

CHROM. 20 916

EFFECT OF RANDOM NOISE AND PEAK ASYMMETRY ON THE PRECISION AND ACCURACY OF MEASUREMENTS OF THE COLUMN EFFICIENCY IN CHROMATOGRAPHY

JENNIFER V. H. SCHUDEL and GEORGES GUIOCHON*

**Department of Chemistry, University of Tennessee, 575 Buehler Hall, Knoxville, TN 37996-1600 (U.S.A.) and Oak Ridge National Laboratory, Analytical Chemistry Division, Oak Ridge, TN 37831-6120 (U.S.A.)*

(First received May 16th, 1988; revised manuscript received August 22nd, 1988)

SUMMARY

The reproducibility and bias of the measurement of column efficiency from chromatographic peaks were studied. Peaks of various signal-to-noise and τ/σ ratios, a measure of the asymmetry of a peak, were generated by computer simulation. A range of experimental peaks were also analyzed. Column efficiencies for a variety of peaks were computer calculated, using eight commonly employed methods. The results are compared.

The moment method gave the best results in almost all instances of noisy or skewed peaks, except in the simulated noisy cases with a signal-to-noise ratio of less than 40, and seems to be the most accurate method for the measurement of column efficiency for those with access to a computer for calculations. The most common method, the half-height method, surprisingly seems fairly sensitive to noise fluctuations, and did not perform as well as some of the other methods, including the 3σ , 4σ and 5σ methods, in the most noisy instances.

INTRODUCTION

There has been a considerable amount of research reported during the last 10 years regarding column efficiency in high-performance liquid chromatography (HPLC) and the influence of many experimental parameters on column performance. It is therefore remarkable that the literature contains so few reports on the precision or accuracy of these determinations, and that only a small number of publications contain any figures on the reproducibility of the data reported¹. The few discussions published that do deal with the advantages and inconveniences of the various methods of determining column efficiency and height equivalent to a theoretical plate (HETP), contain almost no reference to the precision of the methods considered.

This is not a very good sign of the importance placed on these investigations by the scientific community; only important, relevant data have to be precise and accurate. We decided, although belatedly, that it would be useful to study the error generation process in the determination of the plate number and HETP.

The reproducibility and bias of calculated column efficiency values are the result of the influence of many variables, such as pump or temperature fluctuations, detector response and the choice of the calculation or measurement method; the last parameter is one that is very often overlooked. The effect that this choice has on the calculated efficiency values obtained from peaks deviating from the ideal Gaussian profile or those having noisy baselines, as commonly occurs in chromatography, will be discussed here.

Historically²⁻⁴, the calculation of column efficiencies has involved the manual measurement of retention times and peak widths at various peak heights. The analyst, manually analyzing peaks, unconsciously applies averaging procedures in order to find the average signal for the baseline and use it for making determinations. However, these procedures are usually not understood, not reported and probably not linear. The systematic or indeterminate errors introduced by the manual measurement of peak efficiencies have been discussed previously⁵. More recently⁶⁻⁸, however, very sophisticated electronic integrators and procedures of data acquisition and handling by computers have begun to replace the manual methods. However, with skewed peak profiles or noisy baselines, there is still a possibility of errors in these measurements that may introduce biases into the calculation of the column efficiency. Also, in some earlier work the effects of sampling parameters on the calculation of peak areas and retention times were studied⁹⁻¹².

In this work, the effect of the choice of the measurement method on the value of the column efficiency was investigated. The reproducibility and bias of efficiency values obtained by computer simulation using eight different methods were studied for a wide variety of experimental conditions. The results obtained with some experimental chromatographic peaks are compared.

THEORETICAL

The origin and definition of the concept of theoretical plates in chromatography are related to the assumption of a Gaussian profile for the elution band of a very small, narrow, sample plug. A Gaussian profile can be obtained from either the derivation of the elution profile by means of the Craig model, for a large number of plates, or the integration of the mass balance of chromatography, assuming a linear isotherm. The elution profile is then given by

$$y = \frac{A}{\sigma} \exp \left[\frac{-(t - t_R)^2}{2\sigma^2} \right] \quad (1)$$

where y is the solute concentration in the mobile phase at time t , t_R is the retention time (elution time of the peak maximum), A is the peak area and σ is the standard deviation of the peak.

Although the elution peaks observed experimentally are rarely truly Gaussian, this model is the most convenient and remains very useful in discussing problems related to column performance. Most studies of column efficiency use this model while making some minor corrections to account for moderate band asymmetry. The column efficiency is related to the relative thickness of the peak, *i.e.*, to the ratio of the retention time to the standard deviation. The theoretical plate number is the most popular parameter characterizing the column efficiency and is defined as

$$N = \frac{t_R^2}{\sigma^2} \quad (2)$$

Measuring the column efficiency hence requires an estimate of the band variance (σ^2). This can be done by relating the standard deviation of the peak either to its width at some intermediate height, or to its area, using eqn. 1 in both instances.

The band width at a fraction x of the peak height is given by

$$W_x = 2\sigma\sqrt{2\log(1/x)} \quad (3)$$

Depending on whether one is interested in the properties of the nearly symmetrical top part of the band or in those of its tail, one will choose a large or a small value for x .

There are two other methods for estimating the band variance which are not dependent on any part of the profile. The first determines the peak area and the other uses the second central moment of the peak.

If y_0 is the peak height, the peak area is given by

$$A = \sigma y_0 \sqrt{2\pi} \quad (4)$$

The peak area is conveniently measured with an electronic integrator. This method has been extended to non-Gaussian profiles. It then becomes incorrect, however, as its basis is the Gaussian profile defined by eqn. 1.

The second method employs an alternative method for the determination of the variance, which is based on the statistical properties of distributions. The peak profile is the distribution of the residence times of the sample molecules in the column. As the signal rapidly becomes very small at times far from the retention time, the following integrals are finite:

$$\mu_0 = \int_0^{\infty} f(t) dt \quad (5)$$

$$\mu_1 = \int_0^{\infty} tf(t) dt \quad (6)$$

$$\mu_2 = \int_0^{\infty} (t - \mu_1)^2 f(t) dt \quad (7)$$

where μ_0 , μ_1 and μ_2 are the zeroth, first and second central moments of the residence time distribution, respectively. The peak area is equal to the zeroth moment and the retention time or time of the signal mass center is directly related to the first moment. The second central moment is proportional to the variance of the distribution. Between these moments and the classical parameters of chromatography, we have the following relationships:

TABLE I
VARIOUS PEAK HEIGHTS AND CONSTANTS FOR COLUMN EFFICIENCY CALCULATIONS

| Method | Peak height (%) | <i>a</i> |
|-----------------|-----------------|----------|
| 2σ (inflection) | 60.7 | 4 |
| Half-height | 50.0 | 5.54 |
| 3σ | 32.4 | 9 |
| 4σ | 13.4 | 16 |
| 5σ | 4.4 | 25 |

$$t_R = \frac{\mu_1}{\mu_0} \quad (8)$$

$$\sigma^2 = \frac{\mu_2}{\mu_0} \quad (9)$$

thus providing an independent method of determining the retention time and the variance of the distribution. This definition is valid for any distribution, *i.e.*, also for asymmetric peaks. The definition of the plate number still remains based on the assumption that the peak profile is Gaussian, however, and its extension to non-symmetrical peaks is merely empirical.

METHODS OF CALCULATION

We used eight different methods for the determination of the plate number, which we found described and used in the literature. Some of these methods have been used very frequently, others only exceptionally.

Five of these methods are based on the use of the band width measured at different relative heights to estimate the variance of the peak (see eqn. 3), namely the width at half-height and the widths at the relative heights where the band width of a Gaussian peak would be equal to 2σ, 3σ, 4σ and 5σ, respectively. The corresponding values of the relative peak heights are reported in Table I. All of these methods use the retention time of the peak maximum. The sixth method is based on the use of the peak area (see eqn. 4) to derive the standard deviation. Although they may be used, and indeed are used by many workers to account for the column efficiencies obtained with asymmetric bands, these six definitions assume that the bands are Gaussian and, therefore, should not be used when the peak asymmetry is significant.

The seventh method used to determine the column efficiency is based on the ratio of the first and second moments. The last method is asymmetry-based^{8,13}, obtained from the Foley–Dorsey equation¹⁴. A perpendicular is dropped from the apex of the peak to the baseline and the base width at 10% of the peak height, *i.e.*, $W_{0.1}$, is measured, A/B is the ratio of the distances from the perpendicular to the rear side and the front side of the peak, along the 10% horizontal line (see Fig. 1). The ratio of the retention time of the peak maximum to the width at 10% of the peak height and the ratio A/B are combined in a semi-empirical expression for the plate number:

$$N = \frac{41.7 (t_R/W_{0.1})^2}{(A/B) + 1.25} \quad (10)$$

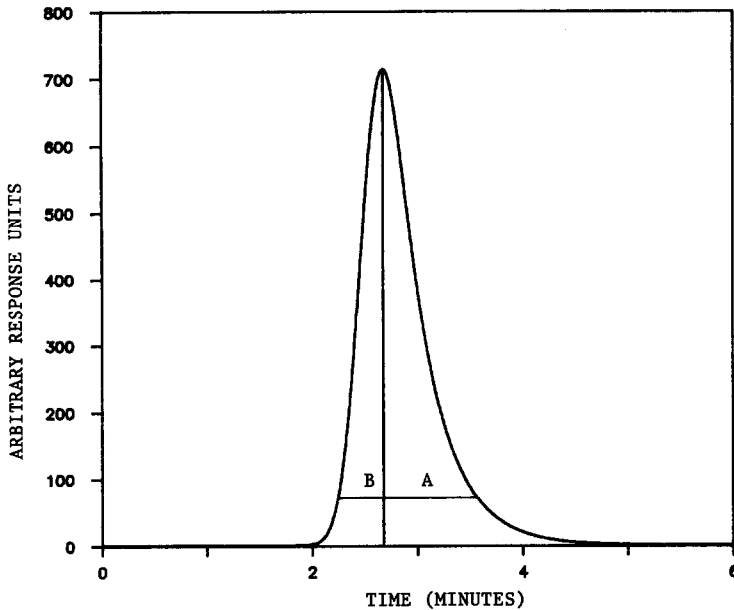


Fig. 1. Determination of peak asymmetry (A/B) at 10% of the peak height for the asymmetry-based method.

A comparison between the results obtained using these eight different methods, and between the errors made in these determinations, permits us to rank their performances.

THE SIMULATION

We simulated the procedure of data acquisition and handling used to determine the column efficiency with these different methods. A signal is generated and a noisy or drifting sequence is added to it to simulate a chromatogram. Then the various algorithms are applied for digital signal acquisition, storage and processing in order to determine the retention time, peak areas and band widths at different heights. The end results are compared to the known values incorporated in the original signal. The repetition of the entire procedure, with the same original signal but different noise sequences, gives the reproducibility of the procedure. Thus, an average bias and a standard deviation can be derived to characterize the performance of a method of column efficiency measurement. The procedure is repeated with the same signal characteristics, using the different methods.

The analog-to-digital (A/D) conversion of the signal is done by selecting the value of the signal generated, at a certain frequency, usually 5 Hz. This duplicates the operation of the most common A/D converters used in HPLC. The analog signal is thus replaced with the set of signal values at the edges of a time grid, kdt . It is important that this time grid be shifted randomly with respect to the true signal maximum, in order to avoid the systematic reproduction of artifacts^{3,10}. There is no reason for a measurement of the signal to occur systematically at the peak maximum, or at the same time away from it. To perform a random shift, the initial time value of each

simulated peak is randomly shifted from the origin of the function by an interval equal to the product of 0.01, the standard deviation and a random integer between -10 and $+10$ (ref. 10). As the interval between points is constant on each chromatogram, this shift in the time of the origin of the signal results in a shift of the calculated peak points along the curve.

The retention times, which are required by all methods except the moment method, are obtained as follows. The upper 2–4% of the peak points, around the peak maximum, are fitted to a second-order polynomial. The coordinates of the maximum of this parabola yield the best estimates of the retention time and the peak height. As even strongly tailing peaks are usually fairly symmetrical near the peak maximum¹⁵, the results should not be significantly altered for an asymmetric peak, except in extreme cases. This procedure permits the elimination of noise spikes or other fluctuations from being considered as potential maxima.

The many types of chromatographic peaks analyzed were computer generated using a PASCAL program. The peak definitions, which include the retention time, the standard deviation, the peak area, the length of the chromatogram, the signal-to-noise ratio (S/N), the random grid shift value, the time constant for the exponentially modified Gaussian^{16,17}, τ , and the data acquisition rate are read from an input file. The ordinate values of the peak are then calculated at the regularly spaced intervals of the time grid. In order not to insert a bias into the calculations, we chose a sufficiently high point density of approximately 20 points per standard deviation¹⁰. Noise is generated randomly by the program and added to each data point. Various retention times were used in order to show if any bias was introduced by increasing retention times.

The peak simulation and evaluation programs were run on a VAX cluster running VAX/VMS version 4.6. Each chromatogram is generated ten times and the same calculation program is run on all of them. The mean value of the column efficiencies is calculated for all measurement methods. The reproducibility is defined as the relative standard deviation of the ten measurements and the bias is the difference between the mean and true values. It should be emphasized that each peak is completely defined by the program user and, hence, the true value of the column efficiency is known. The reproducibility and bias of the calculations were determined using the Lotus Symphony spreadsheet program.

EXPERIMENTAL

Some experimental data were collected and analyzed. The system included a 10 cm \times 4.6 mm I.D. HPLC column packed with 10- μ m ODS particles and the mobile phase was methanol–water (70:30, v/v). The sample used was based on a standard HPLC test mixture and included uracil, ethylbenzene, toluene and benzaldehyde.

Peaks obtained under a variety of experimental conditions were acquired using a Spectra-Physics 4270 computing integrator connected to an IBM PC-AT running Autolab software. The experimental data were then uploaded to the UTCC VAX cluster for use by the efficiency calculation program.

RESULTS AND DISCUSSION

The eight methods were compared by computer evaluation of a series of noisy

TABLE II
PEAK DEFINITION SUMMARY

| Case | Retention time (s) | σ (s) | S/N | τ | Peak area |
|------|--------------------|--------------|--------------------|--------|-----------|
| A | 200 | 5-15 | 25-10 ⁶ | 0-25 | 30 000 |
| B | 1000 | 5-15 | 25-10 ⁶ | 0 | 30 000 |
| C | 2500 | 5-15 | 25-10 ⁶ | 0 | 30 000 |

Gaussian or EMG peaks generated for the analysis. The aim was to use band profiles that are representative of experimentally observed chromatographic peaks. The definitions of the simulated profiles are summarized in Table II.

The primary objective of the comparison was the determination of the accuracy of each method, relative to the "true" or "correct" efficiency value, as defined above, and the reproducibility in the presence of increasing noise or peak asymmetry.

When dealing with noisy Gaussian peaks, the true efficiency value is derived from eqn. 1, as both the retention time and the peak variance are well defined. With tailing peaks, on the other hand, the result of the moment method, applied to noiseless peaks, is assumed to be the correct value.

A systematic study of simulated noisy peaks was conducted. For the entire range of noise values and retention times, the moment method proves to have the lowest bias and the best reproducibility of the eight methods. At S/N in excess of 100, the bias and

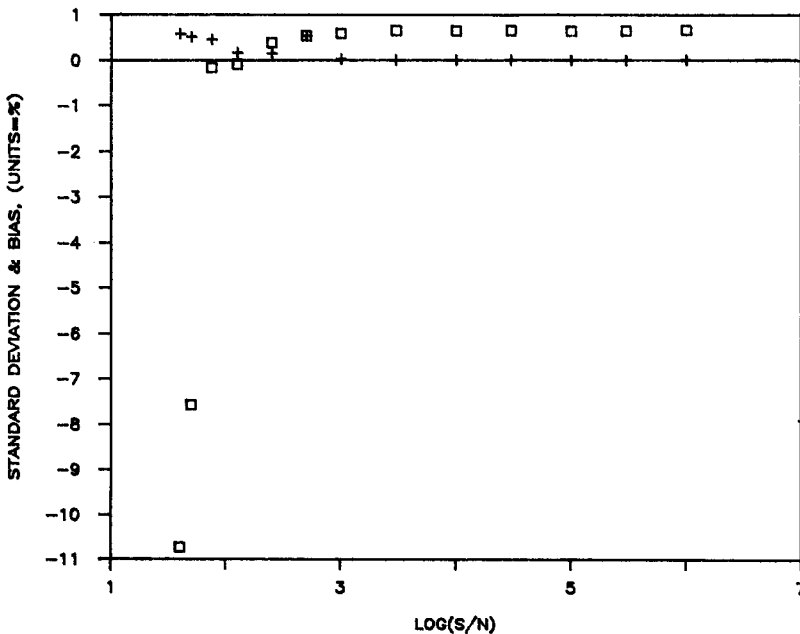


Fig. 2. Standard deviation (+) and bias (□) vs. log (S/N) using the moment method, $\sigma = 15$, $t_r = 200$ s.

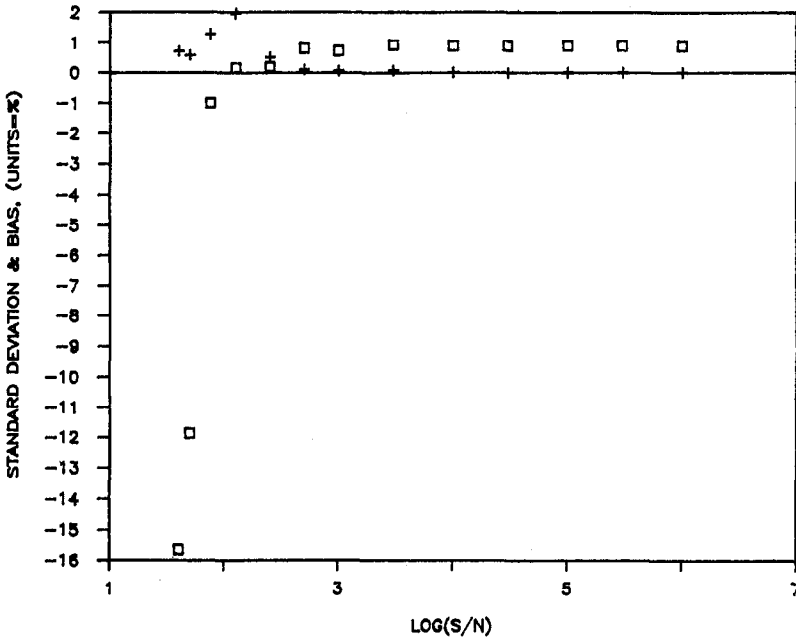


Fig. 3. Standard deviation (+) and bias (□) vs. log (S/N) using the moment method, $\sigma = 10$, $t_r = 200$ s.

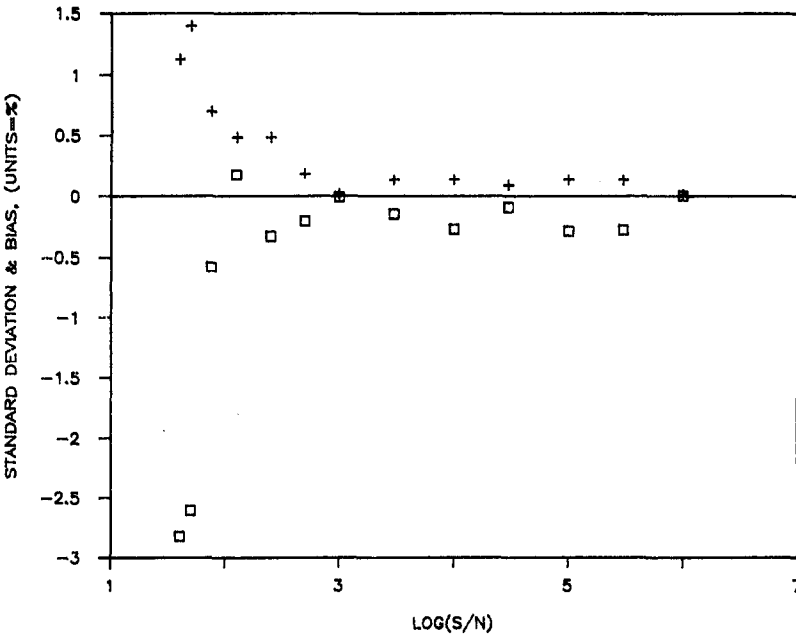


Fig. 4. Standard deviation (+) and bias (□) vs. log (S/N) using the 3σ method, $\sigma = 15$, $t_r = 200$ s.

standard deviation for this method are independent of the noise and remain close to 0.5% (bias) and less than 0.1% (precision), as shown in Figs. 2 and 3. Only at S/N less than 500 does the standard deviation begin to increase slightly. The bias is low and consistent over the entire range of retention times and peak sigma values. Some slight negative deviations in the bias appear at an S/N of 50. This is probably the result of increased difficulty in the determination of the peak thresholds.

The other calculation methods, with the exception of the peak height-area method, give poorer reproducibility and larger bias, even in the least noisy instances, when the peaks are defined to be very narrow. This may be the result of using too few data points per peak sigma value for the calculation of the polynomial fits. However, the moment method seems to be unaffected by this.

For the wider peaks, the general trend is that the 2σ and the half-height methods give consistent results, fairly low standard deviations of the plate number and bias until an S/N of *ca.* 1000, whereas the 3σ , 4σ and 5σ methods give results that are just about indistinguishable at S/N of 500 or larger. For lower S/N values, the standard deviation of the 5σ method increases rapidly whereas the 3σ method seems to follow the results of the moment most closely (see Fig. 4).

In some instances, when S/N is less than 50, the subroutine of the computer program which implements the 5σ method is no longer able to compute an efficiency value, owing to the increased difficulty in determining the peak widths at very low peak heights in the presence of a large amount of noise. Surprisingly, the half-height method seems sensitive to noise fluctuations. This method shows consistently higher and less consistent standard deviations of the plate number and biases when compared with the moment or even the 3σ method, as shown in Fig. 5.

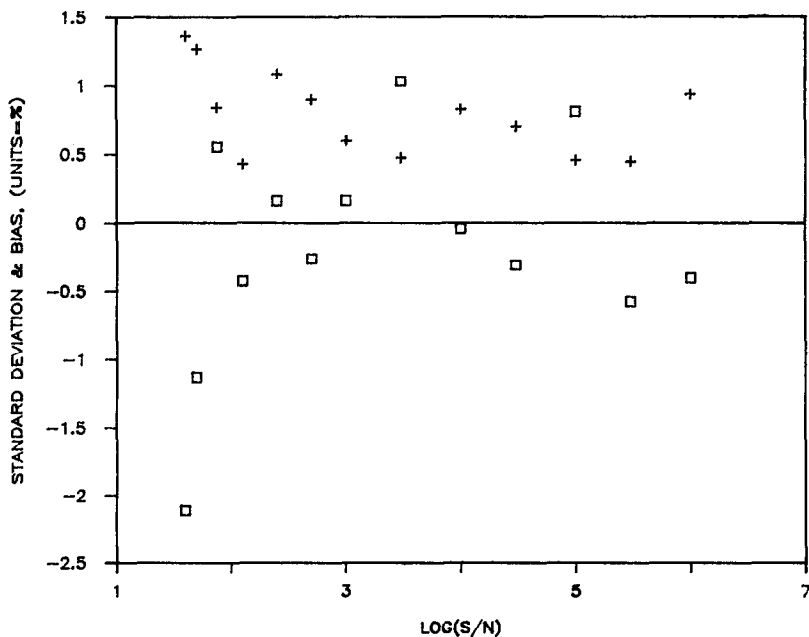


Fig. 5. Standard deviation (+) and bias (□) vs. log (S/N) using the half-height method, $\sigma = 15$, $t_r = 200$ s.

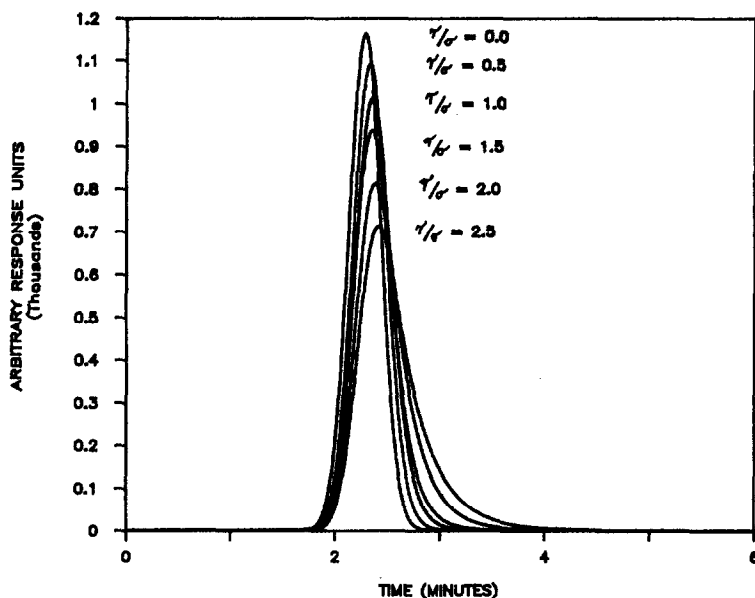


Fig. 6. Variation of peak shape with various τ/σ ratios.

In the study of the asymmetric profiles, the τ/σ ratio is varied from 0.0 to 2.5 (see Fig. 6). The peak standard deviation is held constant at 10 s while the value of τ is varied. In the least skewed case, $\tau = 0$. Increasing the value of τ result in an increase in the amount of tailing in the simulated peaks. The range of τ/σ valves investigated seems to reveal the marked difference between the results of the different calculation methods very well. It is generally observed that the methods that use peak width measurements from the upper portion of the peak profiles give efficiency values that are much higher than those obtained by the moment method⁸. The values obtained by the 5σ and the asymmetry-based methods seem to follow the results of the moment method most closely (see Fig. 7 and ref. 8). This is normal as they can take into account the effect of a larger band width closer to the baseline.

For the computer evaluation of skewed chromatographic peaks, the moment method of calculation is, consistently, the best method. However, as it is not always

TABLE III
EXPERIMENTAL RESULTS

| <i>S/N</i> | <i>Relative standard deviation (%)</i> | <i>Method</i> |
|------------|--|---------------|
| 3000-200 | 0.25-1.60 | Moment |
| 3000-200 | 1.28-2.71 | 3σ |
| 3000-200 | 1.76-3.40 | Half-height |

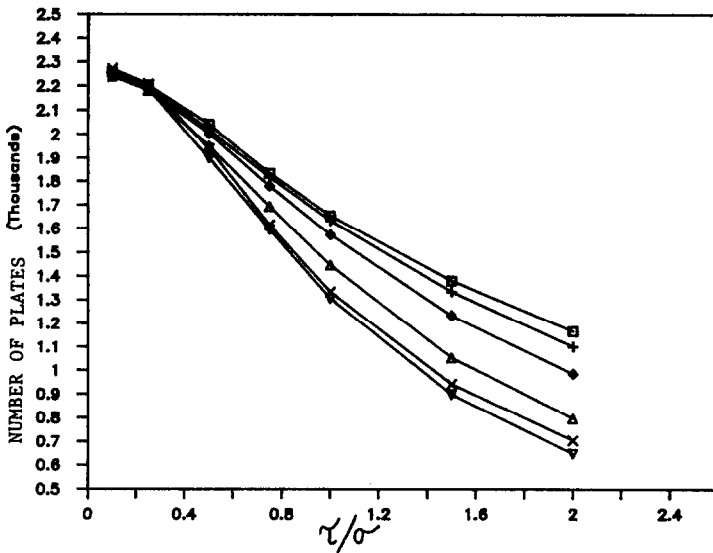


Fig. 7. Effect of the peak asymmetry on calculated column efficiency (number of theoretical plates vs. τ/σ). Method: □, inflection; +, half-height; ◇, 3 σ ; △, 5 σ ; ×, asymmetry; ▽, moment.

possible to use a computer for data acquisition, and as the method of choice for manual methods needs to be sensitive to peak asymmetry, the 5 σ or the asymmetry-based methods seem to be the best suited for this application.

The results of the efficiency calculation program for the experimentally generated peaks followed the same trends as were shown for the study of simulated chromatographic peaks. A summary of the results is given in Table III. Again, the moment method proves to be the best efficiency calculation method. The 3 σ , 4 σ and 5 σ methods give very similar results down to the lowest S/N values. For this set of experimental peaks, the 5 σ method ceased to compute efficiency values at S/N = 200. The half-height method, again, was found to be consistently less precise than the 3 σ , 4 σ and 5 σ methods at higher S/N but it was able to continue computing even when S/N became small.

CONCLUSION

Because the occurrence of peak tailing and random noise is a reality for chromatographers, it is critical that a measurement method gives an accurate value for the column efficiency. This study has determined that the choice of the measurement or calculation method has a major impact on the precision and accuracy of the values obtained. As a result, it is important that all efficiency values be accompanied by a statement of the conditions under which the value is obtained, and the relative accuracy of the calculation method used.

Throughout this work there has been a common result for both the skewed and the noisy profiles; the moment method of calculating column efficiencies from

chromatographic peaks gives the best precision and accuracy. This is not a difficult method to employ with a small computer or even a computing integrator for the calculations. However, if the measurement must be made manually, the 4σ or 5σ method seems to give adequate values for peaks with a moderate to high S/N ratio and skew. Peaks with very high noise levels should probably be measured by the 3σ method or, perhaps, continue to be measured by the standby, half-height method. Worse reproducibility and accuracy should be expected from efficiency determinations when using these methods with noisy peaks.

Finally, it should be emphasized that the determination of column efficiency from signals acquired by a computer is precise and accurate only if proper programs are used. These programs should average out the noise contributions in the calculation of the retention times, the standard deviations and the peak moments. These programs should not simply mimic the behavior of the analyst working on the chromatogram. In that event they do not make good use of the computer possibilities and replicate human errors needlessly.

REFERENCES

- 1 B. L. Karger, H. Barth, E. Dallmeier, G. Courtois and E. Keller, *J. Chromatogr.*, 83 (1973) 289.
- 2 D. Ball, W. Harris and H. Habgood, *Anal. Chem.*, 40 (1968) 1113.
- 3 M. Delaney, *Analyst (London)*, 107 (1982) 606.
- 4 D. Ball, W. Harris and H. Habgood, *Anal. Chem.*, 40 (1968) 129.
- 5 D. Ball, W. Harris and H. Habgood, *Sep. Sci.*, 2 (1967) 81.
- 6 S. Chesler and S. Cram, *Anal. Chem.*, 43 (1971) 1922.
- 7 E. Grushka, M. Myers, P. Schettler and J. C. Giddings, *Anal. Chem.*, 41 (1969) 889.
- 8 B. Bidlingmeyer and F. V. Warren, Jr., *Anal. Chem.*, 56 (1984) 1583A.
- 9 M. Goedert and G. Guiochon, *Anal. Chem.*, 42 (1970) 962.
- 10 M. Goedert and G. Guiochon, *Chromatographia*, 6 (1973) 76.
- 11 M. Goedert and G. Guiochon, *J. Chromatogr. Sci.*, 2 (1973) 326.
- 12 B. Rodgers and J. Oberholtzer, *Anal. Chem.*, 41 (1969) 1234.
- 13 J. Foley and J. Dorsey, *Anal. Chem.*, 55 (1983) 730.
- 14 J. Foley and J. Dorsey, *J. Chromatogr. Sci.*, 22 (1973) 40.
- 15 M. Goedert and G. Guiochon, *Chromatographia*, 6 (1973) 39.
- 16 H. Yau, *Anal. Chem.*, 49 (1977) 395.
- 17 A. Anderson, T. Gibb and A. Littlewood, *J. Chromatogr. Sci.*, 8 (1970) 640.