*The Eleventh Symposium in GRA's*
*Series on Groundwater Contaminants*

# $ClO_4^-$ • 2004

# PERCHLORATE IN CALIFORNIA'S GROUNDWATER

## August 4, 2004 – Glendale, CA

***Co-Sponsors:***

**ARCADIS G&M, INC.**
**DANIEL B. STEPHENS, INC.**
**GEOSYNTEC CONSULTANTS**
**KLEINFELDER, INC.**
**SANTA CLARA VALLEY WATER DISTRICT**

GROUNDWATER RESOURCES ASSOCIATION
OF CALIFORNIA

Groundwater Resources Association
915 L Street , Suite 1000
Sacramento, CA 95814
Phone: (916) 446-3626
Fax: (916) 442-0382
www.grac.org

**Plume Source Identification Using Chemical Ratios and Convex Hull Theory to Measure "Closeness" of Fingerprint Matching**

**Theodore V. Hromadka II[1] and Robert J. Whitley[2]**

A common situation in Hydrogeology is trying to ascertain which of several possible sources is most likely responsible for observed conditions at a given target well. Typically a set of chemicals has been measured at the target well, and similar measurements have been made at one or more wells which represent the background state of unpolluted water, and have been made at one or more wells representing one of the possible sources of pollution. Any of these measurements at a given well, at a given time, can be represented as a vector listing the amounts, in appropriate units, of each of $m$ measured chemicals.

$$V = (v_1, v_2, \ldots, v_m) \tag{1}$$

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 $B^1,\ldots,B^{n_1}$, and the pollution source represented by the vectors $P^1,\ldots,P^{n_2}$. Simplify the notation by letting $V^1,\ldots,V^n$, 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,

$$W = \sum_{k=1}^{n} a_k V^k, \tag{2}$$

where $a_k$ represents the portion of $V^k$ in the mix, and

$$a_k \geq 0, \quad for\ k = 1,\ldots,n \tag{3}$$

and

$$\sum_{k=1}^{n} a_k = 1. \tag{4}$$

This situation can be described in terms of the notion of a convex set, (Peressini et al. 1991 or Bazarra et al. 1993). A subset $S$ in $R^m$ is convex if given the vectors $s$ and $t$ in $S$, and any real number $r$ in $[0,1]$, $rt + (1-r)s$ also belongs to $S$, i.e. the line joining any two vectors in $S$ also lies entirely in $S$. Given a subset $T$ of $R^m$, 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. Therefore, if $T$ is the set

$\{V^1, V^2, \ldots, V^m\}$, then the convex hull of $T$ is exactly the set of all convex combinations of the vectors in $T$. That is, $W$ is a mixture of the vectors $V^1, V^2, \ldots, V^m$ if and only if $W$ belongs to the convex hull of these vectors.

Writing the vectors $\{V^1, V^2, \ldots, V^m\}$ in terms of the $B$ and $P$ vectors,

$$W = \sum_{k=1}^{n_1} a_k B^k + \sum_{k=n_1+1}^{n_2} a_k P^k. \tag{5}$$

Letting $G = \sum_{k=1}^{n_1} a_k$,

$$W = G \sum_{k=1}^{n_1} \frac{a_k}{G} B^k + (1-G) \sum_{k=n_1+1}^{n_2} \frac{a_k}{1-G} P^k. \tag{6}$$

The term multiplied by $G$ is in the convex hull of $B^1, \ldots, B^{n_1}$ while the term multiplied by $(1-G)$ is in the convex hull of $P^1, \ldots, P^{n_2}$, and $W$ is written as a mixture of a background mixture term and a pollution mixture term.

Consider a subset of one coordinate, corresponding to one chemical of interest. 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 $W_1$ lies in the interval with left end point 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 $(b_1^i, b_2^i), i = 1, \ldots, n_1$ and $(p_1^i, p_2^i), i = 1, \ldots, n_2$ 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 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 $R^m$.

The problem of finding out whether or not $W$ belongs to the convex hull $K$ of the vectors $V^1, V^2, \ldots, V^n$ requires, in general, using the full set of coordinates. This can be done as follows: let $A$ be the $m \times n$ matrix. For any point $a_1 V^1 + a_2 V^2 + \ldots + a_n V^n$ in $K$, if $x$ is the vector in $R^n$ with $x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n$, then

$$Ax = a_1 V^1 + a_2 V^2 + \ldots + a_n V^n. \tag{7}$$

Let $(x,y)$ be the inner product of $x$ with $y$ in $R^m$. The square of the distance of $W$ to the point $Ax$ in $K$ is:

$$dist(Ax,W)^2 = (Ax-W, Ax-W) = (Ax, Ax) - 2(W, Ax) + (W,W) \quad (8)$$
$$= (x, A^T Ax) - 2(x, A^T W) + (W,W). \quad (9)$$

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

$$(x, A^T Ax) - 2(x, A^T W) \quad (10)$$

over all $x$ constrained by $x_1 + x_2 + \ldots + x_n = 1$ and all $x_i \geq 0$. As $A^T A$ is a non-negative definite matrix, this is a quadratic programming problem [1,2]. If $x$ is a solution of this minimization, then the distance from $Ax$ 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 $Ax$.

The problem of minimizing a linear function $g(x) = c_1x_1 + c_2x_2 + \ldots + c_nx_n$ over all $x$ constrained by $x_1 + x_2 + \ldots + x_n = 1$ 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, but still solvable using various properties of the functions of the type (13).

Example problem

Consider two measurements $W_1$ and $W_2$ taken at a target well indicating a certain amount of pollution in the levels of six measurements of NA, MG, CA, CL, SO4, HCO3. (In the actual problem there were not two, but more than 30 such measurements.) The background (unpolluted) levels of these measurements are represented by two measurements $B_1$ and $B_2$. (Again, in the actual problem, there were more than 30 measurements in the first set which has been averaged to give $B_1$, and more than 20 in the set which has been averaged to give $B_2$.) There are also measurements taken at two possible sources for the pollution seen at the target well. The first possible source $X$ is represented by two measurements $X_1$ and $X_2$, and the second possible source $Y$ by four measurements $Y_1, Y_2, Y_3$ and $Y_4$. Note that using more points for $Y$ generally reduces the distance. The distance of the target well from a mixture of the source plus the background indicated how well the observed pollution can be approximated by such a mixture. This distance 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 be compared. The distances shown in the table indicate that between the two sources, $X$ is more likely to be the source of pollution than $Y$.

## Conclusions

The "distance" between a set of chemical ratio fingerprints and another set of fingerprints can be computed and used as a measure of closeness between chemical ratio trends of the two sets of fingerprints. Such a measure of closeness can then be used to quantitatively evaluate between possible sources of pollution. Additionally, this measure can be used as an estimate of contribution strength based upon chemical ratio trends.

## References

[1] Peressini, Al, Sullivan, FE and Uhl, JJ Jr. (1991). The Mathematics of Non-linear Programming, Springer-Verlag, New York.

[2] Bazaraa, MS, Sherali, H and Shetty, C (1993). Nonlinear Programming, 2nd edition, Wiley-Interscience, New York.