Quantitation of Foreign Fat in Foreign Fat/Milkfat Mixtures by Multivariate Regression Analysis of Fatty Acid Data

Franz Ulberth*

Department of Dairy Research and Bacteriology, Agricultural University, Gregor Mendel-Strasse 33, A-1180 Vienna, Austria

Multivariate statistical analysis of fatty acid data was used to quantitate tallow in tallow/milkfat mixtures. These techniques proved to be superior in terms of reliability of the estimates in relation to quantitation via the butyric acid (C4:0) content of the fat blend or simple regression analysis of C4:0. Of the three multivariate regression techniques compared, partial least-squares (PLS) regression was best suited for the problem, followed by multiple regression analysis with stepwise variable selection. Calibration functions were derived by computer-simulated fatty acid profiles, and the calibration equations derived were validated by actually analyzed samples with known mixture proportions. The root mean squared difference figure for the validation data set, an indicator for the average error in the analysis, was 1.2% tallow in milkfat in the case of the PLS regression, in contrast to 3.5% for the simple regression solution.

Keywords: *Milkfat; adulteration; fatty acids; multivariate statistical analysis; PLS regression*

1. INTRODUCTION

Fraudulent malpractices in the oils and fats industry have been known for a long time. Highly priced commodities such as virgin olive oil, milkfat (MF), cocoa butter, and other speciality oils are the most attractive targets for the unscrupulous manufacturer or supplier. Moreover, liberalization of food regulations has brought about a legalization of certain fat mixtures, e.g. butter–margarine blends, provided that they comply with given product standards and that the product is correctly labeled. Thus, authentication of genuine products is of concern to the processing industry, food inspectors, and the consumers.

Strategies to detect adulterations include the determination of physicochemical properties, constituents of the unsaponifiable matter, and fatty acid (FA) and triacylglycerol (TG) analysis [for a review see Collomb and Spahni (1991)]. A sample is considered to be adulterated when the determined value for a certain criterion deviates significantly from the range found in genuine products. Over recent years, the combination of high-performance analytical techniques with multivariate statistical evaluation of the results obtained proved to be a very sensitive tool for tracing fat mixtures. Use of these multivariate procedures exploits the information content of, e.g., a chromatographic profile much better than does the application of univariate statistics. Most of the chemometric techniques employed are qualitative in nature; i.e., they are aimed at grouping samples in one of several possible classes (e.g. genuine/adulterated), although a few quantitative approaches have been described. Timms (1980) used multiple regression analysis to fit the percentages of three individual TG fractions of genuine Australian MF to a fixed value of 100 (so-called R value). The R value allows the authentication of pure MF samples on one hand and, in case of an adulteration, the quantitation of the level of the foreign fat admixed with an accuracy of 2 mass %. This applies also to situations when the pure mixture components are not available for analysis. Average deviations of 0.7–0.8% from the known level of admixture (4–7% non-MF in MF) were found with samples of European origin by using a modification of Timm's technique (Precht, 1991, 1992). On the contrary, estimation of the content of MF in fat mixtures via the determination of a single TG fraction (acyl C-34) resulted in unacceptable values when the MF sample used was unavailable for analysis (Schneller and Wullschleger, 1992). On the basis of FA analysis, errors in the order of 1–7 mass % were observed in a study aimed at estimating the mixture proportions of MF/vegetable oil mixtures for which the FA composition of the individual fats was unknown (Jebson and Curtis, 1981).

The International Dairy Federation (IDF) has adopted FA analysis as the basis for the quantitation of MF in fat mixtures (Muuse and Martens, 1993). To increase the reliability of the method, not only is the MF-specific butyric acid (C4:0) taken into account, as is usually done by the Phillips and Sanders (1968) or the Kuzdzal-Savoie and Kuzdzal (1968) method, but the calculation includes also the contents of other FA. In a collaborative study, repeatability figures of 2.0–5.4% MF and reproducibility figures of 4.9–9.8% MF in mixture with other fats were found, provided the individual mixture components were accessible (Muuse and Martens, 1993). Another approach used FA data in a truly multivariate way (Christen, 1989). The proportions of three FA allowed the prediction of the MF content by multiple linear regression (MLR) analysis within 2 mass % of the given value. Mixed integer programming based on FA data was also an effective means for recognizing fat blends (De Jong and de Jonge, 1991).

Partial least-squares (PLS) regression is increasingly used for multivariate calibration problems in a broad range of applications (Thomas, 1994). A combination of FA as well as TG peaks and PLS regression enabled prediction of vegetable oil mixture composition with prediction errors <1 vol % (Kaufmann, 1993).

The aim of this study was to compare different evaluation methods for the quantification of tallow in mixtures with MF using chromatographically determined FA data as input variables. Tallow was chosen

* Fax +43 1 3199193; e-mail h610pb@edv1.boku.ac.at.

0021-8561/95/1443-1556$09.00/0

because its reliable quantitation in MF is known to be difficult and may thus serve as an example for all other non-MF/MF mixtures. The concentration range studied was confined to moderate levels of commingling.

2. MATERIALS AND METHODS

2.1. Milkfat Samples. Genuine butter samples ($n = 352$) were collected in 1987–1991 from different parts of Austria. Butter was melted, the clear oily layer dried with Na_2SO_4, and the anhydrous fat stored under N_2 at −25 °C until analyzed. Tallow was from a local rendering plant (Schachinger, Vienna) and treated in the same way as MF.

2.2. Fatty Acid Analysis. Fats were transesterified according to the Christopherson and Glass (1969) procedure. FA analysis by capillary GLC, calibration of the instrumental setup, and analytical quality control were done as described (Ulberth, 1994).

2.3. Fat Mixtures. Fat mixtures were either prepared gravimetrically from the pure components or simulated arithmetically by the relationship (% FA_i denotes grams of an individual FA/100 g of total FA):

$$\%\ FA_i\ (\text{mixture}) = \%\ MF \times \%FA_i \text{ in MF} + \%\ \text{tallow} \times \%FA_i \text{ in tallow}$$

In this way, each FA profile of the genuine MF samples was blended at randomly selected proportions, covering the range of 2–30%, with 10 different tallow samples. From the resulting data 48 mixtures were chosen by chance to serve as a calibration data set for regression analysis. Seventy-eight mixtures were gravimetrically prepared from randomly selected MF and tallow samples and served as a validation data set.

2.4. Statistics. The PLSPlus version 2.0 software from Galactic Industries Corp. (Salem, NH) was used for PLS regression analysis. Other computations were done with the SAS/STAT release 6.03 program (SAS Institute, Cary, NC). The significance level of a predictor variable for entry and stay in the model was set to 0.15 during execution of the stepwise variable selection in SAS procedure "REG".

3. RESULTS AND DISCUSSION

3.1. Estimation of Tallow in MF by Using Butyric Acid Values. Although C4:0 is found only in MF of ruminant animals, an accurate quantification of the proportions of MF and a non-MF in a mixture by way of its C4:0 content is seriously hampered by the natural variation of this component (Precht, 1991). This dilemma is further aggravated by the fact that volatile fatty acids are somewhat difficult to analyze and quantitate by GLC.

The 352 genuine MF samples analyzed in this study contained 3.71 ± 0.13 g C4:0/100 g of total FA with a maximum value of 4.02 and a minimum of 3.17, respectively. The proportion of MF in an unknown non-MF/MF mixture with, e.g., a C4:0 content of 3.00 g may thus lie between 74.6% and 94.6% if the natural variability of the C4:0 content is completely taken into account. Using published C4:0 values for the computation even worsened the situation. Values as low as 1.61% (Bitman and Wood, 1990) and as high as 6.00% (Iverson and Sheppard, 1989) may be found in the newer literature.

3.2. Estimation of Tallow in MF by Using Simple Regression Analysis of Butyric Acid Data. Preparing gravimetrically a set of non-MF/MF mixtures using MF and non-MF samples from different production areas, seasons, etc. so as to reflect the natural variability of the FA composition and calculating a calibration function by least squares would obviously be a better solution for quantitating the mixture proportions. To guard against overfitting of the function, a synthetic calibration data set was prepared by an arithmetical simulation as was done for the same reason in a survey aimed at predicting the composition of vegetable oil mixtures (Kaufmann, 1993). The appropriateness of the computer-generated FA data for the simulation of MF in mixtures with other fats was proven in another study (Ulberth, 1994). The compositional features of the synthetic calibration set as well as a set of 78 gravimetrically prepared tallow/MF mixtures are listed in Table 1. The latter data set was assigned to validate the calibration functions derived with the computer-generated data. No significant differences between the two data sets were found, as shown by equality of the mean values (Student's t-test, $P > 0.10$), except for C18:0 and C18:2(n−6) ($P < 0.05$). Calibration by simple regression analysis improved the accuracy for predicting mixture compositions, as compared to the approach via C4:0 mean values. The relationship between the C4:0 content and the proportion of tallow in mixture with MF is shown in Figure 1. On the basis of the data in the calibration set, an R^2 value of 0.752 was computed. The function relating the C4:0 content to percent tallow of the calibration set (% tallow = 81.823 − 21.464 × C4:0) was not different from the related function based on the validation data (% tallow = 74.935 − 20.035 × C4:0, R^2 = 0.779). These close results for the two data sets further justified the appropriateness of the computer-generated data for predicting the mixture composition of real samples. Also listed in Table 1 are Pearson's correlation coefficients describing the relationship between the tallow content in the mixtures and other FA besides C4:0. In general, short- and medium-chain FA were negatively correlated, C16:0 was only weakly correlated, and other long-chain FA were positively correlated with the tallow proportion concerning both data sets. This fact suggested that more than one predictor variable should be useful for predicting the mixture composition.

Table 1. Fatty Acid Composition (Grams of Individual FA/100 g of Total FA) of the Calibration and Validation Data Sets

	calibration data			validation data		
	mean	SD	r^a	mean	SD	r
C4:0	3.28	0.313	−0.867	3.38	0.255	−0.882
C6:0	1.97	0.186	−0.908	1.98	0.174	−0.825
C8:0	1.09	0.116	−0.848	1.10	0.103	−0.740
C10:0	2.36	0.287	−0.734	2.42	0.269	−0.640
C12:0	2.68	0.350	−0.677	2.76	0.330	−0.548
C14:0	9.71	0.890	−0.677	9.95	0.787	−0.600
C14:1	0.82	0.087	−0.447	0.84	0.072	−0.559
C15:0ISO	0.33	0.016	−0.068	0.33	0.012	−0.190
C15:0AISO	0.52	0.028	−0.391	0.52	0.020	−0.546
C15:0	1.25	0.076	−0.667	1.25	0.055	−0.700
C16:0	29.94	2.598	−0.105	29.83	2.229	−0.141
C16:1	1.58	0.085	0.569	1.63	0.056	0.583
C17:0ISO	0.56	0.033	−0.215	0.56	0.050	−0.187
C17:0AISO	0.52	0.029	0.603	0.54	0.061	0.462
C17:0	0.86	0.132	0.610	0.57	0.362	0.650
C18:0	10.55	1.290	0.741	10.02	1.073	0.772
C18:1(n−9)	20.45	1.714	0.499	20.52	1.705	0.431
C18:1(n−7)	2.79	0.975	0.049	2.91	0.821	0.070
C18:2(n−6)	1.42	0.185	0.860	1.28	0.158	0.607
C18:3(n−3)	0.81	0.180	0.054	0.78	0.142	−0.050
C18:2conj	0.83	0.353	−0.159	0.89	0.280	−0.128

[a] Pearson's correlation coefficient describing the relationship between percent FA and percent tallow admixed.

3.3. Estimation of Tallow in MF by Using Multiple Regression Analysis of FA Data. Multivariate

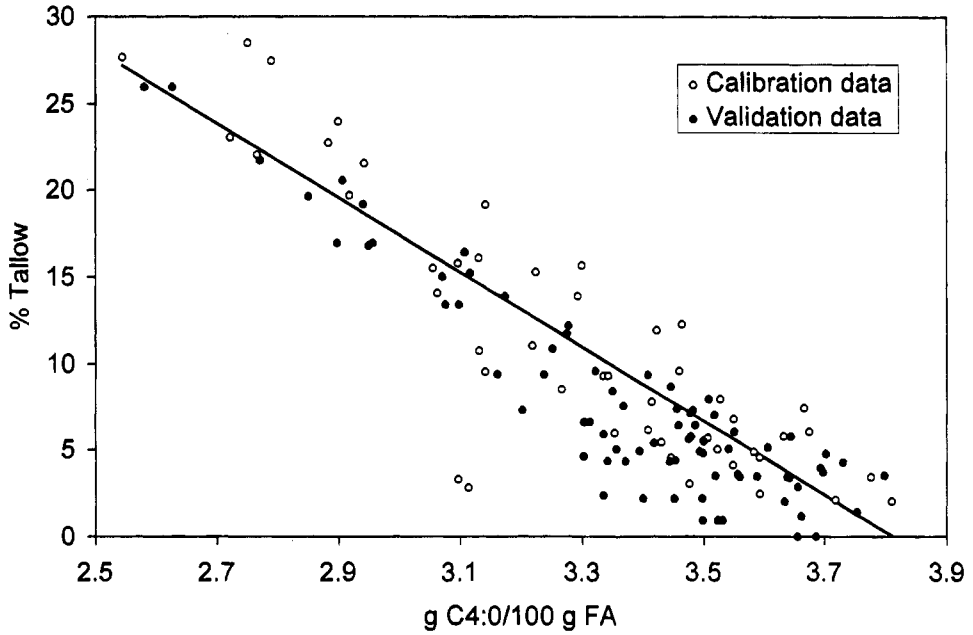

Figure 1. Relationship between the butyric acid content of tallow/MF mixtures and the proportion of tallow admixed. The calibration line shown was based on the calibration data set.

Table 2. Statistical Indices Describing Goodness-of-Fit of the Regression Models

	simple[a]	MLR[b]	SMLR[c]	PLS[d]
R^2	0.779	0.997	0.996	0.988
SSR[e]	666.10	7.29	10.97	12.73
PRESS[f]	717.63	27.05	21.87	32.76
possible outliers excluded				
SSR		5.70	22.55	
PRESS		8.35	15.00	

[a] Simple linear regression. [b] Multiple linear regression. [c] Stepwise multiple linear regression. [d] Partial least-squares regression. [e] Sum of squared residuals. [f] Sum of squares of predicted residual errors.

Table 3. Distribution Indices of Errors As Obtained by Different Regression Methods

	SR[a]	MLR	SMLR	PLS
mean	−2.09	0.18	−0.53	−0.35
SD	2.85	1.48	1.34	1.15
minimum	−9.38	−3.77	−4.28	−3.16
maximum	3.22	3.96	2.79	2.25
SSR	965.10	171.10	161.02	110.87
RMSD[b]	3.52	1.48	1.44	1.19

[a] Abbreviations: see Table 2. [b] Root mean squared difference [calculated as $(\Sigma(\text{known values} - \text{predicted values})/n)^{1/2}$].

calibration is primarily used for various IR spectroscopical techniques in which a quantitative property of the analyte is fitted to light absorption at different wavelengths (Thomas, 1994). By analogy, MLR analysis and MLR with stepwise selection of regressor variables (SMLR) were performed to utilize entirely the information content of the overall FA profiles for predicting the tallow concentrations. Twelve predictor variables (C4:0, C6:0, C14:1, C15:0, C16:1, C17:0, C17:0ISO, C18:0, C18:1($n-9$), C18:1($n-7$), C18:2, C18:3) were selected by a stepwise variable selection algorithm. Likewise, the IDF method recommends also the use of some of the selected FA (the even-numbered saturated FA with 4−16 C atoms and, additionally, C15:0 and C16:0) for the calculation of the mixture proportions (Muuse and Martens, 1993). Only three FA, i.e. C12:0, C14:0, and C15:0, were necessary to predict the butter proportion in butter−margarine blends by MLR analysis (Christen, 1989). The rationale for this simple three-variable solution was obviously the fact that only vegetable oils were used for blend formulation.

Statistical indices calculated for the calibration data set are summarized in Table 2. Both multivariate techniques were superior to the simple regression method as judged by the R^2 value. Moreover, other indicators for "goodness-of-fit" of the data to the regression model, e.g. the sum of the squared residuals (SSR) and the sum of squares of predicted residual errors (PRESS), were much lower for the multivariate models. MLR and SMLR seemed to be equally suited for describing the relationship between tallow content and FA profile of the calibration data set. Two of the 48 observations exhibited comparatively large residuals as indicated by Cook's distance measure (Freund and Littell, 1986), both in the model including all 21 predictor variables and in the model selected by the stepwise selection procedure. These two observations may thus exert a large influence on the regression parameter estimates. Indeed, exclusion of the two observations improved the fit of the models slightly (Table 2) but was, conversely, not advantageous for predictions made on the validation data set. Therefore, the unmodified regression equations were taken for further calculations.

The effectiveness of the regression models for future predictions was checked by means of the validation data set. The outcomes of the tests are summarized in Table 3. As expected from the results of the calibration data, predictions made on the validation data by using the multivariate techniques were much more reliable than by simple regression. The mean error of prediction made by MLR was somewhat lower than by SMLR, but the reverse was true for SSR. The preference for utilizing the SMLR model for predictions made on the validation data is further substantiated by the frequency distributions of the errors (difference between known and predicted values), which are given in Figure 2 for the multivariate regression models.

3.4. Estimation of Tallow in MF by Using Partial Least-Squares Regression Analysis of FA Data. Parameter estimates derived by multiple regression procedures may be unstable and thus misleading when

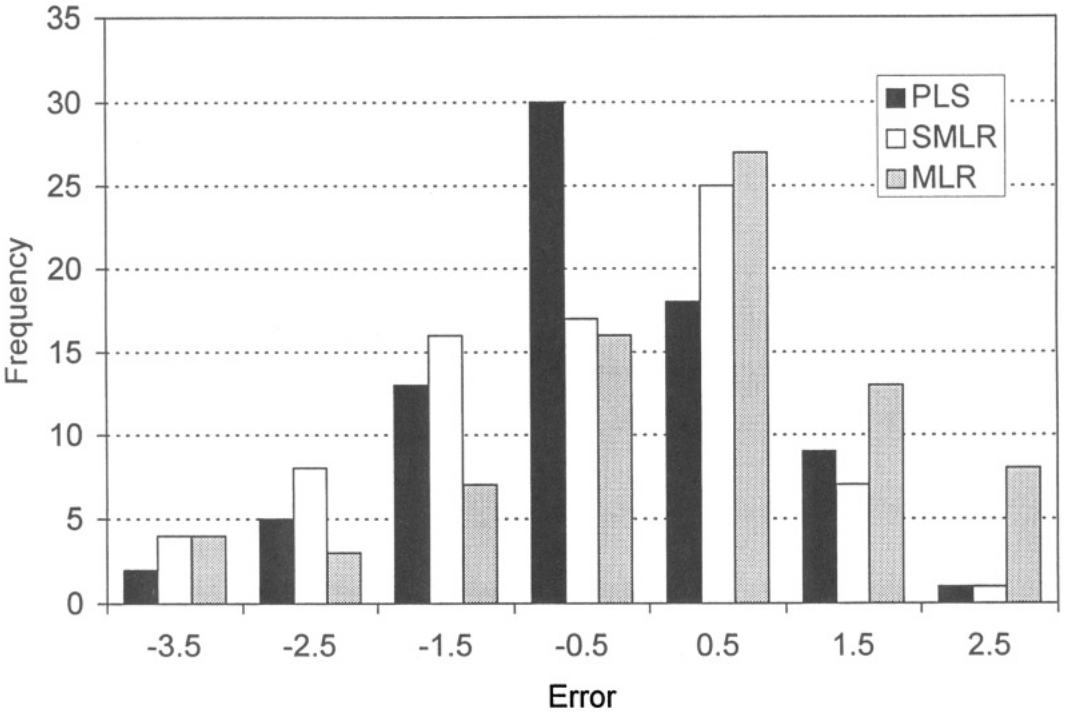

Figure 2. Distribution of errors (known tallow content of the validation data minus predicted value) obtained by different multivariate regression methods.

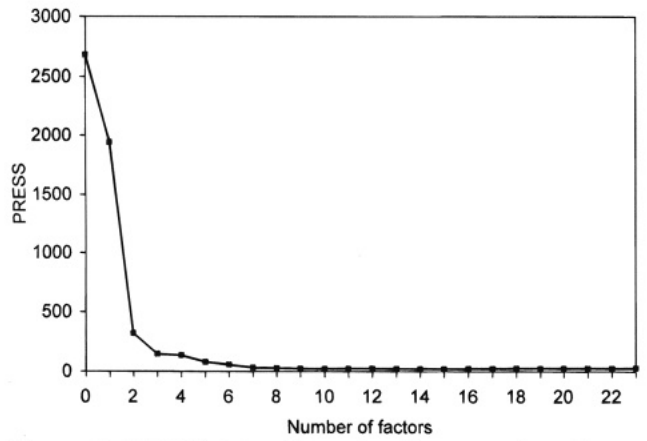

Figure 3. PRESS plot to determine the appropriate dimensionality of the PLS model.

predictor variables are highly intercorrelated (Freund and Littell, 1986). PLS regression is known to be an excellent tool for dealing with such collinear data. Application of PLS regression in connection with cross-validation to select the proper model dimensionality for the calibration data enabled accurate predictions of the mixture composition of the validation set (Table 3). Plotting the PRESS value against the number of factors (dimensions) revealed that seven factors would be a good choice to model the calibration data (Figure 3). Inclusion of more factors brought the PRESS value down from 32.76 (7 factors) to 22.27 (14 factors), but this may well lead to overfitting of the data. Therefore, seven factors were considered satisfactory. This was also justified by an F statistic, which is based on the ratio of the minimum PRESS value to PRESS values involving fewer factors (Haaland and Thomas, 1988). At greater than seven factors the F ratio probability dropped below 0.9.

PLS regression enabled predictions of unknown tallow concentrations with superior accuracy, compared with the other regression techniques considered (Table 3). The root mean squared difference (RMSD) figure, an indicator for the average error in the analysis, was smallest for the PLS solution as was also the spread of the distribution of the errors (Figure 2).

Conclusions. Multivariate evaluation of FA data, in particular PLS regression, proved to be a well-suited technique for quantifying tallow in MF. As a mixture of MF and tallow represents the most challenging situation for an analyst interested in the composition of the mixture, the technique should also be of great value for the analysis of blends based on simpler mixture constituents, e.g. vegetable oil/MF melange products. A further advantage is that FA data already available in databases can be used to generate calibration data sets by computer simulation, thus rendering reanalysis of calibration samples unnecessary. Moreover, multivariate regression procedures are implemented in most commercially available statistical computer programs. This enables the analyst to include certain peculiarities of the samples, due to seasonal and regional variations, breed, etc., in the calculation of individual calibration functions instead of using published criteria or formulas.

ACKNOWLEDGMENT

I thank H.-J. Haider of Haider Scientific Ltd., Kottingbrunn, Austria, for doing the PLS calculations.

LITERATURE CITED

Bitman, J.; Wood, D. L. Changes in milk fat phospholipids during lactation. *J. Dairy Sci.* **1990**, *73*, 1208–1212.

Christen, G. L. A method to quantify butteroil added to butter-margarine blends. *J. Food Qual.* **1989**, *11*, 453–459.

Christopherson, S. W.; Glass, R. L. Preparation of milk fat methyl esters by alcoholysis in an essentially nonalcoholic solution. *J. Dairy Sci.* **1969**, *52*, 1289–1290.

Collomb, M.; Spahni, M. Adulteration of milk products. Review of the analytical criteria for the detection of vegetable and animal fats in milk fat. *Mitt. Geb. Lebensm. Hyg.* **1991**, *82*, 615–662.

De Jong, S.; de Jonge, T. J. R. Computer assisted fat blend recognition using regression analysis and mathematical programming. *Fat Sci. Technol.* **1991**, *93*, 532–536.

Freund, R. J.; Littell, R. C. *SAS System for Regression*; SAS Institute: Cary, NC, 1986.

Haaland, D. M.; Thomas, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **1988**, *60*, 1193–1202.

Iverson, J. L.; Sheppard, A. J. Detection of adulteration in cow, goat, and sheep cheese utilizing gas-liquid chromatographic data. *J. Dairy Sci.* **1989**, *72*, 1707–1712.

Jebson, R. S.; Curtis, H. The analysis if two-component fat mixtures. *N. Z. J. Dairy Sci. Technol.* **1981**, *16*, 121–132.

Kaufmann, P. Prediction of mixture composition by chromatographic characterization, multivariate classification and partial least-squares regression, a comparison of methods. *Anal. Chim. Acta* **1993**, *277*, 467–471.

Kuzdzal-Savoie, S.; Kuzdzal, W. Determination of butterfat in a fat mixture. *Lait* **1968**, *48*, 607–611.

Muuse, B.; Martens, R. Mixtures of milkat with non-milkfat—determination of the milkfat content. *Int. Dairy Fed. Bull.* **1993**, *285*, 65–69.

Phillips, A. R.; Sanders, B. J. Semi-micro determination of butter fat in fat mixtures by gas-liquid chromatography. *J. Assoc. Publ. Anal.* **1968**, *6*, 89–95.

Precht, D. Detection of adulterated milk fat by fatty acid and triglyceride analysis. *Fat Sci. Technol.* **1991**, *93*, 538–544.

Precht, D. Detection of foreign fat in milk fat. II. Quantitative evaluation of foreign fat mixtures. *Z. Lebensm. Unters. Forsch.* **1992**, *194*, 107–114.

Schneller, R.; Wullschleger, R. On the determination of the content of butter fat in fat mixtures by capillary gas chromatography. *Mitt. Geb. Lebensm. Hyg.* **1992**, *83*, 737–744.

Thomas, E. V. A primer on multivariate calibration. *Anal. Chem.* **1994**, *66*, A795–A804.

Timms, R. E. Detection and quantification of non-milk fat in mixtures of milk and non-milk fats. *J. Dairy Res.* **1980**, *47*, 295–303.

Ulberth, F. Detection of milk fat adulteration by linear discriminant analysis of fatty acid data. *J. AOAC Int.* **1994**, *77*, 1326–1334.

Received for review November 1, 1994. Revised manuscript received February 13, 1995. Accepted March 31, 1995.[⊗]

JF9406166

[⊗] Abstract published in *Advance ACS Abstracts,* May 15, 1995.