Journal of Chromatography A, 1218 (2011) 293–302

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

The dimensionality of chromatographic separations

Mark R. Schure*

Theoretical Separation Science Laboratory, The Dow Chemical Company, 727 Norristown Road, Box 0904, Spring House, PA 19477-0904, USA

ARTICLE INFO

Article history:
Received 16 September 2010
Accepted 10 November 2010
Available online 18 November 2010

Keywords:
Liquid chromatography
Gas chromatography
Fractal measure
Peak distribution
Peak spacing
Chromatographic quality
Multidimensional separations
Selectivity
Optimization
Statistical overlap theory
Power law distributions
Orthogonality
Orthogonal separations

ABSTRACT

The box-counting or capacity dimension algorithm, known from the fractal mathematics literature, is used to measure the dimensionality D of chromatographic separation techniques for any number of dimensions. It is shown that D has limit properties that match Giddings' sample dimensionality s. D values are shown to be sensitive to the uniformity of peak spacing. A number of examples are given where D is calculated for various limits in one- and two-dimensional separations and for heart-cutting separations. The use of D as a quantitative measure of multidimensional orthogonality is suggested as D, due to the scale-free nature, is not dependent on the effective separation area. The connection to statistical peak overlap theory is discussed.

1. Introduction

It is often discussed that chromatograms which possess uniform close spacing of peaks would be most advantageous [1] as the number of detectable single component peaks, m, would be equal to the peak capacity n_c. This would place the maximum number of separable components in the minimum amount of separation space or time. We will confine this dialog for separations in time although the concept is general and can be utilized in space, for example with thin-layer chromatography in one or more dimensions. We will return to the distance–time duality occasionally in our discussion.

The peak capacity is defined as [2–4] the number of equispaced single component peaks that fit within a discrete time increment $t_1 \le t \le t_m$ so that

$$n_c = \frac{\Delta t}{4\sigma R_s} \tag{1}$$

where $\Delta t = t_m - t_1$ is the time increment, σ is the Gaussian standard deviation and R_s is the resolution by which peaks are to be resolved. When peaks are equispaced, the separation speed can be increased as the resolution can be lowered somewhat to accommodate more peaks. In real-world separations, this concept of an ordered chromatogram becomes increasingly difficult to implement as the saturation $\alpha = m/n_c$ of the separation increases.

The free energy distribution of solute transfer between mobile and retentive phase controls how efficiently one can order a chromatogram for multicomponent isocratic and isothermal separations. However, when one adds the ability to manipulate the free energy distribution in a chromatogram by temporal changes in temperature and/or in solvent composition, the chromatogram can be more optimally sorted to maximize the number of peaks per operating time. We refer to the modification in the free energy distribution as *selectivity* optimization in the exercise of modifying peak positions in a chromatogram.

A number of metrics have been proposed for evaluating chromatographic performance including the number of peaks, peak capacity, resolution, plate count, plates per unit time, peaks per unit time and more [5]. These metrics are very useful and often define the performance of the chromatogram. However none of these metrics will measure the uniformity of retention time spacing. One set of methods, based on Fourier techniques, [6–11] has been used to identify patterns of ordered peaks in one- and two-dimensional chromatography. These techniques use autocorrelation methods and use the whole peak profile but rely on a specific model to describe this ordering. These methods are not in wide-spread use but still constitute an interesting area of chromatographic research.

* Tel.: +1 215 641 7854; fax: +1 215 619 1616.
E-mail addresses: MSchure@Dow.com, MSchure@Rohmhaas.com

0021-9673/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2010.11.016

In this paper it will be shown how analysis of chromatographic retention times using the box-counting or capacity dimension method, which is often used to calculate the fractal dimension and power law scaling exponent, can be used to measure chromatographic uniformity and the "effective dimension" D of the separation. The dimension D will be shown to make a convenient metric for describing peak uniformity in 1, 2, 3 ... and in general, n-dimensional separations. D will be calculated using synthetic chromatograms and shown to agree with the sample dimensionality s of Giddings [1] at the limit points. Also discussed is the dimensionality of heart-cutting and the limit where $D \to 0$. Quantitative estimates within the limit points are shown for retention time distributions generated by a random (Poisson) peak placement method and a power law renewal process illustrating the sensitivity of D to peak ordering. It is suggested that D is highly useful as a measure of multidimensional peak orthogonality, as it is independent of the scale, area and angle of the multidimensional chromatogram. Examples of using D to characterize experimental measurements are given for one- and two-dimensional data. The mathematical connection of power law scaling to the statistical overlap theory [12–18] is briefly discussed.

2. Basis of the method

2.1. Fractal and power law scaling

An interesting set of tools have been developed in mathematics to characterize the fractal dimension [19–22] of self-similar objects. The connection to chromatographic peaks will be discussed below. These objects are unique as they look the same or have the same scaling of mass with distance on different length scales. Many of these objects have beautiful shapes and appear to be present in many biological systems [19–22]. The fractal dimension [19–22] is defined as:

$$D = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log 1/\varepsilon} \tag{2}$$

where $N(\varepsilon)$ is the number of objects (such as boxes and disks in two dimensions, and spheres and cubes in three dimensions) with object size ε, needed to cover some structure. D need not be an integer number; it describes how the mass or density of an object scales with the length scale that describes the object.

Many variations on this dimension exist when an emphasis is placed on measuring real-world objects and sampled-data temporal signals. These objects and signals are not necessarily self-similar, like most fractal objects, but like fractals possess power law scaling so that [21]:

$$M(\varepsilon) = c\varepsilon^{-D} \sim \varepsilon^{-D} \tag{3}$$

where c is a scaling constant, $\sim$ is read as "scales as," ε is a scale length, for example the length of a covering box, and $M(\varepsilon)$ is the measurement mass or density. Taking the logarithm of both sides of Eq. (3) gives $M(\varepsilon)$ equal to $N(\varepsilon)$ of Eq. (2). It will be shown in the section on estimating D (see Section 2.2 below) that N (and M) are the number of boxes containing at least one data point of the structure. The structure here refers to the retention time of a chromatographic peak.

Another way to state this in chromatographic terms is to state that the density of peaks, as a function of the time scale ε (or time increment ε), $M(\varepsilon)$, scales as ε^{-D}. This indicates that the inter-peak spacing is much more probable to be small (ε^{-D} monotonically decreases as ε increases), yet there is a small probability that the inter-peak spacing can be large, due to the power law (i.e. non-zero) tail.

Because we are not interested in c and only in D, this analysis is often referred to as being "scale invariant" [19–22]. It does not matter if the measurement scale is in seconds, hours, days, etc. This analysis is sensitive to the the ordering of events and how the events occur in time (or position) in a relative sense, but not on the absolute scale of the measurement.

Many natural phenomena, such as the relationship of metabolic rate q_0 and body mass M_b, have power law scaling [23]. In this case $q_0 \sim M_b^{3/4}$ which is obeyed over 27 orders of magnitude. Mammalian lifespan is known [23] to scale as $M_b^{1/4}$. In addition, the connectivity model of nodes on the internet was originally thought to be a Poisson process. This was later shown to be described by a scale-free power law [24]. Network traffic flow was long thought to be a Poisson process with an exponential waiting time density function. This has been found to be far more accurately characterized by a power law distribution [25].

The power law assumption [22,25,26] used in this paper differs significantly from the exponential waiting time distribution of Poisson processes [27–30]. Power law distributions have what is called a "heavy-tailed" distribution [22,25,26] that dies off much slower than an exponential distribution. The persistence of this distribution gives rise to many natural phenomena that are not captured with an exponential waiting time in nature.

One practical aspect of using this approach for measurements on objects and signals, rather than on mathematically constructed objects, such as fractal objects, is that there is a limit to the size of ε in the small limit $\varepsilon \to 0$. For point processes [30], where we accept the retention time as a marker or point, this limit is dictated by the smallest distance between points. For objects, this limit is established by the minimum object size. This is not a problem for objects constructed from formulas which describe fractal objects where self-similarity occurs on all length scales. However, this will be a problem for the determination of D from measurements, as discussed below.

2.2. Box-counting or capacity dimension

One of the most commonly used characterization algorithms to determine D is the so-called box-counting or capacity dimension [20,21,26]. Other algorithms which measure some form of the dimension include the correlation dimension [21,31–33] and the information dimension [21].

The box-counting dimension algorithm is easily implemented as follows. For data taken in one dimension the retention time data are first normalized to be in the range 0–1. This is accomplished by determining the minimum and maximum of the retention times, $t_{\min}$ and $t_{\max}$ respectively. The data is then normalized so that $t'_i = (t_i - t_{\min})/(t_{\max} - t_{\min})$ and the subscript i here runs from 1 to the total number of visible peaks. This approach is also used for n-dimensional data where we define n to be the number of retention axes, as was suggested by Giddings [1]. All data are normalized so that $0 \le t_{ij} \le 1$ in all dimensions where $1 \le j \le n$. A series of intervals is then constructed so that:

$$\varepsilon_i = \frac{t'_{\max} - t'_{\min}}{i} = \frac{1}{i} \tag{4}$$

where the index i here runs from 1 to the maximum number of box divisions. For one-dimensional chromatograms the number of intervals $N(\varepsilon_i)$ needed to cover the data with interval width ε_i is determined by counting. The data are then plotted as $\log N$ versus $\log \varepsilon_i$. The linear least-squares slope with the highest correlation coefficient within a range of the data gives the box-counting dimension times a factor of -1.

The least-squares analysis, depending on the retention time model (e.g. uniform, Poisson, power law, etc.) has a certain range of fitting. For true fractal models, this range is semi-infinite. However, for typical simulated and experimental data, the range of the linear portion of the least-squares analysis data is easily determined

by manual inspection. This will be automated in future work but is not critical in calculating the dimension at the medium-accuracy range requirement used in this paper. Accuracy of the box-counting method has been discussed in various contexts such as the mathematical basis [21] and in applications of image processing [34]. For one and two-dimensional point data, the box-counting algorithm works adequately as long as an adequate number of points is present. It is not clear where the limit is, as it is a function of D. Methods have been devised to correct the box-counting algorithm for finite sample effects [35] under a number of assumptions.

For n-dimensional data the approach is very similar. For $n = 2$, a series of boxes of varying length ε_i are applied to the planar data and the number of boxes $N(\varepsilon_i)$ that can cover the data is recorded as a function of ε_i. A plot of $\log N(\varepsilon_i)$ versus $\log \varepsilon_i$ gives a least-squares estimate of D again by multiplying the slope by -1. The same is true for three-dimensional data where cubes of equispaced sides ε_i are utilized with exactly the same procedure. This procedure applies to any n-dimensional separation with no limit on n although this procedure becomes computationally inefficient for $n \geq 3$. We will refer to the box-counting or capacity dimension simply as *dimension* for the duration of the text.

In all cases the data is divided into segments over the normalized range, as discussed above, and as shown in Eq. (4). In some cases, there is not enough data distributed throughout the range so that a plot of $\log N(\varepsilon_i)$ versus $\log \varepsilon_i$ will appear to be choppy. To remedy this problem, the value of i in Eq. (4) above is not incremented as $i = 1, 2, 3, \ldots, i_{\max}$ but rather a finite step value differing from 1 is often used so that $i = i_{\min}, (i_{\min} + i_{\text{step}}), (i_{\min} + 2 \times i_{\text{step}}), \ldots, i_{\max}$. This is determined by visual inspection of the plot of $\log N(\varepsilon_i)$ versus $\log \varepsilon_i$ and modified when this choppiness becomes apparent. In future work this will be automated so that the properties of the data will be used to determine i_{step} and $i_{\max}$. The correlation coefficient, r, is utilized as a goodness-of-fit least-squares metric throughout these analyses. This is necessary to describe how well power-law scaling fits the data. When comparing D values which vary by small amounts, further statistical analysis may be necessary and this will be discussed in a subsequent paper.

2.3. Synthetic chromatogram generation

Although our interest here is point process retention times, it is instructive to put concentration profiles on the retention time distribution and view these as chromatograms. Furthermore, with this procedure, as the saturation α increases, the number of detectable chromatographic peaks p will decrease due to peak fusion [12] as $p = m \exp(-\alpha)$ for random (Poisson) peaks. By adding zone broadening to the concentration profiles, the number of detectable peaks, as a function of dimensionality, can be investigated.

In all cases, unless indicated, the component peak heights are sampled from an exponential density function [15,36–39] as has been used in other peak simulations. This sampling can be done directly [40,41] as:

$$A_i = -\ln \xi_i \tag{5}$$

where ξ_i is a uniform random number on the interval $0 < \xi_i \leq 1$ and A_i is the amplitude of the ith component peak. The component peak heights are then normalized to $0 < A_i \leq 1$ by dividing all of the component peak heights by the largest component peak height.

2.3.1. Random (Poisson) peaks

For random peak generation in one dimension, peaks can be generated by producing m retention times so that $t_1 \leq t \leq t_m$. This is accomplished simply by generating random numbers ξ_i so that $t_i = t_{\min} + \xi_i(t_{\max} - t_{\min})$ and sorting so that $t_i \leq t_{i+1}$ realizing that the random numbers ξ_i used here are different than that used in Eq. (5). The resulting one-dimensional retention time vector $T_1 = (t_1, t_2, \ldots, t_m)$ is Poisson-distributed [42–44] with component density $\lambda = m/(t_m - t_1)$.

2.3.2. Generating power law peaks of known dimension

A more general procedure to create peaks of known D is to use a specific power law density function. These functions are known as renewal processes[27–29] and describe the probability of having some event occur, such as finding a peak, after some time t or distance x from the last peak.

One of the simplest power law renewal functions is that due to Lowen and Teich [26,45] and this approach will be used here although there are other functions which can also be utilized. This function has been referred to as a fractal renewal process noting that the types of point processes which are produced do not resemble the regular structural features of common fractal objects. However, these functions do have a certain self-similarity over a wide range of length scales, as will be shown in the results section.

The fractal renewal probability density function $P(t)$ has the form [26,45]:

$$P(t) = \frac{D}{A^{-D} - B^{-D}} t^{-(D+1)} \tag{6}$$

where A and B are the time interval lower and upper bound, respectively. A pure power law, like that given in Eq. (3), is unsuitable for a renewal function because the integration of this function over the range 0–∞ would give an infinite result. The form given in Eq. (6) is normalized when used in the interval $A \leq t \leq B$ with $P(t) = 0$ outside this interval.

The random interval spacings between peaks are obtained using the transformation method [40,41] to sample $P(t)$. This works as follows. The function $P(t)$ above is numerically integrated to give

$$F(y) = \int_A^y P(t)dt \tag{7}$$

Then the t value is found where $F(y) = \xi$ and as above, ξ is a uniform random number. This t value is a sample of the $P(t)$ density function. This procedure can be optimized for speed by making a fine-spaced table of y and $F(y)$ and linearly interpolating to get t. After each sample is obtained, it is added to the last sample to make the running sum which gives the retention time vector components t_i. The results are then rescaled to give chromatograms with specified minimum and maximum retention times. When using this procedure, one can specify A, B, and D, and obtain the retention times. D can be retrieved from both the raw and broadened retention times, as discussed below.

The form of Eq. (6) has been criticized [46] as not being a true renewal process because $P(t)$ is not continuous in the limit as $t \to 0$. This is characteristic of power laws where both the lower and upper bound must be defined or else the moments of these probability density functions can diverge [25,26,46] and would not be defined without the bounds. This can create an unphysical situation for use in chromatography because this would preclude two solutes of different structure arriving at the end of the column at exactly the same time and would also demand that all molecules elute within some maximum amount of time. However, this suffices for keeping a simple generator function form for peaks and the lower bound A can be made vanishingly small. More complicated forms for the renewal process function will be discussed in subsequent papers.

2.3.3. Peak broadening

Given the retention time vector T_1, for one-dimensional chromatographic data, we use a simple Gaussian model to produce peak

Table 1
Summary of one-dimensional chromatography results illustrating unique dimensional limits. Symbols and/or units are in parenthesis following column descriptors.

Spacing	Figure	Components (m)	Peaks[a] (p)	$^1\sigma$ (s)	Saturation[b] (α)	Dimension (D)
Ordered	1a	30	30 (100%)	2.50	0.300	1.00
Random (Poisson)	1b	100	67 (67%)	2.00	0.513	0.71
Random (Poisson)	1c	50	45 (90%)	2.00	0.256	0.69
Fractal renewal[c]	1e	100	92 (92%)	2.00	0.513	0.95[e]
Fractal renewal[d]	1f	50	24 (48%)	2.00	0.256	0.20[e]

[a] The number in parentheses are the fraction of visible peaks, p/m, expressed as a percentage.
[b] R_s in Eq. (1), used to calculate α, is set equal to 1.00 here.
[c] $A = 0.100$ and $B = 1.00$.
[d] $A = 0.100$ and $B = 300.0$.
[e] These D values are used as input. The estimated D is discussed in the text.

profiles as:

$$h(^1t) = \sum_{i=1}^{m} A_i \exp\left[-\frac{(^1t - t_i)^2}{2^1\sigma_i^2}\right] \tag{8}$$

where ^1t is the first dimension time, $^1\sigma_i$ is the Gaussian standard deviation of the ith peak and m is the number of components, as before.

For two-dimensional data ($n = 2$) bi-Gaussian peak profiles are produced with first dimension and second times ^1t and ^2t respectively and a retention time matrix

$$T_2 = \begin{pmatrix} t_{11} & t_{21} \\ t_{21} & t_{22} \\ \vdots & \vdots \\ t_{m1} & t_{m2} \end{pmatrix} \tag{9}$$

so that

$$h(^1t, ^2t) = \sum_{i=1}^{m} A_i \exp\left[-\frac{(^1t - t_{i1})^2}{2^1\sigma_i^2}\right] \exp\left[-\frac{(^2t - t_{i2})^2}{2^2\sigma_i^2}\right] \tag{10}$$

where ^2t is the second dimension time and $^1\sigma_i$ and $^2\sigma_i$ are the bi-Gaussian standard deviations in the first and second dimension respectively. In all cases in this paper, the standard deviations $^1\sigma_i$ in Eq. (8) are of equal value. Furthermore, $^1\sigma_i$ in Eq. (10) are of equal value and $^2\sigma_i$ in Eq. (10) are of equal value. Once broadening has been applied to the signal peak profile through by the use of Eq. (8) for 1D chromatograms and Eq. (10) for 2D chromatograms, retention times are found by taking the first derivative of the signal peak profile and determining where the negative-going zero-crossing occurs. These negative-going zero-crossings are utilized (under zero noise conditions) to determine the retention time vector used for dimensional analysis.

2.3.4. Experimental 1D peak analysis

Retention times from an experimental one-dimensional chromatogram are analyzed with in-house written software which detects the negative-going zero-crossing of the first derivative of the amplitude. The first derivative is obtained using a Savitsky–Golay [47–49] first derivative filter with 7-point width. Peaks are determined using an amplitude gate to eliminate the smallest peaks. This gate is adjusted by visual inspection of the detected peaks and eliminating some of the smallest peaks from consideration. The peaks selected for analysis will be shown by marking the selected peaks, as shown below.

3. Results

3.1. Limits in one dimension.

A summary of the results for model one-dimensional systems is given in Table 1 along with corresponding synthetic chromatograms and box-counting curve fits given in Fig. 1a–f. These results indicate that the dimension D of one-dimensional chromatograms is within the limits $D \leq 1$.

For the case where peak spacing is constant across the elution range, shown in Fig. 1a, $D = 1.00$. The least-squares curve fit is linear with negative one correlation coefficient. Although not shown, the dimension does not change if some peaks are omitted from the box-counting dimension calculation. In other words, with missing values the scaling does not change until the data gets sparse and the calculation becomes highly inaccurate.

Random chromatograms, examples of which are shown in Fig. 1b and c, use the random peak-placing Poisson process described above to generate peaks with two saturation levels. In the case of the higher saturation chromatogram, shown in Fig. 1b, 67 peaks are detected (67%), and 45 peaks are detected (90%) from the lower saturation chromatogram shown in Fig. 1c. The estimated dimension D, shown in Table 1 for both cases, are very similar. Two things are suggested with that result in mind. First, D appears to have some insensitivity as a measure from saturation effects. Increasing saturation causes more missing values. Second, from these and other simulations of random chromatograms, values of D have been obtained in the range $0.55 \lesssim D \lesssim 0.76$. These random chromatograms have been generated across the saturation range $0.2 \lesssim \alpha \lesssim 0.6$. This D value range may be characteristic of random peak spacings noting that Poisson processes are not power law distributed.

The curve-fitting least-squares procedures for these two chromatograms are shown in Fig. 1d. In both cases there is a limiting fitting region, as expected. Because the random peak placement does not possess power law scaling, the D values are anomalous, however, the box-counting algorithm gives an idea what the limiting behavior is for these chromatograms. Both datasets have correlation coefficients $r \leq -0.90$. The scatter in the data at the smaller box size is due to the data not being finely spaced enough to have measure significance at that degree of spacing. That data, as the line suggests, is not included within the least-squares analysis.

Two examples of using the fractal renewal process are shown in Fig. 1e and f for high and low-dimensionality chromatograms, respectively. The calculation conditions for these are given in Table 1. Using the box-counting algorithm to analyze the detected retention times produced by Eq. (6), after zone broadening, gives a D value of 0.958 with correlation coefficient r of −0.999. This shows that for a high D case, the specified D is recovered well here.

The high dimensionality chromatogram has 92 peaks that are detected out of 100 (i.e. 92%). This is comparable with the Poisson peaks in Fig. 1c where 90% of the peaks are detected except

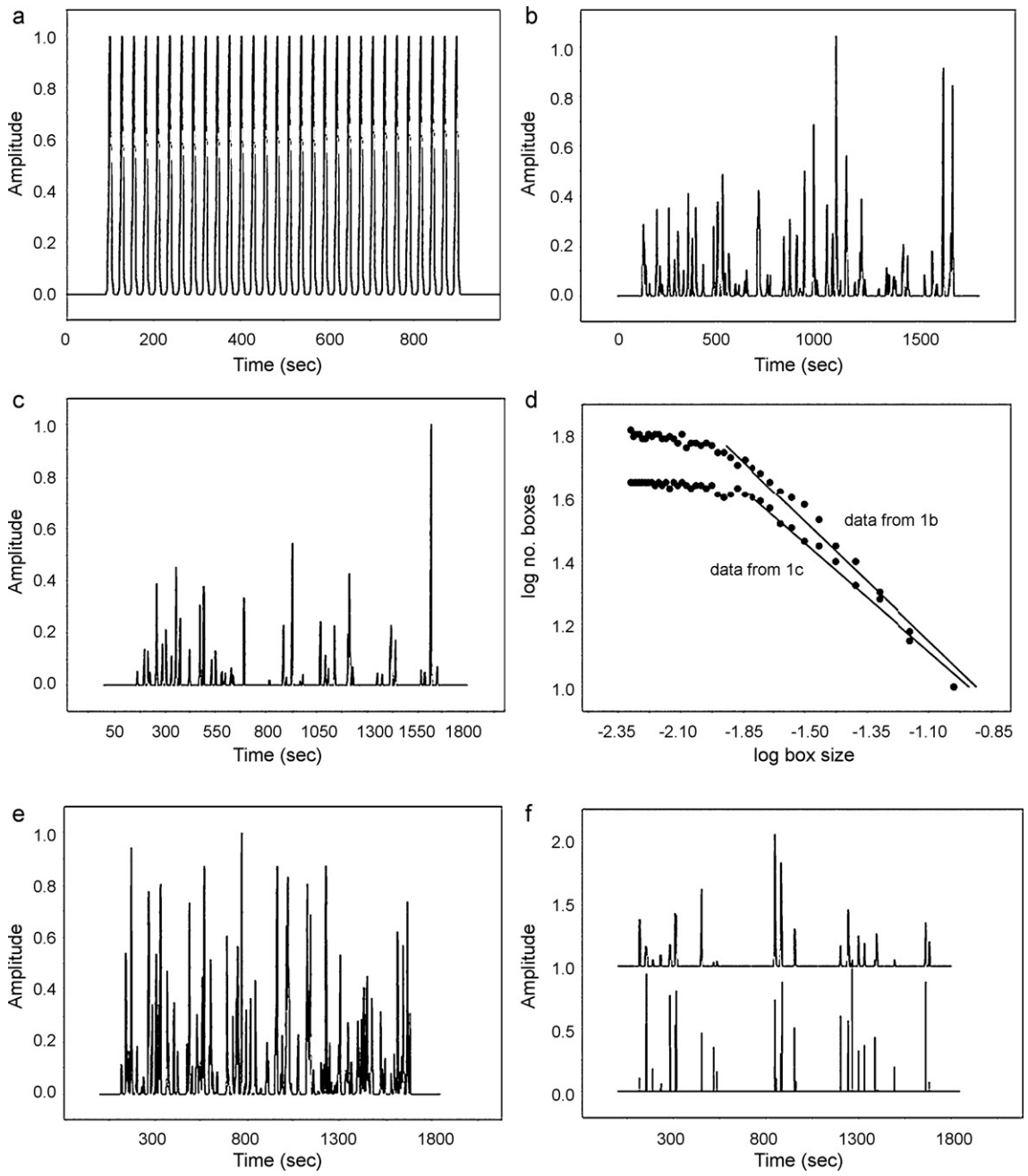

Fig. 1. Synthetic chromatograms used for dimensionality estimation for one-dimensional chromatography. Conditions given in Table 1. (a) Ordered peak spacing. (b and c) Random (Poisson) chromatograms. (d) Box-counting least-squares fits for peaks shown in (b) and (c). (e) High-dimensionality chromatogram. (f) Low-dimensionality chromatogram.

that the random peaks have a saturation $\alpha = 0.256$ as compared to twice that of the high-dimensionality peaks. The high dimensionality chromatogram has more visible peaks at twice the saturation because the peaks are placed more efficiently than with random placement.

This is in contrast to the low-dimensional chromatogram shown in Fig. 1f. This figure shows the chromatogram before broadening (lower) and after broadening (upper). There are many peaks that are exceedingly close together that are lost as the broadening process is applied. In addition, a moderate number of baseline regions of wide duration exist. This type of chromatogram can be expected to be very susceptible to peak loss as broadening is increased due to the close spacing within a clump of peaks. The moderate number of baseline regions shown in Fig. 1f is a clear consequence of a heavy-tailed density function which will not be found in a random chromatogram.

The low-dimensional chromatogram shown in Fig. 1f yields 27 detectable peaks (48%) which is roughly one-half of the number of peaks from the random chromatogram of the same saturation. Clearly, higher D chromatograms display more peaks at constant saturation than chromatograms at lower D and the higher D chromatograms are more resistant to losing peaks (and hence information) from broadening processes than lower D chromatograms.

Estimating D, utilizing the broadened and detected low-dimensional chromatogram, yields estimates of D ranging from $\approx$0.3–0.4. Apparently, the number of detectable peaks, which is 24 in this case, is not enough to give a good estimate of D. However, using the unbroadened 50 retention times also gives poor

Table 2
Summary of two-dimensional chromatography results illustrating unique dimensional limits. Symbols are in parenthesis following column descriptors.

Spacing in each dimension	Figure	Components (m)	Dimension (D)
Uniform in both dimensions	2a	36	2.00
Uniform/on-diagonal	2b	36	1.00
Random/on-diagonal	2c	36	0.71
Random in both dimensions	2d	100	1.78

estimates of D. It may be that at low D values, the box-counting algorithm is quite inaccurate and may require substantially more peaks than that used here, perhaps hundreds of peaks. Nonetheless, the peaks generated by this procedure gives representive low-dimensional chromatograms for comparitive purposes. Because of a high probability of finding the unbroadened peaks, shown as the lower chromatogram of Fig. 1f, a very small distance away from each other, the 50 components are not easily distinguished in the unbroadened chromatogram. This extremely fine spacing appears to give the upper chromatogram, shown in Fig. 1f, with heights that appear to be out of proportion, in some cases, to the lower chromatogram.

3.2. Limits in two dimensions.

A number of results are summarized for two-dimensional chromatography in Table 2 and shown in Fig. 2. For a chromatogram completely ordered in both dimensions, as shown in Fig. 2a, the resulting dimension is calculated to be 2.00 with correlation coefficient $r = -1.00$. This is the most desirable chromatogram and would indicate a "perfectly orthogonal" separation that is most desired in two-dimensional chromatography. This is consistent with Giddings' concept of chemical dimension, as explained below.

The case of having a completely ordered chromatogram on the diagonal, shown in Fig. 2b, yields a dimension of 1.00, with correlation coefficient $r = -1.00$. This would be the case for a chromatogram with identical columns in both dimensions and identical solvent systems in both dimensions when a perfectly ordered chromatogram is present in one of the column systems. Obviously this is a case of a "completely non-orthogonal" separation which due to peak ordering would be most desired in a one-dimensional separation. In two dimensions it is a waste of effort.

The same situation, only for a random chromatogram in one dimension, yields the result shown in Fig. 2c with a dimension of $\approx$0.71, similar to a one-dimensional chromatogram. This result is also completely non-orthogonal and the same result could be obtained with one-dimensional chromatography. This highlights the result that the dimensional analysis gives essentially the same result, independent of the number of independent separation coordinates, when an n-dimensional separation gives a result that could be obtained in less than n dimensions.

For a two-dimensional random chromatogram with Poisson spacing in both dimensions, shown in Fig. 2d, $D \approx 1.78$. Although not shown, varying the saturation α has only a small effect on D for this case. Also not shown is the case where peaks below (or above) the diagonal are eliminated; this case gives markedly lower D values and will be discussed in another paper.

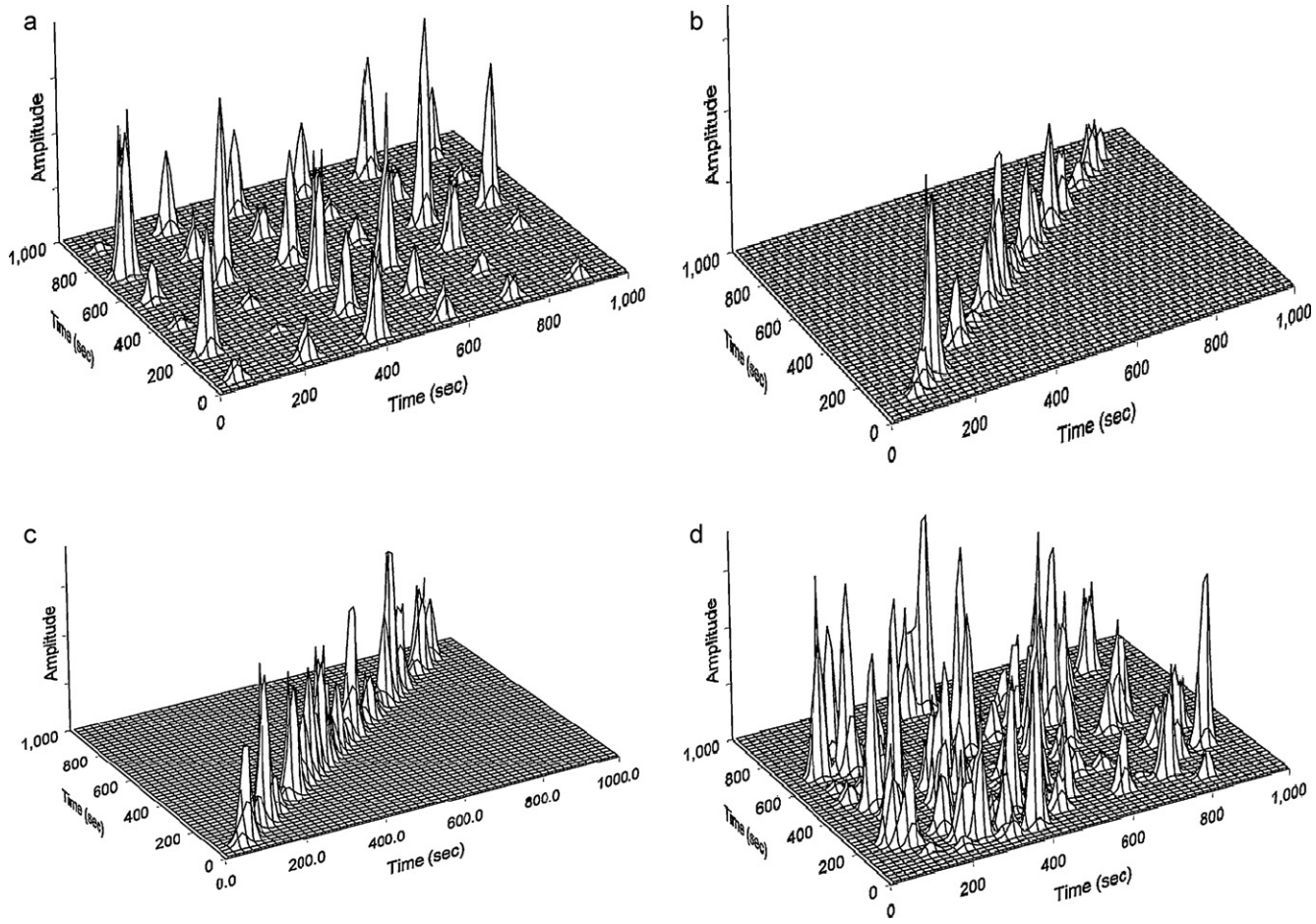

Fig. 2. Synthetic chromatograms used for dimensionality estimation for two-dimensional chromatography. The conditions are given in Table 2.

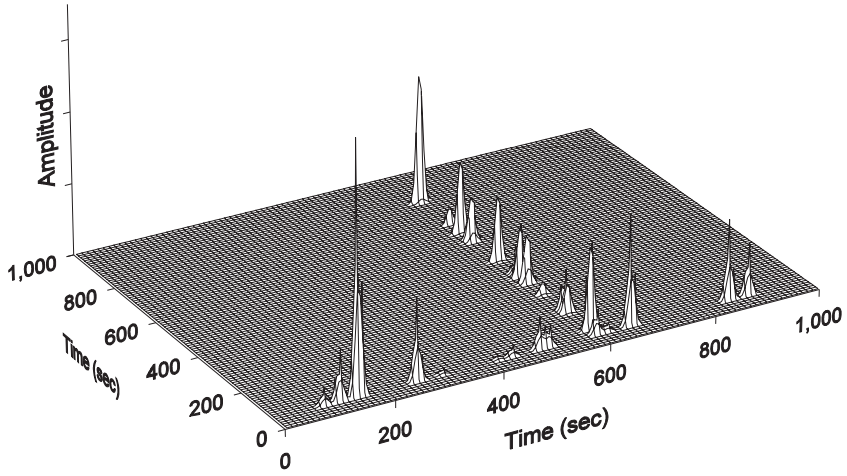

Fig. 3. A heart-cutting chromatogram portrayed as a two-dimensional chromatogram. Each retention axis has a random distribution of peaks. The data for this chromatogram is given in the first row of Table 3.

3.3. The dimensionality of heart-cutting

We have examined the dimensional limits on one and two-dimensional chromatographic systems. Here we examine the so-called "heart-cutting" mode, which takes one or more samples from the first dimension column and injects this on to a second (dimension) column. This is not done continuously, as in the previous examples which emulate comprehensive two-dimensional chromatography [50] where the second dimension separation system is synchronously injected with eluent from the first dimension column. Rather, one can think of this as having one or a few data row vectors (the second dimension column) added to a data column vector (the first dimension column). This can also be a representation of a very sparse two-dimensional chromatogram.

A typical example of a heart-cut chromatogram with random peak spacing in both dimensions is shown in Fig. 3. Using different random numbers, three two-dimensional chromatograms were generated for this example and the subsequent D values were found in the approximate range $1.14 \leq D \leq 1.26$. Other variations on this approach were tried including using ordered chromatograms in one or both of the dimensions and by having two second dimension chromatograms. These results are summarized in Table 3. These results are not rigorous as these ranges reflect the variation in the least-squares curve fits and the variation from using three chromatograms with different random number seeds and with different numbers of peaks for each of the cases studied in Table 3. Nonetheless, the trend shows how D reflects the spatial sparseness of these configurations. From the results presented previously, comprehensive two-dimensional chromatography has higher measured dimensionality than the heart-cutting technique, as expected.

Table 3
Summary of heart-cutting chromatography results illustrating unique dimensional limits.

First dimension spacing	Second dimension spacing	Dimension (D)
Random	Random[a]	≈1.14 – 1.26
Ordered	Random	≈1.19 – 1.23
Ordered	Ordered	≈1.25 – 1.30
Random	Ordered[b]	≈1.32 – 1.52
Random	Random[b]	≈1.41 – 1.59

[a] Shown in Fig. 3.
[b] There are two chromatograms in this dimension.

3.4. One peak limit

It is interesting to ask what the dimension should be for one peak. This can be solved easily with no calculation. With one peak, the box counting algorithm will return one box found, regardless of the box size and the number of boxes in the grid. When one plots the logarithm of the number of boxes needed to "cover" the one peak retention time, as a function of the logarithm of the box size, the resulting curve has a slope of zero. Hence the dimension of one peak, regardless of the number of instrumental dimensions n, is zero. This limit implies that one does not need a separation system to return such a result. It also suggests conveniently that one peak does not have any useful separation information in it. As D decreases, intuitively the complexity of the separation decreases to where the limit $D \rightarrow 0$ implies no separation.

3.5. One-dimensional separation of a tryptic digest

An experimental chromatogram of a peptide separation is presented as an example of estimating D for a one-dimensional separation. This separation, shown in Fig. 4a, is derived from yeast Enolase 1. The Swiss-Prot database identification number [51] is P00924. The sequence length is 437 amino acids. Chromatography was conducted with a Waters ACQUITY UPLC system using a 2.1 mm × 100 mm 1.7 μm particle diameter ACQUITY UPLC BEH300A Peptide Separation Technology column. Solvent conditions were A: 0.02% TFA in water and B: 0.016% TFA in acetonitrile. The gradient ranges from 0 to 50% B over 50 min with temperature 40 °C, and flow of 0.2 ml min^{-1}. The signal utilized for analysis here was the 214 nm signal from a Waters ACQUITY 2996 photodiode array detector.

Fig. 4a shows 57 peaks marked with a small box at the peak top. Analysis of the resulting retention times gives $D = 0.892$ with a correlation coefficient $r = -0.999$. The least-squares fit of the data is shown in Fig. 4b.

It is somewhat surprising that the peptide fragments show a high degree of order as the D value is much higher than random chromatograms, as discussed above. The chromatogram shows very good resolution of the higher amplitude peaks. For this sample, there appears to be very little "dead space" with almost no appreciable amount of baseline placed in random positions. Furthermore, there appears to be little peak fusion with successive retention times appearing away from adjacent peaks as would be expected for an ordered sample. Other one-dimensional exper-

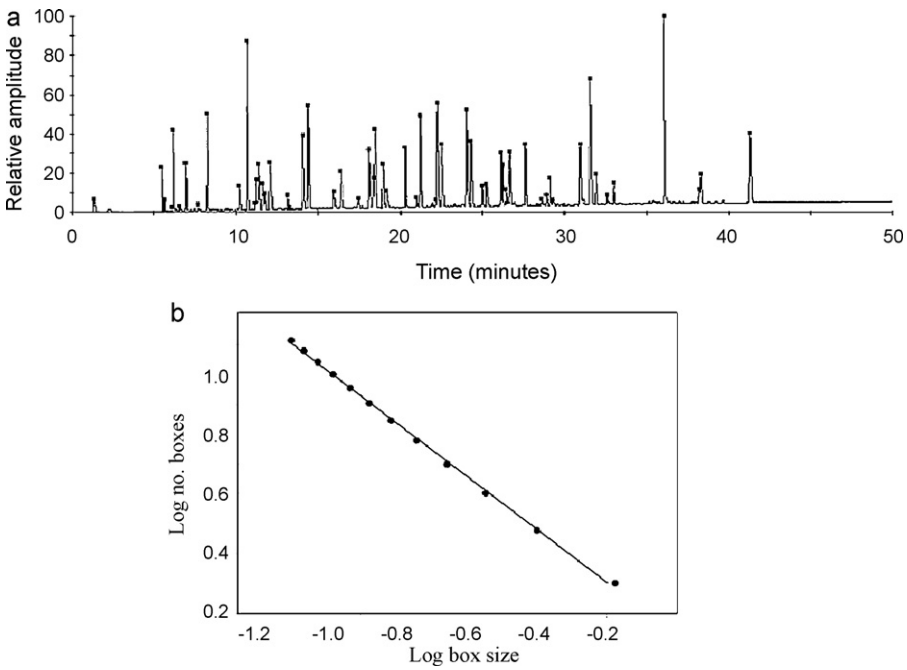

Fig. 4. (a) Experimental one-dimensional chromatogram of a tryptic digest as described in the text. The boxes above the peaks are those retention times chosen for analysis. (b) The least-squares analysis of the dimension of the chromatogram shown in Fig. 4a.

imental chromatograms that have been analyzed appear to be Poisson-like random chromatograms with the dimensions typical of those given above. Some of these were gas chromatograms of natural substances and others were gas chromatograms of alcoholic beverages. A number of peptide mapping separations appear to have ordered chromatograms [52] and this ordering could possibly be a consequence of protein architecture and function.

3.6. Two-dimensional separation of an industrial copolymer

A well-known example of a two-dimensional separation with nearly independent retention mechanisms is the separation of alcohol ethoxylates [53] by liquid chromatography. An experimental two-dimensional chromatogram used for D analysis is shown in Fig. 5a and the experimental conditions are stated in the original publication [53]. Retention times were obtained by measuring peaks manually from enlargements of the original figure.

The alcohol ethoxylates have an alkyl segment of length x and an ethylene oxide (EO) chain of length y. The distribution of polymer length in x and y is important from a performance standpoint. The alkyl chain separation is performed with reversed-phase liquid chromatography using a C_{18} stationary phase column and the EO group separation is performed with normal-phase chromatography. These separations appear to have little correlation between retention mechanisms. This means that the EO segment can be separated regardless of the alkyl chain length and the alkyl chain length separation is only slightly correlated with the length of the EO chain.

The dimension of this separation is $D = 1.85$ with a correlation coefficient of $r = -0.998$. The least-squares curve-fit is shown in Fig. 5b. The power law descriptor is valid for slightly less than one order of magnitude. This chromatogram is very ordered and shows a high dimension value. At full ordering D will be equal to 2, as stated previously.

The concept of orthogonality in multidimensional chromatography is simple yet powerful; it considers that the retention must be uncorrelated between dimensions to achieve the best separation. In that regard, in two-dimensional chromatographic separations, any useful measure of orthogonality must be sensitive to how "spread out" the chromatogram is with respect to peak positioning near the diagonal. This concept also applies for separations where $n \geq 2$ although for $n \geq 3$ these separations are difficult to visualize. Any clustering of the retention times along a diagonal must be avoided.

It appears that D may be a good measure of orthogonality as it is insensitive to the area in two dimensions that is to be measured. This is due to scale invariance discussed above. Previously developed orthogonality metrics [54–58] do not give simple numbers with easily understood limits. These methods typically do not have a dimensionally invariant basis; that is they apply to $n \geq 2$ but not to $n = 1$. This makes D an attractive method for comparison over a wide range of chromatographic experiments. As stated recently [57], the fraction of area covered by a two-dimensional chromatogram [56] is not a good measure of orthogonality; orthogonality has not been rigorously defined in multidimensional chromatography but the concept of having the chromatogram randomly occupy the line ($n = 1$), area ($n = 2$) and volume ($n = 3$) is intuitively satisfying. Even better is to have order in the multiple dimensions as exemplified by higher D values.

Clearly, two-dimensional separations, when peaks are clustered near the diagonal, are not adequately utilizing the separation space. The dimensionality measurement D appears to be a reasonably good metric of orthogonality as it is independent of the scale of the line, area and volume of $n = 1$–3 separations, respectively. This measure is independent of the angle of the retention times with the axes ($n \geq 2$) and is also somewhat independent of missing peaks, as discussed above.

One cannot imply separation speed and separation efficiency based on the measurement of D nor some other measure of orthogonality [57]. Clearly, dimensionality is a highly useful measure for selectivity optimization but not for speed optimization. Speed optimization must involve the component generating rate per dimension and the inherent zone width as parts of the overall speed optimization process. The inclusion of D and these later two scale-dependent metrics into an overall optimization strategy is the subject of current on-going research.

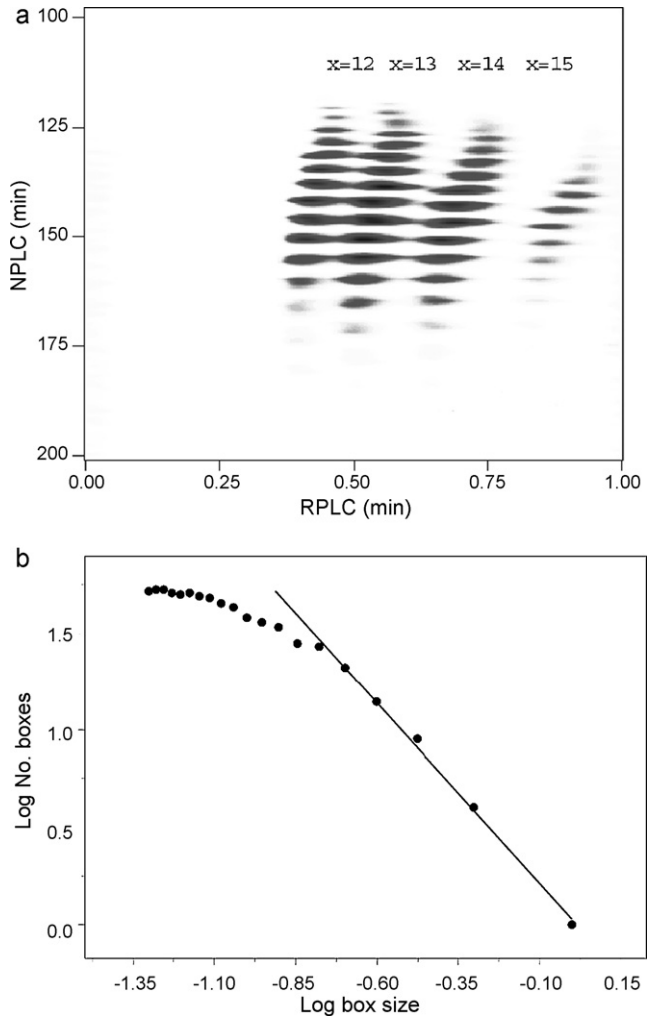

Fig. 5. (a) Experimental two-dimensional chromatogram of an alkyl ethoxylate showing the distribution of alkyl and polyethylene oxide segment lengths. Reprinted from reference 50. (b) The least-squares analysis of the dimension of the chromatogram shown in Fig. 5a.

3.7. Relationship with Giddings' chemical dimension

Giddings [1] suggested the concept that solutes would possess specific groups or ligands that could be separated by specific interaction with a stationary phase for multidimensional separations. The sample dimensionality s would be the number of specific retention mechanisms that could be utilized to specifically separate these molecules. Hence, it is expected that the experimental two-dimensional separation shown in Fig. 5a should have a dimension of $s \leq 2$ using Giddings' sample dimension s. This is in essential agreement with $D = 2$ from the synthetic ordered chromatogram shown in Fig. 2a and with the highly orthogonal alcohol ethoxylate chromatogram shown in Fig. 5a where $D = 1.85$. Although the Giddings sample dimension s is conceptual, it represents an upper bound to the box-counting dimension, i.e. $s \geq D$.

3.8. Statistical peak overlap theory with power law renewal process

The statistical theory of overlap [12–17], considers in the simplest form that the arrival times of component peaks are governed by a Poisson renewal process [27–30]. The Poisson process is based on exponential waiting times:

$$P(x) = \lambda e^{-\lambda x} \tag{11}$$

such that the probability $P(x)$ of finding the next peak some time t or distance x after the last peak, is an exponential waiting time governed by λ where λ is the component density or the number of components per total separation distance or time, as discussed above. Once $P(x)$ is defined, $P(x \geq x_0)$ is found by integration of $P(x)$ between x_0 and ∞, noting that x_0 is some value of distance or time whereby two peaks are still resolved at some stated resolution. The number of singlet peaks, doublets, triplets, ... is then determined by calculation from $P(x \geq x_0)$ and $P(x < x_0)$ and the total peak count summed up across all singlets, doublets, triplets ... via an infinite series to give the number of visible peaks, given the number of pure components. The details of this procedure are given in the original paper [12] and in a paper where other probability density functions are explored [18].

To accommodate the power law scaling used in this paper, Eq. (11) must be modified to accommodate a power law renewal process, such as that given in Eq. (6). There are a number of ways to do this and will be the subject of a separate paper from this laboratory. We will still assume for a power law renewal process that a constant density of peaks occurs across a finite sample of a chromatogram. This was assumed in the original peak overlap theory [12]. However, a constant peak density is not mandatory as an overlap theory has been developed [16] where a variable peak density, as a function of time, can be implemented.

3.9. Discussion

There are a number of investigations that this work has spawned. Due to length constraints and the extent of scope these will not be covered here. First, there are other methods besides the box-counting method that have unique dimensional information. These include the correlation dimension [21,31–33] and the information dimension [21], as mentioned above. Furthermore, wavelet methods [59,60] have been used to deliver higher accuracy estimators [61] than the simple box-counting procedure. These should be compared for this application. Maximum likelihood methods have also been mentioned in the literature [62] as a promising method for the direct estimation of the power law exponent and hence the dimensionality of the data. These need to be tried and contrasted with the box-counting dimension.

The dimensionality approach may have merit for the selectivity optimization of chromatograms in any number of dimensions. If one desires to approach the ideal equispaced chromatogram through selection of columns and optimization of gradient-elution conditions (LC) and temperature gradient conditions (GC), one should be able to measure D in an iterative methods development process.

We have undergone a preliminary investigation of the dimensionality of mass spectral data with the mass-to-charge (m/z) axis being a second separation dimension to the first chromatographic axis in an LC/MS experiment. This is an interesting case because depending on the type of ionization and the sample, the various time-based (mass) chromatograms, as a function of m/z, show lots of correlation across the various m/z ranges. One observation has been, in the case of gas chromatographic analysis of volatile fruit extracts, that the dimension is relatively constant across the m/z axis and these values of D suggest these mass chromatograms are of a Poisson type. More data needs to be taken over a wider class of materials to examine the mass spectrometry results. However, it is expected that some chromatograms will be Poisson in nature and some will obey power law scaling depending on the formation mechanism of the sample. Davis et al. [17] have examined the validity of a number of chromatograms and found

some to be Poisson in character and others did not obey Poisson statistics.

For very complex materials, such as a protein analysis of human serum analyzed by slab gel two-dimensional electrophoresis [63], the use of a single dimension may be greatly misleading. When one looks at this type of two-dimensional separation, for example the map of proteins from colorectal epithelial cells [64], the results are so complex that it is unlikely that one unique dimension would serve to characterize the separation. This has been recognized to occur in highly complex processes where it is thought that there is a distribution of dimensions within the process. To accommodate this requirement, multifractal analysis has been developed [21]. However, single-value dimension numbers appear to be highly useful as a measure of peak uniformity in one-dimensional chromatography and as a measure of orthogonality in two-dimensional chromatography. We expect the use of the dimensionality metric for categorizing separations to increase as this metric is further developed and accepted.

Acknowledgments

I thank Martin Gilar of the Waters Corp. for the use of the peptide chromatogram shown in Fig. 4a. Preliminary GC/MS experiments of volatile natural products were conducted by Francois Huby of the Springhouse Analytical Science group. Discussions with Peter Carr (University of Minnesota) and Joe Davis (Southern Illinois University) are greatly acknowledged.

References

[1] J.C. Giddings, J. Chromatogr. A 703 (1995) 3.
[2] J.C. Giddings, Anal. Chem. 39 (1967) 1027.
[3] E. Grushka, Anal. Chem. 42 (1970) 1142.
[4] E. Grushka, J. Chromatogr. 316 (1984) 81.
[5] P.J. Schoenmakers, Optimization of Chromatographic Selectivity: A Guide to Method Development, Elsevier Science, Amsterdam, 1988.
[6] A. Felinger, L. Pasti, F. Dondi, Anal. Chem. 62 (1990) 1846.
[7] A. Felinger, L. Pasti, F. Dondi, Anal. Chem. 64 (1992) 2164.
[8] F. Dondi, A. Bassi, A. Cavazzini, M.C. Pietrogrande, Anal. Chem. 70 (1998) 766.
[9] M.C. Pietrogrande, M.G. Zampolli, F. Dondi, Anal. Chem. 78 (2006) 2579.
[10] N. Campostrini, L.B. Areces, J. Rappsilber, M.C. Pietrogrande, F. Dondi, F. Pastorino, M. Ponzoni, P.G. Righetti, Proteomics 5 (2005) 2385.
[11] F. Dondi, M.C. Pietrogrande, N. Marchetti, A. Fellinger, in: S.A. Cohen, M.R. Schure (Eds.), Multidimensional Liquid Chromatography, John Wiley and Sons, New York, 2008, p. 59 (chap. 4).
[12] J.M. Davis, J.C. Giddings, Anal. Chem. 55 (1983) 418.
[13] M. Martin, G. Guiochon, Anal. Chem. 57 (1985) 289.
[14] M. Martin, D.P. Herman, G. Guiochon, Anal. Chem. 58 (1986) 2200.
[15] J.M. Davis, in: P.R. Brown, E. Grushka (Eds.), Adv. in Chromatogr., vol. 34, Marcel Dekker, New York, 1994, p. 109.
[16] J.M. Davis, J. Microcol. Sep. 3 (1997) 193.
[17] J.M. Davis, M. Pompe, C. Samuel, Anal. Chem. 72 (2000) 5700.
[18] M.C. Pietrogrande, F. Dondi, A. Felinger, J.M. Davis, Chemometr. Intell. Lab Syst. 5 (1995) 239.
[19] B.B. Mandelbrot, Fractals: Form Chance and Dimension, W. H. Freeman and Co., San Francisco, 1977.
[20] B.B. Mandelbrot, The Fractal Geometry Of Nature, W. H. Freeman and Co., New York, 1983.
[21] K. Falconer, Fractal Geometry: Mathematical Foundations and Applications, 2nd edition, John Wiley and Sons, New York, 2005.
[22] M. Schroeder, Fractals, Chaos Power Laws, Dover Publications, Mineola, New York, 2009.
[23] G.B. West, J.H. Brown, J. Exp. Biol. 208 (2005) 1575.
[24] B. Bellobás, O.M. Riordan, in: S. Bornholdt, H.G. Schuster (Eds.), Handbook of Graphs and Networks, Wiley-VCH, Weinheim, Germany, 2003, p. 1 (chap. 1).
[25] J. Gao, Y. Cao, W.-W. Tung, J. Hu, Multiscale Analysis of Complex Time Series, Wiley-Interscience, New York, 2007.
[26] S.B. Lowen, M.C. Teich, Fractal-based Point Processes, John Wiley and Sons, New York, 2005.
[27] D.R. Cox, Renewal Theory, Methuen Publishing, London, 1970.
[28] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 2, 2nd edition, John Wiley and Sons, New York, 1971.
[29] J.F.C. Kingman, Poisson Processes, Oxford, New York, 1993.
[30] D.R. Cox, V. Isham, Point Processes, Chapman and Hall/CRC Press, New York, 1992.
[31] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, 2nd edition, Cambridge University press, Cambridge, 2004.
[32] P. Grassberger, I. Procaccia, Phys. Rev. Lett. 50 (1983) 346.
[33] P. Grassberger, I. Procaccia, Physica D 9 (1983) 189.
[34] K. Foroutan-pour, P. Dutilleul, D.L. Smith, Appl. Math. Comput. 105 (1999) 195.
[35] S. Borgani, G. Murante, Phys. Rev. E 49 (1994) 4907.
[36] L.J. Nagels, W.L. Creten, P.M. Vanpeperstraete, Anal. Chem. 55 (1983) 216.
[37] F. Dondi, Y.D.A. Kahie, G. Lodi, M. Remelli, P. Reschiglian, C. Bighi, Anal. Chim. Acta 191 (1986) 261.
[38] L.J. Nagels, W.L. Creten, Anal. Chem. 57 (1985) 2706.
[39] M.Z.E. Fallah, M. Martin, Chromatographia 24 (1987) 115.
[40] R.W. Hamming, Numerical Methods for Scientists and Engineers, 2nd edition, McGraw-Hill, New York, 1973.
[41] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in FORTRAN, 2nd edition, Cambridge University Press, New York, 1992.
[42] S. Ross, Stochastic Processes, 2nd edition, Wiley, New York, 1995.
[43] K. Rowe, J.M. Davis, Chemometr. Intell. Lab Syst. 38 (1997) 109.
[44] J.M. Davis, P.W. Carr, Anal. Chem. 81 (2009) 1198.
[45] S.B. Lowen, M.C. Teich, IEEE Trans. Inform. Theory 39 (1993) 1669.
[46] W.M. Lam, G.W. Wornell, IEEE Trans. Signal Process. 41 (1995) 2606.
[47] A. Savitsky, M.J.E. Golay, Anal. Chem. 36 (1984) 1627.
[48] J. Steinier, Y. Termonia, J. Deltour, Anal. Chem. 44 (1972) 1906.
[49] H.H. Madden, Anal. Chem. 50 (1978) 1383.
[50] R.E. Murphy, M.R. Schure, in: S.A. Cohen, M.R. Schure (Eds.), Multidimensional Liquid Chromatography, John Wiley and Sons, New York, 2008, p. 93 (chap. 5).
[51] The swiss-prot protein knowledgebase for p00924, http://www.uniprot.org/uniprot/P00924.
[52] M. Gilar, personal communication, 2010.
[53] R.E. Murphy, M.R. Schure, J.P. Foley, Anal. Chem. 70 (1998) 4353.
[54] Z. Liu, D.G. Patterson, M.L. Lee, Anal. Chem. 67 (1995) 3840.
[55] P.J. Slonecker, X. Li, T.H. Ridgway, J.G. Dorsey, Anal. Chem. 68 (1996) 682.
[56] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, Anal. Chem. 77 (2005) 6426.
[57] N.E. Watson, J.M. Davis, R.E. Synovec, Anal. Chem. 79 (2007) 7924.
[58] M.R. Schure, in: S.A. Cohen, M.R. Schure (Eds.), Multidimensional Liquid Chromatography, John Wiley and Sons, New York, 2008, p. 11 (chap. 2).
[59] D.B. Percival, A.T. Walden, Wavelet Methods for Time Series Analysis, Cambridge University Press, New York, 2000.
[60] S. Mallat, A Wavelet Tour of Signal Processing, 2nd edition, Academic Press, New York, 1999.
[61] C. Yiyu, C. Minjun, W.J. Welsh, J. Chem. Inf. Comput. Sci. 43 (2003) 1959.
[62] H. Bauke, Eur. Phys. J. B 58 (2007) 167.
[63] J.E. Cellis, R. Bravo, Two-dimensional Gel Electrophoresis of Proteins, Academic Press, Inc., New York, 1984.
[64] The swiss-2dpage protein list for p07339 cec_human, http://expasy.org/swiss-2dpage/map=cec_human.