[2]

IDENTIFICATION OF WATER SOURCES USING NORMALIZED CHEMICAL ION BALANCES: A LABORATORY TEST

DAVID A. WOOLHISER, WILLIAM E. EMMERICH and EDWARD D. SHIRLEY

Southwest Rangeland Watershed Research Center, U.S.D.A., A.R.S., Tucson, AZ 85719 (U.S.A.)

ABSTRACT

Woolhiser, D.A., Emmerich, W.E. and Shirley, E.D., 1985. Identification of water sources using normalized chemical ion balances: a laboratory test. J. Hydrol., 76: 205—231.

A normalized chemical ion balance (NCIB) technique, that has been proposed to identify multiple surface and groundwater inflows to a reach of a stream, is tested using two sets of mixtures prepared in the laboratory. The first set consisted of 64 mixtures of from 2—8 natural waters selected from 10 source waters. A second set of 25 mixtures was prepared using 5 source waters. Each of the source waters and each mixture was analyzed for at least 8 ions. Proportions of source waters in each sample were estimated using a quadratic programming solution to minimize the sum of squared error terms of normalized chemical ion-balance equations. Because PO_4^{3-} was present in very low concentrations, and was subject to significant error, it was found that better estimates could be obtained by omitting it from the analyses. For the first mixture set, the probability of failure to detect a source when it was present in the mixture was 0.23, and the probability of identifying a source when it was not present was 0.29. For the second set of mixtures, the probabilities were 0.0 and 0.33, respectively. Identification was improved by dividing each normalized ion-balance equation by the error variance for that ion.

1. INTRODUCTION

To design optimal strategies to control surface or groundwater pollution from point and nonpoint sources, authorities need models that relate observed concentrations of dissolved inorganic or organic species in the stream or in the groundwater to the sources of these species. Nonpoint pollution models, based on the physical, biological and chemical processes involved in the transport of various chemicals from source areas to streams or groundwater, have been developed in the past decade (cf. Knisel, 1980). However, these models are virtually impossible to verify for any but the smallest single-use watersheds because of the diffuse nature of much of the inflow to the stream. Because of large uncertainties in the source-based models, there is a need for techniques which utilize the chemical characteristics of surface or

groundwater to identify contributions from various sources, and thus can serve as a check of model consistency.

The water present in a reach of stream at any time is generally considered to be a mixture of waters that have followed different pathways to the stream. During a surface runoff event, for example, water in the stream might include direct rainfall, runoff from rainfall on saturated areas near the stream, classical Hortonian overland flow, interflow and baseflow. The chemical characteristics of each water source depend upon the chemical quality of the rainfall, the physical and chemical characteristics of vegetation and plant litter, and the physical and chemical composition of the geologic material that it has come in contact with as it moved through the watershed. Although the true number of pathways is very large, we can usually select a rather small number of sources that have unique absolute and relative concentrations of principal and minor anions and cations. Furthermore, it is often assumed that the stream water (or groundwater) is a mixture of solutions derived from different origins with no loss of dissolved species occurring after mixing.

In groundwater studies, the Piper (1944) trilinear diagram has been widely used to graphically test for apparent mixtures of waters from different sources (Morris et al., 1983). More recently, numerical models of chemical transport in groundwater flow have been developed to generate chemical patterns and assist in interpreting field data (Schwartz and Domenico, 1973).

The substantial differences between the dissolved solids content (or electrical conductivity) of surface runoff and groundwater (baseflow) have long been used in a mass-balance approach to estimate the proportion of stream discharge contributed from each source (Pinder and Jones, 1969; Hall, 1970; Visocky, 1970). The same technique has been used with some of the environmental isotopes to differentiate surface and groundwater sources (Sklash et al., 1976). This approach is adequate when only two sources of inflow are considered, but fails when there are more than two sources. Schwartz (1980) apparently used a trial-and-error ion-balance technique to estimate the quantities of surface water contributed from three sources: groundwater, direct precipitation and drainage from a muskeg.

Tsurumi (1982) utilized the multiple ion-balance approach to estimate not only the relative proportions of sources present in a mixture, but also the ranges of chemical composition of the sources of chemical constituents. His method requires several mixtures of the source waters at different proportions and uses an iterative least-squares solution. The source compositions are subject to an anion—cation balance constraint.

Chemical element balances (CEB's) have also been used to identify sources of particulates for many elements which can be associated with specific types of air-pollution sources (Kowalczyk et al., 1982). According to the CEB model, the composition of particles at a receptor is a linear combination of concentration patterns of particles from contributing sources. Kowalczyk et al. (1982) determined source-strength coefficients by a least-squares fit to the observed concentrations of several "marker elements".

Woolhiser et al. (1979, 1982) developed a technique to estimate multiple inflows to a stream reach based on a quadratic programming solution to find the unknown positive inflow quantities that minimized the sum of squares of normalized errors for up to eight ion-balance equations. They used Monte Carlo techniques to investigate the sensitivity of the solution to errors in the chemical analysis and errors in measurements of streamflow. This analysis showed conclusively that the use of the normalized ion-balance equations led to more accurate solutions than those obtained by minimizing the sum of squared errors in the original ion-balance equations. The quadratic programming algorithm constrained all inputs to be equal to or greater than zero.

Gorelick et al. (1983) proposed a method of identifying groundwater pollution sources that includes a linear programming approach to the minimization of errors in ion-balance equations and can accommodate both nonconservative tracers and transient cases.

The technique presented by Woolhiser et al. (1979, 1982) may have practical applications in detecting sources of surface water or groundwater pollution when several potential sources are so located that it should be physically possible for a pollutant to have traveled to the point in question and where each potential source uses the suspected pollutant, but has different relative proportions of other constituents. If water-quality samples are obtained at regular intervals, it may also be possible to estimate the contribution of each source to the annual runoff.

The key assumption in all of these models, except those that include numerical models of chemical transport (Schwartz and Domenico, 1973; Gorelick et al., 1983) is that all elements or dissolved ionic species considered are conservative within the system under investigation. There must be no chemical exchange, deposition, solution, gaseous transport and no uptake by organisms. Woolhiser et al. (1982) listed the following errors that can affect the accuracy of the calculated inflows:

(1) Errors in the chemical analyses (including sampling and storage errors).

(2) Errors in determining the discharge rates at the upper and lower ends of the stream reach.

(3) Nonrepresentative samples of the inflow waters.

(4) Omission of a significant inflow.

They also pointed out that, if the chemical characteristics of two or more inflow sources are nearly identical, it should be very difficult to distinguish between them.

In their empirical sensitivity analysis, they examined the implications of error categories (1), (2) and (4), above, using "theoretical mixtures" perturbed by adding normally distributed, zero mean error terms. The "theoretical mixtures" were obtained by using ion-balance equations to calculate the concentrations of each ion in a mixture of 3—7 water sources sampled near a surface coal mine site in western Colorado, U.S.A. They concluded that the technique appears to be promising as a method of estimating surface and groundwater inflows from several sources in a reach of a stream, but that further testing is required to determine the limitations.

If the normalized chemical ion-balance technique is to be used to identify possible pollution sources, it would be desirable to know something about the probability of making incorrect decisions (i.e. the probability of identifying a source if it is not present, or not identifying a source if it is present) under controlled laboratory conditions. One might infer that, under field conditions, the identification ability would be somewhat worse.

In this paper, we examine the ability of the technique proposed by Woolhiser et al. (1982) to identify the components of laboratory mixtures of waters and the implications of potential interference between sources with small differences in chemical characteristics. The effect of different weighting factors for each ion is also briefly examined.

2. MATERIALS AND METHODS

The first set of 64 mixtures prepared in the laboratory included from 2 to 8 source waters selected from a set of 10 source waters. The 10 source waters included 8 samples that had been collected as part of water-quality studies of the Agricultural Research Service on the Walnut Gulch Experimental Watershed, near Tombstone, Arizona, U.S.A. The other two samples were from the Tucson and Tombstone, Arizona, municipal supplies. A description of the source waters is given in Table I, and the measured ion concentrations

TABLE I

Description of source waters

Source No.	Description
First set:	
1	groundwater sample from Walnut Gulch Experimental Watershed L.G. Ranch well
2	groundwater sample from Walnut Gulch Experimental Watershed well *63*
3	groundwater sample from Walnut Gulch Experimental Watershed well *77*
4	groundwater sample from Walnut Gulch Experimental Watershed well *91*
5	sample from shallow well in alluvium of Walnut Gulch; alluvium saturated by sewage plant effluent
6	surface runoff from main stem of Walnut Gulch
7	surface runoff from main stem of Walnut Gulch
8	rain-water sample, Walnut Gulch Experimental Watershed
9	Tucson, Arizona, municipal water supply (groundwater)
10	Tombstone, Arizona, municipal water supply (groundwater)
Second set:	
1 + 6	Tombstone, Arizona, municipal water supply (groundwater)
2 + 7	groundwater sample from Walnut Gulch Experimental Watershed well *75*
3 + 8	rain-water sample, Walnut Gulch Experimental Watershed
4 + 9	sample from shallow well in alluvium of Walnut Gulch; alluvium saturated by sewage plant effluent
5 + 10	Tucson, Arizona, municipal water supply (groundwater)

TABLE II

Chemical characteristics of first and second set of source samples

Source No.	Ion concentration[*1]									Ion[*2] balance $(meq \, l^{-1})$
	PO_4^{3-} $(\mu g \, l^{-1})$	Cl^-	NO_3^-	SO_4^{2-}	HCO_3^-	Ca^{2+}	Mg^{2+}	Na^+	K^+	
First set:										
1	0	5.2	6.4	3.5	—[*3]	29.2	10.2	30.5	2.3	—
2	0	11.4	0.0	8.5	—	19.5	6.5	10.5	3.2	—
3	0	4.2	0.0	7.2	—	21.6	9.3	20.5	3.6	—
4	0	1.3	2.8	9.8	—	29.1	5.2	8.6	3.7	—
5	12,680	68.8	32.5	57.3	—	68.7	10.8	63.1	13.3	—
6	6	2.5	1.6	9.5	72.0	24.7	1.3	2.1	4.5	0.13
7	63	6.1	0.0	5.2	35.1	12.2	0.6	2.8	4.4	0.08
8	27	1.4	0.5	10.2	—	3.6	0.5	0.3	0.2	—
9	5	32.5	3.1	151.5	—	73.7	14.1	65.6	2.7	—
10	0	3.0	3.2	6.9	—	43.0	17.6	9.4	2.5	—
Second set:										
1	0	45.0	10.8	43.5	244.0	76.3	20.1	32.3	1.7	0.56
2	0	4.4	4.2	14.0	164.7	44.9	8.8	7.7	1.8	0.16
3	0	1.4	0.0	1.8	6.1	1.3	0.4	0.1	0.0	− 0.08
4	3,700	48.0	3.2	39.8	67.1	57.1	7.6	41.9	6.5	2.01
5	0	8.0	3.8	28.4	146.4	42.2	4.6	22.9	0.6	0.19
6	0	42.0	10.6	41.6	244.0	76.0	20.1	31.9	1.4	0.64
7	55	4.4	4.2	13.6	164.7	44.9	8.8	7.8	1.8	0.17
8	5	1.4	0.0	1.3	6.1	1.3	0.4	0.0	0.0	− 0.07
9	3,650	52.0	3.0	38.0	183.0	56.4	7.5	41.4	6.5	− 0.02
10	185	9.0	3.7	27.2	176.9	42.7	4.7	23.3	0.6	− 0.24

[*1] $mg \, l^{-1}$ unless otherwise indicated.
[*2] milli-equivalents cations minus anions.
[*3] — = not analyzed.

are shown in Table II. Surface water samples were centrifuged to remove the sediment, and the other samples were not treated before use. 10-ml mixtures were prepared by combining the sources in the proportions given in Table III. Each proportion used to make the mixture was measured to the nearest 0.01 ml.

A second set of 25 mixtures was prepared using 5 different source waters similar to the first 10, except they were collected at a later date. Mixtures in the second set were 100 ml in volume, and were prepared by combining sources in the proportions given in Table III. Sources *1, 2, 3, 4* and *5* are the same as *6, 7, 8, 9* and *10*, respectively, except they were filtered through a 0.45-μm filter before they were used to make the mixtures. The first five sources were not treated in any way, and all proportions were measured to 0.01 ml. During the preparation and analysis of the second set of mixtures, extreme care was taken at each step to minimize error.

TABLE III

Actual and calculated proportions of first and second set of source waters in laboratory mixtures along with an error index

Mixture No.		Source No.										Error index
		1	2	3	4	5	6	7	8	9	10	
First set:												
1	A[*1]	0.25	0.10	0	0	0	0.25	0.40	0	0	0	
	C1[*2]	0.00	0.30	0.02	0.68	0	0	0	0	0	0	0.10
	C2[*3]	0.13	0	0.04	0.25	0	0	0.51	0.04	0.00	0.04	0.525
2	A	0.50	0	0	0	0	0	0.50	0	0	0	
	C1	0.42	0.05	0.11	0.15	0	0.29	0	0	0	0	0.40
	C2	0.36	0	0	0	0	0	0.47	0.08	0	0.10	0.825
3	A	0.50	0	0	0	0	0	0	0.50	0	0	
	C1	0.36	0.17	0	0.47	0	0	0	0	0	0	0.36
	C2	0.39	0	0	0	0	0	0.05	0.50	0	0.06	0.89
4	A	0.50	0	0	0	0	0	0	0	0.50	0	
	C1	0.73	0.19	0	0	0	0	0	0	0	0.08	0.50
	C2	0.48	0	0	0	0	0	0	0.06	0.47	0	0.945
5	A	0.50	0	0	0	0	0	0	0	0	0.50	
	C1	no convergence[*4]										
	C2	0.39	0	0	0	0.01	0	0	0.12	0	0.49	0.88
6	A	0	0.50	0	0	0	0.50	0	0	0	0	
	C1	0	0.98	0	0.02	0	0	0	0	0	0	0.50
	C2	0	0.44	0.03	0	0	0	0.41	0.11	0	0	0.45
7	A	0	0.50	0	0	0	0	0.50	0	0	0	
	C1	0	0.67	0	0	0	0	0.33	0	0	0	0.83
	C2	0	0.28	0.03	0	0	0	0.67	0.02	0	0	0.78
8	A	0	0.50	0	0	0	0	0	0.50	0	0	
	C1	0	0.97	0	0.03	0	0	0	0	0	0	0.50
	C2	0	0.43	0	0	0	0	0.33	0.24	0	0	0.67
9	A	0	0.50	0	0	0	0	0	0	0.50	0	
	C1	0.21	0.64	0.05	0	0	0	0	0	0	0.10	0.50
	C2	0	0.47	0	0	0	0	0	0.05	0.45	0.02	0.93
10	A	0	0.50	0	0	0	0	0	0	0	0.50	
	C1	0	0.48	0	0.10	0	0	0	0	0	0.41	0.89
	C2	0	0.44	0	0	0	0.05	0	0.07	0	0.43	0.88
11	A	0	0	0.50	0	0	0.50	0	0	0	0	
	C1	0	0.66	0.34	0	0	0	0	0	0	0	0.34
	C2	0	0.10	0.43	0	0	0	0.47	0	0	0	0.43
12	A	0	0	0.35	0	0	0	0.25	0	0.40	0	
	C1	0.04	0.52	0.43	0	0	0	0	0	0	0.01	0.35
	C2	0	0.40	0.49	0	0	0	0	0	0.10	0	0.45
13	A	0	0	0.50	0	0	0	0	0	0.50	0	
	C1	0.17	0.66	0	0	0	0	0	0	0	0.17	0.00
	C2	0	0	0.39	0	0	0	0	0.13	0.47	0.01	0.86

TABLE III (*continued*)

Mixture No.		Source No.										Error index
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
First set (cont.):												
14	A	0	0	0.50	0	0	0	0	0	0	0.50	
	C1	0	0.03	0.47	0.18	0	0	0	0	0	0.31	0.78
	C2	0.02	0	0.44	0.01	0	0	0	0.16	0.01	0.37	0.81
15	A	0	0	0	0.50	0	0.50	0	0	0	0	
	C1	0	0.02	0	0.98	0	0	0	0	0	0	0.50
	C2	0	0	0	0.42	0	0.33	0	0.24	0	0.01	0.75
16	A	0	0	0	0.50	0	0	0.50	0	0	0	
	C1	0	0.07	0	0.93	0	0	0	0	0	0	0.50
	C2	no convergence										
17	A	0	0	0	0.50	0	0	0	0.50	0	0	
	C1	0	0.06	0	0.94	0	0	0	0	0	0	0.50
	C2	0	0	0.05	0.26	0	0.02	0	0.66	0	0	0.76
18	A	0	0	0.10	0.25	0	0	0	0.40	0.25	0	
	C1	0.27	0.70	0	0.03	0	0	0	0	0	0	0.03
	C2	0	0	0	0	0.01	0.12	0	0.57	0.22	0.06	0.63
19	A	0	0	0	0.50	0	0	0	0	0	0.50	
	C1	0	0	0	0.55	0	0	0	0	0	0.45	0.95
	C2	0.03	0	0	0.42	0	0	0	0.12	0	0.43	0.85
20	A	0.025	0.05	0.125	0	0.125	0.125	0.225	0.125	0	0.20	
	C1	0	0	0.17	0	0.12	0.14	0.21	0.17	0	0.18	0.89
	C2	0	0.19	0.13	0.07	0.11	0.26	0	0.13	0	0.12	0.65
21	A	0	0	0	0	0.50	0	0.50	0	0	0	
	C1	0	0	0	0	0.45	0.08	0.39	0.06	0.02	0	0.84
	C2	0.01	0	0	0	0.42	0.11	0.42	0	0.04	0	0.84
22	A	0	0	0	0	0.50	0	0	0.50	0	0	
	C1	0	0	0	0	0.40	0	0.13	0.46	0.01	0	0.86
	C2	0	0	0	0	0.37	0	0.23	0.37	0.03	0	0.74
23	A	0	0	0	0	0.50	0	0	0	0.50	0	
	C1	0	0	0	0	0.44	0	0	0.11	0.45	0	0.89
	C2	0	0	0	0	0.40	0	0.10	0.02	0.48	0	0.88
24	A	0	0	0.033	0.067	0.167	0	0.30	0.267	0.00	0.167	
	C1	0	0	0	0	0.16	0.24	0.11	0.31	0.01	0.18	0.44
	C2	0	0	0	0	0.15	0.24	0.14	0.28	0.01	0.18	0.59
25	A	0.10	0	0.05	0	0	0.40	0	0.45	0	0	
	C1	0	0.15	0	0.85	0	0	0	0	0	0	0.00
	C2	0.05	0	0.08	0	0	0.33	0.02	0.51	0	0	0.89
26	A	0.20	0	0	0	0	0	0.80	0	0	0	
	C1	no convergence										
	C2	no convergence										
27	A	0.20	0	0	0	0	0	0	0.80	0	0	
	C1	0.11	0.28	0	0.61	0	0	0	0	0	0	0.11
	C2	0.18	0	0	0	0	0	0	0.82	0	0	0.98

TABLE III (*continued*)

Mixture No.		Source No.										Error index
		1	2	3	4	5	6	7	8	9	10	
First set (cont.):												
28	A	0.20	0	0	0	0	0	0	0	0.80	0	
	C1	0.53	0.32	0	0	0	0	0	0	0	0.14	0.20
	C2	0.17	0	0	0.15	0	0	0	0	0.51	0.17	0.68
29	A	0.10	0.05	0	0	0	0	0	0	0	0.85	
	C1	0.05	0.05	0	0.08	0	0	0	0	0	0.82	0.92
	C2	0.06	0	0	0	0.01	0.02	0.01	0.08	0	0.82	0.88
30	A	0	0.20	0	0	0	0.80	0	0	0	0	
	C1	0	0.77	0	0.23	0	0	0	0	0	0	0.20
	C2	0	0	0.06	0	0	0.08	0.59	0.22	0	0.04	0.08
31	A	0	0.20	0	0	0	0	0	0.80	0	0	
	C1	0	0.70	0	0.30	0	0	0	0	0	0	0.20
	C2	0	0.17	0	0	0	0	0.02	0.81	0	0	0.97
32	A	0	0.20	0	0	0	0	0	0	0.80	0	
	C1	0.36	0.11	0.45	0	0	0	0	0	0	0.08	0.11
	C2	0.01	0	0.11	0.34	0	0	0	0.10	0.18	0.26	0.18
33	A	0	0	0.20	0	0	0.80	0	0	0	0	
	C1	0	0.29	0.56	0.15	0	0	0	0	0	0	0.20
	C2	0	0	0.26	0	0	0.01	0.24	0.50	0	0	0.21
34	A	0	0	0.20	0	0	0	0.80	0	0	0	
	C1	0	0	0.08	0	0	0	0.79	0.13	0	0	0.87
	C2	0	0	0.07	0	0	0	0.81	0.12	0	0	0.87
35	A	0	0	0.20	0	0	0	0	0	0	0.80	
	C1	0	0.04	0.14	0.08	0	0	0	0	0	0.74	0.88
	C2	0.03	0.01	0.12	0	0	0	0.02	0.13	0	0.68	0.81
36	A	0	0	0	0.20	0	0.80	0	0	0	0	
	C1	0	0.24	0	0.76	0	0	0	0	0	0	0.20
	C2	0	0	0.11	0	0	0.43	0.08	0.38	0	0	0.43
37	A	0	0	0	0.20	0	0	0.80	0	0	0	
	C1	0	0	0.09	0	0	0.09	0	0.83	0	0	0.00
	C2	0	0	0.07	0	0	0	0.07	0.85	0	0	0.07
38	A	0	0	0	0.10	0.05	0	0	0	0.40	0.45	
	C1	0.22	0.15	0	0	0.02	0.08	0	0	0.41	0.12	0.54
	C2	0.09	0.26	0	0	0.05		0	0.04	0.38	0.17	0.60
39	A	0	0	0	0.20	0	0	0	0	0	0.80	
	C1	0	0	0	0.32	0	0	0	0	0	0.68	0.88
	C2	0	0	0	0.24	0	0	0	0.07	0	0.68	0.88
40	A	0	0	0	0	0.20	0.80	0	0	0	0	
	C1	0	0.02	0	0	0.13	0.23	0.48	0	0.03	0	0.36
	C2	0	0.02	0	0	0.12	0.28	0.47	0.08	0.03	0	0.40
41	A	0	0	0	0	0.20	0	0.80	0	0	0	
	C1	0	0	0.01	0	0.15	0.35	0.31	0.17	0	0	0.46
	C2	0.02	0	0	0	0.14	0.38	0.32	0.13	0	0	0.46

TABLE III (*continued*)

Mixture No.		Source No.										Error index
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
First set (cont.):												
42	A	0	0	0	0	0.20	0	0	0.80	0	0	
	C1	0	0	0	0	0.08	0	0.30	0.58	0.04	0	0.66
	C2	0	0	0	0	0.08	0	0.31	0.57	0.04	0	0.65
43	A	0	0	0	0	0.20	0	0	0	0.80	0	
	C1	0	0	0	0	0.09	0	0.19	0.05	0.67	0	0.76
	C2	0	0	0	0	0.10	0	0.16	0.09	0.65	0	0.75
44	A	0	0	0	0	0.20	0	0	0	0	0.80	
	C1	0	0.34	0	0	0.12	0.03	0.08	0.12	0.02	0.28	0.40
	C2	0	0.31	0	0	0.13	0	0.12	0.14	0.02	0.29	0.42
45	A	0.10	0	0	0	0	0.90	0	0	0	0	
	C1	1.00	0	0	0	0	0	0	0	0	0	0.10
	C2	0	0	0.07	0	0	0.57	0.13	0.22	0	0.01	0.57
46	A	0.10	0	0	0	0	0	0	0.90	0	0	
	C1	0	0.27	0	0.73	0	0	0	0	0	0	0.00
	C2	0.07	0	0	0.01	0	0	0	0.91	0	0	0.97
47	A	0.10	0	0	0	0	0	0	0	0.90	0	
	C1	0.42	0.55	0	0	0	0	0	0	0	0.03	0.10
	C2	0	0	0	0	0.01	0	0.11	0.20	0.69	0	0.69
48	A	0.10	0	0	0	0	0	0	0	0	0.90	
	C1	0.03	0.02	0	0.14	0	0	0	0	0	0.81	0.84
	C2	0.06	0	0	0	0	0.05	0	0.09	0	0.79	0.85
49	A	0	0.10	0	0	0	0.90	0	0	0	0	
	C1	0	1.00	0	0	0	0	0	0	0	0	0.10
	C2	0	0	0.01	0	0	0	0.97	0.02	0	0	0.00
50	A	0	0.10	0	0	0	0	0.90	0	0	0	
	C1	0	0	0	0	0	0	1.00	0	0	0	0.90
	C2	0	0	0	0	0	0	0.98	0.02	0	0	0.90
51	A	0	0.10	0	0	0	0	0	0.90	0	0	
	C1	0	0.09	0	0.01	0	0.04	0	0.86	0	0	0.95
	C2	0	0.06	0	0	0	0	0.02	0.91	0	0.01	0.96
52	A	0	0.10	0	0	0	0	0	0	0.90	0	
	C1	no convergence										
	C2	0	0	0	0	0	0	0.07	0.15	0.77	0	0.77
53	A	0	0	0.10	0	0	0.90	0	0	0	0	
	C1	0	0.93	0	0.06	0	0	0	0	0	0	0.00
	C2	0	0	0.10	0	0	0	0.63	0.27	0	0	0.10
54	A	0	0	0.10	0	0	0	0	0	0.90	0	
	C1	no convergence										
	C2	0	0	0	0	0	0	0.05	0.09	0.86	0	0.86
55	A	0	0	0.10	0	0	0	0	0	0	0.90	
	C1	0	0.03	0	0.20	0	0	0	0	0	0.77	0.77
	C2	0.05	0	0	0.05	0	0	0.01	0.15	0	0.74	0.74

TABLE III (*continued*)

Mixture No.		Source No.										Error index
		1	2	3	4	5	6	7	8	9	10	
First set (cont.):												
56	A	0	0	0	0.10	0	0.90	0	0	0	0	
	C1	0	0.33	0	0.67	0	0	0	0	0	0	0.10
	C2	0	0	0.06	0	0	0.27	0.18	0.48	0	0	0.27
57	A	0	0	0	0.10	0	0	0.90	0	0	0	
	C1	0	0	0.01	0	0	0	0.29	0.69	0	0	0.29
	C2	no convergence										
58	A	0	0	0	0.10	0	0	0	0.90	0	0	
	C1	0	0.07	0	0	0	0.07	0	0.86	0	0	0.86
	C2	0	0.03	0.01	0	0	0.01	0.02	0.92	0	0	0.90
59	A	0	0	0	0.10	0	0	0	0	0.90	0	
	C1	0.37	0.40	0	0	0	0	0	0	0	0.23	0.00
	C2	0	0.04	0	0	0.01	0	0.02	0.14	0.79	0	0.79
60	A	0	0	0	0.10	0	0	0	0	0	0.90	
	C1	0	0.05	0	0.35	0	0	0	0	0	0.60	0.70
	C2	0.02	0	0	0	0	0.07	0	0.27	0	0.63	0.63
61	A	0	0	0	0	0.10	0.90	0	0	0	0	
	C1	0	0	0	0	0.02	0.10	0.84	0	0.03	0.01	0.12
	C2	0	0.02	0	0	0.02	0.18	0.75	0	0.03	0	0.20
62	A	0	0	0	0	0.10	0	0.90	0	0	0	
	C1	0	0	0	0	0.04	0.06	0.72	0.16	0	0.01	0.76
	C2	0	0	0	0	0.03	0.13	0.68	0.14	0.01	0.01	0.71
63	A	0	0	0	0	0.10	0	0	0.90	0	0	
	C1	0	0	0	0	0.07	0	0.06	0.86	0.01	0	0.93
	C2	0	0	0	0	0.08	0	0.01	0.90	0	0	0.98
64	A	0	0	0	0	0.10	0	0	0	0.90	0	
	C1	0	0.28	0	0	0.02	0	0	0	0.70	0	0.72
	C2	0	0	0	0	0	0	0.14	0	0.86	0	0.86
Second set:												
1	A	0.5	0.5	0	0	0						
	C1	0.46	0.46	0	0	0.08						0.92
	C2	0.46	0.46	0	0	0.08						0.92
2	A	0.5	0	0.5	0	0						
	C1	0.42	0	0.43	0	0.15						0.93
	C2	0.42	0	0.43	0	0.15						0.93
3	A	0.05	0	0	0.5	0						
	C1	0.47	0	0	0.47	0.06						0.97
	C2	0.48	0	0	0.45	0.07						0.93
4	A	0.5	0	0	0	0.5						
	C1	0.45	0	0	0	0.55						0.95
	C2	0.45	0	0	0	0.55						0.95

TABLE III (*continued*)

Mixture No.		Source No.										Error index
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
Second set (cont.):												
5	A	0	0.5	0.5	0	0						
	C1	0	0.52	0.48	0	0						0.98
	C2	0	0.52	0.48	0	0						0.98
6	A	0	0.5	0	0.5	0						
	C1	0.01	0.46	0	0.50	0.03						0.96
	C2	0	0.46	0	0.50	0.03						0.97
7	A	0	0.5	0	0	0.5						
	C1	no convergence										
	C2	0.01	0.51	0	0	0.48						0.98
8	A	0	0	0.5	0.5	0						
	C1	0.01	0	0.45	0.49	0.05						0.94
	C2	0.01	0	0.45	0.48	0.06						0.93
9	A	0	0	0.5	0	0.5						
	C1	0	0.03	0.49	0	0.48						0.97
	C2	0	0.02	0.49	0	0.48						0.98
10	A	0	0	0	0.5	0.5						
	C1	0.01	0	0	0.51	0.48						0.98
	C2	0.01	0	0	0.51	0.48						0.98
11	A	0.25	0.25	0.25	0.25	0						
	C1	0.19	0.31	0.16	0.24	0.09						0.85
	C2	0.19	0.31	0.16	0.25	0.08						0.86
12	A	0.25	0.25	0.25	0	0.25						
	C1	0.21	0.26	0.20	0	0.33						0.91
	C2	0.21	0.26	0.20	0	0.33						0.91
13	A	0.25	0.25	0	0.25	0.25						
	C1	0.24	0.26	0	0.24	0.26						0.98
	C2	0.24	0.26	0	0.24	0.26						0.98
14	A	0.25	0	0.25	0.25	0.25						
	C1	0.21	0.06	0.19	0.24	0.30						0.89
	C2	0.22	0.07	0.18	0.23	0.31						0.88
15	A	0	0.25	0.25	0.25	0.25						
	C1	0	0.27	0.23	0.26	0.25						0.98
	C2	0	0.27	0.23	0.25	0.25						0.98
16	A						0.5	0.5	0	0	0	
	C1						0.5	0.5	0	0	0	1.00
	C2						0.5	0.5	0	0	0	1.00
17	A						0.5	0	0.5	0	0	
	C1						0.45	0.01	0.42	0	0.12	0.93
	C2						0.45	0.02	0.42	0	0.11	0.93
18	A						0.5	0	0	0.5	0	
	C1						0.46	0	0	0.50	0.04	0.96
	C2						0.48	0	0	0.47	0.05	0.95

TABLE III (*continued*)

Mixture No.		Source No.											Error index
		1	2	3	4	5	6	7	8	9	10		
Second set (cont.):													
19	A						0.5	0	0	0	0.5		
	C1						0.56	0.07	0.15	0	0.22		0.72
	C2						0.46	0.07	0	0	0.47		0.93
20	A						0	0.5	0.5	0	0		
	C1						0.05	0.27	0.69	0	0		0.77
	C2						0	0.50	0.50	0	0		1.00
21	A						0	0.5	0	0.5	0		
	C1						0	0.47	0.01	0.49	0.04		0.96
	C2						0	0.47	0	0.47	0.07		0.94
22	A						0	0.5	0	0	0.5		
	C1						0.25	0.05	0.71	0	0		0.05
	C2						0	0.51	0	0	0.49		0.99
23	A						0	0	0.5	0.5	0		
	C1						0	0.01	0.46	0.48	0.05		0.94
	C2						0	0.01	0.45	0.47	0.07		0.92
24	A						0	0	0.5	0	0.5		
	C1						0.14	0.02	0.83	0	0		0.51
	C2						0	0.02	0.51	0	0.48		0.98
25	A						0	0	0	0.5	0.5		
	C1						0.02	0.02	0	0.46	0.50		0.96
	C2						0.03	0.01	0	0.44	0.52		0.94

[1] Proportions present in laboratory mixture.
[2] Calculated using 8 ions.
[3] Calculated omitting PO_4^{3-}.
[4] Wolfe's (1959) algorithm did not converge after 400 iterations.

All source samples and mixtures were analyzed for eight different ions, and two of the sources in set *1*, and all sources and mixtures in set *2*, were analyzed for HCO_3^-. Concentrations of the cations Na^+, K^+, Mg^{2+} and Ca^{2+} were determined by atomic absorption spectrometry. A Technicon® Auto-analyzer Industrial II System* was used to determine the concentrations of Cl^-, NO_3^-, SO_4^{2-} and PO_4^{3-}. HCO_3^- ion was determined by potentiometric titration.

3. ANALYSES

3.1. Mathematical model

The normalized ion-balance equations can be written in the form:

---

*Citation of brand names is for the readers' information only, and does not imply endorsement by the U.S. Department of Agriculture.

$$\begin{vmatrix} C_{11}^* & C_{12}^* & \ldots & C_{1n}^* - 1 \\ C_{21}^* & C_{22}^* & \ldots & C_{2n}^* - 1 \\ \vdots & \vdots & \vdots & \vdots \\ C_{i1}^* & C_{i2}^* \ldots & C_{ij}^* \ldots C_{in}^* & -1 \\ \vdots & \vdots & \vdots & \vdots \\ C_{m1}^* & C_{m2}^* & \ldots & C_{mn}^* - 1 \end{vmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \\ 1 \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_m \end{pmatrix} \tag{1}$$

where C_{ij}^* is the concentration of the ith ion in the jth inflow divided by the concentration of the ith ion in the mixture; $x_j, j = 1, 2, \ldots, n$, is the decimal fraction of the jth inflow in the mixture; and ϵ_i, $i = 1, 2, \ldots, m$, is the normalized error term.

From mass balance:

$$\sum_{j=1}^{n} x_j = 1 \tag{2}$$

Eq. 2 can be solved for x_n and substituted into eq. 1, resulting in the expression:

$$\begin{vmatrix} C_{11} & C_{12} & \ldots & C_{1n-1} & C_{1n}^* - 1 \\ C_{21} & C_{22} & \ldots & C_{2n-1} & C_{2n}^* - 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{i1} & C_{i2} & C_{ij} & C_{in-1} & C_{in}^* - 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{m1} & C_{m2} & \ldots & C_{mn-1} & C_{mn}^* - 1 \end{vmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_{n-1} \\ 1 \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_m \end{pmatrix} \tag{3}$$

where $C_{ij} = C_{ij}^* - C_{in}^*$. Thus, we have m equations with $(n-1)$ unknowns where $m \geqslant n$ (usually). Because negative x_j's are not physically realistic (in most cases) we wish to have the $x_j \geqslant 0$. Under these circumstances, it is convenient to formulate this as a quadratic programming problem by finding the $x_j, j = 1, 2, \ldots, n - 1$ that minimize the objective function:

$$E = \sum_{i=1}^{m} (\epsilon_i)^2 \tag{4}$$

Squaring each equation in eq. 3, summing and neglecting the constant, we obtain a quadratic form:

$$E = D^{\mathrm{T}} \underline{X} + \underline{X}^{\mathrm{T}} A \underline{X} \tag{5}$$

where D is a vector of coefficients of the linear terms and $\underline{X}^{\mathrm{T}}$ indicates transpose. Thus, E is to be minimized subject to the nonnegativity constraint, $x_j \geqslant 0$; $j = 1, 2, \ldots, n - 1$. It should be noted that, in this form, x_n is not constrained to be $\geqslant 0$, so some care must be used in setting up the data so that x_n is the most likely source. Conditions for existence and uniqueness of solutions can be found in Hadley (1964). In our problem, the matrix A is positive semi-definite. Any local minimum will be a global minimum, but it might not be unique. Since the x_j's are unbounded, there might be no solution. In the latter case, a solution might be found by eliminating another variable. A will be positive definite when its determinant is non-zero, and in this case there is a unique solution.

A version of the Wolfe (1959) quadratic programming algorithm was used to find the best linear combination of source waters for each mixture. The complete source set was considered in the algorithm, and two runs were made for each mixture, one with 8 ions considered and a second with PO_4^{3-} omitted. In the second source and mixture set, mixtures *1—15* and *16—25* were from sources *1—5* and *6—10*, respectively, and were calculated on that basis. Because sources *1—5* and *6—10* were essentially the same except for filtration, all mixtures made from both source groups were considered part of the second mixture set unless otherwise noted. HCO_3^- was omitted from the analyses for both sets. The results of these analyses are shown in Table III.

It should be noted that in set *1*, the number of candidate sources (10) is greater than the number of ion- and water-balance equations (9 with PO_4^{3-} included and 8 without). It is apparent that, in some circumstances, this situation can lead to problems of nonuniqueness. To examine this factor, we constructed a theoretical mixture consisting of 0.8 parts of source *2* and 0.2 parts of source *4*. Two trials were run, the first with sources *1—9* as candidates, and the second with sources *1—10* as candidates. The algorithm converged to the correct answers in both cases. To examine the effect of errors on the quadratic matrix and algorithm convergence, independent, normally distributed error terms, with zero mean and coefficient of variation of 0.05 were added to the concentration of each ion in every candidate source and in the mixture. Ten of these simulated runs were examined for each case. For case *1* (9 candidates and 9 equations), the algorithm converged to correct answers in 7 runs, and did not converge in 400 iterations for 3 runs. For case *2* (10 candidates and 9 equations), the algorithm did not converge 3 times, and converged to the apparently correct answers 7 times. For each run in which the algorithm did not converge, the values of the x_j, after 400 iterations, were very close to the correct values. Although every effort should be made to reduce the number of candidate sources to fewer than the number of ions used in the analyses, it appears that useful results can be obtained if n is slightly larger than $(m + 1)$, provided that there are fewer than $(m + 1)$ sources actually present. For example, in mixture *20* in Table III, 8 sources were included in the

mixture; thus, when PO_4^{3-} was omitted from the analysis, there were as many equations as sources in the mixture. The algorithm identified 7 nonzero sources, and all errors, ϵ_i, were exactly zero. This solution is probably not unique. It is also possible that some of the other solutions for the first set of mixtures are not unique; however, it should be noted that in each case, the sum of squared error terms in eq. 1, calculated by using the x_j identified by the algorithm, was less than if the "correct" x_j were used. The question of uniqueness is explored in more detail in the Appendix.

3.2. Interactions between sources

To determine if a solution with the chemical characteristics of any of the sources could be approximated by a linear combination of any of the other sources, the analyses program was run with each source considered as a

TABLE IV

Best linear combinations of first and second set of sources

Mix-ture No.	Proportion of source in mixture										R^2	Inter-cept	Slope
	1	2	3	4	5	6	7	8	9	10			
First set:													
1	—	0.01	0.48	0	0	0	0	0	0	0.51	0.72	1.32	0.75[*3]
2	0	—	1.0	0	0	0	0	0	0	0	0.69	0.26	1.08[*2]
3	0	0.99	—	0.01	0	0	0	0	0	0	0.69	2.14	0.64[*2]
4	0	0	0.45	—	0	0	0	0	0	0.55	0.88	1.27	1.10[*4]
5	0.56	0	0	0	—	0	0	0	0.44	0	0.15	28.7[*1]	−0.002
6	0	0	0	0.46	0	—	0	0.54	0	0	0.59	2.61	0.58[*2]
7	0	0.96	0	0	0	0	—	0.04	0	0	0.07	8.28[*1]	−0.08
8	0	0	0	0	0	0.34	0.66	—	0	0	0.86	2.15	1.43[*4]
9	0.33	0.30	0	0	0	0	0	0	—	0.36	0.108	7.7	0.06
10	0.16	0.10	0	0.74	0	0	0	0	0	—	0.87	1.76	0.59[*4]
Second set:													
1	—	0	0	0.26	0.74						0.87	5.75	1.40[*3]
2	0.03	—	0.56	0	0.41						0.77	−0.75	1.72[*3]
3	no convergence												
4	0.15	0	0	—	0.85						0.73	7.74	1.16[*2]
5	0.21	0.18	0.61	0	—						0.82	0.92	1.77[*3]
6						—	0	0	0.19	0.81	0.87	5.72	1.43[*3]
7						0.03	—	0.56	0	0.41	0.76	−1.83	1.78[*2]
8	no convergence												
9						0.12	0	0	—	0.88	0.67	9.86	1.12[*2]
10						0.22	0.19	0.59	0	—	0.83	−0.75	1.70[*3]

[*1] Intercept significantly different from zero at 5% level.
[*2] Slope significantly different from zero at 5% level.
[*3] Slope significantly different from zero at 1% level.
[*4] Slope significantly different from zero at 0.1% level.

mixture and the other sources considered as candidates. The results of this analysis are shown in Table IV. This shows, for example, that source _1_ of the first set is best approximated (according to the objective function given by eq. 3) by a mixture of 0.01 parts source _2_, 0.48 parts source _3_, and 0.51 parts source _10_. Simple linear regressions were run between the concentration of each ion as calculated from the proportions shown in Table IV and the measured concentrations of each ion for each source. The coefficients of determination, R^2, the intercept, and the slope for these regressions are shown in the last three columns of Table IV.

An examination of the coefficients of determination, slopes and intercepts reveals several cases when one of the sources can be rather closely approximated by linear combinations of other sources. For example, source _4_ of the first set can be approximated by a combination of 0.45 parts source _3_ plus 0.55 parts source _10_ with $R^2 = 0.88$. The intercept is not significantly different from zero, and the slope is close to one. Therefore, one would expect that it would be difficult to identify these sources, or that they could be identified when they are actually not present.

4. RESULTS AND DISCUSSION

PO_4^{3-} ion was present in very low concentrations ($\mu g\, l^{-1}$ range) in most of the sources used to make the mixtures (Table II). Handling and analyzing samples with low concentrations presented analytical problems that became important in this study. The two problems that were observed, with PO_4^{3-} in particular, were contamination and dilution below the detection limit. Contamination was confirmed in the second set of source samples where PO_4^{3-} ion was found in sources _7_, _8_ and _10_ after filtration. In the first set of sources and mixtures, both contamination and dilution could account for the observed results of PO_4^{3-} ion in a source, but not found in a mixture containing that source. Contamination of the analysis sample used to analyze sources _6_, _8_ and _9_ were suspected in that, when mixtures containing these sources were combined with other sources that did not contain PO_4^{3-}, the resulting mixture contained PO_4^{3-} in only two cases. Contamination and dilution problems were not observed with the other ions, although they probably were present. At higher concentrations, small amounts of contamination are not detectable.

If the NCIB method is to be used simply as a detection technique to identify the presence or absence of a particular source in a mixture, it is useful to examine the probability of not detecting a source when it is present (type-I error) and the probability of detecting a source when it is not present (type-II error). The probabilities of type-I and -II errors for each source, and for all sources combined, are shown in Table V. Although these probabilities apply only to these particular sets of sources and are affected by the mixing strategy used, they do reveal some important points. In general the type-I error was reduced and -II was increased when PO_4^{3-} was omitted in

TABLE V

Probability of type-I and -II errors

Source No.	First set				Second set			
	type I		type II		type I		type II	
	all ions	PO_4^{3-} omitted	all ions	PO_4^{3-} omitted	all ions	PO_4^{3-} omitted	all ions	PO_4^{3-} omitted
1	0.29	0.21	0.14	0.18	0.0	0.0	0.43	0.43
2	0.20	0.47	0.57	0.18	0.13	0.0	0.29	0.29
3	0.50	0.21	0.16	0.24	0.0	0.0	0.0	0.0
4	0.40	0.73	0.39	0.14	0 0	0.0	0.0	0.0
5	0.0	0.07	0.0	0.10	0.13	0.0	0.86	0.86
6	0.80	0.33	0.20	0.22	0.0	0.0	0.70	0.17
7	0.36	0.29	0.16	0.64	0.0	0.0	0.83	0.83
8	0.47	0.0	0.18	0.82	0.0	0.0	0.50	0.0
9	0.69	0.0	0.17	0.17	0.0	0.0	0.0	0.0
10	0.07	0.0	0.20	0.26	0.50	0.0	0.70	0.70
All sources	0.38	0.23	0.20	0.29	0.08	0.0	0.43	0.33

the first mixture set, suggesting that including PO_4^{3-} has some merit in reducing incorrect detection, but this occurs at the expense of increasing the chances of failure to identify a water source that is, in fact, present. For the second set of mixtures, omitting PO_4^{3-} reduced both type-I and -II errors. Overall, the second set of mixtures had a much lower incidence of type-I errors than the first set, but had a larger number of type-II errors. The larger number of type-II errors was probably due to similarity of the sources, as indicated by Table IV. Although there were more type-II errors (Table III), the size of the errors was generally much smaller for the second set of mixtures. The type-I and -II errors are very small for source 5 of the first set and sources 4 and 9 of the second set. These sources are quite different from the others, as indicated by the data in Table II and the results of the linear combination test shown in Table IV. It should be noted that the presence or absence of sources such as these, which indicate water-quality changes due to man's activities, are of primary interest in pollution studies. The surface runoff samples, sources 6 and 7 in set 1, had rather large type-I and -II errors, probably because they are intermediate in composition between the rainfall sample and the Walnut Gulch groundwater samples. It is somewhat surprising that the errors were so low for water 10, which has a high simple correlation with a linear combination of sources 1, 2 and 4.

Another way of portraying results of this laboratory test of the NCIB technique is to construct an error index that accounts for differences in actual and predicted proportions of each source in each mixture. The following index was used for each mixture:

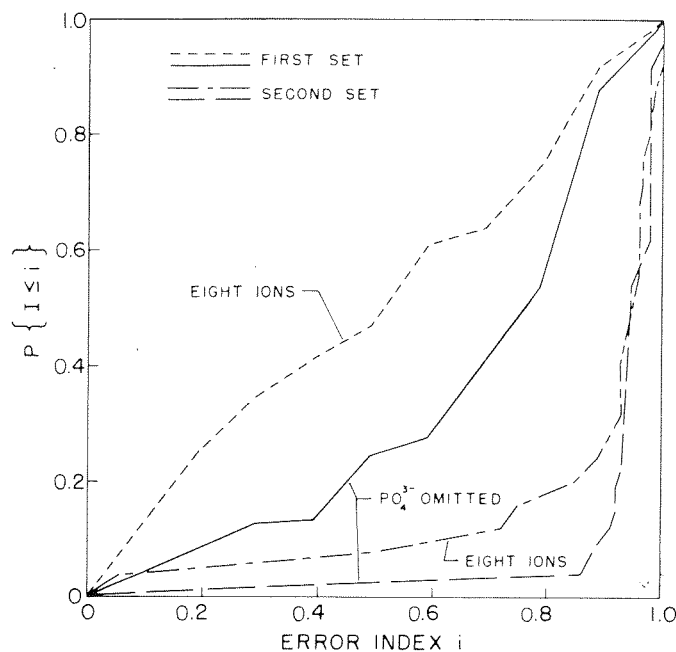$$I_k = 1 - 0.5 \sum_{j=1}^{N_s} |\hat{x}_{jk} - x_{jk}| \tag{6}$$

Fig. 1. Empirical distribution function of error index.

where x_{jk} is the proportion of mixture k made up of source j; $\dot{x}_{jk}$ is the estimated proportion; and N_s is the number of sources considered. If the agreement is perfect, $I_k = 1$. These error indices are tabulated in the last column of Table III, and the sample distribution functions are shown in Fig. 1. Fig. 1 clearly shows that errors in the PO_4^{3-} analyses led to substantial prediction errors, especially in the first mixture set. It is apparent that, although the normalizing procedure compensates for the relative abundance of ions, a set of weighting factors, w_i, should be applied to each equation in eq. 3 to account for uncertainties and accuracy of chemical analysis for each ion. Kowalczyk et al. (1982) weighted concentrations by $1/\sigma_i^2$ where σ_i was the estimated standard deviation of the error in concentration for the ith ion.

Error statistics were calculated for each ion for each of the mixtures shown in Table VI, according to the equation:

$$\epsilon_{ik} = (C_{Tik} - C_{Mik})/C_{Tik} \tag{7}$$

where C_{Tik} is the theoretical concentration of the ith ion in the kth mixture and C_{Mik} is the measured concentration. The theoretical concentration is a linear combination of sources included in the mixture:

$$C_{Tik} = \sum_{j=1}^{N_s} C_{sij} x_{jk} \tag{8}$$

3

TABLE VI

Error statistics

Ion	First set		Second set	
	mean error (%)	standard deviation	mean error (%)	standard deviation
Na^+	21.3	15.8	−0.42	1.63
K^+	9.8	7.5	−0.52	7.10
Ca^{2+}	12.4	11.1	0.11	0.76
Mg^{2+}	14.3	11.1	1.32	3.53
NO_3^-	22.7	52.8	3.44	7.45
PO_4^{3-}(*)	—	—	—	—
SO_4^{2-}	5.3	11.9	5.72	7.00
Cl^-	15.3	25.0	7.06	5.70

*PO_4^{3-} omitted from calculations because of a very large percentage error.

where C_{sij} is the concentration of the ith ion in source j; and x_{jk} is the proportion of source j in the kth prepared mixture.

The mean and standard deviation of the errors for each ion are shown in Table VI. These errors for the first mixture set were much larger than anticipated, and show a significant bias in that the average theoretical concentration is greater than the measured concentration for all ions. The bias ranges from a low of + 5.3% for SO_4^{2-} to + 22.7% for NO_3^-. The high value for NO_3^- is what one could expect from a nonconservative ion; however, no such bias would be anticipated for the other ions. There are six possible sources of bias in water analyses: unrepresentative sampling, instability of samples between sampling and analyses, interference effects, biased calibration, a biased blank correction, and inability to determine all forms of the determinant (Cheeseman and Wilson, 1972). For the first set of mixture samples, it appears that instability, interference effects, and possibly biased calibration might account for the errors shown in Table VI. Also, mistakes in preparing the mixtures and dilution mistakes are, of course, reflected in these data, but are difficult to identify. Identifying the source of these errors was made more difficult because HCO_3^- was measured only for sources 6 and 7 (Table II) in set 1, so we could not check the ion balances for the other sources or mixtures.

These large unexplainable errors served as the impetus to prepare the second set of mixtures and determine if the errors could be reduced. The results indicated that all errors generally were reduced for the second mixture set (Tables III, V and VI). The method of mixture preparation was the likely source of most of the error in the first set of mixtures. In the preparation of the first set of mixtures, the small amounts of source samples used to make the mixtures could account for some of the errors, and this was corrected in the second set of mixtures by using larger samples. The

methods of analysis were the same for both sets of mixtures, although different technicians performed the analyses, and this could account for some error or bias. The second set of mixtures still had an overall bias, with the theoretical concentration being greater than the measured concentration, but the error was significantly smaller, as measured by the mean error and standard deviation (Table VI).

Because of the HCO_3^- analysis of the sources and mixtures, the errors in the second set were investigated further by calculating the ion balance using the major cations and anions (Table II). The difference in milliequivalents per liter between the cations and anions was less than 1 meq for all sources except source 4. Source 4 was from a shallow well influenced by sewage effluent, and microbial activity in the sample was suspected to have shifted the HCO_3^- equilibrium and affected the ion balance. The influence of source 4 on the mixtures was evident in that almost all the mixtures containing it had the higher ion-balance errors (data not presented). Source 9, which was the same as source 4, but filtered through a 0.45-μm filter that should remove microbial activity, had low ion-balance errors throughout. Since ion-balance errors were relatively low and embodied all the analysis errors, a specific source of error could not be determined.

The errors in the computed proportions of source waters, e_x, for all mixtures, were calculated using the relationship:
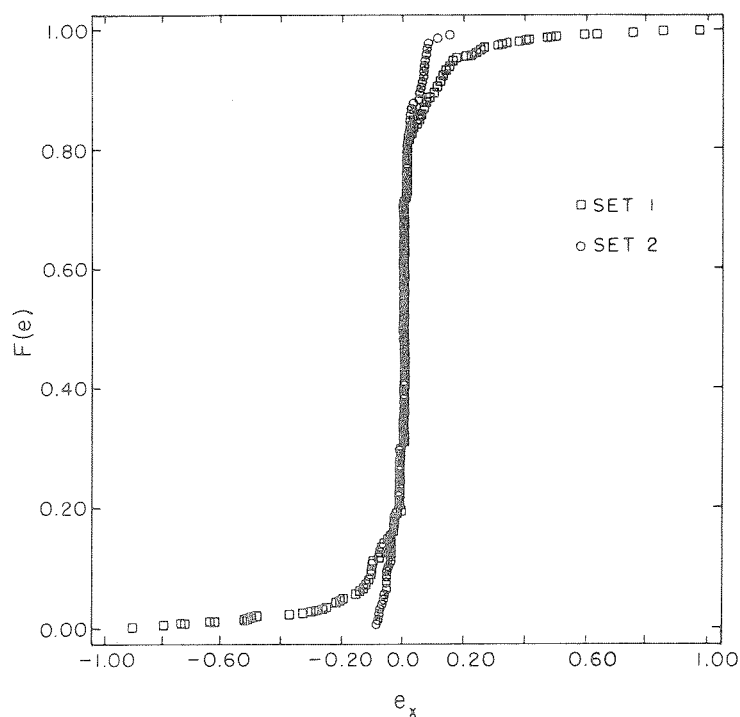
Fig. 2. Distribution of proportion errors, $x_j - \hat{x}_j$.

TABLE VII

Comparison of predictions between variance weighted equations and equally weighted equations

Mixture No.		Source No.										Error index
		1	2	3	4	5	6	7	8	9	10	
	First set:											
1	A[*1]	0.25	0.10	0	0	0	0.25	0.40	0	0	0	
	C2[*2]	0.13	0	0.04	0.25	0	0	0.51	0.04	0	0.04	0.53
	W[*3]	0.21	0	0	0	0	0.13	0.55	0.03	0	0.07	0.75
2	A	0.5	0	0	0	0	0	0.5	0	0	0	
	C2	0.36	0	0	0	0	0	0.47	0.08	0	0.10	0.83
	W	0.46	0	0	0.04	0	0	0.40	0.07	0	0	0.90
3	A	0.5	0	0	0	0	0	0	0.5	0	0	
	C2	0.39	0	0	0	0	0	0.05	0.5	0	0.06	0.89
	W	0.45	0	0	0	0	0	0.02	0.5	0	0.03	0.95
4	A	0.5	0	0	0	0	0	0	0	0.5	0	
	C2	0.48	0	0	0	0	0	0	0.06	0.47	0	0.95
	W	0.28	0	0	0	0	0	0	0.14	0.46	0.12	0.74
6	A	0	0.5	0	0	0	0.5	0	0	0	0	
	C2	0	0.44	0.03	0	0	0	0.41	0.11	0	0	0.45
	W	no convergence										
16	A	0	0	0	0.5	0	0	0.5	0	0	0	
	C2	no convergence										
	W	no convergence										
26	A	0.2	0	0	0	0	0	0.8	0	0	0	
	C2	no convergence										
	W	no convergence										
57	A	0	0	0	0.10	0	0	0.9	0	0	0	
	C2	no convergence										
	W	0	0	0	0	0	0.23	0.52	0.26	0	0	0.52
	Second set:											
11	A						0.25	0.25	0.25	0.25	0	
	C2						0.19	0.31	0.16	0.25	0	0.86
	W						0.24	0.24	0.24	0.24	0.03	0.97
12	A						0.25	0.25	0.25	0	0.25	
	C2						0.21	0.26	0.20	0	0.33	0.91
	W						0.24	0.25	0.25	0	0.27	0.99
14	A						0.25	0	0.25	0.25	0.25	
	C2						0.22	0.07	0.18	0.23	0.31	0.88
	W						0.25	0.03	0.24	0.22	0.26	0.96
17	A						0.5	0	0.5	0	0	
	C2						0.45	0.02	0.42	0	0.11	0.93
	W						0.48	0	0.48	0	0.03	0.97
20	A						0	0.5	0.5	0	0	
	C2						0	0.50	0.50	0	0	1.00
	W						0	0.49	0.51	0	0	0.99
23	A						0	0	0.5	0.5	0	
	C2						0	0.01	0.45	0.47	0.07	0.92
	W						0.01	0	0.50	0.49	0	0.99

[*1] A = actual.
[*2] C2 = equal weight, PO_4^{3-} omitted.
[*3] W = variance weighted, PO_4^{3-} omitted.

$$e_x = x_j - \hat{x}_j \qquad\qquad (9)$$

where x_j is the actual proportion of source j in the mixture; and $\hat{x}_j$ is the amount calculated with PO_4^{3-} omitted. The empirical distribution functions of these errors are shown for both sets of mixtures in Fig. 2. Because of the mass-balance constraint, given by eq. 2, the mean error is zero. There is also a positive probability mass of $e_x = 0$, corresponding to the sources not present in a mixture that were correctly identified as being not present. The falsely identified sources are reflected in the negative errors in Fig. 2, and the tendency to underestimate the proportions of sources present is shown in the positive errors. The reduction in the dispersion of these errors by the improved procedures for mixture set 2 is remarkable.

A set of weighting factors, $w_i = 1/\sigma_i^2$ where σ_i^2 is the error variance for the ith ion, was calculated, and the analysis was repeated for 14 mixtures. The results are compared with the actual proportions and the proportions calculated with the equally weighted equations in Table VII.

These mixtures were selected to represent those for which the estimates, using equally weighted equations, were quite good, and those which were poor, or for which the Wolfe algorithm did not converge. For the first set of mixtures, the variance weighting improved the estimates for mixtures 1, 2 and 3, and allowed convergence for 57. The algorithm still did not converge for 16 and 26, and the weighted estimate for mixture 4 was inferior to the unweighted estimate. The algorithm did not converge for the variance weighted equations for mixture 6. For the second set of mixtures, the variance weighting improved the estimates for mixtures 11, 12, 14, 17 and 23, and had little effect on mixture 20. It also reduced the probability of identifying a source when it was not present.

5. SUMMARY AND CONCLUSIONS

A normalized chemical ion-balance technique (NCIB), to identify water sources in a mixture, was tested using two sets of mixtures prepared in the laboratory. The first set consisted of 64 mixtures of from 2—8 natural waters, selected from 10 sources (surface, groundwater and precipitation). A second set of 25 mixtures was prepared using 5 source waters. Each of the source waters and each mixture was analyzed for at least 8 ions. Proportions of source waters in each sample were estimated using a technique described by Woolhiser et al. (1979, 1982). For the first set of mixtures, the probability of failure to detect a source when it was present in the mixture (type-I error) was 0.23, and the probability of identifying a source when it was not present (type-II error) was 0.29. For the second set of mixtures, the probabilities were 0.0 and 0.33, respectively.

It is possible that some of the solutions for mixture set 1 are not unique. However, the values of the objective functions attained with the estimated x_j were always better than those attained with the theoretically "correct" x_j.

Every effort should be made to include at least as many ions in the analyses as there are candidate sources.

The linear dependence between sources was investigated and found to be significant for several sources. Significant errors, including bias as well as large error variances, were found in the chemical analyses for the first set of mixtures. These errors were substantially reduced for the second set of mixtures, probably because larger sample volumes were used and greater care was taken in preparing and analyzing the samples. The sensitivity of the NCIB technique to errors in chemical analyses dictates that great care should be taken in sample collection and analyses. The NCIB technique, with each normalized ion-balance equation weighted by the inverse of the error variance, showed improved estimates and a reduction in the probability of the type-II error.

It appears that there is little chance of failing to detect a source if it is present in the mixture in relatively large proportions (≥ 0.25), unless there is significant interference between sources. However, sources detected with an estimated proportion smaller than 0.10 are quite likely to be incorrectly identified. Additional investigations would be required to verify or disprove their presence.

ACKNOWLEDGEMENTS

APPENDIX - UNIQUENESS OF QUADRATIC PROGRAMMING SOLUTIONS

The normalized ion-balance equations can be written in the form:

$$\epsilon_i = \sum_{j=1}^{n} c_{ij}^* x_j - 1 = (C^* X)_i - 1, \qquad 1 \leqslant i \leqslant m \qquad (\text{A-1})$$

where $C^* = (c_{ij}^*)$, a $m \times n$ matrix; $c_{ij}^* =$ concentration of the ith ion in the jth inflow divided by the concentration of the ith ion in the mixture; $X = (x_j)$, a $n \times 1$ column matrix; $x_j =$ fraction of the jth inflow in the mixture; and $\epsilon_i =$ normalized error term (due to mismeasurement, etc.).

The normalized inflows satisfy:

$$x_j \geqslant 0 \qquad \text{and} \qquad \sum_{j=1}^{n} x_j = 1$$

so we can establish values from them by minimizing:

$$F(X) = \sum_{i=1}^{m} (\epsilon_i)^2 - m \qquad (\text{A-2})$$

on

$$D = \left\{ X \ \middle| \ \sum_{j=1}^{n} x_j = 1; x_j \geqslant 0, \ 1 \leqslant j \leqslant n \right\}$$

F can be written in matrix notation as:

$$F(X) = X^T A X - 2B^T X \qquad\qquad (A\text{-}3)$$

where $A = C^{*\,T} C^*$, a $n \times n$ matrix; and B = a $n \times 1$ matrix with:

$$B_j = \sum_{i=1}^{m} c_{ij}$$

F is continuous, and D is a closed and bounded set. Thus, F has a minimum on D, though it may not be unique. F is a quadratic form, and:

$$X^T A X = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} c_{ij}^* x_j \right)^2 \geqslant 0$$

so A is a positive semi-definite. Thus, the long form of Wolfe's (1959) algorithm can be used to find a minimum.

If $\|A\| \neq 0$, then A is positive definite and, as noted by Wolfe (1959), F has a unique minimum. When the determinant is zero, as when $m < n$, further analysis is needed to decide if a minimum found by Wolfe's algorithm is unique.

Suppose X_0 is a minimum of F on D. If there is another minimum, say X_1, then consider $X_0 + tV$, $0 \leqslant t \leqslant 1$ where $V = X_1 - X_0$. This is the line segment joining X_0 and X_1. Since X_0 and X_1 are in D, and D is convex, this segment lies in D. The convexity of F gives:

$$F(X_0 + tV) = F[tX_1 + (1-t)X_0] \leqslant tF(X_1) + (1-t)F(X_0)$$

and since $F(X_0) = F(X_1)$ is the minimum value of F on D, we get:

$$F(X_0) \leqslant F(X_0 + tV) \leqslant tF(X_1) + (1-t)F(X_0) = tF(X_0)$$

$$+ (1-t)F(X_0) = F(X_0)$$

This reasoning gives rise to necessary and sufficient conditions for F to have multiple minima:

F has multiple minima if there is a vector V such that:

(1) $V \neq 0$

(2) $X_0 + tV \in D$ for small $t \geqslant 0$ (A-4)

(3) $F(X_0 + tV)$ is constant for small $t \geqslant 0$

For (2) of expressions (A-4) to hold, we need $1 = \Sigma(X_0 + tV) = \Sigma X_0 + t\Sigma V$, and $X_0 + tV \geqslant 0$ for small t. Since $\Sigma X_0 = 1$ and $X_0 \geqslant 0$, this happens just when $\Sigma V = 0$, and $V_j \geqslant 0$ for every j corresponding to zero coordinates of X_0. Since:

$$F(X_0 + tV) = (X_0{}^\mathrm{T}AX_0 - 2B^\mathrm{T}X_0) + t(X_0{}^\mathrm{T}AV + V^\mathrm{T}AX_0 - 2B^\mathrm{T}V)$$
$$+ t^2(V^\mathrm{T}AV)$$

for (3) of expressions (A-4) to hold, we need $X_0{}^\mathrm{T}AV + V^\mathrm{T}AX_0 - 2B^\mathrm{T}V = 0$ and $V^\mathrm{T}AV = 0$. As noted by Wolfe (1959), $V^\mathrm{T}AV = 0$ implies $AB = 0$ and, by taking the transpose, $V^\mathrm{T}A = 0$. Thus, the two equalities reduce to $B^\mathrm{T}V = 0$ and $AV = 0$. In summary, we get:

F has multiple minima if there is a vector V such that:

(1) $V \neq 0$

(2) $V_j \geqslant 0$ for $j \in K = \{j | (X_0)_j = 0\}$

(3) $\displaystyle\sum_{j=1}^{n} V_j = 0$

(4) $AV = 0$

(5) $B^\mathrm{T}V = 0$

(A-5)

Note that if V satisfies expressions (A-5), then so does any scalar multiple of V. In particular, in view of (1) of expressions (A-5), we may divide V by the maximum of the absolute values of the coordinates of V to get an added condition:

(6) $-1 \leqslant V_j \leqslant 1$ for $1 \leqslant j \leqslant n$

A change of variables, $V_j = V_j + 1$, and letting $b = (1, \ldots, 1)^\mathrm{T}$ gives:

F has a multiple minima if there is a vector V such that:

(1) $V \neq b$

(2) $V_j \geqslant 1$ for $j \in K$

(3) $\displaystyle\sum_{j=1}^{n} V_j = n$

(4) $AV = Ab$

(5) $B^\mathrm{T}V = B^\mathrm{T}b$

(6) $V_j \leqslant 2$ for $1 \leqslant j \leqslant n$

(7) $V_j \geqslant 0$ for $1 \leqslant j \leqslant n$

(A-6)

We are now in a position to express our problem in terms of another quadratic programming problem; this time we are only concerned with the value of a maximum, not its uniqueness.

(2)—(7) of expressions (A-6) can be put in a normal form by adding slack and surplus variables as described in Hadley (1964). That is, there are variables W_j, which include the V_j and matrices A^* and b^*, which allow us to write (2)—(7) of expressions (A-6) as $A^*W = b^*$ and $W \geqslant 0$. Now consider the quadratic form:

$$f(W) = \sum_{j=1}^{n} (V_j - 1)^2 - n \qquad\qquad (A-7)$$

The maximum of f on $D^* = \{W | A^*W = b^*, W \geqslant 0\}$ will be $-n$ just when W in D^* gives $V = b$. Thus, we arrive at the conclusion:

F has a unique minimum on D if the maximum of f on D^* is greater than $-n$. We also note that, because of (6) and (7) of expressions (A-6), and the manner in which slack and surplus variables are added, that D^* is closed and bounded. Thus, f will have a maximum on D^*.

REFERENCES

Cheeseman, R.V. and Wilson, A.L., 1972. Manual on analytical quality control for the water industry. Water Res. Cent., Medmanham, Tech. Rep. TR66. (cited by Kirchner, 1983).

Gorelick, S.M., Evans, B. and Remson, I., 1983. Identifying sources of groundwater pollution: An optimization approach. Water Resour. Res., 19(3): 779—790.

Hadley, G., 1964. Nonlinear and Dynamic Programming. Addison-Wesley, Reading, Mass., 484 pp.

Hall, F.R., 1970. Dissolved solids—discharge relationships, 1. Mixing models. Water Resour. Res., 6(3): 845—850.

Kirchner, C.J., 1983. Quality control in water analyses. Environ. Sci. Technol., 17(4): 174A—181A.

Knisel, W.G. (Editor), 1980. CREAMS: A field-scale model for chemicals, runoff and eorsion from agricultural management systems. U.S. Dep. Agric., Conserv. Res. Rep. No. 26, 640 pp.

Kowalczyk, G.S., Gordon, G.E. and Rheingrover, S.W., 1982. Identification of atmospheric particulate sources in Washington, D.C., using chemical element balances. Environ. Sci. Technol., 16: 79—90.

Morris, M.D., Berk, J.A., Krulik, J.W. and Eckstein, Y., 1983. A computer program for a trilinear diagram plot and analysis of water-mixing systems. Ground Water, 21(1): 67—68.

Pinder, G.F. and Jones, J.F., 1969. Determination of the groundwater component of peak discharge from the chemistry of total runoff. Water Resour. Res., 5(2): 438—445.

Piper, A.M., 1944. A graphic procedure in the geochemical interpretation of water-analyses. Trans. Am. Geophys. Union, 25: 914—923.

Schwartz, F.W., 1980. Hydrogeological investigation of Muskeg River Basin, Alberta. Prepared for Alberta Oil Sands Environmental Research Program by Univ. of Alberta, Dep. of Geol., Edmonton, Alta., AOSERP Rep. 87, 97 pp.

Schwartz, F.W. and Domenico, P.A., 1973. Simulation of hydrochemical patterns in regional groundwater flow. Water Resour. Res., 9(3): 707—720.

Sklash, M.G., Farvolden, R.N. and Fritz, P., 1976. A conceptual model of watershed response to rainfall, developed through the use of oxygen-18 as a natural tracer. Can. J. Earth Sci., 13(2): 271—283.

Tsurumi, M., 1982. Self-consistent least-squares method of estimating chemical composition of the sources of chemical constituents in natural waters. Anal. Chim. Acta, 138: 177—182.

Visocky, A.P., 1970. Estimating the groundwater component of storm runoff by the electrical conductivity method. Ground Water, 8(2): 5—10.

Wolfe, P. 1959. The SIMPLEX method for quadratic programming. Econometrica, 27: 382—398.

Woolhiser, D.A., Gardner, H.R. and Olsen, S.R., 1979. Evaluating water quality effects of surface mining, presented at 1979 Winter Meet. A.S.A.E. (Am. Soc. Agric. Eng.), published as Pap. No. 79-2522 by A.S.A.E., St. Joseph, Mich., 32 pp.

Woolhiser, D.A., Gardner, H.R. and Olsen, S.R., 1982. Estimation of multiple inflows to a stream reach using water chemistry data. Trans. A.S.A.E. (Am. Soc. Agric. Eng.), 25(3): 616—622.