CORRESPONDENCE ANALYSIS AND ADSORBATE SELECTION FOR CHEMICAL SENSOR ARRAYS

FERNANDO AVILA

Department of Mathematics, University of Arizona, Tucson, AZ 85721, U.S.A. and Departamento de Matematicas, Universidad de Sonora, Mexico

D. E. MYERS

Department of Mathematics, University of Arizona, Tucson, AZ 85721, U.S.A.

AND

CHRIS PALMER

Department of Chemistry, University of Arizona, Tucson, AZ 85721, U.S.A.

SUMMARY

Carey *et al.* utilized principal components analysis (PCA) to analyze frequency shift data obtained from piezoelectric sensors formed by coating quartz crystals with 27 different GC stationary phases and tested using 14 analytes. The objective of the analysis was to determine an optimal reduced set of coatings for detection of the analytes. The results were correlated with those obtained from cluster analysis. In this paper the data are re-analyzed using correspondence analysis (CA). The advantage of using CA include a symmetric treatment of sensor coatings and analytes and better identification of the representation of the analytes in terms of the detection components. The results obtained by the conjunctive use of PCA, a varimax rotation and cluster analysis were obtained by CA.

KEY WORDS   Correspondence analysis   Cluster analysis   Optimization   Eigenanalysis

INTRODUCTION

Multivariate data sets consist of the sampled (measured) values of a set of variables, often written in matrix form. Usually, columns represent variables and rows represent samples. Multivariate (descriptive) statistical analysis has the goal of obtaining summary descriptions of the data set. This goal can be achieved in various ways; for example, some methods are designed to describe the interrelationships between variables, while others aim at grouping either the variables or the samples on the basis of a similarity measure. The results can then be used to reduce the dimension of the data set without losing essential information.

Two popular multivariate methods are cluster analysis, which is used to classify samples according to a measure of 'closeness' with the results usually shown in a graphical form (a dendogram), and principal components analysis (PCA), which is used to transform a set of correlated variables into a set of uncorrelated variables for the purpose of simplifying the description of the interrelationship between the original variables. Both methods can be used for data reduction, but they are applied under different assumptions and imply a non-symmetrical treatment of samples and variables.

We consider a method, correspondence analysis (CA), that treats rows and columns in a symmetrical fashion although it does not require that row variates and column variates be

similar. We show that it can be used in situations where two different sets of variates label the rows and columns of the input matrix and what is measured is some sort of interaction between the two sets.

In the paper by Carey et al.,[1] PCA followed by a varimax rotation and cluster analysis were used to analyze a data set consisting of the number of frequency shifts obtained from 27 piezoelectric sensors, formed by coating quartz crystal microbalances (QCMs) with 27 different GC stationary phases, tested to detect 14 analytes. The results from both methods were compared and used to categorize the coatings for the purpose of reducing their number.

In this paper we analyze the same data set using CA and obtain results comparable to those obtained by the conjunctive use of PCA, rotation and cluster analysis. CA gives the results for the rows and columns simultaneously, avoiding the need to do separate analysis. CA does not require a rotation of the factor space. CA provides several diagnostics and graphical displays which aid in the interpretation of the results. CA can be used as an alternative, computationally efficient, pattern recognition technique.

THE DATA

The response of a particular coating to a particular analyte is measured as a shift in the fundamental frequency of oscillation of the QCM; the extent of this shift is proportional to the weight gain of the sensor on adsorption of some equilibrium amount of the analyte. The

Table 1. Coatings and their IDs

ID No.	Coating
1	Poly(butadieneacrylonitrile)
2	Poly(p-vinylphenol)
3	Poly(butadiene methacrylate)
4	Polybutadiene hydroxy terminated
5	Poly(vinyl stearate)
6	Poly-1-butadiene
7	Polybutadiene hydroxy terminated liquid
8	Methyl vinyl ether
9	Octadecyl vinyl ether/maleic anhydride
10	Polystyrene
11	Poly(vinyl isobutyl ether)
12	Poly(vinyl chloride)
13	Poly-1-butene
14	Poly(vinylcarbazole)
15	Collodion
16	Poly(vinylbutyral)
17	Poly(methyl methacrylate)
18	Polyethylene
19	Ethyl cellulose
20	Poly(ethylene glycol methyl ether)
21	Poly(caprolactone)
22	Poly(caprolactone)triol
23	Poly(caprolactone)triol 2X
24	Carnuba wax
25	Abietic acid
26	DC 11
27	Phenoxy resin

data are entered as a matrix with 27 rows labelled by the coatings and 14 columns labelled by the analytes.

The coatings and their ID numbers are shown in Table 1.

The analytes are benzene, dodecane, DMMP, DM phosphite, 1-butyl formate, a-pinene oxide, triphenyl phosphite, DIMP, dichloropentane, isopropyl acetate, triamyl phosphite, octane, triphenyl phosphate and water.

PRIOR RESULTS

In their paper, Carey et al.[1] used PCA with the coatings as the variables. After applying a varimax rotation to the eigenvectors, the eight factors shown in Table 2, explaining 94·9% of the variance, were obtained.

After applying a hierarchical cluster analysis, they obtained a good correlation between coatings in the dendogram and in the varimax-rotated factors. On the basis of the results from PCA, they suggested selecting eight coatings—poly(caprolactone) triol, poly(butadiene methacrylate), polybutadiene hydroxy terminated, poly(vinyl isobutyl ether), poly(p-vinylphenol), poly(methyl methacrylate), poly(vinyl chloride) and collodion—as an optimal reduced set of coatings.

Table 2. Factors from PCA after rotation

Factor	ID No.	Coating	Contribution
I	3	Poly(butadiene methacrylate)	27·7
	6	Poly-1-butadiene	19·5
	4	Polybutadiene hydroxy terminated	14·5
	7	Polybutadiene hydroxy terminated liquid	9·6
II	15	Collodion	25·0
	24	Carnuba wax	18·4
	19	Ethyl cellulose	10·4
	25	Abietic acid	10·3
III	14	Poly(vinylcarbazole)	31·5
	16	Poly(vinylbutyral)	19·1
	2	Poly(p-vinylphenol)	17·0
	10	Polystyrene	13·6
IV	19	Ethyl cellulose	34·1
	21	Poly(caprolactone)	23·9
	26	GE DC 11	18·4
V	17	Poly(methyl methacrylate)	39·9
	1	Poly(butadieneacrylonitrile)	9·9
VI	12	Poly(vinyl chloride)	66·5
VII	11	Poly(vinyl isobutyl ether)	84·9
VIII	13	Poly-1-butene	75·1

CORRESPONDENCE ANALYSIS

Correspondence analysis (CA) is a multivariate method which produces a simultaneous graphical representation of the projections of the n rows and p columns of a data matrix

$\mathbf{X} = (x_{ij})_{n \times p}$ onto factorial axes determined by a least squares criterion, using the so-called 'χ^2-metric'. The method is usually applied to the analysis of contingency tables[2-4] but has also been used to analyse geochemical,[5-7] ecological[8] and environmental[9,10] data among other types of data.

CA can be developed from different perspectives;[2,3] we choose a 'geometric' approach along the lines that Stuart[11] suggests for PCA (see also the book by Lebart et al.[3]).

The entries x_{ij} must be non-negative. CA is performed on the normalized matrix $\mathbf{F} = (f_{ij})_{n \times p}$, where

$$f_{ij} = \frac{x_{ij}}{\sum\limits_{j=1}^{p} \sum\limits_{i=1}^{n} x_{ij}}$$

Two diagonal weighting matrices $\mathbf{D}_n = \mathrm{diag}\,(f_{i+})$ and $\mathbf{D}_p = \mathrm{diag}\,(f_{+j})$ are also defined, with

$$f_{i+} = \sum_{j=1}^{p} f_{ij}, \qquad f_{+j} = \sum_{i=1}^{n} f_{ij}$$

The matrices $\mathbf{D}_n$ and $\mathbf{D}_p$ are used to scale $\mathbf{F}$. Think of the rows of $\mathbf{D}_n^{-1}\mathbf{F}$ as vectors in p-dimensional space where the metric is weighted by $\mathbf{D}_p^{-1}$, and of the columns of $\mathbf{F}\mathbf{D}_p^{-1}$ as vectors in n-dimensional space where the metric is weighted by $\mathbf{D}_n^{-1}$. Then the variation of a linear combination of vectors in either space is given by a weighted quadratic form, and the purpose of CA is to find vectors, $\mathbf{u}$ in p-dimensional space and $\mathbf{v}$ in n-dimensional space, having size one in the norms induced by the matrices $\mathbf{D}_p^{-1}$ and $\mathbf{D}_n^{-1}$ and giving the directions of maximum variation.

Mathematically we have the following problems.

1. In 'row space'

$$\text{maximize } \mathbf{u}^{\mathrm{T}}(\mathbf{D}_n^{-1}\mathbf{F}\mathbf{D}_p^{-1})^{\mathrm{T}}\mathbf{D}_n(\mathbf{D}_n^{-1}\mathbf{F}\mathbf{D}_p^{-1})\mathbf{u}$$
$$\text{subject to } \mathbf{u}^{\mathrm{T}}\mathbf{D}_p^{-1}\mathbf{u} = 1$$

2. in 'column space'

$$\text{maximize } \mathbf{v}^{\mathrm{T}}(\mathbf{D}_p^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{D}_n^{-1})^{\mathrm{T}}\mathbf{D}_p(\mathbf{D}_p^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{D}_n^{-1})\mathbf{v}$$
$$\text{subject to } \mathbf{v}^{\mathrm{T}}\mathbf{D}_n^{-1}\mathbf{v} = 1$$

After finding these unit vectors $\mathbf{u}$ and $\mathbf{v}$, we can then search in a sequential way for new solutions orthogonal, with respect to the inner product defined by the weighting matrices, to all previous solutions.

It is easily seen, using Lagrange multipliers, that the CA problem is really an eigenvalue–eigenvector problem. A crucial point, which CA takes advantage of, is that, because of the symmetric scaling done on the rows and columns of the input matrix, there is a relationship between the solutions to the row problem and the solutions to the column problem. This duality is expressed by the equations

$$\mathbf{v} = (\lambda)^{-1/2}\mathbf{F}\mathbf{D}_p^{-1}\mathbf{u}, \qquad \mathbf{u} = (\lambda)^{-1/2}\mathbf{F}^{\mathrm{T}}\mathbf{D}_n^{-1}\mathbf{v}$$

where λ is an eigenvalue for either of the two problems (they have the same set of eigenvalues) and $\mathbf{u}$ and $\mathbf{v}$ are the respective (eigenvector) solutions.

In CA the number of non-trivial solutions is $\min(n, p) - 1$. Because of the initial scaling of the data matrix, there is a 'trivial' eigenvalue equal to one (with its corresponding eigenvector) which is not considered in the analysis, but it is taken into account for reconstruction purposes as shown below.

The factors are defined as the projection operators on the principal axes (the eigenvectors) and are given by

$$\mathbf{g} = \mathbf{D}_p^{-1}\mathbf{u}, \qquad \mathbf{h} = \mathbf{D}_n^{-1}\mathbf{v}$$

The co-ordinates for the plotting of rows and columns are obtained by scaling the projections by a factor $\lambda^{1/2}$.

When CA is used as a dimension reduction technique, there are several diagnostics that will help in the choice of factors (i.e. dimensions) to be retained. These include the following:

(1) A *global measure of fit* when K factors are retained, expressed as a cumulative percentage of explained variation. This is similar to the measure used in PCA and is given in terms of the eigenvalues as the ratio

$$\frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k=1}^{\min(n,p)-1} \lambda_k}$$

(2) The *absolute contributions*, which indicate the composition of the factors as percentages of coatings or analytes. For the kth factor the contribution of coating i and analyte j are computed by the formulae

$$\mathbf{AC}^k(j) = f_{+j}g_{jk}^2, \qquad \mathbf{AC}^k(i) = f_{i+}h_{ik}^2$$

where g_{jk} and h_{ik} are the jth and ith components of factors $\mathbf{g}_k$ and $\mathbf{h}_k$, respectively. Note that the sum of the absolute contributions for a particular factor is one, or 100%.

(3) The *relative contributions*, which indicate the percentage of variation of a coating or analyte explained by each factor. These have also been called 'square correlations'. For a coating i and an analyte j the square correlations with factor k are

$$RC^k(j) = \frac{\lambda_k g_{jk}^2}{\sum_{k=1}^{\min(n,p)-1} \lambda_k g_{jk}^2}, \qquad RC^k(i) = \frac{\lambda_k h_{ik}^2}{\sum_{k=1}^{\min(n,p)-1} \lambda_k h_{ik}^2}$$

Note that the sum of relative contributions for a particular coating or analyte is one, or 100%.

(4) The *reconstruction error*, which measures the size of the residual when the reconstruction formula

$$f_{ij} = f_{i+}f_{+j}\left(1 + \sum_{k=1}^{\min(n,p)-1} \lambda_k^{1/2} g_{jk}h_{ik}\right)$$

is summed only over the retained factors. It indicates how well a coating or analyte is represented with a specified number of factors, When K factors are retained, we can define *error profiles* for the ith coating and jth analyte through the formulae

$$EP^K(j) = \sum_{i=1}^{n} f_{i+}\left(\sum_{k=K+1}^{\min(n,p)-1} \lambda_k^{1/2} g_{jk}h_{ik}\right)^2, \qquad EP^K(i) = \sum_{j=1}^{p} f_{+j}\left(\sum_{k=K+1}^{\min(n,p)-1} \lambda_k^{1/2} g_{jk}h_{ik}\right)^2$$

Note that in the reconstruction formula there is a one adding to the sum. This is a consequence of the trivial solution that occurs because of the initial scaling of the data.

The reconstruction formula also highlights CA as a generalization of the χ^2-test of independence, which is based on the comparison of f_{ij}, thought of as a probability, against the product of the marginals $f_{i+}f_{+j}$.

In this paper we will not explicitly use the error profiles since we are more interested in a global description of the data set than in reconstructing a particular coating or analyte. There are situations, however, where the error profiles may be useful.[10]

RESULTS AND INTERPRETATION

The results from CA can be displayed in a series of tables and graphs. First we have the eigenvalues with their associated percentage of variation explained (Table 3).

Thus the first six factors account for roughly 90% of the total variation.

We obtain a description of the factors from the absolute contributions of the coatings and analytes. In Table 4 we list for each factor the coatings and analytes having the highest absolute contributions (in parentheses, as a percentage of the factor).

Although there are two sets of factors, one for the coatings and one for the analytes, they are connected through the transition formulae.

Consider the quality of representation by each factor. In Table 5 we show the coatings and analytes with the highest relative contributions on each factor. A cut-off value of 40% of relative contribution was chosen to keep the list short.

We should emphasize here that neither of the groupings of coatings or analytes, in terms of absolute or relative contributions, implies a similarity between coatings or analytes grouped together. In fact, a factor could be describing the opposite character of several elements and it is only through a graphical display of the co-ordinates that a clustering may be inferred. In CA a great emphasis is placed on the use of plots to describe the results.

It is appropriate, however, to explore the factors in terms of the chemical characteristics which they may represent. Attempts to make chemical sense of the factors presented here, based on interactions such as Lewis acidity/basicity, polarity, hydrogen bonding, hydrophilicity and dispersion forces, have met with only limited success. Factor II, for example, would appear to represent a scale of hydrophilicity with water giving a strong positive response and octane a strong negative response. However, dodecane defies the pattern with a moderately positive response. Also, an initial inspection of the analytes and adsorbents which define factor I suggests a scale of polarity and polarizability; however, this is inconsistent with the large negative value for collodion.

Table 3. Eigenvalues from CA and variation explained by each

Eigenvalues	Variation (%)
0·1781	41·0
0·0752	17·7
0·0424	10·0
0·0310	7·3
0·0277	6·5
0·0258	6·1
0·0150	3·5
0·0131	3·1
0·0060	1·4
0·0040	1·0
0·0031	0·7
0·0015	0·4
0·0011	0·2

Table 4. Absolute contributions (percentages in parentheses) to the factors

Factor	Coatings	Analytes
I	Collodion (45)	DIMP (30) 1-Butyl formate (15) DMMP (13) Water (10)
II	Poly(vinyl stearate) (19) Poly(vinyl chloride) (16)	Octane (27) Water (25) Benzene (21)
III	Poly(vinyl carbazole) (22)	Dichloropentane (34) Isopropyl acetate (28)
IV	Poly(p-vinylphenol) (24)	DM phosphite (44)
V	Poly(vinyl stearate) (28) Poly(butadieneacrylonitrile) (19)	Benzene (44)
VI	Poly(vinyl isobutylether) (25)	a-Pinene oxide (46) Octane (19)

Table 5. Coatings and analytes with highest relative contributions to the factors

Factor	Coatings	Analytes
I	15, 3, 6, 8, 24	DIMP 1-Butyl formate DMMP
II	20, 27, 12, 22	Octane Water
III	14, 10, 16	Dichloropentane Isopropyle acetate
IV	18, 2	DM phosphate
V	1	None
VI	None	a-Pinene oxide

The quality of representation of rows and columns by the set of the first two factors, comprising roughly 60% of the variation in the data set, may be summarized as follows.

1. Six analytes are well represented by only two factors (sum of relative contributions >60%). These are DIMP, water, 1-butyl formate, DMMP, octane and triphenyl phosphite.
2. Two more analytes are moderately well represented by these factors (sum of relative contributions = 59%). These are benzene and triphenyl phosphate.
3. Eight coatings are well represented by the first two factors: numbers 3, 5, 6, 12, 15, 20, 24 and 27. Collodion, number 15, has the highest variation of all the coatings and has the highest impact on factor I, where almost all of its variation is 'captured'.
4. Three coatings are moderately well represented by the first two factors: numbers 4, 8 and 17. These three coatings do not contribute significantly to the formation of the factors.

The co-ordinates of the analytes for the six factors retained are given in Table 6. A plot of the first two factors is shown in Figure 1, using the first letters of the names as plotting symbols.

It is seen that octane, benzene and water contribute the most to the formation of factor II, but water is opposite the others with respect to this factor. Water and triphenyl phosphite will be displayed as being close together in a plot of the first two factors, but they are opposite on factor III. One should not infer clustering from just one plot unless most of the structure can be captured in one factorial plane.

The co-ordinate of the coatings are given in Table 7.

Several clusters can be detected when looking at a plot of factor I versus factor II as given in Figure 2. These should be examined through the use of the other factors. The clusters are as follows:

(1) Poly(butadieneacrylonitrile), poly(butadiene methacrylate), polybutadiene hydroxy
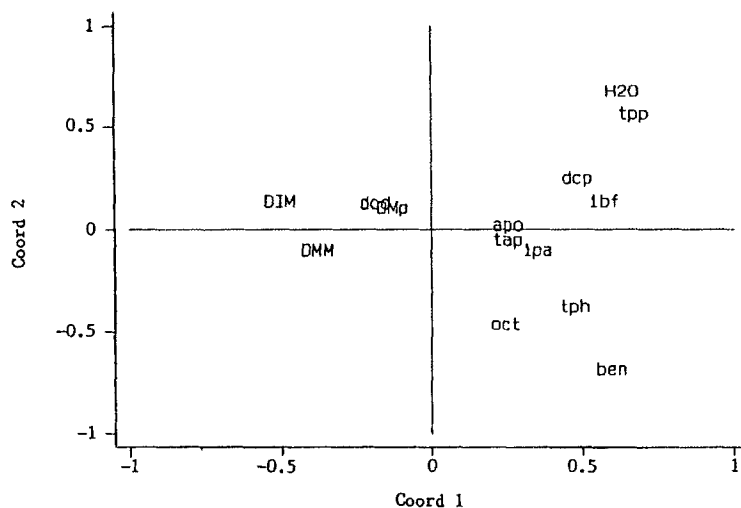

Figure 1. Plot of the first two factors from CA: the analytes

Table 6. Analyte co-ordinates

Analyte	FI	FII	FIII	FIV	FV	FVI
Benzene	0·60	− 0·71	0·38	− 0·11	− 0·62	0·20
Dodecane	− 0·18	0·11	− 0·03	0·11	− 0·00	0·30
DMMP	− 0·37	− 0·13	− 0·07	− 0·11	0·05	0·00
DM phosphite	− 0·12	0·08	− 0·13	− 0·31	− 0·04	− 0·02
1-Butyl formate	0·57	0·12	− 0·13	− 0·00	− 0·01	− 0·07
a-Pinene oxide	0·26	− 0·00	− 0·00	0·21	0·21	0·56
Triphenyl phosphite	0·68	0·54	− 0·41	0·18	− 0·10	− 0·02
DIMP	− 0·50	0·11	0·12	0·16	− 0·07	− 0·04
Dichloropentane	0·49	0·23	0·52	− 0·22	0·34	0·06
Isopropyl acetate	0·36	− 0·13	− 0·53	0·11	0·01	0·11
Tryamil phosphate	0·26	− 0·07	− 0·00	0·10	− 0·03	0·16
Octane	0·25	− 0·49	0·00	0·18	0·22	− 0·24
Triphenyl phosphate	0·48	− 0·41	0·11	− 0·00	− 0·11	− 0·19
Water	0·64	0·66	0·11	0·12	− 0·19	− 0·20

Table 7. Coating co-ordinates

| ID No.	FI	FII	FIII	FIV	FV	FVI
1	0·11	−0·15	−0·03	−0·04	0·33	−0·10
2	0·28	0·02	−0·09	−0·47	−0·02	0·00
3	0·61	−0·26	0·02	0·10	−0·02	−0·16
4	0·29	−0·34	0·00	0·14	0·25	−0·20
5	0·39	−0·39	0·20	−0·05	−0·28	0·14
6	0·56	−0·30	0·03	0·11	0·06	−0·21
7	0·18	−0·25	0·02	0·06	0·29	−0·09
8	0·64	0·09	0·02	−0·07	−0·16	−0·26
9	0·00	−0·21	0·01	0·21	0·15	−0·03
10	0·56	0·36	−0·69	−0·01	−0·11	0·02
11	0·49	0·17	−0·29	0·34	0·02	0·63
12	0·57	0·81	0·11	−0·03	0·02	−0·21
13	0·66	0·28	−0·63	0·45	0·17	0·28
14	0·40	0·27	−0·82	−0·37	−0·11	−0·03
15	−0·46	−0·03	−0·04	−0·04	−0·01	0·01
16	0·46	0·20	−0·52	0·01	−0·17	0·03
17	0·60	0·41	−0·23	0·20	−0·44	−0·26
18	0·13	0·11	−0·20	0·46	0·09	−0·04
19	0·39	0·39	0·39	0·00	0·19	0·35
20	0·40	0·71	0·23	−0·12	−0·09	−0·32
21	0·28	0·40	0·31	−0·07	0·25	0·23
22	0·04	0·27	0·21	0·05	0·10	−0·06
23	0·33	0·51	0·52	−0·50	0·20	0·02
24	−0·29	0·26	0·03	0·09	0·00	0·11
25	−0·65	0·46	0·38	0·52	−0·36	−0·27
26	0·29	0·44	0·17	0·22	0·17	0·52
27	0·37	0·71	0·00	−0·06	−0·21	−0·27
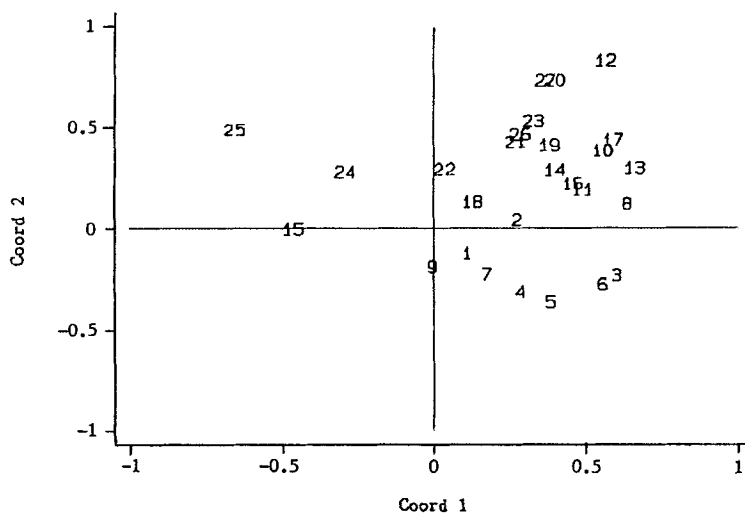
Figure 2. Plot of first two factors from CA: the coatings

terminated, poly(vinyl stereate), poly-1-butadiene and polybutadiene hydroxy terminated liquid (ID numbers 1, 3, 4, 5, 6 and 7). These all plot in the fourth quadrant, having a positive first co-ordinate and negative second co-ordinate. This cluster appears to define adsorbents with polar or polarizable groups and hydrophobic alkane-based backbones.

(2) A large cluster of coatings having positive first and second co-ordinates. This cluster can be broken into the following subclusters: poly(p-vinylphenol), methyl vinyl ether, polystyrene, poly vinyl isobutyl ether, poly-1-butene, poly(vinylcarbazole) and poly(vinylbutyral) (ID numbers 2, 8, 10, 11, 13, 14 and 16), which have a 'moderate' second co-ordinate; poly(methyl methacrylate), ethyl cellulose, poly(caprolactone), poly(caprolactone)triol 2X and DC 11 (ID numbers 17, 19, 21, 23 and 26), which have a 'large' second co-ordinate; and poly(vinyl chloride), poly(ethylene glycol methyl ether) and phenoxy resin (ID numbers 12, 20 and 27), which have a 'huge' second co-ordinate. With the exception of poly-1-butene and DC 11 these adsorbents have polar or polarizable groups and less substantial (vinyl-based) alkane backbones. Poly-1-butene and DC 11 are separated from this cluster by factors III and IV.

(3) A cluster of coatings having a non-positive first co-ordinate: octadecyl vinyl ether/maleic anhydride, collodion, carnuba wax and abietic acid (ID numbers 9, 15, 24 and 25). This cluster can be subdivided into the subclusters (9, 15) and (24, 25) according to the sign of the second co-ordinate. Attempts to make chemical sense of this cluster have met with little success.

A comparison with the varimax factors in Table 2 shows an almost complete agreement between the solutions from PCA and CA, but whereas eight vectors were obtained from PCA and a further rotation was needed, CA produces almost the same groupings using only six factors (vectors) without a rotation.

In the first cluster, poly(vinyl stearate) is somewhat different from the other coatings in terms of the rest of the factors. This leaves only butadiene coatings in this cluster.

In the second cluster, one could exclude methyl vinyl ether from the first subcluster and poly(methyl methacrylate) from the second subcluster. The first subcluster is then the polyvinyl group.

For the purpose of choosing a reduced set of coatings, we would select on the basis of the contributions 'absolute' and 'relative' shown in Tables 4 and 5. We would choose collodium, poly(vinyl chloride), poly(vinyl carbazole), poly(p-vinylphenol), poly(vinyl isobutyl ether) and poly(butadiene eneacrylonitrile) for a set of size six, and add to the list on the basis of other considerations.

The usual practice in CA is to plot the coatings and the analytes on the same diagram. We chose not to do this, in agreement with the arguments advanced by Goodman[12] and Greenacre.[13] If simultaneous plotting is done, care must be taken when trying to give sense to the closeness of an analyte to a cluster of coatings or of a coating to a cluster of analytes, although sometimes this occurrence can be very illuminating.

CONCLUSIONS

CA is a multivariate technique that treats rows and columns of a data matrix with non-negative entries symmetrically, projecting them on to a set of factorial axes. A graphical display of the projections aids in the search for patterns and for an interpretation of underlying relationships.

For the data set given in the paper by Carey *et al.*, CA gave the same results as the combined use of PCA with varimax rotations and cluster analysis, but fewer factors are required to find a sensible set of clusters and CA was shown to be an effective dimension reduction technique. We point out that for this type of data set the choice of PCA for the columns and cluster analysis for the rows is arbitrary, since one may choose to work with the transposed data

matrix and obtain different results; however, CA gives the same results when applied to the transposed matrix.

The advantages of CA include convenient diagnostics that help in the interpretation of the results. These include the absolute and relative contributions of the rows and columns, which were used to find a reduced set of coatings, as well as global measures of the quality of representation.

Another advantage of the CA approach is that the results can be interpreted chemically without the need for rotation of the factors. It is apparent that the clusters observed are influenced by and indicative of the chemistry of the phases. It is interesting to note that the clusters obtained indicate that the polymer backbone is as important as the chemical functionality in determining the chemical behaviour of these stationary phases. The fact that some adsorbents did not appear in chemical clusters where they may seem to fit, or that others may appear in what seem to be inappropriate clusters, may be due to the conformation, uniformity, rigidity and depth of the coatings. These experimental variables are not explicitly incorporated in the analysis.

NOTICE

REFERENCES

1. W. P. Carey, K. R. Beebe, B. R. Kowalski, D. L. Illman and T. Hirschfeld, *Anal. Chem.* **58**, 149 (1986).
2. M. J. Greenacre, *Theory and Applications of Correspondence Analysis*, Academic, London (1984).
3. L. Lebart, A. Morineau and K. M. Warwick, *Multivariate Descriptive Statistical Analysis*, Wiley, New York (1984).
4. D. L. Hoffman and G. E. Franke, *J. Market. Res.* **23**, 213 (1986).
5. F. Valenchon, *Mat. Geol.* **14**, 331 (1983).
6. G. Bergametti, A. L. Dutot, J. P. Quisefit and R. Vie le Sage, *J. Volcano. Geotherm. Res.* **15**, 355 (1983).
7. E. C. Grunsky, *J. Geochem. Explor.* **25**, 157 (1986).
8. T. J. Carleton, *Ecology*, **65**, 469 (1984).
9. A. L. Dutot, G. Bergametti and P. Buat-Menard, *Atmos. Environ.* **22**, 1737 (1988).
10. H. R. Rhodes and D. E. Myers, *J. Chemometrics*, **5**, 273 (1991).
11. M. Stuart, *Am. Stat.* **36**, 365 (1982).
12. L. A. Goodman, *Int. Stat. Rev.* **54**, 243 (1986).
13. M. J. Greenacre, *J. Market. Res.* **26**, 358 (1989).