Journal of Chromatography, 557 (1991) 23–37
Elsevier Science Publishers B.V., Amsterdam

CHROM. 23 296

Quantitative determination limit in chromatography: computer-based simulations

M. ZOUBAÏR EL FALLAH[a] and MICHEL MARTIN*
Ecole Supérieure de Physique et Chimie Industrielles, Laboratoire de Physique et Mécanique des Milieux Hétérogènes (URA CNRS 857), 10 Rue Vauquelin, 75231 Paris Cedex 05 (France)

ABSTRACT

Following the previously developed statistical approach of the peak overlapping phenomenon in the chromatography of complex mixtures, computer-simulated chromatograms were generated to investigate the influence of this phenomenon on the accuracy of quantitative determinations. The probabilities of performing the quantitative determination, with an error smaller than or equal to a given value, of the most abundant component in a peak observed in the chromatogram, and of a given sample component were computed as a function of the relative height or area of the observed peak and of the relative area of the component, respectively, for different values of the accepted error. In addition, these probabilities were shown to depend significantly on the chromatographic saturation factor, which reflects the degree of occupancy of the space available for the separation by the sample. Surprisingly, it appears that small peaks observed in the chromatogram are more likely than large peaks to be pure. However, the probability of the correct quantitative determination of a sample component increases with its concentration in the sample. The determination limit, defined as the minimum relative amount of a component in the sample required to have a given probability of performing its quantitative determination with an error smaller than or equal to a given value, when taking into account the peak overlap phenomenon, was derived.

INTRODUCTION

Recent developments in chromatographic techniques, such as an increase in column efficiency through the use of columns with very small inner diameters in gas chromatography and of very fine particles in liquid chromatography, the preparation of stationary phases with improved stability and reproducibility, the use of highly sensitive detectors and the achievement of high reproducibility in injection and mobile phase delivery systems, have allowed chromatographers to analyse more and more complex matrices. The quantitative determination of particular compounds remains the major goal of such analyses. In a complex matrix, the main source of errors, when performing quantitative determinations, is essentially due to the peak interference phenomenon, which is obviously greater for mixtures containing a large

[a] Present address: Department of Chemistry, University of Tennessee, Knoxville, TN 37996-1600, USA.

0021-9673/91/$03.50

number of solutes. The seriousness of the peak overlap phenomenon and its effect on the analytical information losses have been already pointed out [1–5].

It has been shown that in a crowded chromatogram, the probability of obtaining a single uncontaminated peak, with a reasonable fixed degree of purity, is extremely low. It depends on what can be called the chromatographic saturation, *i.e.*, on the extent to which the space available for the separation, which can be expressed in terms of the peak capacity of the system [6], is to be occupied by all the sample components. Then, one can regard the accuracy of the quantitative determinations in chromatography from a probabilistic point of view. Hence, in a crowded chromatogram, we have only a limited probability of performing the quantitative determination, with an error smaller than or equal to a fixed value, of any solute belonging to the mixture. This probability depends on the degree of saturation of the chromatogram, but also on the relative analytical response of the solute. This relative response may be expressed as the ratio of the area of the so-called parent peak (*i.e.*, the peak which would be obtained if an amount of the solute under consideration equal to that present in the sample was chromatographed alone, under experimental conditions identical with those selected for the analysis of the mixture) to the sum of the areas of all parent peaks, which is also the area of the whole chromatogram. In fact, when the response of the solute parent peak represents an important fraction of the whole chromatogram response, the probability of performing its quantitative determination, with an error smaller than or equal to a fixed value, is relatively high, because the height and area of the largest chromatographic parent peaks are most likely to remain almost unmodified by the interference with the other parent peaks of the mixture. However, when the parent peak relative response is weak, the latter probability will become relatively low because the small parent peaks, when interfering with the others, exhibit a greater probability of being hidden by the largest ones. Accordingly, a small parent peak will most likely belong to a cluster which has a height and an area greatly different from those of the parent peak considered. In the following, the peaks of the envelope of the chromatogram will be called observed peaks. Each observed peak contains one or several parent peaks.

Nagels and co-workers [7,8] were the first to use this probabilistic point of view in their computer simulation studies of the effect of peak interference on the correctness of quantitative chromatographic determinations, when analysing plant extracts by reversed-phase high-performance gradient elution liquid chromatography. In order to simulate a typical plant extract chromatogram, a parent peak area distribution function was required. They used a distribution which was estimated from the observed peak area distribution function, itself obtained by chromatographing 62 plant extracts. They computed the probability that the most abundant component (in terms of detector response) belonging to an observed peak represents more than 50, 90 or 95% of the total relative area of this observed peak. Then, the error associated with the quantitative determination of the most abundant component was smaller than 50, 10 or 5%, respectively. In addition, they computed the probability of performing a quantitative determination, with an error smaller than or equal to a fixed value, of a parent peak with a given relative response. They showed that these two probabilities depend on the peak capacity of the system and on the relative response of the peak. In order to characterize the ability of the chromatographic system to perform quantitative determinations, they introduced a parameter termed the quantitative determina-

tion limit, DL_e^p, which corresponds to the minimum relative response that a parent peak solute must have in the chromatogram in order to have a given probability, p, of performing its quantitative determination with an error smaller than or equal to a fixed value, e. At a fixed probability level of performing the quantitative determination with a fixed relative error, it is clear that the parameter DL_e^p varies with the peak capacity of the system. The higher is the resolving power of the chromatographic system, the lower will be DL_e^p, and hence the lower are the amounts of solutes that can be determined with a fixed probability and a certain error.

It should be pointed out that the conversion of DL_e^p, expressed in terms of solute relative response, to its value in terms of solute amount (weight fraction of solute in the analysed mixture), leads, when response factors are known, to a specific quantitative determination limit parameter SDL_e^p, which characterizes the ability of the system to determine a specific compound. Accordingly, the optimization of quantitative determinations for a complex matrix corresponds to a search of experimental conditions in order to reach a global minimum for all the SDL_e^p corresponding to the components of interest, *i.e.*, a simplex optimization performed on a space dimension equal to the number of components of interest plus the number of the experimental parameters to be optimized.

The parameter SDL_e^p is to be distinguished from the classical "limit of determination" [9,10], which is conventionally defined as the minimum amount of a component producing an analytical signal above which it is considered that the solute can be quantified consistently, within a level of confidence. Of course, in the estimation of the classical limit of determination, apart from the instrumental intrinsic sources of error, none of the other sources of signal perturbations are to be taken into account such as those induced by the presence of the other solutes in the sample. In this study, assuming that the sources of errors in quantitative determinations from complex chromatograms are principally due to peak overlap phenomena, the concept of the quantitative determination limit, DL_e^p, allows one to take into account, in probabilistic terms, the matrix effects as a function of the resolving power of the chromatographic system.

The probability of performing a quantitative determination of a parent peak, with an error smaller than or equal to a fixed value, and thus DL_e^p, depends on the chromatographic saturation factor, *i.e.*, the ratio of the real number of detectable components, m, in the sample to the dimensionless length of the separation space, T. In elution chromatography, T represents the ratio of the distance over which the chromatogram is spread (expressed in time units) to some appropriate reference time. When the standard deviation is the same for all the parent peaks, which is generally acceptable assumption in liquid chromatography with linear solvent strength gradient elution and in gas chromatography with linear temperature programming [11,12], T is linearly proportional to the peak capacity, n_c, of the system. Under these chromatographic conditions, one may express, for convenience, T in standard deviation units. Thus, for a given sample with m detectable components, the above probability and DL_e^p depend on the system peak capacity, n_c. However, Nagels and coworkers [7,8] did not state precisely the number of parent peaks, m, used in their computer simulations in order to estimate the different probabilities of quantitative determinations. Unfortunately, the different curves they presented cannot be used for practical estimations of the quantitative determination limits in real chromatograms,

because the saturation factors to which the resulting curves correspond are not known. In addition, although the quasi-exponential peak-height distribution function used by these authors is similar, for a large number of components, to the theoretical one [13], it is a specific one corresponding only to phenolic plant extracts. Further, the procedure they used to estimate the distribution function of the parent peak areas seems to be very sensitive to the degree of chromatographic saturation, and thus this estimation is not an unequivocal operation.

The objective of this study was to generalize the different probabilities curves of quantitative determinations by utilizing a theoretical parent peak-height distribution function and also by expressing those probabilities and DL_e^p as a function of the chromatographic saturation factor, m/T. The use of m/T is preferred to that of m/n_c as the former, in contrast to the latter, does not imply the selection, on an arbitrary basis, of a value for the critical resolution needed to compute n_c. Indeed, it has been shown that this critical resolution depends on the peak-height distribution of the parent peaks [14]. One can note that the saturation factor, m/T, defined above is one quarter of the saturation parameter, m/n_c, used by Davis and Giddings [1] on the basis of a four standard deviation (4σ) separation between consecutive peaks. In the following, the method selected for quantitative measurements (peak height or peak area) is first discussed, together with its effect on the quantitative determination probabilities. Two different quantitative determination probabilities will be calculated by means of computer simulations: for an observed peak, having a given relative response, we estimated the probability of performing the quantitative determination, with an error smaller than or equal to a fixed value, of the major parent peak belonging to this observed peak. This probability depends on the relative response of the observed peak and on the chromatographic saturation factor. Then, the probability of performing the quantitative determination, with an error smaller than or equal to a fixed value, of a parent peak of a given relative abundance in the chromatogram is estimated. This probability depends on the relative abundance of the parent peak and on the saturation factor. Finally, the estimation of DL_e^p is performed for different values of p and e, using the latter probability curves.

MODEL AND COMPUTER SIMULATION PROCEDURE

In order to simplify the computational procedure of synthetic chromatograms, one assumes that the parent peaks are Gaussian with a constant standard deviation along the retention axis. Under these conditions, the parent peak-height distribution is identical with the distribution of the products of the concentration (in the sample) by the detection response factor of the sample components. One assumes that this distribution is the same as the distribution of the concentrations of the sample components, *i.e.*, that the convolution of the concentration distribution by the component distribution does not affect or affects only slightly the concentration distribution. It was found that there is some support for this hypothesis in liquid chromatography [15]. Thus, under these conditions, the parent peak-height distribution is identical with the component concentration distribution, which is determined from a statistical theory of concentration distribution in complex samples [13].

A program written in Pascal and run on an SPS-7 computer (Bull, Louveciennes, France) allows the simulation of the synthetic chromatograms. To obtain a

chromatogram with a fixed saturation factor, m/T, first two components are selected and the difference between their retention times is set to T standard deviations, then the $m-2$ other components are randomly distributed within this time interval. The random function was available from the Pascal Library and has a period equal to $2^{32}-1$; its sequence can be modified by changing the function argument. This function was used to obtain both the retention times of the randomly positioned parent peaks and their heights from a random selection through a file where all the peaks heights are listed. The arguments of the random functions for the retention time and peak-height selection are different. Different simulated chromatograms are obtained simply by changing the random function arguments. As discussed above, the assumptions of randomness of the distribution of retention times of the sample components and of constant deviation have been shown to be realistic in the case of some analyses using linear solvent strength gradient elution in liquid chromatography or linear temperature programming in gas chromatography, for solutes with no well defined correlation in the molecular structure [11,12,15].

For each simulated chromatogram, a procedure permits the determination of the area and the height of each observed peak, in addition to the retention time of each valley appearing in the chromatogram. An observed peak corresponds to the feature of the chromatogram appearing either between two consecutive valleys or before the first valley or after the last one. Accordingly, there are as many observed peaks as there are maxima in the chromatogram. It is then necessary to identify each observed peak. For this, the following procedure is adopted: any parent peak which has its retention time between the retention times of two consecutive valleys is considered to belong to the observed peak defined by those valleys. Parent peaks whose retention times are either before the retention time of the first valley or after the retention time of the last valley are considered to belong to the first and to the last observed peaks, respectively.

In the following, each parent peak is characterized by its relative height, Hr_p, and its relative area, Ar_p. As the parent peaks are assumed to be Gaussian and to have the same standard deviation, these two quantities are identical. For each simulation, the relative area, Ar_o, and the relative height, Hr_o, of each observed peak are computed. Ar_o is the ratio of the area of the observed peak, as defined above, to the total area of the chromatogram. Hr_o is the ratio of the height of the observed peak to the sum of the heights of all observed peaks. With each observed peak is associated a most abundant parent peak, which corresponds to the parent peak, belonging to this observed peak, which exhibits a relative area larger than that of any other parent peak belonging to the same observed peak. This parent is assumed to be the most abundant one even though a fraction of its area may be outside either limits of the observed peak. Of course, owing to parent peak interference phenomena, the correlation between Ar_o and Hr_o is not obvious, and it depends on the degree of chromatographic saturation.

In all simulations discussed in the following, the number of parent peaks, m, was set to 50. Some simulations, not reported here, were also performed with 25 or 100 parent peaks. The distribution of peak heights, resulting from the adaptation of a theoretical model derived for a very large number of components to finite numbers of components [13], are not exactly the same for the different values of m. Therefore, the differences which appear in the absolute values of the probability curves, to be

described below, for different m can most likely be attributed to this artifact. However, it is remarkable that the curves obtained at different m values but for a given saturation factor, m/T, are all parallel and very close to each other while curves corresponding to a given m value but to different saturation factors are fairly distant from each other and have a different shape. This confirms that the saturation is the key parameter affecting chromatographic performances.

RESULTS AND DISCUSSION

Probability of performing a quantitative determination, with an error smaller than or equal to a given value, of the most abundant parent peak in a given observed peak

Usually, the chromatographer does not have a precise and complete idea about all the components of the mixture under investigation; what he or she has is a chromatogram containing several observed peaks, and he or she may be interested in determining quantitatively one or several particular components for which the retention times are known, under given chromatographic conditions. Generally, the observed peak in the chromatogram which has a retention time (time corresponding to the peak maximum) relatively close to that of the component of interest is attributed to that component. Sometimes, further investigations (such as NMR and spectral analyses of some collected fractions or on-line coupling with mass spectrometry or Fourier transform infrared spectrometry, for instance) are needed to confirm the purity of all the peaks of interest. Nevertheless, in spite of the very high resolving power of some of these combined systems, such as gas chromatography–mass spectrometry, the complete identification of a mixture has been demonstrated, in one case, to be impossible both experimentally and from a statistical point of view [16]. Thus, there is only a limited probability that the most abundant parent peak in an observed peak will correspond to the one the analyst wants to quantify. Therefore, one of the most important questions is to know the probability of performing such quantitative determinations, with an error smaller than or equal to e_0. This probability is denoted $P_{o,h}(e \leqslant e_0)$ or $P_{o,a}(e \leqslant e_0)$ depending on whether the quantitative determination is performed by measuring the height or the area of the observed peak, respectively. Obviously, these probabilities depend on Hr_o and Ar_o, respectively, but also on the saturation factor, m/T.

Here the error, e_0, is computed with respect to the difference between the value of the height or area of the observed peak and that of most abundant peak. Accordingly, when measuring heights, the error is always positive owing to the increase in the peak height when interferences occur, whereas the error associated with area measurements may be either positive or negative. Therefore, in the following, the absolute value of e_0 is used for the area determinations.

For a fixed saturation factor, 10 000 simulations were done ($N_s = 10\,000$). For a given parent peak, exhibiting a given Hr_p (or Ar_p) value, the computation procedure consists in the evaluation of the number of times, $N_{maj,h}(Hr_p, e \leqslant e_0)$, where this particular parent peak was found to be the most abundant one in the observed peak to which it belongs and where the error associated with its height determination is smaller than a given value, e_0. Similarly is evaluated the number of times, $N_{maj,a}$ $(Ar_p, e \leqslant e_0)$, where the given parent peak was found to be the most abundant one in the observed peak to which it belongs and where an error smaller than e_0 is made in

its area determination. One is interested in the variation of the probabilities $P_{o,h}$ and $P_{o,a}$ with Hr_o and Ar_o, respectively. However, the height and area of the observed peak where the given parent peak is found to be the most abundant and where the determination is made with an error $e \leqslant e_0$ are different from one simulation to another. Therefore, the variation of the probabilities $P_{o,h}$ and $P_{o,a}$ with Hr_o and Ar_o can be obtained only indirectly. Each time that the parent peak with a given Hr_p (or Ar_p) is found to be the most abundant in an observed peak with $e \leqslant e_0$, the height, or area, of the observed peak is tabulated and, at the end, the averages, $Hr_{o,av}$ and $Ar_{o,av}$, of all tabulated values are calculated. The values of the probabilities are then given by

$$P_{o,h}(Hr_{o,av}, e \leqslant e_0) = N_{maj,h}(Hr_p, e \leqslant e_0) / N_s \quad (1)$$

$$P_{o,a}(Ar_{o,av}, e \leqslant e_0) = N_{maj,a}(Ar_p, e \leqslant e_0) / N_s \quad (2)$$

Comparison between area and height measurement performances on the basis of the probability computations

It is interesting to compare the two probability values (based on area and height measurements) obtained for a similar mean value of Hr_o and Ar_o. This comparison was made with 10 000 simulations and for both a relatively low-density chromatogram having a saturation factor $m/T = 0.05$ and a relatively dense chromatogram five times more crowded ($m/T = 0.25$). The admitted relative error, e_0, was set to 10%. Fig. 1 shows the variation of $P_{o,h}(e \leqslant 0.1)$ and $P_{o,a}(e \leqslant 0.1)$ as a function of Hr_o and Ar_o, respectively, for these two values of the saturation factor.

Both probabilities decrease with increasing relative area and height of the observed peaks, which leads to the conclusion that the larger is the observed peak, the higher is its probability of being contaminated. The probabilities are lower at high

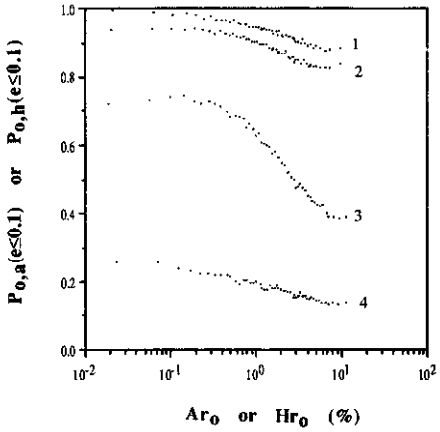

Fig. 1. Variation of the probability of performing a quantitative determination of the height of the major parent peak in an observed peak, with an error smaller than or equal to 0.1, $P_{o,h}(e \leqslant 0.1)$, as a function of the relative height of the observed peak, Hr_o, for a saturation factor $m/T = 0.05$ (curve 1) and 0.25 (curve 4), respectively. Variation of the probability of performing a quantitative determination of the area of the major parent peak in an observed peak, with an error smaller than or equal to 0.1 $P_{o,a}(e \leqslant 0.1)$, as a function of the relative area of the observed peak, Ar_o, for a saturation factor $m/T = 0.05$ (curve 2) and 0.25 (curve 3), respectively.

saturation than at low saturation, which reflects the fact that peaks are more likely to be contamined when the chromatographic saturation is high. At low saturation factors, the decrease in both probabilities is less pronounced than that at high saturation factors. Moreover, the simulations show that at high saturation, the probability of a correct quantitative determination is higher when the peak area rather than the peak height is used for the measurements. Nevertheless, this observation is no longer valid at low saturation factors, for which the probability of correct determination using height measurements was found to be higher than that of area measurements. However, in this instance the difference between the two probabilities is smaller than at high saturation and is expected to decrease as the saturation decreases. Indeed, at very low saturation, *i.e.*, when m/T tends toward 0, both measurements will be equivalent and will lead to determination probabilities which tend toward 1. Therefore, it seems that, for relatively high chromatographic saturations, peak heights will be more seriously affected by interferences than will peak areas. As the problem of quantitative determination is more serious when the saturation is relatively high and as most commercial integrators express quantitative results by mean of area measurements, we shall only consider the area measurements in the following to express the determination probabilities.

Variation of $P_{o,a}(e \leqslant e_0)$ with the relative area of the observed peaks

For a given chromatographic saturation, the probability of performing a quantitative determination of the most abundant peak in an observed peak, with an error smaller than or equal to e_0, depends on the relative area of this observed peak. The general shape for the variation of this probability as a function of Ar_o, for different values of the accepted error, e_0, and for a saturation factor $m/T = 0.25$, is shown in Fig. 2. As expected, for a fixed value of Ar_0, the higher is the tolerated error the higher is the probability of correct determination. However, it may seem surprising that, for a relatively large domain of low Ar_o values, *i.e.*, $Ar_o \leqslant 0.3\%$, the probability remains almost constant, then decreases rapidly as Ar_o increases (however, there is one excep-

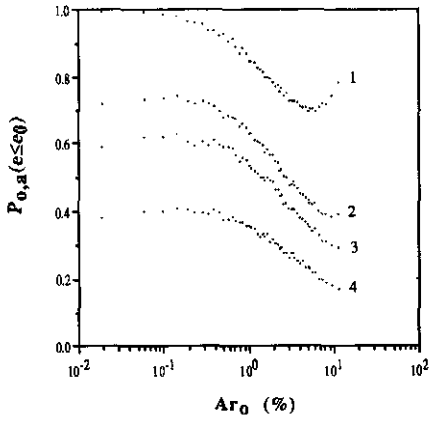

Fig. 2. Variation of the probability $P_{o,a}(e \leqslant e_0)$ as a function of the relative area of the observed peak, Ar_o, for different values of the accepted error, e_0, and for a saturation factor $m/T = 0.25$. e_0: (1) 0.5; (2) 0.1; (3) 0.05; (4) 0.01.

tion to this behaviour for the curve corresponding to $e \leqslant 0.5$, which exhibits a minimum at about $Ar_o = 5\%$). The conclusion drawn from these curves is that if a small observed peak were to be found in a chromatogram, the probability of performing correctly the determination of its most abundant component is relatively high. In other words, a small peak has little chance of being observed in the chromatogram, but once it is observed, it is likely to be relatively pure. The probability $P_{o,a}$ decreases when the observed peak becomes larger. Therefore, larger parent peaks have more chance of representing the most abundant peaks in the observed peak, but they are most likely to be contaminated by the interferences with other parent peaks.

The exception found when $e \leqslant 0.5$ can be explained by the fact that it is rare that the error in the measurement of the area of a large parent peak exceeds 50%. A similar result was obtained by Nagels *et al.* [7], but with a minimum at $Ar_o = 3\%$. This corresponding Ar_o value is most likely to vary with the saturation factor. The present simulation allows the estimation of the probability of correct determination for observed peaks with relative areas as low as 0.01%, which is ten times lower than the results obtained by Nagels *et al.*, probably because they used a relatively small number of parent peaks in their simulations. In fact, owing to the sampling from the parent peak-area distribution, the achievement of lower values of Ar_p relies on the simulation of a larger number of parent peaks, m, in the chromatogram, whereas the achievement of lower Ar_o values depends also on the saturation of the chromatogram.

Fig. 3. shows the variation of $P_{o,a}(e \leqslant 0.1)$ as a function of Ar_o for different values of the saturation factor, m/T. These curves can be used to estimate the probability of performing the quantitative determination of the most abundant peak in an observed peak, for different values of the saturation factor. As expected, at a fixed value of Ar_o, this probability of correct determination increases as the chromatographic saturation decreases. In addition, at a fixed value of Ar_o, the rate of increase of the probability with decreasing saturation factors is larger for larger values of the saturation factor. Thus, for $Ar_o = 1\%$, when the saturation factor decreases from 0.5 to 0.25 the probability $P_{o,a}(e \leqslant 0.1)$ increases by 44.5%, whereas starting with a saturation factor equal to 0.25 and decreasing it by half to 0.125 leads to an increase in the probability of only 23%.

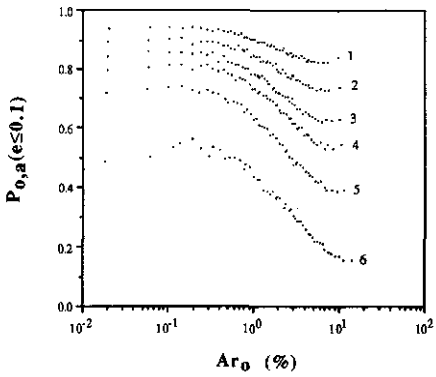

Fig. 3. Variation of the probability $P_{o,a}(e \leqslant 0.1)$ as a function of the relative area of the observed peak, Ar_o, for different values of the saturation factor m/T. m/T: (1) 1/20; (2) 1/12; (3) 1/8; (4) 1/6; (5) 1/4; (6) 1/2.

From these computations we can also represent the variation of $P_{o,a}(e \leqslant 0.1)$ as a function of the saturation factor, at different values of Ar_o. The variation of this probability, for different values of Ar_o, is reported on Figure 4, as a function of the reciprocal of the saturation factor, which is a more convenient representation as we know that the probability of correct determination tends towards zero for an infinite saturation factor. At a fixed Ar_o, as the saturation decreases, *i.e.*, as T increases for m fixed, $P_{o,a}(e \leqslant 0.1)$ increases rapidly at first, then tends more slowly toward 1. As discussed above, one sees that, for a fixed saturation factor, $P_{o,a}(e \leqslant 0.1)$ increases with decreasing Ar_o.

Probability of performing a quantitative determination, with an error smaller than or equal to a given value, of a given parent peak

In addition to the probability of performing the quantitative analysis of the most abundant peak in an observed peak, the chromatographer is also interested in knowing the probability associated with the quantitative determination, with an fixed error, of a given parent peak. Actually we are looking for the quantitative determination of a particular parent peak, in contrast to the previous probability where we were investigating each observed peak in the chromatogram, and computing the probability of determining the most abundant peak. Again, the probability of quantitative determination of a parent peak, with a fixed error, depends also on the degree of saturation in the chromatogram and on the relative area of the parent peak, Ar_p. This probability is referred to as $P_{p,a}(e \leqslant e_0)$. The error, e, is calculated on the basis of the comparison between the observed peak containing the parent peak and the parent peak itself. This probability is computed by accounting the number of times, N_a $(e \leqslant e_0)$, where the parent peak can be determined with an error smaller than or equal to e_0, when N_s simulations are performed. Hence the probability is calculated as

$$P_{p,a}(e \leqslant e_0) = N_a(e \leqslant e_0) \,/\, N_s \tag{3}$$

Variation of $P_{p,a}(e \leqslant e_0)$ with the relative area of a parent peak

For a fixed saturation factor, more than 10 000 simulations are performed in

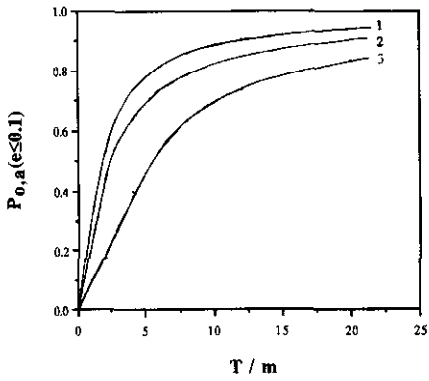

Fig. 4. Variation of the probability $P_{o,a}(e \leqslant 0.1)$ as a function of T/m, the reciprocal of the saturation factor, for different values of the relative area of the observed peak, Ar_o. Ar_o: (1) 0.1%; (2) 1%; (3) 10%.

order to estimate $P_{p,a}$ for different values of the accepted error. Fig. 5 shows the variation of $P_{p,a}(e \leqslant e_0)$ for a saturation factor equal to 0.25 and for different values of e_0. Obviously, for a fixed Ar_p and a given saturation factor, the probability of a correct quantitative determination increases as the accepted error increases. At a fixed saturation factor, it seems that the general shape of the curve of variation of $P_{p,a}$ $(e \leqslant e_0)$ with Ar_p depends on the value of the accepted error. Nevertheless, the global behaviour observed is an increase in the probability with increasing Ar_p, which differs strongly from the behaviour observed in Fig. 2 for $P_{o,a}(e \leqslant e_0)$ *versus* Ar_o. It is expected that the probability of performing a correct quantitative determination of a parent peak increases with its area, as larger parent peaks are relatively less affected by the overlap with smaller parent peaks. Nevertheless, it seems that this observation is not valid when the accepted error is relatively low, as the curves for $e \leqslant 0.01$ and, to a lesser extent, for $e \leqslant 0.05$ appear to become approximately constant when Ar_p exceeds 3% or even to exhibit a maximum at about $Ar_p = 3\%$. No simple explanation can be envisioned to account for this unexpected observation, which is associated with a relatively high saturation factor. This phenomenon is reported here for the first time. The calculations performed by Nagels *et al.* [7] did not show a similar observation, probably because the saturation factor corresponding to their computations was not high enough.

Fig. 6a and b show the variation of $P_{p,a}$ as a function of Ar_p for different values of the saturation factor, m/T, at two different accepted errors, $e_0 = 0.1$ and 0.5, respectively. Again, for a given saturation factor, the probability of correct quantitative determination is seen to increase as the area of the parent peak increases. This increase is almost linear for the curves in Fig. 6a corresponding to $e_0 = 0.1$, which means that the probability of correct determination varies exponentially with the relative area of the parent peak, whereas the rate of increase of the probability with Ar_p for $e_0 = 0.5$ (Fig. 6b) is much more pronounced, especially at relatively high saturation factors. For a fixed parent peak, the probability of its correct quantitative determination increases as the saturation factor decreases, because interferences between par-

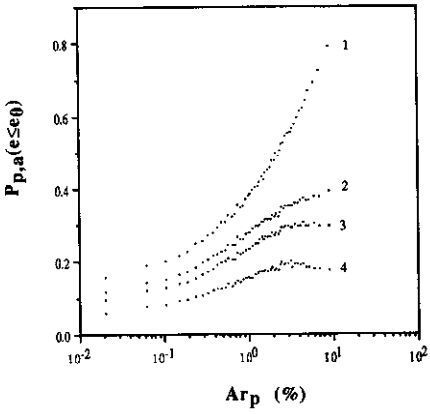

Fig. 5. Variation of the probability, $P_{p,a}(e \leqslant e_0)$, of performing a quantitative determination of the area of a parent peak, with an error smaller than or equal to e_0, as a function of the relative area of the parent peak, Ar_p, for a saturation factor $m/T = 0.25$, and for different values of the accepted error, e_0. e_0: (1) 0.5; (2) 0.1; (3) 0.05; (4) 0.01.

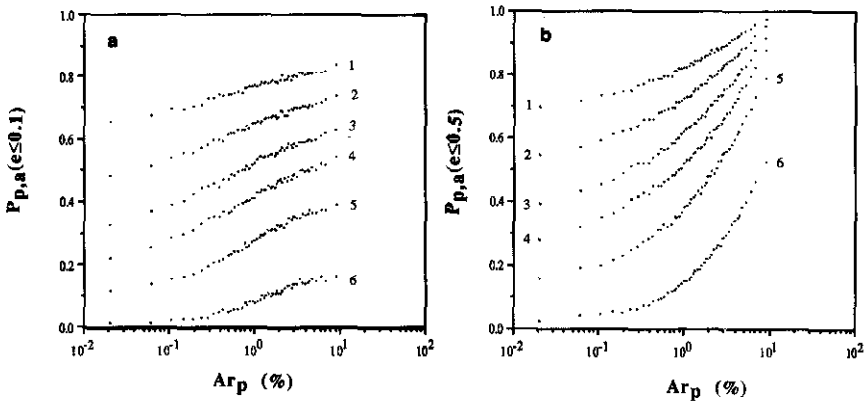

Fig. 6. (a) Variation of $P_{p,a}(e \leqslant 0.1)$ as a function of the relative area of the parent peak, Ar_p, for different values of the saturation factor. m/T: (1) 1/20; (2) 1/12; (3) 1/8; (4) 1/6; (5) 1/4; (6) 1/2. (b) Variation of $P_{p,a}(e \leqslant 0.5)$ as a function of the relative area of the parent peak, AR_p, for different values of the saturation factor. m/T: (1) 1/20; (2) 1/12; (3) 1/8; (4) 1/6; (5) 1/4; (6) 1/2.

ents peaks diminish when the degree of chromatographic saturation decreases.

These curves can be used in order to estimate the probability of correct determination for a given parent peak. In addition, they provide a means of estimating the upper limit of the saturation factor, *i.e.*, the minimum resolving power, required to achieve a given probability of performing correctly a quantitative determination within some tolerated error.

Computation of the quantitative determination limit

The quantitative determination limit, DL_e^p, corresponds to the minimum relative abundance of a parent peak, expressed as a percentage of the whole chromatographic response, which leads to a probability, p, of performing its quantitative determination with an error smaller than or equal to some fixed value e. Since the variation of the probability, $P_{p,a}(e \leqslant e_0)$, of quantitative determination, with an error smaller than or equal to e_0, of a parent peak of given relative area, Ar_p, has already been studied, it is then possible to estimate DL_e^p, for a given saturation factor, by using one of the curves in either Fig. 6a or b, depending on the chosen value of the accepted error. Hence, DL, for a given accepted error, corresponds to the value of Ar_p, estimated from the curve corresponding to the appropriate saturation factor, which leads to the desired probability of correct quantitative determination. Reversing the coordinates in Fig. 6a and b shows that, for a fixed accepted error and for a given saturation factor, DL_e^p increases very steeply with the required probability, p. In addition, for a given required probability, p, the smaller is the accepted error, the greater is the value of the determination limit.

The domain of the saturation factor, m/T, investigated in this study lies between 0.05 and 0.5. In other words, this corresponds to a degree of occupancy of the chromatographic space between 10% and 100% of the peak capacity, if the peak capacity is calculated on the basis of a 2σ separation. The corresponding interval becomes 14.2–142% of n_c, if n_c is calculated on the basis of a 2.84σ separation, as it should be in order to see as many maxima as there are peaks in the chromatogram when taking

into account the distribution of peak heights [14]. Within this domain of saturation investigated, it appears from Fig. 6a that it is impossible to achieve a probability of determination of the parent peak equal to 0.9 for an accepted error smaller than or equal to 10%. This means that in order to achieve a probability of determination as high as 0.9 with an error smaller than or equal to 0.1, the chromatographic saturation factor must be well below 0.05, even for the largest parent peaks in the mixture. However, if one chooses a probability of determination equal to 0.5, with an associated error smaller than or equal to 0.1, the quantitative determination limit, $DL_{0.1}^{0.5}$, is then equal to 8% and 1%, for $m/T = 0.167$ and 0.125, respectively. Fig. 6b shows that, with an accepted error of 0.5, it is possible to achieve probabilities of quantitative determinations as high as 0.9, especially at low saturation factors.

Fig. 7 shows the variation of two quantitative determination limit parameters, $DL_{0.1}^{0.5}$ and $DL_{0.5}^{0.9}$, as a function of the saturation factor. Both DL are seen to increase as the saturation increases, but this increase is more pronounced for $DL_{0.1}^{0.5}$, which is certainly due to the fact that the probability of correct determination is more sensitive to the increase in the saturation factor for low rather than for high values of the accepted error. Ultimately, whatever e and p, all DL curves will tend toward 100% at very large saturation factors.

In practice, for a given complex chromatogram, the saturation factor can be calculated by means of one of the recently developed procedures for estimating the number of sample components [1–5]. Then the values of various DL_e^p parameters can be estimated from curves such as those in Fig. 7. Alternatively, they can be used to evaluate the reduction of the DL values resulting in the improvement of the separation power of a chromatographic system or from switching to another more efficient separation system.

CONCLUSION

The problem of quantitative analysis from chromatograms of complex mixtures appears to be very serious. From a statistical point of view, even for the largest

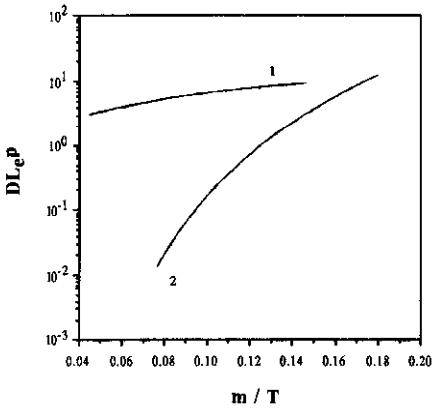

Fig. 7. Variation of the quantitative determination limit, DL_e^p, corresponding to a probability p of performing a quantitative determination of a parent peak with an error smaller than of equal to e as a function of the saturation factor m/T. (1) $p=0.9$ and $e=0.5$; (2) $p=0.5$ and $e=0.1$

parent peaks the probability associated with their quantitative determination, with an error smaller than or equal to given value, can be very low, especially if the chromatographic saturation factor is relatively high. In order to achieve a reasonable probability of correctly performing a quantitative determination, the chromatographic system must exhibit a very high resolving power, *i.e.*, it must have a peak capacity greatly exceeding the number of components in the sample. In addition, it was found that small observed peak are most likely to be pure but on the other hand, the probability of finding them in a typical complex chromatogram is relatively low. One must be more careful when quantifying large observed peaks, as they are most likely to be the result of two or more parent peaks lumped together.

The use of the quantitative determination limit, DL_e^p, permits the quantification, in probabilistic terms, of the effect of peak overlap phenomena on quantitative determinations. This parameter is expressed in terms of a fraction of the whole chromatographic response. Thus, for a component of interest, it is easy, when its response factor is known, to convert DL into a specific parameter, SDL_e^p expressed as the weight fraction of solute in the analysed mixture, which allows one to quantify the ability of the chromatographic system to analyse this particular solute, in the studied matrix. Indeed, the parameter SDL_e^p, which is associated with the "horizontal" sources of error (the interferences between parent peaks), can be compared with the classical limit of determination, which reflects the effect of the "vertical" source of error (the influence of noise in the signal). As they most likely are independent, the largest of these two parameters will impose the overall limit of determination. In addition, the curves of the variation of DL_e^p (or SDL_e^p) as a function of the saturation factor allow one to estimate the efficiency of the chromatographic system required to obtained a reasonable probability of performing the quantitative determination of a given parent peak within some tolerated error.

Of major importance here is the fact that the larger is the response factor of a solute with a particular detector, with respect to the whole chromatographic response, the lower is the amount of this component which can be determined with a reasonable probability and an accepted error, that is, the more likely is the chance that this solute produces a chromatographic signal which is above the *DL* level. In addition to seeking an increase in the resolving power of the chromatographic system when analysing complex mixtures, it seems that one of the major tasks for the analyst interested in the quantitative determination of one or a few components is certainly the optimization of the detection, in order to make it more selective for the components of interest. This step in any development of an analytical procedure can be decisive for its success. In this respect, the present statistical study allows one to express quantitatively the improvement brought into the validity of the quantitative determination of a component by any increase in its relative response factor.

This study has been presented within the context of chromatography, but it is clear that its results can be extended to any kind of zonal separation method provided that the hypotheses underlying the simulation model (randomly distributed Gaussian zones with constant standard deviation) can correspond to realistic attributes of the method.

ACKNOWLEDGEMENTS

The authors are very grateful to P. Cousot of the Computer Centre of the Ecole Polytechnique (Palaiseau, France) for allowing them free access to the facilities of the Centre. They are also grateful to Rhône-Poulenc Recherches (Courbevoie, France) for partially supporting this research. M.Z.E. thanks the Ecole Polytechnique for financial support during his Thesis work. This work was presented in part at the *13th International Symposium on Column Liquid Chromatography, Stockholm, June 25–30, 1989.*

REFERENCES

1 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 55 (1983) 418.
2 M. Martin, D. P. Herman and G. Guiochon, *Anal. Chem.*, 58 (1986) 2200.
3 M. Z. El Fallah and M. Martin, paper presented at the *17th International Symposium on Chromatography, Vienna, September 25–30, 1988.*
4 J. M. Davis, *J. Chromatogr.*, 449 (1988) 41.
5 A. Felinger, L. Pasti and F. Dondi, *Anal. Chem.*, 62 (1990) 1846.
6 J. C. Giddings, *Anal. Chem.*, 39 (1967) 1072.
7 L. J. Nagels, W. L. Creten and P. M. Vanpeperstræte, *Anal. Chem.*, 55 (1983) 216.
8 L. J. Nagels and W. L. Creten, *Anal. Chem.*, 59 (1987) 822.
9 L. A. Currie, *Anal. Chem.*, 40 (1968) 586.
10 P. W. J. M. Boumans, *Spectrochim. Acta Part B*, 33 (1978) 625.
11 L. R. Snyder, *High performance Chromatography, Advances and Perspectives*, Vol. 1, Academic Press, New York, 1980, p. 207.
12 H. W. Habgood and W. E. Harris, *Anal. Chem.*, 32 (1960) 450.
13 M. Martin, to be published.
14 M. Z. El Fallah and M. Martin, *Chromatographia*, 27 (1987) 115.
15 M. Martin and G. Guiochon, *Anal. Chem.*, 57 (1985) 289.
16 D. Rosenthal, *Anal. Chem.*, 52 (1982) 63.