Separation Science and Technology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713708471

Application of Factor Analysis to the Prediction of Activity Coefficients of Nonelectrolytes

Philip T. Funke[a]; Edmund R. Malinowski[a]; Daniel E. Martire[ab]; Luigi Z. Pollara[a]
[a] Department of Chemistry and Chemical Engineering, Stevens Institute of Technology, Hoboken, New Jersey [b] Chemistry Department, Georgetown University, Washington, D.C

To cite this Article Funke, Philip T. , Malinowski, Edmund R. , Martire, Daniel E. and Pollara, Luigi Z.(1966) 'Application of Factor Analysis to the Prediction of Activity Coefficients of Nonelectrolytes', Separation Science and Technology, 1: 6, 661 — 676

To link to this Article: DOI: 10.1080/01496396608049473

URL: http://dx.doi.org/10.1080/01496396608049473

SEPARATION SCIENCE, 1(6), 661–676 (1966)

Application of Factor Analysis to the Prediction of Activity Coefficients of Nonelectrolytes

---

PHILIP T. FUNKE, EDMUND R. MALINOWSKI, DANIEL E. MARTIRE,* and LUIGI Z. POLLARA
DEPARTMENT OF CHEMISTRY AND CHEMICAL ENGINEERING,
STEVENS INSTITUTE OF TECHNOLOGY, HOBOKEN, NEW JERSEY

Summary

The mathematical technique of factor analysis is employed in the prediction of infinite-dilution solute activity coefficients for solutions of nonelectrolytes.

Activity coefficients at infinite dilution in the liquid phase are of interest and importance to chemists and chemical engineers. Physical chemists study them because they represent a direct macroscopic measure of the interaction between a solute molecule and its solvent environment. Other chemists are interested in them because of their utility in the selection of an optimum solvent in extractive distillation, in liquid-liquid chromatography, or in gas-liquid chromatography. Chemical engineers are interested in them because of their utility in the prediction of vapor-liquid equilibrium for use in the design of heterogeneous reactors, absorbers, distillation columns, and other petroleum or chemical processes in which desirable products are separated from undesirable components.

The ability to make accurate quantitative predictions about deviations from ideal behavior in solutions of nonelectrolytes is virtually essential for intelligent separation-process evaluation.

---

* Present address: Chemistry Department, Georgetown University, Washington, D.C.

It is not always feasible to determine activity coefficients experimentally for this purpose. Furthermore, the theory of solutions is at present insufficiently advanced to provide an exact quantitative treatment with which to predict activity coefficients, particularly for mixtures containing one or more polar components.

Consequently, it was decided to attempt the development of a completely general and empirical method for predicting activity coefficients from a limited amount of experimental data. To this end, the mathematical technique of factor analysis and activity coefficient data from gas-liquid chromatography were employed.

FORMALISM OF FACTOR ANALYSIS

Since 1930, psychologists have been using the mathematical technique of factor analysis to help simplify many complex problems concerning animal behavior. Recently, Malinowski (*1*) derived the necessary relationships for the potential application of factor analysis to physical and chemical problems. The present study represents the first application of his proposed treatment to the solution of an actual problem. Only the pertinent derivations will be given here. For a comprehensive survey of the various other mathematical treatments of factor analysis one is referred elsewhere (*2–4*).

An over-all schematic representation of the key steps involved in factor analysis is presented in Fig. 1. Essentially, factor analysis divides a measurement into a linear sum of fundamental variables. By mathematical manipulation a matrix of the experimental data is converted into a correlation matrix, which is then broken up into linear factors. The linear factors are mathematically rotated into physically significant linear parameters (such as bond length, electronegativity, etc.), which in turn account for the experimental data.

Our aim in this study is to express the natural logarithm of the activity coefficient of the solute component at infinite dilution ($\ln \gamma^{\infty}$) with a minimum number of independent factors that combine as weighted sums. Let n_{ij} be the contribution of the jth factor of solute i, and let q_{jk} be the contribution of the jth factor of solvent k. Assuming that the activity coefficient of solute i in solvent k ($\ln \gamma_{ik}^{\infty}$) can be written as a linear combination of the products of

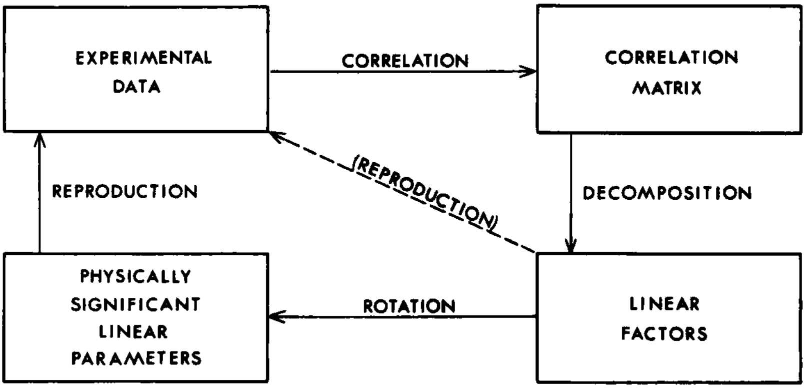

FIG. 1. The key steps of factor analysis.

the independent solute and solvent factors, we have

$$\ln \gamma_{ik}^{\infty} = \sum_{j=1}^{j=r} n_{ij} q_{jk} \tag{1}$$

The problem now is to find the minimum number (r) of solute and solvent factors which are necessary to reproduce the data, and then to find the nature and value of these factors. The solution follows; see Table 1.

We select activity coefficient data for t solutes in s solvents and construct a $t \times s$ matrix, $[P]$, called the *data* or *property matrix*; in our case the property is $\ln \gamma^{\infty}$. If we assume that $s < t$, then, for any scheme to work, the minimum number of factors (r) must be less than the smallest dimension of the data matrix (s). In other

TABLE 1

	Solvent				
Solute	1	2	3	· · ·	s
1	P_{11}	P_{12}	P_{13}	· · ·	P_{1s}
2	P_{21}	P_{22}	P_{23}	· · ·	P_{2s}
3	P_{31}	P_{32}	P_{33}	· · ·	P_{3s}
·	·	·	·	· · ·	·
·	·	·	·	· · ·	·
·	·	·	·	· · ·	·
t	P_{t1}	P_{t2}	P_{t3}	· · ·	P_{ts}

words, we must have more data than we have factors; i.e., we must start with a set of data that is linearly dependent.

If we write $[\ln \gamma_{ik}^{\infty}]$, $[n_{ij}]$, and $[q_{jk}]$ as the matrices whose elements are $\ln \gamma_{ik}^{\infty}$, n_{ij}, and q_{jk}, respectively, then, considering Eq. (1), we have

$$[P] = [\ln \gamma_{ik}^{\infty}] = [n_{ij}][q_{jk}] \tag{2}$$

where $[n_{ij}]$ is a $t \times r$ matrix which we shall call the *solute-factor matrix*, and $[q_{jk}]$ is an $r \times s$ matrix which we shall call the *solvent-factor matrix*. Our first objective is to determine the numerical values of the solute- and solvent-factor matrices strictly from a knowledge of the matrix of the experimental data.

Accordingly, we construct a square symmetric *correlation matrix*, $[C]$, of dimension $s \times s$, by taking the product of the data matrix premultiplied by its transpose,

$$[C] = [P^T][P] \tag{3}$$

The matrix $[C]$ is then diagonalized by finding a matrix $[B]$ such that

$$[B^{-1}][C][B] = [\lambda_j \partial_{jk}] \tag{4}$$

where ∂_{jk} is the Kronecker delta and the λ_j's are the eigenvalues of the set of equations

$$[C]\{B_j\} = \lambda_j\{B_j\} \tag{5}$$

with $j = 1, 2, 3, \ldots, s$, and where the $\{B_j\}$ are the corresponding eigenvectors. These eigenvectors are orthogonal and therefore linearly independent. Thus they can be used as the basis set to express our data.

Now, from Eq. (6),

$$\begin{aligned}[B^{-1}][C][B] = [B^{-1}][P^T][P][B] &= [B^T][P^T][P][B] \\ &= [D^T][D] = [\lambda_j \partial_{jk}]\end{aligned} \tag{6}$$

we see that the property matrix can be expressed in terms of $[B]$ and $[D]$:

$$[D] = [P][B]$$

or

$$[P] = [D][B^T] \tag{7}$$

Comparing the above result with Eq. (2), it is obvious that

$$\begin{aligned}[D] &= [n_{ij}] \\ [B^T] &= [q_{jk}]\end{aligned} \tag{8}$$

Now, to find the minimum number of linearly independent vectors necessary to reproduce $[P]$ within experimental error, we start with the eigenvector B_1 associated with the largest eigenvalue λ_1, and perform the following matrix multiplication using the appropriate D:

$$[P] = [D_1][B_1] \tag{9}$$

where $D_1 = \{n_{i1}\}$, a column vector, and $B_1 = \{q_{1k}\}$, a row vector. We continue by employing the next largest eigenvector,

$$[P] = [D_1, D_2]\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \tag{10}$$

and the next one, and so on, until we have a sufficient number to reproduce the data. If B_m (the last eigenvector needed to reproduce the data) is equal to B_s, then either we have not introduced enough data (i.e., our data do not span the factor space) or all the data cannot be expressed in terms of the same factors. On the other hand, if $m = r < s$, this stage of the factor analysis is complete, and r is determined by the number of eigenvectors needed.

From the standpoint of a chemist, the treatment should not end here. The solute and solvent factors in their present forms are not necessarily recognizable as physical or chemical parameters. As Malinowski (*1*) points out, the reference axes were chosen to yield true mathematical solutions to the problem and to span the factor space. For completeness, we should now seek out chemically recognizable factors from the orthogonal set of vectors; i.e., we must rotate the reference axes so that they become aligned with fundamental properties of the molecules.

We can accomplish this by the introduction of a matrix $[R]$, such that

$$\begin{aligned}[P] &= [D][R^{-1}][R][B^T] \\ &= \quad [F] \qquad [A]\end{aligned} \tag{11}$$

where $[F] = [D][R^{-1}]$ and $[A] = [R][B^T]$.

If $[F]$ is known, then $[R]$ can be found, because from Eqs. (6)

and (11),

$$[D^T][F] = [D^T][D][R^{-1}] = [\lambda_j \partial_{jk}][R^{-1}]$$

or

$$[R^{-1}] = [(1/\lambda_j)\partial_{jk}][D^T][F] \tag{12}$$

The object is to "guess" at $[F]$ and then to perform the indicated operations to see if $[P]$ can be reproduced and if the $[A]$ matrix can be identified with physically meaningful quantities. In this manner the last step, called *reproduction* (see Fig. 1, solid arrow), may be completed.

The last part of the factor-analysis scheme is the weak point of the technique. It is quite difficult (and frequently impossible) to obtain the functional form of all the physical parameters. However, as we shall soon see, the "rotation" need not be successful to employ factor analysis as a data-predicting tool. In that case one would short-circuit the scheme, going directly from the mathematical factors to reproduction of the experimental data (see Fig. 1, dashed arrow).

APPLICATION OF FACTOR ANALYSIS TO THE PROBLEM

Gas-liquid chromatography (GLC) was used to obtain infinite-dilution solute activity coefficients for 273 binary nonelectrolytic solutions (39 solutes, 7 solvents) at three temperatures. These data have been published elsewhere (5).

To demonstrate the utility of the technique, a factor analysis has been performed on the data at one temperature (74.1°C). The property matrix $[P_{ik}]$ was formed by taking 0.1 ln γ_{ik}^{∞} as the indi-

TABLE 2

Solvent-Factor Matrix $[q_{jk}]$ and Solute-Factor Matrix $[n_{ij}]$ at 74.1°C

	Solvent factors				
k	q_{1k}	q_{2k}	q_{3k}	q_{4k}	q_{5k}
1	−.071605	−.176336	.595539	−.173640	−.108781
2	−.158792	−.590313	.440760	.158289	.501329
3	.216734	.202399	.491188	−.491170	−.325187
4	.384875	−.527745	−.320597	−.587322	−.036614
5	.341108	−.368229	.082487	.565865	−.642885
6	.518916	−.122168	−.143592	.076422	.330693
7	.623638	.387863	.282120	.180428	.327242

TABLE 2 (*Continued*)

	Solute factors				
i	n_{i1}	n_{i2}	n_{i3}	n_{i4}	n_{i5}
1	−.049532	.080031	−.016950	−.030155	.011960
2	−.031515	.075496	−.013552	−.032131	.017319
3	.063266	.062745	−.040139	.005527	−.012960
4	.096945	.052620	−.021702	.002344	−.005247
5	.085898	.054676	−.029760	.002724	−.008225
6	.032719	.066479	−.038576	−.010303	−.001649
7	.073024	.060777	−.018255	−.003278	−.000224
8	.088923	.051212	−.008069	−.001563	−.002775
9	.113290	.044398	.004615	−.002661	.001036
10	.094281	.052666	−.007068	−.002757	.001698
11	.106589	.043923	.002006	−.001892	−.000342
12	.123814	.036319	.009636	−.003104	.001662
13	.144596	.042294	.002396	.003880	−.002081
14	.139090	.041981	−.001155	.006069	−.002824
15	.161545	.038652	.012901	.002729	.003260
16	.149849	.042421	.004470	.002111	.003744
17	.162178	.034110	.012065	.003936	.000580
18	.156472	.033952	.007427	.004796	−.000769
19	.181708	.027776	.022532	.003812	.003219
20	.184340	.026231	.024389	.003057	.003255
21	.159319	.037687	.009357	.001633	.005566
22	.169263	.031101	.014214	.002589	.003314
23	.172659	.027314	.015229	.004382	.001494
24	−.046560	.060575	−.009647	−.010391	−.011116
25	−.118989	.099813	.017557	.005154	−.001302
26	−.146262	.139314	.004821	.036839	.007793
27	−.017236	.083470	−.031787	−.012789	.003038
28	−.083876	.060155	.045690	−.006422	−.002676
29	−.148330	.126639	.010556	.030385	.006065
30	−.048304	.084704	−.022993	−.010214	.002406
31	−.112681	.103747	.014924	.005773	.002067
32	−.063686	.096355	−.023005	.005457	−.005140
33	−.059999	.099096	−.022161	.005255	−.000555
34	−.067355	.055338	.057004	−.015885	−.006359
35	−.060357	.077225	.024815	−.005502	−.002753
36	−.058253	.046510	.053413	−.018423	−.005967
37	−.013807	.062212	.003175	−.008637	−.010849
38	.005135	.055531	.010673	−.010211	−.007647
39	.015129	.055841	.014282	−.011493	−.005565

vidual elements. The dimensions of this matrix were 39×7 ($t = 39$, $s = 7$), with each column representing the property of all the solutes in a particular solvent.

The computation was performed on an IBM 1620 Mark II computer using an original program which carried out the matrix manipulations outlined in the previous section. The resulting numerical values for the solvent-factor matrix $[q_{jk}]$ and the solute-factor matrix $[n_{ij}]$ are given in Table 2. The solute and solvent numbering code is given in Table 3. The correlation matrix and the eigenvalues are not listed. It was found that five factors ($r = 5$) are

TABLE 3

Solvent and Solute Code

k	Solvent	*k*	Solvent
1	*n*-Eicosane	5	Di-*n*-butyltetrachlorophthalate
2	Squalane	6	Di-*n*-butylphthalate
3	1-Hexadecanol	7	Benzyldiphenyl
4	N,N-Dimethylmyristamide	8	Dinonylphthalate
i	Solute	*i*	Solute
1	Benzene	23	*n*-Heptane
2	Toluene	24	Methyl iodide
3	Cyclopentane	25	Dichloromethane
4	Methylcyclopentane	26	Chloroform
5	Cyclohexane	27	Carbon tetrachloride
6	Cyclohexene	28	Dibromomethane
7	2-Methylbutene-2	29	Bromodichloromethane
8	Pentene-1	30	Bromotrichloromethane
9	4- Methylpentene-1	31	*cis*-1,2-Dichloroethylene
10	2-Methylpentene-1	32	*trans*-1,2-Dichloroethylene
11	Hexene-1	33	Trichloroethylene
12	Heptene-1	34	1,2-Dichloroethane
13	Isopentane	35	1,1-Dichloroethane
14	*n*-Pentane	36	1,2-Bromochloroethane
15	2,2-Dimethylbutane	37	1-Chloropropane
16	2,3-Dimethylbutane	38	1-Chlorobutane
17	2-Methylpentane	39	2-Chlorobutane
18	*n*-Hexane	40	Methylcyclohexane
19	2,2-Dimethylpentane	41	2-Methylbutene-1
20	2,4-Dimethylpentane	42	3-Methylpentane
21	2,3-Dimethylpentane	43	Bromochloromethane
22	3-Methylhexane	44	2-Chloropropane

needed to reproduce the measured activity coefficient data to within 2% (taken as the experimental error) for all 273 systems. Note that, since $s = 7$, the factor space has been spanned ($s > r$).

At this point we have developed a consistent mathematical scheme for 39 solutes in 7 solvents. If we have chosen these binary systems wisely, so that all the conceivable types of intermolecular forces have been incorporated, we should be able to fit new solvents and new solutes into our scheme. We should be able to characterize each of these new compounds by five independent factors.

To fit a new solvent ($k = 8$) into the scheme, i.e., to find a new set of q values, we must first experimentally determine the activity coefficients of five solutes in this solvent. The solutes, carefully chosen to span the solute-factor space, are, in order, $i = 1$, $i = 11$, $i = 20$, $i = 26$, and $i = 34$. The solvent being characterized ($k = 8$) is di-(3,3,5-trimethylhexyl)phthalate, also known as dinonylphthalate or DNP. The activity coefficients were measured by GLC in the usual manner (5) and are listed in Table 4.

We can now write

$$(0.1 \ln \gamma_{i8}^{\infty}) = [N_{ij}](q_{j8}) \tag{13}$$

where $(0.1 \ln \gamma_{i8}^{\infty})$ is a column vector composed of the measured

TABLE 4

Characterization of a New Solvent ($k = 8$), Dinonylphthalate, at 74.1°C

A. Experimental activity coefficients in DNP

Solute	$i = 1$	$i = 11$	$i = 20$	$i = 26$	$i = 34$
γ_{ik}^{∞}	0.552	0.956	1.281	0.331	0.494

B. The matrix $[N_{ij}^{-1}]$

−0.485303	3.712916	2.075084	−0.849182	−1.090963
0.935736	13.618815	−4.938518	2.604648	1.691631
−0.893750	−24.876668	18.518088	0.834583	10.158809
−13.707298	−6.535990	6.606945	11.735748	−7.665114
39.513498	−127.489240	84.542771	9.826367	−20.766745

C. Solvent factors for DNP (q_{j8})

q_{18}	q_{28}	q_{38}	q_{48}	q_{58}
.234338	−.646417	−.184668	.251518	.695656

values for the five solutes, $[N_{ij}]$ is a 5 × 5 matrix composed of the n_{ij}'s for these solutes, and (q_{j8}) is a column vector (the unknown) composed of the five factors for the new solvent.

Premultiplying both sides of Eq. (13) by $[N_{ij}^{-1}]$, we obtain

$$[N_{ij}^{-1}](0.1 \ln \gamma_{i8}^{\infty}) = [N_{ij}^{-1}][N_{ij}](q_{j8}) = (q_{j8}) \tag{14}$$

Thus to determine (q_{j8}) from Eq. (14) the inverse of $[N_{ij}]$ must first be calculated. The numerical values for both $[N_{ij}^{-1}]$ and (q_{j8}) are tabulated in Table 4.

We should now be able to predict the activity coefficients of all 39 solutes in DNP by employing Eq. (1), the n_{ij} values in Table 2, and the q_{j8} values in Table 4. The predicted values are compared with the experimental GLC values in Table 5. The agreement is excellent. Therefore, on the basis of five measurements one can predict the activity coefficients for 39 binary systems. It is particularly significant that both positive and negative deviations from Raoult's law have been quantitatively predicted with an accuracy that approaches experimental accuracy.

Let us now attempt the characterization of five new solutes ($i = 40, 41, 42, 43, 44$); i.e., let us try to find a set of n values for each of these solutes. To accomplish this, we must first measure the activity coefficients for each solute in five solvents which span the solvent-factor space. The solvents chosen are, in order, $k = 2$, $k = 3$, $k = 4$, $k = 5$, and $k = 7$. The solute activity coefficients were again measured by GLC and are tabulated in Table 6, along with the pure-solute-saturated vapor pressures (p°) at 74.1°C, which were used in the calculation of γ_{ik}^{∞} from the GLC specific retention volume data (5).

We can now write

$$(0.1 \ln \gamma_{ik}^{\infty}) = (n_{ij})[Q_{jk}] \tag{15}$$

where $(0.1 \ln \gamma_{ik}^{\infty})$ is a row vector composed of the measured values for solute i in the five solvents, (n_{ij}) is a row vector (the unknown) composed of the five factors for solute i, and $[Q_{jk}]$ is a 5 × 5 matrix composed of the q_{jk}'s for the five chosen solvents.

Postmultiplying both sides of Eq. (15) by $[Q_{jk}^{-1}]$, we obtain

$$(0.1 \ln \gamma_{ik}^{\infty})[Q_{jk}^{-1}] = (n_{ij})[Q_{jk}][Q_{jk}^{-1}] = (n_{ij}) \tag{16}$$

Hence $[Q_{jk}^{-1}]$ must first be calculated to determine (n_{ij}) from Eq.

TABLE 5

Predicted versus Measured Activity Coefficients for Solutes in DNP at 74.1°C

Solute (i)	Predicted 0.1 ln γ^∞	Predicted γ^∞	Measured γ^∞
1	−.059475	0.552	0.552
2	−.049718	0.608	0.598
3	−.025947	0.771	0.759
4	−.010601	0.899	0.901
5	−.014755	0.863	0.871
6	−.031921	0.726	0.718
7	−.019784	0.820	0.798
8	−.013100	0.877	0.866
9	−.002952	0.971	0.975
10	−.010158	0.903	0.877
11	−.004499	0.956	0.956
12	.004133	1.042	1.036
13	.005631	1.058	1.076
14	.005232	1.054	1.043
15	.013443	1.144	1.164
16	.010003	1.105	1.098
17	.015121	1.163	1.160
18	.014020	1.151	1.132
19	.023663	1.267	1.276
20	.024771	1.281	1.281
21	.015528	1.168	1.153
22	.019892	1.220	1.205
23	.022133	1.248	1.229
24	−.058632	0.556	0.543
25	−.095254	0.385	0.363
26	−.110533	0.331	0.331
27	−.053229	0.587	0.609
28	−.070455	0.494	0.498
29	−.106709	0.344	0.357
30	−.052723	0.590	0.581
31	−.096211	0.381	0.362
32	−.075164	0.471	0.471
33	−.073089	0.481	0.474
34	−.070501	0.494	0.494
35	−.071945	0.486	0.475
36	−.062364	0.536	0.546
37	−.053756	0.584	0.568
38	−.044552	0.640	0.622
39	−.041951	0.657	0.650

TABLE 6

Characterization of Five New Solutes at 74.1°C

A. Experimental activity coefficients in five solvents

Solute (i)	Solvent $k=2$	$k=3$	$k=4$	$k=5$	$k=7$	Solute vapor pressure $p°$, mm Hg
40	0.564	1.264	1.178	1.143	2.187	333.2
41	0.596	1.257	1.091	1.030	2.032	2750.4
42	0.644	1.499	1.395	1.461	3.080	1055.2
43	0.806	1.082	0.388	0.531	0.756	918.5
44	0.732	1.291	0.803	0.837	1.339	2438.6

B. The matrix $[Q_{jk}^{-1}]$

−0.170418	−0.685937	0.737094	0.073541	0.461771
−0.051894	0.121105	0.996284	−0.652751	−0.543622
0.644670	−0.557772	−0.485273	−0.522776	0.139124
0.357808	−0.390718	0.133168	0.522658	−0.638311
0.984574	0.341193	0.068098	0.265921	0.569777

C. Solute factors (n_{ij})

Solute (i)	n_{i1}	n_{i2}	n_{i3}	n_{i4}	n_{i5}
40	.100962	.054493	−.019803	.000124	−.000863
41	.084103	.056467	−.014386	−.002829	.003408
42	.151158	.040053	.004470	.003842	.002245
43	−.107927	.083791	.027570	.000313	−.002927
44	.012190	.053721	.012709	−.009593	−.003366

(16). The computed numerical values for $[Q_{jk}^{-1}]$ and (n_{ij}) are given in Table 6.

Finally, to verify that the n_{ij} values found are consistent with the entire factor scheme, the activity coefficients are predicted for the five solutes in the remaining three solvents ($k = 1$, $k = 6$, and $k = 8$) by employing Eq. (1) and the data in Tables 2, 4, and 6. The favorable comparison of the predicted values with the experimental GLC values can be seen in Table 7. Again, on the basis of five measurements, we have been able to characterize a new solute component. Prediction appears possible over a wide range of activity coefficients.

TABLE 7

Predicted versus Measured Activity Coefficients for the Five Solutes in Three Solvents at 74.1°C

	Values in *n*-eicosane ($k = 1$)		
Solute (i)	Predicted 0.1 ln γ^∞	Predicted γ^∞	Measured γ^∞
40	−.028560	0.751	0.754
41	−.024426	0.783	0.795
42	−.015999	0.852	0.860
43	.009636	1.101	1.109
44	−.000745	0.999	1.021

	Values in di-*n*-butylphthalate ($k = 6$)		
Solute (i)	Predicted 0.1 ln γ^∞	Predicted γ^∞	Measured γ^∞
40	.048301	1.622	1.740
41	.039720	1.488	1.372
42	.073939	2.096	2.081
43	−.071144	0.491	0.503
44	−.013909	0.870	0.843

	Values in dinonylphthalate ($k = 8$)		
Solute (i)	Predicted 0.1 ln γ^∞	Predicted γ^∞	Measured γ^∞
40	−.008478	0.919	0.955
41	−.012477	0.883	0.819
42	.011234	1.119	1.102
43	−.086504	0.421	0.412
44	−.038971	0.677	0.621

DISCUSSION

At this point, one still cannot be absolutely certain that the factor space has been completely spanned for all nonelectrolytic solutes and solvents. Although it is true that the original compounds (the 39 solutes and 7 solvents) included in the factor scheme were quite diversified, there is still the distinct possibility that more than five factors are involved. Any doubts can be eliminated only by successfully introducing into the scheme new compounds with different functional groups and by continued success in generating accurate predictions.

On the other hand, there may even be less than five factors, because the experimental error could be greater than 2%. This is not unlikely when one considers, for example, our inability to correct the GLC data for gas-phase imperfections.

It should be mentioned that we have failed so far in our attempts to determine the functional form of the true solute and solvent parameters. We are presently working on this aspect of the problem, and are endeavoring to resolve the above questions.

In any event, factor analysis does show great promise for predicting many activity coefficients on the basis of a few measurements. This is very useful in separation-process evaluation. To illustrate, let us consider a problem of interest—the selection of an effective solvent to bring about a difficult separation by GLC.

Most analysts appreciate the fact that the separation of a two-component mixture can be effected by a nonpolar stationary liquid phase if the components have appreciably different boiling points. They also can appreciate that one must resort to "polar" or "selective" liquid phases (solvents) to separate a mixture in which the compounds have nearly identical boiling points, but different dipole moments. Unfortunately, the solvent-selection procedure is not always this simple. Often, the solutes to be separated have such subtle differences that unforeseen factors can influence their relative retention. Thus it would be highly desirable that a quantitative basis be given to the selection of a liquid phase.

Consider, for example, two compounds, x and y, that we wish to separate by GLC. The degree of separation in a particular solvent, i.e., the relative volality (α_{xy}), is given by the ratio of their retention times (t_R):

$$\alpha_{x,y} = \frac{(t_R)_y}{(t_R)_x} \tag{17}$$

It can be shown (6) that in a given column under specified operating conditions

$$\alpha_{x,y} = \frac{(V_g^\circ)_y}{(V_g^\circ)_x} = \frac{\gamma_x^\infty p_x^\circ}{\gamma_y^\infty p_y^\circ} \tag{18}$$

where V_g° is the specific retention volume of the solute component, γ^∞ its activity coefficient, and p° the saturated vapor pressure of the pure solute at the column operating temperature.

A problem often encountered is the separation of two close-boiling compounds ($p_x^\circ \approx p_y^\circ$). In this case, Eq. (18) reduces to

$$\alpha_{x,y} = \gamma_x^\infty/\gamma_y^\infty \equiv S_{x,y} \tag{19}$$

where $S_{x,y}$ is the selectivity of the solvent toward solutes x and y. An effective solvent is then characterized by a value of $S_{x,y}$ far removed from unity. If one is now able to predict the activity coefficients for these two solutes in a number of solvents, one can select the most effective solvent on the basis of its yielding the activity coefficient ratio furthest removed from unity.

As an example, consider the separation of hexene-1 (bp, 63.49°C) from 3-methylpentane (bp, 63.28°C). Because their boiling points and molar volumes are approximately the same, one would expect no separation by a nonpolar liquid phase such as *n*-eicosane or squalane. Indeed, $S_{x,y}$ is very close to unity in both of these solvents at 74.1°C. To find a suitable solvent, we can now employ the results of the factor analysis. At this stage, the best that can be done is to predict which of the remaining six solvents will give the finest separation for the solute pair.

Using the numerical values previously tabulated for n_{ij} and q_{jk}, we can predict [through Eq. (1)] the activity coefficients for the two solutes and [through Eq. (19)] the selectivities in all six solvents. The results, listed in Table 8, indicate that benzyldiphenyl should be the best solvent.

To test these predictions the actual relative volatilities (which should be close to the selectivities) were determined experimentally from GLC retention data [see Eq. (17)]. The $\alpha_{x,y}$ results, also listed in Table 8, agree very well with the predicted $S_{x,y}$

TABLE 8

Predicted Selectivities ($S_{x,y}$) versus Actual Relative Volatilities ($\alpha_{x,y}$) for Hexene-1 (y) and 3-Methylpentane (x) in Six Solvents at 74.1°C

Solvent (k)	$S_{x,y}$	$\alpha_{x,y}$
3	1.06	1.07
4	1.16	1.17
5	1.20	1.21
6	1.28	1.27
7	1.34	1.34
8	1.17	1.16

values. They show that benzyldiphenyl is indeed the most effective liquid phase for separating the given solute pair.

To handle a separation of three or more components, the same technique, but with a few more calculations, can be applied. Although, at present, our scheme contains 8 solvents and 44 solutes, more solvents and solutes can be adapted to the scheme by the procedure that has been demonstrated, thus expanding the number of systems which can be treated. If desired, further expansion can be achieved by carrying out factor analyses on data at other temperatures.

Acknowledgments

This work was supported by the National Aeronautics and Space Administration under Contract NSG-494. The computations were carried out at the Stevens Computer Center, which is supported in part by a grant from the National Science Foundation.

REFERENCES

1. E. R. Malinowski, Ph.D. Thesis, Stevens Institute of Technology, 1962; *Dissertation Abstr.*, **23**(8), Abstr. 62-2027 (1963).
2. K. L. Holzinger and H. H. Harman, *Factor Analysis*, Chicago Univ. Press, Chicago, 1941.
3. L. L. Thurston, *Multiple Factor Analysis*, Chicago Univ. Press, Chicago, 1947.
4. R. B. Catell, *Factor Analysis*, Harper & Row, New York, 1952.
5. D. E. Martire and L. Z. Pollara, *J. Chem. Eng. Data*, **10**, 40 (1965).
6. D. E. Martire, *Anal. Chem.*, **33**, 1143 (1961).

Received by editor July 14, 1966
Submitted for publication August 3, 1966