

## A multivariate chemometric approach to fluorescence spectroscopy

Lars Nørgaard

Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Food Technology, Thorvaldsensvej 40, DK-1871 Frederiksberg, Denmark

Received 29 December 1994; revised 7 March 1995; accepted 10 March 1995

### Abstract

A multivariate approach to the solution of problems often encountered in the spectrofluorometry of natural samples, utilising information from whole spectra is presented. (a) Piecewise direct standardisation is implemented and employed to transfer emission spectra measured with two different xenon lamps of different ages as if the spectra were measured with the same lamp. (b) It has been shown using a multivariate analysis approach that it is possible to use the raw data points instead of the smoothed data based on an algorithm included in the instrument software by the manufacturer. (c) It is documented that Raman scattering does not hamper the performance of multivariate calibration; on the contrary, in an experiment with sugar samples the concentration prediction errors become about five times lower by including the whole emission spectrum in the analysis instead of using a univariate calibration based on an emission wavelength that only reflects the analyte of interest. (d) An algorithm for variable selection is implemented and employed in the selection of optimal excitation wavelengths. Among 13 emission spectra recorded for a sugar sample at different excitation wavelengths, four of these are chosen that describe 98.51% of the total variance in the original data. (e) Finally the combination of fluorescence spectroscopy and multivariate calibration with conventional chemical data according to the near-infrared black box model is presented. The refined sugar quality parameter, the ash content and the fluorescence emission spectra are correlated by a partial least-squares regression model. Five experiments employing different monochromator slit widths and sugar concentrations are performed, and the best correlation obtained by full cross-validation of the 15 sugar samples is  $R = 0.98$ .

### 1. Introduction

Fluorescence spectroscopy has been used for decades as a powerful analytical tool in all sorts of chemical, biochemical, food and environmental laboratories due to its two principal properties: excellent sensitivity and specificity. The use of fluorescence excitation and emission spectra has so far mostly been qualitative, with the purpose of finding a pair of excitation–emission wavelengths where the analyte of interest is the only chemical component giving rise to the recorded signals [1,2].

On searching the literature, very few papers are encountered dealing with the application of

multivariate chemometric methods in fluorescence spectroscopy. This seems especially striking bearing in mind the enormous number of publications using chemometrics in combination with traditional spectroscopic techniques like near-infrared (NIR) spectroscopy [3–6]. In the majority of the papers dealing with fluorescence spectroscopy and chemometrics, only synthetic samples or natural samples containing very few chemical species are analysed with the aim of resolving a measured excitation–emission matrix (EEM) into pure excitation and emission spectra of the chemical components in the samples, and with the aim of

predicting analyte concentrations by rank annihilation methods [7–12]. When the number of signal-producing chemical components in a sample becomes larger than 5–6 it is very difficult to perform this resolution within acceptable limits of accuracy and precision [9].

In studies by Lindberg et al. [13] and Sjöström et al. [14] concentrations of two- and three-component synthetic mixtures are successfully predicted from fluorescence emission spectra by the method of partial least-squares regression. In Refs. [15–18] fluorescence spectroscopy in combination with partial least-squares regression is used for predicting botanical tissue components of complex wheat flour samples. The latter example is analogous to the widespread use of multivariate calibration in NIR spectroscopy [3–6], and illustrates how the chemical information for “dirty” natural samples can be enhanced by fluorescence spectroscopy in combination with multivariate statistical methods.

In this paper, the focus is on the analysis of spectra from refined sugar samples in order to investigate different intrinsic instrumental parameters as well as the combination of fluorescence spectroscopy of given samples and multivariate calibration. The study both serves as a general investigation of how to enhance the potential of fluorescence spectroscopy by chemometrics as well as a special investigation of how fluorescence spectroscopy of sugar samples can be approached in a multivariate sense. Sugar samples are real samples and all topics treated reflect the problems of measuring fluorescence spectra on real samples.

Refined sugar is a very pure foodstuff containing only small concentrations of impurities (below 1%). These impurities, such as phenolic compounds, amino acids, melanoidins, and melanins [19], stem from the sugar beet delivered to the sugar factory superimposed the chemical compounds produced at the individual unit operations of the factory. It is the long-term purpose of the project, of which this publication is a part, to use fluorescence spectroscopy for the on-line prediction of (a) refined sugar quality [20] and (b) optimal adjustment of process parameters. The fluorescence measured is autofluorescence (or primary fluorescence), i.e. the native fluorescence of the impurities in the dissolved sugar sample. The beauty of this approach rests in the fact that nature and processing combined respond with a multivariate fluorescent fingerprint [21].

Using several examples, the potential of multivariate analysis in fluorescence spectroscopy will be demonstrated. The topics addressed are as follows.

(i) Transferring spectra between different instrumental set-ups. Spectrofluorimeters are all different due to differences in lamps, monochromators and photomultipliers [2]. When analysing large sample sets, which is the case in on-line/at-line process applications, it is essential in the case of instrument break-down to be able to compare samples measured with different set-ups or even on different instruments. It is outlined how this problem can be solved by the piecewise direct standardisation algorithm originally developed for the standardisation of NIR instruments [22–24].

(ii) Smoothing of spectral data. The spectrofluorimeter employed for analysis automatically smooths the measured emission spectra. Raw data will be compared to data smoothed with binomial and Savitzky-Golay smoothing algorithms.

(iii) Raman scattering. This often overlaps the analyte emission spectra. The classical way of circumventing the problem of Raman scattering is to employ suitable filters [25] (if available), to subtract a blank emission spectrum from all the samples [25], or to change the excitation wavelength [25]. An example will be given of how multivariate methods are capable of overcoming the problem of the Raman scattering peaks by including whole emission spectra in the mathematical analysis.

(iv) Selection of wavelengths. The measurement of fluorescence emission spectra at a large number of excitation wavelengths is a time-consuming operation. It will be demonstrated how the principal variable algorithm [26] chooses a small but optimal number of excitation wavelengths describing the main variations in the excitation–emission data matrix. The wavelength selection is especially important in the multivariate calibration of large sample sets as well as for fast on-line prediction of analyte concentrations in future samples.

(v) Multivariate calibration. Multivariate calibration [27] of the sugar quality parameter ash content in refined sugar samples is performed. Experiments with different excitation monochromator and emission monochromator slit widths and sugar concentrations are performed in order to investigate their influence on the multivariate calibration models. The influence of pH and the use of corrected/not

corrected emission spectra in connection with multivariate modelling will be touched on.

## 2. Materials and methods

### 2.1. Instrumentation

All experiments are performed on a Perkin-Elmer LS 50B spectrometer. The computer controlling the instrument through an RS232C interface is an IBM-compatible 486/50 MHz PC. Uncorrected fluorescence spectra samples with 0.5 nm intervals are recorded in all experiments.

### 2.2. Programs

Calculations are performed with Matlab for Windows version 4.2c.1 (MathWorks, Inc.) and Unscrambler version 5.5 (CAMO A/S). A chemometric toolbox made by B.M. Wise [28] is used for piecewise direct standardisation. The Perkin-Elmer LS50 FLDM Instrument program (version 4.00) is used for controlling the instrument. Spectral data are converted to ascii files by a program made in the OBEY language furnished by Perkin-Elmer (OBEY, version 3.50). An OBEY program made by S. Huckins, Perkin-Elmer, is applied to obtain the raw fluorescence emission spectra from the instrument.

### 2.3. Measurement conditions

In experiments including several excitation wavelengths, the measurement always starts with the largest excitation wavelength and ends with the lowest excitation wavelength in order to minimise photodecomposition of the sample. This problem needs no further concern as long as each sample in a given experiment is subject to exactly the same measurement cycle. In all experiments the sample holder is thermostatted to  $24\text{ }^{\circ}\text{C} \pm 0.1$ , which is the average room temperature during a normal working day.

### 2.4. Samples

Double ion exchanged water was used for all sugar solutions and dilutions. There was a high degree of repeatability of the sample measurements (see Figs. 12 and 15). The buffer capacity of refined sugar is extremely low and it is impossible to obtain reliable pH measurements

of such samples. In the pH experiment, phosphate buffers with  $C = 0.1\text{ M}$  were prepared from sodium dihydrogen phosphate. The glassware and cuvettes were kept extremely free from contamination by washing with methanol and a 2% Extran solution (Merck). The sample sets analysed are arranged as follows (including relevant instrumental parameters).

(i) Sample set A (Piecewise Direct Standardisation) comprises solid standard blocks nos. 1 (anthracene and naphthalene mixture), 2(ovalene), 3(*p*-terphenyl), 4(tetraphenylbutadiene), and 5(compound 610) from the Perkin-Elmer C 520-7440 standard set, and a refined sugar sample (15.00 g of refined sugar added to 15.00 ml of water); six samples in total. Instrumental parameters were excitation wavelength, 300 nm; emission range recorded, 320–540 nm; excitation monochromator slit width, 5.0 nm; emission monochromator slit width, 5.0 nm; and emission scan velocity,  $500\text{ nm min}^{-1}$ . The spectra are binomial smoothed (filter factor 9) by the FLDM software.

(ii) Sample set B (smoothing of spectra) comprises standard block 2 (ovalene) from the Perkin-Elmer set. Instrumental parameters were excitation wavelength, 342 nm; emission range recorded, 440–600 nm; excitation monochromator slit width, 3.0 nm; emission monochromator slit width, 5.0 nm; and emission scan velocity,  $500\text{ nm min}^{-1}$ .

(iii) For sample set C (Raman scattering), a stock solution consisting of 25.00 g of refined sugar added to 50.0 ml of water was prepared. From this stock solution, 10 samples were prepared containing from 0.5 ml to 5.0 ml (with 0.5 ml intervals). All solutions were diluted to 10.0 ml with water, i.e. the concentrations are given as 5.0, 10.0, . . . , 50.0% v/v of the stock solution. A blank solution (water) was prepared. Instrumental parameters were excitation wavelength, 300 nm; emission range recorded, 315–540 nm; excitation monochromator slit width, 8.0 nm; emission monochromator slit width, 5.0 nm; and emission scan velocity,  $1000\text{ nm min}^{-1}$ . The emission spectra are not smoothed.

(iv) Sample set D (multivariate calibration of fluorescence spectra) was as follows. (I) Fifteen different refined sugar samples consisting of 15.00 g of sugar added to 15.00 ml of  $\text{H}_2\text{O}$  (approximately 50% (w/w)). (II) Each sample in sample set 1 was diluted 1:1 (10.00 ml of sugar solution added to 10.00 ml of  $\text{H}_2\text{O}$ ). (III) The same original 15 sugar samples but with

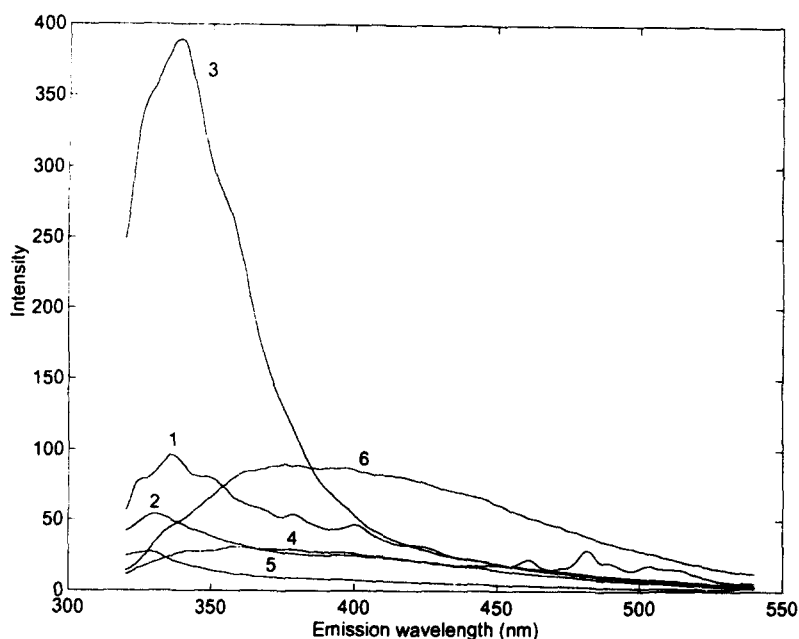


Fig. 1. Emission spectra of the 6 samples used in piecewise direct standardisation measured with lamp 1 mounted. The numbers correspond to sample set A.

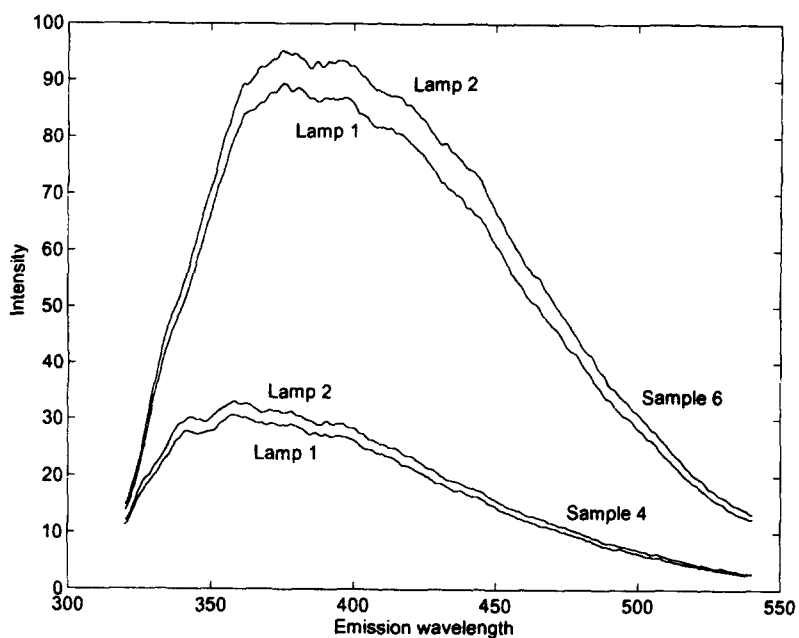


Fig. 2. Emission spectra of samples 4 and 6 measured with two different lamps mounted. Lamp 1 is older than lamp 2.

1.50 g of sugar added to 15.00 ml of H<sub>2</sub>O (approximately 10% (w/w)). Instrumental parameters were excitation wavelengths, 230 nm (emission, 245–435 nm), 240 nm (emission, 255–455 nm), 290 nm (emission, 305–555 nm), and 330 nm (emission, 345–635 nm), (1864 data points in total); and emission scan velocity, 1000 nm min<sup>-1</sup>. The emission spectra are not smoothed.

### 2.5. Chemometric methods

The chemometric methods employed are described in detail in the literature, so only a brief description including the relevant literature references will be given here. Nomenclature: capital, italic bold face letters symbolise matrices, small, italic bold letters symbolise vectors and small italic characters are scalars. The number

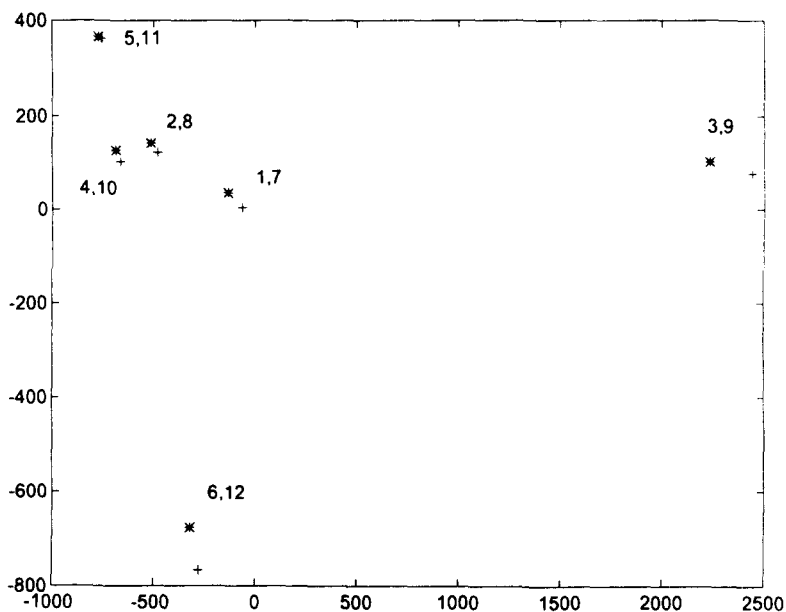


Fig. 3. Score plot (first two PCs) of 12 emission spectra before standardisation. Samples marked \* (1–6) are measured with lamp 1 and samples marked + (7–12) are measured with lamp 2. Corresponding pairs of samples are 1, 7; 2, 8; ...; 6, 12.

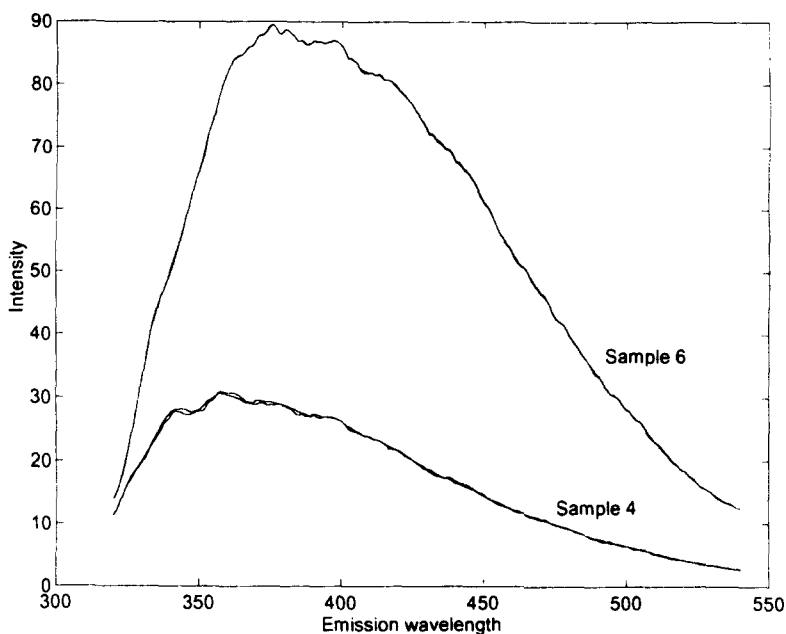


Fig. 4. Emission spectra of samples 4 and 6 measured with lamp 1 and the transformed spectra of the same samples originally measured with lamp 2. The spectral differences after transformation are very small.

of samples, the wavelength, and factors are denoted by  $s$ ,  $w$  and  $f$ , respectively.

#### Principal component analysis (PCA)

Principal component analysis finds the main variation in a multidimensional data set by creating new linear combinations of the raw data [27,29]. In matrix form we have

$$X = TP'$$

where  $X$  is the analysed data matrix with dimensions  $s \times w$ ,  $T$  is the score matrix with dimensions  $s \times \min(s, w)$ , and  $P'$  is the loading matrix with dimensions  $w \times \min(s, w)$ . Only the significant number of principal components (PCs),  $f$ , equal to the chemical rank of the  $X$

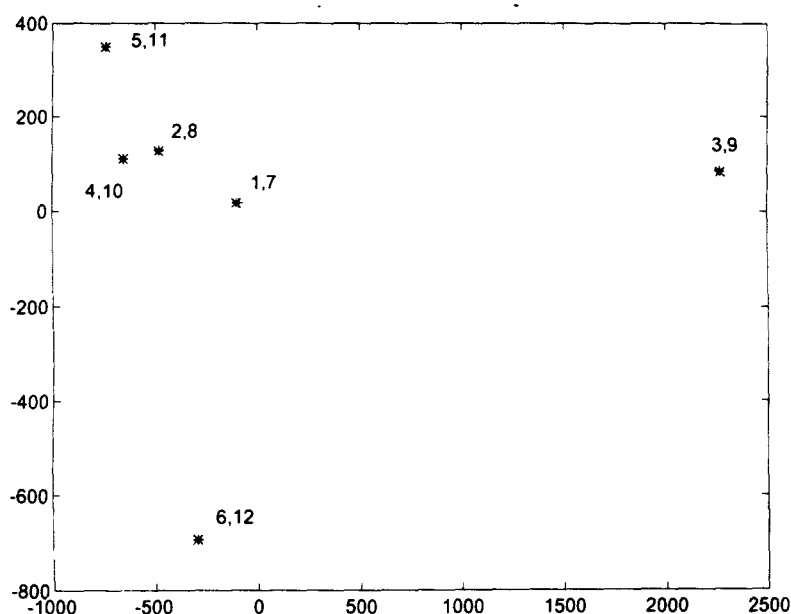


Fig. 5. Score plot (first two PCs) of 12 emission spectra after standardisation. Samples marked \* (1–6) are measured with lamp 1 and samples marked + (7–12) originally measured with lamp 2 are transformed as if they were measured with lamp 1. Corresponding pairs of samples are 1, 7; 2, 8; ...; 6, 12. Samples 2, 3, and 6 are used as transformation standards.

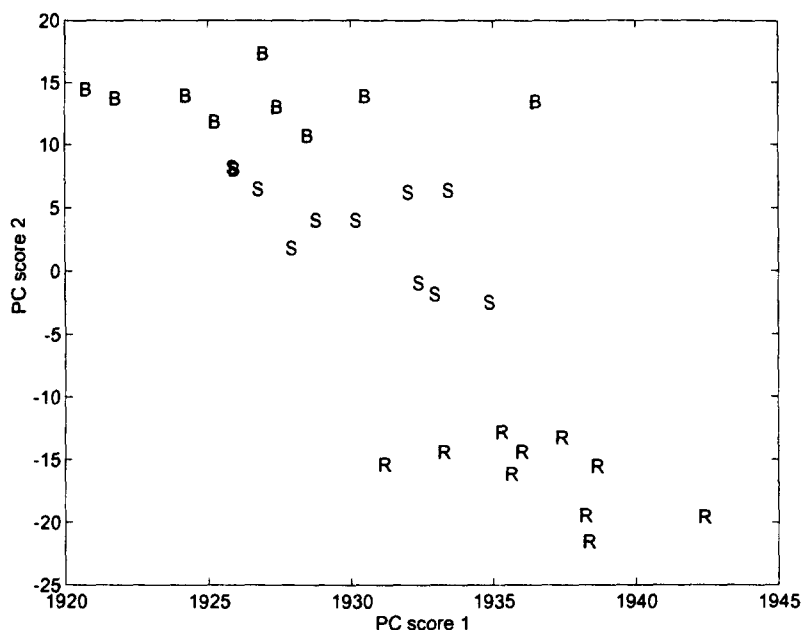


Fig. 6. Score plot (PC1 versus PC2) of 30 emission spectra of standard block 2. Three clusters are observed: binomial (B) smoothed, Savitzky-Golay (S) smoothed and raw (R) data. Two S and B samples are overlapping.

matrix is relevant in describing the information in  $X$ . This leads to the following decomposition of  $X$

$$X = T_f P_f' + E$$

where  $T_f$  is the score matrix with dimensions  $s \times f$ ,  $P_f$  is the loading matrix with dimensions  $w \times f$ , and  $E$  is a residual matrix with the same dimensions as  $X$ .

#### Partial least-squares regression (PLS)

The purpose of regression is to build a model between  $X$  and  $y$ , where  $X$  (see above) contains the fluorescence spectra and  $y$  ( $s \times 1$ ) is a vector containing the property of interest (if several properties have to be modelled, we have a matrix  $Y$  with dimensions  $s$  multiplied by the number of properties). In principal component regression [27,30] (PCR) a multiple linear re-

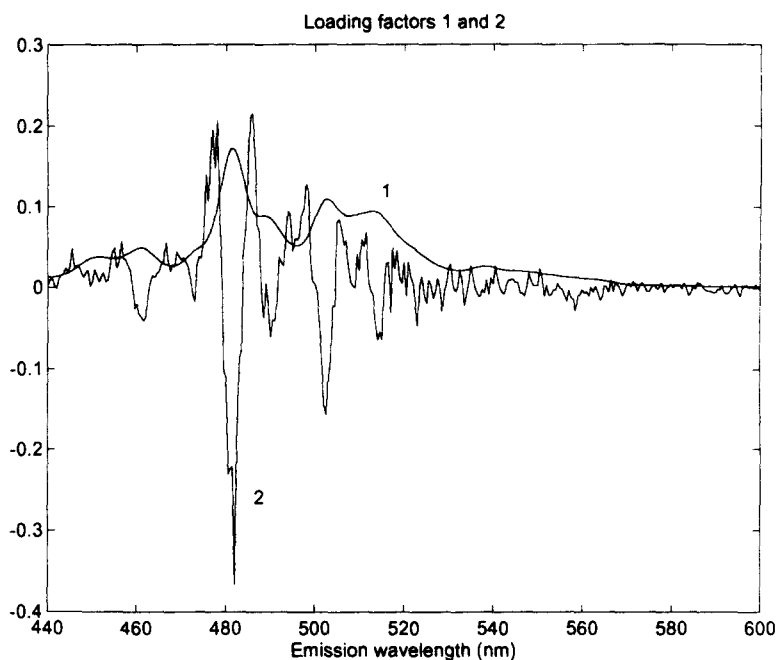


Fig. 7. Loadings from the PCA on 30 emission spectra (smoothed and raw). Loading vector 1 is smooth, while loading vector 2 indicates the spectral differences between the three data types.

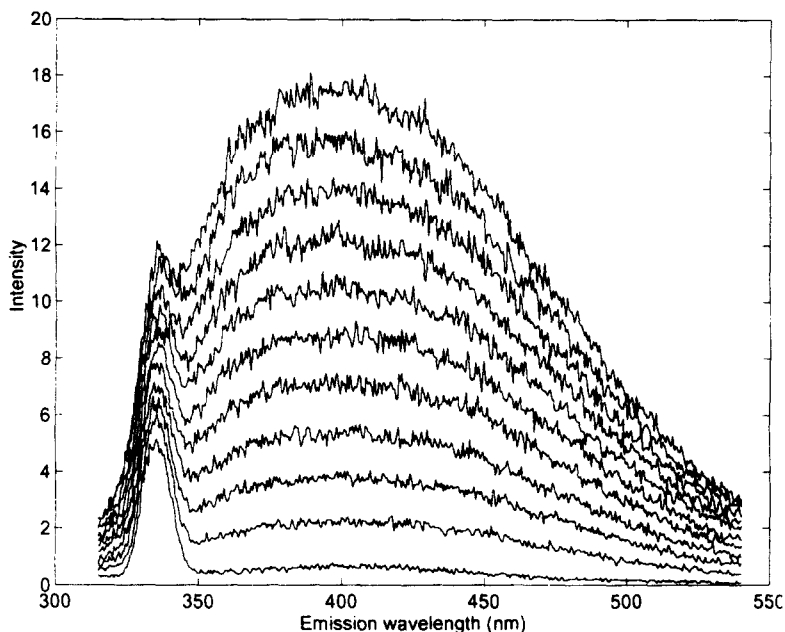


Fig. 8. Emission spectra of 11 samples with different concentrations of the same refined sugar sample (including a blank sample). The Raman scattering overlaps the analyte spectrum in the region around 334 nm.

gression model is built between the significant PCA scores ( $T_j$ ) and  $y$ . PCR maximises the variance in  $X$  followed by a regression to  $y$ . In partial least squares the significant score values  $T_j$  are found in a slightly different way, taking into account the variation in  $y$  during the decomposition of  $X$ , i.e. in PLS the covariance between  $X$  and  $y$  is maximised [27, 30–33].

The number of factors (principal components or partial least-squares components) to include when applying the above mentioned methods is found by test set validation [27]. In the case of small data sets, cross-validation [27,34] is an alternative way of validating the model performance.

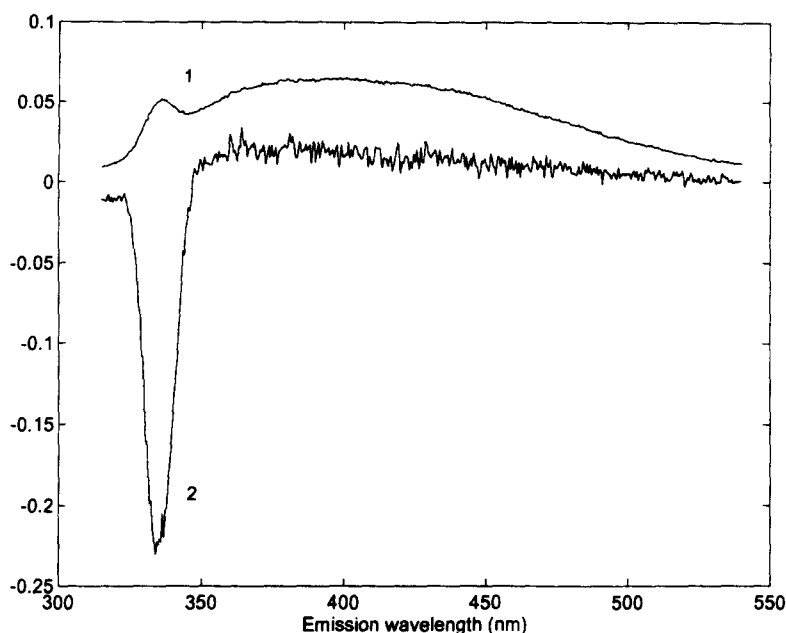


Fig. 9. Loading weights 1 and 2 for the PLS prediction of sugar concentration. The Raman scattering contributes to both loading vectors.

Table 1

Prediction of sugar concentrations

(A) Results of PLS calibration (with and without mean centring) and univariate calibration based on raw data

	With mean centring RMSEP <sup>a</sup>	R <sup>b</sup>	Without mean centring RMSEP <sup>a</sup>	R <sup>b</sup>
One PLS component	0.017	0.9999	0.134	1.000
Two PLS components	–	–	0.020	0.9999
Univariate calibration	–	–	0.105	0.9988

(B) Results of PLS calibration (with and without mean centring) based on Savitzky-Golay smoothing

	With mean centring RMSEP <sup>a</sup>	R <sup>b</sup>	Without mean centring RMSEP <sup>a</sup>	R <sup>b</sup>
One PLS component	0.025	0.9999	0.144	0.9999
Two PLS components	–	–	0.029	0.9998

<sup>a</sup> RMSEP =  $\sqrt{\sum_{i=1}^N (C_i^{\text{predicted}} - C_i^{\text{reference}})^2 / N}$ , where  $C_i^{\text{predicted}}$  is the model estimated concentration,  $C_i^{\text{reference}}$  is the reference value and  $N$  is the number of samples.

<sup>b</sup> Correlation coefficient.

### Piecewise direct standardisation (PDS)

Piecewise direct standardisation [22–24] works on two data matrices  $A$  and  $B$ . These matrices contain spectra of the same standards measured under two different instrumental conditions  $A$  and  $B$  (dimensions are number of standards multiplied by  $w$ ). No data preprocessing is performed prior to the standardisation. A multiple of local multivariate regression models between each spectral point (wavelength) in  $A$  and a window of spectral points (including the wavelength chosen in  $A$ ) in  $B$  are

then built. This yields a banded transfer matrix  $M$  ( $w \times w$ ) with a window of non-zero regression elements along the matrix diagonal and zeros elsewhere [22]. In matrix notation, we have

$$A = BM$$

Future measurement of a sample under instrumental condition  $B$  can now be transformed as if it were measured under instrumental condition  $A$  by

$$a'_{\text{rec}} = b'_{\text{new}}M$$



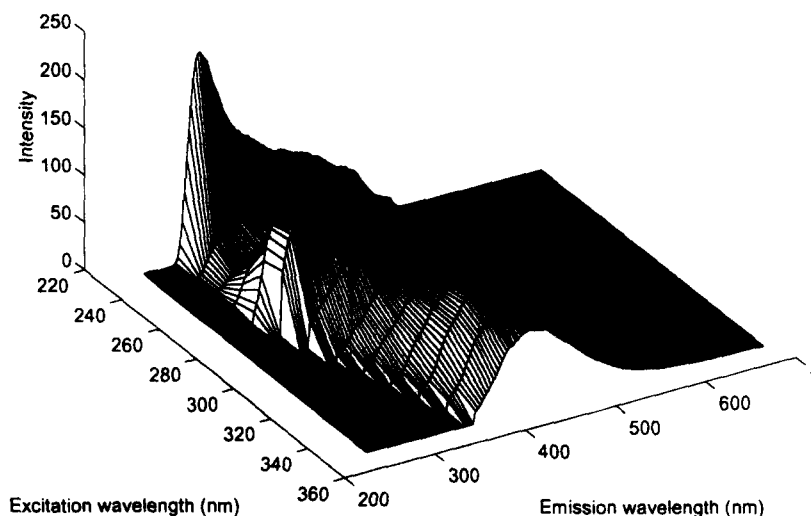


Fig. 10. A three-dimensional plot of an excitation–emission matrix (EEM) measured on a refined sugar sample. (Every 2.5 nm of the emission spectra are used in the plot).

where  $b_{\text{new}}$  is a column vector containing the measured spectrum and  $a_{\text{rec}}$  is the reconstructed spectrum. The selection of appropriate standards is important [22].

#### Principal variables (PV)

Assume that we have a matrix  $X$  ( $s \times w$ ). The method of principal variables [26] aims at finding few variables (in this case columns corresponding to wavelengths) that describe as much of the total variance in the data matrix as possible. The method is based on finding the largest diagonal value of

$$X'XX'X \quad (1)$$

with the dimension  $w \times w$ . The wavelength variable indexed by this value is the first principal variable, designated  $w$  ( $s \times 1$ ). In the next step  $X$  is orthogonalised with respect to the chosen variable

$$k = (X'w)/(w'w) \quad (2)$$

$$X_{\text{new}} = X - wk' \quad (3)$$

where  $w$  ( $s \times 1$ ) and  $k$  ( $w \times 1$ ) are column vectors. The next PVs are found by repeating the procedure from step 1 to 3 with  $X_{\text{new}}$  found in step 3. The number of PVs to choose depends on the actual problem to be solved. By analysing the matrix  $X'yy'X$  instead of  $X'XX'X$ , the method of principal variables is extendible to deal with the selection of wavelength variables to obtain optimal correlations with properties of interest ( $y$  values).

### 3. Results and discussion

#### 3.1. Piecewise direct standardisation

When using multivariate methods including many calibration samples, it is very important to be able to transfer spectra between different instrumental set-ups (lamps, slit widths, scan velocities) of the same spectrofluorimeter and, if necessary, between different instruments. Furthermore, the xenon lamp in the spectrofluorimeter also needs continuous standardisation due to a change of spectral characteristics during ageing. If, by accident, for example, the lamp breaks down after the measurement of 100 calibration samples, it is of the utmost importance to be able to use the calibration models based on these samples for the concentration prediction of samples measured with a new lamp with other spectral characteristics.

Table 2

Experiments performed to investigate the multivariate calibration of sugar samples with respect to ash content

		Sample set <sup>a</sup>		
		DI	DII	DIII
Ex 3.5	Em 3.5	Expt. 1	–	–
Ex 8.0	Em 5.0	Expt. 2	Expt. 3	Expt. 4
Ex 15.0	Em 20.0	–	–	Expt. 5 <sup>b</sup>

<sup>a</sup> Description of the sample sets is given in the text.

<sup>b</sup> Emission ranges are 260–420 nm (ex. 230 nm), 270–440 nm (ex. 240 nm), 320–540 nm (ex. 290 nm), 360–620 nm (ex. 330 nm) (1624 data points in total).

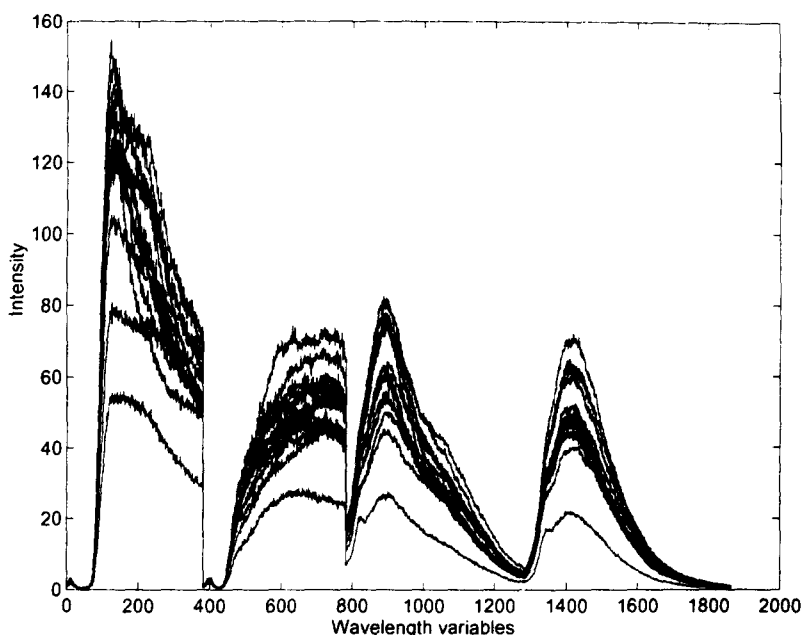


Fig. 11. Concatenated emission spectra of 15 different sugar samples (experiment 3, sample set DII) measured at four excitation wavelengths (230, 240, 290 and 330 nm). Emission variables 1–381 correspond to 245–435 nm; variables 382–782 correspond to 255–455 nm, variables 783–1283 correspond to 305–555 nm, variables 1284–1864 correspond to 345–635 nm.

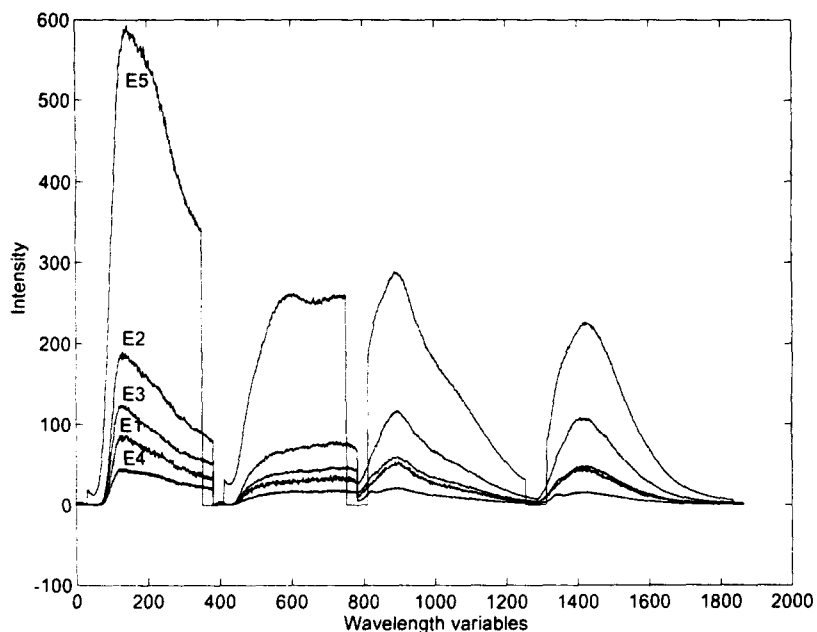
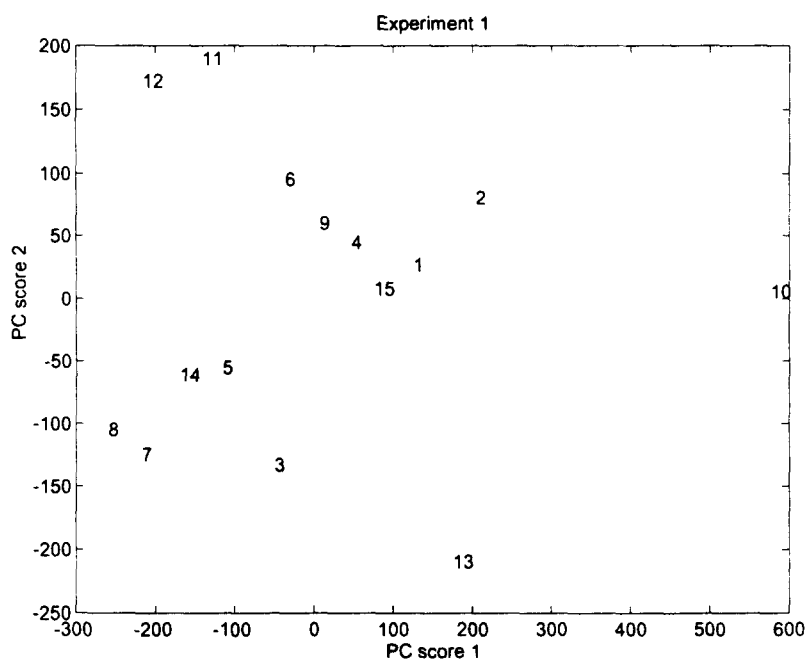


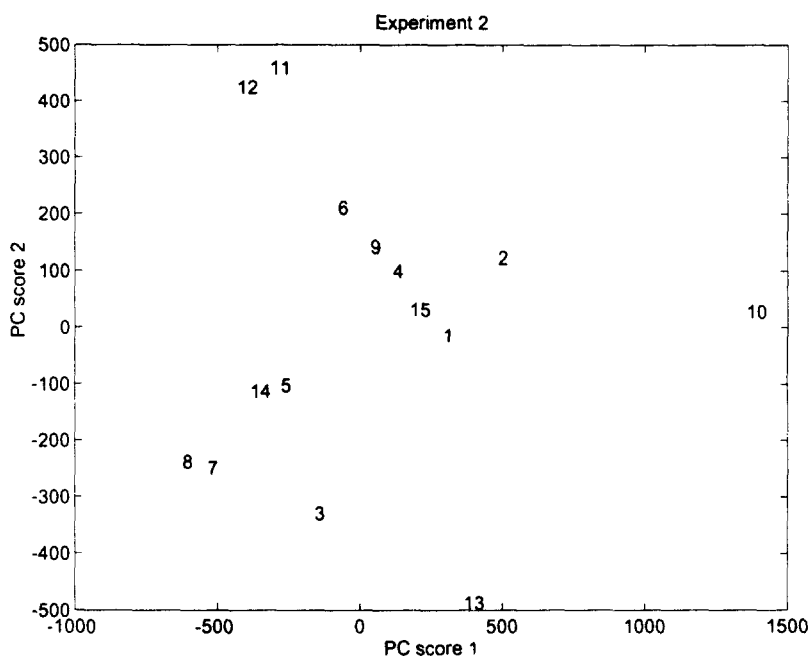
Fig. 12. Concatenated emission spectra of sugar sample 4 from experiments 1–5 (E1–E5). A replicate made in experiment 4 is included to illustrate the small replicate and measurement variation.

In order to make a preliminary investigation of the potential of using piecewise direct standardisation (PDS) on the fluorescence spectra, two experiments were performed with the same fluorescence spectrometer. In the first experiment, six samples (sample set A) are measured with xenon lamp 1 (old) mounted on the apparatus, while in the second experiment the

samples are measured with xenon lamp 2 (new) fitted. The raw fluorescence emission spectra of all six samples measured with excitation at 300 nm and with xenon lamp 1 are shown in Fig. 1. In Fig. 2 the emission spectra of samples 4 and 6 measured with both lamps are shown. A distinct difference between the spectra is observed and principal component analy-



(a)



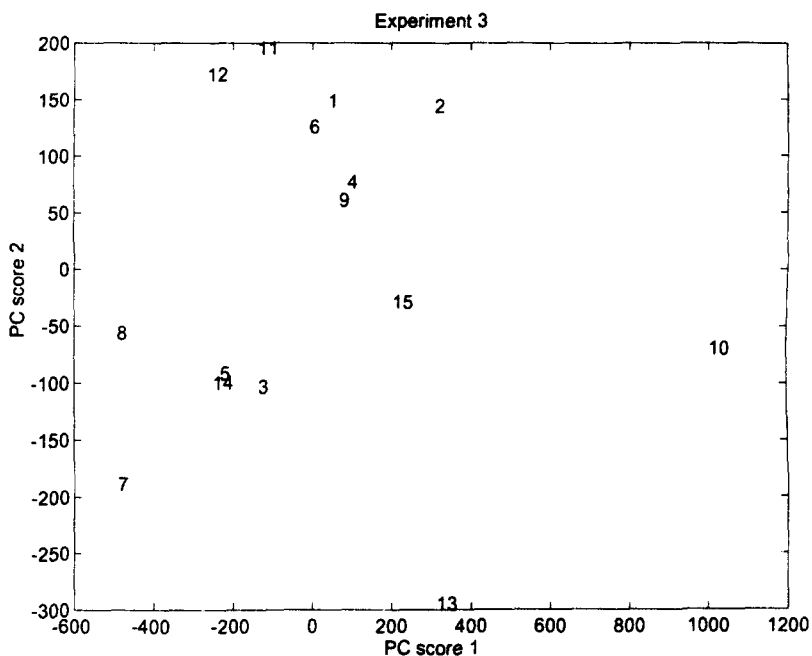
(b)

Fig. 13 (a) and (b).

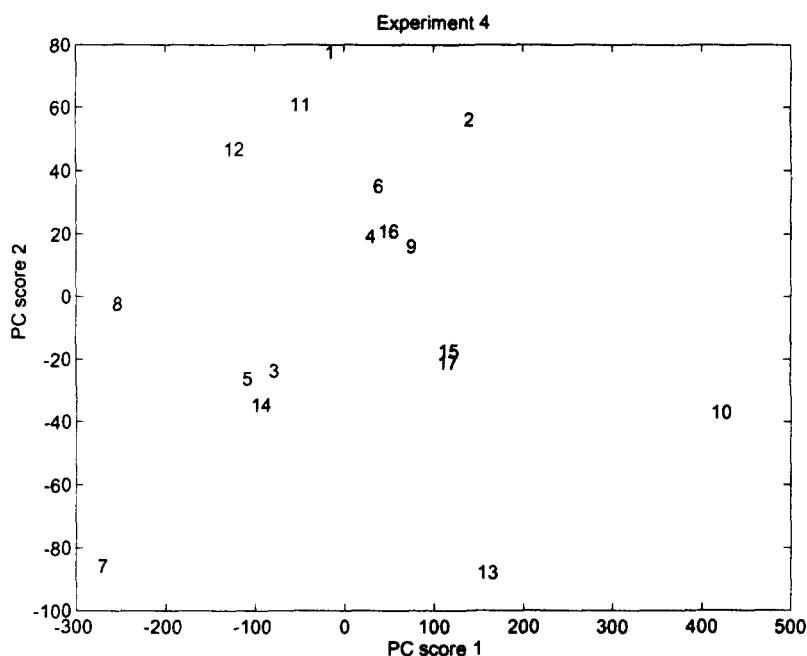
sis of all 12 emission spectra shows the displacement of identical samples as seen in the score plot in Fig. 3. Samples 2, 3 and 6 are chosen as transfer standards in the PDS algorithm (how to select appropriate standards is described in Ref. [22]). The PDS algorithm calculates a matrix capable of transforming the spectra measured with lamp 2 as if they were measured with lamp 1.

$$A_{\text{rec}} = \mathbf{B}M_{\text{transfer}}$$

where  $A_{\text{rec}}$  is a matrix with 6 objects and 441 variables containing the reconstructed spectra,  $\mathbf{B}$  is a  $6 \times 441$  matrix containing the spectra measured with lamp B and  $M_{\text{transfer}}$  is a  $441 \times 441$  transfer matrix. In this case the optimal window size was found to be 7, i.e. 7 spectral points in  $\mathbf{B}$  (this corresponds to an emission range of 3 nm) are used to model the chosen wavelength in  $A$ .



(c)

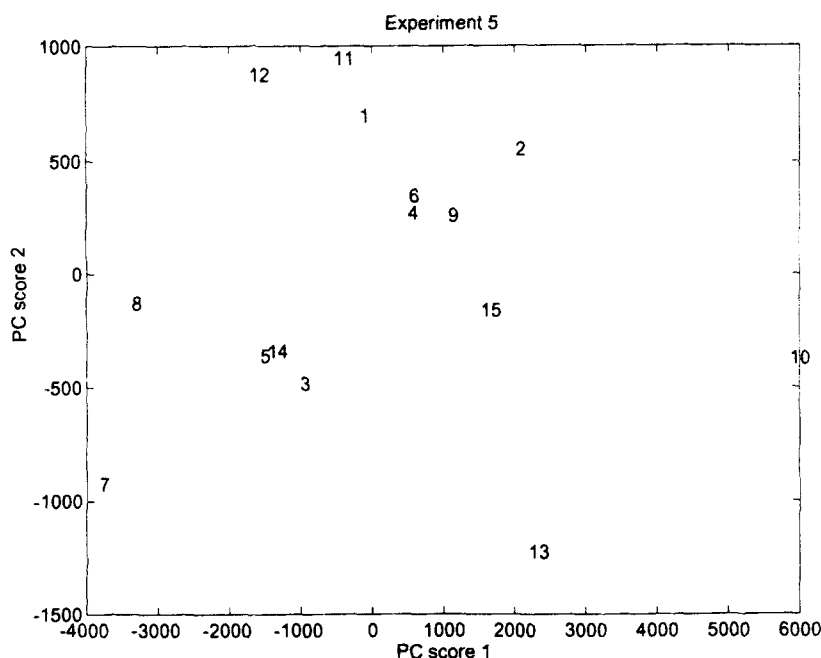


(d)

Fig. 13 (c) and (d).

In order to estimate the efficiency of PDS transformation, the true spectra measured with lamp 1 are compared with the transformed (reconstructed) spectra (Fig. 4). A PCA score plot of the 6 original lamp 1 spectra and the 6 transformed spectra is given in Fig. 5. As expected, the spectra of the three standards used in calculating the transfer matrix compare very well, but the three samples not included in the transfer calculation also compare satisfactorily.

This example indicates that the PDS is a valuable tool in transferring fluorescence spectra between different instrumental set-ups for the same fluorescence instrument and between different instruments. An extensive investigation of the performance of the PDS algorithm in combination with multivariate calibration methods of fluorescence spectra (including a larger sample set) is presented in Ref. [35].



(e)

Fig. 13. Score plots (PC1 against PC2) for (a) experiment 1; (b) experiment 2; (c) experiment 3; (d) experiment 4; and (e) experiment 5. In experiment 4 replicates of samples 4 and 15 are included (nos. 16 and 17, respectively).

### 3.2. Smoothing of spectra

The Perkin-Elmer LS 50B instrument is sold with two types of software smoothing filters, i.e. the chemist must choose either binomial or Savitzky-Golay smoothing of the raw data. To investigate the smoothing effect on spectral shape an OBEY program was written to obtain the raw unsmoothed data as output from the instrument (see Section 2). Ten emission spectra were recorded on standard block 2 (sample set B) for each smoothing type. A PCA of the 30 spectra shows three clusters each for one of the data types (Fig. 6). The data were not preprocessed in any way in order to illustrate the sizes of the two significant principal components (the first PC explains 99.988% of the total variance and the second PC explains 0.004% of the total variance). In Fig. 7 the corresponding loadings are plotted. The first loading vector primarily describes the intensity level, while the second loading substantiates that the differences indeed are very small but still significant. Furthermore, the second derivative shape of the second loading vector at peak locations might indicate peak broadening. This agrees with the broadening effect of the smoothing algorithms.

In the section below (3.3) it is shown that instrumental Savitzky-Golay smoothing of

data during recording of the spectra produces slightly larger prediction errors than when raw data are used as spectral input into a multivariate calibration model of the sugar concentration. It is concluded that no special smoothing algorithms are needed when the multivariate data approach is used.

### 3.3. Raman scattering

Raman scattering appears as a solvent-dependent emission occurring at longer wavelengths than the excitation wavelength [25]. Furthermore, the Raman frequency shifts are independent of the excitation frequency [25]. This can be a problem in classical univariate calibration if the spectral region of the analyte of interest is overlapped by the Raman scattering caused by the actual solvent (in this case water). A way of circumventing the influence of the Raman scattering is [25] (a) to subtract the blank spectrum from each of the sample spectra, assuming that the Raman scattering is independent of the analyte concentration and that the blank measurement is possible; (b) to employ appropriate filters (if it is possible in the actual case); or (c) to use another excitation wavelength with the risk of losing relevant analyte information.

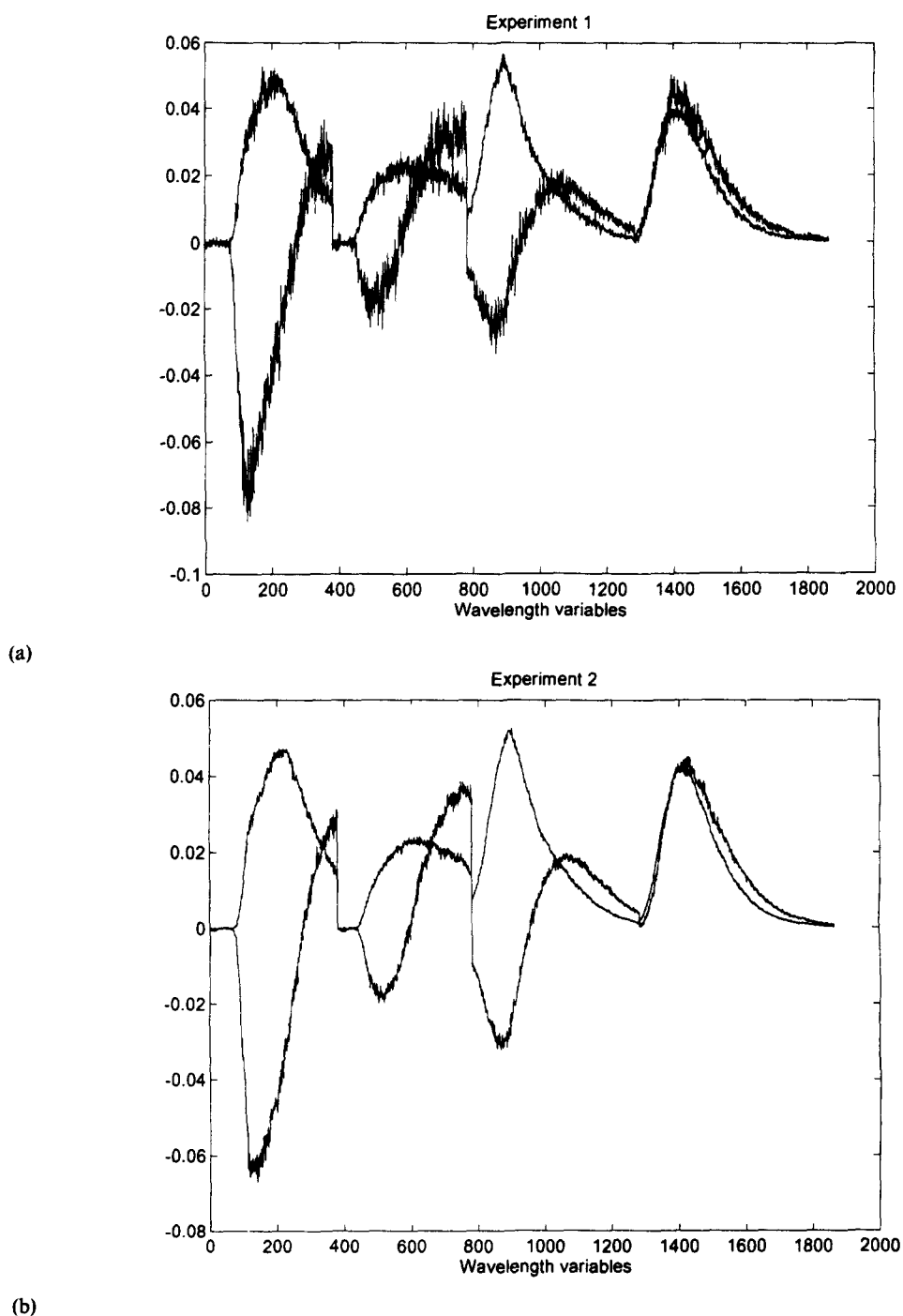
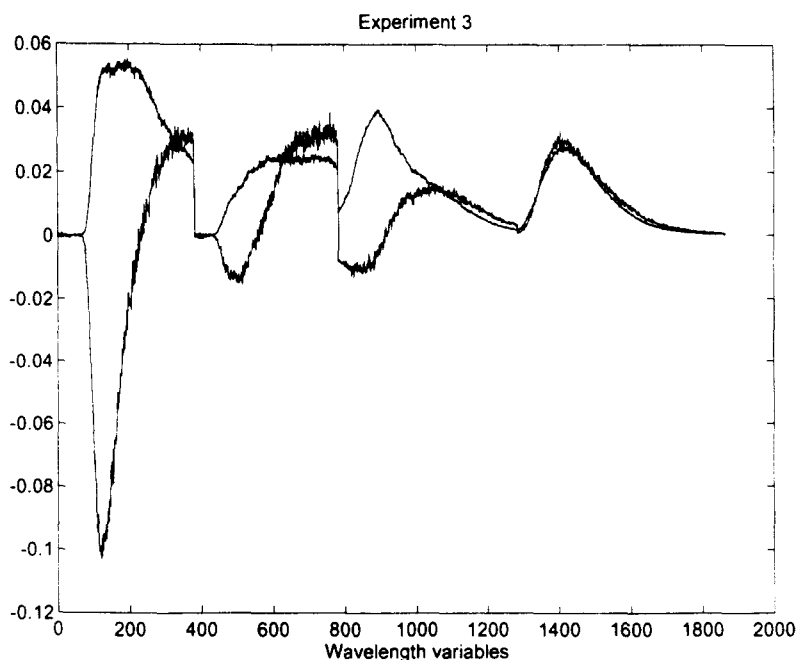


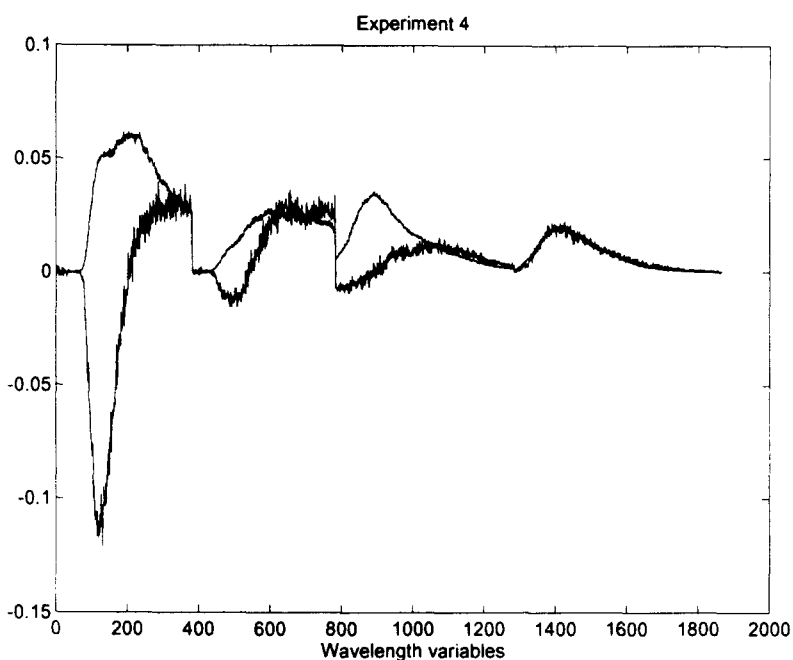
Fig. 14 (a) and (b).

A multivariate approach to the problems of Raman scattering is to analyse the whole emission spectrum including the Raman scatter. In multivariate calibration the Raman signals are treated as interferent(s) and cause no problems for the multivariate modelling of the analyte emission spectra. This is illustrated by the following examples.

The emission spectra of 11 solutions of a given refined sugar sample having different concentrations are measured at an excitation wavelength of 300 nm (sample set C). At this excitation wavelength, Raman scattering will occur at around 334 nm, overlapping the emission spectrum of the sample. The fluorescence emission spectra are depicted in Fig. 8. A full cross-validated PLS model with the fluo-



(c)

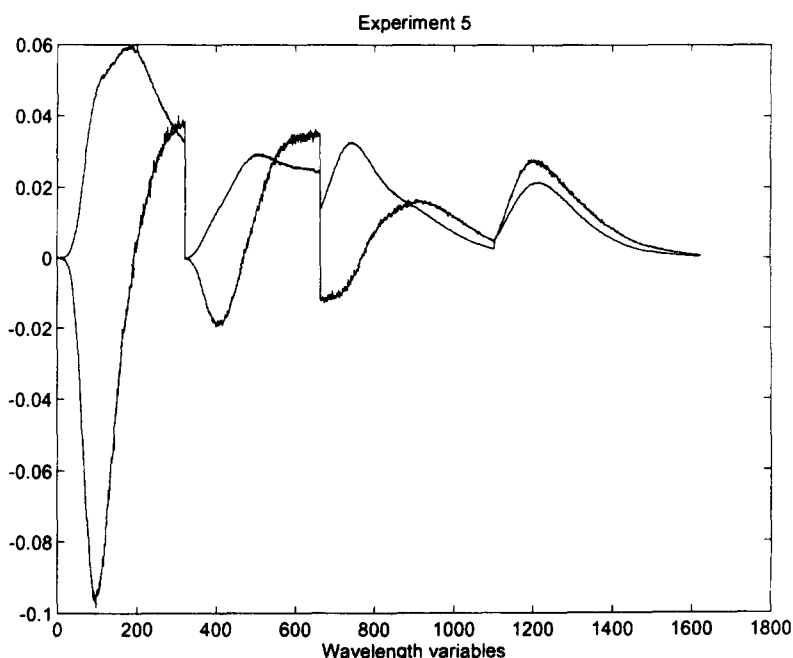


(d)

Fig. 14 (c) and (d).

cence emission spectra in  $X$  ( $11 \times 451$ ) and the concentrations in  $y$  ( $11 \times 1$ ) shows two significant components; the loading weights of the PLS analysis are given in Fig. 9. The Raman effect is seen in both loading vectors, i.e. both components are necessary in a full description of the data set (see Table 1). The PLS model was built without any preprocessing of the data to illustrate how PLS takes into account the

Raman scattering. If the data are mean centred, a one-factor solution is achieved due to the fact that the Raman scattering has the same intensity in all the samples. The results are compared with full cross-validated univariate calibration (at emission wavelength, 395 nm) in Table 1(A). The prediction errors obtained when using multivariate models are approximately five times smaller than the error



(e)

Fig. 14. The first two loading vectors for (a) experiment 1; (b) experiment 2; (c) experiment 3; (d) experiment 4; and (e) experiment 5. See comments in the text.

obtained in univariate calibration, i.e. even in the case of single analyte calibration multivariate modelling should be used. The prediction errors obtained with and without mean centring are of comparable size. In the case of no centring two factors are included in the model to handle the interfering Raman scattering.

The effect of the instrumental Savitzky-Golay smoothing of the raw data during recording of the spectra as input to the PLS model is shown in Table 1(B). The prediction errors obtained by this data preprocessing are slightly larger than when using raw data in the PLS model (see smoothing of spectra, section 3.2.).

In this one-analyte example (the same sugar solution diluted several times) PLS is capable of predicting the sugar concentrations of the samples without removal of the Raman signal in advance; in the case of several analytes being present and possible interaction between the Raman signals and the analyte signals, it is a necessity to use multivariate methods.

### 3.4. Selection of excitation wavelengths

The selection of a subset of excitation wavelengths is essential in order to minimise the time necessary to analyse a chemical sample. It takes approximately 40 s automatically to

record one emission spectrum by the OBEY software at a given excitation wavelength (including the adjustment of the excitation monochromator), i.e. to be able to measure at least 20 samples per hour for on-line/at-line purposes, four to five excitation wavelengths have to be chosen. One way of selecting the excitation wavelengths from a large excitation–emission matrix (EEM) is the principal variable [26] (PV) algorithm, which is illustrated by the analysis of a refined sugar sample (sample 1 from sample set DI). An EEM is recorded at the excitation wavelengths 230–350 nm with 10 nm intervals, i.e. 13 emission spectra are obtained in the emission range 245–685 nm (excitation and emission monochromator slit widths were 8.0 nm and 5.0 nm, respectively). Rayleigh scattering peaks are removed from the emission spectra by filling with zeroes. In Fig. 10 a three-dimensional plot of the EEM is depicted. The PV algorithm applied to this EEM determines the optimal subset of excitation wavelengths to be 230 nm, 330 nm, 290 nm and 240 nm. The variances computed explain 60.28%, 33.91%, 3.57% and 0.75%, respectively, of the total variance. In total 98.51% of the original variance is explained. “Optimal” signifies a compromise between variance explained and the precision (see Ref. [26]).



### 3.5. Multivariate calibration of fluorescence spectra

The multivariate calibration of the ash content in sugar samples is important in developing fast on-line/at-line methods for detecting the end-product quality, which among others, is described by the ash content. Clear solutions of 15 refined sugar samples with measured levels of ash content determined at the laboratory of The Danish Sugar Factories (International Commission for Unified Methods of Sugar Analysis, ICUMSA) are analysed by spectrofluorimetry without pH correction.

Table 3  
Prediction of ash content. Performance of PLS modelling in the five experiments performed

	10 <sup>3</sup> RMSEP <sup>a</sup>	10 <sup>3</sup> Bias <sup>b</sup>	R <sup>c</sup>
Expt. 1 (a)	1.06	-0.067	0.903
(b)	1.11	-0.071	0.763
Expt. 2 (a)	1.06	-0.200	0.916
(b)	1.07	0.000	0.788
Expt. 3 (a)	0.775	0.067	0.951
(b)	1.04	0.071	0.826
Expt. 4 (a)	0.516	0.000	0.979
(b)	0.463	-0.071	0.966
Expt. 5 (a)	0.683	-0.20	0.967
(b)	0.707	-0.21	0.933

Key: (a), All 15 samples; (b), sample 10 left out.

<sup>a</sup> RMSEP =  $\sqrt{\sum_{i=1}^N (C_i^{\text{predicted}} - C_i^{\text{reference}})^2 / N}$ , where  $C_i^{\text{predicted}}$  is the PLS estimated concentration,  $C_i^{\text{reference}}$  is the reference value and  $N$  is the number of samples.

<sup>b</sup> Bias =  $\sum_{i=1}^N (C_i^{\text{predicted}} - C_i^{\text{reference}}) / N$ .

<sup>c</sup> Correlation.

Table 4  
Ash content reference value and predicted ash content, from PLS modelling of fluorescence spectra measured on sugar samples in experiment 3

Sample	Reference	Predicted
1	0.012	0.011
2	0.009	0.010
3	0.012	0.012
4	0.010	0.010
5	0.011	0.011
6	0.010	0.011
7	0.010	0.011
8	0.014	0.014
9	0.009	0.010
10	0.003	0.003
11	0.012	0.011
12	0.012	0.011
13	0.008	0.007
14	0.010	0.011
15	0.008	0.008

Based on the results in the previous paragraph, the chosen excitation wavelengths are 230 nm, 240 nm, 290 nm and 330 nm, with the respective emission wavelength ranges 245–435 nm, 255–455 nm, 305–555 nm and 345–635 nm. The four emission spectra are concatenated, i.e. one sample is associated with 1864 spectral data points.

Five experiments were performed to investigate the influence of excitation monochromator and emission monochromator slit widths as well as the sugar concentration on the modelling of the ash content (Table 2). Rayleigh scattering becomes very broad when the slit widths are fully open, so the region of each recorded emission spectrum in experiment 5 is reduced by 30 nm (see Table 2).

Raw emission spectra from the 15 measurements in experiment 3 are illustrated in Fig. 11, and raw emission spectra of sample 4 from all five experiments, including a replicate from experiment 4, are given in Fig. 12. For qualitative comparison of the spectral shapes produced under different experimental conditions, a PCA was performed on each of the data sets. In Figs. 13a–13e, score plots (score 1 against score 2) including two replicates in experiment 4, are shown. The orders of samples are much alike in experiments 1 and 2 and in experiments 4 and 5, and it is seen that experiment 3 fits into the shift of the samples between experiments 2 and 4. The amount of variance explained by the first two significant factors are 81.6%, 86.4%, 89.9%, 91.3% and 93.7%, respectively. Figs. 14a–14e shows the first two loading vectors of each experiment. The shapes of the loading vectors are much alike, indicating that the experimental conditions give rise to the same kind of measured emission spectra. Moreover, it is observed that the noise level increases with decreasing slit widths and that it increases with decreasing sugar concentration.

A quantitative measure is obtained by looking at PLS models of the ash content. For each experiment, a full cross-validated [27,34] PLS model of the ash content is performed. RMSEP, Bias and the correlation coefficient  $R$  for all experiments with and without sample number 10 are given in Table 3. Reference values and predictions obtained in experiment 3 are given in Table 4. Sample 10 has a low level of ash content but it is still very well predicted by the PLS models. It should be stressed that the prediction of sample 10 is based on models excluding this sample, i.e. sample 10 is a

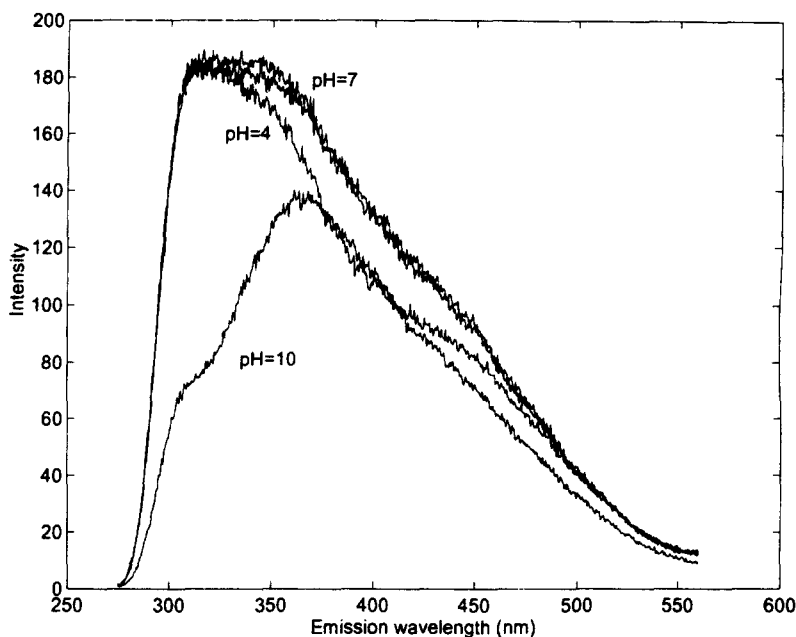


Fig. 15. Emission spectra of a refined sugar sample dissolved in buffer solutions at pH 4, 7, 10. The excitation wavelength is 230 nm. A replicate is made of the sample buffered to pH 7.

“good” outlying sample stabilising the models, as seen when comparing RMSEP and  $R$ .

From the prediction results it is seen that the selective wavelength information lost by using fully open slit widths is recovered in the signal-to-noise ratio by the employment of multivariate modelling. Furthermore, low concentrations of sugar yield lower prediction errors, indicating some kind of quenching or perhaps a viscosity effect. The conclusion is that the sugar concentration should be within the range 10–20% and the slit widths should be in the middle region of their respective extremes (excitation slit, 2.5–15 nm; emission slit, 2.5–20 nm). However, it is a very small difference in correlation ( $R$ ) from the middle region slits to the fully open slits.

It is demonstrated how the multivariate method PLS is capable of utilising the information in fluorescence spectra despite very different conditions of measurement with respect to slit widths and sugar concentrations. These promising results are confirmed on a larger scale when about 90 refined sugar samples with and without pH correction are analysed [20]. Similar correlations to the ash content ( $R = 0.93$ ) are documented as well as high correlations to quality parameters like amino-nitrogen and colour.

### 3.6. The pH and corrected spectra

#### pH

It is well known that the spectral shapes of fluorescing components are pH-dependent. This is confirmed in Fig. 15 where the difference in spectral shape obtained at pH 4.0, 7.0 and 10.0 for a given refined sugar sample is illustrated. The measurement at pH 7.0 has been repeated to show the uncertainty of a sample replicate. In ref. [20] comparisons of prediction errors from a large sample set of refined sugar samples (about 90) dissolved in pure water and in buffer solution at pH 7.0 are presented. The value of  $R$  with respect to the ash content is shown to be 0.92 with pH correction and 0.93 without correction. Even though the difference in prediction error is insignificant, the spectral shapes are systematically different in the two experiments, indicating that the information content is changed when going from water solutions to buffered solutions. It may be that some sample and process information is lost on pH correction of the solutions.

#### Corrected emission spectra

Throughout this presentation, uncorrected spectra are recorded in all experiments. When the fluorescence spectra are corrected [36] all spectra are wavelength multiplied by the same correction spectrum which stems from a stan-

standard measurement of, for example, a quinine sulphate solution, which introduces a reproducible small bias because of the inherent limitations in the choice of the standard. In multivariate calibration methods like PCR and PLS, however, a more general optimal weighting of wavelengths is found by the algorithms, so the present choice is not to correct the fluorescence spectra before they are introduced to the multivariate calibration model. In the area of transferring spectra between instruments, the method of piecewise direct standardisation using several transfer standards is much more preferable than using corrected spectra based on only one standard sample.

It should be noted that if resolution methods like rank annihilation factor analysis [9] (RAFA) and the generalised rank annihilation method [37] (GRAM) are employed and a library search of the resolved fluorescence spectra is performed, the measured spectra have to be corrected to compare with library spectra.

#### 4. Conclusions

The outlined multivariate problems with regard to the evaluation of different data, all due to complex natural samples, individual spectrofluorimeters, Raman scattering, smoothing algorithms, and selection of excitation/emission wavelengths, may in part explain why fluorescence spectroscopy is a tool infrequently chosen by the analytical chemist as opposed to other spectroscopic methods. It has been demonstrated that by the use of multivariate chemometric methods, these problems are circumvented and fluorescence spectroscopy can be an important tool for the analytical chemist.

The NIR analogy of predicting reference values from spectra has been transferred to spectrofluorimetry, by determining the ash content in refined sugar samples from concatenated emission spectra. The potential of autofluorescence in this context is enormous for the analysis of all kinds of chemical, biological and environmental samples including on-line/at-line process analyses [21]. A further enhancement of spectrofluorimetry can be obtained by utilising the instrumental methods of polarisation and synchronous fluorimetry in combination with multivariate statistics and by the development and implementation of N-way algorithms [37–40] capable of dealing with the higher-order data structures produced by a full excitation–emission scan.

#### Acknowledgements

The author thanks Ole Hansen, Lars Bo Jørgensen and John Jensen, The Danish Sugar Factories (DDS Development Centre, Maribovej 2, Post Box 119, DK-4900 Nakskov, Denmark) for a stimulating cooperation and for providing the sugar samples. Professor Lars Munck is acknowledged for valuable and inspiring discussions during the experimental work and during the preparation of the manuscript. The investigation is sponsored by funds to Professor Lars Munck from the Danish Research Councils 13-4804-1 (agriculture) and 16-5180-1 (technology) and from the Nordic Industry Foundation project P93149.

#### References

- [1] G.G. Guilbault, *Practical Fluorescence*, 2nd edn., Marcel Dekker, New York, 1990.
- [2] C.A. Parker, *Photoluminescence of Solutions*, Elsevier, 1968.
- [3] S.D. Brown, T.B. Blank, S.T. Sum and L.G. Weyer, *Anal. Chem.*, 66 (1994) 315R.
- [4] W.F. McClure, *Anal. Chem.*, 66 (1994) 43A.
- [5] T. Isaksson, *Doctoral Thesis*, Chalmers Tekniska Högskola, 1990.
- [6] D. Bertrand and C.N.G. Scotter, *Appl. Spectrosc.*, 46 (1992) 1420.
- [7] C.-N. Ho, G.D. Christian and E.R. Davidson, *Anal. Chem.*, 50 (1978) 1108.
- [8] C.-N. Ho, G.D. Christian and E.R. Davidson, *Anal. Chem.*, 52 (1980) 1071.
- [9] C.-N. Ho, G.D. Christian and E.R. Davidson, *Anal. Chem.*, 53 (1981) 92.
- [10] C.J. Appellof and E.R. Davidson, *Anal. Chim. Acta*, 146 (1983) 9.
- [11] S.L. Neal, E.R. Davidson and I.M. Warner, *Anal. Chem.*, 62 (1990) 658.
- [12] D.S. Burdick, X.M. Tu, L.B. McGown and D.W. Millican, *J. Chemometrics*, 4 (1990) 15.
- [13] W. Lindberg, J.-Å. Persson and S. Wold, *Anal. Chem.*, 55 (1983) 643.
- [14] M. Sjöström, S. Wold, W. Lindberg, J. Persson and H. Martens, *Anal. Chim. Acta*, 150 (1983) 61.
- [15] B. Pedersen and H. Martens, in L. Munck (Ed.), *Fluorescence Analysis in Foods*, Longman, Singapore, 1989, Chapter 13.
- [16] S.A. Jensen, L. Munck and H. Martens, *Cereal Chem.*, 1982, 59, 477.
- [17] S.A. Jensen and H. Martens, *Cereal Chem.*, 60 (1983) 171.
- [18] S.A. Jensen and H. Martens, in H. Martens and H. Russwurm (Eds.), *Food Research and Data Analysis*, Applied Science, London, 1982, pp. 253–69.
- [19] R.F. Madsen, W. Kofod Nielsen, B. Winstrøm-Olsen and T.E. Nielsen, *Sugar Technol. Rev.*, 6 (1978/79) 49.
- [20] L. Nørgaard, *Classification and prediction of quality and process parameters of beet sugar and thick juice by fluorescence spectroscopy and chemometrics*, *Zuckerindustrie*, in press.

- [21] L. Munck, Chapter 1 in *Fluorescence Analysis in Foods*, Edited by L. Munck, Longman, 1989.
- [22] Y. Wang, D.J. Veltkamp and B.R. Kowalski, *Anal. Chem.*, 63 (1991) 2750.
- [23] Y. Wang and B.R. Kowalski, *Anal. Chem.*, 65 (1993) 1174.
- [24] Y. Wang, M.J. Lysaght and B.R. Kowalski, *Anal. Chem.*, 64 (1992) 562.
- [25] C.A. Parker, *Analyst*, 84 (1959) 446.
- [26] A. Höskuldsson, *Chemometr. Intell. Lab. Syst.*, 23 (1994) 1.
- [27] H. Martens and T. Næs, *Multivariate Calibration*, 2nd edn., Wiley, New York, 1993.
- [28] B.M. Wise, *Chemometrics Toolbox*, Version 1.3 (<ftp://ra.nrl.navy.mil/MacSciTech/chem/chemometrics/PLSToolbox13/>).
- [29] S. Wold, K. Esbensen and P. Geladi, *Chemometr. Intell. Lab. Syst.*, 1987, 2, 37.
- [30] P. Geladi and B.R. Kowalski, *Anal. Chim. Acta*, 185 (1986) 1.
- [31] A. Höskuldsson, *J. Chemometrics*, 2 (1988) 185.
- [32] P.J. Brown, *Anal. Proc.*, 27 (1990) 303.
- [33] P. Geladi and B.R. Kowalski, *Anal. Chim. Acta*, 185 (1986) 19.
- [34] S. Wold, *Technometrics*, 20(4) (1978) 397.
- [35] L. Nørgaard, *Direct standardisation in multi wavelength fluorescence spectroscopy*, *Chemometrics and Intelligent Laboratory Systems*, in press.
- [36] J.W. Hofstraat, and M.J. Latuhihin, *Appl. Spectrosc.*, 48 (1994) 436.
- [37] E. Sanchez and B.R. Kowalski, *Anal. Chem.*, 58 (1986) 496.
- [38] S. Wold, P. Geladi, K. Esbensen and J. Öhman, *J. Chemometrics*, 1 (1987) 41.
- [39] H. Henrion, *Chemometr. Intell. Lab. Syst.*, 25 (1994) 1.
- [40] A.K. Smilde and D.A. Doornbos, *J. Chemometrics*, 5 (1991) 345.