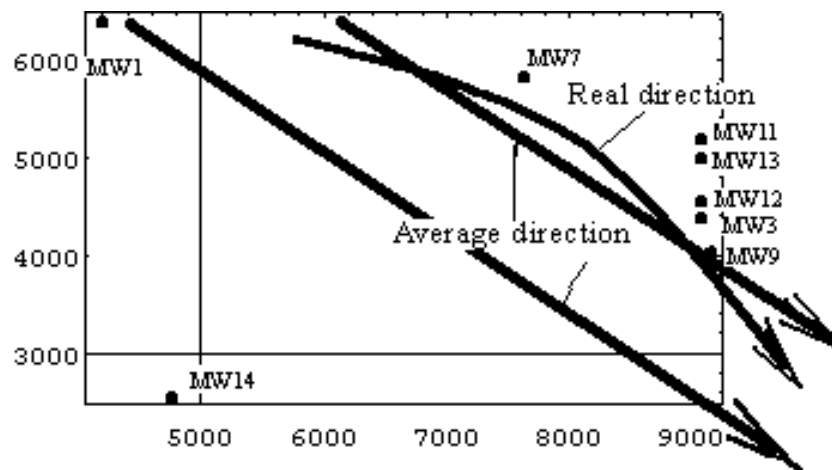
In the expanded model, the value for  $\theta$  is 7% larger. We think that there is some deflection of the direction of the groundwater flow, as shown in **Figure 4**.

**Table 4.**  
Effects of perturbations of the parameter values.

Parameter	Change in location (ft)	Change in time (yr)
$\theta$	70	0.2
$\alpha_L$	<10	<0.1
$\alpha_T$	10	<0.1
$V_d$	<10	<0.1

**Table 5.**  
Comparison of 4- and 5-dimensional models.

Dimension	$\theta$	$x_0$ (ft)	$y_0$ (ft)	$t_0$ (yr)	$S \times 10^6$
4	0.785	7750	6040	93.75	2.1
5	0.84	7750	6100	93.60	2.2



**Figure 4.** Suspected deflection of groundwater flow.

## Problem Two

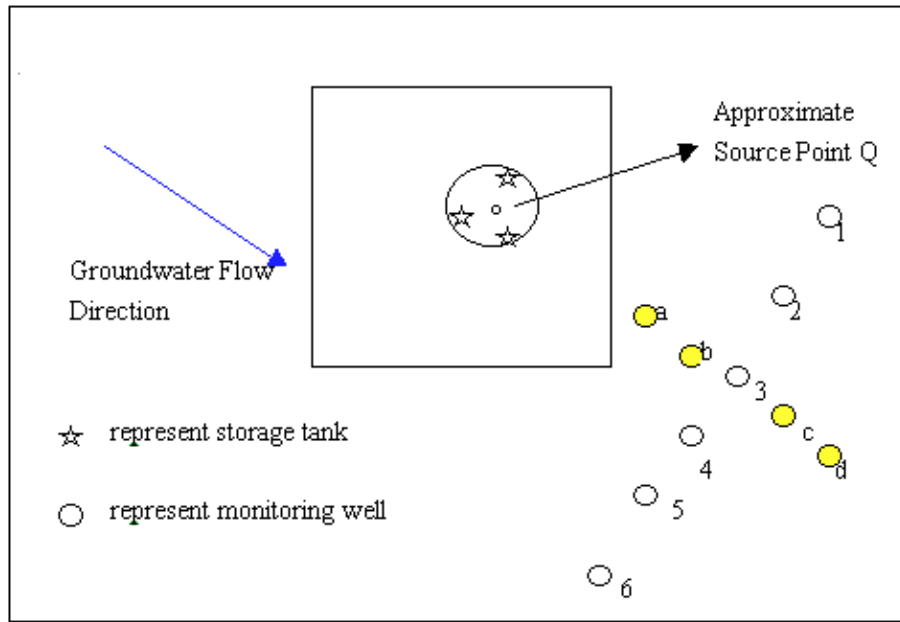
### Local Assumptions

- The storage tanks are located underground in the saturated zone.
- The direction of the groundwater flow remains the same.
- The saturated zone is semi-infinite.
- The leak process is continuous, since the primary cause of leaks in steel underground storage systems is corrosion.

## Model Design

To detect the pollutant rapidly and accurately, we develop a three-step method.

1. According to the shape and size of the storage and the direction of the groundwater flow, we determine the number and location of the first group of wells. Provided that the storage is a square  $S$  m on a side, the number of the first group of wells is  $N = S/20$ . That is, we drill a well every 20 m in a line perpendicular to the direction of the groundwater, as shown in **Figure 5**. We monitor the data from the wells.



**Figure 5.** Locations of monitoring wells. Empty circles represent the initial monitoring wells; filled circles are the wells drilled after pollution is detected and found to affect well 3 most of all.

2. Once there is some evidence of pollution, we determine which well is most affected by the pollutant. Near this well, we drill a series of wells (perhaps five or more) along the direction of the groundwater flow. Thus, we can construct a three-dimensional formulation to calculate the fluctuation of the pollutant concentration. Here the area occupied by the storage facility may not be very large (with side less than 1000 ft), so we cannot use (1). We employ the three-dimensional equation

$$R_d \frac{\partial C}{\partial t} + V_d \frac{\partial C}{\partial x} = D \left( \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2} \right) + \frac{m}{n}.$$

Because the leaking is a continuous process, we assume that the pollution

source is applied as a step function (continuously) with boundary conditions

$$\begin{aligned} C(x, y, z, 0) &= 0, \quad (x, y, z) \neq (0, 0, 0), \\ m(x, y, z, t) &= qC_0\delta(x, y, z), \\ C(\pm\infty, y, z, t) &= C(x, \pm\infty, z, t) = C(x, y, \pm\infty, t) = 0, \quad t \geq 0. \end{aligned}$$

For an instantaneous point source at time  $t = 0$ , this equation possesses an analytical solution of the form

$$\begin{aligned} C(x, y, z, t) &= \frac{R_d q C_0}{8\pi n D r} \exp\left(\frac{V_d}{2D}\right) \\ &\times \left\{ \exp\left(\frac{V_d x}{2D}\right) \operatorname{erfc}\left[\frac{r + V_d t}{2} \left(\frac{R_d}{D_t}\right)^{1/2}\right] \right. \\ &\quad \left. + \exp\left(\frac{-V_d}{2D}\right) \operatorname{erfc}\left[\frac{r - V_d t}{2} \left(\frac{R_d}{D_t}\right)^{1/2}\right] \right\}, \end{aligned}$$

where  $r = (x^2 + y^2 + z^2)^{1/2}$ .

When  $t \rightarrow \infty$ , a steady-state equation results:

$$C(x, y, z, t) = \frac{R_d q C_0}{4\pi n D r} \exp\left(\frac{V_d(r - x)}{2D}\right). \quad (3)$$

For convenience, we employ the symbol  $C_m(x, y, z, t)$  to represent the right side of the (3).

For constant  $V_d$ ,  $R_d$ ,  $n$ ,  $q$ , and  $D$ , we can draw an equal-concentration plane with the concentration value  $0.01C_0$ , as in **Figure 6**.



**Figure 6.** Large dose.

Let Height be the maximum height of the equal-concentration plane. We transform the Cartesian coordinates in the same way as Problem One.

For a monitoring well at  $(x, y)$  and aquifer thickness  $b$ , we consider the concentration in the well for three situations:

- If  $b \ll \text{Height}$  or  $b \ll \text{Size of the storage facility}$ , we can transform (3) into a two-dimensional equation like (1).
- If  $b \geq \text{Height}/2$ , it is reasonable to consider  $b = \infty$ . Thus, the problem can be simplified. We assume that the substance in the aquifer cannot enter the unsaturated zone except at the source point. Thus,  $\partial C / \partial z|_{z=0} = 0$ .

Moreover, for every point  $(x, y, z)$  under the water table, the concentration is double that depicted by (3) in the case of semi-infinite space:

$$C(x, y, z, t) = 2C_m(x, y, z, t).$$

- Otherwise, for  $(x, y, z)$  on the upper or lower surface of the aquifer (see Figure 7), we have

$$\left. \frac{\partial C}{\partial z} \right|_{z=0} = \left. \frac{\partial C}{\partial z} \right|_{z=-b} = 0.$$

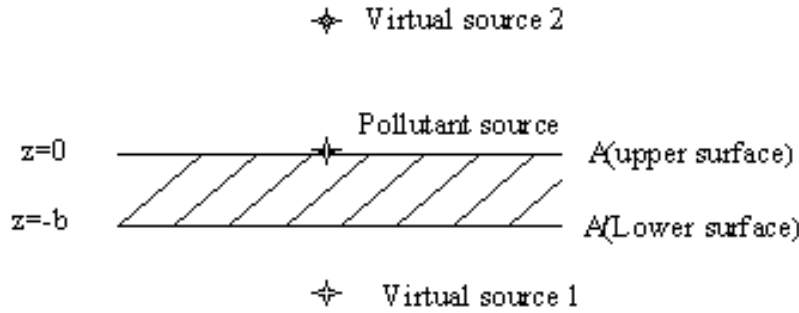


Figure 7. Pollution source on the aquifer .

Draw virtual source 1 symmetric to the pollutant source with the axis being A' (the lower surface). Thus, condition  $\left. \frac{\partial C}{\partial z} \right|_{z=-b} = 0$  is satisfied, while the condition  $\left. \frac{\partial C}{\partial z} \right|_{z=0}$  is not satisfied. In the same way, we draw virtual source 2 symmetric to virtual source 1 with A (the upper surface). Repeating this process, we get virtual source 3, and so on.

The concentration on the upper or lower surface can be considered as the result of accumulation of all the concentration values of all the sources (including the virtual sources). That is,

$$C_t(x, y, z) = 2 \sum_{i=0}^{\infty} C_m(x, y, z + 2(-1)^{i+1} \lfloor \frac{i+1}{2} \rfloor b, t).$$

Actually, we need to consider only the former three virtual sources, for the following reasons:

- The distances from these three sources to A(A') are the smallest, so they have the most effect on  $C_t$ . Other virtual sources are very far away from A(A') and generally the distance between them is larger than the value Height. Therefore, we neglect these virtual sources.
- The pollutant discharged from the virtual sources far from the surfaces of the aquifer needs a long time to reach the aquifer.

Finally, we transform (3) into

$$C_t(x, y, z) = 2 [C_m(x, y, z, t) + C_m(x, y, z + 2b, t) + C_m(x, y, z - 2b, t)].$$

Thus, we get the final analytical solution of  $C(x, y, z, t)$ . Then we use the same computer-based method as in Problem One to calculate the approximate location and the time of the pollutant source.

3. In the last step, we draw a circle with center the approximate source point  $Q$  and diameter 25 m (or more). Inside this circle, we sample some soil from the surface and analyze its chemical constituents to find the maximum. Thus, we can accurately identify the location of the pollutant source.

## Numerical Integration Scheme

To calculate leakage, it is necessary to integrate the values of the dependent variable ( $C$ ) over space. Unfortunately, the integral of (3) does not possess an analytical representation and must therefore be integrated numerically. We employ a three-dimensional integration scheme for this model. The molar mass ( $M$ ) of leaked liquid is computed as

$$M = \int \int \int C(x, y, z, t) dx dy dz \approx \sum_{i,j,k} C_{ijk} \delta x \delta y \delta z,$$

where  $C_{ijk}$  refers to the computed concentration in “differential” element  $(i, j, k)$ . We employ uniform spatial steps of  $\delta x = \delta y = \delta z = 1$  m.

## A Better Method for Mass Estimation

While processing the data by computer program, we minimize the variance to get a quasi-optimal solution. Meanwhile, we have estimated the  $m$  value, so we can compute the molar mass of leaked liquid more conveniently and efficiently as

$$M = mt.$$

## Strengths and Weaknesses of the Model

### Strengths

- The model has quite good practicality, and the given algorithm has little time complexity. For the given problem size, our C program for the grid-search algorithm runs in less than 2 min on a Pentium-166 computer.
- The model gives good agreement of predicted values and data. It is fast, efficient, and stable.
- As the given data are refined to simplify the computation, the accuracy does not decrease. For illustration, we list data for calcium in **Table 6**.

**Table 6.**  
Effect of refining to simplify computation.

Number of data points	$x_0$ (ft)	$y_0$ (ft)	$t_0$ (years since 1900)
60 (primitive)	7750	6060	93.70
36 (after refining)	7750	6040	93.75

## Weaknesses

- If the detected area is not large enough, there is some error. As the distance between the pollution source and the monitoring well grows larger, computation accuracy increases, measurement accuracy decreases, and response time increases.
- To decrease the complexity of the computation, we simplify the groundwater flow net, which may affect the accuracy of the results.
- Not taking statistical factors into our model makes the result of our model not fit the crude data exactly.

## References

- Abriola, Linda M., editor. 1989. *Groundwater Contamination*. Wallingford, U.K.: International Association of Hydrological Sciences.
- Barcelona, Michael, Allen Wehrhann, Joseph F. Keely, and Wayne A. Pettyjohn. 1990. *Contamination of Ground Water: Prevention, Assessment, Restoration*. Park Ridge, NJ: : Noyes Data Corp.
- Guswa, J.H., W.J. Lyman, A.S. Donigian, T.Y.R. Lo, and E.W. Shanahan. 1984. *Groundwater Contamination and Emergency Response Guide*. Park Ridge, NJ: Noyes.
- Ward, C.H., W. Giger, and P.L. McCarty, editors. 1985. *Ground Water Quality*. New York: Wiley.