

Quantitative Analysis by Derivative Spectrophotometry (III) — Simultaneous quantitation of vitamin B group and vitamin C in by multiple linear regression analysis —

ManKi PARK and JungHwan CHO

Department of Pharmacy, College of Pharmacy, Seoul National
University, Seoul, 151-742, KOREA

(Received December 14, 1987)

Abstract □ The feature of resolution enhancement by derivative operation is linked to one of the multivariate analysis, which is multiple linear regression with two options, all-possible and stepwise regression. Examined samples were synthetic mixtures of 5 vitamins, thiamine mononitrate, riboflavin phosphate, nicotinamide, pyridoxine hydrochloride and ascorbic acid. All components in mixture were quantified with reasonably good accuracy and precision. Whole data processing procedure was accomplished on-line by the development of three computer programs written in APPLESOFT BASIC language.

Keywords □ Quantitation, Simultaneous; Derivative Spectrophotometry; Multiple Linear Regression; Computer Program.

Derivative spectra have been used with various purposes in quantitative analysis and qualitative analysis, showing great usability in many aspects. The applications are of wide variety, in the fields of UV-VIS, fluorescence, luminescence, and IR spectrophotometry. Since the first description on the application of derivative spectrophotometry in the region of ultraviolet in 1953, it has been applied extensively to qualitative and/or quantitative analysis.

The most attractive aspect of derivative operation is the enhancement of resolution of spectral features. In other words, derivative operation can resolve some peaks, which may represent the concentration of each component in a given mixture. In more detail, a series of peaks resolved by derivative operation can be related to the concentrations of interested compound in complex mixtures by all-possible or stepwise multiple regression analysis, which is well established mathematical and statistical theory for the multivariate analysis. A combination of derivative UV spectrophotometry and multiple regression analysis was examined. To accomplish this combination, the author developed three well organized computer programs, [D-UV-DATA] program which manipulates UV spectral data to prepare data sets for multiple regression analysis, [MULTIPLE] program which performs multiple regression analysis with two alternative options, all-possible or stepwise regression, and [D-UV-RSLT] program which calculates actual concentrations of interested component in given mixtures, using regression equation generated by

[MULTIPLE] program. Using these programs, derivative UV spectrophotometry could be applied to actual aspects of quantitative analysis.

THEORETICAL BACKGROUNDS

Multivariate analysis has been used extensively in various aspects of chemical, chromatographic and spectrophotometric systems, with the quantitative and even qualitative purposes. Multivariate statistics is a collection of mathematical tools that can be applied to a kind of analysis in which more than one measurement are acquired for each sample. Several multivariate methods such as multiple linear regression (4-12), principal component regression (12-23), partial least squares (12,24,25), rank annihilation (26,27), robust regression (28), multiple nonlinear regression (29) and weighted least squares (9,30) have been examined in relation with various system. Among these, least squares multiple regression and stepwise multiple regression were examined in connection with derivative UV spectrophotometry for the purpose of quantitative analysis. Its application was accomplished via following procedure.

At first, a series of standard mixtures were prepared and spectral data of these mixtures were acquired, using computer-controlled UV spectrophotometer. A set of spectral points (wavelengths) were selected to be used as independent variables for multiple regression analysis and then data set (matrix) was prepared from derivative

values at selected spectral points and manipulated to fit the algorithm of multiple regression program.

At second, using above-prepared data set, all-possible and stepwise multiple regression analysis were performed on the basis of least squares optimization. Data set subjected to multiple linear regression can be divided into two sets, data matrix (X) constructed from the derivative values at selected spectral points for a given set of standard mixtures and a matrix of concentration values (Y) of a given standard in mixtures. In estimation of Y from X, linear combination of variables in X is found, which minimizes the errors in reproducing Y. This means a relationship between X and Y such as $Y = XC + E$, where C is a matrix of regression coefficients and E is a matrix of errors associated with regression model. Matrix C is estimated by linear regression, and the term

$$\text{Error} = \sum_{k=1}^K \sum_{i=1}^I (y_{ik} - \hat{y}_{ik})^2 = \sum_{k=1}^K \sum_{i=1}^I \varepsilon_{ik}^2$$

is minimized, where \hat{y}_{ik} is the actual concentration element in the *i*th and *k*th column of Y, \hat{y}_{ik} is the regression estimate of the same element calculated using C, and ε_{ik} is the corresponding element of the matrix E. All-possible regression means that all of the matrix X are used to construct the model, regardless of whether or not it is relevant in describing the true model. Stepwise regression means that analysis is repeated by inserting variables of matrix X in turn until the model (regression equation) is satisfactory. The order of insertion is determined by using the partial correlation coefficient as a measure of the importance of variables not yet in the model.

At third, concentrations of a given component in unknown sample mixtures are calculated using above prepared model (regression equation) by replacing the variables with actual derivative values at corresponding wavelengths in the UV spectral data of unknown samples.

COMPUTER PROGRAMS

All computer programs developed or modified and used for this work were written in APPLE-SOFT BASIC. All programs are compiled to machine codes, using APPLESOFT BASIC Compiler.

[D-UV-DATA] Program

[D-UV-DATA] program prepares the data set which is used by [MULTIPLE] program, multiple linear regression program. This program calculates

derivative values and extracts a set of values at pre-selected spectral points and rearranges these values and concentration data to the form of input file for [MULTIPLE] program.

[START] module initializes array variables and reads convoluting integers for derivative operation. [MAIN] module determines the derivative order and convoluting block size and selects spectral points for data extraction. Spectral points can be selected at equal interval or one by one according to user's request.

[PREP] module controls the sequence of data preparation.

[CONC INPUT] module deals with input of concentration values of a certain component in mixtures and verification of entered values.

[COMMAND] module directs the flow of processing after data preparation as the user requests.

[VALUES] module extracts derivative values at preselected spectral points.

[DERIVATIVE] module calculates derivative values at all spectral points of spectrum.

[REGISTER] module registers the names of files for the spectra of standard mixtures, which will be used for modeling (construction of regression equation).

[RECALL] module reads spectral data of standard mixtures into memory.

[RECORD] module writes prepared data for regression on the floppy diskette with user-defined file name with the extension of ".MRD".

[MULTIPLE] Program

[MULTIPLE] program is made by wholesome modification of [STEPWISE] program made for IBM/PC series computer by Wolfe and Koelling (31). Modifications are of the file management method, result display pattern and mode, data processing sequence and some minor ones. Resulting regression equation can be recorded on the floppy diskette by additional module, which is not implemented in [STEPWISE] program. More detailed descriptions on the basic procedure of regression analysis is available in a text book written by Draper and Smith (32).

[START] module prepares array variables for data matrix.

[COMMAND] module directs the flow of processing and determines result display options, output device and whether result is printed at each iteration or not.

[REGRESSION] module performs multiple linear regression with two options, all-possible and stepwise regressions.

[ESTIMATE] module calculates regression estimates using matrix C.

[SAVE] module records resulting regression equation (matrix C) on the floppy diskette.

[REGISTER] module registers the name of file containing regression data set prepared by [D-UV-DATA] program.

[RECALL] module reads regression data set into memory.

[D-UV-RSLT] Program

[D-UV-RSLT] program calculates concentrations of a certain component in sample mixtures to be quantified using above-mentioned regression equation.

[START] module prepares array variables and reads convoluting integers for the derivative operation.

[COMMAND] module directs the flow of processing as user requests.

[MAIN] module controls the sequence of concentration calculation by calling subroutines in order, and prints out the result of calculation.

[DERIVATIVE] module performs derivative calculation with the derivative order and convoluting block size, which are represented in regression equation file.

[REGISTER] module registers the names of files for spectral data of sample mixtures and the regression equation file.

[SPEC RECALL] module reads the spectral data of sample mixtures to be quantified.

[EQ RECALL] module reads the regression equation to be used for the calculation of concentrations of sample mixtures.

EXPERIMENTAL METHODS

Instrument

UV spectrophotometer used for this work is Ultrospec 4050 of LKB Biochrome (England) connected to and controlled by APPLE][plus personal computer of APPLE Co. (U.S.A.) via RS-232c communication card. All data acquired are in digital form and recorded on floppy diskette in ASCII files. All programs are written in APPLE-SOFT BASIC and stored on floppy diskette.

Reagents

Hydrochloric acid buffer (pH 2.1) was prepared as prescribed in USP XX. Thiamine mononitrate, riboflavin phosphate, nicotinamide, pyridoxine hydrochloride and ascorbic acid were used as supplied by Sigma Co. (U.S.A.).

Acquisition of UV Spectral Data

All spectral data used in this work were acquired with 1 nm interval, using LKB Ultrospec 4050 controlled by APPLE][plus personal computer.

Multiple Linear Regression

Five vitamin standards used to construct standard mixtures were thiamine mononitrate, riboflavin phosphate, nicotinamide, pyridoxine hydrochloride and ascorbic acid. These vitamins were dissolved in hydrochloric acid buffer (pH 2.1). Twenty mixtures were prepared to contain from 2.0 to 9.6 $\mu\text{g/ml}$ of each of 5 vitamins with the interval of 0.4 $\mu\text{g/ml}$ between the mixtures. Each mixture contained a given vitamin of different concentration from those of other mixtures. Therefore, 20 mixtures contained 5 vitamins of different concentrations.

UV spectral data were acquired using computer-controlled UV spectrophotometer. Least squares polynomial smoothing was performed to these spectral data with the convoluting integers derived from transfer function for cubics and convoluting block size of 9 points (2). The selected wavelengths for data extraction were 226.0, 233.6, 250.1, 256.0, 267.7, 273.0, 275.0, 281.0, 291.5, 299.2nm, which are the maxima and minima of 1st derivative spectra of 5 vitamins found and calculated by [PEAK] module of [UV-SPECMAN] program (3). The 1st derivative values at these wavelengths were extracted and assembled with concentration values of a given vitamin in standard mixtures for multiple linear regression analysis by [D-UV-DATA] program. Using these data sets, multiple linear regression analysis was performed to construct regression equation for each vitamin in mixtures on the basis of all-possible and stepwise regression. Another 3 mixtures were prepared using above 5 vitamins of known concentrations. UV spectral data were acquired for these mixtures, also. And then these spectral data were manipulated by [D-UV-RSLT] program to fit the regression equation derived by above procedure and the resulting calculated concentrations of 5 vitamins were compared with the known prepared concentrations.

RESULTS AND DISCUSSIONS

Absorbance spectra of 5 vitamins are presented in Fig. 1. As shown in figure, all peaks of vitamins are overlapped with each other, which means that the concentrations of each vitamin in mixtures cannot be estimated only by the absorbance values at λ_{max} of each vitamin in mixtures with reasonable ac-

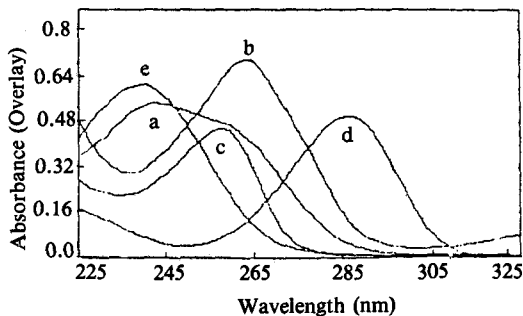


Fig. 1. Absorbance Spectra of 5 Vitamins

- a. thiamine mononitrate
- b. riboflavine phosphate
- c. nicotinamide
- d. pyridoxine hydrochloride
- e. ascorbic acid

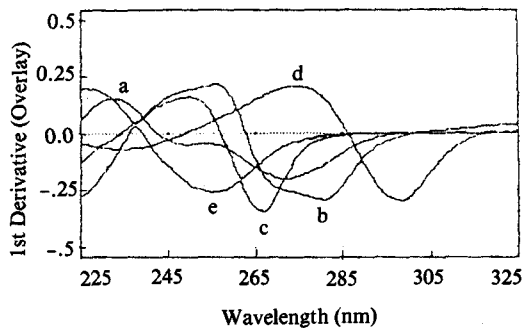


Fig. 2. 1st Derivative Spectra of 5 Vitamins

- a. thiamine mononitrate
- b. riboflavine phosphate
- c. nicotinamide
- d. pyridoxine hydrochloride
- e. ascorbic acid

curacy and precision. 1st derivative spectra of 5 vitamins are presented in Fig. 2. As shown in figure, minor spectral features are resolved relatively well, but multiple linear regression analysis is thought to be needed to relate the concentrations of vitamins in mixtures with the 1st derivative values.

All-Possible Regression

10 Wavelengths of characteristic maxima or minima in the 1st derivative spectra of 5 vitamins are incorporated in the regression equation for each vitamin in the standard mixtures by the concept of all-possible regression, which is shown in Table I. As shown in table, all coefficients of determination

Table I. Regression Equations derived by All-Possible Regression

Vitamin	Coefficient of Regression	STD Deviation of Estimate
Thiamine mononitrate	0.9997	0.061
Riboflavine phosphate	0.9997	0.072
Nicotinamide	0.9997	0.088
Pyridoxine HCl	0.9999	0.039
Ascorbic acid	0.9983	0.137

except that for ascorbic acid are greater than 0.999, which implies that each regression equation constructed thereby is well-corresponding to actual concentration of that vitamin in the mixtures. In the case of ascorbic acid, there is only one characteristic minimum point of derivative value in the selected spectral range, which resulted in the relatively poor coefficient of determination. Therefore, there should be as many characteristic wavelengths for each compound to be quantified as possible, to get the more accurate and precise model for the concentrations in mixtures. In this sense, it is reasonable to use the higher order derivative spectra for the quantitation of a certain compound in much more complex and complicated mixtures. But in this case, the high susceptibility of derivative operation to noises is the most serious limiting factor. Therefore, the proper order of derivative spectra should be selected in consideration of the efficiency of digital filtering for noise reduction and the degree of complexity of mixture to be quantified.

Stepwise Regression

As a different side of discussion, stepwise regression is a useful tool for the construction of the

Vitamin	Const.	226.0nm	233.6nm	250.1nm	256.0nm	267.7nm	273.0nm	275.0nm	281.0nm	291.5nm	299.2nm
Thiamine mononitrate	-3.74	-0.43	80.22	13.67	20.82	10.76	-29.45	25.58	-7.04	-22.69	-6.43
Riboflavine phosphate	3.27	15.46	-41.97	-5.16	1.75	-10.14	63.45	-76.20	15.34	-76.29	38.32
Nicotinamide	2.87	-0.44	-17.81	-1.39	5.09	-14.87	-8.86	-28.79	52.99	6.23	-0.09
Pyridoxine HCl	2.15	-8.45	12.62	0.63	0.18	4.11	-6.60	10.16	13.16	-37.20	-3.58
Ascorbic acid	38.05	-72.15	-118.91	-30.52	5.77	-5.62	40.03	-23.90	-9.65	26.59	68.36

Table II. Regression Equations derived by Stepwise Regression.

Vitamin	Const.	226.0nm	233.6nm	250.1nm	256.0nm	267.7nm	273.0nm	275.0nm	281.0nm	291.5nm	299.2nm
Thiamine mononitrate	-4.74	—	84.46	10.48	27.06	2.32	—	—	—	-36.63	—
Riboflavine phosphate	-0.49	—	-31.75	—	-14.95	—	—	—	-25.11	-21.25	—
Nicotinamide	0.17	-14.51	—	—	—	-20.77	—	-19.22	42.44	—	—
Pyridoxine HCl	-0.02	—	—	—	—	—	—	—	1.34	—	-32.61
Ascorbic acid	3.51	—	-48.58	-58.18	—	-22.00	—	—	—	—	—

Vitamin	Coefficient of Regression	STD Deviation of Estimate
Thiamine mononitrate	0.9993	0.067
Riboflavine phosphate	0.9987	0.088
Nicotinamide	0.9994	0.064
Pyridoxine HCl	0.9994	0.057
Ascorbic acid	0.9791	0.308

“best” regression equation. The regression equations for each vitamins are shown in Table II. As mentioned above in the section of theoretical backgrounds, the selection algorithm of stepwise regression excluded some variables from the final regression equations, which are represented with the sign of ‘-’ in table. These variables may not be significant in the aspect of model construction. But this does not mean that these variables are totally un-

necessary for the quantitation of vitamin concerned, but that these variables are excluded for the construction of the best equation if and only if the given data set is used, in which the informations on the concentrations of 5 vitamins are included. In other words, the excluded variables may be incorporated in regression equation for another mixture made of different set of constituents. As in the cases of all-possible regression, the coefficients of determination for the stepwise regression equations are greater than 0.999 except that for ascorbic acid, which is thought to be resulted from the same origin as mentioned for all-possible regression.

Application of Regression Equations

Results of quantitation are shown in Table III, which were obtained by the application of regression equations derived by two concepts of regression. Three mixtures were quantified. As shown in table,

Table III. Results of Calculation of Concentrations of 5 Examined Vitamins in Synthetic Mixtures by Two Regression Equation Types.

Vitamin		Mixture 1	Mixture 2	Mixture 3
Thiamine mononitrate	PREP	3.200	6.800	9.200
	ALL	3.205 (100.16)	6,841 (100.60)	9.211 (100.12)
	STEP	3.204 (100.13)	6.843 (100.63)	9.299 (101.08)
Riboflavine phosphate	PREP	4.800	8.400	2.800
	ALL	4.813 (100.27)	8.417 (100.20)	2.802 (100.07)
	STEP	4.736 (98.67)	8.373 (99.68)	2.813 (100.46)
Nicotinamide	PREP	5.200	3.200	4.000
	ALL	5.213 (100.25)	3.158 (98.69)	4.012 (100.30)
	STEP	5.266 (101.27)	3.127 (97.72)	4.024 (100.60)
Pyridoxine hydrochloride	PREP	6.800	4.800	5.600
	ALL	6.815 (100.22)	4.845 (100.94)	5.555 (99.20)
	STEP	6.854 (100.79)	4.884 (101.75)	5.592 (99.86)
Ascorbic acid	PREP	9.600	5.200	7.600
	ALL	9.500 (98.96)	5.229 (100.56)	7.613 (100.17)
	STEP	9.408 (98.00)	5.417 (95.83)	7.691 (101.20)

*‘PREP’ means prepared concentration in $\mu\text{g}/\text{ml}$.

‘ALL’ means concentration calculated by all-possible regression.

‘STEP’ means concentration calculated by stepwise regression.

*Numbers in parenthesis are recoveries in percentage.

the calculated concentrations are well-corresponding to the prepared concentrations within 1% error in a whole, which again means that regression equations were well-constructed. On the bases of these results, it is concluded that the feature of resolution enhancement of derivative operation can be successfully linked to the multivariate analysis for the purpose of the quantitative analysis of complicated mixtures.

CONCLUSION

Derivative UV spectrophotometry is successfully linked to multiple linear regression analysis by the development of computer programs. The whole processes of data acquisition and data manipulation and data analysis (multiple regression) and result presentation are linked on-line by the utilization of the auxiliary memory device, magnetic floppy diskette, on which all spectral data and intermediate data are recorded by the corresponding modules of computer programs made for this work. Using these programs, 5 vitamins in mixtures are quantified simultaneously with reasonably good accuracy and precision.

LITERATURES CITED

- Park, M.K., Cho, Y.H., and Cho, J.H.: Quantitative analysis by derivative spectrophotometry (I): Simultaneous quantitation of pyridoxine HCl and nicotinamide in the mixture. *Yakhak Hoeji*, **30**, 185 (1986).
- Park, M.K., and Cho, J.H.: Quantitative Analysis by Derivative Spectrophotometry (II): Derivative Spectrophotometry and Methods for the Reduction of High Frequency Noises. *Arch. Pharm. Res.*, **10**, 1 (1987).
- Park, M.K., Cho, J.H., and Park, Y.H.: Qualitative Analysis by Derivative UV Spectrophotometry (I): Identification of Penicillins and Cephalosporins. *SNU J. Pharm. Sci.*, in press (1987).
- Sternberg, J.C., Stillo, H.S., and Schwendeman, R.H.: Spectrophotometric Analysis of Multicomponent Systems Using the Least Squares Method in Matrix Form: The Ergosterol Irradiation System. *Anal. Chem.*, **32**, 84 (1960).
- Barnett, H.A., and Bartoli, A.: Least-Squares Treatment of Spectrometric Data. *Anal. Chem.*, **32**, 1153 (1960).
- Brown, C.W., Lynch, P.F., Obremski, R.J., and Lavery, D.S.: Matrix Representations and Criteria for Selecting Analytical Wavelengths for Multicomponent Spectroscopic Analysis. *Anal. Chem.*, **54**, 1472 (1982).
- Kisner, H.J., Brown, C.W., and Kavarnos, G.J.: Simultaneous Determination of Triglycerides, Phospholipids, and Cholesteryl Esters by Infrared Spectrometry. *Anal. Chem.*, **54**, 1479 (1982).
- Kisner, H.J., Brown, C.W., and Kavarnos, G.J.: Multiple Analytical Frequencies and Standards for the Least-Squares Spectrometric Analysis of Serum Lipids. *Anal. Chem.*, **55**, 1703 (1983).
- Haaland, D.M., Easterling, R.G., and Vopicaka, D.A.: Multivariate Least Squares Methods Applied to the Quantitative Spectral Analysis of Multicomponent Samples. *Appl. Spectro.*, **39**, 73 (1985).
- Birth, G.S.: Evaluation of Correlation Coefficients Obtained with a Stepwise Regression Analysis. *Appl. Spectro.*, **39**, 729 (1985).
- Burkhard, L.P., and Weininger, D.: Determination of Polychlorinated Biphenyls Using Multiple Regression with Outlier Detection and Elimination. *Anal. Chem.*, **59**, 1187 (1987).
- Beebe, K.R., and Kowalski, B.R.: An Introduction to Multivariate Calibration and Analysis. *Anal. Chem.*, **59**, 1007A (1987).
- Blackburn, J.A.: Computer Program for Multicomponent Spectrum Analysis Using Least-Squares Method. *Anal. Chem.*, **37**, 1000 (1965).
- Macnaughtan, D., Jr., Rogers, L.B., and Wernimont, G.: Principal-Component Analysis Applied to Chromatographic Data. *Anal. Chem.*, **44**, 1421 (1972).
- Ohta, N.: Estimating Absorption Bands of Component Dyes by Means of Principal-Component Analysis. *Anal. Chem.*, **45**, 553 (1973).
- Warner, I.M., Christian, G.D., Davidson, E.R., and Callis, J.B.: Analysis of Multicomponent Fluorescence Data. *Anal. Chem.*, **49**, 564 (1977).
- Painter, P.C., Rimmer, S.M., Snyder, R.W., and Davis, A.: A Fourier Transform Infrared Study of Mineral Matter in Coal: The Application of a Least Squares Curve-Fitting Program. *Appl. Spectro.*, **35**, 102 (1981).
- Sharaf, M.A., and Kowalski, B.R.: Quantitative Resolution of Fused Chromatographic Peaks in Gas Chromatography/Mass Spec-

- trometry. *Anal. Chem.*, **54**, 1291 (1982).
19. Sasaki, K., Kawata, S., and Minami, S.: Constrained Nonlinear Method for Estimating Component Spectra from Multicomponent Mixtures. *Appl. Opt.*, **22**, 3599 (1983).
 20. Hemel, J.B., van der Voet, H., Hindriks, F.R., and van der Slik, W.: Stepwise Deletion: A Technique for Missing-Data Handling in Multivariate Analysis. *Anal. Chim. Acta*, **193**, 255 (1987).
 21. Rutan, S.C., and Motley, C.B.: Factor Analysis and Kalman Filter Studies of Severely Overlapped Amino Acid Derivatives in Thin-Layer Chromatography. *Anal. Chem.*, **59**, 2045 (1987).
 22. Robert, P., Bertrand, D., Devaux, M.F., and Grappin, R.: Multivariate Analysis Applied to Near-Infrared Spectra of Milk. *Anal. Chem.*, **59**, 2187 (1987).
 23. Friedrich, H.R., and Yu, J.P.: Combinations of Orthogonal Spectra to Estimate Component Spectra in Multicomponent Mixture. *Appl. Spectro.*, **41**, 227 (1987).
 24. Lindberg, W., Persson, J.A., and Wold, S.: Partial Least-Squares Method for Spectrofluorimetric Analysis of Mixtures of Humic Acid and Ligninsulfonate. *Anal. Chem.*, **55**, 643 (1983).
 25. Frank, I.E., Kalivas, J.H., and Kowalski, B.R.: Partial Least Squares Solutions for Multicomponent Analysis. *Anal. Chem.*, **55**, 643 (1983).
 26. Ho, C.N., Christian, G.D., and Davidson, E.R.: Application of the Method of Rank Annihilation to Fluorescent Multicomponent Mixtures of Polynuclear Aromatic Hydrocarbons. *Anal. Chem.*, **52**, 1071 (1980).
 27. Gampp, H., Maeder, M., Meyer, C.J., and Zuberbuehler, A.D.: Quantification of a known Component in an Unknown Mixture. *Anal. Chim. Acta*, **193**, 287 (1987).
 28. Phillips, G.R., and Eyring, E.M.: Comparison of Conventional and Robust Regression in Analysis of Chemical Data. *Anal. Chem.*, **55**, 1134 (1983).
 29. Maris, M.A., Brown, C.W., and Lavery, D.S.: Nonlinear Multicomponent Analysis by Infrared Spectrophotometry. *Anal. Chem.*, **55**, 1694 (1983).
 30. De Levie, R.: When, Why, and How to Use Weighted Least Squares. *J. Chem. Educ.*, **63**(1), 10 (1986).
 31. Wolfe, P.M., and Koeling, C.P.: "BASIC Engineering and Scientific Programs." A Prentice-Hall publishing and Communications Company, 71 (1983).
 32. Draper, N.R., and Smith, H.: "Applied Regression Analysis". 2nd ed. John Wiley & Sons, Inc., New York (1981).