CHROM. 6644

MAJOR COMPONENTS INFLUENCING RETENTION INDICES IN GAS CHROMATOGRAPHY

SVANTE WOLD and KURT ANDERSSON

Organic Chemistry Department, Umeå University, S 901 87 Umeå (Sweden)

(Received February 2nd, 1973)

SUMMARY

A principal-components analysis of the McReynolds' retention index matrix (10 solutes on 226 liquid phases) showed that in order to reproduce the data within about 30 retention index units, three components have to be introduced in the model. The first component is connected with the polarity of the liquid phase, the second depends almost solely on the solute, and the third is related to interactions with hydroxylic groups in the solute.

The theoretical foundation of principal-components and factor models is discussed. It is shown that the results of a data analysis according to these models cannot be given a unique interpretation. The results of the present analysis are still, however, very useful for the practical classification of liquid phases.

INTRODUCTION

The characterization of the separation properties of liquid phases (LPs) with respect to different types of solutes is of great importance for practical work in gas chromatography. At present, much interest is directed towards classification systems based on the retention indices (RIs) of a few test compounds[1-4]. Moreover, a number of investigators have made efforts to obtain correlations between RIs and the structural and physical properties of the solutes[5-10].

An ideal classification system for the LPs, specifically the RIs of different solutes of these LPs, would indeed be based on easily determined physical properties or easily derived theoretical parameters of the LPs and the solutes. However, the interactions between LPs and solutes are usually of the weak type, such as Van der Waal's forces, charge transfer interactions or hydrogen bonding. At present, weak interactions are much less understood than strong interactions such as those developed in chemical reactions. Hence, the selection of appropriate parameters of LPs and solutes in order to establish the basis of a classification system must at present be founded on a purely empirical search of the trial and error type. For such a search, the methods of principal-components (PC) analysis, and the very similar factor analysis, are parti-

cularly suitable. These methods give an indication of how many components (factors*) are needed to reproduce the observed data within specified limits. Moreover, these analyses give a numerical value that can be assigned to each component, and these values can be used as a basis for a classification system. Alternatively, these component values can be further analysed in terms of physical parameters in order to obtain a physical interpretation of the results.

Weiner and co-workers[5-8] have recently applied factor analysis to a number of RI data matrices**. These authors showed that in order to reproduce the RIs within the experimental accuracy (about three RI units), five to eight components were required, depending on the range of the data***. Such a "complete" analysis is however, in our view, less suitable as a starting point for the construction of a classification system of LPs. Firstly, the large number of components makes the system difficult to handle in practice, and secondly, a "complete" system is, by definition, unable to distinguish between normal behaviour (weak interactions) and abnormal behaviour of an LP with respect to one or several solutes because all types of interactions are in the "complete" system from the beginning.

Accordingly, in an attempt to investigate the possibility of obtaining a classification system for LPs that does not have these drawbacks, we have performed a PC analysis of the RI matrix of McReynolds[3] (10 solutes on 226 LPs)§. In this investigation, we deliberately restricted the number of components in order to obtain results that can be used in practice. As a consequence, we have not considered the problem of solute identification on the basis of correlations between structural and physical properties and RIs, which demands elaborate models based on a complete analysis[5-10]. On the contrary, the explicit purpose of this investigation has been to continue the course set by McReynolds, namely to elucidate similarities and differences between LPs, which are of importance for the selection of LPs in practical separation problems.

PRINCIPAL COMPONENTS (PC) ANALYSIS

(a) The model

The analysis of data according to the PC model corresponds to the assumption that the dependent observed variable y_{ki} (in the present case, the RI of the ith solute on the kth LP) can be linearly related to a number (M) of components (x_{kj}) and loadings (u_{ji})§§. The components are related only to the LPs and the weights§§ are

---

* For simplicity, only the PC nomenclature with components, which correspond very closely with the factors in factor analysis, is used.
** The data for PC and factor analysis are expressed in the form of a matrix where, for example, the data in one row correspond to a single LP and the data in one column correspond to a single solute.
*** During the course of this investigation, Szentirmay et al.[11] published an analysis of the McReynolds' data based on a five-component model.
§ McReynolds published the data in the form of ΔRI values referring to the standard LP of squalane. Our PC analysis is based on the RI matrix corresponding to the McReynolds matrix with the squalane RIs added (see the RESULTS section).
§§ To prevent misunderstandings, we shall use the term weight instead of loading. The statistical loading is not to be confused with the chemical term loading, which is often used for the liquid phase (LP) in gas chromatography.

related only to the solutes. Thus:

$$y_{ki} = x_{k1}u_{1i} + x_{k2}u_{2i} + \cdots + x_{kM}u_{Mi} + e_{ki} = \sum_{j=1}^{M} x_{kj}u_{ji} + e_{ki} \tag{1}$$

The residuals (e_{ki}) are used to describe the part of the data that is not explained by the model. They are assumed to be mutually independent (uncorrelated) between the separate rows (fixed ks) and between the separate columns (fixed is), respectively. Hence, a consistent method for estimating the unknown parameters (x_{kj} and u_{ji}; $k = 1, 2, \ldots, m$; $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, M$) is to simultaneously minimize the row and column sums of squared residuals[12-16]:

$$\sum_{k=1}^{m} e_{ki}^2 = \text{min.} \qquad \text{for } i = 1, 2, \cdots, n$$

$$\sum_{i=1}^{n} e_{ki}^2 = \text{min.} \qquad \text{for } k = 1, 2, \cdots, m \tag{2}$$

In order to make the numerical solution unique, a normalization of each component vector (x_j) is necessary; we have used the usual condition:

$$\sum_{k=1}^{m} x_{kj}^2 = 1 \qquad \text{for } j = 1, 2, \cdots, M \tag{2a}$$

The factor model is the same as eqn. 1, but the assumptions about the residuals (e_{ki}) are somewhat different[16]. With small residuals, however, the resulting parameters (x and u) are virtually the same, and in the following we will not explicitly distinguish between PC and factor analysis. Nevertheless, given the choice of PC and factor analysis, we have preferred the former in view of the more operative specification of the parameters in this model, as is clear from the connection between the PC model and ordinary least-squares theory[12-16].

(b) Relation to approximation theory

Following the treatment of Palm[17], we can derive the PC model from the behaviour of an arbitrary function in a small area around an arbitrary fixed point. Let us first consider a function in two variables only, say $F(z,w)$. A Taylor expansion of the function around the point (z_0, w_0) gives:

$$F(z,w) = F(z_0,w_0) + F_z'\Delta z + F_w'\Delta w + \tfrac{1}{2} F_{zz}''\Delta z^2 + F_{zw}''\Delta z\Delta w + \\ + \tfrac{1}{2} F_{ww}''\Delta w^2 + R(3) \tag{3}$$

Here F_z' denotes the value of $\partial F/\partial z$ at the point (z_0, w_0), the variable Δz denotes $(z - z_0)$, and so on. The remainder, $R(3)$, contains only terms of the third order or higher. Rearranging the terms and borrowing from the third-order terms gives (note that $R(3)$ differs in eqns. 3 and 4):

$$F(z,w) = F(z_0,w_0) - F_z'F_w'/F_{zw}'' + (F_w' + \Delta z F_{zw}'' + \Delta z^2 F_{zz}''F_{zw}''/2F_z')$$
$$(F_z'/F_{zw}'' + \Delta w + \Delta w^2 F_{ww}''/2F_w') + R(3) = C + h(z)g(w) + R(3) \tag{4}$$

In this way, an arbitrary function of two variables can be approximated in a local area to the second order provided that derivatives up to the second order exist and are not zero. Continuing along the same lines, it can be shown (see ref. 18 for a more complete derivation) that a third-order approximation can be obtained by adding a second product term to eqn. 4:

$$F(z,w) = C + h_1(z)g_1(w) + h_2(z)\,g_2(w) + R(4) \tag{5}$$

In general $F(z,w)$ can be approximated by a product sum of functions in the separate variables as

$$F(z,w) = C + \sum_{j=1}^{M} h_j(z)g_j(w) + R(M+2) \tag{6}$$

Turning to functions of several variables, let us first consider the case when N of these variables are grouped together in the vector z, while one of the variables, denoted by w, is treated separately. For a function of these variables, a second-order approximation is, under the same conditions as before:

$$F(z,w) = f(z) + h(z)g(w) + R(3) \tag{7}$$

This can, in turn, be generalized to the case of a function of several variables, N of which are grouped into the vector z and M into the vector w. Around a point (z_0, w_0), this function can be approximated to the second order by the product sum ($M \leqslant N$):

$$F(z,w) = C + \sum_{j=1}^{M} h_j(z)g_j(w) + R(3) \tag{8}$$

In order to establish the relationship between eqns. 4–8 and eqn. 1, we now let the function F correspond to the observed variable y. This observed variable is, at least in chemical applications, a function of several macroscopic variables such as temperature, partial pressures and substrate concentrations. These macroscopic variables in turn influence y via a large number of microscopic variables such as energy levels in distribution functions and electron distributions in molecules. While the macroscopic variables can cover any range, we shall distinguish between situations where the variation of the microscopic variables is either small or large.

If two macroscopic variables, denoted by t and v, are now varied discretely ($t = t_1, t_2, \ldots, t_k, \ldots, t_m$; $v = v_1, v_2, \ldots, v_l, \ldots, v_n$) so that y is always observed at a grid point (t_k, v_l), the observed data (y) can be arranged in the form of a matrix and consequently be treated by means of PC analysis. Thus, we see that $F(t_k, v_l)$ corresponds directly to y_{kl}. However, since we usually do not know whether the variation of t and v corresponds to the variation of only two or to the variation of several microscopic variables, we cannot judge which of models 4–8 is applicable in a particular case. A number of possibilities can be seen, as follows.

(1) The macroscopic variables t and v each affect only one independent microscopic variable. If, in addition, the variation of the microscopic variables (z and w) over the range of t and v is sufficiently limited to make the second-order approximation of eqn. 3 valid with $R(3) \approx 0$, then eqn. 4 will be a good approximation. This model, translated to the form of a PC model, rewriting $h(z_k)$ as x_k and $g(w_l)$ as u_l, gives

$$y_{kl} = C + x_k u_l + e_{kl} \tag{9}$$

In this case, there is obviously a strong connection between the macroscopic and microscopic variables; hence t and v will be called fundamental variables[18]. It is readily seen that this first case includes the case when t and v each influence a number of microscopic variables proportionally, all of a limited range, which can be expressed

as the variation of only one independent microscopic variable per macroscopic variable.

(2) One macroscopic variable, denoted by t, affects several independent microscopic variables, while the second macroscopic variable, v, affects only one. Furthermore, as in case (1), the variation of the microscopic variables over the experimental range of t and v is so small that the second-order approximation is good. The PC model will still be of the one component type, but with variable location; from eqn. 7 we get:

$$y_{kl} = b_k + x_k u_l + e_{kl} \tag{10}$$

The macroscopic variable v is still fundamental, while t is not. This case is very common with linear free energy relationships (LFERs); the Hammett equation is of this type, with the fundamental variable v corresponding to the variation of substituent[18].

(3) The macroscopic variables t and v still affect only one independent microscopic variable each, but to such an extent that the second-order approximation is no longer good. Hence the pertinent model is eqn. 6, which, translated into the PC form, will give a polycomponent model:

$$y_{kl} = C + \sum_{j=1}^{M} x_{kj} u_{jl} + e_{kl} \tag{11}$$

(4) The macroscopic variables t and v both affect several independent microscopic variables. Independently of whether the second-order approximation is valid or not, eqn. 8 or its expansion will lead to a polycomponent PC model, as given by eqn. 11.

It is therefore phenomenologically (from the behaviour of the data) impossible to distinguish between cases (3) and (4). If the PC analysis of a certain data matrix shows that the adoption of a polycomponent model is necessary, the reason can be either (i) that the variation in the two macroscopic variables corresponding to the axes of the matrix affects only two independent microscopic variables, but to such an extent that the second-order approximation breaks down; (ii) that a number of independent microscopic variables are influenced by each macroscopic variable; or (iii) a mixture of these two cases. This shows that a certain amount of care is necessary in the interpretation of the results of a polycomponent PC analysis.

Computational procedures

The problem of finding the numerical values of the components (x) and weights (u) which fulfill the least-squares criteria (2) for a specified data matrix (y) can be solved by means of several procedures[12–15]. We have used the NIPALS method developed by Wold and co-workers[13–16], programmed in Ansi-Fortran on the CD 3300 computer of Umeå University Computer Centre. This method computes one component vector with corresponding weight vector at a time, which makes it possible to inspect intermediate results obtained with arbitrarily many components. The problem how many components should be included in the model in order to obtain satisfactory agreement between the data and the model has been treated here in a rather *ad hoc* manner, which is discussed below.

DATA

The matrix of observed data (y_{kt}) used in this investigation consists of the retention indices (RIs) of 10 different solutes on 226 different liquid phases (LPs) (see Table I). The data were published by McReynolds[3] with the aim of being representative both of types of solutes and of LPs in practical use.

TABLE I

COMPONENT VALUES z_1, z_2 AND x_3 FOR LPs 1–226 (INDEX k), SD OF RIs FOR EACH LP, INDEX (i) FOR SOLUTE HAVING LARGEST DEVIATION AND VALUE ($e_{max.}$) OF THIS DEVIATION

Blank rows indicate abnormal LPs; see text.

k	LP	z_1	z_2	x_3	SD(RI)	i	$e_{max.}$
1	Squalane	−0.10212	0.06199	0.02716	127.5	4	−19.40
2	Squalane	−0.10154	0.06069	0.02775	127.1	4	−19.09
3	Hexatriacontane	−0.10330	0.08744	0.03303	129.3	4	−23.31
4	Nujol	−0.10144	0.07592	0.03173	127.7	4	−20.90
5	Mineral oil	−0.10119	0.07791	0.03410	127.6	4	−21.78
6	Liquid paraffin	−0.10168	0.08287	0.03362	128.2	4	−21.58
7	Convoil 20	−0.09875	0.07761	0.03074	126.0	4	−19.69
8	Apiezon M	−0.09921	0.12176	0.04426	128.3	8	−22.87
9	Apiezon L	−0.09979	0.13216	0.04647	129.2	8	−24.32
10	Apiezon L, treated	−0.09875	0.12248	0.04403	128.2	8	−23.68
11	Polybutene 32	−0.09325	0.08111	0.04346	123.6	8	−20.20
12	Montan wax	−0.08939	0.03710	−0.00824	119.4	4	−29.28
13	Polybutene 128	−0.09578	0.10681	0.05216	126.4	8	−24.84
14	Apiezon L	−0.09763	0.12446	0.04438	127.5	8	−24.75
15	DC-330	−0.08067	−0.01412	0.02295	110.8	7	−12.35
16	SF-96	−0.08033	−0.01615	0.02080	110.7	7	−12.28
17	Apiezon J	−0.09447	0.12208	0.04628	125.6	8	−24.17
18	Apiezon N	−0.09327	0.11615	0.04283	124.6	8	−23.89
19	SE-30	−0.08051	−0.00697	0.02578	111.2	7	−11.93
20	E-301	−0.08015	−0.00757	0.02291	111.0	7	−12.32
21	OV-1	−0.08041	−0.00450	0.02440	111.2	7	−11.85
22	UC L 46	−0.07992	−0.00914	0.02245	110.6	7	−12.23
23	SE-31	−0.08020	−0.00582	0.02647	111.1	7	−12.86
24	W-982	−0.08013	−0.00537	0.02493	111.0	7	−12.08
25	SE-33	−0.08013	−0.00508	0.02662	111.0	7	−12.22
26	M and B silicone oil	−0.07863	−0.01844	0.02328	109.8	7	−12.68
27	DC-200	−0.07926	−0.01244	0.02318	110.2	7	−12.41
28	OV-101	−0.07945	−0.00955	0.02381	110.4	7	−11.97
29	DC-410	−0.07963	−0.00413	0.02564	110.6	7	−12.38
30	DC Silastic 401	−0.07909	−0.00948	0.02403	110.2	7	−12.65
31	Versilube F-50	−0.07874	−0.00798	0.02713	109.9	7	−10.81
32	DC-11	−0.07375	−0.03725	−0.00121	106.5	7	−13.91
33	DC-510	−0.07567	−0.00204	0.03501	108.3	7	−9.34
34	SE-52	−0.07454	0.01093	0.03970	108.2	7	−8.50
35	SE-54	−0.07456	0.01373	0.03980	108.2	7	−8.33
36	DC-560	−0.07319	0.00426	0.04008	106.9	3	7.93
37	DC-556	−0.06987	0.00381	0.05108	105.1	8	13.87
38	Butyl stearate	−0.06569	−0.02885	−0.02070	100.3	6	9.67
39	OV-3	−0.06980	0.02320	0.04981	106.2	1	−5.74
40	Beeswax	−0.06608	−0.00547	−0.01569	105.5	5	35.05

TABLE I *(continued)*

k	LP	z_1	z_2	x_3	$SD(RI)$	i	$e_{max.}$
41	Fluorolube HG 1200	−0.05836	−0.04549	0.09663	100.0	7	−32.37
42	Kel F wax	−0.05945	−0.03068	0.09749	100.1	7	−31.03
43	Apiezon H	−0.06492	0.01328	0.07609	110.0	5	23.34
44	Butoxyethyl stearate	−0.05816	−0.03128	−0.02373	95.5	5	−7.73
45	Halocarbon wax	−0.05813	−0.03748	0.09817	99.6	7	−31.29
46	OV-7	−0.06138	0.05345	0.06924	103.1	1	−8.75
47	DC-550	−0.06005	0.06020	0.07553	102.5	1	−9.14
48	Apiezon W	−0.06330	0.08951	0.03014	107.1	8	−27.74
49	Dinonyl sebacate	−0.04670	−0.05448	−0.03085	89.2	5	−17.43
50	Octoil S	−0.04696	−0.04483	−0.02999	89.4	5	−14.75
51	Dioctyl sebacate	−0.04645	−0.04826	−0.03025	89.0	5	−15.33
52	Diethex sebacate	−0.04638	−0.04745	−0.02937	89.2	5	−13.93
53	DC-703	−0.05647	0.04921	0.07532	99.8	1	−11.60
54	DC-702	−0.05610	0.04887	0.07540	99.5	8	11.90
55	DC-550	−0.05785	0.06682	0.07801	101.5	1	−8.76
56	Diisodecyl adipate	−0.04459	−0.05312	−0.02630	88.2	5	−15.45
57	DINA	−0.04345	−0.05583	−0.02701	87.2	5	−17.40
58	Ditridecyl phthalate	−0.04747	−0.01998	0.01005	92.4	4	13.82
59	Diethex tetrachlorophthalate	−0.05026	0.00644	0.06102	94.2	1	21.75
60	DEG stearate	−0.03939	−0.08137	−0.06180	91.5	5	42.38
61	Octyldecyl adipate	−0.04218	−0.05392	−0.02780	86.0	5	−17.24
62	Dilauryl phthalate	−0.04629	−0.02050	0.01125	92.7	9	−13.44
63	Diisooctyl adipate	−0.04069	−0.06443	−0.03080	85.6	5	−18.63
64	TMP tripelargonate	−0.04149	−0.04427	−0.02255	86.3	5	−14.16
65	Diisooctyl adipate	−0.03805	−0.07204	−0.02978	84.1	5	−19.23
66	Diisodecyl phthalate	−0.04044	−0.03613	0.01046	88.3	4	17.35
67	OV-11	−0.05345	0.10475	0.09643	101.6	1	−9.36
68	Dinonyl phthalate	−0.03541	−0.05974	0.00789	85.3	4	20.72
69	Triton X-400	−0.02802	−0.11803	−0.27417	98.3	5	−24.05
70	Triethex phosphate	−0.02439	−0.13948	−0.20412	88.8	5	−35.16
71	DC-710	−0.05046	0.10394	0.09994	100.2	1	−10.80
72	Flexol GPE	−0.03309	−0.05608	−0.03428	82.2	5	−18.73
73	Dioctyl phthalate	−0.03587	−0.03940	0.01167	86.0	4	20.28
74	Diethex phthalate	−0.03609	−0.03723	0.01159	86.2	4	20.39
75	Dioctyl phthalate	−0.03567	−0.03508	0.01212	85.7	4	18.13
76	Hallcomid M-18	−0.02721	−0.10892	−0.15340	86.0	5	−26.30
77	Diisooctyl phthalate	−0.03277	−0.05031	0.01005	84.4	4	20.21
78	Buoctyl phthalate	−0.03320	−0.04132	0.01198	84.7	4	21.52
79	OV-17	−0.04900	0.12505	0.10706	100.3	6	−10.93
80	Hallcomid M-18 OL	−0.01572	−0.13151	−0.09774	78.1	9	49.20
81	Flexol 8N8	−0.02118	−0.09635	−0.07548	79.6	5	−25.21
82	SP-392	−0.04541	0.13845	0.11610	98.5	6	−10.10
83	Span 60	−0.01685	−0.11649	−0.08080	85.0	5	46.08
84	Versamid 930	−0.03096	0.01635	−0.17131	94.3	8	−30.75
85	Hercoflex 600	−0.02258	−0.06686	−0.02224	77.2	5	−20.04
86	Versamid 930	−0.03002	0.01727	−0.17236	93.8	8	−32.26
87	Versamid 940	−0.02951	0.01384	−0.17263	93.4	8	−31.87
88	Zinc stearate						
89	Ucon LB-550-X	−0.02426	−0.04051	−0.05955	79.0	6	−11.55
90	Span 80	−0.01547	−0.09648	−0.06704	83.7	5	43.67
91	UCON 50-HB-1800-X	−0.02406	−0.02739	−0.05706	79.4	6	−13.14
92	Castorwax	−0.01966	−0.05295	−0.05481	81.1	4	−28.45
93	Flexol B-400	−0.01878	−0.05657	−0.06496	76.9	8	12.71

TABLE I *(continued)*

k	LP	z_1	z_2	x_3	SD(RI)	i	$e_{max.}$
94	OV-22	−0.04344	0.19518	0.13211	100.8	6	−16.84
95	Triton X-200	−0.01112	−0.09641	−0.04165	81.5	4	23.52
96	PPG-2000	−0.01743	−0.04656	−0.06746	76.7	8	10.90
97	Estynox	−0.02044	−0.01294	−0.01965	79.0	4	12.01
98	Trimer acid						
99	Pluracol P-2010	−0.01718	−0.04520	−0.06717	76.6	8	11.16
100	Atpet 200	−0.01037	−0.09062	−0.06283	82.7	5	45.10
101	UCON LB-1715	−0.01472	−0.05160	−0.06187	75.8	8	10.86
102	Dibutoxyet adipate	−0.00957	−0.07861	−0.03096	71.0	5	−18.53
103	Thanol PPG-1000	−0.01087	−0.06526	−0.08149	75.2	8	11.20
104	Acetyltribu citrate	−0.00764	−0.09575	−0.01306	71.7	5	−23.26
105	Diethex phthalate	−0.01235	−0.05390	0.01470	75.9	4	26.40
106	Didecyl phthalate	−0.01221	−0.05403	0.01326	75.6	4	26.20
107	Elastex 50-B	−0.01530	−0.02309	0.01513	77.8	4	24.81
108	Dicyclohexyl phthalate	−0.01819	0.01062	0.01660	80.0	4	21.93
109	OV-25	−0.03996	0.22704	0.14506	101.2	6	−19.97
110	Pluronic L81	−0.01211	−0.03443	−0.06552	75.7	6	−11.86
111	OS-124	−0.03020	0.17007	0.11252	90.8	4	−11.99
112	Tributyl citrate	0.00009	−0.12007	−0.01465	71.3	8	14.07
113	GE SR 119	−0.02517	0.12910	0.09950	91.3	5	19.17
114	OS-138	−0.02953	0.18513	0.11473	91.9	4	−12.48
115	Diethoxyet sebacate	−0.00072	−0.08954	−0.04106	70.4	8	14.01
116	Dibutoxyet phthalate	−0.00551	−0.04869	0.00899	73.9	4	19.78
117	Dibutoxyet phthalate	−0.00318	−0.04693	0.00809	73.0	4	21.26
118	Tri(butoxyethyl) phosphate	0.01121	−0.15537	−0.15041	76.9	4	23.56
119	Zonyl E-91						
120	NPG sebacate	0.00344	−0.03955	−0.02919	76.6	5	15.68
121	Squalene	0.01082	−0.10641	−0.03643	77.2	5	30.82
122	UCON 50-HB-280-X	0.00481	−0.04290	−0.06221	71.6	8	12.72
123	Polytergent J-300	0.00873	−0.07532	−0.07850	72.8	8	6.66
124	Tricresyl phosphate	−0.00102	0.01070	0.00141	75.2	4	22.46
125	SAIB	0.00978	−0.07571	−0.01139	69.2	5	−26.90
126	Paraplex G-25	0.00203	0.00066	−0.00680	73.3	4	10.29
127	Ethomeen 18/25	0.01058	−0.06247	−0.08872	74.3	6	−4.07
128	Polytergent J-400	0.01028	−0.05728	−0.07153	72.6	4	8.17
129	Oronite NIW	0.00983	−0.03860	−0.05124	72.6	6	−5.48
130	QF-1						
131	PPG sebacate	0.00705	−0.00350	−0.01234	72.5	4	7.99
132	UCON 50-HB-660	0.00919	−0.02531	−0.05758	71.6	8	12.58
133	OV-210						
134	UCON 50-HB-3520	0.00836	−0.01036	−0.05294	72.0	6	−13.67
135	Ethofat 60/25	0.01376	−0.05299	−0.06090	72.2	4	7.01
136	Ethomeen S125	0.01392	−0.04980	−0.08492	74.4	6	−5.06
137	Igepal CO-360	0.01400	−0.03991	−0.04867	72.5	8	5.68
138	LSX-3-0295						
139	Pluronic P-85	0.01193	−0.01926	−0.05470	72.4	6	−13.59
140	Pluronic P-65	0.01361	−0.02179	−0.05535	72.3	6	−12.29
141	Tergitol NPX						
142	UCON 50-HB-2000	0.01432	−0.02793	−0.05560	71.7	8	12.66
143	Cresyl diphenyl phosphate	0.00861	0.02628	0.01509	73.4	4	20.75
144	Emulphor ON-870	0.01565	−0.03539	−0.05892	72.6	4	9.48
145	Emulphor ON-870	0.01604	−0.03784	−0.06046	72.7	4	9.14
146	Polytergent B-350	0.01609	−0.02823	−0.04815	72.4	6	−6.38
147	Pluronic L35	0.01663	−0.02650	−0.06436	72.3	6	−11.33

TABLE I *(continued)*

k	LP	z_1	z_2	x_3	$SD(RI)$	i	$e_{max.}$
148	Polytergent G-300	0.01899	−0.03798	−0.04891	72.1	8	7.58
149	Igepal CO-710	0.01835	−0.03142	−0.04671	72.4	6	−5.68
150	Triton X-100	0.01949	−0.03958	−0.04823	72.3	8	7.42
151	Polyglycol 15-200	0.01800	−0.02726	−0.06371	72.1	6	−11.37
152	Stepan DS 60						
153	Diethoxyet phthalate	0.02263	−0.03305	0.01393	71.4	4	23.92
154	UCON 50-HB-5100	0.02458	−0.04344	−0.05280	72.3	8	11.99
155	Siponate DS-10						
156	Renex 678	0.02293	−0.01412	−0.04293	73.2	6	−9.18
157	Igepal CO-730	0.02287	−0.01242	−0.04409	72.7	6	−8.89
158	XE-60	0.03335	−0.08987	0.04141	81.8	4	47.27
159	OV-225	0.02327	0.02816	0.07712	81.9	4	38.26
160	Bis(ethoethoet) phthalate	0.02904	−0.01881	0.00360	71.9	4	22.68
161	NPGA	0.03468	−0.03787	−0.01650	76.5	6	10.48
162	NPGA	0.03700	−0.04497	−0.01633	78.1	5	14.44
163	UCON 75-H-90000	0.02966	0.01905	−0.04092	73.5	4	15.49
164	Pluronic F-88	0.03308	0.01826	−0.03654	74.7	6	−16.36
165	Igepal CO-880	0.03523	0.00542	−0.03779	74.7	4	11.92
166	Surfonic N-300	0.03540	0.00830	−0.03565	74.7	4	12.05
167	Pluronic F-68	0.03459	0.01613	−0.03704	75.0	4	16.62
168	Triton X-305	0.03675	0.00446	−0.03827	75.1	6	−13.07
169	HI-EFF 8 BP	0.03827	0.03086	0.01347	78.8	3	−14.18
170	CHDMS	0.04057	0.01541	0.00766	81.1	5	23.66
171	CW-4000 monostearate	0.04151	0.01703	−0.03720	76.7	4	16.24
172	Zonyl E-7						
173	Paraplex G-40	0.04424	0.02232	0.01630	75.3	5	−15.82
174	CW-4000 monostearate	0.06298	−0.16278	−0.03398	75.2	10	−40.93
175	Quadrol	0.06658	−0.17223	−0.20040	99.2	1	−33.03
176	NPGS	0.05106	−0.02102	0.01298	80.2	4	14.12
177	NPGS	0.05155	−0.02327	0.01167	80.4	4	13.33
178	NPGS	0.05330	−0.02369	0.01439	82.6	1	−12.43
179	Igepal CO-990	0.04745	0.03429	−0.03030	78.5	4	16.94
180	EGSP-Z	0.04875	0.10746	0.07256	86.1	5	28.18
181	Carbowax 20M	0.05456	0.05780	−0.02486	81.4	6	−18.57
182	Carbowax 20M TPA	0.05616	0.04834	−0.02358	83.5	3	−21.13
183	Epon 1001						
184	Carbowax 6000	0.05630	0.04991	−0.02686	82.2	3	−19.41
185	MER-21	0.05580	0.05586	−0.02924	81.9	3	−18.43
186	PEG 4000	0.05921	0.04433	−0.03480	82.9	3	−17.97
187	Ethylene glycol isophthalate	0.06436	0.05479	0.07256	90.1	8	−13.09
188	XF-1150	0.07767	−0.05382	0.07001	100.7	4	52.78
189	Sorbitol hexa-acetate	0.08275	−0.04708	0.03105	88.7	5	−39.46
190	FFAP	0.07585	0.00477	−0.02379	100.4	5	56.50
191	STAP	0.07694	0.01239	−0.02211	99.8	5	50.20
192	Carbowax 1000	0.07849	0.01722	−0.04902	92.4	3	−13.63
193	Sucrose octa-acetate	0.08570	−0.01599	0.03770	93.4	5	−34.09
194	MER-2	0.06886	0.15538	0.11545	92.1	9	17.25
195	PEG-600	0.08509	−0.00433	−0.07478	95.9	5	11.01
196	Butanediol succinate	0.07955	0.06020	0.03666	93.3	3	−11.02
197	EGA	0.08090	0.06295	0.03813	92.8	3	−10.38
198	EGA	0.08188	0.05882	0.03964	93.5	3	−9.92
199	Ethylene glycol adipate	0.08436	0.04251	0.03785	96.1	5	15.66
200	Butanediol succinate	0.08601	0.03526	0.01670	96.6	6	10.78
201	PDEAS	0.07617	0.14369	0.11044	102.2	8	−27.05

TABLE I *(continued)*

k	LP	z_1	z_2	x_3	SD(RI)	i	$e_{max.}$
202	Reoplex 400	0.09329	−0.01134	−0.01877	105.0	5	46.10
203	LAC IR-296	0.08912	0.03306	0.01879	99.7	5	26.37
204	DEG adipate	0.09029	0.02894	0.01942	100.2	5	26.67
205	Carbowax 1540	0.09190	0.00886	−0.03631	99.9	5	16.00
206	Resoflex R 296	0.09179	0.02902	· 0.01651	101.8	5	30.85
207	LAC-2-R-446	0.09267	0.04365	0.01905	101.5	5	22.92
208	EGSS-Y	0.09290	0.04908	0.07757	101.3	9	10.91
209	Hyprose SP-80						
210	ECNSS-M	0.12603	−0.02046	0.04825	123.3	3	26.04
211	Diglycerol						
212	DEGS Supelco 1045	0.12489	0.04407	0.05756	119.3	4	−22.04
213	EGSS-X	0.12619	0.09655	0.09269	121.2	9	16.26
214	DEGS	0.12867	0.09588	0.06039	121.8	3	−14.29
215	Ethylene glycol phthalate						
216	DEGS Supelco 1303	0.13622	0.07212	0.05860	129.9	5	30.41
217	DEGS	0.14025	0.05655	0.06474	135.3	5	46.14
218	LAC-3-R-728	0.13931	0.07416	0.06221	133.4	5	35.93
219	Glycol succinate	0.14740	0.10420	0.09936	·136.2	9	19.28
220	THEED						
221	Tetracyanoethoxy PE	0.15595	0.02558	0.09810	137.7	3	27.00
222	EGS	0.15345	0.08163	0.09596	142.0	4	−24.61
223	TCEP	0.17967	0.05746	0.13363	156.2	5	−39.33
224	TCEP	0.17925	0.07195	0.13298	155.6	5	−40.50
225	Cyanoetylsucrose	0.19052	0.11065	0.12488	160.4	3	33.14
226	BCEF						

The RI of the ith solute on the kth LP is defined by the corresponding retention time T_{ki} in conjunction with the retention times of two consecutive normal alkanes, a_{kn} and $a_{k(n+1)}$, on the same LP. The definition is based on the empirical fact that the logarithmic retention times of alkanes on a specific LP vary almost linearly with the number of carbon atoms (n) in the alkanes; this relationship has the following form[1]:

$$RI_{ki} = 100n + 100(\log T_{ki} - \log a_{kn}) / (\log a_{k(n+1)} - \log a_{kn}) \qquad (12)$$

Thus, in fact, the RIs are pre-processed data that mainly express the relationships between the behaviour of solutes and normal alkanes on different LPs. Since the retention times of the alkanes themselves (a_{kn}) are influenced by different combinations of interactions for different LPs, this pre-processing obscures the absolute causes of the solute behaviour. In our view, this fact makes RI data less suitable for the theoretical analysis of LPs and/or solutes, if an exhaustive classification is aimed at, and especially so if efforts are made to relate the classification to structural and physical parameters of the LPs and solutes. However, RI data remain suitable for the practical classification of LPs with respect to the overall separation power for different types of solutes. In such an application, the emphasis is placed on differences in practical behaviour and not on absolute causes; the pre-processing obscures the latter, but not the former. Hence, in the present investigation, the RI data of McReynolds have been used with the specific condition that the numerical results should not be further analysed in terms of physical concepts but be used only as an aid in practical separation problems.

RESULTS

Several PC analyses were performed on both the total (226 × 10) data matrix and on several different smaller submatrices.

(a) To study the effects of transformations of the data, PC analyses were made according to eqn. 1 with the dependent variable (y_{kl}) given by different transformations of the RI values as defined by eqn. 12:

$$\text{(i)} \quad y_{kl} = RI_{kl} \tag{13}$$

$$\text{(ii)} \quad y_{kl} = \log RI_{kl} \tag{14}$$

$$\text{(iii)} \quad y_{kl} = RI_{kl} - RI_{1l} = \varDelta RI_{kl} \tag{15}$$

$$\text{(iv)} \quad y_{kl} = \log(RI_{kl} - RI_{1l}) = \log \varDelta RI_{kl} \tag{16}$$

Here RI_{1l} denotes the RI value of the lth solute on the LP of squalane, the original standard of McReynolds[3] (LP No. 1 in Table I), who actually tabulated $\varDelta$RI values corresponding to eqn. 15. However, the use of these $\varDelta$RI values as y_{kl} (eqns. 15 and 16) in the PC analyses gave inconsistent results with much larger residuals for LPs of large polarity than for those of low polarity (high and low k indices, respectively, in Table I). The logarithmic transformation (eqns. 14 and 16) gave the same relative goodness of fit as the non-transformed data. Hence the RIs were used directly as the dependent variable in further analyses (eqn. 13).

(b) The total RI matrix (226 × 10) was analysed according to a PC model with three components (eqn. 1 with $M = 3$). The standard deviation (SD) of the original data (RI_{kl}) was 190.2 and of the residuals (e_{kl}) 21.1[*]. LPs showing larger residuals than 63 (3 SD_e) were classified as abnormal and set aside for a special analysis.

(c) The remaining "normal" LPs (213 in number) were re-analysed. It was found that two components ($M = 2$ in eqn. 1) described 89% of the variation (SD) in the original RI matrix (213 × 10). However, for two of the solutes, butanol and methylpentanol, the fit was considerably worse than for the other solutes. This lack of fit was removed by the addition of a third component in the model, obviously describing some kind of hydroxylic specificity (i.e., hydrogen-bonding ability) of the LPs. Three LPs that gave abnormally large residuals (> 63) were removed and added to the abnormal group (209, 220 and 226). The remaining 210 LPs were classified as being normal. For these, the three first components described 93% of the original RI variation. The fourth component described less that an additional 3%. Accordingly, the three first components with corresponding weights have been taken to be sufficient for a classification of normal LPs. The SD of the residuals is 14.3, which means that with a three-component PC model, the residuals are smaller than 30 RI units in about 95% of the normal cases.

The resulting component and weight vectors can be represented in infinitely

---

[*] The SD was estimated according to the standard equation[19,20]:

$$SD_y = \left[\sum_k \sum_l (y_{kl} - \bar{y})^2 / (N - P) \right]^{\frac{1}{2}}$$

where $\bar{y}$ is the total mean value of y and P is the number of parameters.

many ways because any rotation of the following type is equivalent to the original representation (x_1, x_2 and x_3 are the original components and z_1, z_2 and z_3 the rotated components):

$$z_1 = a_1x_1 + a_2x_2 + a_3x_3$$
$$z_2 = a_2x_1 - a_3x_2 - a_1x_3$$
$$z_3 = a_3x_1 + a_1x_2 - a_2x_3$$
$$a_1^2 + a_2^2 + a_3^2 = 1 \qquad (17)$$

The weights are, of course, rotated according to the new weights, denoted by v. Since, in the present investigation, the third component (x_3) was connected directly to hydrogen bonding ability, this component was left untransformed. The first two components, however, both increased with polarity (x_1 and x_2). Hence, linear combinations ($z_1 = a_1x_1 + a_2x_2$ and $z_2 = a_2x_1 - a_1x_2$; $a_1^2 + a_2^2 = 1$) were formed having the property that the first (z_1) was maximally correlated with x_1 and x_2. The resulting final components are shown in Table I (z_2 being transformed by the subtraction of its mean value, 3.15583, to show better the variation with the LP) together with the row SDs of the original RI data (corresponding to the variation of the RIs for each specific LP) and the index and the value of the largest residual for each LP. The corresponding weights (v_1, v_2 and u_3) are given in Table II. These tabulated values therefore describe the observed RI data by the model[*](in the "normal" cases):

$$RI_{kl} = z_{k1}v_{1l} + (z_{k2} + 3.15583) v_{2l} + x_{k3}u_{3l} + e_{kl} \qquad (18)$$

(d) The abnormal LPs (16 in number) were re-analysed with the result that a PC model was of little relevance for these data. This is probably due to specific strong LP–solute interactions in some cases (at least one strong interaction for each

TABLE II

WEIGHT VALUES CORRESPONDING TO THE FIRST TWO ROTATED COMPONENTS (v_1 AND v_2) AND THE THIRD UNROTATED COMPONENT (u_3)

Index i denotes solute number.

i	Solute	v_1	v_2	u_3
1	Benzene	2070	264.4	133.2
2	1-Butanol	3193	289.1	−462.0
3	2-Pentanone	2502	274.1	151.4
4	1-Nitropropane	3487	318.0	157.7
5	Pyridine	3389	321.0	95.7
6	2-Methyl-2-pentanol	2522	295.7	−313.1
7	1-Iodobutane	1785	309.9	25.8
8	2-Octyne	1305	306.0	8.9
9	1,4-Dioxane	2977	293.0	239.8
10	cis-Hydrindane	986	343.7	− 36.2
Percent of total SD described by component		65	24	3.5

[*] For example, for LP 55 and solute 2 (1-butanol), the RI calculated by eqn. 18 becomes RI_{calc} = −0.05785 × 3193 + (0.06682 + 3.15583) × 289.1 − 0.07801 × 462.0 = 710.9, which can be compared with the observed value of 714, leaving a residual ($RI_{obs} - RI_{calc}$) of 3.1 RI units.

abnormal LP). If these strong interactions are of different types for different LPs and solutes, it is a direct consequence that the PC analysis is meaningless. Table III shows the larger residuals for these abnormal LPs from analysis (b) described above, and can be used if an LP with specific interactions for some type of solute is searched for. It should be noted that the only abnormal LPs in common use are QF-1 and OV-210 (numbers 130 and 133, respectively). These two LPs have abnormally large RIs for ketones (large positive residuals for solute 3 in Table III), which indicates strong interactions with carbonyl groups. In fact, this is a property that has made these two LPs popular in steroid analysis.

TABLE III

LARGEST RESIDUALS FOR "ABNORMAL" LPs (INDEX k) FROM THREE-COMPONENT PC ANALYSIS OF THE TOTAL (226 × 10) DATA MATRIX

Names of LPs as in Table I.

k	i									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
88				−46	288	−53				86
98		−55		−66	120					
119	−46		66			44	−44			
130	−59		66					−41		39
133	−60		66					−40		39
138	−60		66					−44		41
141								−66		32
152	−78		59						94	
155	−80		71		−59				105	
172	−57		79		50	−61	52			
183	−36			−39	68					
209	−40			−65	62					
211				−100	62				70	
215					88					
220				−77	44			−49		
226	44		53		−67					

DISCUSSION

A three-component PC model describes 93 % of the variation of the RIs of the majority (210 of 226) of the LPs investigated by McReynolds[3]. As already discussed, the incorporation of more components would make the model difficult to use in practice and also of little meaning, since the improvement in fit given by the fourth and following components is very small. On the other hand, the adoption of a two-component model would have resulted in the classification of all hydrogen-bonding LPs as abnormal, which, in our view, is less desirable. Hence, we feel that the three-component PC model is a reasonable compromise between generality and practicality, having most of the desirable features of a classification system and few undesirable drawbacks.

Despite the difficulty in interpreting the numerical results of the PC analysis, we might make the following tentative assessment of the component values:

The first, most important, component (z_1 in Table I) corresponds well with

the commonly used "polarity" concept and will accordingly be called the polarity component. The corresponding weight (v_1 in Table II) describes the sensitivity of the solutes to this polarity component; the larger the polarity of the solute, the larger is the v_1 value. The first component accounts for 65 % of the variation in the RI matrix.

The second component (z_2) with corresponding weight (v_2) is harder to assess. It is almost constant (3.15583 ± 0.20). The largest deviation from the mean value (3.15583) occurs for $k = 94$ (OV-22, $z_2 = 0.19518 + 3.15583$) and gives a contribution of 67 RI units for $i = 10$ (cis-hydrindane, $v_2 = 343.7$), apart from the mean contribution of 1085 RI units (3.15583 × 343.7). This component accounts for about 24 % of the variation of the original RI matrix.

The third component (x_3) describes, as indicated above, an interaction that is large only for alcoholic solutes (note that neither primary nor secondary amines are included among the solutes). The corresponding weight (u_3) has large negative values only for alcoholic solutes; accordingly, the third component will be called the hydrogen-bonding ability of the LP, and the third weight hydrogen bond donating ability. Large negative values of x_3 will correspond to a large ability of the LP. The largest interaction is 127 RI units, namely for $k = 69$ (Triton X-400, $x_3 = -0.27417$) and $i = 2$ (1-butanol, $u_3 = -462.0$). The total contribution to the RI variation is about 3 %.

In addition to the three components z_1, z_2 and x_3, there is a property of the LPs that is of importance for the separation ability, namely, the standard deviation of the RIs for a particular LP (denoted SD(RI) in Table I). Thus, a large SD(RI) for a specific LP will indicate a good separation ability for this LP, regardless of the component values. It is interesting that this parameter, SD(RI), is large both for LPs with very low polarity and for LPs with very high polarity. This implies that in most cases (separation of non-hydrogen-bonding solutes) the first 10 and last 10 LPs are the most suitable, unless an abnormal LP can be found that is particularly suitable for the actual separation problem.

One of the more significant results of the present analysis is, in our view, the division of the LPs into the two classes of normal and abnormal LPs. The latter class can profitably be searched for LPs with specific interactions suitable in particular separation problems, as exemplified by the two LPs 130 and 133 (QF-1 and OV-210), which are well suited for separation of ketones (large positive residuals for solute 3 in Table III).

To conclude, we wish to emphasize once more that the resulting component and weight values of the present analysis must not be strictly interpreted in terms of physical concepts such as polarity or hydrogen-bonding ability. A PC analysis can seldom, if ever, due to the general approximation properties of the PC model shown in eqns. 9–11 and due to the non-uniqueness of the results with respect to rotation, give results that can be uniquely interpreted, neither in terms of microscopic terms such as "polarity" or hydrogen-bonding ability, nor in terms of macroscopic physical properties such as boiling points or dipole moments. Hence, the component names are introduced more as tentative indicators of the probable main causes that underlie the component and weight values.

Moreover, since the scope of an empirical model is determined by the analysed data we wish to stress the "local properties" of the PC model. Interpretation of results in terms of variables and phenomena outside the framework set by the empirical data

used in the analysis, and predictions of the outcome of measurements that differ considerably from those represented in the data framework, are of little meaning. However, the PC model can be used with great confidence within the data framework, both for predictions and "interpretations" in terms of "pattern recognition" and similar analogy reasoning. This makes PC analysis especially valuable for use in situations such as the one in the present investigation, for which knowledge of the basic physical variables is diffuse or completely lacking but for which one wishes to obtain results of practical utility.

APPENDIX

Classification of new phases

When a phase (LP) not included in the data matrix of McReynolds[3] is to be classified in the same manner as described above, it can be achieved in the following way.

Run a number of test solutes plus some suitable n-alkanes on the new LP. A classification compatible with the one presented here, should be based on the same 10 test solutes as those used by McReynolds. However, for a rough preliminary classification, fewer test solutes can be used, but not less that six, however (see example below). Thereafter, the RIs are calculated by eqn. 12, followed by a standard multiple regression analysis[19,20] with the RIs as the dependent variable and the weights in Table II as the independent variables:

$$RI_{new, i} = c_1 v_{1i} + c_2 v_{2i} + c_3 u_{3i} + e_i \qquad (19)$$

The regression coefficients c_1, c_2 and c_3 are calculated so as to minimize the sum of squared residuals (e_i). When the resulting residuals (e_i) are all less than about 60, these coefficients (c_1–c_3) can be assumed to correspond to the component values z_1, $z_2 + 3.15583$ and x_3 in eqn. 18. Otherwise, the new LP is classified as abnormal.

If larger numbers of new LPs are to be classified (10 or more), it is probably best to make a completely new PC analysis, using the normal RI data of McReynolds, together with the new data[*].

As an illustrative example of the procedure of a preliminary classification of a new LP, we have determined RIs for McReynolds' first six solutes ($i = 1, 2, \ldots, 6$ in Table III) on the LP of halocarbon K-352. These RIs were 700, 660, 700, 890, 845 and 810 RI units, respectively.

A multiple regression analysis according to eqn. 19 gave the coefficient values $c_1 = 0.026$, $c_2 = 2.885$ (3.156 − 0.271) and $c_3 = 0.105$. The second value (c_2) is unrealisticly large when compared with the corresponding value in Table I (z_2). This result, less useful for a classification, arises because too few data were used in the multiple regression, thus making the analysis unstable.

Hence, when fewer than 10 values are used in the preliminary classification, the following modified procedure is recommended. First analyse the data according

---

[*] This can be carried out by sending the new RI data of one of the authors (S.W.) at this laboratory in the form of punched standard cards in the same form as in the McReynolds table (squalane RIs subtracted), one card for each phase with the ten $RI_{new, i}$ values punched in the Fortran format 10 F8.1.

to a two-component model (eqn. 19 with the term c_3x_3 excluded)[*]. In the present example, this gives the more reasonable values $c_1 = 0.0346$ and $c_2 = 2.956$ ($3.156 - 0.200$). The residuals were -10, -84, -24, 71, 13 and 23, respectively. Secondly, if the residuals corresponding to 1-butanol ($i = 2$) and 2-methyl-2-pentanol ($i = 6$) are large and of the same sign, a third component can be included. The data are then analysed according to the full eqn. 19. In the present case, it can be seen that the inclusion of a third component is unwarranted. Furthermore, since the residuals after the two-component analysis are large (two are greater than 60), the LP of halocarbon K-352 is preliminary classified as being abnormal.

Classification of other solutes

Solutes not used in the McReynolds matrix can be classified in a manner similar to the classification of a new phase described above. Thus, the new solutes plus suitable n-alkanes and some standard solutes as checks are run on a representative sample of LPs (minimum 10) having high and low values of all three components. Ordinary multiple regression analysis of the RIs (calculated from eqn. 12) with the component values of Table I (z_1, $z_2 + 3.15583$ and x_3) as independent variables will give regression coefficients corresponding to the weights v_1, v_2 and u_3, respectively (very large residuals for some LP will indicate abnormally strong interactions). However, in our opinion, classification of solutes that differ considerably in type from those used in the McReynolds matrix is better made by means of a totally new analysis with the inclusion of several new representative test solutes. We wish to emphasize once more that the present investigation has been made in order to obtain a classification of LPs, not a classification of solutes.

ACKNOWLEDGEMENTS

REFERENCES

1 E. Kováts, *Helv. Chim. Acta*, 41 (1958) 1915.
2 L. Rohrschneider, *J. Chromatogr.*, 22 (1966) 6.
3 W. O. McReynolds, *J. Chromatogr. Sci.*, 8 (1970) 685.
4 W. R. Supina and L. P. Rose, *J. Chromatogr. Sci.*, 8 (1970) 214.
5 P. H. Weiner and D. G. Howery, *Can. J. Chem.*, 50 (1972) 448.
6 P. H. Weiner, C. J. Dack and D. G. Howery, *J. Chromatogr.*, 69 (1972) 249.
7 P. H. Weiner and J. F. Parcher, *J. Chromatogr. Sci.*, 10 (1972) 612.
8 P. H. Weiner and D. G. Howery, *Anal. Chem.*, 44 (1972) 1189.
9 J. Takács, Zs. Tálas, I. Bernáth and Gy. Czakó, *J. Chromatogr.*, 67 (1972) 203 and references therein.
10 G. Castello and G. D'Amato, *J. Chromatogr.*, 58 (1971) 127.
11 Zs. Szentirmai, G. Tarján and J. Tákacs, *J. Chromatogr.*, 73 (1972) 11.

---

[*] The exclusion of the third component is possible because it is orthogonal to the first two components. It is, however, not correct to use only the first component singly in an analysis, since the first component is not orthogonal to the second owing to the rotation performed on the original two first components.

12 P. Whittle, *Skand. Aktuarietidskr.*, (1952) 223.
13 H. Wold, in *Proceedings of an International Symposium in Multivariate Analysis, Dayton, Ohio, 1965*, Academic Press, New York, 1966.
14 H. Wold, in F. N. David (Editor), *Festschrift for Jerzy Neyman*, Wiley, New York, 1966.
15 A. Christoffersson, *Thesis*, University Institute of Statistics, Uppsala, Sweden, 1970; Summary in
   · *Bull. Internat. Stat. Inst., Proc. 37th Session, London, 1969*, Vol. 1, p. 31.
16 E. Lyttkens, in *Proceedings of an International Symposium in Multivariate Analysis, Dayton, Ohio, 1965*, Academic Press, New York, 1966.
17 V. A. Palm, *Osnovy Kolitchestvennoi Teorii Organitcheskikh Reaktsii*, Izdatelstvo Khimiya, Leningrad, 1967 (German translation: Akademie Verlag, Berlin, G.D.R., 1971).
18 S. Wold, *Chem. Scr.*, to be published.
19 P. D. Lark, B. R. Craven and R. C. L. Bosworth, *The Handling of Chemical Data*, Pergamon Press, London, 1968.
20 N. R. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York, 1966.