ELSEVIER

Journal of Chromatography A, 677 (1994) 307–318

JOURNAL OF CHROMATOGRAPHY A

Influence of integrator parameters on estimates calculated with the statistical model of overlap

DeWayne Bowlin, Christian Hott, Joe M. Davis*

Department of Chemistry and Biochemistry, Southern Illinois University at Carbondale, Carbondale, IL 62901, USA

First received 22 September 1993; revised manuscript received 26 April 1994

Abstract

The effect of the slope and area thresholds of a digital integrator on the calculation of component numbers with the statistical model of overlap (SMO) is critiqued. As these thresholds increase, the numbers of maxima recognized by the integrator and components calculated with the SMO both decrease. With modest changes in threshold, the maxima and component numbers decrease proportionally, such that their ratio remains constant. For larger changes in threshold, the number of components usually decreases more rapidly than the number of maxima. Computer simulations are developed that confirm the general trend of these observations. The implications of these variations are discussed.

1. Introduction

The statistical model of overlap (SMO) has been used by this [1–3] and other groups [4–8] to estimate the quality of separation in chromatograms of complex mixtures. These estimations are based on either an interpretation of the numbers of maxima in a series of chromatograms of different peak capacity [2,4,6,8], or an interpretation of the distribution of intervals between adjacent maxima in a single chromatogram [1–3,5–7]. By either interpretation, several statistical parameters can be calculated, including the expected number of mixture components, the expected numbers of singlet and multiplet peaks, and the probability of resolving components as single-component peaks. The SMO and its applications have been reviewed recently [9].

Because these parameters are calculated from the numbers of, or intervals between, chromatographic maxima, they depend intimately on the mechanics of peak detection and processing. Commonly, both the numbers of peaks and their retention times (from which intervals between peaks are calculated easily) are determined by digital integrators. Regardless of their commercial origin, these integrators process chromatographic signals in a similar manner. Furthermore, many integrator settings are adjustable by the scientist. Thus, the scientist has considerable influence over the number of peaks determined.

The subject of this paper is the dependence of the most important statistical parameter, the number of mixture components, on the processing of peaks by a commercial integrator. The stimulus for this work was a previously reported observation based on gas chromatograms of a coal-tar extract [3]. When retention times were

* Corresponding author.

0021-9673/94/$07.00
SSDI 0021-9673(94)00412-3

determined by an integrator, the numbers of maxima and estimates of component number increased with increasing flow-rate. In contrast, when retention times were determined by visual inspection and manual digitization, the numbers of maxima decreased slightly with increasing flow-rate and estimates of component number were largely independent of flow-rate. It was suggested (but not proven) that the slope threshold of the integrator was responsible for this difference. The most important observation was that the means of peak processing —in one case by a digital integrator and in the other by eye— had a pronounced effect on the value of statistical parameters.

To our knowledge, a systematic experimental study of the effects of peak processing on SMO predictions has not been reported. The closest related work is by Dondi et al. [6], who interpreted computer-simulated chromatograms containing 50 major components and from 0 to 500 minor components. The latter were added to determine if their presence affected the statistical interpretation of the number of major components (the effect was small). In that study, however, all maxima below a certain amplitude were ignored. This differs from the study here, in which maxima are detected in some chromatograms but not in others.

Integrators have qualitative and especially quantitative limitations, which have been discussed by several authors [10–13]. Because integrators are fairly ubiquitous in the separation laboratory, we have examined the effect of two major parameters on the predictions of the SMO. In brief, 30 gas chromatograms of lime oil were developed under conditions appropriate to interpretation by the SMO. These chromatograms were processed differently, however, by varying systematically the electrometer amplification of the gas chromatograph and the slope and area thresholds of the integrator. The principal effect of this processing was to cause the retention times and areas of small maxima to be determined in some chromatograms but not in others. Thus, the sequence of determined retention times varied from chromatogram to chromatogram. The numbers of components detected then were estimated from these retention-time sequences by a well established procedure. Because these sequences differed for different processing conditions, the number of detected components also differed. In total, 480 different sequences of retention times were interpreted.

To lend credence to the interpretation of these sequences, we also determined retention times in 480 computer-simulated chromatograms using an in-house algorithm superficially similar to that of the integrator. The results of those interpretations closely, although not identically, paralleled those determined by experiment.

It should be noted that the word "detected" is used loosely here to indicate signals that are recognized and interpreted by the integrator. Signals that were truly detectable (i.e., statistically above the noise level) consequently will be described as undetected, if they were not so recognized and interpreted.

2. Theory

The expected number $\overline{m}$ of single-component peaks (SCPs) detected by the integrator was determined by the so-called single-chromatogram method, which is based on the SMO and is described elsewhere [5,14]. In brief, the differences between the retention times of adjacent maxima in an interval of temporal span X were compared to a series of arbitrary times. The numbers p' of intervals between adjacent maxima greater than x_0' were determined and graphed against x_0'/X as

$$\ln p' = \ln \overline{m} - \overline{m} x_0'/X \quad (1)$$

For SCPs distributed in accordance with Poisson statistics (i.e., distributed randomly throughout the chromatogram), values of $\ln p'$ are constant for small x_0'/X values but then decrease linearly as x_0'/X increases [14]. In this case, a line fit to the linear region has a slope equal to $-\overline{m}$ and an intercept equal to $\ln \overline{m}$. The two $\overline{m}$ estimates so determined are designated m_{sl} and m_{in}, respectively, and are pooled to obtain a weighted estimate of $\overline{m}$, which is designated m_{ave}. The

algorithm for pooling m_{sl} and m_{in} is outlined elsewhere [1]. The resultant estimate, m_{ave}, is an estimate to the actual number m of detectable SCPs.

To obtain accurate estimates, m_{sl} and m_{in} should agree closely, the graph of ln p' vs. x_0'/X should be linear (except for small x_0'/X values), the density of maxima in region X should be constant, and α should be less than 0.5 or so, when determined as

$$\alpha = -\ln \frac{p_m}{m_{ave}} \quad (2)$$

where p_m is the number of maxima detected by the integrator [2,14]. Unless noted otherwise, these criteria were satisfied fairly well in this study.

3. Procedures

3.1. Protocol

A cold-pressed lime oil (A.M Todd Co., Kalamazoo, MI, USA) was diluted 10-fold with HPLC-grade methylene chloride (Fisher Scientific, St. Louis, MO, USA). The mixture was stored in the refrigerator when not in use.

Chromatograms of this mixture were developed on a 30 m × 0.320 mm I.D. DB-1701 fused-silica capillary (J & W Scientific, Folsom, CA, USA) having a phase thickness of 0.25 μm. The capillary was incorporated into a Shimadzu GC-9AM modular gas chromatograph (Columbia, MD, USA) equipped with flame ionization detection (FID) and a SPL-G9 split–splitless injector. The electrometer output was sent to a CR-6A integrator (Shimadzu) for digital processing, as detailed below.

Two of four possible amplifications of FID current by the electrometer were used in this study. The full-scale electrometer currents at these amplifications were 10 and 100 pA; the chromatograms so developed will be designated 10-pA and 100-pA chromatograms. The other two amplifications generated either too few or too many maxima for interpretation. In the former case, too few (e.g., 20 or less) were detected to obtain reliable statistics; in the latter case, the saturation α substantially exceeded 0.5.

For both amplifications, 15 replicate chromatograms were developed as outlined below over a two-day period. The short interval of time favored the development of highly reproducible chromatograms, which was essential to our purpose. Thus, a total of 30 chromatograms (15 chromatograms per amplification × 2 amplifications) was generated.

Each of these chromatograms then was processed repetitively by integrator software to generate files of retention times for statistical interpretation. Two of four adjustable integrator parameters were varied. The width parameter was fixed for all chromatograms at 5 s to reduce noise, and the baseline-drift parameter was set to its default value. In contrast, the area threshold A_t was varied over the four values, 5, 50, 125 and 625 μV s, and the slope threshold S_t was varied over the four values, 200, 400, 600 and 800 μV/min. For any chromatogram, the retention time of a maximum was recorded, only if the experimental slope and area exceeded S_t and A_t, respectively. For each amplification, each of the 15 replicate chromatograms was processed at all 16 possible combinations of S_t and A_t, such that $16 \times 15 = 240$ retention-time files were obtained. These combinations of S_t and A_t will be represented here by the coordinates, (S_t, A_t). Thus, a total of 480 retention-time files were generated (240 files per amplification × 2 amplifications).

Precautions were taken to minimize the detection of spurious maxima. To ensure that solvent impurities were not detected as maxima, the solvent periodically was chromatographed at the most sensitive coordinate, (200, 5). No maxima were detected, other than the solvent peak. In addition, for any new coordinate, the capillary first was programmed through its temperature sequence to verify that no maxima were detectable. This precaution also was carried out at the beginning of each day.

The data acquired by the integrator were converted to ASCII files of the digitized chromatogram and retention times by Chromatopac data archiving utility (Shimadzu). These files

were downloaded to a 386 PC compatible computer (Northgate Computer Systems, Minneapolis, MN, USA) and then transferred to a Macintosh SE (Apple Computer, Cupertino, CA, USA) for editing, graphing and analysis. The communications software between the PC and Macintosh computers were PROCOMM PLUS (DataStorm Technologies, Columbia, MO, USA) and VersaTerm (Synergy Software, Reading, PA, USA).

3.2. Chromatography

The chromatographic conditions were: carrier, helium at average flow-rate of 4.25 ml/min (range, 4.23–4.32 ml/min); split flow, 42 ml/min; purge flow, 2.1 ml/min; make-up gas, 40 ml/min; air flow-rate to FID, 320 ml/min, hydrogen flow-rate to FID, 35 ml/min; injector and detector temperatures, 270°C; and sample size, 0.2 μl. The split–splitless ratio was controlled by and measured from the chromatograph. The carrier flow-rate was reproduced closely to ensure that the maximum slope of a given peak varied little among replicate chromatograms; such variation might have affected the number of detected peaks.

An empirically determined series of heating rates was used to develop chromatograms containing randomly distributed SCPs at the most sensitive (S_t, A_t) coordinate, (200, 5), and the 10-pA amplification. The temperature program developed at flow-rate $F = 4.25$ ml/min was: isothermal period, 65°C for 2.82 min; first ramp, 10.6°C/min to 95°C; second ramp, 21.4°C/min to 120°C ; third ramp, 3.6°C/min to 140°C; isothermal period, 140°C for 4.71 min; fourth ramp, 40°C/min to 220°C; isothermal period, 220°C for 9.41 min. The purpose of the final ramp was to flush high-boiling components from the capillary. The fraction of the chromatogram interpreted statistically is shown in Fig. 1.

We note in passing that theory was developed recently to circumvent the need of developing elution patterns consistent with constant-density Poisson statistics [15]. The theory shows much promise in statistically interpreting chromatograms more simply than performed here. However, the present study was initiated prior to this development.

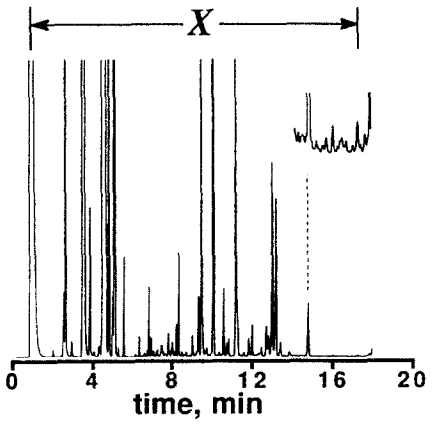


Fig. 1. Region X of lime-oil chromatogram interpreted by graphs of ln p' vs. x_0'/X.

3.3. Computer simulations

To add credibility to our interpretation, computer-simulated chromatograms were generated and interpreted by an algorithm superficially similar to that of the integrator. The retention-time sequences so generated then were interpreted by Eq. 1 as were the experimental sequences. The standard deviation of SCPs in these simulations was set arbitrarily to 2 s. The amplitudes of chromatograms were computed at intervals of 0.5 s to mimic the integrator's storage of digitally smoothed data at intervals equal to 1/10th of the width parameter (equal to 5 s; see above). Other details are deferred to the Results and discussion section.

It is important to recognize that we did not attempt to reproduce by simulation the experimental chromatograms exactly. Such a reproduction is not possible. In the experimental chromatograms, for example, SCP widths change systematically with changes in heating rate, and detailed theory is needed to model this variation. Even if this modeling were implemented, the SCP amplitude distribution is unknown, and consequently the relative amplitudes one should assign SCPs are unknown. Furthermore, baseline drift and noise are present in experimental

chromatograms but not the simulated ones. Instead, what we sought to verify by simulation was that the experimental variation of p_m and m_{ave} with S_t and A_t was an expected and reasonable result and not an artifact of our interpretation. In other words, we sought trends, not an exact agreement.

The following "rules" were adopted to mimic an integrator:

(a) A maximum was judged as detectable if and only if the slope on its front edge exceeded S_t, if the absolute value of the slope on its back edge exceeded S_t, and if the slope at one or more subsequent points on its back edge exceeded $-S_t$. Maxima that did not satisfy all these conditions were ignored.

(b) For each maximum, the integration limits were chosen from among the 10 amplitudes preceding (on the front edge) or following (on the back edge) the point at which the slope threshold was crossed. Among those values, the limit was identified with a valley between two maxima, a return to baseline, or the 10th value. The use of 10 values was not arbitrary but mimicked the integrator. Areas were evaluated with Simpson's rule.

(c) If the area of the maximum exceeded A_t, then the retention time was recorded. The retention time was determined by parabolic interpolation of the local maximum amplitude and the two adjacent amplitudes. If more than one maximum was found (a rare occurrence), then the retention time of the largest maximum was recorded (this action also mimicked the integrator).

The recorded retention times comprised the retention-time file for that simulated chromatogram.

3.4. Analysis of retention-time files

The retention-time files of both experimental and simulated chromatograms were interpreted by a series of FORTRAN programs written in Language Systems FORTRAN (Language Systems Corp., Herndon, VA, USA) and executed on an SE Macintosh (Apple) or Outbound Macintosh. The first program generated the data pairs, $(x_0'/X, \ln p')$, for graphs of $\ln p'$ vs. x_0'/X; the second computed values of m_{sl}, m_{in}, and m_{ave}; and the third generated simulated chromatograms for purposes of comparison to experimental ones. The third program was executed only periodically to verify a resemblance between experimental and simulated chromatograms.

For both experimental and simulated chromatograms, graphs of $\ln p'$ vs. x_0'/X and plots of simulated chromatograms were generated with KaleidaGraph (Synergy Software). Each graph of $\ln p'$ vs. x_0'/X was examined to exclude from the fitting those coordinates at small x_0'/X values that did not fall on a straight line. Three-dimensional graphs were generated with DeltaGraph Professional (DeltaPoint, Monterey, CA, USA).

4. Results and discussion

4.1. Dependence of ln p' graphs on A_t and S_t

Fig. 2 shows graphs of $\ln p'$ vs. x_0'/X developed from the 10-pA chromatograms at various coordinates (S_t, A_t). The numbers p_m of maxima recognized by the integrator varied from 58 to 85; the numbers m_{ave} of components detected by the integrator varied from 102 to 171. The circles represent experimental data; the filled circles represent data fit to Eq. 1 by least-squares theory; the solid lines represent theoretical fits. Although chromatographic conditions were chosen to develop linear graphs only at the most sensitive coordinate, (200, 5), the graphs are fairly linear for all coordinates. From the graphs alone, one cannot tell that approximately one-third of the maxima detectable at the smallest (S_t, A_t) coordinate are not detected at the largest (S_t, A_t) coordinate.

4.2. Experimental variation of m_{ave} with S_t and A_t

In Fig. 3a and c three-dimensional graphs are shown of the average numbers $\bar{p}_m$ of maxima determined from 10- and 100-pA chromatograms, respectively, vs. S_t and A_t. Here, $\bar{p}_m$ represents the mean number of maxima determined from 15 replicate chromatograms by inte-

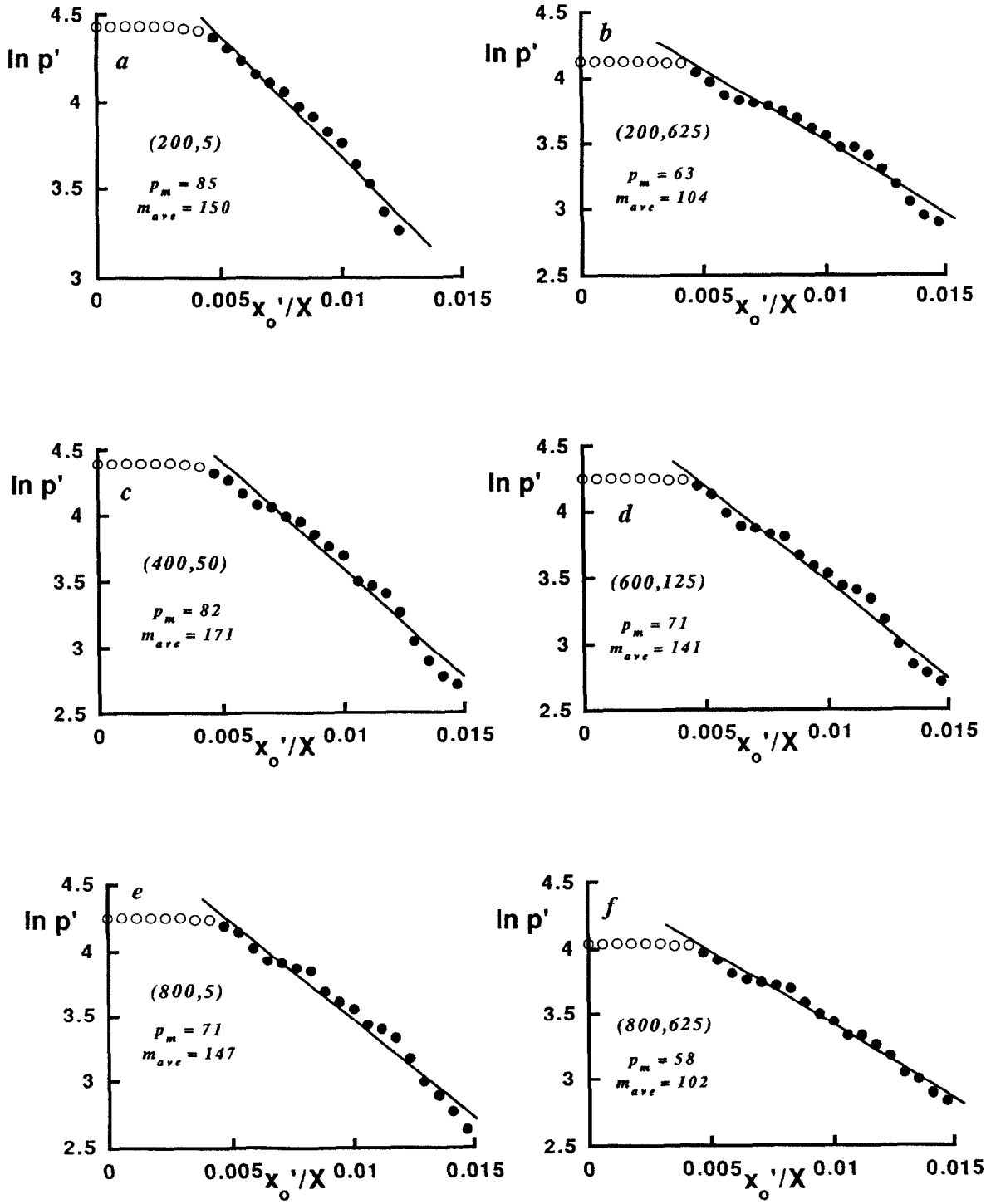


Fig. 2. Graphs of ln p' vs. x_0'/X developed from 10-pA chromatograms interpreted at different (S_t, A_t) coordinates. All symbols represent experimental data; filled symbols represent data fit to Eq. 1; solid lines represent theory.

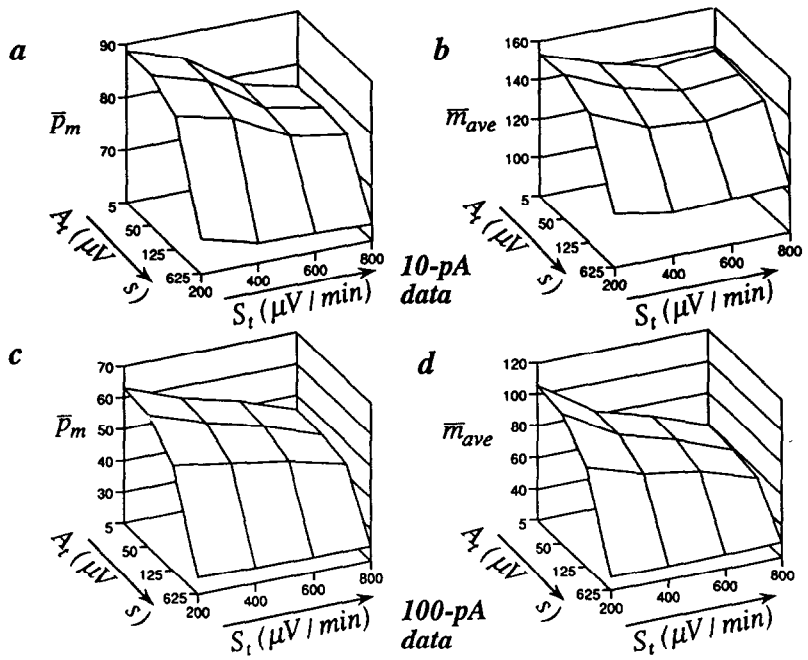


Fig. 3. Graphs of the average number $\bar{p}_m$ of maxima vs. S_t and A_t in the (a) 10-pA and (c) 100-pA chromatograms. Graphs of the average number $\bar{m}_{ave}$ of components calculated from Eq. 1 vs. S_t and A_t for the (b) 10-pA and (d) 100-pA chromatograms. Each datum represents the average of p_m or m_{ave} determined from 15 replicate chromatograms.

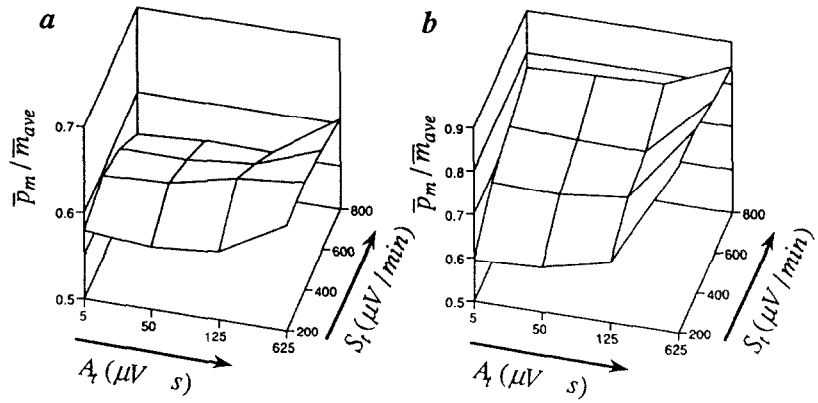
grator software at any coordinate, (S_t, A_t). Unsurprisingly, this number decreases with increasing S_t and A_t and with increasing full-scale current output. In Fig. 3b and d three-dimensional graphs are shown of the average numbers $\bar{m}_{ave}$, computed from these chromatograms with Eq. 1, vs. S_t and A_t. Here, $\bar{m}_{ave}$ represents the mean number of components estimated from the 15 retention-time files processed at the coordinate, (S_t, A_t). In general, these estimates track the variation of $\bar{p}_m$ with S_t and A_t, but this tracking is only roughly proportional.

For example, in Fig. 4a and b graphs are shown of the ratio $\bar{p}_m/\bar{m}_{ave}$, computed from the data comprising Fig. 3a–d. In both figures, this ratio is about 0.58–0.60 for small values of S_t and A_t but varies with increasing S_t and A_t (it usually increases). In particular, the $\bar{p}_m/\bar{m}_{ave}$ ratio at $A_t = 625$ is quite large for the 100-pA chromatograms. In general the rate of increase is larger for the 100-pA chromatograms than for the 10-pA chromatograms.

4.3. Generation and analysis of computer-simulated chromatograms

To lend credibility to these observations, computer simulations were generated and interpreted by an algorithm mimicking the digital integrator. These simulations were generated to gauge if the graphs in Fig. 3 reflected realistic and expected experimental behaviors, as opposed to artifacts.

A brief outline of these simulations' generation is given here. The simulatory equivalents of S_t and A_t corresponding to the experimental coordinate, (800, 625), were determined for both the 10- and 100-pA chromatograms, such that the numbers of maxima in the simulatory and experimental chromatograms were identical at this coordinate. All remaining simulatory S_t and A_t values then were scaled relative to these values and their experimental counterparts. To implement the scaling, however, the true number of components, corresponding to what one



10-pA data *100-pA data*

Fig. 4. Graphs of the ratio $\bar{p}_m/\bar{m}_{ave}$ vs. S_t and A_t for the (a) 10-pA and (b) 100-pA chromatograms.

would detect at $S_t = 0$ and $A_t = 0$, was required. This number was estimated very simply from the data by appropriate expansions.

The normalization of simulations for the 10-pA chromatograms is described here in some detail. A first-order estimate of m was made by extrapolating via two-dimensional Taylor-series expansions [16] the $\bar{m}_{ave}$ values in Fig. 3b to the coordinate, (0, 0), the coordinate at which all components would be detected. A similar extrapolation was made for the $\bar{p}_m$ values in Fig. 3a. The extrapolated $\bar{m}_{ave}$ and $\bar{p}_m$ values were 163.2 and 92.2, respectively. This value of $\bar{m}_{ave}$ then was corrected for the deterministic error resulting from saturation. This error depends on the SCP amplitude distribution; evidence exists that this distribution is approximately exponential [17]. Relative to the coordinate, (0, 0), α was calculated from Eq. 2 as $-\ln(\bar{p}_m/\bar{m}_{ave}) = -\ln(92.2/163.2) = 0.571$. As shown elsewhere, the approximate percentage error PE expected in m under these conditions is [3]

$$PE = 100\left(\frac{m_{ave}}{m} - 1\right)$$
$$= 108.8\alpha^2 - 173.1\alpha + 49.7 \quad (3)$$

From Eq. 3, a corrected estimate of m, equal to 189, was calculated. This value was interpreted as the best estimate of m as S_t and A_t approach zero.

To determine the simulatory thresholds corresponding to the experimental coordinate, (800, 625), the average number of maxima in 50 simulations containing $m = 189$ SCPs with exponential amplitudes and a mean amplitude of 10 units was determined for various arbitrary but proportional S_t and A_t values. These thresholds were increased systematically from zero, until the average number of maxima (here, equal to 63.0) determined from the simulations equalled that determined at the experimental coordinate, (800, 625). These thresholds were increased proportionally, such that the ratio A_t/S_t equalled 46.9 s^2, the ratio determined by the coordinate, (800, 625) [i.e., 625 μV s/(800 μV/min · 1 min/60 s) = 46.9 s^2]. It was found that 63 maxima were detected in the simulations, when $A_t = 43.75$ unit · s and $S_t = 0.933$ unit/s (or 56.0 unit/min). The remaining simulatory A_t and S_t values then were scaled relative to these values, e.g., the simulatory area threshold corresponding to $A_t = 50$ μV s was calculated as (50/625)43.75 = 3.5 unit · s.

In Fig. 5a and b three-dimensional graphs are shown of the average numbers $\bar{p}_m$ of maxima and $\bar{m}_{ave}$ of detected components, respectively, determined from 15 computer simulations vs. the S_t and A_t simulatory thresholds. The number of simulations interpreted at any (S_t, A_t) coordinate was identical to the number of interpreted experimental chromatograms. Unlike for the

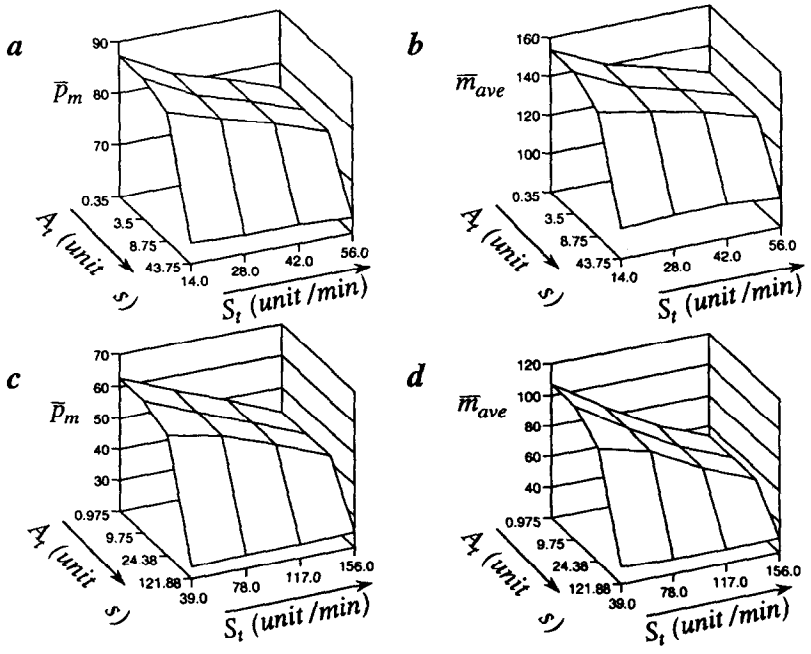


Fig. 5. Graphs of the average number $\overline{p}_m$ of maxima vs. S_t and A_t in scaled computer simulations of the (a) 10-pA and (c) 100-pA chromatograms. Graphs of the average number $\overline{m}_{ave}$ of components calculated from Eq. 1 vs. S_t and A_t for scaled computer simulations of the (b) 10-pA and (d) 100-pA chromatograms. Each datum represents the average of p_m or m_{ave} determined from 15 computer simulations. Values of S_t and A_t were determined by simulation.

experimental chromatograms, however, the simulations were not replicates but rather were generated from different sequences of retention times and amplitudes (replicate simulations simply would have generated identical p_m and m_{ave} values). The (S_t, A_t) coordinates reported in the figure differ from their experimental counterparts, because of scaling. However, the $\overline{p}_m$ and $\overline{m}_{ave}$ axes in Figs. 3a and 5a, and 3b and 5b, are scaled identically to facilitate easy comparison. This comparison shows that experimental and simulatory graphs of $\overline{p}_m$ and $\overline{m}_{ave}$ vs. S_t and A_t are qualitatively identical and quantitatively similar. Some differences do exist; in particular, m_{ave} values for the experimental 10-pA data increase slightly for large S_t and small A_t values, in contrast to the results from simulation. Mostly, however, a good agreement exists.

Indeed, in light of all the assumptions we were forced to make in generating the simulations, and also of the fact that $\overline{p}_m$ was "fixed" at only one of the 16 (S_t, A_t) coordinates and $\overline{m}_{ave}$ was "fixed" at none of them, the agreement is actually quite profound. The agreement strongly suggests that the trends illustrated in Fig. 3a and b are real.

The 100-pA chromatograms were mimicked similarly. The first-order estimates of $\overline{m}_{ave}$ and $\overline{p}_m$ at the coordinate, (0, 0) were again determined by Taylor-series expansions and were 130.2 and 70.0, respectively, with $\alpha = 0.619$. This estimate was corrected for saturation to the value, $m = 154$. The simulation thresholds S_t and A_t corresponding to the experimental coordinate, (800, 625), were determined as before and were 156.0 unit/min and 121.88 unit · s. These thresholds are larger than for the 10-pA data, because more maxima (and thus more maxima having small areas) existed in that case. Fig. 5c and d are three-dimensional graphs of $\overline{p}_m$ and $\overline{m}_{ave}$ determined from these simulations. As before, a strong correlation between the results of simulation and experiment (cf. Fig. 3c and d) is observed.

We note that it would have been desirable to interpret at least a fraction of these simulations by the integrator software itself to test the integrity of our in-house algorithm. Unfortunately the integrator software generates a proprietary binary code, whose encryptation we could not break.

4.4. Interpretation of results

Two observations can be made with respect to these experiments, analyses, and simulations. Prerequisite to these observations are the postulates that detectable SCPs are Poisson distributed, α is sufficiently small that reliable m_{ave} values can be calculated from graphs of ln p' vs. x_0'/X, and detectable SCPs may not be detected, if integrator thresholds are sufficiently large. Further, we presume that the value of neither p_m nor $\overline{m}$ is as important as the ratio, $p_m/\overline{m}$, which is a measure of the extent of separation [18].

The first observation is that the probability $p_m/\overline{m}$ that any detectable maximum is an SCP is identical for all maxima and that this probability is not altered by the exclusion of maxima from detection by high thresholds. Thus, the actual $p_m/\overline{m}$ ratio for a chromatogram does not change, even as maxima are excluded from detection; both maxima and components are excluded proportionally. The second observation is that approximation of $p_m/\overline{m}$ by p_m/m_{ave} can be erroneous, if the maxima distribution from which m_{ave} is calculated via ln p' vs. x_0'/X graphs is distorted by the selective exclusion of maxima. Thus, even though $p_m/\overline{m}$ is constant, the measurable ratio $p_m/\overline{m}_{ave}$ can vary, and misleading conclusions can be drawn about the extent of separation.

These behaviors are evident in our experimental results. At sufficiently small (S_t, A_t) coordinates, $\bar{p}_m/\overline{m}_{ave}$ is about 0.6 for both 10- and 100-pA chromatograms, even though $\bar{p}_m$ differs by about 26 between the two chromatogram sets. For larger coordinates, however, $\bar{p}_m/\overline{m}_{ave}$ varies systematically (it usually increases) for both the 10- and 100-pA chromatograms. The reasons for this behavior are not understood at this time. In general, the behavior occurs because the detectable maxima density, which is Poisson at small (S_t, A_t) coordinates, is distorted by the exclusion of maxima at large (S_t, A_t) coordinates.

The above assertion that $p_m/\overline{m}$ is identical for all maxima applies, if one considers only the Poisson distribution of components along the time axis. As shown by others [17,19], the probability that a peak is chemically pure actually depends on amplitude; in particular, small peaks are more likely to be pure than peaks of intermediate size. This finding does not explain our data, however, since the integrator's failure to detect small pure (or nearly pure) peaks should cause p_m to decrease more rapidly than $\overline{m}$, when $p_m \ll \overline{m}$. In fact, the opposite behavior is observed, i.e., $\overline{m}$ decreases more rapidly than p_m.

Thus it appears that experimental conditions consistent with development of Poisson-distributed retention times are restrictive not only in terms of chromatographic specifics (e.g., temperature program and flow-rate) but also detective specifics (e.g., S_t and A_t thresholds). In other words, the alteration of integrator thresholds can distort significantly a Poisson distribution that was established at other thresholds. This study shows that only minor threshold adjustments can be made without distortion.

From this study, one can conclude that integrator thresholds cannot be adjusted extensively, once separation conditions are found to develop a Poisson distribution. The converse also is likely; i.e., separation conditions cannot be substantially altered, once integrator thresholds consistent with a Poisson distribution are adopted. This imposes some difficulties on systematic investigations by the SMO and other statistical theories.

An illustration of the converse scenario is reported here. To test a new theory of overlap that relaxes the constraint of randomness from a global to a local domain [15], gas chromatograms were developed at simple heating rates r between 2 and 10°C/min [20]. Instead of systematically decreasing because of non-equilibrium effects, p_m initially increased with r and decreased only when r exceeded 7°C/min. The

anomalously detected maxima all had small areas (e.g., <500 μV s) and most probably were detected, because they eluted rapidly enough at high heating rates to cross an S_t threshold not crossed at lower heating rates. Since $\overline{m}$ increases as p_m increases, the anomalous detection of these maxima complicated the testing.

It is ironic that very small maxima, which often are of little or no interest, cause these difficulties. Fortunately, they cause complications only if one systematically varies separation conditions. At any single (S_t, A_t) coordinate, conditions appropriate to the SMO can usually be found. One should bear in mind, though, that those conditions are detective as well as separatory.

5. Conclusions

This paper shows that the number of components predicted by the SMO is a relative and not absolute quantity, because it is estimated from only those maxima recognized by the integrator. This number, in turn, varies with conditions of signal processing. Maxima that are not detected, or that are detected but are ignored by processing software for one or more reasons, make no contribution to the interpretation.

It is clear that a systematic protocol for adjusting integrator thresholds must be developed, if their effects on estimates of $\overline{m}$ and other statistical parameters are to be isolated from effects due to separation conditions. This may prove to be a formidable challenge. Nevertheless it is necessary for systematic studies of the kind pursued by this group.

It should be noted that other integrator parameters, e.g., baseline drift, width, etc., also are adjustable. The importance of these parameters, however, is secondary to that of slope and area thresholds, as indicated by the close agreement between experiment and simulation.

Other factors may be relevant in some instances. If one were interested in trace analysis, for example, one might normalize all peak areas to the area of the largest peak. Relative to the study reported here, fewer maxima would be detected at a given (S_t, A_t) coordinate, since peak area is smaller. This coordinate itself also could be scaled, however, such that the area normalization had no effect. Clearly, the results one obtains from any experiment will depend on the magnitudes of the signal, S_t, and A_t. However, only a relative, and not absolute, magnitude is important, as also was demonstrated by the computer simulations.

Finally, this study calls into question a guideline suggested in previous publications for gauging chromatographic saturation by comparing experimental chromatograms to published simulated ones [14]. In such simulations, all visible maxima contribute to saturation, but in an experimental chromatogram, only detected maxima contribute to saturation. Specifically, maxima that are visible to the eye but are not detected by an integrator create a misleading impression of high saturation. The converse also is true; maxima that are detected by an integrator but are not visible to the eye create a misleading impression of low saturation. The only way this guideline can be applied correctly is to attenuate or amplify a chromatogram to a level where maxima that are not detected by the integrator are not visible and maxima that are detected are visible; for practical reasons, this is not always possible.

Acknowledgements

The authors thank Lee Olszewski and Jim Mott of the Delta Instrument Company for instructive comments on the operation of the GC-9AM chromatograph and CR-6A integrator. The assistance of Eric Fletcher in processing the computer-generated chromatograms is gratefully acknowledged. This work was supported by the National Science Foundation (CHE-9215908).

References

[1] J.M. Davis, *J. Chromatogr.*, 449 (1988) 41.
[2] S.L. Delinger and J.M. Davis, *Anal. Chem.*, 62 (1990) 436.

[3] F.J. Oros and J.M. Davis, *J. Chromatogr.*, 550 (1991) 135.
[4] D.P. Herman, M.-F. Gonnord and G. Guiochon, *Anal. Chem.*, 56 (1984) 995.
[5] J.M. Davis and J.C. Giddings, *Anal. Chem.*, 57 (1985) 2178.
[6] F. Dondi, Y.D. Kahie, G. Lodi, M. Remelli, P. Reschiglian and C. Bighi, *Anal. Chim. Acta*, 191 (1986) 261.
[7] S. Coppi, A. Betti and F. Dondi, *Anal. Chim. Acta*, 212 (1988) 165.
[8] F. Dondi, T. Gianferrara, P. Reschiglian, M.C. Pietrogrande, C. Ebert and P. Linda, *Anal. Chim. Acta*, 485 (1990) 631.
[9] J.M. Davis, *Adv. Chromatogr.*, 34 (1994) 109.
[10] L.R. Snyder, *J. Chromatogr. Sci.*, 10 (1972) 200.
[11] J.P. Foley, *J. Chromatogr.*, 384 (1987) 301.
[12] E. Grushka and D. Israeli, *Anal. Chem.*, 62 (1990) 717.
[13] N. Dyson, *Chromatographic Integration Methods*, Royal Society of Chemistry, Cambridge, 1990.
[14] J.M. Davis and J.C. Giddings, *Anal. Chem.*, 57 (1985) 2168.
[15] J.M. Davis, *Anal. Chem.*, 66 (1994) 735.
[16] J.M. Davis, F.-R.F. Fan and A.J. Bard, *J. Electroanal. Chem.*, 238 (1987) 9.
[17] L.J. Nagels, W.L. Creten and P.M. Vanpeperstraete, *Anal. Chem.*, 55 (1983) 216.
[18] M. Martin and G. Guiochon, *Anal. Chem.*, 57 (1985) 289.
[19] M.Z. El Fallah and M. Martin, *J. Chromatogr.*, 557 (1991) 23.
[20] M. Johll, D. Bowlin and J.M. Davis, *Anal. Chim. Acta*, submitted for publication.