Journal of Chromatography A, 654 (1993) 143–149
Elsevier Science Publishers B.V., Amsterdam

CHROM. 25 425

Least-squares analysis of gas chromatographic data for polychlorinated biphenyl mixtures

Jordan L. Spencer*

Department of Chemical Engineering, Materials Science and Mining Engineering, Columbia University, New York, NY 10027 (USA)

James P. Hendricks and D. Kerr

Astoria General Laboratories, Consolidated Edison Company of New York, 31-01 20th Avenue, Long Island City, NY 11105 (USA)

(First received July 14th, 1993; revised manuscript received July 9th, 1993)

ABSTRACT

A method for processing the peak area *vs.* retention time data obtained in the gas chromatographic analysis of polychlorinated biphenyl (PCB) mixtures is described. The method is based on a least-squares procedure for representing the chromatogram of the unknown (sample) as a linear combination of the chromatograms of the base PCB mixtures of which the sample is assumed to be composed. A factor used to adjust for minor variations in retention times is also determined by an optimization procedure.

INTRODUCTION

In laboratories responsible for analyzing samples containing polychlorinated biphenyls (PCBs), it is often necessary to process samples made up of PCBs from several sources. (In the following we will call a mixture of polychlorinated biphenyl congeners of fixed composition a PCB, e.g. Arochlor 1260 is a PCB. A combination of PCBs will be called a blend.) A given PCB may contain 10 to 30 or more distinct chemical species, since 209 PCB congeners exist [1]. Thus a PCB or a blend, when analyzed using a suitable column and chromatographic procedure, produces a chromatogram made up of multiple peaks, each with a corresponding area and retention time.

Methods for separating PCB congeners by gas chromatography are well developed and have been used for many years [1–6]. If each PCB shows one or more large peaks which do not occur in any of the others, it is easy to calculate the amount of each PCB in a blend. However this is not always the case, because each PCB is in fact made up of the same distinct chemical species (congeners). Thus a more sophisticated method for processing the peak area *vs.* retention time data is needed. In this paper we report the development and testing of such a method, and show that it performs well in analyzing PCB blends.

THEORY

The data analysis involves two stages. In the first, the peak area *vs.* retention time data are used to reconstruct a continuous function of time that approximates the original signal obtained from the chromatograph detector as the sample is eluted from the column. This continuous function is then sampled at evenly spaced times to produce a discrete-time function, or vector.

* Corresponding author.

0021-9673/93/$06.00

We denote the value of the kth function at the jth time as C_j^k.

Assuming that the chromatograph operates in the linear range, the signal produced when analyzing a PCB mixture is a linear combination of the signals produced when running each of the pure PCB samples. The coefficients of the linear combination are proportional to the amounts of pure PCB which make up the sample being analyzed. In the second phase of the processing these coefficients are determined by requiring that they produce a least-squares fit to the experimental chromatogram. This method has the advantage that all of the data are used, rather than limiting the analysis to a small number of peaks. If, however, data corresponding to certain ranges of retention times are known to be especially reliable (or unreliable), these data can be weighted more (or less) heavily in determining the coefficients.

Elution curve reconstruction

The reconstructed chromatogram, $C(t)$, is the sum of Gaussian functions, each Gaussian centered at the corresponding retention time and having an area proportional to the area of the peak occurring at that time. Thus

$$C(t) = \sum_{i=1}^{L} p_i \, e^{-[(t-t_i)/a]^2}$$

Here p_i is the area of the ith peak, t is time, t_i is the retention time for the ith peak, a is the peak width, and L is the number of peaks. Note that, in the absence of other information, each peak is assumed to have the same width.

The peak width a can be chosen arbitrarily. If, however, a is too small, slight variations in the retention time of a peak will have a major effect on the coefficients, and thus on the results. On the other hand, if a is too large, peaks located close to each other will appear to coalesce and accuracy will be compromised. In practice, a is easily chosen so as to reproduce reasonably well the experimental spectra.

Finally, the discrete-time version of the chromatogram, C_j, is determined by sampling the continuous function $C(t)$ at M evenly spaced points, *e.g.*

$$C_j = C(j\,\Delta t)$$

where j indexes the time ($j = 1, 2, \ldots, M$) and Δt is the time increment between points.

Parameter estimation

The problem now becomes that of expressing the chromatogram of the sample, corresponding to the M-dimensional vector V, as a linear combination of the K M-dimensional chromatogram vectors C^k exactly. That is, we want

$$V = \sum_{k=1}^{K} b_k C^k$$

where the K parameters b_k are to be determined. Note that on physical grounds the parameters b_k must be non-negative, since it is not possible to use a negative amount of a mixture in preparing a sample. We choose, however, not to impose an inequality constraint on the parameters. Instead the occurrence of a significantly negative parameter value will be taken as the sign of a problem in fitting the data.

Since V and C^k are M-dimensional vectors, with M greater than K, this problem cannot in general be solved exactly. Thus we employ the widely used method of least-squares, *i.e.* we determine the parameters b_k by minimizing the objective function J, with

$$J = r^{\mathrm{T}} r$$

The residual vector r is given by

$$r_j = V_j - \sum_{k=1}^{M} b_k C_j^k$$

If we define the $M \times M$ matrix B as

$$B_{jk} = C_j^k$$

the minimum of J is found by solving the K linear equations (often called the normal equations)

$$B^{\mathrm{T}} B b = B^{\mathrm{T}} r$$

for the parameter vector b. Here superscript T denotes transposition.

The standard error, s_k, of the kth parameter can be found as

$$s_k = \sqrt{s_m^2 [(B^{\mathrm{T}} B)^{-1}]_{kk}}$$

where s_m^2 is the variance of the measured data, which can be estimated by running the same sample a number of times and determining the run-to-run variation in the chromatogram.

The parameter correlation matrix P is found from the matrix $W = (B^T B)^{-1}$ as follows:

$$P_{jk} = W_{jk}/(\sqrt{W_{jj}}\sqrt{W_{kk}})$$

In general, when all the off-diagonal elements of P (which are necessarily less than or equal to 1 in absolute value) are smaller than, say, 0.9, the parameters will be accurately determined. This occurs when the standards (PCBs) differ widely from each other in their congeneric content, and thus exhibit very different chromatograms. If, on the other hand, one of the standards is very close to being a mixture of the other standards, one of the off-diagonal elements of P will approach unity in absolute value, and the parameters corresponding to the row and column of this element will not be well-determined. This situation will reveal itself also in large parameter standard errors.

In the actual practice of chromatography, the retention time of a given species depends inversely on the flow rate of the eluting gas and also on the temperature of the column, both of which are subject to variations which are in general small, but not exactly zero. When the peaks are narrow and well separated, even a small shift in the retention time can lead to errors in the estimation of the composition of the sample using the above algorithm. Thus we have added as a parameter to be determined a factor, denoted f, which multiplies the observed retention times of the sample. This factor was de-

TABLE I

RETENTION TIMES AND PEAK AREAS FOR AROCHLOR STANDARDS

Peak number	Arochlor 1242		Arochlor 1254		Arochlor 1260	
	Time	Area	Time	Area	Time	Area
1	1.27	0.140	1.40	0.299	1.40	0.141
2	1.44	4.247	1.69	0.064	1.69	0.034
3	1.70	6.605	2.09	0.209	2.09	0.093
4	1.89	2.409	2.44	2.509	2.45	0.163
5	2.09	18.359	2.93	1.776	3.55	1.095
6	2.31	5.748	3.27	0.497	3.73	1.622
7	2.47	12.975	3.55	6.790	4.30	0.089
8	2.69	0.345	3.73	7.553	4.87	4.036
9	2.94	6.238	4.32	2.202	5.34	5.531
10	3.06	6.006	4.52	2.148	5.92	9.551
11	3.27	4.430	4.83	6.875	6.30	1.371
12	3.56	14.717	5.20	12.476	6.89	6.213
13	4.25	8.035	5.85	16.808	7.64	18.509
14	4.83	1.956	6.92	5.492	8.51	1.930
15	5.20	2.320	7.68	16.875	9.35	5.221
16	5.85	3.317	8.39	1.290	9.83	4.557
17	6.89	0.148	9.29	1.630	11.10	18.201
18	7.55	1.781	10.14	3.260	13.26	7.439
19	18.21	0.105	11.18	5.165	14.59	7.713
20	25.30	0.122	13.21	0.552	17.75	1.658
21	–	–	14.59	1.681	21.00	3.664
22	–	–	17.80	0.202	23.63	1.168
23	–	–	21.03	0.390	–	–
24	–	–	23.68	0.109	–	–
25	–	–	27.96	0.064	–	–

termined by calculating the RMS residual for three values of f, namely 0.990, 1.000, and 1.010. Then a quadratic was fitted to these points and the next f, denoted f^*, was calculated as that value which minimized the quadratic. Repeating this procedure for $f^* - 0.002$, f^*, and $f^* + 0.002$ gave the final value of f, which minimized almost exactly the RMS residual. In the results reported below, this procedure was used. A search over a range of values of f close to 1.000 gave essentially the same results, but was somewhat slower.

EXPERIMENTAL

The analyses were carried out using Perkin-Elmer 8400 Series gas chromatographs, according to a procedure described previously [2]. These chromatographs are temperature-programmed single channel units using Ni^{63} electron capture detectors. The conditions were as follows: column packing, Chromosorb W HP; column load, 4% OV-225; mesh size, 180–250 μm; column size, 1.8 m × 4 mm I.D.; carrier gas, argon–methane (95:5, v/v); carrier gas flow, 30 ml/min; oven temperature, 225°C; auxiliary temperature, 375°C.

Table I contains the retention times and peak areas for the three standards used, namely Arochlor 1242, Arochlor 1254, and Arochlor 1260. The samples were made up from these standards. Note that in all cases the solvent peak, which occurs at less that 1 min, has been removed.

While Arochlor 1242 has large peaks at lower retention times, the other two standards show peaks over a wide range of retention times.

Samples used to test the method were prepared by accurately weighing specified amounts

TABLE II

RETENTION TIMES AND PEAK AREAS FOR SAMPLES 1, 2, AND 3

Peak number	Sample 1		Sample 2		Sample 3	
	Time	Area	Time	Area	Time	Area
1	1.42	0.530	1.42	0.297	1.39	0.158
2	1.70	0.411	1.70	0.278	1.69	0.062
3	1.89	0.109	1.89	0.078	2.09	0.125
4	2.10	1.185	2.10	0.796	2.45	0.344
5	2.45	3.333	2.31	0.231	2.93	0.243
6	2.94	2.203	2.47	0.620	3.29	0.063
7	3.28	0.680	2.94	0.255	3.55	1.551
8	3.56	7.260	3.07	0.272	3.74	2.115
9	3.74	7.306	3.28	0.151	4.31	0.319
10	4.33	2.492	3.56	1.575	4.54	0.204
11	4.53	2.076	3.74	1.625	4.87	4.201
12	4.84	6.743	4.28	0.256	5.34	6.162
13	5.21	12.196	4.88	3.843	5.93	10.075
14	5.87	16.468	5.35	5.457	6.30	1.355
15	6.93	5.329	5.94	9.396	6.90	6.202
16	7.70	16.524	6.31	1.262	7.65	18.430
17	8.40	1.263	6.91	5.942	8.53	1.930
18	9.30	1.529	7.66	18.222	9.36	4.976
19	10.16	3.357	8.54	1.702	9.85	4.461
20	11.21	5.526	9.38	4.953	11.12	17.303
21	13.26	0.876	9.86	4.373	13.26	6.730
22	14.64	1.813	11.14	17.809	14.62	6.969
23	17.79	0.230	13.29	6.989	17.76	1.480
24	21.06	0.394	14.64	7.219	21.05	3.450
25	23.65	0.103	17.78	1.613	23.72	1.090
26	27.12	0.063	21.07	3.632	–	–

of the standards, which were supplied as 1000 ppm (w/w) solutions in isooctane by Supelco, Bellefonte, PA (USA). Table II shows the retention times and peak areas obtained by analyzing the first three of these samples.

RESULTS AND DISCUSSION

In order to demonstrate the effect of the peak width parameter, denoted a, that is used in reconstructing the chromatograms, the peak area *vs.* retention time data for Arochlor 1260 (Table I) were used. Peak width values of 0.100 min, 0.250 min, and 1.000 min were employed to generate the chromatograms (for Arochlor 1260) shown in Fig. 1. It is evident that $a = 0.100$ min produces very sharp and well separated peaks, such that the value of the chromatogram at a given time can be extremely sensitive to the retention time of the peak. In contrast, for $a =$ 1.000 the peaks are much broader and tend to overlap strongly. In fact the nine peaks with retention times between 4 and 10 min coalesce into a single peak with one shoulder. When a peak width of 0.250 min is used the peaks are somewhat less sharp, but remain well separated except for the coalescence of a few peaks between 5 and 8 min. Thus a peak width of 0.250 was chosen for the cases discussed below. This is also justified by the good qualitative agreement between the original chromatogram (Fig. 2, in which the detector signal in arbitrary units is plotted for 660 values of the retention time) and the reconstructed chromatogram, for $a = 0.250$, in Fig. 1. Note that in the original chromatogram

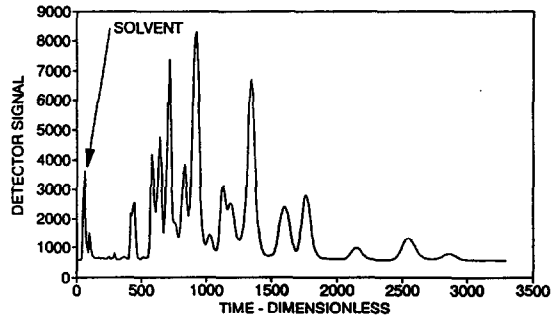

Fig. 2. Original chromatogram for Arochlor 1260, with every fifth of 3300 original points plotted.

the solvent peak at a retention time of about 1 min has been retained, while the solvent peak has been removed in the reconstructed chromatograms.

Fig. 3 shows the reconstructed chromatograms for the three PCB mixtures from which the samples were prepared. Note that they all overlap to some extent, particularly Arochlors 1254 and 1260. Nevertheless they differ sufficiently that it is reasonable to assume that a given blend can be accurately resolved into its components. This in fact is true, as the results below show.

The data in Table II were used to reconstruct discrete-time chromatogram vectors of dimension 501, *i.e.* vectors corresponding to 501 evenly-spaced points 0.050 min apart and thus covering a 25-min interval. From these vectors the composition of each sample was calculated by setting up and solving the normal equations. In Table III the calculated compositions (in mass percent) are compared with the compositions of

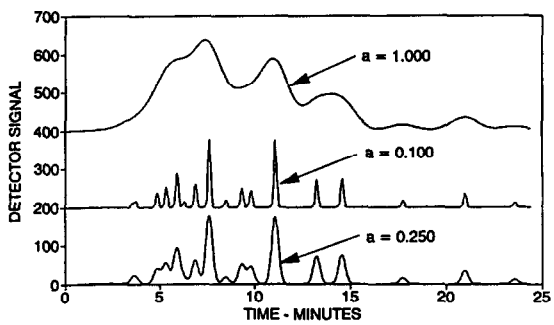

Fig. 1. Reconstructed chromatograms for Arochlor 1260 for various values of peak width parameter: $a = 0.100$, $a = 0.250$, and $a = 1.000$. Solvent peaks have been removed.

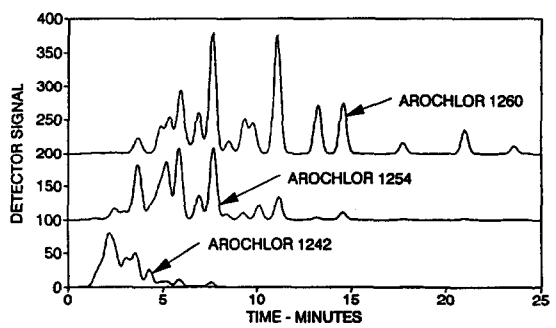

Fig. 3. Reconstructed chromatograms for Arochlors 1242, 1254, and 1260. Peak width parameter $a = 0.250$. Chromatograms shifted upward to avoid overlap.

TABLE III

CALCULATED AND KNOWN COMPOSITIONS IN MASS PERCENT FOR SAMPLES 1–9

Sample	Known Arochlor			Calculated Arochlor		
	1242	1254	1260	1242	1254	1260
1	10.00	0.00	90.00	11.14	−1.01	88.38
2	0.00	10.00	90.00	0.95	10.50	90.44
3	0.00	90.00	10.00	−0.23	91.86	11.09
4	10.00	90.00	0.00	9.48	88.11	1.57
5	1.00	99.00	0.00	0.39	97.92	−0.31
6	0.00	1.00	99.00	0.11	1.48	99.25
7	0.00	99.00	1.00	−0.20	98.62	0.49
8	0.00	1.00	99.00	0.80	1.05	97.94
9	99.00	0.00	1.00	98.70	0.59	1.43

the samples produced by accurate volumetric mixing of the standards.

In general the agreement is excellent. The maximum difference between the known and calculated mass percents is 1.89%, and the average difference is 0.26%. Note also that in only four cases was a negative mass percent calculated, even though the estimated parameters were not constrained to be non-negative. And the largest negative mass percent was −0.26%, the other three being −0.11, −0.09, and −0.02%.

The variance–covariance matrix was used to estimate the standard error of the parameter estimates, based on the RMS residuals. In all cases the standard errors lay between 0.10 and 0.40% (w/w), which is consistent with the observed differences. The parameter correlation matrix P,

$$P = \begin{bmatrix} 1.0000 & -0.5736 & 0.3891 \\ -0.5736 & 1.0000 & -0.8133 \\ 0.3891 & -0.8133 & 1.0000 \end{bmatrix}$$

shows that the standards are sufficiently different in congener content to give relatively low parameter correlation, and thus permit accurate estimates of the parameters. Note that parameters 2 and 3 are the most strongly correlated, corresponding to the occurrence of peaks (see Fig. 3) with retention times between 3 and 12 min in the chromatograms of Arochlors 1254 and 1260.

As another indication of the results, Fig. 4a–c shows the reconstructed and best-fit chromatograms for samples 1, 2 and 3, respectively. Also shown are the residuals. These are quite small, indicating again that the best-fit chromatograms fit the reconstructed experimental chromatograms well.

As a further test, new standards were prepared and run, and four samples with Arochlor ratios of 1:1:0, 1:0:1, 0:1:1 and 1:1:1 were run

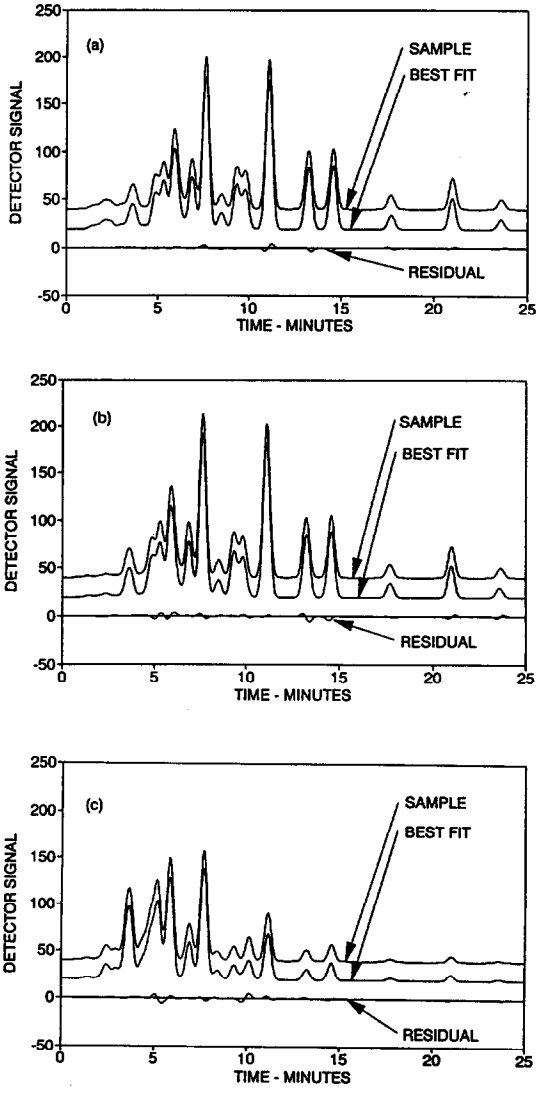

Fig. 4. Reconstructed and best-fit chromatograms for samples 1, 2 and 3, and the corresponding residuals. (a) Sample 1; (b) sample 2; (c) sample 3.

TABLE IV

CALCULATED AND KNOWN COMPOSITIONS IN MASS PERCENT FOR ADDITIONAL SAMPLES

Sample	Known Arochlor			Calculated Arochlor		
	1242	1254	1260	1242	1254	1260
1	50.00	50.00	0.00	50.21	49.78	0.00
2	50.00	0.00	50.00	50.33	0.98	48.69
3	0.00	50.00	50.00	0.52	47.80	51.69
4	33.33	33.33	33.33	34.39	32.79	32.82

using the data from the standards. The results are shown in Table IV. Negative mass percents (one value only, −0.78%) were set to zero, and the rest were scaled to bring the sum of the mass percents to 100. Again agreement was good, with a maximum absolute error of 2.2% (w/w), and an average error of 1.06%.

In summary, the least-squares procedure, as applied to 13 known blends each made up from three PCBs (Arochlors 1242, 1254, and 1260), worked well. The results were quite close to the known compositions of the samples. The use of an optimization method to determine a best value for the retention time factor also contributed to the accuracy of the method, and produced good agreement between chromatograms reconstructed from peak area *vs.* retention time data and the best-fit chromatogram as determined by the least-squares calculation.

REFERENCES

1 D.E. Schulz, G. Petrick and J.C. Duiker, *Environ. Sci. Tech.*, 23 (1989) 852–858.
2 S.I. Lerman, H. Gordon and J.P. Hendricks, *Amer. Lab.*, 42 (1981) 176–181.
3 C. Strang, S.P. Levine, B.P. Orlan, T.A. Gouda and W.A. Saner, *J. Chromatogr.*, 314 (1984) 482–487.
4 R.P. Kozloski, *J. Chromatogr.*, 318 (1985) 211–219.
5 S. Sabbah and M.L. Bouguerra, *J. Chromatogr.*, 552 (1991) 206–212.
6 M.R. Driss, S. Sabbah and M.L. Bouguerra, *J. Chromatogr.*, 552 (1991) 213–222.