CHROM. 6022

THE PRECISION AND OTHER ASPECTS OF GAS CHROMATOGRAPHIC ANALYSIS BY THE LINEAR RELATIONSHIP METHOD

ALEKSANDER JANIK
Institute of Heavy Organic Synthesis, Blachownia Sl. (Poland)

(Received January 11th, 1972)

SUMMARY

Gas chromatographic analysis by the linear relationship method and the least-squares procedure reveals important features of the multicomponent system investigated. This enables the precision of qualitative and quantitative analyses and the chromatographic behaviour of the system to be determined.

INTRODUCTION

The area factors related to any standard can be determined by the linear relationship method (L.R.M.), even when the pure components of analysed mixtures are not available[1]. The usefulness of L.R.M. has been shown in the quantitative analysis of three- and four-component mixtures[2]. The use of a gas density detector makes it possible to determine the molecular weights of unknown components from values of their area factors[3]. Other specific properties of components of complex mixtures, such as refractive index and heat of combustion, could be determined by L.R.M. by using detectors that respond to these properties.

The response of the thermal conductivity and flame ionisation detectors, and also other detectors, is usually some function of a specific property of chemical compounds. Nevertheless, the area factors determined could be used not only for quantitative analysis but also for qualitative analysis of specific characteristics of the components of complex mixtures together with retention times or Kováts retention indices.

The quantitative composition of complex mixtures can easily be determined by L.R.M. and this gives a new basis for the evaluation of the infrared spectra of individual components from the spectra of complex mixtures[4].

The use of L.R.M. is not restricted to gas chromatography, but can also be applied to other chromatographic and separation techniques that obey accepted assumptions.

The scope of application of L.R.M. seems to be very wide and it is desirable to discuss the precision of the method. In this paper, L.R.M. is discussed in terms of the

least-squares procedure[5] and some special cases of gas chromatographic analysis are also discussed.

THEORETICAL

Let us assume that all assumptions that form the basis for L.R.M. are obeyed and that we are dealing with peak areas biased with random errors. Then, for procedure 1 (ref. 1), we can write

$$\sum_{t=1}^{n} k^0_{is} P_{il} = k^0_{sl} P_{sl} + \varepsilon_{al} \qquad (1)$$

or

$$\sum_{t=1}^{n} k^0_{is} \frac{P_{il}}{k^0_{sl} P_{sl}} = 1 + \varepsilon_{bl} \qquad (2)$$

and for procedure 3 (ref. 1)

$$\sum_{t=1}^{n} k^0_i P_{il} = 1 + \varepsilon_{cl} \qquad (3)$$

k^0_{is} = area factor of component i related to the area factor of standard s;

k^0_{sl} = (weight of sample l)/(weight of the standard s added to it). We may accept this value as being unbiased with error as the weighing accuracy is high;

k^0_i = area factor of component i related to the constant amount of sample injected into the chromatograph (by volume or weight);

P_{il}, P_{sl} = peak areas of component i and standard s of sample l, respectively;

ε_{al}, ε_{bl} and ε_{cl} = total errors of the corresponding equation for sample l (or injection, if the sample has been injected many times).

The errors in eqns. 1, 2 and 3 depend, first of all, on the errors in the area measurement. The error in the area measurement includes the reliability of the detector, electrical circuits, and all parameters that influence the recording of the signal, as well as the error in the area measurement itself.

All the equations of procedures 1 and 3 can be expressed in matrix notation by

$$PK^0 = Y + \varepsilon \qquad (4)$$

where

P = the rectangular matrix $m \times n$ of peak areas P_{il} for eqns. 1 and 3, and values $P_{il}/(k^0_{is}P_{is})$ for eqn. 2 (matrix of independent variables);

K^0 = $n \times 1$ column vector of the area factors of parameters being estimated;

Y = $m \times 1$ column vector of $k^0_{is}P_{is}$ or 1 (vector of observations);

ε = $m \times 1$ vector of errors;

m = number of samples;

n = number of components.

The exact values of the column vector of errors are not known, so we are not able to calculate the true values of the area factors k^0_{is} or k^0_i but only their best estimation by the least-squares procedure:

$$K = (P'P)^{-1}P'Y \quad (5)$$

with variances

$$V(K) = (P'P)^{-1}\sigma^2 \quad (6)$$

where

K = estimation of vector K^0;
$\sigma^2 \approx s^2$ = the residual sum of squares divided by the residual degrees of freedom;
P' = transpose of the matrix P;
$(P'P)^{-1}$ = variance–covariance matrix.

The variance of $\hat{Y}$ for given vector I can be estimated from the equation:

$$V(\hat{Y}) = P'_I (P'P)^{-1}P_I \sigma^2 \quad (7)$$

where

P_I' and P_I are the versus and column vectors, respectively;
$\hat{Y}$ = the value calculated from the regression equation for given vector P_I.

$V(\hat{Y})$ represents the precision of the quantitative determination and can be used for the estimation of the confidence limits for the sum of the concentrations of all components of the sample or the sum of normalized peak areas. Thus, the ratio $V(\hat{Y})/\hat{Y}$ can be applied as the criterion for the choice of the best detector for the analysis of a given sample.

Eqn. 6 characterizes the precision of the determination of the related area factors k_{is} as characteristics or physical constants of the components for qualitative analysis.

Eqns. 6 and 7 are valid for any sample that was used for estimation of the area factors. Therefore, if the calculations are performed with the aid of a computer, then it is reasonable to add data for newly determined samples to a previously accumulated set of data to calculate a new variance–covariance matrix and a new set of the area factors with higher precision.

DISCUSSION

To be certain that the area factors and other results based on them are reliable, the fit of the experimental values into eqn. 1, 2 or 3 must be checked, and to do this the contribution of the "pure error" sum of squares and the "lack of fit" sum of squares to the residual sum of squares must be found[5]. To estimate the "pure error" sum of squares, exactly equal amounts of some samples should be injected into the column. This can easily be achieved for procedure 3, but such a requirement is not possible in procedure 1. In the latter case, if many repeated injections of the same amount of sample do not give a response within the linear range of the detector, the "pure error" sum of squares may contain a contribution from the "lack of fit" sum of squares.

If the accepted models, *i.e.*, eqns. 1, 2 and 3, were strictly obeyed, no "lack of fit" sum of squares should be observed. The existence of the "lack of fit" sum of squares indicates that the assumptions accepted for L.R.M. are not strictly valid for the multicomponent system analysed.

In practice, we may meet at least one of the following situations that result in the "lack of fit" sum of squares making a large contribution to the residual sum of squares, or the lack of fit even become significant.

I. Peaks are composed of more than one component

The "lack of fit" sum of squares will be observed if the area factors of the components that are eluted as a single peak differ in value. As this situation occurs the most often, it is quite simple to decide whether or not all the components of the sample have been separated.

II. One or more components of the sample are absorbed in the column

In this case, no "lack of fit" sum of squares will be observed if the concentrations of absorbed components are constant for all the samples, but this has very little probability of occurring. If the "lack of fit" sum of squares is due to the absorption of any component in the column, then L.R.M. cannot be applied directly. If the range of linearity of the detector is small, then L.R.M. cannot be applied in the determination of very small or trace concentrations of components. Both instances are similar and they can therefore be treated together. In first instance, from a given set of samples, the set of samples that contain all the components except those absorbed in the column must be obtained, and in the second the set of samples that contain the trace components as principal ones must be obtained. Distillation can be used for this purpose.

The area factors of components of a set of samples obtained in this manner can be determined by procedure 1. If the same standard is added to the original set of samples, the percentage fractions of eluted or trace components can be determined from the relation

$$x_{il} = \frac{k_{is} P_{il}}{k^0_{sl} P_{sl}} \tag{8}$$

III. The detector does not work in a linear mode

This situation has been discussed previously[1]. It can be added that a higher-order relation than two should be checked.

IV. The existence of a systematic error

A systematic error in peak area measurement seems to be difficult to avoid. The peaks are not strictly symmetrical and the "evaluation" of any peak from the zero line and its neighbouring peaks usually causes the systematic over- or under-estimation of its area. In such a situation, we can accept some assumptions concerning the values of systematic errors for all or some of the peaks. If these assumptions are reasonable, they will cause a significant decrease in the "lack of fit" sum of squares.

The standard should be chromatographically pure and give a symmetrical, well

resolved peak. Its degree of impurity will influence the values of $k^0_{sl}P_{sl}$ and is therefore a serious source of systematic error.

This discussion leads to the conclusion that the application of L.R.M. and the least-squares procedure reveals very important features of the multicomponent system investigated.

It can also be said that the utility and the scope of application of L.R.M. will depend on the range of linearity of detectors, their sensitivity and reliability, reaction for some specific property of chemical compounds, and the accuracy of peak area measurements.

REFERENCES

1 A. JANIK, *J. Chromatogr.*, 54 (1971) 321.
2 A. JANIK AND J. HETPER, *J. Chromatogr.*, 54 (1971) 327.
3 J. HETPER AND A. JANIK, *J. Chromatogr.*, 64 (1972) 162.
4 A. JANIK, *Chem. Anal.*, submitted for publication.
5 N. R. DRAPER AND H. SMITH, *Applied Regression Analysis*, John Wiley, New York and London, 1967.