A new mathematical technique for identifying potential sources of groundwater contamination

W. R. Laton · R. J. Whitley · T. V. Hromadka II

Abstract The observed hydrogeochemical condition of groundwater at a particular well is usually represented as a mixture of various sources of pollution and background conditions and is given in terms of measurements of multiple dissolved inorganic water contaminants such as total dissolved solids (TDS). Concentrations from a given set of wells can be compared against one another in a variety of ways, but the consideration as to which chemical concentrations are best related to one another is limited. In this analysis, an example is given to show that if there are a total of N concentration values, all N must be considered simultaneously in order to ascertain whether the observed conditions at the well can be explained as a mixture, and this can be done by solving a quadratic programming problem-convex hull.

Résumé L'état hydrogéochimique des eaux souterraines observé au droit d'un puits particulier est usuellement représenté comme un mélange de diverses sources de pollutions et de bruits de fond et est donné en terme de mesures de plusieurs contaminants inorganiques dissous, comme les solides totaux dissous par exemple (STD). Les concentrations issues d'un groupe de puits peuvent être comparées entre elles de différentes manières, mais le souhait de corréler au mieux telle concentration chimique avec une autre est limité. Dans cette étude, l'exemple traité montre que s'il existe un nombre total N de valeurs de concentration, ces Nvaleurs doivent être considérées simultanément afin de déterminer si les conditions observées au puits peuvent s'expliquer en terme de mélange ; ceci peut être réalisé en résolvant un problème quadratique programmable-une enveloppe convexe.

Resumen La condición hidrogeoquímica observada en un pozo particular es generalmente resultante de la mezcla de varias fuentes de contaminación y las condiciones antecedentes y se traduce en términos de medidas de múltiples contaminantes inorgánicos disueltos tal como los sólidos totales disueltos (STD). Las concentraciones de un conjunto de pozos pueden ser comparadas unas contra otras por una variedad de métodos pero la consideración de las concentraciones químicas que mejor se relacionan es limitada. En este análisis se introduce un ejemplo que muestra que si existe un total de N valores de concentración, todos los N deben ser considerados simultáneamente a fin de establecer si las condiciones observadas en el pozo pueden ser explicadas como una mezcla. Esto se realizada por resolución de un problema cuadrático de envolvente convexa.

Keywords Hydrochemical modeling · Numerical modeling · Quadratic programming · Convex hull

Received: 5 April 2004 / Accepted: 22 August 2006
Published online: 28 October 2006

W. R. Laton (✉)
Department of Geological Sciences,
California State University, Fullerton,
800N. State College Blvd., MH-208,
Fullerton, CA 92834, USA
e-mail: wlaton@fullerton.edu

R. J. Whitley
Department of Mathematics,
University of California, Irvine,
103 Multipurpose Science & Technology Bldg.,
Irvine, CA 92697-3875, USA
e-mail: rwhitley@uci.edu

T. V. Hromadka II
Department of Mathematical Sciences,
United States Military Academy,
West Point, NY 10996, USA
e-mail: ted@phdphdphd.com

Introduction

Trying to ascertain which of several possible sources is most likely responsible for observed conditions at a given target well is a common problem faced by hydrogeologists. Typically, a range of chemical constituents, including total dissolved solids (TDS), is measured at the target well. Similar measurements may have been made at one or more wells which represent the background or unpolluted state of the groundwater, and have also been made at one or more wells that most closely represent possible sources of pollution. The determination of the hydrogeochemical state is generally approached in either a graphical (e.g., stiff or Piper diagrams) or statistical manner (Fetter 2001; Dominico and Schwartz 1998).

Any of these measurements at a given well, at a given time, can be represented as a vector (**V**) listing the

 DOI 10.1007/s10040-006-0106-4

amounts, in appropriate units, of each of 'm' measured dissolved solids: $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m)$.

One way of formalizing this problem is to ask whether the observed conditions at the target well, represented by the vector **W**, can be explained as a mixture of the background conditions, represented by the vectors $\mathbf{B}^1$,..., $\mathbf{B}^{n1}$, and the pollution source represented by the vectors $\mathbf{P}^1$,..., $\mathbf{P}^{n2}$. The notation is simplified by letting $\mathbf{V}^1$,..., $\mathbf{V}^n$ be the collection of the **B** vectors and the **P** vectors. The question then is how well the **W** vector can be represented by a mixture, a perfect representation being:

$$\mathbf{W} = \sum_{k=1}^{n} \alpha_k \mathbf{V}^k \tag{1}$$

where α_k represents the portion of $\mathbf{V}^k$ in the mix, and so we must have

$$\alpha_k \geq 0, \; for \; k = 1, \ldots\ldots, n \tag{2}$$

and

$$\sum_{k=1}^{n} \alpha_k = 1. \tag{3}$$

Mathematical formulation

This multi-source problem can be described in terms of the notion of a convex set, a simple but powerful notion, (see for example Peressini et al. 1991; Bazaraa et al. 1993). A subset S in $\mathbf{R}^m$ ($\mathbf{R}^m$ is the space of m-dimensional vectors, each of which has the form (a_1, a_2,..., a_m), where as a_1, a_2, .., a_m are real numbers) is "convex" if given the vectors **s** and **t** in S, and any real number α in [0,1] (α is a real number satisfying, $0 \leq \alpha \leq 1$), $\alpha\mathbf{t} + (1-\alpha)\mathbf{s}$ also belongs to S, i.e. the line joining any two vectors in S also lies entirely in S. Given a subset T of $\mathbf{R}^m$, the "convex hull" of T is the smallest convex set containing T. If, for example, T consists of two vectors, its convex hull is the line joining those points; if T consists of three vectors, its convex hull is the triangle having those vectors as vertices (Fig. 1). A basic result (Peressini et al. 1991, theorem 2.1.4) or (Bazaraa et al. 1993, Theorem 2.1.6) is that if T is the set $\{\mathbf{V}^1, \mathbf{V}^2, ..., \mathbf{V}^m\}$, then the convex hull of T is exactly the set of all convex combinations of the vectors in T, i.e. it consists of all the sums (Eq. 1) with alphas satisfying the conditions of Eqs. (2) and (3). That is to say that **W** is a mixture of the vectors $\mathbf{V}^1$, $\mathbf{V}^2$,..., $\mathbf{V}^m$ if and only if **W** belongs to the convex hull of these vectors.

Writing the vectors $\{\mathbf{V}^1, \mathbf{V}^2, ..., \mathbf{V}^m\}$ in terms of the **B** and **P** vectors, if Eq. 1 holds, then

$$\mathbf{W} = \sum_{k=1}^{n_1} \alpha_k \mathbf{B}^k + \sum_{k=n_1+1}^{n_2} \alpha_k \mathbf{P}^k \tag{4}$$

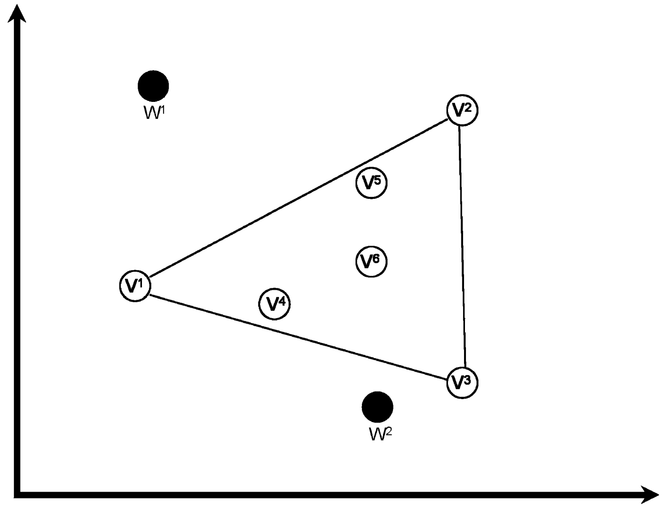

Fig. 1 Example of two-dimensional convex hull; $\mathbf{W}_1$ and $\mathbf{W}_2$ are not in the convex hull bounded by $\mathbf{V}^1$, $\mathbf{V}^2$ and $\mathbf{V}^3$

Letting $\gamma = \sum_{k=1}^{n_1} \alpha_k$, Eq. (4) can be written

$$\mathbf{W} = (\gamma) \sum_{k=1}^{n_1} \frac{\alpha_k}{\gamma} \mathbf{B}^k + (1-\gamma) \sum_{k=n_1+1}^{n_2} \frac{\alpha_k}{1-\gamma} \mathbf{P}^k \tag{5}$$

The term multiplied by γ is in the convex hull of $\mathbf{B}^1$,..., $\mathbf{B}^{n1}$ while the term multiplied by ($1-\gamma$) is in the convex hull of $\mathbf{P}^1$,..., $\mathbf{P}^{n2}$, and **W** is written as a mixture of a background mixture term and a pollution mixture term.

Since vectors in $\mathbf{R}^m$, m>3 are difficult to visualize, answering the question of whether **W** is a mixture of background and a specific pollution source, is addressed by considering subsets consisting of two or three of the m measurements, for if **W** is exactly (or is approximately) a convex combination of the **B** and **P** vectors, then equation (4) holds (or holds approximately) for all components and so for any subset of the components.

Theoretical example

The simplest case is to consider a subset of one coordinate, corresponding to a single TDS reading. Any one component of **W**, say the first, is in the convex hull of the first components of the **B** and **P** vectors if and only if $\mathbf{w}_1$ lies in the interval with the left endpoint the minimum of the first components of the **B** and **P** vectors and with right endpoint the maximum. For two coordinates, say the first and the second, the convex hull of the first two components of the **B** and **P** vectors is a convex polygon in the plane with "vertices" at the points (b_1^i, b_2^i), i=1,..., n_1 and (p_1^i, p_2^i), i=1,..., n_2, where some of these "vertices" may fall inside the polygon and so are not true vertices. For subsets of three coordinates, ternary diagrams can be used to visualize whether or not a vector made up of three coordinates of **W** lies in the convex hull of the vectors

 DOI 10.1007/s10040-006-0106-4

consisting of the corresponding three coordinates of the **B** and **P** vectors. If any of these tests show that the vector made up of a subset of 1, 2, or 3 coordinates of the **W** vector does not lie in the convex hull of the corresponding reduced **B** and **P** vectors, then we know that **W** cannot lie in the convex hull of the **B** and **P** vectors in $\mathbf{R}^m$.

Here, an example is provided to show that the problem of deciding whether or not **W** is in the convex hull of a set cannot be simplified by looking at subsets of fewer coordinates. For the first example consider these vectors in $\mathbf{R}^2$: $\{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2\}$, where $\mathbf{e}_0$=(0, 0), $\mathbf{e}_1$=(1, 0), $\mathbf{e}_2$=(0, 1). The convex hull of these vectors is a two-dimensional triangle. Let **W** be a point belonging to the rectangle which has three corners at the given points, but **W** not in the triangle. Visually it is clear that **W** passes both one-coordinate tests for belonging to the convex hull of the three given vectors, but does not pass the two-coordinate test, i.e., it is not in the convex hull.

To construct an example for the case of *m* coordinates (*m* TDS values), let *T* be the subset of $\mathbf{R}^m$ consisting of the vectors

$$T = \{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m\}, \tag{6}$$

where $\mathbf{e}_0$=(0, 0,..., 0), $\mathbf{e}_1$=(1, 0, 0,..., 0), $\mathbf{e}_2$=(0, 1, 0, 0,..., 0), . . ., $\mathbf{e}_m$=(0, 0,..., 1). Choose

$$\mathbf{W} = (\gamma, \gamma, \ldots, \gamma) \tag{7}$$

where γ is chosen so that $\frac{1}{m} < \gamma < \frac{1}{m-1}$. If we consider any subset of coordinates, with no loss of generality the first *m*−1,

$$(w_1, w_2, \ldots, w_{m-1}) = \sum_{k-1}^{m-1} \gamma\mathbf{e}_{\mathrm{k}} + [1 - (m-1)\gamma]\mathbf{e}_0, \tag{8}$$

and thus, when restricted to the first *m*−1 coordinates, **W** lies in the convex hull of the similarly restricted set of vectors from *T*; and then it is true that **W** will be in the convex hull for any proper subset of the m coordinates. However, **W** does not lie in the convex hull of *T*, indeed, using Lagrange multipliers,

$$\frac{1}{m\gamma}\mathbf{W} = \frac{1}{m}\sum_{k=1}^{m} \mathbf{e}_{\mathrm{k}}, \tag{9}$$

is the vector in the convex hull of *T* which is closest to **W**, and since $\frac{1}{\gamma m} < 1$, **W** is not itself in the convex hull of *T*.

With the example above in mind, the problem of finding out whether or not **W** belongs to the convex hull *K* of the vectors $\mathbf{V}^1$, $\mathbf{V}^2$,..., $\mathbf{V}^n$ requires, in general, using the full set of coordinates. This can be done as follows: let *A* bethe *m* by *n* matrix with entries a_{ij}=$\mathbf{V}^i(j)$. For any point $\alpha_1\mathbf{V}^1$+$\alpha_2\mathbf{V}^2$+. . .+ $\alpha_n\mathbf{V}^n$ in *K*, if **x** is the vector in $\mathbf{R}^n$ with $\mathbf{x}_1$=α_1, $\mathbf{x}_2 = \alpha_2$,..., $\mathbf{x}_n$=α_n, then

$$Ax = \alpha_1\mathbf{V}^1 + \alpha_2\mathbf{V}^2 + \ldots + \alpha_n\mathbf{V}^n. \tag{10}$$

Let (**x**, **y**) be the inner product of **x** with **y** in $\mathbf{R}^m$. The square of the distance of **W** to the point *A***x** in *K* is:

$$\begin{aligned}\mathrm{dist}(A\mathbf{x}, \mathbf{W})^2 &= (A\mathbf{x} - \mathbf{W}, A\mathbf{x} - \mathbf{W}) \\ &= (A\mathbf{x}, A\mathbf{x}) - 2(\mathbf{W}, A\mathbf{x}) + (\mathbf{W}, \mathbf{W})\end{aligned} \tag{11}$$

$$= \left(\mathbf{x}, A^T A\mathbf{x}\right) - 2\left(\mathbf{x}, A^T\mathbf{W}\right) + (\mathbf{W}, \mathbf{W}). \tag{12}$$

To find the closest point to **W** in *K*, minimize the expression:

$$\left(\mathbf{x}, A^T A\mathbf{x}\right) - 2\left(\mathbf{x}, A^T\mathbf{W}\right) \tag{13}$$

over all **x** constrained by x_1+x_2 +. . .+x_n=1 and all x_i=0. As A^TA is a non-negative definite matrix, this is a quadratic programming problem (Peressini et al. 1991; Bazaraa et al. 1993). If **x** is a solution of this minimization, then the distance from *A***x** to **W** is the distance from **W** to *K*, and it can be seen if this is zero or not. In addition, the magnitude of this distance, if it is not zero, gives a numeric measure of how close **W** is to being the nearest mixture *A***x**. It is often the case that the measurements of the totally dissolved solids, i.e., the components of the vectors, are subject to significant errors, in which case **W** being close to *K* is perhaps adequate evidence.

The problem of minimizing a linear function **g(x)** = c_1x_1+c_2x_2+. . .+c_nx_n over all **x** constrained by x_1+x_2+. . . + x_n=1 and all $x_i \geq 0$ is an example belonging to the well-known class of linear programming problems. Changing the function to a quadratic function makes the problem more difficult to solve; however, simple programming subroutines are readily available. The following sample problem was solved via a Fortran 95 program compiled using a Lahey-Fujitsu Fortran 95 compiler and Visual Numerics IMSL quadratic programming subroutine (R. Whitley, University of California Irvine, unpublished data, 2006).

Application of the convex hull approach

In the following is discussed an application of the convex hull approach to a situation where TDS levels are increasing in groundwater. An aquifer near Riverside, California, USA, with unconfined conditions, comprised primarily of sands and gravels with occasional interbedded silts, was chosen. Depth to groundwater is 60–90 m below ground surface and is flowing in a southwesterly direction (Fig. 2). The study area has elevated total dissolved solid (TDS) levels within several wells of which only one, well W will be looked at. Well W is an intermittent pumping well that has seen an increase in TDS with time. The possible source(s) were identified to be one of several up-gradient industries, which for the

DOI 10.1007/s10040-006-0106-4

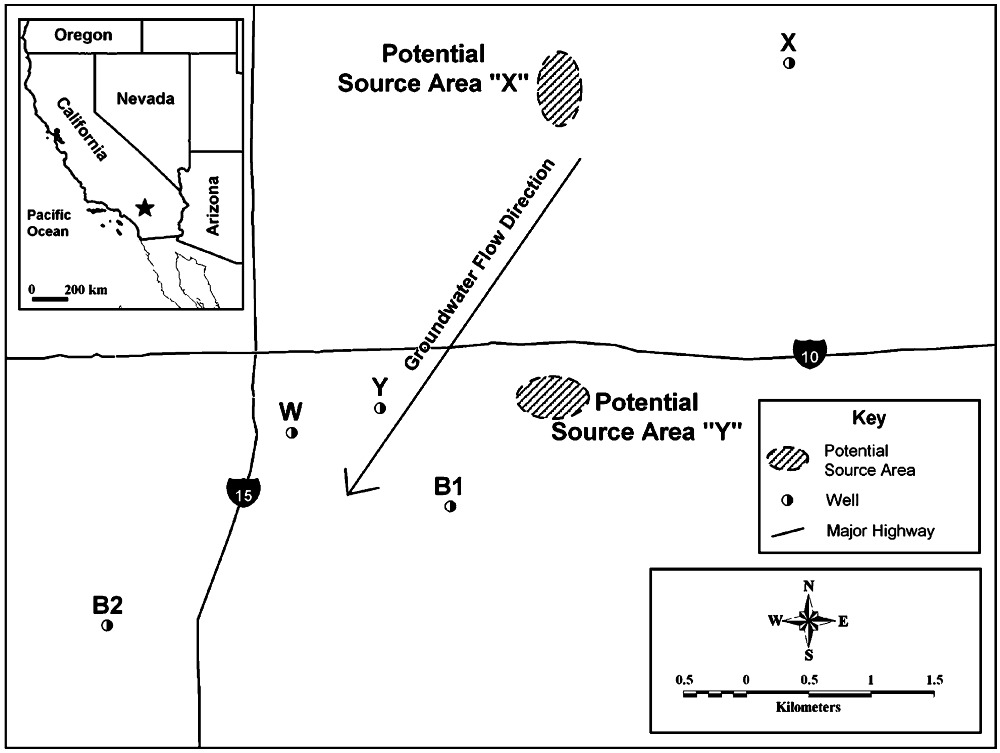

Fig. 2 Location map of the study area near Riverside, California, USA (marked with a *star*). Two areas are identified as potential sources of pollution (Source area *X* and source area *Y*)

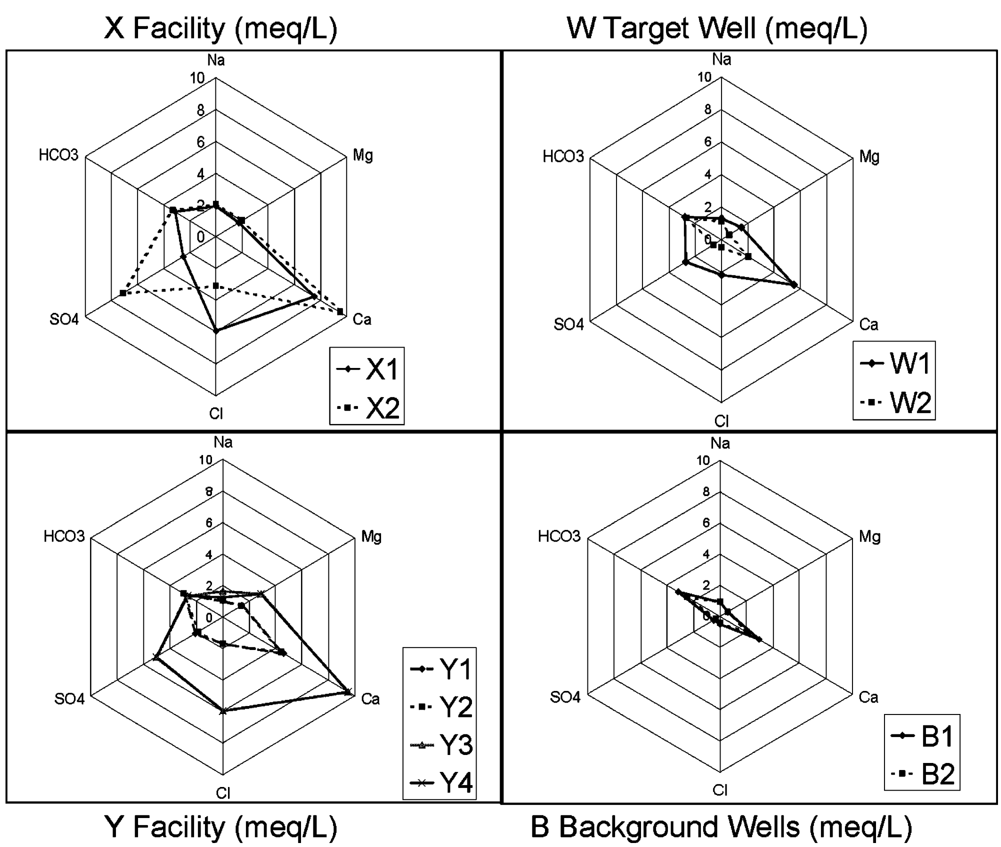

Fig. 3 Radial diagram of concentrations

 DOI 10.1007/s10040-006-0106-4

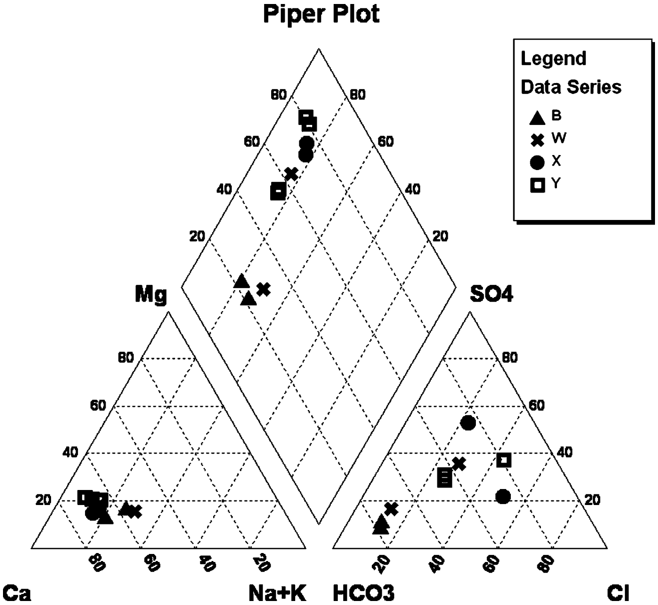

Fig. 4 Piper diagram of concentrations (mg/L)

purpose of this paper will be referred to as possible source areas X and Y. Based on modeling results of the groundwater regime, it was deemed possible that either facility could be the cause of the elevated TDS issue. Using a simple graphical approach (Figs. 3 and 4) proved difficult in determining which facility is the dominant contributing source. Through the use of the convex hull modeling technique, however, it was determined that facility X was the most likely source of the elevated TDS readings. Based on proximity of facility to the impacted well, one would have thought the most likely candidate would be facility Y. Based on flow modeling, facility X was shown to be the source of the high TDS waters.

Using available data sets, two up-gradient wells (X and Y), one designated for each potential source facility, were compared to the target well, W, and background readings. In the example there are two measurements $\mathbf{W}_1$ and $\mathbf{W}_2$ taken at the target well for which this comparison was established. Each sample contains values for the following constituents: Na, Mg, Ca, Cl, SO_4, HCO_3 (Table 1); note this makes the dimension of the vectors n=6 (in the actual problem there were not two, but more than 30 such

Table 1 Sample data (mg/L)

Sample ID	Na	Mg	Ca	Cl	SO_4	HCO_3
X_1	44	21	150	210	120	190
X_2	47	24	190	110	340	200
Y_1	23.5	17.2	92.4	60	100	182
Y_2	23.6	16.8	89	61	90	182
Y_3	38	35	190	211	242	163
Y_4	28	35	190	211	242	163
$\mathbf{W}_1$	30	18	110	77	130	170
$\mathbf{W}_2$	26	7	41	17	29	160
$\mathbf{B}_1$	21	7.1	58.6	18.1	21.9	193
$\mathbf{B}_2$	21.5	7.1	40	14.7	13.3	149.1

Table 2 Results of convex hull model

Convex hull model	
CX=convex hull of X_1, X_2, $\mathbf{B}_1$ and $\mathbf{B}_2$	
Distance of $\mathbf{W}_1$ to CX	7.26
Distance of $\mathbf{W}_2$ to CX	9.01
CY=convex hull of Y_1, Y_2, Y_3, Y_4, $\mathbf{B}_1$ and $\mathbf{B}_2$	
Distance of $\mathbf{W}_1$ to CY	14.39
Distance of $\mathbf{W}_2$ to CY	10.44

measurements). The background (unpolluted) levels of these measurements are represented by two measurements $\mathbf{B}_1$ and $\mathbf{B}_2$ (again, in the actual problem, there were more than 30 measurements in the first set which have been averaged to give $\mathbf{B}_1$, and more than 20 in the set which has been averaged to give $\mathbf{B}_2$). There are also measurements taken at the two possible up-gradient sources (X and Y) for the elevated TDS seen at the target well, W (Fig. 2).

The possible source X is represented by two measurements X_1 and X_2, and the possible source Y is represented by four measurements Y_1, Y_2, Y_3, and Y_4 (Table 1). Note: using more points for Y generally reduces the distance to the convex hull. The distance to the convex hull of the target well samples from a mixture of the source plus the background indicates how well the observed pollution can be approximated by such a mixture (Table 2). This distance to the convex hull is seldom equal to zero because of measurement errors, measurements decreasing due to recharge in the aquifer, differential transport of the pollutants, and multiple sources of pollution, but these distances can still be compared. The distances to the convex hull between well **W** and the X facility were calculated to be between 7.26, using $\mathbf{W}_1$ and 9.01, using $\mathbf{W}_2$, whereas the distance between well W and the Y facility is calculated to be between 10.44, using $\mathbf{W}_1$ and 14.39, using $\mathbf{W}_2$. This indicates that in comparing the two sources, facility X is more likely to be the source of elevated TDS than that of facility Y.

Conclusion

The observed hydrogeochemical conditions at a particular well are usually represented as a mixture of various sources of pollution and background conditions and are given in terms of measurements of multiple dissolved inorganic water contaminants such as total dissolved solids (TDS). The convex hull approach to mathematically separate out different source potentials will aid in this quest to define source areas in problems associated with complex hydrogeology and potential multiple confusing sources. It should be acknowledged that groundwater chemistries are complex and simple dilution scenarios are easy to model; however, this approach can shed light on the statistically most likely candidates for the source(s). As seen in the above example, using quadratic programming, the problem of identifying potential source areas from one another can be rapidly determined and allow for multiple points of interest. Although the provided appli-

 DOI 10.1007/s10040-006-0106-4

cation was a simple problem with only two candidate sources, the method can be applied to problems involving multiple wells or sources. The method is also not limited to one constituent and can be used for a multitude of chemical components (simultaneously or independent of one another), as experienced in most hydrogeologic problems. A fully functioning computer program is available. If desired, the present Fortran 95 source code can be requested from the authors via email.

References

Bazaraa MS, Sherali H, Shetty C (1993) Nonlinear programming, 2nd edn. Wiley-Interscience, New York

Dominico PA, Schwartz FW (1998) Physical and chemical hydrogeology, 2nd edn. Wiley, New York

Fetter CW (2001) Applied hydrogeology, Prentice-Hall, Englewood Cliffs, New Jersey

Peressini AL, Sullivan FE, Uhl JJ Jr (1991) The mathematics of nonlinear programming. Springer, Heidelberg Berlin New York

DOI 10.1007/s10040-006-0106-4