CHROMSYMP. 2236

Effect of stationary phase on predictions of the statistical model of overlap from gas chromatograms

FRANCES J. OROS and JOE M. DAVIS*

Department of Chemistry and Biochemistry, Southern Illinois University. Carbondale. IL 62901 (USA)

ABSTRACT

The numbers of detectable components in each of four complex mixtures, which vary widely in polarity, were estimated with the statistical model of overlap from several dozen gas chromatograms. These chromatograms were generated at several heating rates and flow-rates on each of four capillary columns, with stationary phases that varied widely in polarity. For each mixture, the estimates so determined were grouped according to stationary phase. These groups then were compared by analysis of variance to establish any dependence of the model's predictions on stationary phase. These analyses show that internally consistent estimates can be calculated, unless the polarities of the mixture and stationary phase are highly mismatched or physical constraints on the chromatography prevent the establishment of random elution orders. It is also shown that the numbers of maxima detected in chromatograms generated by using a simple temperature program are comparable to those predicted by the statistical model of overlap, even when the elution order of the components is not deliberately randomized. Some problems inherent in measuring the retention times of peak maxima, which are needed to apply the model, by data processors are also addressed.

INTRODUCTION

This paper addresses the dependence on the stationary phase of parameters estimated with the statistical model of overlap (SMO) from gas chromatograms. The SMO $[1]$ is one of several proposed statistical theories $[1-6]$ that quantify the degree of peak overlap in chromatograms. The principal conclusions one draws from these theories is that the chromatograms of complex, multi-component mixtures contain a surprisingly large fraction of multiplet peaks and that the chromatography of such mixtures on a single column is inadequate for the resolution of the mixture components. Several applications of the SMO to both computer-generated $[7-10]$ and experimental $[9,11-15]$ chromatograms have been reported. In particular, the SMO has been confirmed experimentally by its application to gas chromatograms of synthetic mixtures containing known numbers of detectable components [14,15]. The theory also has been reinterpreted $[16-18]$ and extended $[19]$ by others.

In accordance with the Poisson statistics on which the SMO is based, the basic prerequisite to the model's application is the random elution of the various mixture components from the chromatographic column. Gas chromatographic studies have shown that, in some instances, random elution orders can be established by using

a single linear temperature program [9,11,14]. In other instances, several contiguous linear programs (which are largely determined by trial and error) are required to force components into random elution orders [13,15]. By and large, the minimal condition necessary for the model's application can be found.

Nevertheless, several practical constraints on the model do exist. One such constraint is that the saturation, or relative component density, of the chromatogram must not exceed a certain limit, because the statistics do not account for the amplitudes of the component peaks $[7-10]$. (A recent statistical theory $[6]$ shows considerable promise [20] in dealing with these amplitude effects, at least to some extent.) Another constraint, which is usually minor, is that the resolution factor, which resolves single-component peaks into separate observable peaks, depends on the signal-tonoise ratio [8]. Both of these constraints exist because the model does not address the fundamental chromatographic attributes of peak amplitude and noise. In fact, the only chromatographic attribute incorporated into the model is the peak capacity of the column [l]. In general, the basic independence of the model from specifics of the chromatography is one of its most attractive features, because the model can be applied to chromatograms of widely varying attributes. At the same time, one must be cautious in interpreting the model's predictions in cases where other chromatographic attributes, which have not been studied in detail by simulation or experiment, are or could be important.

The gas chromatographic applications of the SMO reported in the literature show the same basic trends as those determined from computer simulations [7-lo]. In most of these studies, the stationary phase of the column was carefully chosen to be compatible with the mixture. In general, the compatibility of the stationary phase with a given mixture is a complex chromatographic attribute, whose influence on the predictions of the SMO has not been addressed in any detail. The only work of this kind, of which we are aware, is that of Coppi *et al.* [13], who gas chromatographed a camomile extract on an OV-1 capillary and two Carbowax 20M capillaries of different length. They found that the statistical parameters estimated from these chromatograms were internally consistent for several temperature programs. While these findings are certainly encouraging, they perhaps comprise too small a data set to be able to draw any general conclusions.

In particular, any detailed study of this attribute must answer the question of the care with which one must choose a stationary phase if one wishes to apply the SMO to a multi-component gas chromatogram. Many factors affect the interactions of mixture components with the stationary phase, but perhaps the most important is the phase polarity, which controls the solubility of the components in the phase. If one can simply match loosely the polarity of the mixture with that of the stationary phase $(e.g.,)$ if one can choose any polar stationary phase for the partial resolution of a polar mixture), then one can be guided by experience in the choice of that phase and can apply the SMO with little or no concern about this attribute. On the other hand, if the predictions of the model depend strongly on the stationary phase, then the utility of the SMO is called severely into question, and one will have little or no incentive to use it to characterize multi-component separations. To answer the above question more thoroughly, one must estimate statistical parameters from many chromatograms, which are generated from numerous mixture and stationary phase combinations, and check for internal consistency among the estimates.

We report here results based on the application of the SMO to several dozen gas chromatograms generated by the partial resolution of four mixtures on four capillary columns having different stationary phases. Both the various mixtures (a synthetic mixture of hydrocarbons, a coal tar extract, lime oil and peppermint oil) and the stationary phases (DB-I, DB-1701, RTX-225 and DB-WAX) span a wide range of polarity. In brief, several chromatograms were generated at several heating and flow-rates for each stationary phase-mixture combination. Several criteria were subsequently applied to evaluate the suitability of these chromatograms for analysis by the SMO. If these criteria were satisfied, then statistical parameters, including approximations to the number of detectable components, were estimated from these chromatograms. These approximations then were grouped into data sets corresponding to different stationary phase-mixture combinations. The data sets so developed from each mixture were compared by a one-way analysis of variance to determine if any statistical differences existed among them.

As an alternative to the experimental program outlined above, one could perhaps carry out a study of this type by theoretically estimating retention indices of various components on various stationary phases and constructing from these indices a series of computer-simulated chromatograms, which then could be analyzed. We chose to investigate the problem experimentally because, as observed elsewhere [151, important systematic effects, which cannot be anticipated by computer simulation, are often found in studies of this kind. Indeed, such effects are detailed below. Further, we anticipate that such a study will be more "believable" than one based on computer simulation.

THEORY

Application of the SA40

The theory underlying the SMO is detailed elsewhere [1]; only its application is reviewed here. The mean number *p* of detectable "peaks" in a chromatogram containing m detectable single-component peaks (SCPs) is

$$
p = \bar{m} \exp(-\bar{m}/n_c) = \bar{m} \exp(-\alpha) \tag{1}
$$

where \bar{m} is a statistical approximation to m , n_c is the peak capacity of the chromatogram (or column) and $\alpha = \bar{m}/n_c$ is the chromatographic saturation. With this definition of α , one can express eqn. 1 in the dimensionless form

$$
p/n_{\rm c} = \alpha \exp(-\alpha) \tag{2}
$$

By fitting data from experimental chromatograms to these expressions, one can estimate the statistical parameters \bar{m} and α , which are measures of the quality of separation. Several approaches have been suggested for this fitting $[1,7-10,12]$. In perhaps the simplest of these, the so-called single-chromatogram method [151 by which one estimates these parameters from a single chromatogram, one expresses eqn. 1 in the equivalent form [lo]

$$
\ln p = \ln \bar{m} - \bar{m}x_0/X \tag{3}
$$

where x_0 is any arbitrarily chosen span, X is the span of the chromatogram to which the SMO is applied and p is interpreted as the number of gaps between adjacent chromatographic maxima which exceed x_0 . In accordance with theory, a plot of ln p vs. x_0/X is a line whose slope $(-\bar{m})$ and intercept (ln \bar{m}) yield two statistical approximations to *m*, which are termed m_{sl} and m_{in} , respectively [10]. By assigning the appropriate statistical weights to data in this plot, one also can calculate the standard deviations of $m_{\rm{sl}}$ and $m_{\rm{in}}$ [10]. From $m_{\rm{sl}}$, $m_{\rm{in}}$ and their standard deviations, one then can calculate a pooled or average approximation to m [14], which is designated m_{ave} . The saturation α then can be estimated from eqn. 1 as [9]

$$
\alpha = -\ln(p_m/m_{\rm ave}) \tag{4}
$$

where $p = p_m$ is the number of chromatographic maxima. The identification of the "peak" number p in eqn. 4 with the number p_m of chromatographic maxima implies that the saturation α , as defined by the minimum resolution R_{α}^{*} required to resolve adjacent SCPs, is less than about 0.5, when $R_s^* \approx 0.5$ [7-10,15].

As intimated in the Introduction, several criteria must be satisfied if one is to obtain accurate estimates of *mave* by using this method. These criteria are justified elsewhere [10] and are simply stated here: m_{sl} must equal m_{in} within statistical error, the graph of $\ln p$ vs. x_0/X must be linear, the graph of the number of chromatographic maxima eluting per unit interval of time vs. time must be uniformly constant and α must be less than 0.5, when $R_*^* = 0.5$. In addition, the overall appearance of the chromatogram must be consistent with this *a* value, as judged by a comparison of the experimental chromatogram with computer-generated chromatograms of the same saturation [11,15]. Unless noted otherwise, the results reported here were estimated from chromatograms which satisfied these criteria. The specific application of these criteria to this study is detailed below.

Analysis of variance (ANOVA)

Details on the Model I ANOVA used here can be found in appropriate references [21]. By using ANOVA, one can compare two or more groups of data and evaluate (within limitations) the statistical equivalence or non-equivalence of these groups. The basis of comparison is the Fisher ratio F_{v_1,v_2} , where v_1 and v_2 are the numbers of degrees of freedom among and within the compared groups, respectively.

In this study, each data group is composed of a series of *mave* values. Each value of m_{ave} in any of these groups was determined by applying the SMO to a single chromatogram of a specific mixture. This chromatogram, in turn, was developed on a specific capillary at a unique heating and flow-rate. Each group, therefore, corresponds to a specific mixture-stationary phase combination, and the various m_{ave} values comprising that group correspond to determinations made under different experimental conditions (which are detailed below). For each mixture, the groups of m_{ave} values corresponding to different stationary phases were compared by ANOVA. A schematic outline of the experimental generation and statistical comparison of such groups is depicted in Fig. 1.

We shall interpret the statistical equivalence of these groups as evidence that the m_{ave} values estimated from chromatograms of a given mixture, as developed on capillaries with different stationary phases, are identical. In other words, the stationary

Fig. 1. Schematic outline of the experimental generation and statistical interpretation of groups of m_{ave} values in this study. Quantity $m_{\text{avg}}^{\theta q}$ is the qth member of the group of m_{avg} values corresponding to stationary phase θ . The number q of members per group can vary. Each group member is evaluated from a chromatogram generated at a unique heating rate r and flow-rate F, such that the ratio r/F is constant for all members of that group. Different groups, however, are associated with different *r/F* ratios.

phase has no statistically measurable effect on the estimates. This equivalence is quantitatively measured by the inequality $F_{v_1,v_2} < F_{v_1,v_2}^*$, where F_{v_1,v_2}^* is a critical Fisher ratio for a given confidence level (here the 95% confidence level is used). In contrast, unless noted otherwise, we shall interpret the statistical non-equivalence of these groups (as measured by $F_{v_1,v_2}^* < F_{v_1,v_2}$) as evidence that the numbers m_{ave} depend on the stationary phase. This interpretation initially may seem too simplistic, because the dependence of these numbers on other attributes, e.g., saturation α , is well established [lo]. As stated above, however, extensive criteria have been developed by which to judge if these other attributes affect the estimates, and these criteria are employed in this study.

By our use of the Model I ANOVA, we are assuming implicitly that the m_{ave} values comprising each group are drawn from the same statistical population. As stated above, the *mave* values comprising each group were estimated from a series of chromatograms of a given mixture, as developed on a given capillary. Each chromatogram in this series, however, differs in the heating rate of, and volumetric flow-rate through, the capillary (see Fig. l), although the ratio of these rates is held constant. Clearly, our implicit assumption that a single statistical population describes these m_{ave} values will not be valid, unless the m_{ave} values so determined are independent of these variations of heating and flow-rates. Previous studies have suggested that this independence indeed exists [9,13,15].

As an alternative to this protocol, these groups of *mave* values could have been generated more simply by selecting a single heating rate and flow-rate for each mixture-stationary phase combination and by generating replicate chromatograms at these fixed rates. We believe that our variation of these rates provides a significantly larger body of information for interpretation. Further, our variation of these rates enables us to confirm the independence of *mave* of these variations and also to discover any unexpected trends. By and large, the results presented below verify this independence, although some subtle effects are also discovered.

EXPERIMENTAL

Preparation of mixtures

The synthetic hydrocarbon mixture, which previously was used in a detailed experimental verification of the SMO [15], was prepared from 54 C_6-C_{10} alkane and alkene standards, ethylbenzene and cyclopentanone. The preparation of this 56-component mixture is detailed elsewhere [151; here, we note that the mean concentration of the mixture components was 1% in the solvent, tetradecane. The coal tar extract was obtained from NET as SRM 1597 (Complex Mixture of Polynuclear Aromatic Hydrocarbons) in toluene solvent and was used as purchased. Peppermint and lime oils were kindly donated by the A. M. Todd Co. (Kalamazoo, MI, USA). One part of each of these oils was dissolved in ten parts of methylene chloride.

Chromatography

Fused-silica capillaries bonded to DB-1, DB-1701, RTX-225 and DB-WAX stationary phases were purchased from commercial sources (J & **W** Scientific, Folsom, CA, and Restek, Bellefonte, PA, USA). Some physical properties of these capillaries and phases are reported in Table I. The capillaries were incorporated into a Shimadzu (Columbia, MD, USA) GC9-AM modular gas chromatograph equipped with a Model SPL-9 split-splitless injector and flame ionization detector and interfaced to a Shimadzu C-R6A Chromatopac data processor. Aliquots of 1.0 μ l were injected, with a 2O:l split of helium carrier gas. Helium was also passed to the detector as make-up gas at a flow-rate of 37 ml/min, except for the hydrocarbon mixture (in the latter instance, we simply did not realize that the make-up gas was turned off). The air and hydrogen flow-rates to the detector were 320 and 36 ml/min, respectively. Flowand heating rates and injector and detector temperatures were varied for different mixture-stationary phase combinations and are reported in Table II. To minimize the possible contamination of any chromatogram by the elution of high-boiling components from previously injected mixtures (especially the coal tar extract), the oven temperature was held at the maximum temperatures reported in Table II for 30 min after the apparent completion of each chromatogram.

' Maximum recommended temperature for stationary phase.

 \mathbf{r} $\overline{1}$ $\overline{1}$

PREDICTIONS OF OVERLAP FROM GAS CHROMATOGRAMS

TABLE II

Data acquisition and use

The voltages generated by the flame ionization detector were amplified to the same level in developing chromatograms of a given mixture on different capillaries, although this level was varied for different mixtures. Also, the threshold peak area, below which maxima were ignored by the data processor, was held constant in developing chromatograms of a given mixture on different capillaries. Both these precautions were taken to minimize the spurious detection or oversight of maxima in chromatograms of a given mixture. In addition, the flow-rate of the make-up gas was set to a relatively high value (37 ml/min) to minimize any changes in the detector response (e.g., peak height) at different capillary flow-rates.

The retention times of chromatographic maxima were measured by the data processor. For reasons discussed below, the retention times of maxima in chromatograms of the coal tar extract also were measured by manually digitizing the relative positions of peak maxima with a True Grid 1011 Digitizer (Houston Instruments, Austin, TX, USA) interfaced to an Apple IIe microcomputer. Regardless of the mode of generation, the differences between adjacent pairs of retention times were calculated to generate plots of $\ln p$ *vs.* x_0/X . The procedural details underlying the generation of these plots are detailed elsewhere [11].

Ideally, all maxima should be included in span X to avoid discrepancies in comparing statistical parameters determined from one capillary to those determined from another. This ideal was achievable with the hydrocarbon mixture and oils but not the coal tar extract, because the relatively low temperature limit T_{max} of the DB-1701 capillary precluded the rapid elution of some of its high-boiling components (the temperature limits of the stationary phases are reported in Table I). These components could not be included in span X , because their slow elution rate was incompatible with the rapid elution rate of the more volatile components. In other words, their inclusion in span X would destroy an otherwise random elution order. A compromise was required, in which a common span X was chosen in the chromatograms developed on this and the DB-1 capillary. This selection was not complicate by slight shifts in the component elution orders, which were minor.

Criteria for evaluation of statistical estimates

As stated in the Introduction, the basic prerequisite for the calculation of good statistical parameters with the SMO is the random elution of mixture components from the chromatographic column. As observable maxima are not necessarily SCPs, however, this randomness cannot be determined by simply inspecting the chromatogram. Consequently, several previously developed tests were used to gauge this randomness, including the linearity of the $\ln p$ vs. x_0/X plot, the uniformity (or constancy) of the number of maxima eluting per unit time and the consistency between the appearance of the chromatogram and the α value predicted by eqn. 4.

The application of these tests was straightforward. Each mixture was first chromatographed on the capillary whose stationary phase most closely matched its polarity. This initial chromatography was similar to that routinely used for complex mixtures (e.g., flow-rates of 2–3 ml/min and heating rates of 2–6 \degree C/min). For the coal tar extract [22] and lime and peppermint oils [23], reference chromatograms were available for purposes of comparison. Next, the uniformity of the elution rate of the maxima in this chromatogram was evaluated by plotting the number of maxima eluting per unit time $(i.e.,$ the density of the maxima) against elution time. Computer simulations have shown that this number is fairly constant when the underlying SCP distribution is random [lo]. The uniformity of this distribution was not evaluated statistically but by qualitative inspection only. The chromatography was then adjusted, if needed, to achieve the best possible compliance with this inspection. These adjustments principally consisted of changing the single heating rate into two or more rates, which were applied to the capillary at different times during the chromatography. In some instances, only partial compliance could be achieved. The lime and peppermint oils, in particular, produced small regions of maxima density which greatly exceeded that throughout the remainder of the chromatogram. These small nonuniformities do not seem to affect the prediction of statistical parameters, however, as detailed below.

Once the uniformity of the elution rate of the maxima was nearly optimum, a plot of $\ln p$ vs. x_0/X was constructed and checked for linearity. These plots were usually linear, because the existence of a uniform maxima density implies that the SCPs are randomly distributed in time [10]. On some occasions, however, the chromatography required some minor adjustments for one to attain linearity in these plots. Following these changes, the maxima density was again checked for its uniformity. When all necessary criteria were satisfied, then values of m_{sl} , m_{in} , m_{ave} and α were determined. In some instances, the α value for a given chromatogram exceeded the threshold value of 0.5. The statistical estimates from such chromatograms were rejected, unless noted otherwise.

Additional estimates of these four statistical parameters were then calculated from additional chromatograms of the mixture, which were generated by proportionally varying the heating rate r of, and volumetric flow-rate F through, the capillary. To a first approximation, this proportional variation maintained the retention temperatures of the mixture components [9,15,24] and consequently preserved the empirically determined elution order. For each chromatogram so developed, plots of maxima density vs. time and ln p vs. x_0/X were generated and checked for uniformity and linearity, respectively. Usually, these criteria were satisfied, although occasionally some minor changes in the chromatography were required (interestingly, these criteria were far more difficult to satisfy at low flow-rates, e.g., 0.5-1.0 ml/min, than at higher values). Once these criteria were satisfied, then new estimates of m_{sl} , m_{in} , m_{ave} and a were calculated.

Some representative *r/F* programs are reported in Table II for the different mixture-stationary phase combinations in this study. These programs are representative only; the minor adjustments to these programs are not listed here. Only twelve programs, instead of sixteen, are reported, because random elution orders could not be established for four mixture-stationary phase combinations. The arrangement of these combinations is such that the least and most polar combinations are found in the lower left and upper right positions in the table, respectively.

A final criterion was then applied to gauge the acceptability of the calculated parameters. Each chromatogram was compared with the computer-simulated chromatograms in ref. 10 to gauge if the calculated α values were consistent with the expected appearance of the chromatogram. In almost all instances, this consistency was found. As discussed below, however, the **mave** values estimated from a few chromatograms were rejected because of such inconsistency.

STATISTICAL PARAMETERS m_{new} AND α DETERMINED FROM, AND MAXIMA NUMBERS p_{m} DETECTED IN, CHROMATOGRAMS DEVELOPED FROM THE MIXTURE-STATIONARY PHASE COMBINATIONS IN THIS STUDY DEVELOPED FROM THE MIXTURE-STATIONARY PHASE COMBINATIONS IN THIS STUDY

STATISTICAL PARAMETERS mave AND a DETERMINED FROM, AND MAXIMA NUMBERS pm DETECTED IN, CHROMATOGRAMS

TABLE III

 $S_{\rm r}$. The different capital $E_{\rm r}$ is a set of \sim . " 10-m capillary. ^b 15-m capillary. $\frac{15-m}{2}$ capillary.

These procedures, with minor refinements, were then repeated to develop chromatograms of the same mixture on the other capillaries. Instead of using a simple temperature ramp to develop trial chromatograms, however, the previously developed *r/F* programs were used as guidelines.

Analysis of variance (ANOVA)

The estimates m_{ave} computed as detailed above for each mixture-stationary phase combination were grouped. For a given mixture, the various groups corresponding to different stationary phases were compared by one-way ANOVAs, formulae for which are given in standard references [21].

RESULTS AND DISCUSSION

Interpretation of ANOVAs

Table III reports the means and standard deviations of the statistical parameters m_{ave} and α estimated from chromatograms developed from twelve of the sixteen mixture-stationary phase combinations in this study. Also reported are the means and standard deviations of the numbers p_m of maxima detected in these chromatograms by the data processor (for reasons given below, the numbers p_m reported for the coal tar extract were determined by manual digitization and not the data processor). Several Fisher ratios F_{v_1,v_2} and F_{v_1,v_2}^* (the latter for the 95% confidence level) are reported at the bottom of the table.

Four of the sixteen positions in Table III are empty, because various constraints on the chromatography prevented the establishment of random elution orders for these mixture-stationary phase combinations. For example, the lower temperature limit of the RTX-225 stationary phase $(40^{\circ}C)$ precluded sufficient cooling of the capillary to adequately retain the low-boiling constituents of the hydrocarbon mixture. In addition, the components of this mixture eluted from the DB-WAX capillary as several small clusters of maxima with large intervening gaps. Finally, the relatively low upper temperature limits, T_{max} , of the RTX-225 and DB-WAX phases (reported in Table I) prevented the rapid elution (and inclusion in a random elution order) of a substantial fraction of the high-boiling components in the coal tar extract. While our inability to develop random elution orders for these cases is disappointing, it is perhaps not surprising, because the polarities of these mixtures and stationary phases are poorly matched. Indeed, our ability to develop random elution orders for other cases of poorly matched polarities (e.g., for peppermint oil on a $DB-1701$ phase) was more surprising than these failures.

In contrast to the above cases, the components of the hydrocarbon mixture could be forced into random elution orders on the DB-1 and DB-1701 capillaries (the chromatography of this mixture on the DB-1 capillary has been detailed elsewhere [151). As shown by the Fisher ratios at the bottom of the second column in Table III, no statistical difference exists between the m_{ave} values estimated from the chromatograms developed on these capillaries $(F_{1,24} < F_{1,24}^*)$. The estimates are not only statistically equivalent but also accurate; as reported in Table III, the mean estimates differ from the true number of detectable components [56] by ≤ 1 . The p_m values detected in, and the α values estimated from, chromatograms developed on the DB-1701 capillary are larger and smaller, respectively, than their DB-1 counterparts, probably because the former capillary was longer and more efficient than the latter (see Table I).

Unlike for the other mixtures considered here, the number *m* of detectable components in the hydrocarbon mixture, which was prepared from analytical standards [15], is known. This knowledge enables us to interpret further the data summarized in the second column of Table III. Fig. 2 is a plot of p/n_c vs. α constructed from these data. The solid curve is a graph of eqn. 2, whereas the points represent experimental results. A similar plot of data developed from 10- and 30-m lengths of a DB-1 capillary was presented elsewhere [15]; here, only data from the 10-m capillary are displayed. As detailed in that work [15], p was approximated as p_m , α was calculated from eqn. 4 and n_c was calculated as $56/\alpha = -56/\ln(p_m/m_{ave})$. By such calculations, an exact agreement between experiment and theory is expected only if $p = p_m$ and $m = m_{\text{ave}} = 56$ [15]. The close agreement between experiment and theory again affirms the accuracy of the statistical parameters estimated with the SMO.

Fig. 2. Dimensionless plot of p/n_c vs. α ($R_s^* = 0.5$). The solid curve is a graph of eqn. 2; the data were estimated from chromatograms of the hydrocarbon mixture developed on the (\blacksquare) DB-1 (10-m) and (\lozenge) DB-1701 capillaries.

In contrast to the hydrocarbon mixture, statistical parameters for the lime oil mixture were determined from chromatograms developed on all four capillaries. The Fisher ratios at the bottom of the fourth column in Table III indicate that no statistical difference exists among the various *mave* values determined from these capillaries $(F_{3,27} < F_{3,27}^*),$ despite their marked differences in stationary phase polarity. In particular, the mean *mave* values estimated from the three most polar capillaries agree within the value 2. Fig. 3 depicts examples of typical chromatograms developed from this mixture on these capillaries. In each chromatogram, the span X is indicated. In contrast to one's first impressions, a substantial number of small maxima were detected in the intermediate region of the DB-1 chromatogram.

For reasons outlined above, the statistical parameters for the coal tar extract were estimated from chromatograms developed on the DB-1 and DB-1701 capillaries only. In contrast to the finding reported above, the Fisher ratios at the bottom of the third column in Table III indicate a statistical difference exists between these *mave* values $(F_{1,13} > F_{1,13}^*)$. This finding is disquieting, because this mixture is relatively non-polar and both capillaries have low polarities. However, the finding is not crippling, because the means of the two distributions differ by only ca . 10%. The reason why the ANOVA is significant is that the standard deviations associated with

Fig. 3. Typical chromatograms of the lime oil mixture developed on the DB-1, DB-1701, RTX-225 and DB-WAX capillaries. The span X to which the SMO was applied is indicated. Flow-rates: DB-1 capillary (10-m segment), 4.11 ml/min; DB-1701 capillary, 2.78 ml/min; RTX-225 capillary, 3.39 ml/min; DB-WAX capillary, 3.57 ml/min. Representative *r/F* programs are reported in Table II.

these means are even smaller than the small difference between the means. For example, the coefficients of variation for the *mave* values determined from the DB- 1 and DB-1701 chromatograms are only 6.2 and 7.3%, respectively.

In the Theory section, we argued that significant ANOVAs would indicate a dependence of m_{ave} on the stationary phase, unless other well known systematic effects were present. In this specific case, we believe that such effects are indeed present. One reason why the *mavc* values estimated from the DB-1 chromatograms are significantly smaller than their DB-1701 counterparts is that relatively high α values $(\alpha = 0.47 \pm 0.06)$, some of which clearly exceed the threshold value, $\alpha = 0.5$, are associated with the DB-1 chromatograms. As documented elsewhere, values of m_{sl} and (to a lesser extent) m_{in} are underestimated by the single-chromatogram method employed here, as α approaches this threshold [10].

To determine if this underestimation could account for the significant depression of the m_{ave} values determined from the DB-1 capillary, we reinterpreted some previously published results determined from the analysis of computer-generated

chromatograms. Table III in ref. 10 reports the means and standard deviations of the percentage errors in $m_{\rm sl}$ and $m_{\rm in}$ for the α values, 0.333, 0.500 and 0.667, as determined from the analysis of computer-generated chromatograms containing 100,200 and 300 SCPs distributed exponentially in amplitude. By computing m_{ave} from these m_{el} values, m_{in} values and standard deviations as detailed elsewhere [14], one can show that the quadratic polynomial which describes the pooled percentage error, *PE,* in these computer-generated chromatograms at these *a* values is

$$
PE = 100 \left(\frac{m_{\text{ave}} - m}{m} \right) = 108.8 \alpha^2 - 173.1 \alpha + 49.7 \tag{5}
$$

This expression for PE should not be used with α values outside the range, $0.333 \le \alpha \le 0.667$, from which it was determined. If one now corrects for the expected percentage errors in the *mave* values reported in Table III by using eqn. 5 and the reported α values, one anticipates that the number m of components in the coal tar extract is 182, as determined by analysis of the DB-1701 chromatograms, and 173, as determined by analysis of the DB-1 chromatograms. These values differ by only about 5%, unlike the uncorrected values, which differ by about 10%. The close agreement between these numbers lends credence to the idea that the saturation α is too high for the estimates determined from the DB-1 chromatograms to be considered fully reliable. However, these arguments indicate a possible trend at best and cannot be interpreted as rigorous proof of deterministic error.

The statistical parameters determined from chromatograms of the highly polar peppermint oil mixture and reported in the final column in Table III are most revealing. The first set of Fisher ratios reported at the bottom of this column indicates that no statistical difference exists among the *mave* values determined from the DB-WAX, RTX-225 and DB-1701 capillaries ($F_{2,24} < F_{2,24}^*$), whose polarities are intermediate to high. If the *mave* values determined from the 15-m length of the non-polar DB-1 capillary are now included in the ANOVA, however, significant differences among the *mave* values are found. These differences are indicated by the second set of Fisher ratios reported at the bottom of the column $(F_{3,29} > F_{3,29}^*)$. Clearly, this mixture-stationary phase combination leads to estimates which, for reasons that are not readily apparent, differ substantially from the others. Our only comfort is the realization that no experienced chromatographer would attempt to separate a mixture of this high polarity on such a non-polar capillary.

The parameter m_{ave} was also estimated from four chromatograms of this mixture, as developed on a 10-m length of DB-1 capillary. In contrast to an earlier study, which indicated that m_{ave} was fairly independent of capillary length when the polarities of the mixture and stationary phase were well matched [151, the *mave* values determined from the 10-m capillary ($m_{\text{ave}} = 69 \pm 3$) differ most significantly from those reported in Table III for the 15-m capillary ($m_{\text{ave}} = 104 \pm 6$). Interestingly, plots of ln p vs. x_0/X and maxima density vs. time were linear and uniform, respectively, for both capillary lengths. The only test, which suggested that the m_{ave} values determined from the 10-m capillary were spurious, was the inconsistency between the α values calculated from eqn. 4 and the subjective appearance of the chromatograms. In all instances, the chromatograms appeared to be far more saturated than indicated by the α values (α = 0.27 \pm 0.05). The overall results determined from this mixturestationary phase combination are puzzling.

Other findings of merit

During the course of this study, some interesting observations were made, the implications of which lie beyond the objectives outlined in the Introduction, We conclude this section with a brief description of them.

Comparison of maxima numbers developed from simple and complex r/F programs. The senior author occasionally has been criticized for portraying the chromatography of complex mixtures in too grim a light. Here, we simply wish to anticipate the objection that the overlap resulting from the use of the complex *r/F* programs reported in Table II may be worse than that resulting from the use of the simple temperature programs commonly used in analytical and clinical laboratories. To develop a proper perspective on these comparative degrees of overlap, each mixture for which a random elution order was developed on a given capillary was also chromatographed on that capillary by using a simple program consisting of a brief isothermal period, a single temperature ramp and a final isothermal period. Flow-rates were set slightly above the optimum value, e.g., $2-3$ ml/min. Table IV reports the details of these programs and the numbers of maxima, *pms,* detected by the data processor in these simple chromatograms (for reasons detailed below, the numbers of maxima, p_{ms} , in the coal-tar chromatograms were determined by manual digitization). In addition, the number of maxima, p_m , detected in the SMO-type chromatogram developed at the flow-rate closest to that adopted for the simple chromatogram is also reported. In all instances, the numbers p_{ms} are comparable to the numbers p_m . In other words, the gas chromatography of these mixtures, as carried out under typical conditions, generated chromatograms that were little, if any, better than those predicted by the SMO (and those are bad enough!). These findings should raise an unambiguous warning about the integrity of routine gas chromatographic separations of complex mixtures on single capillary columns.

This warning is also reinforced by the mean value of the ratio $p_m/m_{\text{ave}} = 0.69 \pm$ 0.05, which was calculated from the mean values of p_m and m_{ave} reported in Table III. On average, about one SCP in three is not detectable in the SMO-type chromatograms considered here. Because the numbers $p_{\rm ms}$ are comparable to the numbers $p_{\rm ms}$, a similar decrease in efficiency is also found in these simple chromatograms.

Potential variation of statistical parameters with flow-rate. The numbers of maxima, p_m , detected in, and numbers of components, m_{ave} , estimated from, chromatograms of the coal tar extract developed on the DB- 1 and DB- 1701 capillaries were found to increase systematically with increasing flow-rate, *F,* when the maxima numbers were measured by the data processor. In contrast, these numbers were essentially independent of *F* for chromatograms of the hydrocarbon mixture and lime and peppermint oils when measured by the same processor. These trends are illustrated in Fig. 4a-d, which are graphs of p_m and m_{ave} vs. *F*, as constructed from chromatograms of all four mixtures, as developed on the DB-1701 capillary. If one were to predict any such variation, one would anticipate a decrease in p_m (but not m_{ave}) with increasing *F*, because of increased non-equilibrium dispersion. Indeed, one could argue that Fig. 4a-c show such a trend, although it is slight. The trend shown in Fig. 4d for the coal tar extract, however, is opposite to this expectation. Were the m_{ave} values depicted in Fig. 4d to be used in this study (they are not), this dependence would be of concern for at least two reasons. First, it would invalidate the Model I ANOVAs used here, in which one assumes that *mave is* independent of *F.* Second, and more important,

150

TABLE IV

SIMPLE TEMPERATURE PROGRAMS FOR THE MIXTURE-STATIONARY PHASE COMBINATIONS IN THIS STUDY AND THE NUMBERS $\rho_{\rm ms}$

 4 10-m capillary.
 b 15-m capillary. ' 10-m capillary. * 15-m capillary.

Fig. 4. Plots of maxima numbers detected in, and component numbers estimated from, chromatograms developed from various mixtures on the DB-1701 capillary νs . flow-rate F . Flow-rates are reported in ml/min. The maxima numbers were measured by the data processor, unless noted otherwise. $O = m_{\text{ave}}$; $\bullet = p_m$.

it would call strongly into question the integrity of statistical parameters determined for this mixture.

A close inspection of the coal tar chromatograms showed that this variation in p_m was due principally to the detection by the data processor of very small maxima at high but not at low flow-rates. This observation suggested the variation was related more to systematic errors in the detection of maxima by the data processor than to properties of the mixture. To investigate this behavior, the retention times of all maxima in chromatograms of the coal tar extract, as developed on the DB-1 and DB-1701 capillaries, were manually digitized as detailed under Experimental. From these digitized sets of retention times, new sets of statistical parameters were calculated. In contrast to the findings reported above, both the maxima numbers p_m and component numbers *mave so* determined were largely independent of *F,* as shown in Fig. 4e for results determined from the DB-1701 capillary. Fig. 4e shows the slight decrease in p_m with F that is observed in Fig. 4a–c, instead of the substantial increase in p_m with *F* that is observed in Fig. 4d. Further, the various m_{ave} values in Fig. 4e are scattered about a constant value, in marked contrast to their counterparts in Fig. 4d. These trends also were observed for results similarly determined from the DB-1 capillary.

The manual digitization of the retention times of maxima should be an acceptable means of data acquisition, as it was used successfully in an earlier verification of the SMO [14]. To verify this procedure further, the retention times of maxima in chromatograms of the lime oil, as developed on the DB-1701 capillary, were also manually digitized. From these retention times, a new set of statistical parameters was estimated and compared with that estimated by using the data processor. Fig. 4f depicts the p_m values and m_{ave} values so determined, which are statistically equivalent to those depicted in Fig. 4b. In other words, the maxima numbers generated by, and the component numbers estimated from, the lime oil are independent of the means by which retention times were measured. However, this independence does not extend to the coal tar extract, as shown by Fig. 4d and e.

Interestingly, both uniform plots of maxima density vs. time and linear plots of In p vs. x_0/X were generated from chromatograms of the coal tar extract, as developed on both capillaries, regardless of the means by which the retention times were measured. This finding implies that the small peaks, which are not detected by the data processor at low flow-rates but are detected by the eye, are randomly distributed throughout the chromatogram, such that their oversight by the data processor still leaves a random (but less dense) distribution. Indeed, one should anticipate their distribution to be random, as no *apriori* reason exists to expect more small peaks in the beginning of a chromatogram than at its end.

What is the origin of this behavior? One characteristic of the coal tar extract, in comparison to the other mixtures, is the low concentration of its components. These concentration differences were so large that the signals generated by the flame ionization detector, in response to components of the coal tar extract, required substantially higher amplification (by a factor of 10 or more) than those generated by components of the other mixtures. Herein may lie the key to the behavior of the coal tar extract. We believe the most likely origin is the slope sensitivity of the data processor. The conservation of SCP area by the flame ionization detector, which is a masssensitive detector, requires that SCPs eluting at high flow-rates be taller and narrower than those eluting at low flow-rates [25,26]. The initial slope of a peak consequently is greater at high than at low flow-rates, and a small peak is more easily detected at high than at low flow-rates. In our system, the flow of make-up gas to the detector attenuated but did not eliminate this behavior; the highest capillary flow-rates were smaller than the make-up gas flow-rate by only a factor of ca . 3. If one were to expect shortcomings in the detection of maxima by a data processor, one would anticipate this behavior more from weak signals $(e.g., from the coal tar extract)$, which may or may not cross the slope-sensitivity threshold, than from strong signals (e.g., from the other mixtures), which always cross the threshold.

Other than to propose this explanation, we leave the origin of this behavior unresolved here. The reason that we do so is that its resolution has no fundamental bearing on the issues raised in this particular study. The only motive for our present investigation of this behavior was to find means, which were independent of *F,* for estimating *mave* from chromatograms of the coal tar extract. The acquisition of retention times by manual digitization seems to fulfill this goal. More specifically, the use of this digitization procedure enabled us to apply the SMO to the coal-tar chromatograms and the Model I ANOVA to the *mave* values so determined, and the results of these analyses are reported in Tables III and IV.

Another error associated with the detection of maxima by the data processor merits brief discussion. In chromatograms of the lime oil mixture, as developed on the DB-1 and DB-1701 capillaries, and of the peppermint oil, as developed on the DB- 1701 capillary, the data processor detected an off-scale signal as several maxima at low flow-rates but as only one maximum at higher flow-rates. When the attenuation was increased to bring this signal on-scale, only one maximum was observed. In our applications of the SMO, these signals at low flow-rates were interpreted as one maximum, with a retention time equal to the average of the various detected retention times. Little error was incurred by this averaging, because the *mave* values so determined are similar to those determined at higher flow-rates.

CONCLUSIONS

The principal conclusion of the first part of the study is that internally consistent statistical parameters can be calculated by applying the SMO to single gas chromatograms, provided that the polarity of the mixture and the stationary phase are reasonably well matched. In other words, the concern stated in the Introduction about the compatibility of the stationary phase with a mixture can largely be dismissed, if one is guided by common sense in choosing a stationary phase. This conclusion was also reached by Coppi *et al.* [13], but we have provided a substantially larger body of data to support it. Based on these data, we feel safe in proposing the general conclusion that no particular stationary phase is required to determine statistical parameters, provided that random elution orders can be established. Nor, for that matter, is a particular *r/F* program or flow-rate required. In some instances one can even obtain internally consistent estimates when the mixture and stationary phase polarities are not particularly well matched $(e.g.,\$ lime oil on the DB-1 capillary and peppermint oil on the DB-1701 capillary). As a rule of thumb, however, our results do suggest that one should avoid extreme mismatches in polarity, which can lead either to failed efforts to establish a random elution order (e.g., hydrocarbon mixture on the DB-WAX capillary) or to the calculation of estimates of questionable integrity (e.g., peppermint oil on the DB-1 capillary). This conclusion was not reached by Coppi *et al.* [13], principally because these behaviors were not encountered in their limited study. Certainly, these latter results emphasize the need to apply the SMO judiciously and to evaluate the estimates so calculated by several criteria.

The principal conclusion one draws from the (admittedly limited) comparison of simple and complex temperature programs is that the routine gas chromatographic separations of multi-component mixtures are relatively poor. In other words, one does not need the highly elaborate *r/F* programs sometimes prerequisite to the application of the SMO to develop deliberately separations of low quality. We develop these low-quality separations routinely in our laboratories, regardless of whether or not we want them. The principal conclusion one draws from the latter part of this study is that, on occasion, m_{ave} can depend strongly on the method of data acquisition. This conclusion is a simple consequence of the dependence of the SMO's predictions on what a sensor (e.g., a mechanical device or the human eye) detects: the distribution of peak maxima in a chromatogram. In general, the attainment of reliable estimates may require more labor-intensive efforts than those based on the use of processor-measured retention times, especially when the mixture is composed of many trace components.

Because this method of data acquisition is highly convenient, further study of its limitations would be meritorious.

ACKNOWLEDGEMENTS

The authors thank Milton Lee of Brigham Young University for his suggestion of SRM 1597 as a test mixture, William Faas of the A. M. Todd Co. (Kalamazoo, MI, USA) for his donation of lime and peppermint oils and Lee Olszewski and Steve Roepke of Delta Instrument (Waterloo, IL, USA) for helpful discussions. This study was supported in part by the Office of Research Development and Administration at Southern Illinois University.

This study was presented at the *21st Annual Ohio Valley Chromatography Symposium, Hueston Woods State Park Lodge, OH, June 2&22,1990,* and at *FACSS XVII, Cleveland, OH, October 7-12, 1990.*

REFERENCES

- J. M. Davis and J. C. Giddings, *Anal. Chem., 55* (1983) 418.
- K. A. Connors, *Anal. Chem., 46* (1974) *53.*
- D. Rosenthal, *Anal. Chem., 54* (1982) *63.*
- L. J. Nagels, W. L. Creten and P. M. Vanpeperstraete, *Anal. Chem., 55* (1983) 216.
- M. Martin, D. P. Herman and G. Guiochon, *Anal. Chem., 58* (1986) *2200.*
- A. Felinger, L. Pasti and F. Dondi, *Anal. Chem., 62* (1990) 1846.
- J. C. Giddings, J. M. Davis and M. R. Schure, in S. Ahuja (Editor), *Uhrahigh Resolution Chromatography (ACS Symposium Series, No. 250),* American Chemical Society, Washington, DC, 1984, p. 9.
- 8 J. M. Davis and J. C. Giddings, *J. Chromatogr., 289* (1984) *277.*
- 9 D. P. Herman, M. F. Gonnard and G. Guiochon, *Anal. Chem., 56* (1984) *995.*
- 10 J. M. Davis and J. C. Giddings, *Anal.* Chem., 57 (1985) 2168.
- 11 J. M. Davis and J. C. Giddings, *Anal. Chem., 57* (1985) 2178.
- 12 F. Dondi, Y. D. Kahie, G. Lodi, M. Remelli, P. Reschiglian and C. Bighi, *Anal. Chim. Acta, 191 (1986) 261.*
- 13 *S.* Coppi, A. Betti and F. Dondi, *Anal. Chtm. Acta, 212 (1988) 165.*
- 14 J. M. Davis, *J. Chromatogr., 449* (1988) 41.
- 15 S. L. Delinger and J. M. Davis, *Anal. Chem., 62* (1990) *436.*
- 16 M. Martin and G. Guiochon, *Anal.* Chem., 57 (1985) 289.
- 17 W. L. Creten and L. J. Nagels, *Anal. Chem., 59* (1987) *822.*
- 18 M. R. Schure, *J. Chromatogr.,* in press.
- 19 M. Martin, presented at the *17th International Symposium on Chromatography, Vienna, September 25- 30, 1988.*
- 20 A. Felinger, L. Pasti, P. Reschiglian and F. Dondi, *Anal. Chem., 62* (1990) 1854.
- 21 R. R. Sokal and F. J. Rohlf, *Biometry,* Freeman, San Francisco, 1981.
- 22 S. A. Wise, B. A. Benner, H. Liu and G. A. Byrd, *Anal.* Chem., 60 (1988) 630.
- 23 *Chromatography Catalog, No. 28,* Supelco, Bellefonte, PA, 1990, pp. 32-33.
- 24 W. E. Harris and H. W. Habgood, *Programmed Temperature Gas Chromatography,* Wiley, New York, 1966.
- 25 I. Halasz, *Anal.* Chem., 36 (1964) 1428.
- 26 N. Dyson, *Chromatographic Integration Methods,* Royal Society of Chemistry, Cambridge, 1990.