Section 1. Probabilistic aspects of the point-statistics SMO.

The probability that n adjacent SCs form a n-tet cluster peak ($n \geq 1$) is given[1,13] by:

$$P_n(\alpha) = F^2(\alpha)[1 - F(\alpha)]^{n-1} \tag{1s}$$

with the properties:

$$\sum_{n=1}^{\infty} P_n(\alpha) = F(\alpha) \tag{2s}$$

and:

$$\sum_{n=1}^{\infty} nP_n(\alpha) = 1 \tag{3s}$$

n is called the peak multiplicity value. Eqs. 2s and 3s (as well as others in the following) were evaluated by using mathematical symbolic package (Mathematica® 2.2.1 for Windows[22]). From Eq. 3s one can write:

$$\sum_{n=1}^{\infty} mnP_n(\alpha) = m \tag{4s}$$

and $P_n(\alpha)$ is thus the following limiting frequency[1]:

$$P_n(\alpha) = \lim_{m \to \infty} \left(\frac{N[n]}{m} \right) \tag{5s}$$

where $N[n]$ is the number of peaks of multiplicity n in a chromatogram of m SCs. Note that the limit to an infinitely great number of SCs means considering an infinitely long chromatogram. However, in a finite length piece of this infinitely long chromatogram, m is a random quantity[1] because the SC sequence is determined by the IM defined by Eqs. 5 and 6 of Principal Paper (equation numbers without "s" refer to Principal Paper). Thus, the SMO theory essentially gives expected quantities, i.e. mean quantities. In this instance Eq. 5s can be rewritten as:

$$P_n(\alpha) = \frac{\overline{N[n]}}{\overline{m}}$$ (6s)

In Eq. 6s, the average quantities are computed over a set of repeated finite length chromatograms all having the same statistical properties, i.e. the same IM and the same value of α. The total number of peaks in the chromatogram, p, is thus:

$$p = \sum_{n=1}^{m} N[n]$$ (7s)

The sum is here extended to m since this is the theoretical maximum multiplicity.

For the subsequent treatment it can prove useful to interpret the peak appearance within the chromatogram as an event associated with its outcome, peak multiplicity, and to introduce its probability distribution. The event of n multiplicity value appearance is represented as $\{M_n\}$. One is interested in the peak multiplicity probability, $\Pr\{M_n\}$, i.e. the limiting frequency of the number of peaks of given multiplicity, $N[n]$, in reference to the total number of peaks p, i.e.:

$$\Pr\{M_n\} = \lim_{m \to \infty}\left(\frac{N[n]}{p}\right)$$ (8s)

Note that the limit in Eq. 8s means considering an infinitely long chromatogram. From Eqs. 6s-8s one has:

$$\Pr\{M_n\} = \frac{P_n(\alpha)}{\sum_{i=1}^{\infty} P_i(\alpha)}$$ (9s)

and by introducing Eqs. 1s and 2s into Eq. 9s one has:

$$\Pr\{M_n\} = F(\alpha)[1 - F(\alpha)]^{n-1}$$ (10s)

Moreover since:

$$\sum_{n=1}^{\infty} F(\alpha)[1-F(\alpha)]^{n-1} = 1 \qquad \text{(11s)}$$

one has:

$$\sum_{n=1}^{\infty} \Pr\{M_n\} = 1 \qquad \text{(12s)}$$

According to Eq. 12s, the collection of the events $\{M_1\}$, $\{M_2\}$, ...$\{M_n\}$ have the important property of being exhaustive. Moreover, they are also mutually exclusive. In fact, every peak in the chromatogram will belong to one, and only one, of the $\{M_n\}$ multiplicity cases. In terms of Probability Theory, $\{M_1, M_2, ..., M_n\}$ constitutes a "complete and countable system of mutually exclusive sets of events" and is, thus, a "partition"[1s]. This is also expressed as:

$$S = \uplus M_i \qquad \text{(13s)}$$

where S means the sure event and $\uplus$ means the union of mutually exclusive events $\{M_i\}$. These properties are a basic requirement for subsequent development. The point is important since this provides the basis for building the quantitative SMO model.

It is of interest to compute the statistics of the multiplicity value, n. By considering Eq. 10s, $\Pr\{M_n\}$ is recognized as a geometric type distribution[1s]:

$$r_s = z(1-z)^s \qquad s = 0,1,2.....\infty \qquad \text{(14s)}$$

with: ⁻

$$z = F(\alpha) \qquad \text{(15s)}$$

$$s = n-1 \qquad \text{(16s)}$$

where z is the distribution parameter ($0 \le z \le 1$).

The mean and variance, defined as:

$$\bar{n} = \sum n \Pr\{M_n\} \qquad \text{(17s)}$$

4

$$\sigma_n^2 = \sum (n - \bar{n})^2 \Pr\{M_n\} \tag{18s}$$

are[14], respectively:

$$\bar{n} = \frac{1}{F(\alpha)} \tag{19s}$$

$$\sigma_n^2 = \frac{1 - F(\alpha)}{F^2(\alpha)} \tag{20s}$$

By combining Eqs. 4s, 5s, 8s, and 17s one can show that the meaning of $\bar{n}$ is:

$$\bar{n} = \lim_{m \to \infty} \left(\frac{m}{p} \right) \tag{21s}$$

Eq. 21s can be rewritten as:

$$\bar{n} = \frac{\bar{m}}{\bar{p}} \tag{22s}$$

where $\bar{p}$ is the expected number of peaks or the set average quantity, where the set is made by repeated finite length chromatograms.

Section 2. Probabilistic aspects of the pulse-point statistics SMO.

Let us first derive the amplitude frequency function, $g_n(y)$, of multiplet peaks of the same n multiplicity value. The amplitude frequency function of the singlet peaks, $g_1(y)$, will be simply:

$$g_1(y) = g(y) \tag{23s}$$

The amplitude frequency function of doublet peaks is a case of addition of identically distributed, independent random variables and is given by the convolution products of the frequency functions[16]:

$$g_2(y) = \int_0^y g(x)g(y - x)dx \tag{24s}$$

This convolution integral will be symbolically written as:

$$g_2(y) = g(y) * g(y) = g^{2*}(y) \tag{25s}$$

5

In general, for an n-tet peak, there will be:

$$g_n(y) = g^{n*}(y) \tag{26s}$$

which is related to the probability, $\Pr\{A(y)\}$, of observing a peak within the range between y and $y + dy$, given a multiplicity value equal to n, $\{M_n\}$:

$$\Pr\{A(y)\| M_n\} = g^{n*}(y)dy \tag{27s}$$

where $\Pr\{A(y)\| M_n\}$ is the conditional probability.

In the observed chromatogram a given amplitude value equal to y can correspond either to a singlet or a multiplet peak. The problem of determining its frequency is handled by rigorous probabilistic methods. This well founded approach is suitable for any future development of the theory which aims at deriving the more complex chromatogram attributes such as determination limits[9-12]. The unconditional event of peak amplitude occurrence of a value between y and $y+dy$, $\{A(y)\}$, has an occurrence probability defined as follows:

$$\Pr\{A(y)\} = g_{obs}(y)dy \tag{28s}$$

$\{A(y)\}$, implies $(\subset)$ one of the multiplicities M_i; i.e. one can write:

$$A(y) \subset \cup M_i \tag{29s}$$

In terms of set theory, then, since $\{M_1, M_2,... M_n\}$ is a "partition" (see Eq. 13s) and because of Eq. 29s, the following equation can be written (see Theorem P8 in ref. 1S):

$$\Pr\{A(y)\} = \sum_{n=1}^{\infty} \Pr\{A(y)M_n\} \tag{30s}$$

where $\{A(y)M_n\}$ means the intersection of the two events, i.e. the simultaneous occurrence of both events $\{A(y)\}$ and $\{M_n\}$. According to the conditional probability rule [1S], one can set:

$$Pr\{A(y)|M_n\} = \frac{Pr\{A(y)M_n\}}{Pr\{M_n\}} \qquad (31s)$$

Thus, by combining Eqs. 30s and 31s, one has:

$$Pr\{A(y)\} = \sum_{n=1}^{\infty} Pr\{A(y)|M_n\} Pr\{M_n\} \qquad (32s)$$

By combining Eqs. 32s, 27s and 28s, one has:

$$g_{obs}(y) = \sum_{n=1}^{\infty} Pr\{M_n\} g^{n*}(y) \qquad (33s)$$

It must be underlined that $g_{obs}(y)$ depends on IM and α, because $Pr\{M_n\}$ is dependent on both IM and α (see Eq. 10s). Eq. 33s is the most general representation of SC peak overlapping on peak amplitude.

Section 3. Characteristic Function and Statistical Attributes of the observed peak amplitude distribution.

Eq. 33s is the most general representation of SC peak overlapping on peak amplitude. This type of distribution belongs to an important and well exploited class of distributions, the compound distributions[2s], representing a random sum of identically distributed independent random variables. The random variable added here is the SC amplitude, represented by its frequency function $g(y)$, $Pr\{M_n\}$ being the weighting factor for this process. For these distributions the so-called Characteristic Function (CF) — the Fourier Stiltjes Transform — can be generally expressed and from this both the statistical attributes and, in certain cases, the full expression of $g_{obs}(y)$ can be obtained [17,18,2s,3s]. The CF of a continuous frequency function, such as $g(y)$ expressing the type of AM, is defined as[1s]:

7

$$\Phi_{AM}(\xi) = \int\limits_{-\infty}^{+\infty} \exp(i\xi y)g(y)dy \qquad\qquad (34s)$$

and that of a discrete distribution, such as $\Pr\{M_n\}$ is expressed as:

$$\Phi_M(\xi) = \sum_{n=1}^{\infty} \exp(i\xi n)\Pr\{M_n\} \qquad\qquad (35s)$$

where i is the imaginary unit and ξ is the auxiliary variable. The CF of $g_{obs}(y)$ (see Eq. 33s) can be generally expressed by using a ln-exp transformation procedure reported elsewhere in handling similar cases of compound distribution[17,18]:

$$\Phi_{obs}(\xi) = \Phi_M\left[\frac{\ln\Phi_{AM}(\xi)}{i}\right] \qquad\qquad (36s)$$

where Φ_M and Φ_{AM} are the CFs of the multiplicity distribution and of the AM, respectively, defined in Eqs. 34s and 35s. The distinct feature in Eq. 36s is that the argument of Φ_M is the complex function $\ln\Phi_{AM}(\xi)/i$. By using the relationship between the derivatives in $\xi=0$ of CF and the moments of the corresponding frequency function, the statistical attributes of $g_{obs}(y)$ can be obtained[17,18]. The result is:

$$\bar{y}_{obs} = \bar{n}\,\bar{y} \qquad\qquad \{IM = all, AM = all\} \qquad (37s)$$

for the mean peak amplitude value in the chromatogram. As one can see, $\bar{y}_{obs}$ is simply the product of the mean multiplicity value, $\bar{n}$, and the SC mean area value, $\bar{y}$. $\{IM = all, AM = all\}$ specifies that this expression holds for all the IM and AM types. Likewise, by applying the same procedure for obtaining variance, one has:

$$\sigma_{obs}^2 = \sigma_n^2\bar{y}^2 + \bar{n}\sigma_y^2 \qquad\qquad \{IM = all, AM = all\} \qquad (38s)$$

where σ_y^2 is the variance of $g(y)$.

The α dependence of either $g_{obs}(y)$ or its statistical attributes can be singled out if one assumes that the subsequent SC positions can be described by a continuous IM function. First of all the CF of $\Pr\{M_n\}$ can be computed by noting that it has a shifted geometric distribution, the shifting being equal to $+1$ (see Eqs. 14s-16s). The CF of $\Pr\{M_n\}$ is thus that of the geometric distribution[14] multiplied by $\exp(i\xi)$, the latter factor accounting for the shifting effect [3s]:

$$\Phi_M(\xi) = \frac{F(\alpha)\exp(i\xi)}{\left[1-(1-F(\alpha))\exp(i\xi)\right]} \qquad (39s)$$

By combining Eq. 36s and 39s, one obtains the general CF expression for the peak amplitude distribution:

$$\Phi_{obs}(\xi) = \frac{F(\alpha)\Phi_{AM}(\xi)}{\left[1-(1-F(\alpha))\Phi_{AM}(\xi)\right]} \qquad \{IM = all,\ AM = all\} \qquad (40s)$$

holding for any type of IM and AM, provided that they are independent of one another. Eq. 40s will be exploited later on for specific AM and IM types.

The statistical attributes of $g_{obs}(y)$ are obtained by introducing Eqs. 19s and 20s into Eqs. 37s and 38s:

$$\bar{y}_{obs} = \frac{\bar{y}}{F(\alpha)} \qquad \{IM = all,\ AM = all\} \qquad (41s)$$

$$\left(\frac{\sigma_{obs}}{\bar{y}_{obs}}\right)^2 = 1 + F(\alpha)\left[\left(\frac{\sigma_y}{\bar{y}}\right)^2 - 1\right] \qquad \{IM = all,\ AM = all\} \qquad (42s)$$

where the dependence on α is singled out.

Section 4. Model Exploitation: Multicomponent Chromatograms with Poissonian Retention Pattern.

In the case of Poissonian Retention Pattern, the frequency function of the interdistance is:

$$f(x) = \frac{1}{T} \exp\left(-\frac{x}{T}\right) \qquad\qquad \{IM = E\} \qquad\qquad (43s)$$

and from Eqs. 5 and 7 one has:

$$F(\alpha) = \exp(-\alpha) \qquad\qquad \{IM = E\} \qquad\qquad (44s)$$

Under such conditions Eq. 41s becomes:

$$\bar{y}_{obs} = \bar{y} \exp(\alpha) \qquad\qquad \{IM = E, \ AM = all\} \qquad\qquad (45s)$$

or, using Eqs. 2 and 4 and taking the logarithm, one has

$$\ln(\bar{y}_{obs}) = \ln(\bar{y}) + \bar{m}\left(\frac{x_0}{X}\right) \qquad\qquad \{IM = E, \ AM = all\} \qquad\qquad (46s)$$

where m was replaced by its average value $\bar{m}$.

The dependence of peak amplitude relative dispersion, $(\sigma_{obs}/\bar{y}_{obs})$, on both α and $(\sigma_y/\bar{y})$ is given by Eq. 42s. This is made explicit in Table I for a series of AM types: the Exponential (E), Uniform (U), Constant (C) and Mixed Exponential (ME), all under $\{IM = E\}$. Provided $\sigma_y / \bar{y} \neq 1$, i.e. with the exclusion of $\{AM = E\}$, by introducing Eqs. 2, 4 and 44s and taking the logarithm, Eq. 42s can be rewritten as:

$$\ln\left[1 - \left(\frac{\sigma_{obs}}{\bar{y}_{obs}}\right)^2\right] = \ln\left[1 - \left(\frac{\sigma_y}{\bar{y}}\right)^2\right] - \bar{m}\frac{x_0}{X} \qquad \{IM = E, \ AM: \ \sigma_y / \bar{y} < 1\} \quad (47s)$$

$$\ln\left[\left(\frac{\sigma_{obs}}{\bar{y}_{obs}}\right)^2 - 1\right] = \ln\left[\left(\frac{\sigma_y}{\bar{y}}\right)^2 - 1\right] - \bar{m}\frac{x_0}{X} \qquad \{IM = E, \ AM: \ \sigma_y / \bar{y} > 1\} \quad (48s)$$

Section 5. Model Exploitation: Multicomponent Chromatograms with Exponentially distributed SC amplitudes.

The case of the multicomponent chromatogram where the $\{AM = E\}$ is fully exploited under both general and specific IM type conditions, by using the Characteristic Function method.

10

In this case the amplitude frequency function is:

$$g(y) = \frac{1}{\bar{y}} \exp\left(-\frac{y}{\bar{y}}\right) \qquad \{AM = E\} \qquad (49s)$$

The peak area distributions within the chromatogram, $g_{obs}(y)$, is obtained here by using the CF approach outlined above (see Eqs 36s, 39s, 40s). The CF for the exponential distribution (Eq. 49s) is[3s]:

$$\Phi_{AM}(\xi) = \frac{1}{1 - i\bar{y}\xi} \qquad \{AM = E\} \qquad (50s)$$

By introducing Eq. 50s into Eq. 40s, the CF of $g_{obs}(y)$ is obtained:

$$\Phi_{obs}(\xi) = \frac{1}{1 - \left(\dfrac{\bar{y}}{F(\alpha)}\right)i\xi} \qquad \{AM = E, IM = all\} \qquad (51s)$$

By introducing Eq. 41s into Eq. 51s one has:

$$\Phi_{obs}(\xi) = \frac{1}{1 - \bar{y}_{obs}i\xi} \qquad \{AM = E, IM = all\} \qquad (52s)$$

By comparing Eq. 52s with Eq. 50s one can identify $\Phi_{obs}(\xi)$ as a CF of an exponential type distribution and thus:

$$g_{obs}(y) = \frac{1}{\bar{y}_{obs}} \exp\left[-\frac{y}{\bar{y}_{obs}}\right] \qquad \{AM = E, IM = all\} \qquad (53s)$$

Section 6. Entropy Function Approach to pulse-point statistics SMO.

The entropy functions for $g_{obs}(y)$ and $g(y)$ will be respectively:

$$H(Y_{obs}) = -\int g_{obs}(y) \log_2 g_{obs}(y) dy \qquad (54s)$$

$$H(Y) = -\int g(y) \log_2 g(y) dy \qquad (55s)$$

Separation can be accounted for though the relative entropy function:

11

$$H_{rel}(Y_{obs}) = H(Y_{obs}) - H(Y) \qquad\qquad (56s)$$

i.e. by estimating the entropy function changes. This way of considering the separation process is close to what was presented by Martin and Guiochon[4s] who established a parallel between separation and depolymerization. Evaluation of $H_{rel}(Y_{obs})$ requires that both $g_{obs}(y)$ and $g(y)$ be expressed in analytical form and the integrals of Eqs. 54s and 55s be evaluated.

The entropy function for the exponential SC amplitude distribution, $\{AM = E\}$, is[19]:

$$H(Y) = \ln e\bar{y} \qquad \{AM = E\} \qquad (57s)$$

where e is the natural logarithm base. Moreover, one important property of E distribution is that it is the maximum entropy distribution under the positive constraint value for the variable ($y>0$) and for the distribution mean constant[19]. In practical terms one can conceive that for a series of mixtures, all having the same number of components and with the same mean concentration, $\bar{y}$, those having components exponentially distributed in the concentration can be arranged in the highest number of ways, i.e. the E distribution is the most likely one. Under such conditions, the exponential amplitude distribution is most suitable for theoretical investigation since it appears to be the most common, just as observed in practical experience (Refs 9 and 23). In addition, the point emphasizes the interest in studying the entropy function properties. It has been observed that if $g(y)$ is of type E, $g_{obs}(y)$ remains the same under all α separation conditions and for all IM type cases (see Eq. 53s) and thus its entropy function is:

$$H_{obs}(Y) = \ln e\bar{y}_{obs} \qquad \{AM = E, IM = all\} \qquad (58s)$$

It is remarkable that separation follows a condition of maximum entropy function for the peak amplitude distribution.

By considering Eqs. 41s, 53s and 57s, one has:

$$H(Y_{obs}) = \ln e\bar{y} - \ln F(\alpha) \qquad \{AM = E, IM = all\} \qquad (59s)$$

and thus by introducing Eqs. 57s and 59s into Eq. 56s and further introducing Eqs. 29s and 22s, one has:

$$H_{rel}(Y_{obs}) = \ln \bar{n} = \ln \bar{m} - \ln \bar{p} \qquad \{AM = E, IM = all\} \qquad (60s)$$

for an infinitely long chromatogram, i.e. for $m \to \infty$. The relative entropy is thus a positive quantity related to the logarithmic difference between components, $\bar{m}$, and peaks, $\bar{p}$.

By combining Eq. 56s, 57s, 59s and 44s, the relative entropy function under $\{IM = E\}$:

$$H_{rel}(Y_{obs}) = \alpha \qquad \{AM = E, IM = E\} \qquad (61s)$$

By introducing Eqs. 2 and 4 one has:

$$H(Y_{obs}) = \ln e + \ln \bar{y} + \bar{m}\frac{x_0}{X} \qquad \{AM = E, IM = E\} \qquad (62s)$$

Eq. 61s establishes a meaning to α.

Section 7. Equivalence between the Davis-Giddings and the present approach for determining the number of Single Components.

The present approach base on Eq. 46s for determining the number of SCs is similar to that previously proposed by Davis and Giddings[1,3] based on the peak number:

$$\ln \bar{p} = \ln \bar{m} - \bar{m}\left(\frac{x_0}{X}\right) \qquad \{IM=E\} \qquad (63s)$$

In fact, Eq. 46s and Eq 63s can be derived from one another since both $\bar{y}_{obs}$ and p, and $\bar{y}$ and $\bar{m}$ are interrelated throughout the total area y_{tot} of the chromatogram:

$$y_{tot} = \sum_{i=1}^{p} y_{i,obs} \qquad (64s)$$

$$y_{tot} = \sum_{i=1}^{\bar{m}} y_i \qquad (65s)$$

$$\bar{y}_{obs} = y_{tot} / p \qquad\qquad (66s)$$

$$\bar{y} = y_{tot} / \bar{m} \qquad\qquad (67s)$$

and thus it is understood that Eq. 63s must hold true for all the AM types. The i index in Eqs. 63s and 65s refer to individual SCs or to individual peaks in the chromatogram. Note that the mean symbol refers to expected quantities as discussed above.

Any deviation in the logarithmic plot of Eq. 63s, when applied to a real chromatogram, will only be due to discrepancies (random or systematic) between the true interdistance distribution in the real multicomponent chromatogram and the assumed exponential IM of the Davis-Giddings statistics. The point is even more clearly elucidated by introducing Eqs. 56s and 59s into Eq. 12:

$$f(x) = -\frac{1}{\bar{m}} \frac{d(\bar{p})}{dx} \qquad \{IM=all, AM=all\} \qquad (68s)$$

where the dependence of the peak number on x is unambiguously referred to the IM type under most general conditions. Eq. 68s corresponds to Eq. 12 previously derived by using the pulse-point statistics SMO. Because of this close correspondence, Eq. 68s holds under the same general conditions of $\{IM=all, AM=all\}$.

14

ADDITIONAL GLOSSARY FOR SUPPORTING INFORMATION

$A(y)$	peak area of value equal to y
$\mathbf{i}$	imaginary unit
M	peak multiplicity
n	value of the peak multiplicity
$N[\ ]$	Number of cases of type $[\ ]$
$P(\alpha)$	Probability of the n-tet cluster as a function of α
r_s	geometric distribution in the s variable
S	sure event
z	parameter ($0 < z < 1$)
ξ	auxiliary variable
Φ	Characteristic Function
$\cup$	event union
$\subset$	inclusion ($A \subset B$, event A implies event B)
$\mid$	conditioning
$*$	symbol for convolution integral

ADDITIONAL LITERATURE CITED FOR SUPPORTING INFORMATION

Bibliography numbers without "s" refer to Principal Paper.

[1s] Pfeiffer, P. E. *Concept of Probability Theory*, 2nd revised ed.; Dover Publications: New York, 1978.

[2s] Cramér, H. *Random Variables and Probability Distributions*; Cambridge University Press: Cambridge, 1970.

[3s] Feller, W. *An Introduction to Probability Theory and its Applications*, Vol. II; J. Wiley and Sons: New York, 1966.

[4s] Martin, M.; Guiochon, G. *Anal. Chem.*, **1985**, *57*, 289-295.
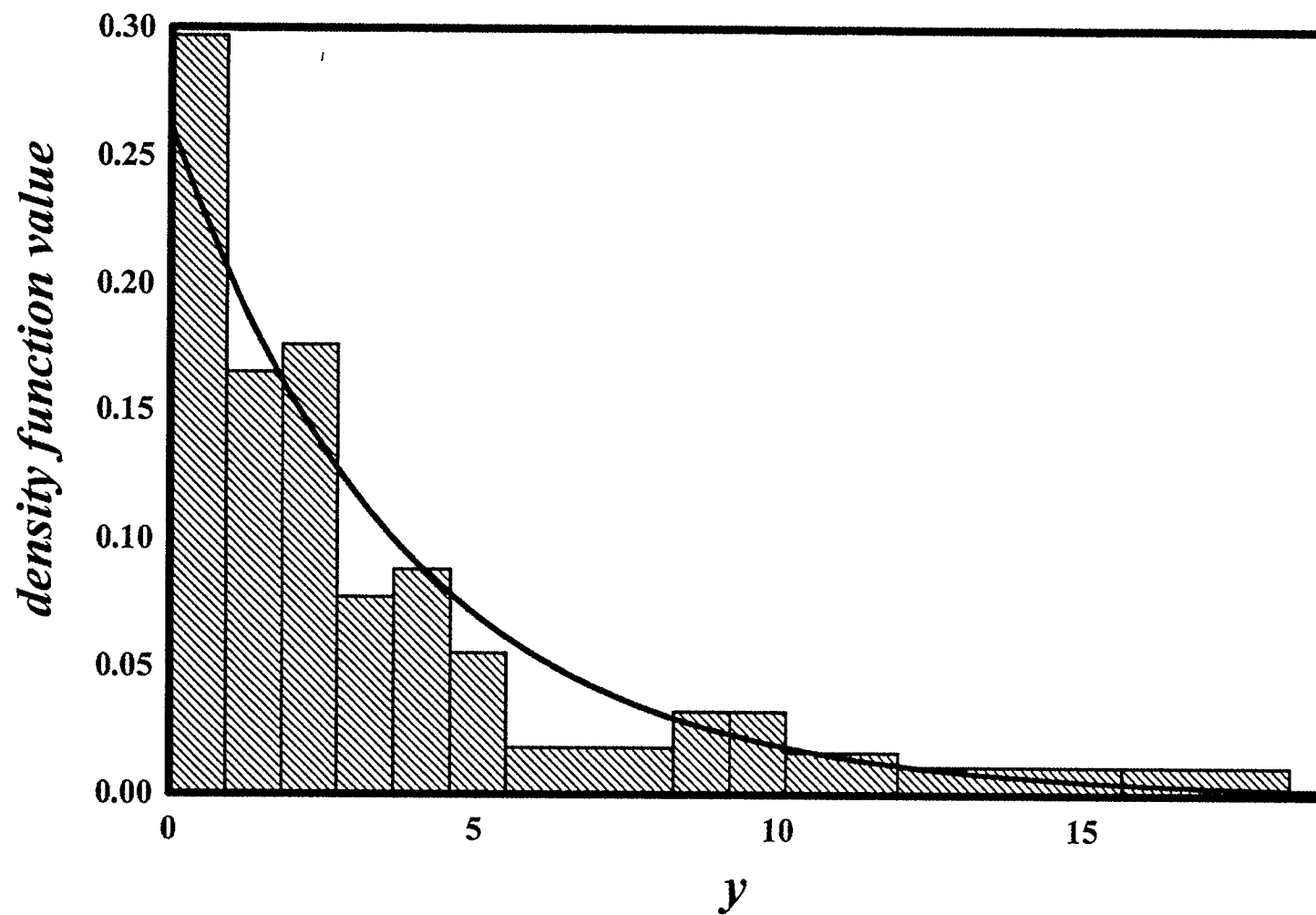
Fig. 1s. Amplitude distribution of PBC in Aroclors. Source: Ref. 29. Histogram: original data.
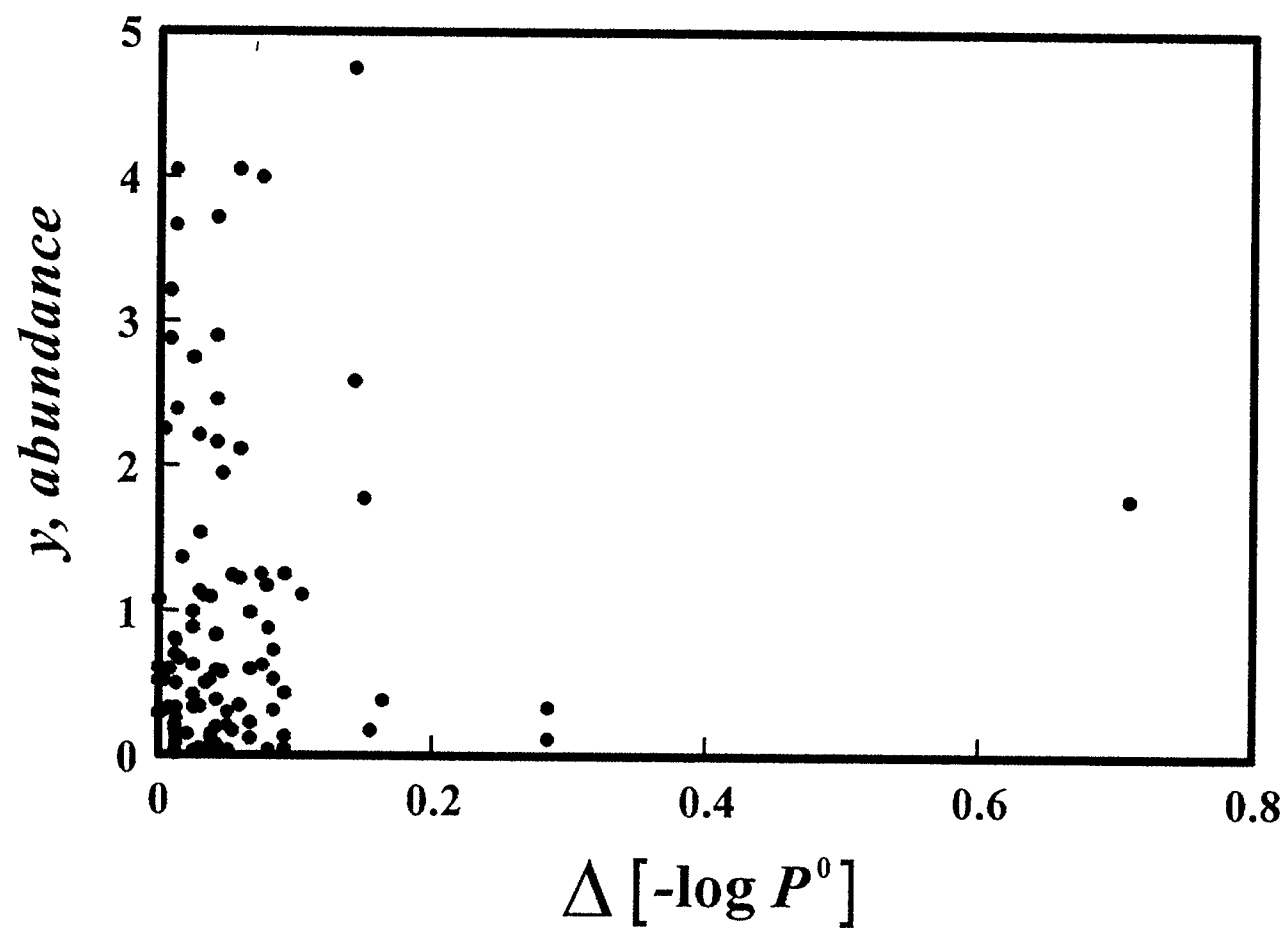Continuous function: Exponential approximating function of mean equal to the histogram mean.

Fig. 2s. Amplitude vs differences between subsequent -log P^0 values of PCB in Aroclors. P^0, vapour pressure at 25 °C (mmHg). Source: Ref. 28.