Journal of Chromatography, 289 (1984) 277–298
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 228

ORIGIN AND CHARACTERIZATION OF DEPARTURES FROM THE STATISTICAL MODEL OF COMPONENT-PEAK OVERLAP IN CHROMATOGRAPHY

JOE M. DAVIS and J. CALVIN GIDDINGS*
Department of Chemistry, University of Utah, Salt Lake City, UT 84112 (U.S.A.)

SUMMARY

A statistical model of component-peak overlap in complex chromatograms is reviewed. Procedures for the estimation of the number of components in an analyte from its chromatograms by means of this model are restated. We note that the statistical model does not account for the effects of certain realistic chromatographic attributes. The influences of component-peak density, amplitude range, asymmetry, and noise levels on the estimation of the average component number are determined empirically with computer-generated chromatograms and are quantified by analyses of variance. We find that small departures from the model arise from variations in the magnitude of the amplitude range, density and the noise level. A large departure from theory arises from an application of the model to chromatograms containing highly asymmetric component-peaks. In spite of these departures, the estimation of the component number from chromatograms containing randomly distributed Gaussian component-peaks is uniformly more accurate with the use of the model than from a counting of peak maxima in chromatograms of extraordinarily high resolving power.

INTRODUCTION

In a recent paper, the authors presented a simple statistical model predicting the degree of component-peak overlap in complex chromatograms containing randomly spaced components[1]. In a later publication, the essential correctness of the results predicted by the model were verified by a computer generation of complex, synthetic chromatograms[2]. In particular, the simulation results confirmed the validity of a proposed procedure for estimating the number of components in an analyte from a series of its chromatograms, all having substantial component overlap.

The model of component-peak overlap is derived from Poisson point statistics with the implicit assumption that each component maximum (or center of gravity) may be represented by a point positioned on an elution volume or time axis[1]. This assumption does not provide any theoretical description of certain complicating but realistic chromatographic attributes and of their effects on the results predicted by our model. These attributes are principally the relative amplitude range of the com-

0021-9673/84/$03.00

ponents, density, noise, baseline stability and component-peak asymmetry. The severity of their influence on the predicted results will determine the overall usefulness of the model to chromatographic science.

We have successfully established two methods for the estimation of the number of components (many of them hidden by overlap with other components) in synthetic chromatograms[2]. We shall now augment this work and determine empirically the effects of noise, amplitude range, density and component-peak asymmetry on the results predicted by the application of these two procedures. We shall assume that the determination of the number of components in a complex analyte is one of the primary applications of our model and shall, in this paper, restrict our study to the authenticity of this number. Computer simulations are used in lieu of experimental chromatograms because, with the former, exact control over the statistical distribution and the properties of the component-peaks is possible[2]. To determine quantitatively the magnitude and significance of any departures from the model, we shall use the method of analysis of variance (ANOVA).

THEORY

Peak overlap

The number of observed peaks ("peaks"), p, in a chromatogram containing randomly distributed components is related to the mean or expected number of components, $\bar{m}$, and to the peak capacity, n_c, by[1,2]

$$p = \bar{m}e^{-\bar{m}/n_c} \tag{1}$$

or, in reduced form,

$$p = n_c \, \alpha e^{-\alpha} \tag{2}$$

in which case the saturation factor, α, equal to $\bar{m}/n_c$, is a measure of the saturation of the separation space by the components. The peak capacity, n_c, is the maximum number of component-peaks which can be accommodated in the chromatogram with some minimum resolution between adjacent component-peaks. Eqn. 1 may be linearized by taking logarithms of both sides, *viz.*,

$$\ln p = \ln \bar{m} - \bar{m}/n_c \tag{3}$$

We have suggested[1] and confirmed[2] that plots of the logarithm of the number of visually determined peaks *versus* reciprocal peak capacity provide reasonable estimates of $\bar{m}$ from both slope and intercept. Our data were drawn from the interpretation of computer-generated chromatograms containing randomly spaced Gaussian components.

We shall assume that our hypothesis of random component distribution is valid in this paper. We have noted that this hypothesis is not applicable in all chromatograms and that some chromatograms are highly ordered[1,2]. The distribution of peaks in many complex chromatograms is nevertheless unstructured. We must characterize the departures from our model which arise from the chromatographic attributes previously cited before we can separate them from departures which arise from

a nonrandom distribution of components.

The number of peaks, p, is a function of the peak capacity, n_c, which is determined by the total separation space, X, over which the model applies, the average standard deviation, σ, of two adjacent components and the critical resolution, R_s^*, which is needed to discriminate between peaks:

$$n_c = \frac{X}{x_0} = \frac{X}{4\sigma R_s^*} \tag{4}$$

where x_0 is the minimum distance between adjacent peaks yielding the minimum acceptable resolution, R_s^*.

In the context of our model, a peak is "a detected cluster of one or more components in which the first and last components in the cluster are separated by components adjacent to the cluster by a resolution greater than or equal to R_s^* and in which each member of the cluster is separated from adjacent members of the cluster by a resolution of less than $R_s^{*\prime}$ (ref. 2). High-resolution chromatographic maxima may be peaks in this context with a proper choice of resolution. One of our peak-counting procedures is based on this possibility. It is important to recognize, however, that this choice is only one of many possible alternatives.

The dependence of the magnitude of R_s^* on relative component amplitude has been noted[2]. Our success in the estimation of $\bar{m}$ from baseline-resolved peak counts, with $R_s^* = 1.5$, was reported[2]. While these counts are virtually independent of the relative component amplitude, the problems of baseline drift, noise and the statistics of peak counting preclude baseline resolution as the optimal counting procedure[2]. We also reported that a meaningful empirical value of R_s^* was calculable with the proper fitting of maxima peak counts to a least-squares procedure[2]. In this procedure, an effective saturation, equal to $4\sigma\bar{m}/X$, was calculated from the known mean component number. A plot of the logarithm of the maxima count *versus* this effective saturation was made. An examination of eqns. 3 and 4 reveals that the resolution factor, R_s^*, is calculated as a slope from this procedure. The factor R_s^* is the mean resolution with which maxima are distinguishable and is reasonably postulated to be a function of relative component-peak amplitude, density range, symmetry and noise levels. We then confirmed that this empirical value could be used to determine n_c and, by eqn. 3, $\bar{m}$ from slope and intercept over a given component amplitude and density range and for a given level of noise.

Analysis of variance

The analysis of variance (ANOVA), as developed by Fisher, is a systematic means by which one may determine if one or more groups of data statistically differ from one or more other groups of data. This methodology is an indispensable statistical tool which, in the hands of the chemist, physicist and biologist, provides an added dimension and insight into the nature of variation in natural events[3].

In this work, we will determine $\bar{m}$ from synthetic chromatograms generated with different amplitude ranges, densities (α), asymmetries and noise levels. We will then use ANOVA techniques to decide if the determined $\bar{m}$ values are significantly affected by variations in these attributes. We will utilize two forms of the Model I ANOVA. In a Model I ANOVA, it is assumed that differences in the means of the

data groups under study arise from fixed and clearly definable differences among the data groups[3]. The simplest procedure is the single classification or one-way ANOVA in which the significance of one source of variation is determined with respect to all other sources of variation. As an example, one can determine with a one-way ANOVA if the values of an analyte level reported from different laboratories are statistically equivalent. The second procedure is a one-way nested ANOVA in which the members of each data group also vary due to random (and usually uncontrollable) factors. One may use this procedure to determine if variations within groups, arising from random factors, are significant relative to the variations among the main groups arising from the fixed differences. As an example, one might study the percentage weight increase in two groups of mice which are fed two different food supplements. Since the mice are genetically different, some will inherently grow more rapidly than others on either food supplement. The experimentalist has no control over the genetic factors which influence growth rate. With a nested ANOVA, however, the magnitudes of the variation in growth within and between the groups can be properly compared to determine if one supplement significantly influences the growth rate over the other.

Our groups are composed of data calculated from applications of the two counting procedures. Five classes of data may be defined. Estimates of $\bar{m}$ may be computed as the slope (a) and the intercept (b) from an application of the baseline counting procedure and as the intercept (c) from the application of the maxima counting procedure. Empirical resolution factors (d) are computed as a slope from the application of the maxima counting procedure. A procedure (e) for the estimation of $\bar{m}$ based on point chromatograms will be briefly discussed.

The physical meaning of the one-way ANOVAs computed in this paper is now discussed. We shall propose a null hypothesis that the data within all groups are drawn from the same parent population. The variance of the parent distribution is then calculable in two ways. One method is to compute a weighted sum of squares (SS) based on the individual and calculable variances of each group. The SS is the sum of the squared deviations of the members of the group with respect to the group mean. The weighting factor for each group variance is the number of degrees of freedom in the group. The second method is to compute a SS from the mean values of each group with respect to the mean value of the entire data. When the SS are properly weighted by the number of degrees of freedom in the calculations, they become estimates of the variance of the parent population. These quantities will not necessarily be equal but will be similar in magnitude if our assumption of a single parent distribution is correct. If any measurable systematic variation among the groups is present, the group means will be dissimilar. The variance calculated from the means of each group, or the variance among groups, is then larger than the weighted variance, or variance within groups, calculated from the variances of each group.

If the ratio of these variances exceeds a critical Fisher ratio F, we may state that, within some quantitative degree of certainty expressed as a confidence limit, the two variances are not measures of the breadth of the same parent population, the fixed and definable differences among the groups are measurable, and differences among the groups exist. The test is labeled significant if this condition is met.

We will conform our symbolism to the standard conventions of the literature.

The mean squares (MS) are the sums of squares of a group divided by the degrees of freedom (df) in the calculation. The calculated Fisher ratio (F) is computed by the ratio of the mean squares among groups to the mean squares within groups. The critical value from the F distribution, against which comparisons are made, is $F_\gamma[\nu_1, \nu_2]$, in which ν_1 and ν_2 are the degrees of freedom among groups and within groups, respectively, and γ is the area of the normalized distribution to the right of the F value (the symbol α is normally used for this area, but we have adopted this symbol for our saturation factor). A simple symbolism for the 95% confidence level which we shall use is F^*. Ref. 4 may be consulted for tabulated values of F^*. All ANOVA computations utilize one-tailed F tests, and the calculated Fisher ratios may be less than unity. An ANOVA which is not significant (NS) confirms the null hypothesis that the data are drawn from the same population.

A prerequisite for the correct and meaningful application of ANOVA is homogeneity among the variances of the compared groups or homoscedasticity. Bartlett's test for homogeneity is applied to all ANOVAs presented in this paper. For those ANOVAs which are inherently heteroscedastic, Snedecor's procedure is used to approximate an ANOVA. The reader may find details on all these procedures in ref. 3.

PROCEDURES AND METHODS

The calculation of Gaussian component-peaks, the generation of Gaussian noise and the production of the synthetic chromatograms with a Versatek plotter were discussed in detail in ref. 2. In addition, the generation of tailing asymmetric component-peaks is conveniently made via exponentially convoluted Gaussians. The amplitude $h(t)$ of a computer-generated chromatogram of tailing component-peaks is given by the following sum of convolution integrals[5]:

$$h(t) = B + \sum_{n=1}^{m} \frac{A_n}{\sqrt{2\pi}\,\sigma_n\tau_n} \int_0^\infty \exp\left[\frac{-(t - t_{rn} - t')^2}{2\sigma_n^2}\right] \exp\left(\frac{-t'}{\tau_n}\right) dt' \qquad (5)$$

An equivalent and computationally convenient equation is

$$h(t) = B + \sum_{n=1}^{m} \frac{A_n\,\sigma_n}{\sqrt{2}\,\tau_n} \cdot \exp\left[\frac{1}{2}\left(\frac{\sigma_n}{\tau_n}\right)^2 - \frac{(t - t_{rn})}{\tau_n}\right] \int_{-\infty}^{z_n} \exp(-x^2)dx \qquad (6)$$

$$z_n = \left(\frac{t - t_{rn}}{\sigma_n} - \frac{\sigma_n}{\tau_n}\right) / \sqrt{2}$$

With reference to the nth component, A_n is the amplitude of the pure Gaussian component at time $t = t_{rn}$, t_{rn} is the component retention time, σ_n is the pure component standard deviation in time units, τ_n is the exponential dilution time constant, B is the baseline offset amplitude and m is the total number of components.

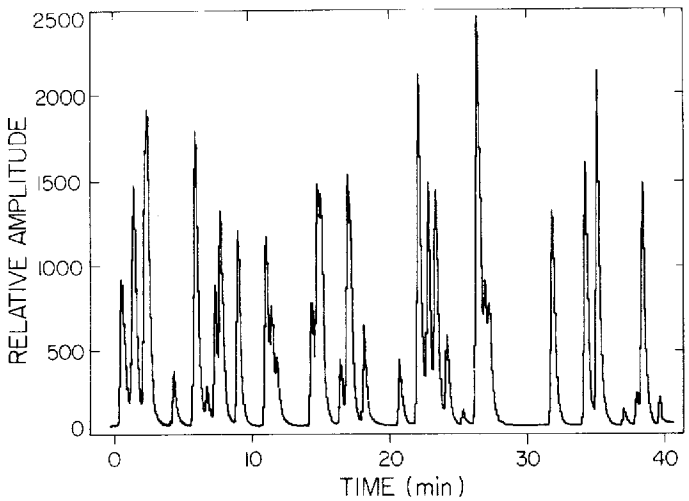


Fig. 1. Fifty exponentially convoluted Gaussian components with $\sigma/\tau = 0.35$. Amplitude range: 30–4500 ADCs. $\sigma = 4$ sec.

Fig. 1 is a computer simulation of 50 exponentially convoluted Gaussian components which are randomly spaced over 40 min with $\sigma = 4$ sec and $\sigma/\tau = 0.35$. The departure of the overlapping components from Gaussian shape is significant and increases as the ratio σ/τ decreases. The severity of component-peak overlap is illustrated via the visual detection of only 34 maxima.

The amplitudes of the components were chosen randomly to simulate chromatograms in a very general manner. (While the component-peak amplitude range is a variable of interest, only the amplitudes of the components themselves can be conveniently controlled.) The standard deviations, σ, of the components were identical in a given simulation to model the approximately constant component-peak widths observed in programmed analyses. The σ/τ ratio was held constant in the computation of convoluted Gaussians to model a column or detector inefficiency which acts as a dilution chamber. The kinetics of adsorption chromatography are not well modeled by this procedure[6].

The methodology for counting baseline-resolved peaks and peak maxima has been published[2]. Additional synthetic chromatograms were produced to augment this earlier study. To obtain estimates of $\bar{m}$ via eqn. 3, five chromatograms, each of which contained the same component distribution resolved to a different level, were generated to produce five data points for the ln p *versus* $1/n_c$ plot. The data consisted of the number of visually determined peaks at a given peak capacity. The five chromatograms were composed of components with respective standard deviations of 12, 10, 8, 6 and 4 sec.

The fifth class of data groups (e) is composed of estimations of $\bar{m}$ from "point chromatograms". A series of point chromatograms was generated to determine the effects of amplitude on our predicted results. These were composed of randomly chosen retention times only. No amplitudes were assigned. The number of "observed" peaks was determined by counting the number of adjacent retention time differences which exceeded the value x_0 representing the given level of efficiency.

The data were fitted to eqn. 3 by the procedure of least squares. The weight of each peak count, in accordance with Poisson statistics and the required transformation from the exponential dependence of eqn. 1 to the linear dependence of eqn. 3, was the peak count itself. The variances in slope and intercept were calculable from least-squares equations[4].

The statistical uncertainties in both slope and intercept do not enter into our ANOVA computations. The validity of the ANOVA procedure rests on the stipulation that each datum carries the same statistical weight. Each slope, intercept or empirical resolution value carried a weight of unity in our calculations.

Characterization of the simulated chromatograms

Table I contains the information required to characterize the generated synthetic chromatograms. The contents of the second column identify the set members as synthetic component-peak chromatograms or point chromatograms. The third column contains the number of simulation series, each of which is composed of five chromatograms, created for that particular set. Although the intercept of the plot stemming from eqn. 3 has the value ln $\bar{m}$, we shall loosely use the word "intercept" to refer to $\bar{m}$ directly. The reported standard deviations in the final columns are based on the arithmetic means of the calculated $\bar{m}$ and R_s^* values.

The remaining columns are self-explanatory, but the units need clarification. The noise level is the standard deviation of Gaussian noise in analog-to-digital (ADC) units as discussed in ref. 2. The minimum and maximum component amplitudes are also given in ADC units. The minimum amplitude value of 10 ADCs corresponded to 0.02 in. in our synthetic chromatograms. The values of α are measures of the component-peak density and are defined by eqns. 2 and 4 with the arbitrary value $R_s^* = 1.5$.

The data sets that were partially analyzed in ref. 2 are indicated by an asterisk.

Description of ANOVA procedures

Table II contains the nine general ANOVA procedures conducted on the data in Table I. The letters of the compared sets lie to the left of a colon in the third column. The data examined were estimates of $\bar{m}$ from slope and intercept and empirical resolution factors. These are symbolically indicated as Sl, In and R_s^*, respectively. These symbols lie to the right of the colon in the third column. The remaining columns are self-explanatory.

Baseline peak counts using the critical resolution $R_s^* = 1.5$ are not completely independent of amplitude. The relative minimum value of the amplitude envelope between two components separated by six average standard deviations may not coincide with the baseline. The degree of coincidence is determined by the relative amplitude range, baseline stability, noise, component-peak asymmetry and visual acumen.

ANOVA I was performed to determine if the presence of amplitude differences influences the number of visually determined peaks as postulated in ref. 2. Any departure from the model arising from the effects of amplitude differences should increase with high levels of noise because the error in the determination of the number of baseline-resolved peaks increases. Therefore, ANOVA II was computed to determine firstly if a visual estimation of the number of baseline peaks is influenced by

TABLE I

COMPOSITION OF SYNTHETIC CHROMATOGRAMS

Set	*Type*	*No. of simulations*	*Noise*	*Amplitude range*	σ/τ	*Baseline* α	*No. of components*	*Slope from baseline counts*	*Intercept from baseline counts*	*Intercept from maxima counts*	*Empirical resolution*
A*	Peak	9	0	100–1800	∞	0.18–0.55	80	82.18 ± 14.86	77.79 ± 5.45	77.91 ± 3.18	0.5226 ± 0.1757
B	Point	9	0	0	–	0.18–0.55	80	88.01 ± 18.42	82.69 ± 5.75	–	–
C*	Peak	12	0	100–1800	∞	0.37–1.10	160	163.60 ± 16.95	153.54 ± 7.75	149.94 ± 7.73	0.5121 ± 0.0784
D	Point	12	0	0	–	0.37–1.10	160	159.14 ± 19.31	158.02 ± 12.02	–	–
E*	Peak	9	30	100–1800	∞	0.37–1.10	160	147.02 ± 14.96	146.16 ± 8.42	145.62 ± 5.05	0.4963 ± 0.0900
F*	Peak	9	50	100–1800	∞	0.37–1.10	160	152.44 ± 8.21	150.07 ± 5.66	141.70 ± 6.28	0.4250 ± 0.0455
G	Peak	8	0	100–1800	0.35	0.37–1.10	160	465.10 ± 64.44	139.35 ± 26.26	147.46 ± 7.71	0.7211 ± 0.0680
H	Peak	9	0	10–900	∞	0.37–1.10	160	–	–	155.98 ± 7.21	0.5182 ± 0.1094
I	Peak	9	0	10–1800	∞	0.37–1.10	160	–	–	154.73 ± 6.72	0.5095 ± 0.1078
J	Peak	9	0	100–1800	∞	1.03–2.13	200	210.55 ± 14.55	189.62 ± 24.19	183.17 ± 18.88	0.5184 ± 0.0811
K	Point	9	0	0	–	1.03–2.13	200	201.25 ± 20.24	204.36 ± 26.33	–	–
L*	Peak	9	0	100–1800	∞	0.55–1.65	240	266.28 ± 14.35	246.57 ± 18.15	220.15 ± 10.79	0.4682 ± 0.0764
M	Point	9	0	0	–	0.55–1.65	240	242.05 ± 21.85	243.66 ± 20.78	–	–

Gaussian noise sufficiently to yield significantly different estimates of $\bar{m}$. The possible acceptance of the null hypothesis, on the basis of the Fisher ratio, in ANOVA II does not imply that the results are consistent with the point model. The possible insignificance of the Fisher ratio calculated from ANOVA II implies only that the results predicted from a counting of noiseless and noisy Gaussian baseline-resolved peaks are comparable. Therefore, ANOVA III was calculated to determine if the results predicted from a series of noisy simulations are comparable to results predicted from the point chromatograms. Finally, ANOVA IV is an assessment of the magnitude of departure introduced by component-peak tailing (convolution).

The magnitude of the empirical value R_s^* is determined and limited by the chromatographer's ability to distinguish maxima. ANOVAs V, VI and VII are assessments of the respective effects of component-peak density, amplitude range and noise on the magnitude of R_s^*. ANOVA VIII is an evaluation of the effects of component-peak tailing on R_s^*.

Finally, a comparison of the intercepts predicted by both counting methods

TABLE II

ANOVAs PERFORMED ON SETS A–N

ANOVA number	*Type*	*Variable*	*ANOVA groupings*	*Homoscedasticity*	*Significance, 95% confidence level*
I	One-way	Gaussian amplitude	(a) A,B:Sl,In	(a) Yes	(a) NS
			(b) C,D:Sl,In	(b) Yes	(b) NS
			(c) J,K:Sl,In	(c) Yes	(c) NS
			(d) L,M:Sl,In	(d) Yes	(d) For slope: $F = 7.69$; $F_{1,16}^{*} = 4.50$
II	One-way; nested	Noise	C,E,F:Sl,In	Yes	For slope: $F = 3.73$, $F_{2,27}^{*} = 3.36$
III	One-way; nested	Irreducible amplitude and noise	(a) D,E:Sl,In	(a) Yes	(a) NS
			(b) D,F:Sl,In	(b) No	(b) For intercept: $F = 5.12$; $F_{1,16}^{*} = 4.50$
IV	One-way; nested	Convoluted Gaussian amplitude	C,G:Sl,In	No	For slope: $F = 167.74$; $F_{1,8}^{*} = 5.32$
V	One-way	Component density	A,C,K,L:R_s^*	No	NS
VI	One-way	Component amplitude	C,H,I:R_s^*	Yes	NS
VII	One-way; nested	Noise	C,E,F:R_s^*	Yes	For slope: $F = 4.05$; $F_{2,27}^{*} = 3.36$
VIII	One-way; nested	Convolution	C,G:R_s^*	Yes	For slope: $F = 33.60$ $F_{1,18}^{*} = 4.43$
IX	One-way	Intercept calculation	(a) A:In	(a) Yes	(a) NS
			(b) C:In	(b) Yes	(b) NS
			(c) E:In	(c) Yes	(c) NS
			(d) F:In	(d) Yes	(d) For intercept: $F = 8.85$; $F_{1,16}^{*} = 4.50$
			(e) G:In	(e) No	(e) NS
			(f) J:In	(f) Yes	(f) NS
			(g) L:In	(g) Yes	(g) For intercept: $F = 14.08$; $F_{1,16}^{*} = 4.50$

is made in ANOVA IX. Since the maxima counting procedure is not fully justified theoretically, no rigorous reason exists for the expectance of comparable mean intercepts.

Some elaboration is necessary to justify the ANOVA choices and purposes. ANOVA I could be performed as a two-way ANOVA in which the component density of the four groupings is the second variate, but little useful information would be gained beyond what we have extracted.

A description of the random generation of component amplitudes and retention times by an allocation of two random number seeds was given in ref. 2. The sets C, E and F; A and B; C and D; and K and L were generated, through a lack of foresight, with different seeds and different component distributions. The interpretation of any significant ANOVA from these groups was questionable. If a test was significant, we also performed a nested ANOVA on the identical data to verify that the variations observed arose from the main variate and not from the different seed choices. With this procedure, we decomposed randomly the members of each group into smaller groups and treated the variations due to seed choice as random factors. In all of the ANOVAs for which this approach was required, the variation arising from seed choice was insignificant relative to the variation in the selected variate. This approach both circumvented the need to generate vast amounts of additional data and removed any doubts about our conclusions.

The variables of ANOVA III are the combined effects of noise and amplitude on the results predicted from the point model. The control group is a set of point simulations. Ideally, this analysis should be performed as a two-way ANOVA since these two variates are independent. There is, however, no meaningful way to introduce noise into a point simulation.

ANOVAs V, VI and VII can be combined into a three-way ANOVA in which the combined effects of density, amplitude and noise on the empirical values of R_s^* are determinable. An extensive generation of additional chromatograms and data would be required. This additional work was not undertaken here.

RESULTS AND DISCUSSION

The standard deviations calculated from our values of $\bar{m}$ are larger for estimations from point simulations than from chromatographic peak simulations. The respective simulations were calculated on two different computers with slightly different random number generator library functions[2]. The variances are not statistically different with the one exception of the compared groups in ANOVA IIIb.

Significant differences in the slopes from Sets L and M are reported for ANOVA Id in Table II. The variable is amplitude. The attribute which distinguishes Set L from Sets A, C and J in ANOVA I is the baseline saturation range (α). The range in Set L varies from a baseline α value less than unity to a value greater than unity, whereas the principal range of the remaining sets is either less than unity (A, C) or greater than unity (J).

In ref. 2, we suggested that the additive amplitude of overlapping component-peaks extends the width of a peak beyond its theoretical value. We then argued that the number of visually determined baseline peaks in highly saturated chromatograms should be less than the theoretical prediction. This "third body amplitude"

hypothesis[2] is justified by computer simulation and explains the significance of ANOVA I. The confirmation of this hypothesis is discussed briefly below.

The baseline-resolved peaks in our chromatograms were determined by visual inspection[2]. Any components contained in an amplitude response departing from and returning to the baseline were counted as one peak. The error in the estimation of baseline resolution in noiseless simulations, which arises from a loss of visual acumen and principally from a lack of uniformity in the baseline printed by our plotter, was conservatively estimated as 0.01 in. This value corresponded to 5 ADC units in the chromatograms produced.

In a series of twelve simulations computed at eight different degrees of saturation, 200 components were distributed randomly over an 18-fold (100–1800 ADCs) amplitude range. The number of peaks was then computed by scanning the calculated amplitudes. Any amplitude range lying between the two extremes of 5 units was counted as a peak. In addition, the number of theoretical peaks based on the Poisson model was calculated, as previously described for the point chromatograms, from the same distribution of retention times. The peak counts from each procedure were then averaged and plotted for comparison with eqn. 2. Fig. 2 is a composite graph of the data and clearly demonstrates that fewer actual baseline peaks than theoretical baseline peaks are observed when the baseline α value roughly exceeds the value unity. The bar bracketing each point is the standard deviation of the p/n_c ratio.

The departure found in ANOVA I is now explicable. The baseline peak counts obtained from the highly efficient ($\alpha < 1$) chromatograms in Set L agree well with the predictions of the point model, whereas the peak counts from the poorly efficient ($\alpha > 1$) chromatograms are less than predicted by the model. A least-squares fit of these data to eqn. 3 predicts a slope considerably higher than the theoretical value. Because the severity of the "third body" effect decreases continuously with increasing separation efficiency (decreasing α), the mean values of the intercepts for Sets L and M exhibit no statistical difference.

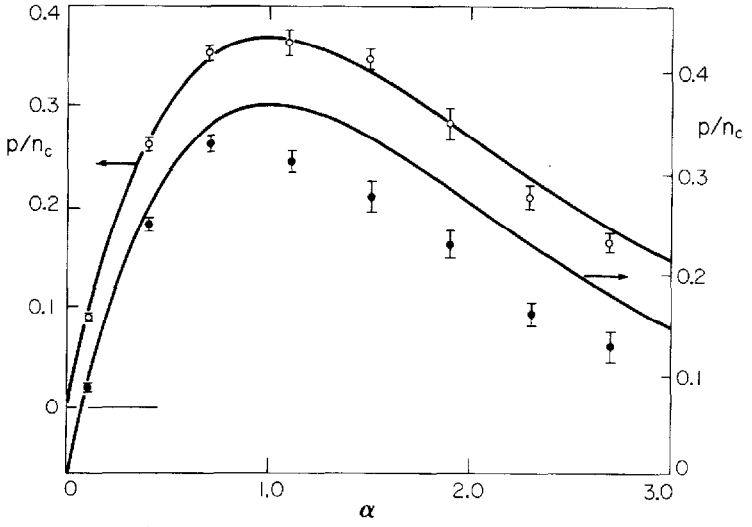
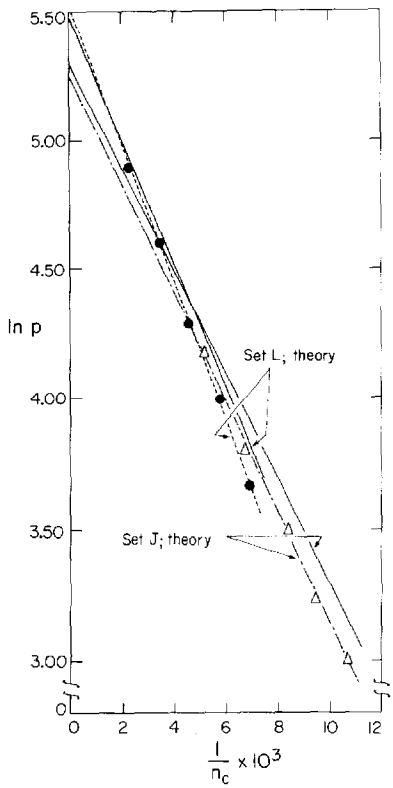


Fig. 2. Comparison of eqn. 2 and simulation results from baseline peak counting. Upper curve: theoretical peaks. Lower curve: peaks counted with error tolerance of 0.01 in.



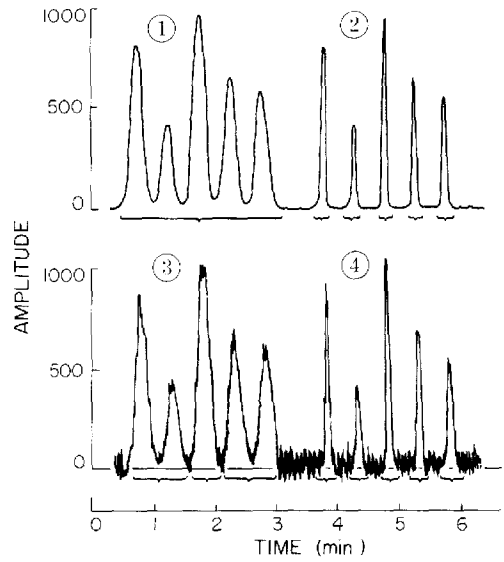


Fig. 3. Logarithm of baseline peak counts from Sets J and L *versus* reciprocal peak capacity with $R_s^* = 1.5$. Solid lines, theoretical; dashed lines, least-squares fits.

Fig. 4. Ordered distributions of highly resolved components at two levels of efficiency and noise. Amplitude range: 400–1000 ADCs. The noise is Gaussian with $\sigma = 50$ ADCs.

The logarithm of the mean value of each of the peak counts from Set L is plotted against reciprocal baseline peak capacity in Fig. 3. The standard deviations in the logarithm of the peak counts are not depicted, and will not be depicted in similar graphs, for reasons of clarity. A departure in the value of ln p (and therefore p) with decreasing peak capacity is observed relative to the theoretical line based on eqn. 3. In contrast, no significant difference between the slopes of Sets J and K is observed. In these saturated chromatograms, the baseline α range exceeds unity. The suppression of the magnitude of the baseline peak counts appears roughly constant for α values greater than 1 and less than 3. This observation is drawn from the distribution of data in Fig. 2. A plot of the logarithm of each of the mean peak counts from Set J *versus* reciprocal baseline peak capacity is also shown in Fig. 3. The value of the intercept is lower than the value predicted by theory owing to the lost peak counts. The least-squares slope, however, roughly parallels the theoretical line.

The significant results from the computations of ANOVAs II and III both arise from the influence of noise on the correctness of baseline peak counts. The mean estimations of $\bar{m}$ from both slope and intercept are lower in Sets E and F than

in Sets C and D. A small difference in the mean slopes of noisy and noiseless chromatograms is established by ANOVA II. An *a posteriori* analysis[3] clearly established a difference between the two noisy sets and the one noiseless set. A difference in the mean value of the intercepts calculated from the more noisy series and from point simulations is established by ANOVA III. Both departures arise from the same origin and will be discussed together.

The difficulties in the visual determination of baseline resolution are multiplied by low signal-to-noise ratios. The difficulties increase because the eye cannot interpret the signal near the baseline independently of the noise, and false judgments of the departure from and return to baseline are easily made. This concept may be easily conveyed graphically.

Fig. 4 is an ordered distribution of five components resolved at two levels of efficiency. The distributions were produced by our plotter and traced for clarity. Sub-figures 1 and 3, and 2 and 4, differ only in the presence or absence of noise. Using a criterion of baseline resolution, the numbers of baseline-resolved peaks in sub-figures 1–4 are respectively one, five, three and five. Two additional baseline-resolved peaks are falsely counted in sub-figure 3 because of the random excursions of noise below the estimated baseline. There is no analogous problem with the counting of baseline-resolved peaks in the two sub-figures of high efficiency because the components are well-resolved.

Fig. 5 is a plot of the logarithm of each of the mean peak counts from Sets C, E and F against reciprocal peak capacity. The data are slightly shifted from the actual reciprocal capacities to accommodate all the values clearly. The arrows indicate the correct position of the points, and the solid line is the theoretical prediction from eqn. 3 with $R_s^* = 1.5$. It is apparent that a greater number of poorly resolved baseline peaks were visually detected in the two sets with Gaussian noise than in the noiseless set. This trend is not observed in the peak counts determined from the chromatograms of highest efficiency. The most probable explanation for this observation is the counting of false low-resolution baseline peaks as discussed above. The peak counts from chromatograms of high resolving power are similar in number,

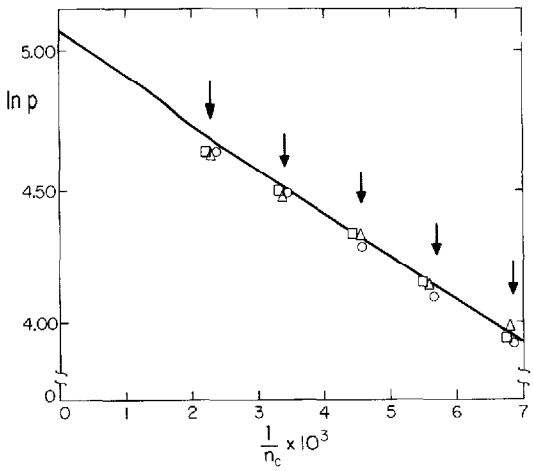


Fig. 5. Logarithm of baseline peak counts from Sets C (○), E (△) and F (□) *versus* reciprocal peak capacity. Solid line, theoretical with $R_s^* = 1.5$.

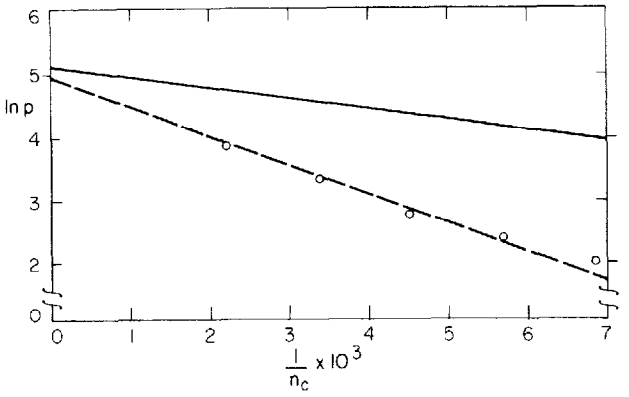


Fig. 6. Logarithm of baseline peak counts from Set G *versus* reciprocal peak capacity. Solid line, theoretical with $R_s^* = 1.5$; dashed line, least-squares fit.

whereas the counts from noisy chromatograms of low efficiency exceed the noiseless peak numbers. The combined effects depress both the slope and the intercept values estimated from the noisy simulations relative to the noiseless one.

An examination of the peak counts from Sets C, E and F by a one-way ANOVA reveals that none of the peak counts differs statistically from the others at a given efficiency level[2]. The optimal values of both the slope and intercept of eqn. 3 are nevertheless dependent on any correlated behavior of the peak counts as a function of efficiency. The mean slopes and intercepts from Sets E and F are therefore understandably significantly different from the results of Sets C and D even though the peak counts themselves are statistically identical at a given efficiency.

We note that the increase in noise power from 30 ADCs in Set E to 50 ADCs in Set F is accompanied by a slight increase in the mean values of both slope and intercept. This increase is not statistically significant. As previously stated, an *a posteriori* ANOVA established the significant differences among Sets C, E and F to be between Sets E and F and Set C.

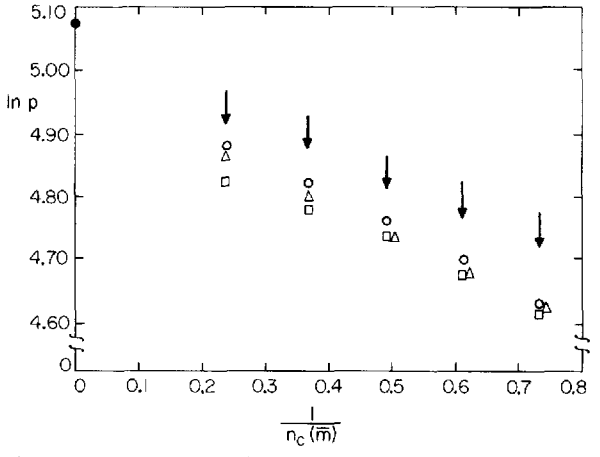


Fig. 7. Logarithm of maxima peak counts from Sets C (○), E (△) and F (□) *versus* the effective saturation, $1/n_e(\bar{m})$. The theoretical intercept is indicated (●).

The magnitude of the Fisher ratio from ANOVA IV is indicative of the large differences in the mean slopes which are calculated from Gaussian and convoluted Gaussian baseline-resolved peak counts. The single variable is convolution. The extreme dissimilarity in slopes is unsurprising because the true representative component-peak width at baseline, x_0, is not equal to $4\sigma R_s^*$ with $R_s^* = 1.5$. The peak capacity calculated from this value of x_0 and eqn. 4 is not correct. The surprising observation is the statistical equivalence of the intercepts between Sets C and G. The rigor of the ANOVA is partially mitigated by the heteroscedasticity of and the required weighting of the data. The difference is nevertheless insignificant. A modestly accurate intercept is calculated for reasons analogous to those presented in the discussion of the "third body" effect. The magnitude of the perturbation introduced by the component-peak tailing decreases continuously with increasing efficiency. The peak counts obtained from relatively highly efficient convoluted Gaussian component-peak chromatograms are in rough agreement with the point model, and a least-squares fit of baseline-convoluted peak counts to eqn. 3 provides a poor slope but modest intercept.

Fig. 6 is a graph of the logarithm of each of the mean convoluted Gaussian peak counts against reciprocal baseline peak capacity with $R_s^* = 1.5$. As stated, it is observable that the influences of component-peak convolution on the results predicted by our model diminish with increasing separation efficiency.

We now devote our remaining discussion on the departures from our model to the influences of real chromatographic attributes on the calculated values of empirical resolution and intercepts from maxima counts. It will be recalled that the resolution is computed as a slope and is the mean resolution factor with which maxima may be differentiated. If the number of maxima observed at a given efficiency changes in response to changes in the chromatographic attributes, we will observe variations in R_s^*.

We first examine the effects of Gaussian noise. Fig. 7 is a plot of the logarithm of each of the average peak maxima counts against the effective saturation which is based on the known component number[2]. Fig. 7 is presented in a manner analogous to the presentation in Fig. 5. It is observable that the number of maxima peak counts decreases in noisy simulations of high efficiency.

It was suggested in ref. 2 that the detection of each of two closely overlapping maxima would be more difficult in the presence than in the absence of noise. This concept is illustrated in Fig. 8, in which a component in each of two well-resolved pairs of maxima is lost with the addition of noise. These ordered distributions were generated by our plotter and traced for clarity. It is clear from Fig. 8 that noise

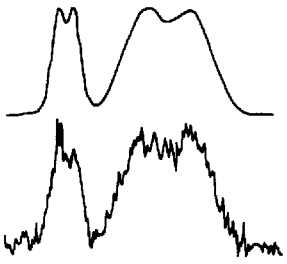


Fig. 8. Loss of maxima with noise. Amplitude: 500 ADCs. The noise is Gaussian with $\sigma = 50$ ADCs.

diminishes the resolubility of maxima in both lowly and highly efficient chromatograms, whereas the loss of noisy maxima is found principally in the highly efficient chromatograms.

A second possible origin of the low maxima count is the loss of graphic quality and resolution from our Versatek plotter. A brief description of the plotting characteristics of this unit will clarify this hypothesis. The amplitude in the plots of the overlapping components is represented by a vertical bar of appropriate size at each of many equally spaced positions along the plot. In Fig. 1, the discreteness of and the disjointedness between the vertical amplitude segments can be faintly observed. The distance between amplitude segments is constant and independent of the length of the plot. In our generation of the chromatograms, we arbitrarily kept the scaling along the abscissa at 10 min per inch. Therefore, the components with small standard deviations in the chromatograms of high resolving power are composed of few amplitude bars per component, whereas the components with large standard deviations are composed of many amplitude bars per component. The plotting resolution in the highly efficient chromatograms is therefore less than the resolution in the poorly efficient chromatograms. This disjointedness between segments in the chromatograms of highest efficiency was most noticeable in the presence of noise. The visual identification of maxima is therefore more difficult because of poor resolution.

Another consequence of the disjointedness of the chromatograms must also be considered. It is possible that the number of maxima in the highly efficient noiseless simulations is underestimated. If the counts are significantly underestimated, then the empirical resolutions for the noiseless simulations will be in error.

The resolution of the plotter and noise are two independent variates of the synthetic chromatograms. To determine the possible significance of each attribute, a two-way ANOVA was performed with additional computer-generated chromatograms. In a two-way ANOVA, each of two variates is changed independently among different groups to determine the significance of each[3]. In our study, six chromatograms with different component distributions were generated over an 18-fold ampli-

TABLE III

TWO-WAY ANOVA ON NOISE AND PLOTTING RESOLUTION

Set	*Maxima (mean ± standard deviation)*	*Source of variation*	*Degrees of freedom*	*SS*	*MS*	*F*
(a) No noise, high resolution	134.00 ± 3.95	Noise	1	145.04	145.04	$F_{1,20} = 10.63$
(b) No noise, low resolution	131.00 ± 2.83	Resolution	1	222.04	222.04	$F_{1,20} = 16.28$
		Interaction	1	22.05	22.05	$F_{1,20} = 1.62$
(c) Noise, high resolution	129.83 ± 3.31	Within subgroups	20	272.83	13.64	
		Total	23	661.96		
(d) Noise, low resolution; $F_{1,20} = 4.35$	123.00 ± 4.47					

tude range (100–1800 ADCs) for each of four groups. Two of the groups were composed of distributions which were identical in plotting resolution and density (α) with the chromatograms of highest efficiency from Sets C, E and F. These groups differed only in the presence or absence of Gaussian noise with σ = 50 ADCs. The graphical resolution in the chromatograms of the other two groups was increased by lengthening the plots. The density within these groups was also identical with the density in the chromatograms from Sets C, E and F. These groups also differed in the presence or absence of Gaussian noise with σ = 50 ADCs. The ANOVA is presented in Table III. The procedure for this computation is given in ref. 3.

Each of the two variates is significant. The inferior resolution of the shorter plots and the presence of noise both contribute to the undercounting of peak maxima composed of highly resolved component-peaks. An examination of the mean maxima counts reveals that the principal deviation is found in the counting of noisy maxima from low-resolution plots. The insignificance of the interaction Fisher ratio indicates that no synergism or interference between the two variates exists[3]. The additivity of the sums of squares (SS) and the degrees of freedom is indicated by the totals in the fourth and fifth columns.

The means of the maxima peak counts from the two sets of noiseless chromatograms are insignificantly different, and the logarithms of the mean values differ by only 0.5%. We therefore have confidence in the integrity and correctness of the R_s^* values calculated from our noiseless simulations. An error exists in the R_s^* computed from Set F. We have chosen not to re-evaluate its magnitude for the practical reason that the minimum signal-to-noise ratio is two. This situation rarely exists in chromatography because of sophisticated pre-concentration techniques. The numerical value is not a highly important quantity.

It might be argued that we have added an unrealistic amount of Gaussian noise to our synthetic chromatograms. They were generated prior to the conception of the maxima counting procedure. We postulated that large amounts of noise would be required to observe any perturbation in the values of $\bar{m}$ obtained from noiseless baseline peak counts. We then applied the maxima counting procedure to the same chromatograms to circumvent the time and cost of generating additional plots. In general, however, we may conclude that the effect of noise on the magnitude of the empirical resolution factor is small. The similarity between the magnitude of the R_s^* factor calculated from the data of Set E and all of the other empirical values, except the ones from Sets F and G, is indicative of this small influence. The minimum signal-to-noise ratio in Set E is three, a large value, but the resolution factor is essentially the same as the value computed from the data of Set C.

A significant difference in the empirical resolutions calculated by the utilization of the maxima counting procedure with Gaussian and convoluted Gaussian component-peak chromatograms was established by ANOVA VIII. The explanation for the deviation is similar to the one previously presented to account for the significantly high slope calculated with baseline convoluted Gaussian peak counts. The number of maxima observable in the highly efficient separations is comparable for both component shapes, whereas many maxima are hidden under the tailing edge of the low-efficiency convoluted component-peaks and are not observed. A fraction of these hidden component-peaks is observable in the low-efficiency Gaussian chromatograms. The mean intercept calculated from application of the maxima counting pro-

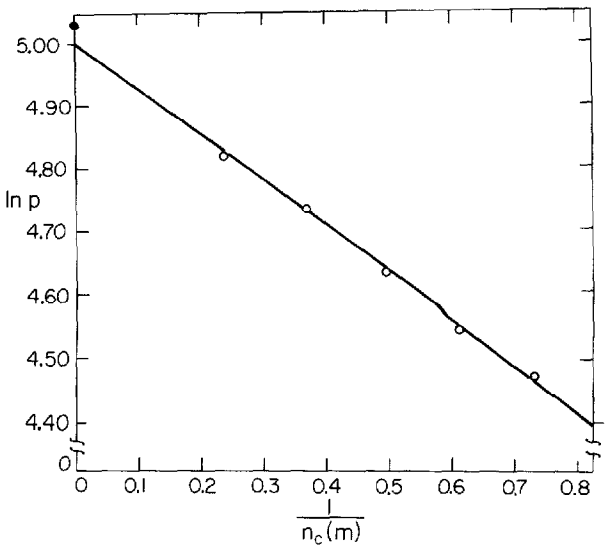


Fig. 9. Logarithm of maxima peak counts from Set G *versus* the effective saturation. Solid line, least-squares fit to data. The theoretical intercept is indicated (●).

cedure is statistically equivalent to the mean intercept calculated from the baseline counting procedure (ANOVA IXe) because the perturbations arising from the component-peak tailing decrease with increasing efficiency.

Fig. 9 is a graph of the logarithm of each of the mean maxima counts from Set G *versus* the saturation factor. The resolution factor is significantly different from the value computed from the data of Set C, and the increased magnitude has a physical meaning. Maxima are more easily obscured at a given efficiency by the tailing edges of the convoluted Gaussian component-peak amplitudes than by the Gaussian component-peak amplitudes. On the average, a greater separation between two maxima is required for visual differentiation. The average resolution factor from Set G is therefore larger than the factor from Set C.

The final two significant ANOVAs may be briefly presented. These both depict the differences in mean intercept values calculated from the applications of the two different counting procedures. The results of ANOVA IXd are indicative of the statistical difference between the intercepts of the noisiest chromatographic series. The reasons for the depression of the empirical resolution value, and therefore the intercept, were previously discussed and need no further comment.

In general, the baseline and maxima counting procedures both predict intercepts which are slightly smaller than the theoretical values. This result occurs in the application of the baseline-resolved peak counting procedure because a return of the amplitude envelope to baseline between two components separated by six average standard deviations is not observed over all amplitude ranges. The counted values lie slightly below theory for this reason and, as a result, the intercept lies below theory. The intercepts fall short of theory with the application of the maxima counting procedure most probably because all peak maxima which are separated by the empirical distance x_0 are not observable and are hidden under the envelope of the overlapping component-peak amplitudes.

This general observation does not apply to Set L. The departure from the model of the mean calculated slope in Set L, which arises from the "third body" effect previously discussed, raises the mean intercept value of Set L above the theoretical value as shown in Fig. 3. The mean value from the maxima counting procedure lies below theory, and a significant difference is observed.

We conclude our study with a commentary on the meaning and importance of our observations. Most importantly, we observe that, in spite of several significant ANOVA procedures, the utilization of our overlap model with both counting procedures is a source of good estimates of the mean component number. We mean by the word "good" that more accurate information is available on this number with the application of our model than is available from a counting of peak maxima in synthetic chromatograms of extraordinarily high efficiency.

We note that the amplitude range of Gaussian component-peaks, the density and noise do not result in large erroneous estimations of $\bar{m}$ from the slope and intercept values which were calculated from the application of the baseline-resolved peak counting procedure. The identification of the mean component number with the slope calculated from convoluted Gaussian baseline-resolved peak counts is erroneous, but a reasonable estimation is accessible via the intercept. The baseline counting procedure is consequently established as a potentially successful means by which one may apply the statistical model.

We do not consider the methodology unequivocably established for the following reason. One important attribute which we have not examined is baseline drift. This variable is important, and its effects on our predictions of $\bar{m}$ need characterization. We have not undertaken this work for this publication. We will confine our remaining discussion to our findings on the peak maxima counting methodology.

The number of observed maxima is dependent on the scientist's ability to distinguish between maxima in a chromatogram. We have established by the two-way ANOVA that this number may be underestimated, especially in the presence of noise, from a chromatogram of poor graphical resolution. Therefore, a chromatogram from which maxima are counted should be of sufficient length for one to observe all the maxima clearly and distinctly. This point is perhaps intuitive and superfluous, but we now recognize its importance from our own work.

In general, we have established a series of empirical resolution factors with which the mean component number may be estimated from both slope and intercept of a ln p *versus* $1/n_c$ plot. These factors initially were postulated to be functions of the realistic attribute of chromatograms under investigation in this paper. The numerical values of the factors are surprisingly independent of the limited amplitude and density range examined and small levels of noise but do exhibit a dependence on component-peak asymmetry. The insensitivity to density may be roughly understood. As the number of components per unit separation space increases, the potential number of observable maxima increases. However, the number of maxima which overlap and are lost in the amplitude envelope also increases. A counterbalance is apparently established over the examined range. The insensitivity of the resolution factors to the amplitude range is not understood. It is reasonable to hypothesize that the number of observable maxima will decrease with increasing component amplitude range because small maxima will be obscured by amplitude envelopes of increasing size. Our computations refute this hypothesis.

We must emphasize that the observed insensitivity of the resolution factors to the amplitude range may arise from the uniform amplitude distributions with which we have generated our synthetic chromatograms. The resolution factors computed from the maxima peak counts of overlapping components which are non-uniformly distributed in amplitude may exhibit a dependence on the non-uniform amplitude range.

We infer from the magnitudes of the resolution factors calculated from the data of Sets C and G that, in general, empirical resolutions which are calculated from counts of observable maxima in convoluted component-peak chromatograms vary as a function of the σ/τ ratio. A family of R_s^* values is calculable as a function of σ/τ. We have not undertaken the extensive work required to compute this family. We shall argue, however, that the numerical value of the resolution factor is immaterial if one is willing to estimate the mean component number from the intercept alone.

We may write the reciprocal capacity in eqn. 4 as

$$\frac{1}{n_c} = \frac{4\sigma R_s^*}{X} = \frac{\beta\sigma}{X} \tag{7}$$

where $\beta = 4R_s^*$. In this equation, R_s^* is an empirical resolution value which may be unknown. The least-squares equation for the intercept a of a linear function is related to the abscissa (x) and ordinate (y) values of the fitted data and to the weighting factor w for each ordinate value and is[4]

$$a = \frac{\sum w_i x_i^2 \sum w_i y_i - \sum w_i x_i \sum w_i x_i y_i}{\sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2} \tag{8}$$

If we substitute the ith reciprocal capacity for variable x_i in eqn. 8, we observe that the quantity β cancels in the calculation. The numerical value of the resolution factor is not required to calculate the intercept. Only the relative, and not the absolute, reciprocal capacities are required for this computation. The choice of the reciprocal capacity axis is arbitrary, but it is essential to scale the spacing between the points for the $\ln p$ *versus* $1/n_c$ plot relative to the component standard deviation per separation space ratios σ/X. We note that the statistical equivalence between the mean intercepts of Sets C and G (ANOVA IV) is now understandable in light of the dependence of the intercept value upon only the relative reciprocal capacity values.

The average component standard deviations, which are required to calculate both the absolute and relative peak capacities, are in principle calculable from the chromatograms. Practically, one does not usually know if the observed maxima are single component-peaks or singlets. An accurate estimation of these quantities can be made by averaging the standard deviations of known pure components which are representative of the analyte in question[1]. The chromatographic conditions under which these measurements are made should, of course, be identical with the conditions under which the analyte is partially resolved.

If the logarithm of each of the maxima counts is plotted against a relative reciprocal capacity and significant departures from linearity are observed, then the resolution factor is changing as a function of chromatographic efficiency and β is not constant. We did not observe significant departures from linearity in our simulations.

In general, a poor estimate of $\bar{m}$ should be expected from nonlinear data.

It was stated earlier that the intercepts computed from the maxima counting procedure were less than the theoretical values. The average percentage error with respect to theory is calculated from the nine appropriate values of intercept compiled in Table I and is 0.0627 ± 0.0312. This value may be used, if desired, to calculate a first-order correction to the experimentally determined intercept.

In the application of our model to real chromatograms, in which the integrity of our hypothesis of random component distribution will be questioned, the value of the slope may prove more useful than the intercept. In the light of the uncertainty of this hypothesis, the reader may question the usefulness of an extensive characterization of synthetic chromatographic data on which we have imposed the condition of random component distribution. Our studies have revealed many departures from theory which arise from realistic properties of chromatograms which we have not modeled. Only with a knowledge of these attributes can we separate their effects from departures from the statistical hypothesis.

The general magnitudes of the empirical resolutions calculated from the Gaussian component-peak chromatograms may disturb some physical scientists. The overall magnitude of the R_s^* factors calculated from the Gaussian component-peak chromatograms is approximately $R_s^* = 0.5$. The minimum x_0 with which two Gaussian components of equal amplitude are resolvable is 2σ, with $R_s^* = 0.5$. A larger value of x_0, and a greater R_s^* value, is required to differentiate between the maxima of two Gaussian components with unequal amplitudes. One might argue that our empirical values should therefore be somewhat larger.

The least-squares intercepts computed from the maxima counting procedure lie below the theoretical values, as previously discussed. We allowed the intercept to assume its optimal value in our computations of the resolution factors. A rigorous fit of the data to theory, with the intercept fixed as $\ln \bar{m}$, would result in the calculation of higher values of empirical resolution factors ($R_s^* \approx 0.7$ for Set C), but the quality of the fit to data would be poor. If this resolution factor were then used to compute a peak capacity for the fitting of maxima, the magnitude of the intercept calculated would be independent of the resolution factor. However, the slope calculated from the least-squares procedure would be smaller than a slope computed from the same peak counts but a peak capacity determined by $R_s^* \approx 0.5$. The mean component number would therefore be underestimated from the slope. We arbitrarily chose to compute the resolution factors so that estimates of $\bar{m}$ could be made from both slope and intercept.

In the light of the independence of the intercept value and the empirical resolution, it is reasonable to ask if our extensive calculation of empirical resolution values was an unnecessary undertaking. We believe the proper answer to this question is no. Our work has definitely established that approximately the same number of maxima peaks, which are composed of pure Gaussian components, will be resolved at a given level of system efficiency irrespective of the density and amplitude range examined and small amounts of noise. This statement is correct because the empirical resolution, which determines the system's peak capacity, is largely independent of these variables. The number of peaks is then exclusively determined by this well defined peak capacity and the mean component number. These are the only attributes on which eqn. 1 is dependent. The simulated chromatograms are well-modeled by

Poisson statistics and point processes in which the peak maxima are represented by points.

We have recently developed successful procedures for the estimation of the mean component number from the slope and the intercept of a single "point chromatogram". It is therefore not unreasonable to propose a study in the feasibility of the determination of the mean component number from a single chromatogram of Gaussian component peaks via a counting of peak maxima. The success or failure of this approach will be presented in a future publication.

ACKNOWLEDGEMENTS

This work was supported by Public Health Service Grant GM10851-26 from the National Institutes of Health. One of us (J.M.D.) acknowledges the support of an ACS Analytical Division Fellowship sponsored by T. L. Chester of the Procter and Gamble Company.

REFERENCES

1 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 55 (1983) 418.
2 J. C. Giddings, J. M. Davis and M. R. Schure, in S. Ahuja (Editor), *ACS Symposium on Ultra-High Resolution Chromatography*, American Chemical Society, Washington, DC, 1983, in press.
3 R. R. Sokal and F. J. Rohlf, *Biometry*, W. Freeman, San Francisco, 1st ed., 1969.
4 P. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 1969.
5 E. Grushka, *Anal. Chem.*, 44 (1972) 1733.
6 J. C. Giddings, *Dynamics of Chromatography, Part I, Principles and Theory*, Marcel Dekker, New York, 1965.