Talanta, Vol. 41, No. 1, pp. 59–66, 1994
Printed in Great Britain.

0039-9140/94 $6.00 + 0.00

SPECTROPHOTOMETRIC DETERMINATION OF MIXTURES OF 2-, 3-, AND 4-HYDROXYBENZALDEHYDES BY FLOW INJECTION ANALYSIS AND UV/VIS PHOTODIODE-ARRAY DETECTION

LARS NØRGAARD* and CARSTEN RIDDER†
Chemistry Department A, Building 207, The Technical University of Denmark, Lyngby, Denmark

(*Received* 22 *February* 1993. *Revised* 16 *June* 1993. *Accepted* 16 *June* 1993)

Summary—A multivariate approach to the quantitative determination of 2-hydroxybenzaldehyde (salicylaldehyde), 3-hydroxybenzaldehyde, and 4-hydroxybenzaldehyde in their mixtures is described. The method is based on second order data generated in a flow injection analysis system with a pH gradient and photodiode-array detection. Each injection gave rise to an 89 (times) × 101 (wavelength) matrix, containing both the acidic and the basic characteristics of the sample injected. A least-squares algorithm based on Lambert–Beers law was used for the prediction of concentrations in unknown samples. No assumptions concerning the qualitative mixture composition of the hydroxybenzaldehydes were necessary to perform concentration predictions. The following four data types were used in the least-squares modelling: (1) unfolded raw data, (2) acidic spectra, (3) basic spectra, and (4) first spectral derivative of the raw data. The prediction errors obtained were comparable to literature results. A graphic method, based on the model residuals for detecting erroneous samples, was developed.

Quantitative multicomponent analysis of samples containing compounds with severely overlapping spectra is a fundamental problem in spectrophotometric analysis. The problem of quantitative analysis can be divided into two classes: (a) all components in the sample are known, *i.e.* the compound spectra are attainable; and (b) unknown interferences are present in the sample. In this paper the focus will be on the former.

Several univariate approaches have been developed in order to perform quantitative analysis and resolution of mixtures of compounds with overlapping spectra. A method for the quantitative analysis of binary mixtures by derivative spectroscopy based on zero-crossing measurements has been presented by O'Haver and Green.[1,2] A method based on the first derivative of the ratio spectra, at each wavelength, between the absorption spectrum of the binary mixture and the absorption spectrum of a standard solution of one of the components, was developed by Salinas *et al.*[3] This method determines the concentration of the second component from a calibration graph at a single wavelength, as demonstrated in Refs 4, 5, and 6, where the quantitative analysis of the three possible binary mixtures of 2-hydroxybenzaldehyde (2HBA), 3-hydroxybenzaldehyde (3HBA), and 4-hydroxybenzaldehyde (4HBA) was performed. This approach has been further extended by a combination of Salinas' method and the zero-crossing method, to analyze ternary mixtures, as described in Ref. 6, where ternary mixtures of 2HBA, 3HBA, and 4HBA were resolved.

The two main disadvantages in the above approaches are (a) that the analyst has to know the mixture composition in advance, *i.e.* to analyze a binary mixture of hydroxybenzaldehydes it is necessary to know that the sample contains 2HBA and 3HBA and not 2HBA and 4HBA; and (b) the capability of detecting erroneous samples is lacking, *i.e.* if a sample in advance is assumed to contain 2HBA and 3HBA, but actually contains 2HBA and 4BHA, the results will be erroneous and there is no way of detecting this error. These disadvantages are inherent in the univariate approach.

By combining the versatile analytical technique flow injection analysis (FIA)[7] with photodiode-array (PDA) spectrophotometric

*Present address: Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Food Technology, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark.
†Author for correspondence.

detection, multivariate analytical signals are achieved. During the last decade the development of chemometric methods have facilitated the treatment of the vast amounts of information gathered in such FIA systems with PDA detection.[8,9]

In this work mixtures of hydroxybenzaldehydes are quantified by injection of the sample into an FIA system with a pH gradient. The pH gradient increases the information content about the samples, in exploiting the facts that hydroxybenzaldehydes possess both acidic and basic properties, and that the pK_a values are different. The pK_a values of 2HBA, 3HBA, and 4HBA are 6.79, 8.00, and 7.66, respectively.[10] A PDA spectrophotometer is applied to collect data produced by a sample injection. The produced data are used in a least-squares direct calibration method[11] based on unfolding of the sample matrices.[12] This method allows all mixture combinations to be analyzed, and moreover, erroneous samples can be detected looking at the model residuals.

EXPERIMENTAL

Apparatus

In all experiments a single-line manifold incorporating a six-port valve was used. The manifold and zone structure in the FIA system is shown in Fig. 1 and is generated by a six-port valve.[7] The sample is sandwiched by the carrier stream (low pH) and the reagent plug (high pH). In this way a reproducible pH gradient is created over the sample plug, provided that the sample volume is sufficiently small. The pH gradient is established by the use of a buffer system to give a smooth change from low to high pH.[13] All tubes employed were polypropylene (0.70 mm i.d.), which was used to avoid adsorption of the analytes in question. Both the carrier stream and the injected reagent are Britton–Robinson buffers with a pH of 4.5 and 11.4, respectively.

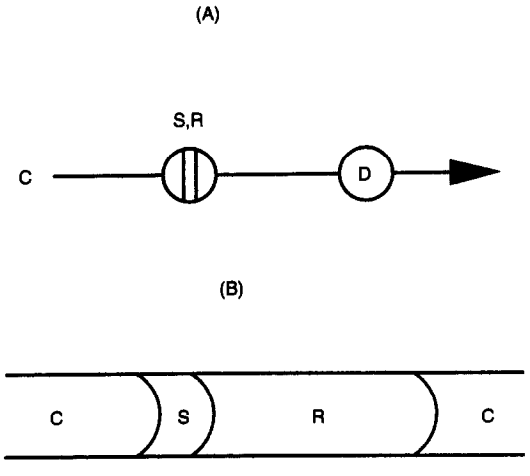

Fig. 1. (A) FIA manifold. C = carrier stream pumped at 0.375 ml/min (Britton–Robinson buffer, pH = 4.5); S = sample injection volume (77 μl); R = reagent injection volume (770 μl, Britton–Robinson buffer, pH = 11.4); and D = photodiode-array spectrophotometer. (B) Zone structure.

A Hewlett-Packard HP 8452A PDA spectrophotometer furnished with an 8-μl flow cell was used. The carrier stream was propelled by an ABU 80 autoburet from Radiometer A/S (0.375 ml/min). The automatic injection valve and the program for controlling it were developed in this laboratory.

The sample (77 μl) and the reagent solution (770 μl) are injected simultaneously by the six-port valve.[7] The PDA starts scanning 20 sec after the injection, and the scanning continues with a 1.0 sec interval in 88 sec. The wavelength range recorded in each scan is 250–450 nm (every 2 nm), *i.e.* one sample injection results in a data matrix consisting of 89 × 101 absorbances. The chosen scanning period ensures that the recorded spectra stem from both a pure basic and a pure acidic sample and all combinations in between these two extremes. The three-dimensional structure obtained by an injection of a 150.0 μM 2HBA solution is shown in Fig. 2. In Fig. 3, the pure acidic and the pure basic spectra corresponding to 2HBA, 3HBA, and 4HBA are depicted.

Reagents

Ethanol–water solutions were used in preparing the carrier, reagent, and standards, so that the final solutions were 1:9 in ethanol–water content. The Britton–Robinson buffer used in the carrier and reagent contained citric acid, potassium dihydrogenphosphate, boric acid, and tris-(hydroxymethyl)aminomethane (TRIS).[14] TRIS is used instead of 5,5-diethylbarbituric acid as given in the original recipe for the Britton–Robinson buffer, because 5,5-diethylbarbituric acid shows considerable absorption in the ultraviolet region, while TRIS does not possess this characteristic. To lower the blank signal to a minimum, the concentrations were chosen to be 1/8th of the one given in Ref. 14; *i.e.* the concentrations in the final carrier and reagent solutions were 1.788 mM for all four substances. The buffer should not be further diluted in order to ensure a suitably large

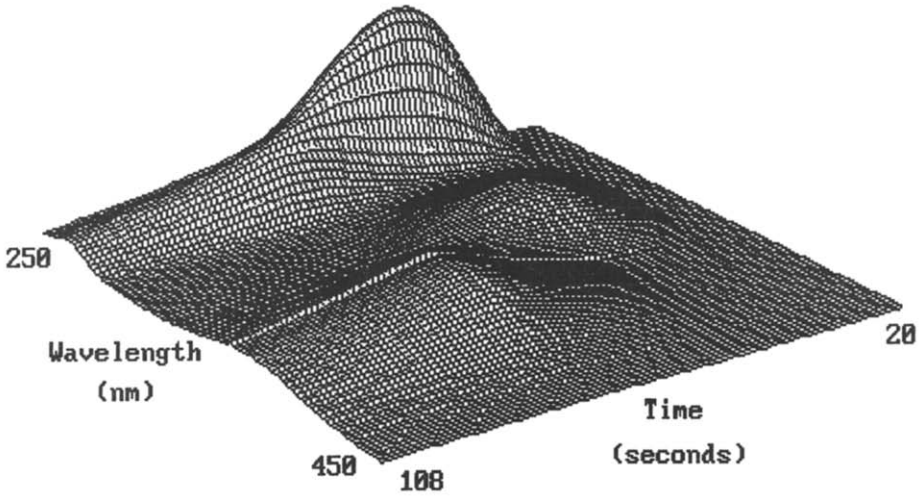

Fig. 2. Three-dimensional plot of a 150 μM 2HBA injection. Absorbances are in the range 0–0.6. Note the gradual transition from the acidic form to the basic form caused by the pH gradient.

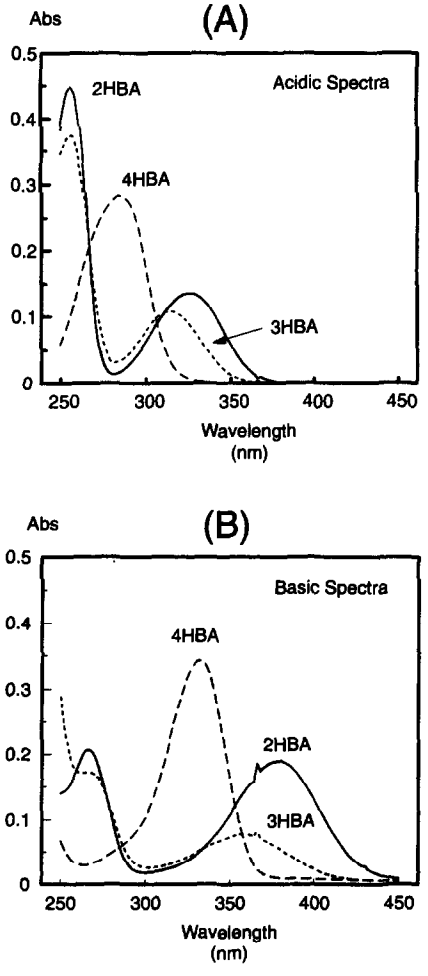

Fig. 3. (A) Pure acidic spectra of 2HBA, 3HBA and 4HBA. (B) Pure basic spectra of 2HBA, 3HBA and 4HBA.

buffer capacity. Furthermore, the reagent solution was made 12.5 mM in sodium hydroxide, to give a pH of 11.4. All standards were prepared from 1:1 ethanol–water 10 mM stock solutions of 2-hydroxybenzaldehyde (salicylaldehyde), 3-hydroxybenzaldehyde, and 4-hydroxybenzaldehyde.

Programs

The PDA spectrophotometer was connected to an IBM-compatible PC (80486 processor/50 MHz clock frequency) through HP 89531A UV–visible operating software, HP Part. No. 89531-9000 (Hewlett-Packard). Programs for file manipulations were made in Turbo Pascal version 6.0 (Borland International). 386-Matlab, MathWorks, Inc., was used for all calculations.

RESULTS AND DISCUSSION

The model used in describing the data is a simple linear least-squares model, often referred to as the direct linear calibration model.[11] The model requires knowledge of the pure data matrices of 2HBA, 3HBA, and 4HBA solutions with known concentrations, and assumes that Lambert–Beers law is obeyed. The pure data matrices are attained by sample injection into the FIA system followed by recording of the second-order data matrix, corresponding to each of the pure samples. By relating these three data matrices to the data matrix of an unknown

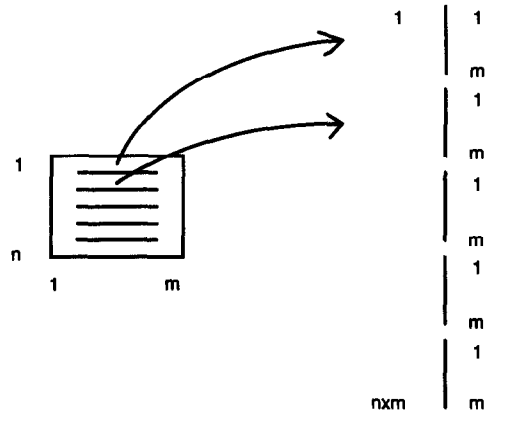

Fig. 4. The (n, m)-matrix is unfolded into a $(1, n \cdot m)$-column vector.

sample in a least-squares sense, it is possible to predict the respective concentrations simultaneously. A mathematical formulation is

$$\mathbf{S} = k_A \cdot \mathbf{2HBA} + k_B \cdot \mathbf{3HBA} + k_C \cdot \mathbf{4HBA} + \mathbf{E} \quad (1)$$

where **2HBA**, **3HBA**, and **4HBA** are the measured matrices of known concentrations, **S** is the unknown sample matrix, **E** is the error matrix (the size of which is minimized in the least-squares algorithm), and k_A, k_B, and k_C are scalars proportional to the concentrations of the hydroxybenzaldehydes in the unknown sample. The dimensions of all the given matrices are 89 (times) × 101 (wavelengths).

Previous to the least-squares calculation, the matrices are unfolded into vectors, *i.e.* the 89 × 101 matrices are unfolded into 8989 × 1 vectors, as depicted in Fig. 4. With the unfolded matrices the mathematical formulation is

$$S = k_A \cdot 2HBA + k_B \cdot 3HBA + k_C \cdot 4HBA + E \quad (2)$$

where S, $2HBA$, $3HBA$, $4HBA$, and E are now column vectors.
Equation (2) equals

$$S = (2HBA\ 3HBA\ 4HBA) \cdot \begin{pmatrix} k_A \\ k_B \\ k_C \end{pmatrix} + E$$

$$= \mathbf{HBA} \cdot \mathbf{k} + E \quad (3)$$

where **HBA** is a 8989 × 3 matrix and **k** is a 3 × 1 column vector.

The least-squares solution to equation (3) is given by

$$\mathbf{k} = (\mathbf{HBA}^T \cdot \mathbf{HBA})^{-1} \cdot \mathbf{HBA}^T \cdot S \quad (4)$$

where superscript^T means transposed.

The actual concentrations in an unknown sample are obtained by multiplying the known concentrations of the pure standards by the calculated proportional factors given by **k**.

Prediction by unfolding

The predicted concentrations in different mixtures are given in Table 1. It should be noted that all the predictions are based on only three pure standard injections, namely, a 150 μM 2HBA, a 150 μM 3HBA, and a 70 μM 4HBA solution. In the cases where the output from the least-squares calculation gave (very small) negative concentrations, the results are given as 0.0.

Table 1. True and predicted concentrations obtained by unfolding of the data matrices. The predictions are based on injections of 150 μM 2HBA, 150 μM, 3HBA, and 70 μM 4HBA solutions. The standard error of prediction is defined as $\mathrm{SEP} = \sqrt{\sum_{i=1}^{N} (C_i^{\mathrm{Predicted}} - C_i^{\mathrm{True}})^2 / N}$, where N is the number of predicted concentrations ($N = 15$)

	True			Predicted			True-predicted		
	2HBA	3HBA	4HBA	2HBA	3HBA	4HBA	2HBA	3HBA	4HBA
1	50	50	60	51.5	50.9	59.5	−1.5	−0.9	0.5
2	50	100	40	50.5	100.5	40.4	−0.5	−0.5	−0.4
3	50	100	60	49.3	99.5	58.9	0.7	0.5	1.1
4	100	50	40	100.8	51.2	41	−0.8	−1.2	−1.0
5	100	100	40	101.9	93.3	41.6	−1.9	6.7	−1.6
6	100	100	60	97.4	99.3	58.6	2.6	0.7	1.4
7	0	100	40	0	103.6	40.7	0	−3.6	−0.7
8	0	100	60	0	107.5	58.3	0	−7.5	1.7
9	50	0	60	48.8	3.8	60.7	1.2	−3.8	−0.7
10	100	0	60	97.3	3.4	59.6	2.7	−3.4	0.4
11	50	100	0	49.9	102.5	0	0.1	−2.5	0
12	100	50	0	101.4	49.5	0	−1.4	0.5	0
13	80	0	0	82.1	0	0.3	−2.1	0	−0.3
14	0	80	0	0	82.9	0.3	0	−2.9	−0.3
15	0	0	40	0	1.1	40.9	0	−1.1	−0.9
						Bias:	−0.9	−19.0	−0.8.
						SEP:	1.39	3.26	0.90.

Table 2. Predicted concentrations by use of the pure acidic spectra (measured at 20 sec after scan start). SEP is defined in Table 1

	True			Predicted			True-predicted		
	2HBA	3HBA	4HBA	2HBA	3HBA	4HBA	2HBA	3HBA	4HBA
1	50	50	60	59.1	41.3	59.9	−9.1	8.7	0.1
2	50	100	40	61	74.7	39.6	−12.0	25.3	0.4
3	50	100	60	64.4	66.9	57	−14.4	33.1	3.0
4	100	50	40	111.3	25.3	39.8	−11.3	24.7	0.2
5	100	100	40	124.3	49.6	42.7	−24.3	50.4	−2.7
6	100	100	60	119.3	41.8	55.4	−19.3	58.2	4.6
7	0	100	40	1.7	99.4	41.6	−1.7	0.6	−1.6
8	0	100	60	4.3	88.8	56.5	−4.3	11.2	3.5
9	50	0	60	53.5	2.1	62.8	−3.5	−2.1	−2.8
10	100	0	60	105.8	0	58.9	−5.8	0	1.1
11	50	100	0	52.7	89.7	0	−2.7	10.3	0
12	100	50	0	107	45.9	0.1	−7	4.1	−0.1
13	80	0	0	82.7	8.4	0	−2.7	−8.4	0
14	0	80	0	0	87.3	0	0	−7.3	0
15	0	0	40	0	2.5	41.9	0	−2.5	−1.9
						Bias:	−118.1	206.3	3.8.
						SEP:	10.51	24.13	2.08.

As seen from Table 1 the prediction errors are low, and comparable to the errors presented in Ref. 6. Some of the samples 7–15 are analyzed to contain a few millimoles of a substance used in the least-squares calculation, but not present in the sample. To decide whether a substance is present or not, in the case of a low concentration prediction, is it necessary with further investigation of the system with regard to detection limits. This can be a complicated matter and is beyond the scope of this work.

Prediction by using the acidic or the basic spectrum

To investigate the effect of reducing the information content of each sample, predictions are made by using only one spectrum at a given time. Equation (3) is still valid, but instead of unfolded matrices, *2HBA*, *3HBA*, and *4HBA* are simply vectors containing the spectrum at the chosen time. The results obtained by using a single acidic spectrum (at 20 sec after scan start), and a single basic spectrum (at 65 sec after scan start) are shown in Tables 2 and 3, respectively. The predictions obtained by using the basic spectrum are comparable to the predictions with unfolded matrices, while the acidic predictions give somewhat larger errors. This corresponds with the results given in Ref. 6, where a model based on acidic spectra gave unacceptable predictions.

Derivative spectrophotometry

Derivative spectrophotometry often improves concentration predictions, when dealing with mixtures and univariate modelling.[1,2] In order to

Table 3. Predicted concentrations by use of the pure basic spectra (measured at 65 sec after scan start). SEP is defined in Table 1

	True			Predicted			True-predicted		
	2HBA	3HBA	4HBA	2HBA	3HBA	4HBA	2HBA	3HBA	4HBA
1	50	50	60	48	47.3	58.2	2.0	2.7	1.8
2	50	100	40	49.8	98.6	39.9	0.2	1.4	0.1
3	50	100	60	46.4	93.9	57.3	3.6	6.1	2.7
4	100	50	40	100.8	47.6	41.1	−0.8	2.4	−1.1
5	100	100	40	96.2	92.3	39.7	3.8	7.7	0.3
6	100	100	60	97.2	91.4	57.2	2.8	8.6	2.8
7	0	100	40	0	100.4	39.4	0	−0.4	0.6
8	0	100	60	0	100.9	57.7	0	−0.9	2.3
9	50	0	60	47.6	0	59.7	2.4	0	0.3
10	100	0	60	96	0	59.2	4.0	0	0.8
11	50	100	0	51.9	102.1	0	−1.9	−2.1	0
12	100	50	0	98.5	49.4	0	1.5	0.6	0
13	80	0	0	81.1	3.1	0.7	−1.1	−3.1	−0.7
14	0	80	0	0.6	87.9	0.4	−0.6	−7.9	−0.4
15	0	0	40	0.5	0	41.5	−0.5	0	−1.5
						Bias:	15.4	15.1	8.0.
						SEP:	2.15	4.19	1.39.

Table 4. Predicted concentrations based on first-order derivative unfolded matrices. SEP is defined in Table 1

	True			Predicted			True-predicted		
	2HBA	3HBA	4HBA	2HBA	3HBA	4HBA	2HBA	3HBA	4HBA
1	50	50	60	52.6	49	60.1	−2.6	1.0	−0.1
2	50	100	40	51.9	97.7	41.3	−1.9	2.3	−1.3
3	50	100	60	49.7	96.5	59.5	0.3	3.5	0.5
4	100	50	40	101.6	49.8	42.1	−1.6	0.2	−2.1
5	100	100	40	97.7	94.9	42.2	2.3	5.1	−2.2
6	100	100	60	96.9	93.7	59.5	3.1	4.3	0.5
7	0	100	40	1.8	100.3	41.5	−1.8	−0.3	−1.5
8	0	100	60	1.3	98.8	59.2	−1.3	1.2	0.8
9	50	0	60	52.1	0	61.3	−2.1	0	−1.3
10	100	0	60	100.6	0	60.4	−0.6	0	−0.4
11	50	100	0	51.8	100.7	0.3	−1.8	−0.7	−0.3
12	100	50	0	100.3	50.8	0	−0.3	−0.8	0
13	80	0	0	82.3	0.6	0	−2.3	−0.6	0
14	0	80	0	0.6	82.8	0.1	−0.6	−2.8	−0.1
15	0	0	40	1	0	41.1	−1.0	0	−1.1
						Bias:	−12.2	12.4	−8.6.
						SEP:	1.78	2.22	1.08.

investigate the influence of first-order derivation, each spectrum in each data matrix was differentiated by the Savitzky–Golay algorithm.[15] The first derivatives were calculated for the 89 spectra individually using five experimental points. The subsequent data treatment was equal to the data treatment of the raw data matrices, that is, unfolding followed by least-squares predictions. The prediction results are given in Table 4, and the total standard error of prediction (SEP) values for all the methods investigated are given in Table 5. It is seen that the SEP value for the acidic spectra predictions is larger than the others by several orders of magnitude. The lowest SEP value is obtained by the first-order derivative method, however, no firm conclusions can be made regarding which method is the best in a predictive sense.

Erroneous samples

A very important feature of the multivariate approach is the ability of detecting outliers among the samples. When the **k** vector is determined for a given unknown sample, it is possible to reconstruct the data matrix of the unknown sample by

$$\mathbf{S}_{rec} = \mathbf{HBA} \cdot \mathbf{k} \tag{5}$$

where subscript rec means reconstructed.

Table 5. Total standard error of prediction (SEP) for all the methods investigated. The SEP values are in each case based on the total number of predicted concentrations ($N = 3 \times 15 = 45$)

	Raw data	Acidic spectra	Basic spectra	First derivative
SEP_{total}	2.11	15.24	2.86	1.76

In the ideal case of no measurement noise and no interfering species present in the sample, the difference between **S** and $\mathbf{S}_{rec}$ would exactly be a zero matrix. In the case of measurement noise and no interfering species present in the sample, the difference matrix would contain low numerical values reflecting the random variance in the measurement noise, and finally in the case of both measurement noise and the presence of interfering species, the difference matrix would show large systematic deviation from the zero matrix.

By utilizing the above-mentioned facts, it is possible to discover outlying samples. Assuming that we only want to resolve binary mixtures of *2HBA* and *3HBA*, *i.e.* only *2HBA* and *3HBA* standards are used in the prediction of unknown samples. If, by accident, the analyzed sample contains *4HBA*, univariate methods would give a wrong result without warning. By looking at the residuals ($\mathbf{S} - \mathbf{S}_{rec}$) for a good sample and for an outlying sample (Fig. 5), it is seen that the fluctuations in the residuals stemming from the latter are very large. A quantitative measure of the fit is the total sum of squares (SS) of the residuals:

$$SS_E = \sum_{i=1}^{89} \sum_{j=1}^{101} E_{ij}^2, \tag{6}$$

where $\mathbf{E} = \mathbf{S} - \mathbf{S}_{rec}$.

The concentrations calculated for the erroneous samples are useless, but the analyst is warned. It should be noted that this residual check is also applicable to vectorial data.[9]

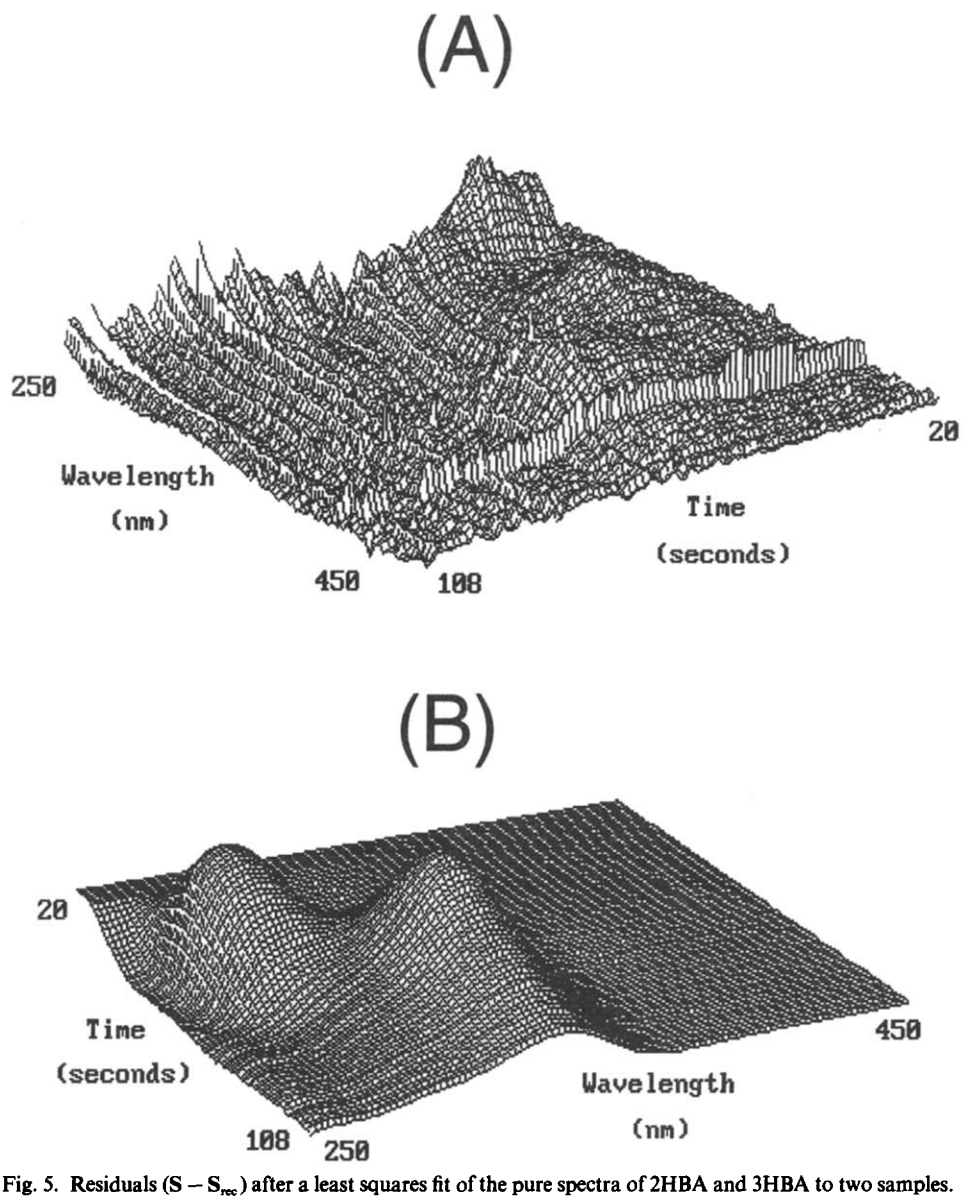

Fig. 5. Residuals ($S - S_{rec}$) after a least squares fit of the pure spectra of 2HBA and 3HBA to two samples. (A) Residuals from sample #12, which is fitted with the correct number of pure analyte spectra. The absorbances are in the range from -0.006 to $+0.008$, and they are small and random ($SS_E = 0.03$). (B) Residuals from sample #3, containing both 2HBA, 3HBA and 4HBA. The absorbances are in the range -0.16–0.36 and they are large and systematic ($SS_E = 70.13$).

CONCLUSIONS

A simple least-squares method for second-order FIA data is used in the quantitative analysis of mixtures of 2-, 3-, and 4-hydroxybenzaldehydes. The method is capable of detecting erroneous samples and of analyzing the mixtures without any previous knowledge of the mixture composition.

In a later paper[16] we investigate the method of Rank Annihilation Factor Analysis (RAFA) in order to perform concentration predictions based on two injections only.

Acknowledgement—The Danish Council for Industrial and Scientific Research is acknowledged for financial support.

REFERENCES

1. T. C. O'Haver and G. L. Green, *Anal. Chem.*, 1976, **48**, 312.
2. T. C. O'Haver, *Clin. Chem.*, 1979, **25**, 1548.
3. F. Salinas, J. J. Berzas Nevado and A. Espinosa Mansilla, *Talanta*, 1990, **37**, 347.
4. J. J. Berzas Nevado, C. Guiberteau Cabanillas and F. Salinas, *Anal. Lett.*, 1990, **23**, 2077.
5. J. J. Berzas Nevado, C. Guiberteau Cabanillas and F. Salinas, *Bull. Soc. Chim. Belg.*, 1991, **100**, 79.

6. J. J. Berzas Nevado, C. Guiberteau Cabanillas and F. Salinas, *Talanta*, 1992, **39**, 547.
7. J. Ruzicka and E. H. Hansen, *Flow Injection Analysis*, 2nd. ed. Wiley, New York.
8. I. Lukkari and W. Lindberg, *Anal. Chim. Acta*, 1988, **21**, 1.
9. C. Ridder and L. Nørgaard, *Chemometr. Intell. Lab. Syst.*, 1992, **14**, 297.
10. *Handbook of Biochemistry*, (H. A. Sober (ed.)) The Chemical Rubber Co., Cleveland, Ohio, 1968.
11. H. Martens and T. Naes, *Multivariate Calibration*. Wiley, New York, 1989.
12. S. Wold, P. Geladi, K. Esbensen and J. Öhman, in: Nordisk Symposium i Anvendt Statistik (N. E. Raun, ed.) pp. 249–277. Danmarks edb-center for forskning of uddannelse, København, 1986 (in English).
13. L. Nørgaard, *Anal. Chim. Acta*, 1991, **255**, 143.
14. D. D. Perrin and B. Dempsey, *Buffers for pH and Metal Ion Control*. Chapman and Hall, London, 1974.
15. A. Savitzky and M. J. E. Golay, *Anal. Chem.*, 1964, **36**, 1627.
16. L. Nørgaard and C. Ridder, *Proc. of the 3rd Scandinavian Symposium on Chemometrics*, June 1993, *Chemometr. Intell. Lab. Syst.* (submitted).