Functional Analysis of Gas Chromatographic Data for C-4 through C-18:2 Fatty Acids

C.A. WULFF, Department of Chemistry, University of Vermont, Burlington, Vermont 05405;
A.H. DUTHIE and **S. WULFF**, Department of Animal Sciences, Agricultural Experiment Station, University of Vermont, Burlington, Vermont 05405

ABSTRACT

The problem of finding a functional relationship between concentration and output data from GC has been considered for a series of fatty acids (FA). The concentrations used are relative values, i.e., the actual concentration of a FA divided by that for an internal standard. The GC output used is relative peak areas, i.e., the integrated chromatogram area of the FA divided by that for the internal standard. Five functions have been investigated, each in reasonable accord with the concept of relative response factors being constant for a homologous series of compounds. We conclude that one of the five, $C_j = aA_j^b$, is clearly superior and recommend its use in agreement with a literature report for different compound types (5). In the equation above, "C" and "A" are the relative concentrations and areas, respectively, and "a" and "b" are fitting constants determined from the data by least-squares minimization. We found it was impossible to deal with a matrix of 10 acids and 9 concentrations in a manner that would reproduce the data within the experimental uncertainty (6%), and we concluded that an acid-by-acid analysis was preferable.

INTRODUCTION

A quantitative procedure for the analysis of C-4 through C-18:2 nonesterified fatty acids (NFA) is being investigated in the Flavor Chemistry Laboratory at the University of Vermont. The procedure involves extraction and isolation of NFA as potassium salts, and then injection of these salts into a gas chromatograph (GC).

In 1975, Cochrane (1) stated that esters of C-2 through C-6 acids could not be analyzed quantitatively by GC. He recommended the method of Ackman and Burgher (2). Our procedure required the quantitative analyses of butanoic and hexanoic acids, so Cochrane's statements interested us. Ackman and Burgher (2) first reported that when a GC was modified to allow formic acid vapors in the carrier gas to pass into the injection port, C-2 through C-6 acids were separated, and ghosting and peak tailing were not problems. Cochrane (3) applied Ackman and Burgher's method for the quantitative analysis of C-2 through C-12 acids. In his review article, Cochrane (1) reported that quantitative analysis of C-2 through C-6 acids could not be accomplished unless the carrier gas contained formic acid vapors. Gray (4) used Cochrane's approach to analyze C-2 through C-18:1 acids from Cheddar cheese. He reported a mean recovery of 97.2 ± 4.8% for 10 acids added to Cheddar cheese.

In our procedure, before an extraction method and an isolation method can be selected, it is essential that each NFA injected into the GC be separated completely and the output (GC data) calculated accurately. Initial calibration curves prepared in our laboratory with standard C-4 through C-18:2 acids injected into the GC as potassium salts gave unacceptable precision at low concentrations in the linear dynamic range. We suspect that these deviations from linearity resulting from interactions with sample column and/or sample GC will be most apparent at low concentrations, independent of the type of sample analyzed.

In a recent paper, Shatkay and Flavian (5) discuss some potential errors inherent in quantitative analysis by GC and make recommendations concerning volumes of sample injected and the construction of calibration curves for systems containing internal standards. Most interesting, from our point of view, is their conclusion – based on results for esters and hydrocarbons – that an exponential relationship between chromatographic peak area and sample concentration provides a better data fit than does the conventional linear relationship implied by assumed detector linearity. Our results, gathered over a 3-year period on several different columns, support Shatkay and Flavian's conclusion as will be shown from the results obtained for several different model-fitting functions.

EXPERIMENTAL

Provenance and Preparation of Standards

Butanoic acid (99 + %), C-4, was obtained from the Aldrich Chemical Co. (Milwaukee, WI); hexanoic acid (99 + %), C-6, from Fisher Scientific, Inc., (Fair Lawn, NJ); and octanoic acid (99 + %), C-8, from Eastman Organic Chemicals (Rochester, NY). All other acids (chromatographic quality) were obtained from Applied Science Labs., Inc. (State College, PA). Standard stock solutions of potassium salts of the fatty acids (FA) under investigation were prepared in 80% ethanol in concentrations of either 0.05 mg/cm^3 or 1.0 mg/cm^3 by neutralization with Fisher potassium hydroxide. Nine standard solutions were prepared from these stock solutions. These standards were evaporated to dryness, and the recovered potassium salts were dissolved in one cm^3 of the internal standard solution. This solution, consisting of the potassium salts of C-7 and C-17, was prepared as a large stock solution with concentrations of ca. 0.25 mg/cm^3 of C-7 and 1.00 mg/cm^3 of C-17. This constant amount, one cm^3, of internal standard solution applies to all samples used in the calibration study.

Apparatus

The FA were separated by GC on a Hewlett-Packard Model 402 chromatograph equipped with a dual flame ionization detection system. A glass column (1.8 m x 2 mm i.d.) packed with 5% DEGS-PS (diethylene glycol succinate-phosphoric acid added) on 100/120 mesh Supelcoport (Supelco, Supelco Park, PA) was used. The helium carrier gas was passed through a glass trap containing formic acid (99 + %, ICN Pharmaceuticals, Plainview, NY). The helium was saturated with formic acid vapor by passing the gas over the liquid surface prior to its entering the injection port of the chromatograph. The carrier gas flow rate was ca. 40 cm^3 min.$^{-1}$ The chromatograph was programmed from 100 to 200 C at a rate of 7.5 C min.$^{-1}$ Peak areas were determined with a Hewlett-Packard Model 3380A Reporting Integrator. Hewlett-Packard 67/97 Programmable Calculators were used for all numerical analyses.

Data Analysis

The calibration information was contained in two corresponding matrices, C_{ij}/and A_{ij}, where C is the dimensionless ratio of the concentration of a particular FA to the concentration of a reference acid and A is the dimensionless

TABLE I

Compositions of Calibration Solutions, j = 1 - 9

Potassium Salts of FA (mg/cm^3); C-7: 0.25, C-17; 1.00 throughout									
FA	j=1	j=2	j=3	j=4	j=5	j=6	j=7	j=8	j=9
C-4,6,8,10,12,14,18, 18:2	0.01	0.05	0.10	0.25	0.50	0.75	1.00	1.50	2.00
C-16, 18:1	0.10	0.20	0.50	1.00	2.00	3.00	4.00	5.00	6.00

ratio of their areas as determined by GC. The index "i" covers the sequence of acids analyzed for even carbon number acids from butanoic, C-4, to octadecanoic, C-18, and the two unsaturated acids, 9-octadecenoic, C-18:1, and 9, 12-octadecadienoic acid, C-18:2. The index "j" covers nine concentration ranges studied that spanned factors of 0.04 to 8.00 multiples of the reference acid. The reference acids were heptadecanoic, C-17, for C-17 and C-18:1, and heptanoic, C-7, for the remainder. The concentrations are given in Table I. Replicate analyses (N = 7 to 10) were performed for each of the nine standardizing solutions to permit an assessment of the experimental uncertainty in the GC result. The most convenient expression of this uncertainty is as a coefficient of variation, γ_{ij}, defined as γ_{ij} , = (s_{ij}/A_{ij}) x 100, where s_{ij} is the standard deviation of the several results for the value of A_{ij}. Of the γ_{ij}, 50 were less than 4%, 19 were between 4% and 7%, 11 were between 7% and 10%, and 7 were between 10% and 17%. The larger values were encountered for the most dilute solutions. The combination of 10 acids being considered and 9 concentrations gave rise to 87 entries in each of the C_{ij} and A_{ij} matrices. These 87 pairs shall be our data in the following discussion. (Only 87, rather than 90, pairs are available because data could not be obtained for C-18 at the lowest C_{ij} or for C-18:2 at the two lowest C_{ij}.)

The following requirements were set for the functional relationship that would correlate the C_{ij} with the A_{ij}: (a) the representation should be analytical rather than graphical for easier data handling; (b) recovery of the original data should be within the uncertainties imposed by the γ_{ij}; (c) deviations between the recovered and original data should *not* show a trend with concentration; (d) the functionality should be as simple as possible and should be compatible with our knowledge of GC behavior.

Five functional forms were considered.

(1) $(C_{ij}) = a_1(A_{ij})$. This simplest form is based on the assumption that the homologous series of FA could be described by a single relative response factor. If this assumption is true, the value of a_1 should be unity. (For convenience, the C_{ij} values have been scaled to the concentration of the internal standard to permit output of actual unknown concentrations rather than multiples thereof. Since the concentration of C-7 is 0.25 mg/cm^3, the expected values of a_1 becomes 0.25. Data for C-16 and C-18:1 have been divided by four also for convenience.) The numerical value of a_1 can be obtained from the data by a minimization of the function:

$$D_1 = \Sigma\,[(C_{ij}) - a_1\,(A_{ij})]^2. \qquad 1\text{-}1$$

Setting (dD/da) = 0, leads to the working relationship,

$$a_1 = \left\{\Sigma(C_{ij})(A_{ij})\right\}/\Sigma(A_{ij})^2. \qquad 1\text{-}2$$

(2) $(C_{ij}) = a_2(A_{ij})$. The principal difficulty with the relation described by equation 1-2 is that the quantity minimized in equation 1-1 is the sum of the squares of the absolute deviations. No accounting is made for the fact that a deviation of 0.01 in a calculated A_{ij}, while representing less than 1% of the total at high concentrations, might amount to 30% of the A_{ij} value at the lowest concentration. This situation can be relieved by minimizing not the absolute derivation, but rather the percentage deviation. The equations corresponding to 1-1 and 1-2 are:

$$D_2 = \Sigma\left[\left\{(C_{ij}) - a_2(A_{ij})\right\}/(C_{ij})\right]^2, \qquad 2\text{-}1$$

and

$$a_2 = \left[\Sigma\left\{(A_{ij})/(C_{ij})\right\}\right] \div \left[\Sigma\left\{(A_{ij})/(C_{ij})\right\}^2\right]. \qquad 2\text{-}2$$

(3) $(C_{ij}) = b_3 + a_3(A_{ij})$. Previous authors (6,7) have noted that better fits of GC data could be obtained if their relationship were not forced through the origin of a C vs. A plot. The nonzero intercept can be attributed to one of several apparatus artifacts. From a data-handling point of view, the easing of a restriction by introducing a second parameter would allow for a better fit. The most usual approach is a straightforward linear least-squares technique based on the minimization of:

$$D_3 = \Sigma\,[(C_{ij}) - b_3 - a_3(A_{ij})]^2 \qquad 3\text{-}1$$

through the relations $(\partial D_3/\partial b_3) = 0$, and $(\partial D_3/\partial a_3) = 0$, to yield

$$a_3 = \left\{\Sigma(C_{ij})\cdot\Sigma(A_{ij}) - N\Sigma(C_{ij})(A_{ij})\right\} \div \left[\left\{\Sigma(A_{ij})\right\}^2 - N\Sigma(A_{ij}{}^2)\right], \qquad 3\text{-}2$$

and

$$b_3 = \left\{\Sigma(C_{ij}) - a_3\Sigma(A_{ij})\right\} \div N. \qquad 3\text{-}3$$

The N appearing in equations 3-2 and 3-3 is the number of paired data entries, with 87 for the entire matrix fit.

(4) $(C_{ij}) = b_4 + a_4(A_{ij})$: The results of equations 3-1 to 3-3 suffer from the same drawback as do 1-1 and 1-2, i.e., no accounting is made of the fact that a deviation of given size has a greater effect for the low concentrations than it does for the larger ones. In the same fashion that (1) was modified to (2), we can minimize the percentage deviation by treating

$$D_4 = \Sigma\left[\left\{(C_{ij}) - b_4 - a_4(A_{ij})\right\}/(C/(C_{ij})\right]^2. \qquad 4\text{-}1$$

Minimization through the differentials $(\partial D_4/\partial b_4) = 0$ and $(\partial D_4/\partial a_4) = 0$, leads to

$$a_4 = \frac{\left\{\Sigma(C_{ij})^{-1}\right\}\left\{\Sigma(A_{ij})(C_{ij})^{-2}\right\} - \left\{\Sigma(A_{ij})(C_{ij})^{-1}\right\}\left\{\Sigma(C_{ij})^{-2}\right\}}{\left\{\Sigma(A_{ij})(C_{ij})^{-2}\right\}\left\{\Sigma(A_{ij})(C_{ij})^{-2}\right\} - \left\{\Sigma(C_{ij})^{-2}\right\}\left\{\Sigma(A_{ij})^2(C_{ij})^{-2}\right\}} \qquad 4\text{-}2$$

and

$$b_4 = [\{\Sigma(C_{ij})^{-1}\} - a_4\{\Sigma(A_{ij})(C_{ij})^{-2}\}] \div \Sigma\{(C_{ij})^{-2}\}. \quad 4\text{-}3$$

(5) $(C_{ij}) = a_5(A_{ij})b_5$. A final two-parameter fit, based on the assumption of a linear logarithmic fit between (C_{ij}) and (A_{ij}), was attempted. The pertinent sum of squares to be minimized is

$$D_5 = \Sigma[\ell n(C_{ij}) - \ell n(a_5) - b_5\ell n(A_{ij})]^2, \quad 5\text{-}1$$

and the differential equations are $(\partial D_5/\partial \ell n\ a_5) = 0$ and $(\partial D_5/\partial b_5) = 0$. The simultaneous solution of these equations yields the following working relationships for a_5 and b_5:

$$b_5 = \frac{\{\Sigma\ell n(A_{ij})\}\{\Sigma\ell n(C_{ij})\} - N\Sigma\{\ell n(A_{ij})\}\{\ell n(C_{ij})\}}{\{\Sigma\ell n(A_{ij})\}^2 - N\Sigma\{\ell n(A_{ij})\}^2} \quad 5\text{-}2$$

and

$$\ell n(a_5) = [\Sigma\ell n(C_{ij}) - b_5\Sigma\ell n(A_{ij})] \div N, \quad 5\text{-}3$$

where N is, once again, the number of paired data entries.

Programs were written for the Hewlett-Packard 67/97 calculator series that would evaluate all the needed sums and fitting constants with a single pass of data entry and that, with a second pass of data entry, would evaluate the root-mean-square percentage deviations, δ, for each of the five fitting functions. Throughout the following discussion we will use δ as a measure of the "goodness of fit," defining it by

$$\delta = \left[\Sigma\left\{\frac{(A_{ij})^{obs} - (A_{ij})^{fit}}{(A_{ij})^{obs}} \times 100\right\}^2 \div N\right]^{1/2}$$

The criterion for "goodness of fit" shall be that $\delta \cong \gamma_{ij}$.

The results for the data base of 87 paired entries are the following:

(1) $(C_{ij}) = 0.2499(A_{ij})$	$\delta_1 = 33\%, 19\%$
(2) $(C_{ij}) = 0.2568(A_{ij})$	$\delta_2 = 31\%, 17\%$
(3) $(C_{ij}) = 0.0608 + 0.2379\ (A_{ij})$	$\delta_3 = 193\%, 200\%$
(4) $(C_{ij}) = 0.0014 + 0.2526\ (A_{ij})$	$\delta_4 = 31\%, 17\%$
(5) $(C_{ij}) = 0.2762(A_{ij})^{0.9586}$	$\delta_5 = 28\%, 17\%$

Two values have been recorded for δ. The first is for all data points and the second is recomputed with the elimination of the results for C-18, the substance with the largest deviations. These values of δ are to be compared with an overall uncertainty of $\gamma_{ij} = 6\%$. It should be noted that the five values of the linear coefficient are all close to the expected value, 0.25, and that the intercepts in relations (3) and (4) are close to zero, and that the exponent in relation (5) is close to unity. These results can be interpreted as a general confirmation of the concept of a relative response factor with only small deviations from a linear (C_{ij}) to (A_{ij}) relationship. The relatively large values of δ, however, make the derived functions useless for the purpose of setting an analytical calibration scale. The failure of the functions can be demonstrated by plots of individual values of Δ as a function of (C_{ij}), where Δ is defined by

$$\Delta = \frac{(A_{ij})^{obs} - (A_{ij})^{fit}}{(A_{ij})^{obs}} \times 100.$$

Such plots are shown in Figures 1 and 2 for the acids

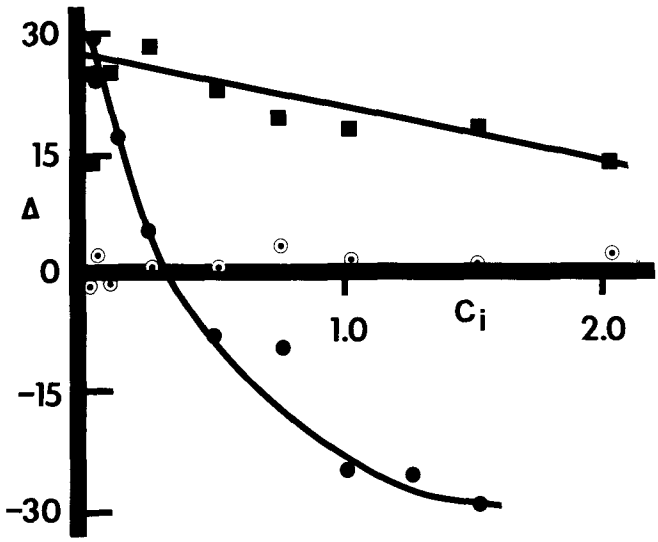

FIG. 1. Deviation plot for selected acids against function (4) as developed for all data. ■ = C-4 ◉= C-12 ● = C-18:1.

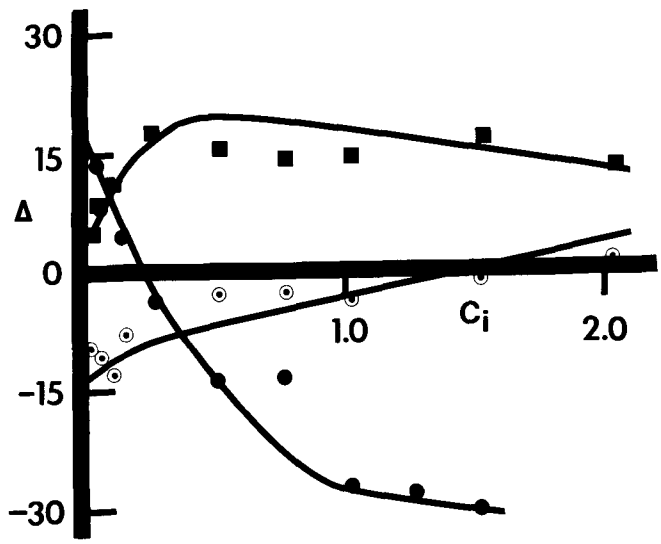

FIG. 2. Deviation plot for selected acids against function (5) as developed for all data. ■ = C-4 ◉= C-12 ● = C-18:1.

C-4, C-12, and C-18:1 for relations (4) and (5). C-12 was chosen as an example of a "good" fit, whereas C-4 and C-18:1 were chosen to demonstrate the existence of concentration dependent trends. It is also interesting to note that the largest deviations are found for the simple linear least-squares fit. (3). On the basis of these results, we concluded that further attempts to fit all 87 data pairs with a single functional relationship would, most likely, prove futile and our efforts were directed to single fits for each acid using the same five functions developed above. The results are presented in Table II.

From these results it appears that functions (4) and (5), a linear fit designed to minimize the *percentage* deviations and an exponential fit, respectively, are the most promising. For both of these the averaged δ values, 7.0 and 5.3%, are comparable to the $\gamma_{ij} = 6\%$. The intercepts in relation (4) were in the range -0.0014 to 0.0518 and the exponents in relation (5) ranged from 0.8587 to 1.0154. These ranges are both in reasonable accord with the linearity implied by relative response factors. The choice between the two relationships can be simplified when the results are depicted as

TABLE II

A Comparison of δ Values for the Individual Acids

	C-4	C-6	C-8	C-10	C-12	C-14	C-16	C-18	C-18:1	C-18:2	<δ>
δ_1	7.9	4.4	2.1[a]	2.4[a]	4.4	12.0	16.7	11.3	45.2	12.1	17.0
δ_2	4.8[a]	4.4	1.9[a]	2.4[a]	4.1	10.4	11.4	10.8	28.2	11.7	10.2
δ_3	70.7	23.3	1.9[a]	15.2	6.1	12.8	33.2	39.3	158.1	25.6	59.3
δ_4	3.7[a]	1.9[a]	0.9[a]	1.1[a]	1.3[a]	2.0[a]	7.7	7.4	16.8	8.7[a]	7.0
δ_5	1.9[a]	3.7[a]	1.3[a]	1.7[a]	2.6[a]	5.4[a]	3.5[a]	9.9	5.3[a]	10.6[a]	5.3[a]
γ_i	5	4	3	3	4	6	4	7	6	11	6

[a]Value of $\delta < \gamma$.

deviation plots. Figures 3 and 4 are such plots for the same acids used in Figures 1 and 2. Figure 3, the deviations from individual fits to function (4), still shows an apparent systematic trend for C-18:1. The deviations from relation (5), depicted in Figure 4, show no obvious trends and are smaller in magnitude. It should be emphasized again that the particular acids chosen for graphical representation demonstrate both "good" (C-12) and "poor" (C-4, C-18:1) results. On the basis of these figures, relation (5) seems to provide a more precise data fit. Further evidence supporting our preference for relation (5) can be seen in Figure 5. Plotted here are the C_j vs. A_j data for the lowest concentrations of C-18:1 with the experimental uncertainties ($\pm s_{ij}$) in the data indicated by the vertical bars. This large scale plot shows that there is definite curvature in the data that should not be hidden in a linear least-squares fit.

A final confirmation of our choice of relation (5) lies in its utility as an extrapolation function. While we recommend that calibration curves cover the spectrum of expected results, we recognize that the unforeseen does occur and that extrapolations may be necessary. We have tested the two relations, (4) and (5), using the data for C-18:1. Nine C_j-A_j pairs are available for this acid, and we have developed fitting functions using j = 1-9(a), 1−8(b), 1-7(c), 2-9(d) and 3-9(e). Modes (b) - (e) were then used to calculate values of A_j for the missing entries for comparison with the actual values. These results are shown in Table III and indicate clearly that relation (5) is refereable to relation (4).

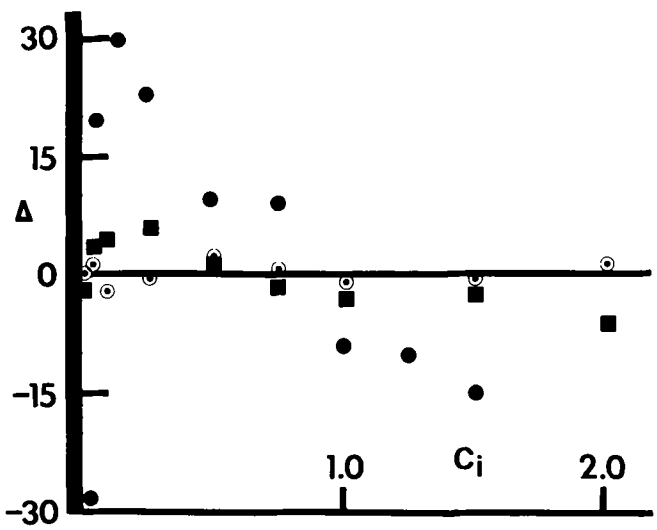

FIG. 3. Deviation plot for selected acids against individual acid functions (4). ■ = C-4 ◉= C-12 ● = C-18:1.

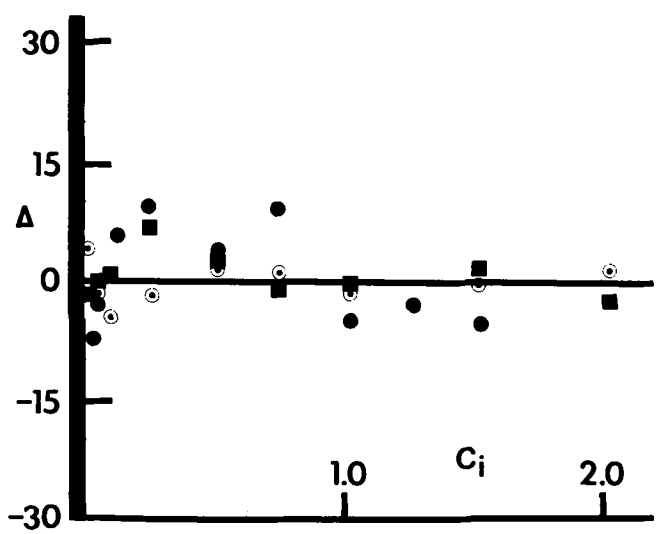

FIG. 4. Deviation plot for selected acids against individual acid functions (5). ■ = C-4 ◉= C-12 ● = C-18:1.

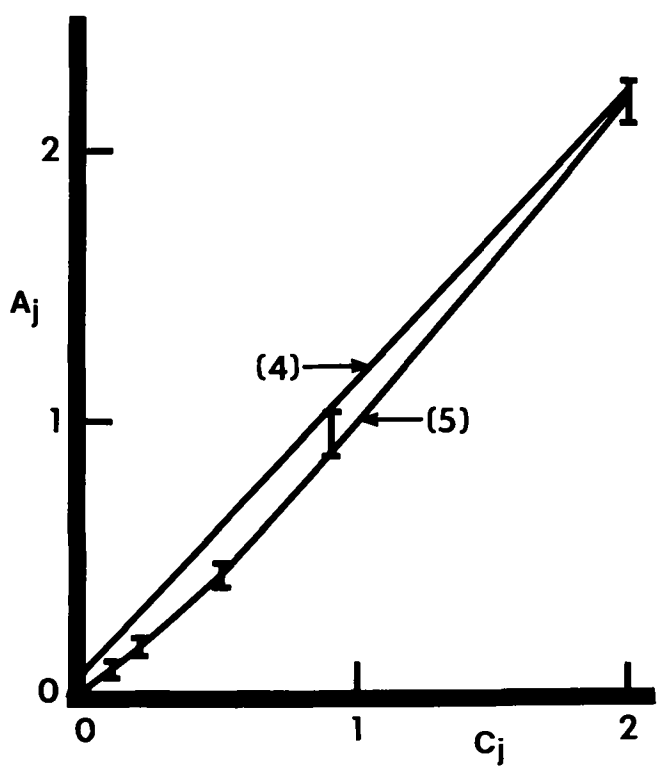

FIG. 5. Relative peak area, A_j, as a function of relative concentration, C_j for C-18:1; curves for functions (4) and (5).

TABLE III

Extrapolated Values Using Functions (4) and (5)

Function (4), acid C-18:1											
Cj	A(obs)[a]	A(4a)	δ(4a)	A(4b)	δ(4b)	A(4c)	δ(4c)	A(4d)	δ(4d)	A(4e)	δ(4e)
1	0.0729	0.0596	-18					(0.0071)	-90	(-0.159)	-319
2	0.1568	0.182	16					0.133	-15	(-0.025)	-116
3	0.4244	0.551	30					0.515	21	0.318	- 10
7	5.323	4.84	- 9	4.65	-13	4.45	-16				
8	6.748	6.07	-10	5.83	-14	(5.58)	-17				
9	8.543	7.29	-15	(7.00)	-18	(6.69)	-22				

Function (5), acid C-18:1											
Cj	A(obs)	A(5a)	δ(5a)	A(5b)	δ(5b)	A(5c)	δ(5c)	A(5d)	δ(5d)	A(5e)	δ(5e)
1	0.0729	0.0687	- 6					(0.0651)	-11	(0.0593)	- 19
2	0.1568	0.1536	- 2					0.147	- 6	(0.136)	- 14
3	0.4244	0.4479	6					0.436	6	0.413	3
7	5.323	5.04	- 5	4.96	- 7	4.86	- 9				
8	6.748	6.54	- 3	6.42	- 5	(6.29)	- 7				
9	8.543	8.08	- 5	(7.92)	- 7	(7.74)	- 9				

[a]Extrapolated values in parentheses.

We have had occasion to calibrate three sets of GC columns for FA analysis. In each case the most reliable relationship has been the one we refer to as (5), the exponential,

$$(C_{ij}) = a(A_{ij})^b.$$

DISCUSSION

Investigators (6) use correction factors (7) (relative response factors, RRF) to convert GC output to concentrations. When relative response ratios are plotted against concentrations, a linear relation commonly is sought and is considered to be most desirable. Proportionality is assumed between the relative response ratios and the concentrations of the sample. However, in practice, sample-column and sample-instrument interactions may cause nonproportionality or even nonlinearity. The nonlinearity observed at low concentrations (opposed to high concentration sample overloading) can be compensated for, "since an actual calibration is established, eventual nonlinearity, support interaction, etc. does not reduce the accuracy of the determination because these also happen during the establishment of the calibration curve" (8). The RRF method is valid when used within the bounds of the linear dynamic range (LDR) of the system under investigation. The LDR must be established before the method can be applied for quantitative analyses. It limits the investigator to analysis of samples within the LDR. Extrapolation beyond the lower limits may cause serious errors in calculating concentrations (see Table III).

We conclude that the relationship

$$C_j = aA_j^b$$

provides the best fit for the data collected under conditions existing in our laboratory. The functional relationship is convenient in being: (a) analytical rather than graphical, (b) precise within the limits of the uncertainties of the data, and (c) suitable for handling by programmable desk calculators.

ACKNOWLEDGMENTS

Vermont Agricultural Experiment Station Journal Article No. 419. This study was supported by Vermont Agricultural Experiment Station project Hatch 227 and the Walker Research Fund.

REFERENCES

1. Cochrane, G.C., J. Chromatogr. Sci. 13:440 (1975).
2. Ackman, R.G., and R.D. Burgher, Anal. Chem. 35:647 (1963).
3. Cochrane, G.C., Proc. Soc. Anal. Chem. 10:212 (1973).
4. Gray, I.K., New Zealand J. Dairy Sci. Technol. 10:158 (1975).
5. Shatkay, A., and S. Flavian, Anal. Chem. 49:2222 (1977).
6. Harte, B.R., and C.M. Stine, J. Dairy Sci. 60:1266 (1977).
7. Bills, D.D., L.L. Khatri, and E.A. Day, Ibid 46:1342 (1963).
8. Ettre, L.S., "Practical Gas Chromatography," Perkin-Elmer, Norwalk, CT, 1973, pp. 6-29.

[Received March 5, 1979]