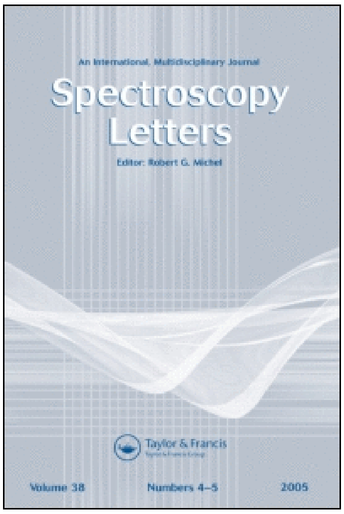

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

MULTIVARIATE QUALITY CONTROL APPLIED TO DETECT THE SOYBEAN OIL OXIDATION USING FOURIER TRANSFORM INFRARED SPECTROSCOPY

Rosangela C. Barthus[a]; Ronei J. Poppi[a]
[a] Instituto de Química, Universidade Estadual de Campinas, Campinas, Brazil

To cite this Article Barthus, Rosangela C. and Poppi, Ronei J.(2002) 'MULTIVARIATE QUALITY CONTROL APPLIED TO DETECT THE SOYBEAN OIL OXIDATION USING FOURIER TRANSFORM INFRARED SPECTROSCOPY', Spectroscopy Letters, 35: 5, 729 — 739

To link to this Article: DOI: 10.1081/SL-120014943

URL: http://dx.doi.org/10.1081/SL-120014943

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

SPECTROSCOPY LETTERS
Vol. 35, No. 5, pp. 729–739, 2002

MULTIVARIATE QUALITY CONTROL APPLIED TO DETECT THE SOYBEAN OIL OXIDATION USING FOURIER TRANSFORM INFRARED SPECTROSCOPY

Rosangela C. Barthus and Ronei J. Poppi*

Universidade Estadual de Campinas, Instituto de
Química C. P. 6154,13083-970, Campinas, Brazil

ABSTRACT

Multivariate quality control was used in conjunction with Fourier transform infrared spectroscopy to detect the oxidation process in soybean oils. The multivariate approach was based on principal component analysis (PCA) to develop two control charts: a T^2 chart using the most significant principal components and a Q chart with the remaining PCs not used in the first one. From these two charts it was possible to identify oxidation levels in oil samples which were out of the pattern normally present in edible soybean oils.

---

*Corresponding author. E-mail: ronei@iqm.unicamp.br

DOI: 10.1081/SL-120014943
0038-7010 (Print); 1532-2289 (Online)

www.dekker.com

Key Words: Multivariate quality control; PCA; Vegetable oils

INTRODUCTION

The oxidation of lipids is a spontaneous and inevitable phenomenon. It is a major deteriorative reaction affecting edible fats and oils and several factors can contribute to this process, such as oil type and oxidation conditions. This process can be divided into initiation, propagation and termination steps and several products are formed in these three steps such as hydroperoxides, alcohols, aldehydes, hydrocarbons and free fatty acids with esters, ketones, lactones, furans and other minor products.[1]

This oxidation process can be characterized by numerous methods based on physical, chemical or physical-chemical monitoring. The American Oil Chemists' Society (AOCS) has a number of official methods to provide an indication of the oxidative status of an oil,[2] the most common being the peroxide value (PV), the anisidine value (AV) and the thiobarbituric acid value. These chemical methods measure basically the primary or secondary products of oxidation, such as hydroperoxides or carbonyl-type compounds.

In recent years, research has been developed to apply Fourier transform infrared spectroscopy (FTIR) in quality control methods for the food industry.[3] Methods for the determination of iodine value,[4] and free fatty acids[5] in vegetable oils have been developed. Fourier Transform Infrared spectroscopy is a suitable tool to perform these determinations because it is generally rapid, can be automated and can reduce the need for solvents and toxic reagents associated with wet chemical methods for fats and oils analyses.

In this work, an approach based on Fourier Transform Infrared Spectroscopy in conjunction with Multivariate Quality Control was developed to monitor the quality of soybean oil and to verify the existence of deteriorated edible oil samples. No quantitative analysis was performed as no specific primary or secondary products of oxidation were monitored. The multivariate approach was based on principal component analysis (PCA) of the oil spectra in the infrared region (Figs. 1 and 2). These spectra were used to develop two control charts: a T^2 chart using the most significant principal components and a Q chart with the remaining PC not used in the first one. From these two charts it was possible to identify oxidation levels in oil samples which were out of the pattern normally present in edible soybean oils.

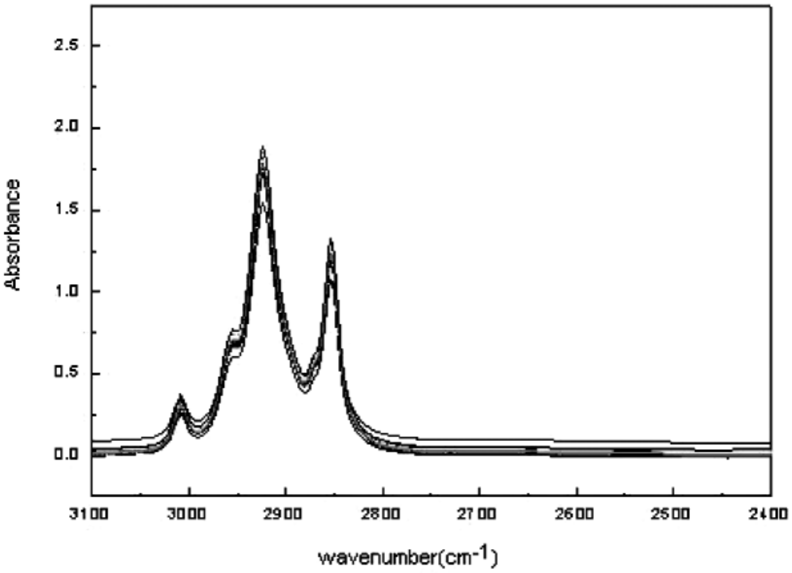

Figure 1. Infrared spectra of the soybean oils in the range of 3100 to 2400 cm^{-1}.

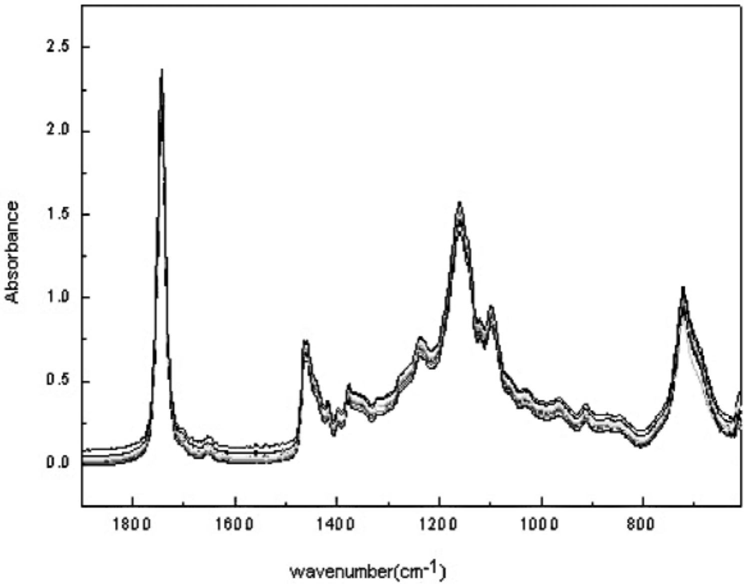

Figure 2. Infrared spectra of the soybean oils in the range of 1900 to 600 cm^{-1}.

MULTIVARIATE QUALITY CONTROL

Normally in quality control, Shewart charts[6] are used to monitor a small number of variables to detect an event having a 'special cause' that makes the process out-of-control. This special cause can be found and eliminated, helping to improve the process and the product quality.

A univariate control chart is constructed based on a plot of sample number or time on the abscissa and three important values on the ordinate: mean value $\bar{x}$ and control limits obtained from standard deviation which are $\pm 2\sigma$ (95% confidence regions) or $\pm 3\sigma$. The process is out-of-control when there is one observation outside the control limit or two observations near a control limit, called warning limit ($\pm 2\sigma$). The out-of-control limit is characterized when there are unusual or nonrandom patterns in the data. The control limits are usually determined by analyzing the variability in a reference set of process data collected when variabilities from only 'normal' or 'common' causes are present.

A disadvantage in the use of univariate control charts appears when there are a great number of variables to be monitored. In this case "out of control" samples can be missed due correlations in the data set. Also, many times it is impractical to control a process with univariate control charts when many variables are involved. To overcome this problem multivariate control charts have been used.[7–9]

In the multivariate approach, two control charts, which are based on principal component analysis (PCA),[10] are used. Principal Component Analysis (PCA) is a tool for data compression and information extraction. PCA finds combinations of variables that describe major trends in the data. The objective of PCA is to obtain the significant information of a process using fewer factors from the original variables. In effect, PCA decomposes a matrix **X** (matrix obtained from sample spectra in rows and measure of absorbance in several wavelengths in columns) as:

$$\mathbf{X} = \mathbf{T}_{\mathrm{K}}\mathbf{P}_{\mathrm{K}}^{\mathrm{T}} + \mathbf{E} = \sum_{\mathrm{i}=1}^{k} \mathbf{t}_{\mathrm{i}}\mathbf{p}_{\mathrm{i}}^{\mathrm{T}} + \mathrm{E} \tag{1}$$

where **T** is the matrix of scores, **P** the matrix of loadings, **E** is a residual matrix and K is the number of principal components included in the model. Sometimes a pretreatment of data is necessary before the principal component analysis can be carried out.

The first step in building the multivariate control chart is the determination of the number of principal components, which will be used in each chart. The method adopted was based on the evaluation of the eigenvalues of the **X** matrix for each principal component. The eigenvalues explain the quantity of variance explained by each principal component and a number of principal components were chosen that explain practically all the system variance.

The significant principal components are used to construct the T^2 Chart and the remaining principal components are used to construct the Q Chart.

The T^2 has its origin in the work of Hotelling's and is estimated from the scores of PCA in a space formed by the most significant principal components. The T^2 is a measure of the variation in each sample within the PCA model and is defined as [7]:

$$T_i^2 = \mathbf{t}_i \lambda^{-1} \mathbf{t}_i^T = \mathbf{x}_i \mathbf{P}_K \lambda^{-1} \mathbf{P}_K^T \mathbf{x}_i^T \tag{2}$$

where $\mathbf{t}_i$ refers to the ith row of $\mathbf{T}_K$ (the matrix of K scores vectors from the PCA model), λ is the diagonal matrix containing the eigenvalues associated with the eigenvectors included in the PCA model, $\mathbf{x}_i$ is the ith row in **X** and $\mathbf{P}_K$ is the matrix of the K loadings vectors retained in the PCA model.

The confidence limit for T^2 is calculated by the F distribution and is given as:

$$T^2 = \frac{(m-1)K}{m-K} \cdot F_{K,m-K,\alpha} \tag{3}$$

where m is the number of samples utilized to develop a PCA model, K is the number of significant principal components and α corresponds to a confidence limit region. The T^2 limit defines a plane within which the data are normally projected.

The Q chart corresponds to a lack of fit statistic for PCA models and is simply the sum of squares of the residual matrix (**E**) of each row (sample) and it is defined as:

$$Q_i = \mathbf{e}_i \mathbf{e}_i^T = \mathbf{x}_i (\mathbf{I} - \mathbf{P}_K \mathbf{P}_K^T) \mathbf{x}_i^T \tag{4}$$

where $\mathbf{e}_i$ is the ith row of **E** and **I** is the identity matrix.

Q is a scalar that measures the amount of variation not accounted for by the PCA model. If the process is controlled, the Q values must be small. The confidence limit can be calculated according to:

$$Q\alpha = \Theta_1 \left[\frac{c_\alpha \sqrt{2\Theta_2 h_o^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0 (h_o - 1)}{\Theta_1^2} \right]^{\frac{1}{h_o}} \tag{5}$$

$$\Theta_i = \sum_{j=K+1}^{A} \lambda_j^i \quad \text{for } i = 1, 2, 3 \tag{6}$$

$$h_o = 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2} \tag{7}$$

where c_α in Eq. 5 is the normal standard deviation corresponding to the upper (1-α) percentile. In Eq. 6, K is the number of principal components retained in the model, A is the total number of principal components and $\lambda_j$is the eigenvalue associated with the jth principal component.

After the development of multivariate control charts from samples under control, new multivariate observations of new samples can be projected onto a plane defined by the PCA loading vectors to obtain their new scores, T^2 and Q. If this new values of T^2 and Q are inside the confidence limit in the multivariate control charts the sample is ‘‘in control’’, otherwise the sample is ‘‘out of control’’.

EXPERIMENTAL

Sample Set

In this work, the oxidative processes of soybean oil were monitored. For this purpose, thirty-five commercial soybean oils and one rice oil were used. The oil samples came from different regions of cultivation and different epochs of the growing year to guarantee a certain heterogeneity among the samples.

These oils were split into 2 groups; one group with twenty-one soybean oil samples was used to build the multivariate control charts. These samples represent fresh oils or samples analyzed before the expiration date and the normal variation between samples is expected to be extracted from this group. The other group with fifteen samples had eight fresh soybean oil samples, four samples of soybean oil heated at different temperatures to induce oxidation: 90° for 4 h, 120°C for 1 h (two samples from different suppliers) and 200°C for 40 min, one soybean oil used for several days, one sample that was kept in a refrigerator for eight months and one rice oil. This last group was used to validate the control charts developed.

Also, univariate control charts were developed to compare the results with the multivariate charts.

Instrumentation

The instrument used in this work was a Nicolet 520 FTIR. To minimize water and CO_2 interference, the instrument was purged with nitrogen. A horizontal ATR sampling accessory equipped with a ZnSe crystal was used to obtain the spectra of the oil samples in neat form. For the ATR data acquisition, 1 mL of oil sample was pipetted onto the crystal and its spectra was recorded. An air spectrum was used as reference in absorbance calculations and they were collected from 254 scans with $4\,cm^{-1}$ resolution.

In the development of multivariate control charts spectra regions were used which represent characteristic bands such as 3100 to $2400\,cm^{-1}$ and 1900 to $600\,cm^{-1}$, as shown in Figs. 1 and 2. To build the univariate control charts variables were used representing the most important characteristic absorbances of the vegetable oils: $2923.5\,cm^{-1}$ (asymmetric aliphatic C-H stretch), $2852.2\,cm^{-1}$ (symmetric aliphatic C-H stretch), $1743.3\,cm^{-1}$(C=O stretch in ester), $1151.1\,cm^{-1}$ (C-O stretch in ester) , $719.4\,cm^{-1}$ (=C-H in plane bending).

The control charts were made using the Matlab 5.2 for Windows and the PLS Toolbox for use with Matlab,[11] with mean centered data.

RESULTS AND DISCUSSION

The first step in building multivariate control charts is the correct determination of the number of significant principal components. This number was determined from an eigenvalue vs. principal component plot, as shown in Fig. 3. From Fig. 3, is possible to see that three principal components are an adequate choice because the eigenvalue is low and no appreciated variation is observed after this number. By using three principal components, 96.53% of data variance was explained.

After the determination of the adequate number of principal components, the multivariate control charts could be constructed. The T^2 chart used the most significant principal components (three in this case) and the remainders were used to build the Q chart. In the calculations of T^2 and Q Eqs. 2 and 4 were used and the control limit for these multivariate control charts was found using the Eqs. 3 and 5.

The T^2 chart and the Q chart were developed based on data obtained from fresh oils (first group), where only common causes of variations were present. The control limit for the T^2 chart was 10.5 and, for the Q chart, it was 0.019 with 95% confidence. All samples used in the development of the model fell inside the control limits, denoting no abnormal variations among the oils.

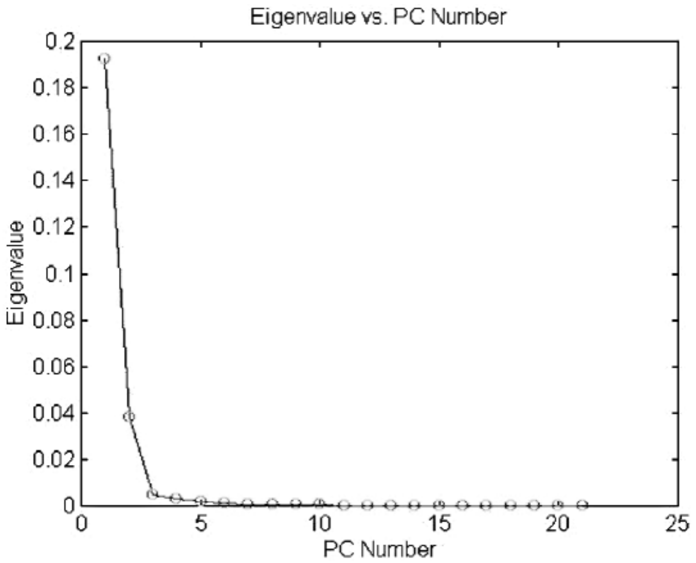

Figure 3. Number of principal components vs. eigenvalue.

The behavior of the other fifteen samples (second group), including the oxidized samples, rice oil and the oil kept in the refrigerator, could be referenced against the ''in control'' model. The samples of the second group were projected onto loading vectors found with the model made from the samples of the first group (''in control'' model) and their scores, Q and T^2 were obtained.

The T^2 and Q value could be compared with control limits (95% confidence) established in the multivariate control charts and, therefore, ''in control'' and ''out of control'' samples could be visualized in these plots. These results are shown in Figs. 4 and 5. A sample is considered ''out of control'' when its T^2 or Q values were found outside of the control limit. In Fig. 4, the T^2 chart shows that samples 5, 6, 8, 10 and 11 are outside control limits and the Q chart (Fig. 5) shows that samples 5 to 11 are outside of the control limits. The samples numbered 5 to 11 correspond to the deteriorated oils and the rice oil. Sample number 5 corresponds to the rice oil, the samples 6, 8, 9 and 11 correspond to oils submitted to oxidation at different temperatures, sample number 7 is the used soybean oil and sample number 10 is the soybean oil kept in refrigerator after the package was opened. The others samples, 1, 2, 3, 4, 12, 13, 14 and 15 represent fresh oils and they are ''in control'', as expected. It is important to note that samples 7 and 9, classified as ''in control'' on the T^2 chart, but ''out of control'' on the Q

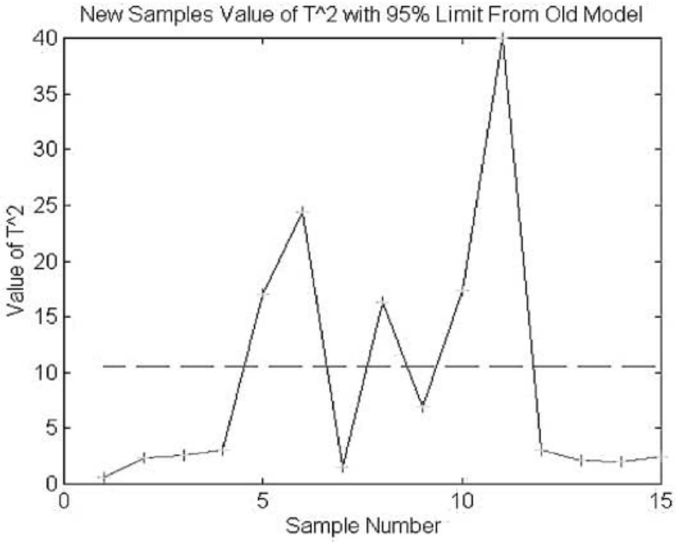

Figure 4. T^2 chart.

chart are less oxidized oils. In sample number 9 the temperature of oxidation was 90°C and sample 7 was oil used for several days.

Another study was carried out using univariate control charts. The objective was to perform a comparison with multivariate control charts. The variables chosen were the absorbance values of characteristic peaks of vegetable oils: $2923.5\,\mathrm{cm}^{-1}$, $2852.2\,\mathrm{cm}^{-1}$, $1743.3\,\mathrm{cm}^{-1}$, $1151.2\,\mathrm{cm}^{-1}$ and $719.4\,\mathrm{cm}^{-1}$. Then, five different univariate control charts were developed.

To build the univariate control charts the first group of samples was used. The mean (mean value of absorbance) and the standard deviation for these selected frequencies were calculated to developed the warning limits of $\pm 2\sigma$ (95% confidence) and also $\pm 3\sigma$ (99% confidence) for each chart. Again, the samples of the second group were used to validate the control charts developed.

In the first univariate chart (using the absorbance value at $2923.5\,\mathrm{cm}^{-1}$) sample number 10 was classified as "out of control". On the second chart ($2852.2\,\mathrm{cm}^{-1}$), samples 9, and 10 were "out of control". On the third chart ($1743.3\,\mathrm{cm}^{-1}$), samples 5 and 10 were "out of control". On the fourth chart ($1151.2\,\mathrm{cm}^{-1}$), sample 9 was "out of control" and on the last chart ($719.4\,\mathrm{cm}^{-1}$), samples 5, 6, 9, 10 and 11 were "out of control". Even using all univariate control charts in conjunction, it was not possible to identify all expected "out of control" samples.

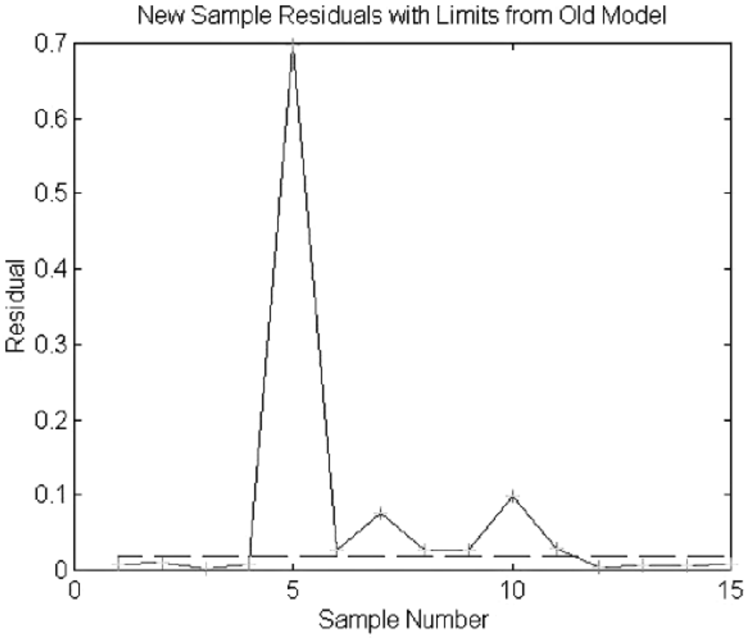

Figure 5. Q chart.

CONCLUSION

In this paper the feasibility of the use of Fourier transform infrared spectroscopy in conjunction with multivariate statistics, based on principal component analysis, is demonstrated to develop a quality control strategy for the detection of vegetable oil oxidation. This procedure is simple, fast and adaptable to oil process monitoring and requires only the infrared spectra spectrum acquisition of a set of normal samples to develop the control charts.

The multivariate control charts were capable of identifying all expected "out of control" samples. On the other hand, all univariate control charts together were not able to do this. The main advantage in using the multivariate strategy is that it is only necessary to monitor two different charts at the same time. Also, correlations among variables are eliminated and they are no problem in chart development. In univariate control charts is needed the selection of correct variables as well as the simultaneous analysis of several charts.

ACKNOWLEDGMENTS

The authors thank CNPq, Brazil, for a fellowship to R. C. Barthus and Prof. C. H. Collins for manuscript revision.

REFERENCES

1. van de Voort, F.R.; Ismail, A.A.; Sedman, J.; Emo, G. J. Am. Oil Chem Soc. **1994**, *71*, 243.
2. *Official Methods and Recommended Practices of the American Oil Chemists' Society*; 4th Ed.; American Oil Chemists' Society: Champaign, 1989.
3. Wilson, R.H.; Tapp, H.S. Trends Anal. Chem. **1999**, *18*, 85.
4. Ma, K.; van de Voort, F.R.; Ismail, A.A.; Sedman, J. J. Am. Oil Chem. Soc. **1998**, *75*, 1095.
5. van de Voort, F.R.; Ismail, A.A.; Sedman, J. J. Am. Oil Chem. Soc. **1995**, *72*, 873.
6. Montgomery, D.C. *Statistical Quality Control: An Introduction*; 2nd Ed.; Wiley: New York, 1991.
7. Johnson, A.R.; Wichem, W.D. *Applied Multivariate Statistical Analysis*; 4th Ed.; Prentice Hall: Upper Saddle River, New Jersey, 1998.
8. Nijhuis, A.; Jong, S.; Vandeginste, B.G.M. Chemom. Intell. Lab. System. **1997**, *38*, 51.
9. Nijhuis, A.; Jong, S.; Vandeginste, B.G.M. Chemom. Intell. Lab. System. **1999**, *47*, 107.
10. Beebe, K.R.; Kowalski, B.R. Anal. Chem. **1987**, *59*, 1007A.
11. Wise, B.M.; Gallagher, N.B. *PLS Toolbox for Use with Matlab*, ver. 2.1; Eigenvector Technologies: Manson, 1998.

Received February 6, 2002
Accepted June 17, 2002