ELSEVIER

Chemometrics and Intelligent Laboratory Systems 45 (1999) 39–53

Chemometrics and intelligent laboratory systems

Using contrasts as data pretreatment method in pattern recognition of multivariate data

W. Wu, Q. Guo, D. Jouan-Rimbaud, D.L. Massart *

ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium

Received 7 January 1998; accepted 4 April 1998

Abstract

A contrast method originally proposed by Spiegelman [C.H. Spiegelman, Calibration: a look at the mix of theory, methods and experimental data, presented at Compana '95, Wuerzburg, Germany.] is modified to pretreat multivariate data for classification. Three NIR data sets and one pollution data set are used as examples. Our results show that the contrast method greatly improves the ratios of between- to within-class variance. It is more powerful than offset correction, SNV, first- and second-derivative methods in the cases studied. This conclusion does not depend on the type of classifier used. Regularised discriminant analysis (RDA) and partial least squares (PLS2) with univariate feature selection based on Fisher's ratio were applied here. There is a risk that chance correlations occur after the contrast pretreatment. The chance correlation decreases after first eliminating un-informative variables using the modified Uninformative Variable Elimination (UVE)-PLS method.

Keywords: Data pretreatment; Contrast; NIR; Pattern recognition; Regularised discriminant analysis (RDA); Partial least squares (PLS)

Contents

* Corresponding author. Tel.: +32-2-477-4737; Fax: +32-2-477-4735; E-mail: fabri@bub.vub.ac.be

0169-7439/99/$ - see front matter
PII: S0169-7439(98)00088-4

1. Introduction

In the analysis of multivariate data such as NIR data, data pretreatment plays a very important role. Proper pretreatment can greatly improve the results, since it removes or minimises the multiplicative interferences due to scatter, particle size effects, baseline shifts, instrument noise, etc. Many data pretreatment methods have been proposed in the literature such as multiplicative scatter correction (MSC), first derivative, second derivative, second-derivative/logarithm (SDL), standard normal variate (SNV) and offset correction [1–4]. Offset correction can only correct parallel shifts in the baseline, and cannot correct for the slope changes in the baseline. The derivative and SDL methods can remove both parallel shifts and slope changes, but at the same time enhance the noise in the spectra. MSC is based on using the mean spectrum of the data set, so that one cannot treat spectra individually. The SNV method can be applied to individual spectra [5] and yields good results, but is subject to a closure problem [6]. Recently, Spiegelman suggested a new method based on contrasts to pretreat spectral data prior to calibration [7]. In this paper, this method is adapted and applied for classification purposes. Other data pretreatment methods such as offset correction, first derivative, second derivative and SNV are also applied for comparison by using partial least squares (PLS2) and regularised discriminant analysis (RDA) as classifiers.

2. Theory

2.1. Using contrasts as data pretreatment for classification

Spiegelman determines all pairwise contrasts, i.e., all pairwise differences of absorbance at different wavelengths [7]. In this way, he includes, in a certain sense, the offset-correction method and the first derivative. Indeed, in the offset correction method, one subtracts a given value such as the mean value of the absorbance at the first few wavelengths from the spectrum to correct for the baseline shift. The first derivative method is based on subtraction of absorbance at the neighbouring wavelengths. Moreover, in this way one includes automatically the comparison of certain important features, e.g., two peaks.

There are two steps in Spiegelman's contrast method: the first one is to create the new variables, and the second one is to select a subset of those new variables. Suppose $\mathbf{X}_{n\times m}$ is a spectral data matrix with n objects and m wavelengths. In the first step of this method, a new matrix is formed by taking the ith column minus the jth column provided that j is greater than i. The m original columns of the $\mathbf{X}$ matrix are also included in this new matrix. A new matrix with $(m(m-1)/2+m)$ columns is formed.

In the second step, Spiegelman selects m columns from this matrix. His application was calibration, while ours is classification. Therefore, in this second step, we calculate Fisher's criterion, i.e., the ratio of between- to within-class variance for each column of the new matrix:

$$e_i = \sum_{j=1}^{k} n_j\left(\bar{y}_{ji} - \bar{y}_{\cdot i}\right)^2 \Bigg/ \sum_{j=1}^{k} \left(n_j - 1\right) s_{ji}^2 \qquad (1)$$

where $j = 1, 2, 3, \ldots, k$, k is the number of classes, n_j is the number of objects in class j, $\bar{y}_{ji}$ denotes the mean absorbance of the objects belonging to class j at the ith wavelength, $\bar{y}_{\cdot i}$ denotes the mean absorbance of the objects belonging to all classes at the ith wavelength and s_{ji} is the standard deviation of the

absorbance of the objects belonging to class j at this wavelength. For each variable, the Fisher's weight is calculated, and the variables are selected which have the highest Fisher's weights. The m columns with the highest Fisher's criterion are selected in the new matrix, $\mathbf{X}_{\text{new}}$.

When the number of variables (m) of the original data is large, such as for NIR data, the number ($m(m+1)/2$) of columns of the new matrix created after the first step is extremely large, which requires (too) much computer memory. To solve this problem, we propose to perform the second step during the first step instead of after the first step. We start with $\mathbf{X}_{n\times m}$ containing the original variables as a temporary new matrix, and estimate their Fisher's ratios. During the first step, after a few new variables (instead of all new variables) have been created, their Fisher's ratios are calculated. Those new variables together with the variables in the temporary matrix are ranked according to their Fisher's ratios, and the first largest m variables are selected to replace the variables in the temporary matrix. In the end, the variables remaining in the matrix are the same as those obtained using a two-step procedure. The modified algorithm is as follows:

1. Obtain the initial matrix $\mathbf{X}_{\text{new}}$ with n rows and m columns by using all the variables in $\mathbf{X}$, and estimate their Fisher's ratios.
2. For a given variable i ($i = 1$ to $m-1$), calculate the differences of variables of $\mathbf{X}$ between i and j ($j = i+1$ to m) as the new variables, estimate Fisher's ratios for these new variables and add these new variables to $\mathbf{X}_{\text{new}}$, yielding an extended matrix $\mathbf{X}_{\text{new-ext}}$. Values of the new variables can be positive, negative or zero.
3. Rank the variables in $\mathbf{X}_{\text{new-ext}}$ according to their Fisher's ratios in decreasing order.
4. Update the $\mathbf{X}_{\text{new}}$ matrix by keeping only the first m variables in $\mathbf{X}_{\text{new-ext}}$, i.e., deleting the last variables.
5. Repeat steps 2–4, until i reaches $m-1$, then the final $\mathbf{X}_{\text{new}}$ is the output of the pretreated $\mathbf{X}$.

The resulting $n \times m$ matrix is exactly the same as the one that would be obtained with the two-step approach. This algorithm can easily be run on a personal micro-computer without memory problem. However, it needs slightly more computing time than the two-step approach.

2.2. Classifier

In this study, RDA and PLS2 are used as classifiers. RDA has been described in the paper of Friedman [8] and in our previous work [9]. In PLS, if there is only one dependent variable, PLS1 is used; otherwise one uses PLS2. The difference is that PLS2 can simultaneously model several correlated columns in matrix $\mathbf{Y}$, while PLS1 builds separate models for each column in matrix $\mathbf{Y}$. In PLS2 [10], the dependent variables of matrix $\mathbf{Y}$ are defined as binary values with 1 for the corresponding class and 0 for the other classes. For instance, [0 1 0] is used as $\boldsymbol{y}$ vector of an object from class 2. The correct classification rate (CCR) based on leave-one-out cross-validation is used as the classification result. In PLS2, the CCR is calculated as described in Refs. [11,12]. For each object, the predicted vector of $\boldsymbol{y}$ is calculated. The object is assigned to the class for which the predicted value is the only one higher than 0.5. For instance, if the predicted $\boldsymbol{y}$ is [0.1, 0.9, 0.2], the object is assigned to class 2. When the prediction is, e.g., [0.1, 0.9, 0.8] or [0.1, 0.4, 0.2], the object is regarded as a misclassified object. When only two classes are studied, one can also use PLS1. We prefer to use PLS2 because it is more strict. For instance, if the predicted $\boldsymbol{y}$ is [0.1, 0.2], the object is more reasonably regarded as a misclassified object. When PLS1 is applied, the object is always assigned to one of the two classes.

2.3. UVE-PLS and its modification

Uninformative Variable Elimination method (UVE-PLS) was recently developed to eliminate uninformative variables for calibration of NIR data [13]. The original data are used to build the model in this method. Artificial random variables are added to the data as a reference so that those variables which play a less important role in the model than the random variables are eliminated. The importance of variables is estimated by the ratios of b-coefficients to their standard deviations in the PLS1 model, where b-coefficients refer to coefficients in the final PLS regression vector. The standard deviations of b-coefficients are obtained through leave-one-out cross-validation. Several versions of UVE-PLS were described in Ref. [13]. Here the simplest version is used and adapted. PLS2 is used instead of PLS1 for clas-

sification of data with more than two classes, and CCR is used instead of root mean square error of prediction (RMSEP). The modified algorithm is as follows:

1. With all variables, use PLS2 to estimate CCR based on leave-one-out cross-validation. The optimal number (n_f) of factors is considered to be the one for which CCR reaches the highest value.
2. The data matrix is expanded by adding the same number of random variables multiplied by 10^{-10} so that these variables do not influence the model.
3. With the expanded data matrix, n b-coefficient matrices $\mathbf{B}_{2m\times p}$ are obtained by PLS2 with n_f factors based on the leave-one-out cross-validation, where n is the number of objects, m the number of variables, p the number of classes.
4. For the first column of **B**, calculate the ratio of the absolute value of the b-coefficient to its standard deviation for each row (variable) separately.
5. Use the maximum value of the ratios for rows $m+1$ to $2m$ (random variables) as the threshold value, and select a subset of variables whose ratios are higher than the threshold value.
6. If there are more than two classes, repeat steps 4–5 $p-1$ times for each of the first $p-1$ columns of **B**. In the end, obtain $p-1$ subsets of variables, and find the common variables in these subsets.
7. Use these common variables to build the optimal PLS2 model based on leave-one-out cross-validation, and provide the CCR for this model.
8. Repeat steps 2–7 with $n_f = n_f - 1$, until the optimal CCR does not increase. The CCR and subset of the common variables are the final outputs of UVE-PLS.

The modified UVE-PLS is used to eliminate the uninformative variables. One can perform UVE-PLS before or after data pretreatment. Here, UVE-PLS is applied after data are pretreated by the contrast method.

3. Experimental

3.1. Data

Four data sets were studied. Data sets 1–3 are NIR data sets which have been used in Refs. [11,14]. They

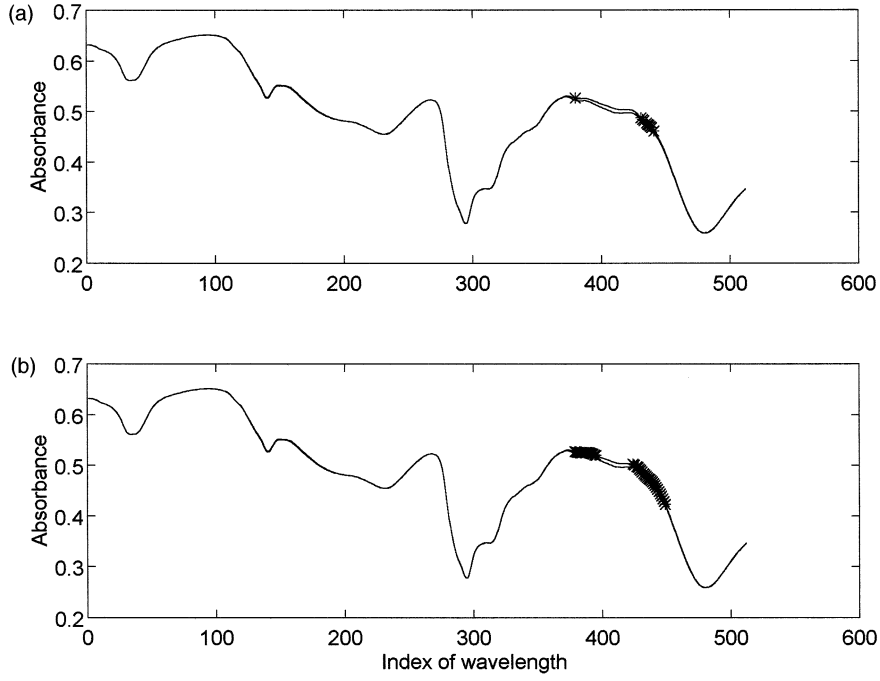

Fig. 1. The mean spectra for the two classes of data set 1. ∗Denotes the selected variables by (a) PLS with univariate feature selection and (b) UVE-PLS.

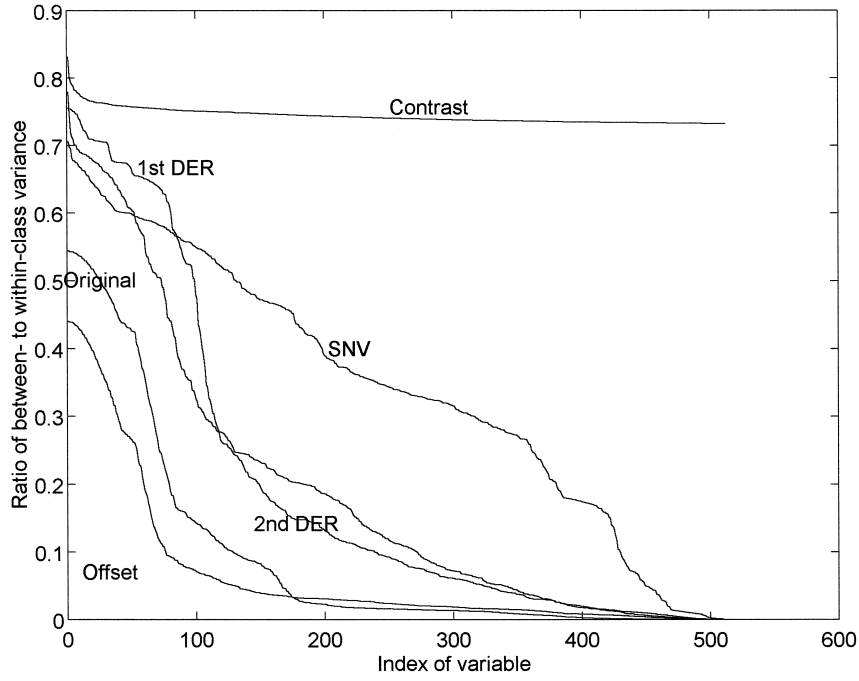

Fig. 2. The ratio of between- to within-class variance versus the index of the variables after ranking with different pretreatment methods including the original data; data set 1.

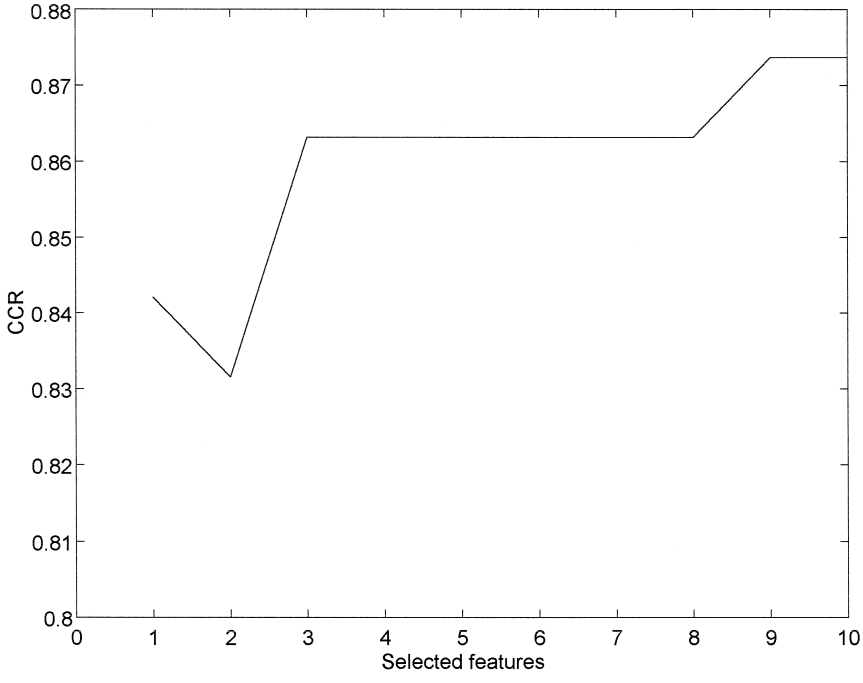

Fig. 3. The classification results (CCR) for RDA based on cross-validation versus number of selected features.

Table 1
The classification results (CCR) based on leave-one-out cross-validation for RDA with the optimal number of selected variables after different pretreatments; data set 1

Method	Variables	(λ, γ)	CCR
Contrast	9	(0, 1)	0.874
Original	10	(0.4, 0)	0.684
Offset	10	(0, 0.2)	0.674
SNV	15	(0.6, 0)	0.768
First DER	25	(0.6, 0.2)	0.621
Second DER	19	(1, 1)	0.653

were measured with a NIR instrument Bran + Luebbe. The spectra are presented as $\log(1/R')$ absorption values, where R' is the reflectance of the sample versus that of a white ceramic reflectance. For convenience, the wavelength is expressed by the index in the resulting data matrix. Data set 4 is an industrial pollution data set; it has also been used in earlier publications [14,15].

Data set 1 contains 95 spectra (1130–2152 nm, 512 wavelengths) of pure and impure butanol. The pure class consists of 42 spectra, and the impure class consists of 53 spectra containing different concentrations of water (from 0.02% to 0.32%). The goal is to identify if a sample is pure or impure in order to control the manufacturing process in industrial practice.

Data set 2 is an artificial data set. The basis are the 42 spectra of pure butanol of data set 1. To this, another spectrum was added to produce a data set for which the degree of difficulty can be adapted. Both spectra are measured in the same wavelength range (512 wavelengths). The 42 butanol spectra are randomly divided into two subsets a and b with the same number of objects. Subset a is used as class 1. Class 2 is made of 99% subset b and 1% of the other spectrum of polymer.

Data set 3 consists of 60 spectra (1376–2398 nm, 512 wavelengths) of three batches of excipients which are made by mixing cellulose, mannitol, sucrose, sodium saccharin and citric acid in different proportions. Each class contains the spectra measured for 20 samples of the same batch of excipients.

Data set 4 contains 49 objects related to two different kinds of pollution (galvanisation and steelworks sludges) with 10 variables including the pH, Cd, Cr, Pb, Cu, residual at 105°C, residual at 660°C, Zn, CN^- and Ni. Three outlier objects were elimi-

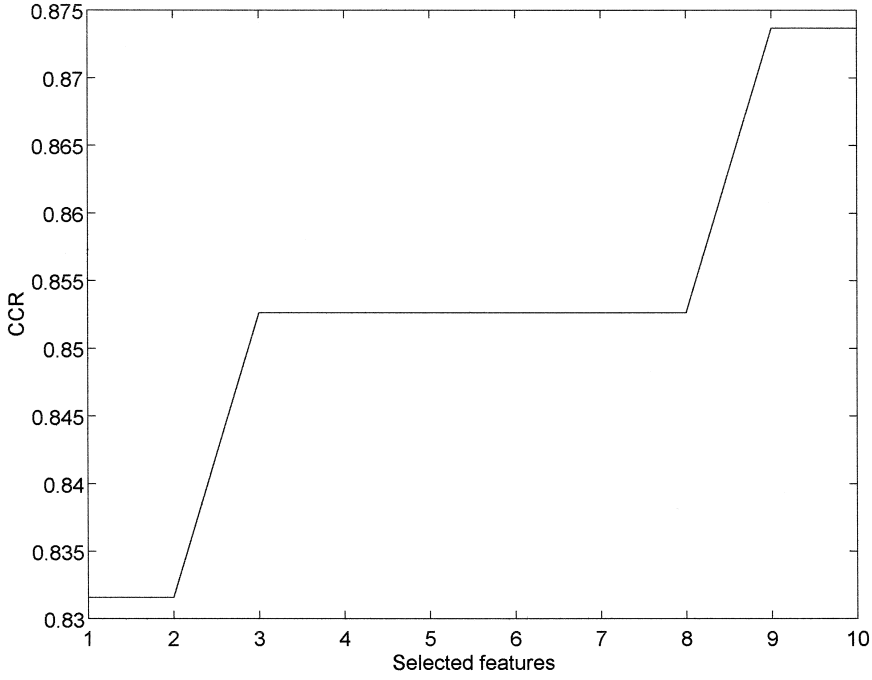

Fig. 4. The optimal classification result (CCR) for PLS based on cross-validation versus number of selected features.

nated [15]. Class 1 consists of 30 objects and class 2 consists of 19 objects.

4. Results and discussion

4.1. Data set 1

Fig. 1 shows the mean spectra for the two classes for this data set. There is a difference between the two classes only in the region of wavelengths 380–440 which corresponds to the characteristic wavelength of the impurity of water.

As described earlier, Spiegelman's contrast method is an extension of derivative and offset correction methods. One expects that the contrast method will give similar results to these other methods. To compare different pretreatment methods, one first calculates the ratio of between- to within-class variance for each variable separately, then ranks the ratio from large to small. After ranking, the ratio is plotted as a function of the index of the variables for the original data and transformed data with different methods: contrast, first derivative, second derivative, offset correction and SNV. Fig. 2 shows that the contrast method gives higher Fisher ratios than all other methods.

Table 2
The classification results (CCR) based on leave-one-out cross-validation for the PLS classifier with the optimal number of selected variables after different pretreatments; data set 1

Method	Variables	Factors	CCR
Contrast	9	1	0.874
Original	24	2	0.579
Offset	17	17	0.600
SNV	21	4	0.790
First DER	24	4	0.632
Second DER	24	1	0.615

We classify the objects first using RDA with selected features according to Fisher's criterion, since RDA cannot be directly applied to NIR data because of the singularity of the variance–covariance matrix [9]. Fig. 3 shows the results for RDA with the data set pretreated by the contrast method. As the number of selected variables changes from 1 to 25, CCR first

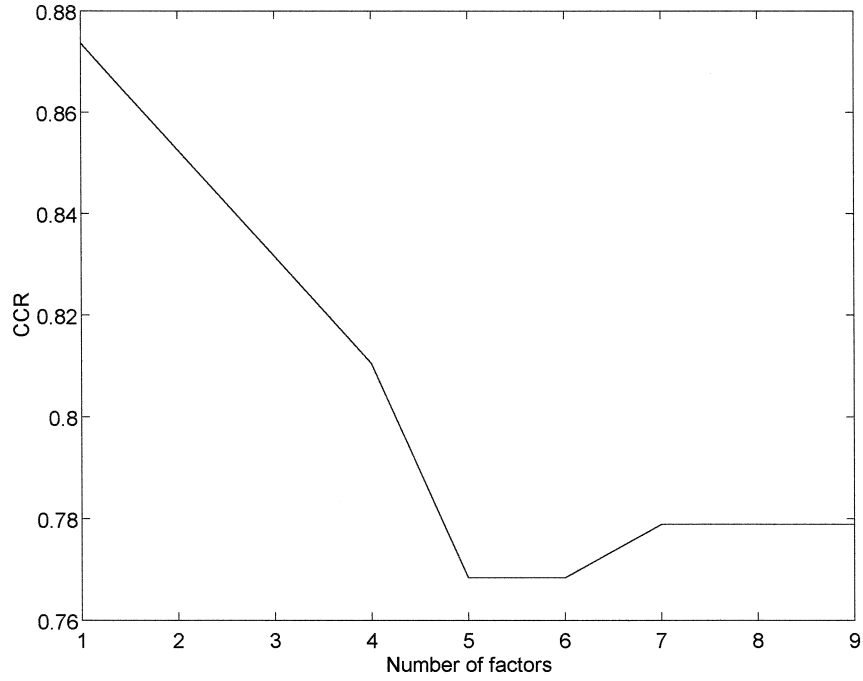

Fig. 5. The classification result (CCR) for PLS based on cross-validation versus number of factors with the optimal number of selected features.

Table 3
The classification results (CCR) for PLS and UVE-PLS when an independent test set is used; LOO—the results of leave-one-out cross-validation for the training set

Data set	Method	No. of objects (training set)	No. of objects (test set)	No. of variables	No. of factors	CCR (LOO)	CCR (test set)
1	PLS	74	21	5	1	0.851	0.810
1	UVE-PLS	74	21	166	1	0.851	0.857
2	PLS	30	12	1	1	1.000	1.000
3	PLS	45	15	all	7	1.000	1.000
3	UVE-PLS	45	15	4	2	1.000	1.000
4	PLS	39	19	4	4	0.923	0.900
4	UVE-PLS	39	19	6	4	0.897	0.900

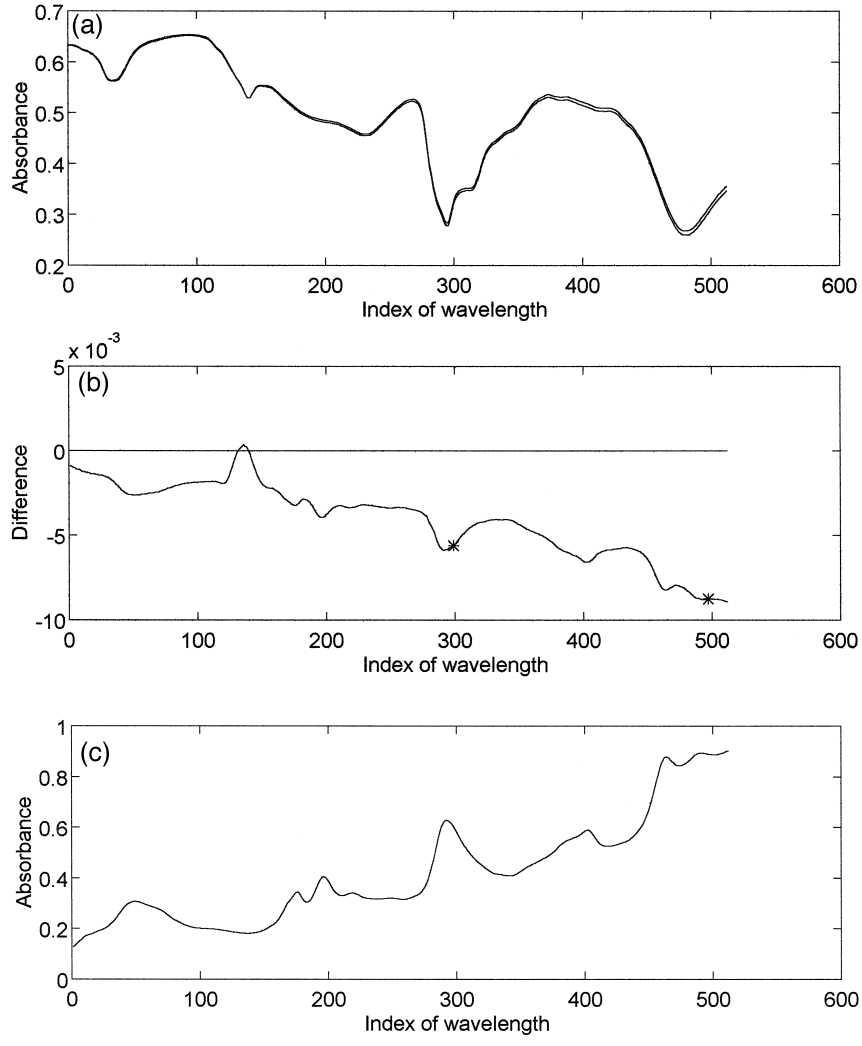

Fig. 6. (a) The mean spectra for the two classes of data set 2; (b) the difference between the two mean spectra for classes 1 and 2 versus the index of the wavelengths. ∗Denotes the selected variables by PLS; (c) the mean spectra of polymer.

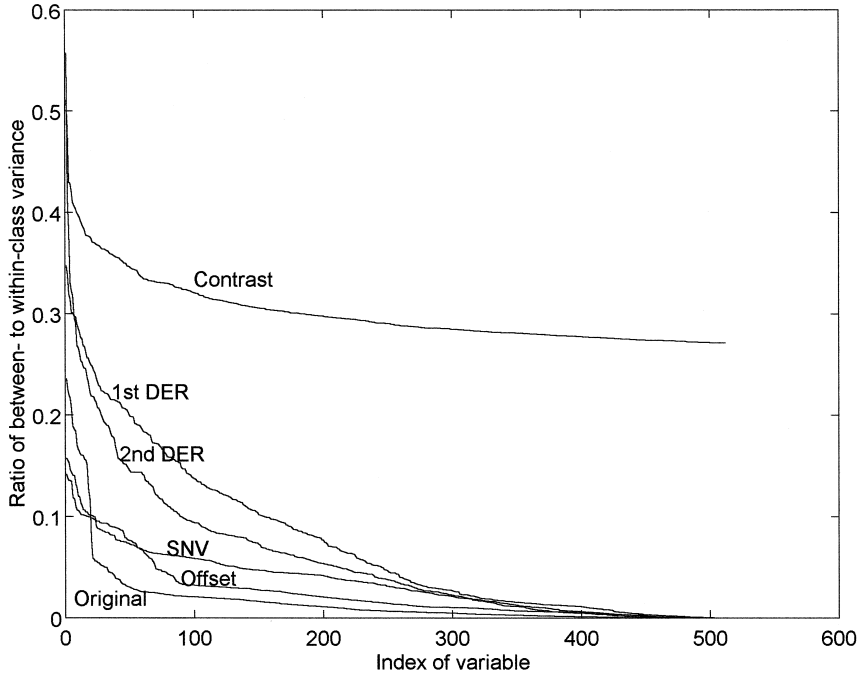

Fig. 7. The ratio of between- to within-class variance versus the index of the variables after ranking with different pretreatment methods including the original data; data set 2.

increases and then becomes constant at the maximum value of 0.874 when more than eight selected variables are used. The optimal results for RDA with the studied pretreatment methods (including no pretreatment) are listed in Table 1. The results show that the contrast method gives the best result of all pretreatment methods.

PLS2 is then applied as classifier. The number of features and number of factors are systematically changed from 1 to 25 after pretreatment by the contrast method. When the number of factors is larger than the number of variables, the CCR is set to 0. For a fixed number of features, the optimal CCR is calculated.

Fig. 4 demonstrates the optimal CCR as a function of the number of features. We can see that when nine features or more are selected, PLS2 gives the best result (CCR = 0.874). With the nine features, PLS2 gives the best result with one PLS component (Fig. 5). The nine features are the differences of the nine paired original wavelengths: (431–433), (431–432), (431–438), (431–436), (431–437), (431–439),

Table 4
The classification results (CCR) based on leave-one-out cross-validation for RDA with the optimal number of selected variables after different pretreatments; data set 2

Method	Variables	(λ, γ)	CCR
Contrast	1	(1, 0)	1.000
Original	6	(1, 0)	0.905
Offset	3	(0.8, 0)	0.857
SNV	4	(0.4, 0)	0.810
First DER	5	(0, 0.8)	0.881
Second DER	4	(0, 0.2)	0.714

Table 5
The classification results (CCR) based on leave-one-out cross-validation for the PLS classifier with the optimal number of selected variables after different pretreatments; data set 2

Method	Variables	Factors	CCR
Contrast	1	1	1.000
Original	2	1	1.000
Offset	3	1	0.976
SNV	2	1	1.000
First DER	2	1	1.000
Second DER	2	1	1.000

(431–441), (431–435) and (380–441). These wavelengths are situated in the region of the characteristic peak of the water impurity (Fig. 1a).

The best results of PLS with all studied pretreatment methods are listed in Table 2. The results show that the contrast method gives a better result than all other studied methods, and that the best CCR is the same as that for RDA.

Being similar in this case to a derivative pretreatment method, the noise in the signal should be enlarged in the contrast method. Also, in the first step of Spiegelman's method, $m(m+1)/2$ new variables are formed from combinations of the difference of the original m variables. For instance, if $m = 500$, 125 250 new variables are created. The number of new, somewhat noisy, variables is so large that the probability of chance correlation is increased; this means that some noise variables which have high Fisher ratios by chance, therefore, may be selected in the second step of the Spiegelman's method. Al-

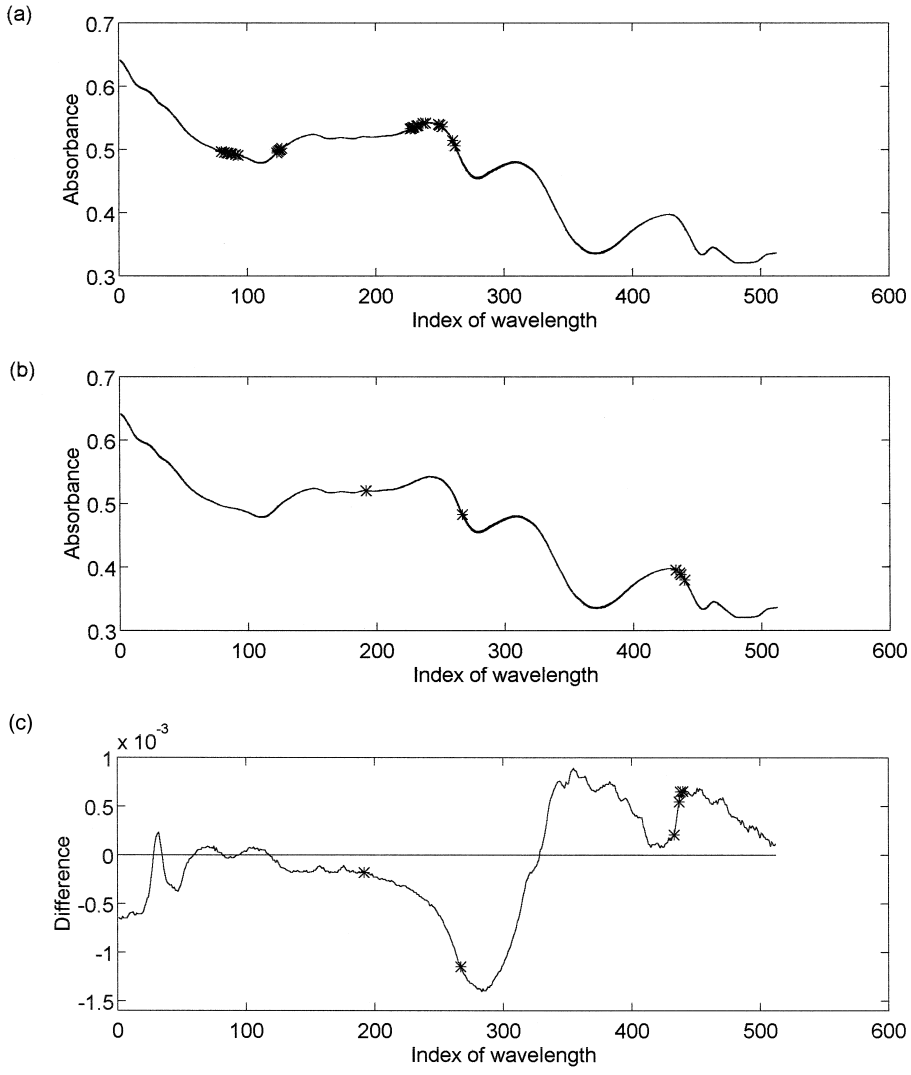

Fig. 8. The mean spectra for the three classes of data set 3. ∗Denotes the selected variables by (a) PLS with univariate feature selection and (b) UVE-PLS; (c) the difference between the two mean spectra for classes 1 and 2 versus the index of the wavelengths; ∗Denotes the selected variables by UVE-PLS.

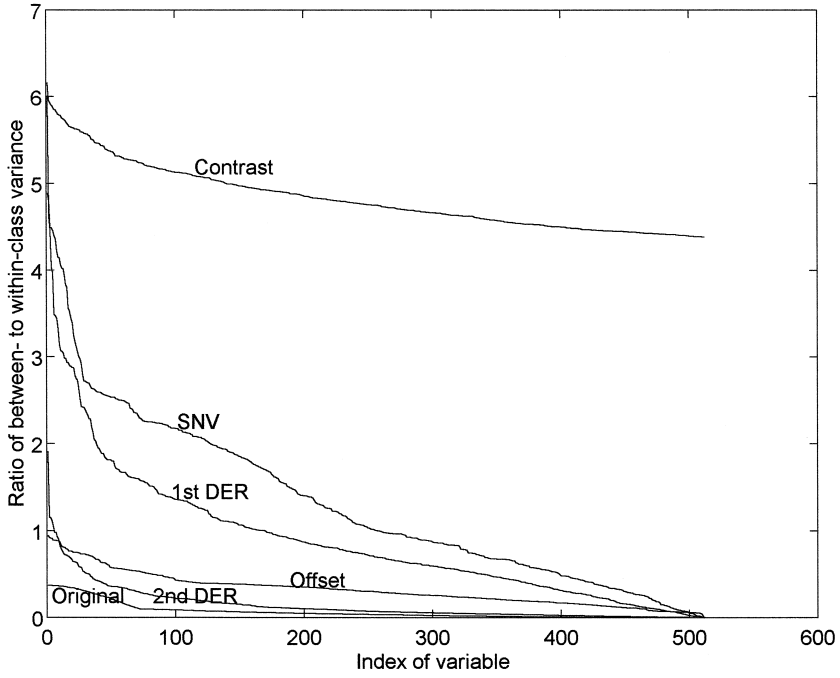

Fig. 9. The ratio of between- to within-class variance versus the index of the variables after ranking with different pretreatment methods including the original data; data set 3.

though leave-one-out cross-validation is used in the classification step, chance correlations still may lead to over-optimistic results. To test this, an independent test set was used to validate the results [14]. The data sets were divided into training sets and test sets, using a Kennard–Stone selection procedure [11]. The principle of Kennard–Stone method is to select a subset of objects that are uniformly distributed over the whole object space. The data are first pretreated by the contrast method. With the selected variables, PLS is applied to the training set to obtain the model by cross-validation, and the final model is evaluated using the independent test set. The results are shown in Table 3.

For PLS, the CCR from cross-validation of the training set is 0.851, while the CCR of the independent test set is somewhat lower (0.810). These results indicate that chance correlations may occur. For this reason, we applied UVE-PLS to delete the un-informative variables after the contrast pretreatment. With the remaining informative variables, PLS is used as classifier. The results in Table 3 show that UVE-PLS improves the CCR for the independent test set compared to PLS, and the difference between the training set and the independent test set is reduced, because more wavelengths in the region of the characteristic peak are selected (Fig. 1b) than for PLS (Fig. 1a). This indicates that UVE-PLS decreases the chance correlation.

4.2. Data set 2

Fig. 6a shows the mean spectra for the two classes for this data set. There are some differences between

Table 6
The classification results (CCR) based on leave-one-out cross-validation for RDA with the optimal number of selected variables after different pretreatments; data set 3

Method	Variables	(λ, γ)	CCR
Contrast	2	(0.2, 0)	0.967
Original	23	(0.8, 0)	0.633
Offset	24	(0.6, 0.2)	0.667
SNV	20	(1, 0.8)	0.733
First DER	23	(0.4, 0.2)	0.800
Second DER	24	(0.8, 0.4)	0.633

Table 7
The classification results (CCR) based on leave-one-out cross-validation for the PLS classifier with all variables and the optimal number of selected variables after different pretreatments; data set 3

Method	Variables	Factors	CCR	Variables	Factors	CCR
Contrast	all	4	1.000	6	3	0.967
Original	all	8	1.000	25	5	0.600
Offset	all	7	1.000	12	7	0.650
SNV	all	7	1.000	20	5	0.767
First DER	all	6	1.000	7	3	0.933
Second DER	all	5	1.000	20	4	0.900

Table 8
The classification results (CCR) based on leave-one-out cross-validation for RDA with the optimal number of selected variables after the contrast and no pretreatments; data set 4

Method	Variables	(λ, γ)	CCR
Contrast	3	(0, 0)	0.939
Original	7	(0.2, 0)	0.939

the absorbances in the region of wavelengths 300 to 500, due to the fact that class 2 contains 1% of the impurity.

Fig. 7 shows that the contrast method gives higher Fisher ratios than all other methods. Table 4 lists the comparison of the results with different pretreatment methods when RDA is used. It shows that the contrast method gives a higher CCR than all other pretreatment methods. When PLS is used as classifier (Table 5), the contrast method gives higher CCR than the offset correction, and the same CCR as the other methods. However, the contrast method performs better than the other methods in the sense that it uses the lowest number of variables.

To test chance correlation, the data set was divided into training sets and test sets by the Kennard–Stone algorithm. The data are pretreated by the contrast method. The comparison results with PLS are also listed in Table 3. For PLS, the CCR from cross-validation of the training set is 1.000 with one selected variable and one factor, and the CCR of the independent test set is also the same (1.000). Those results indicate that there is no chance correlation. PLS (Fig. 6b) selects two wavelengths where there

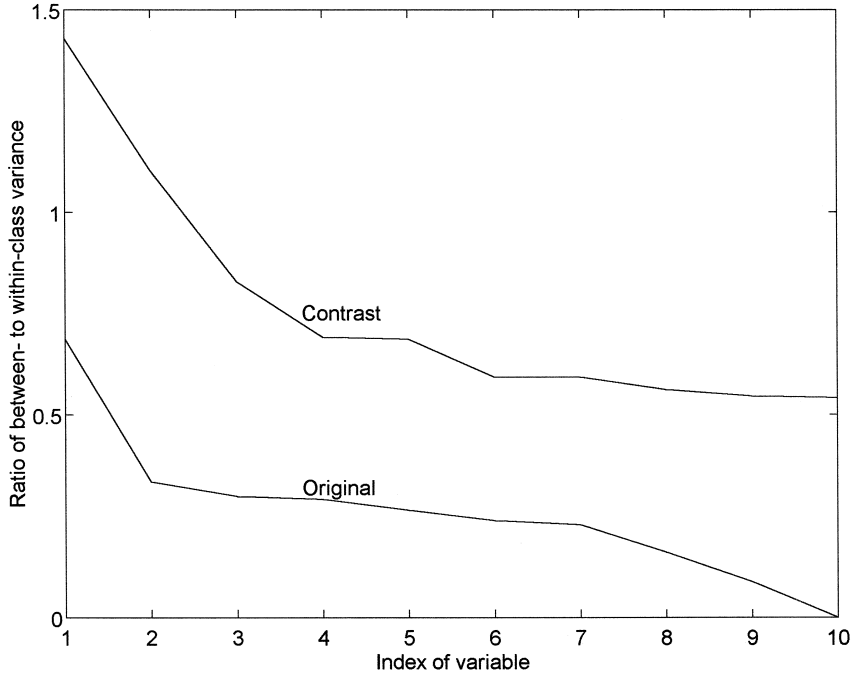

Fig. 10. The ratio of between- to within-class variance versus the index of the variables after ranking with the contrast-pretreated data and original data; data set 4.

Table 9
The classification results (CCR) based on leave-one-out cross-validation for the PLS classifier with all variables and the optimal number of selected variables after the contrast and no pretreatments; data set 4

Method	Variables	Factors	CCR	Variables	Factors	CCR
Contract	all	4	0.918	7	3	0.959
Original	all	2	0.898	3	3	0.918

are relatively large absorbance differences between the two classes, and they correspond to the two characteristic peak regions of polymer near wavelengths 300 and 500 (Fig. 6c).

4.3. Data set 3

This data set is used as an example with more than two classes. Fig. 8a demonstrates that the differences among the mean spectra for the three classes are difficult to be observed. Fig. 9 shows that the contrast method gives the highest Fisher ratios of all the methods. Table 6 shows that the contrast method leads to a higher CCR for RDA than all other methods. Table 7 lists the comparison of the results with different pretreatment methods for PLS. It shows that the contrast method leads to a higher CCR for PLS with feature selection than all other methods. It also demonstrates that, when all variables are used, all methods (including no pretreatment) give the same CCR (1.0000). However, the contrast method requires the lowest number of factors. When 1 to 25 features are selected as the input of PLS, the CCR is less good, which indicates that 25 features do not contain enough useful information. Most of the wavelengths corresponding to these selected variables are not situated in the characteristic peak region of wavelength 430 (Fig. 8a). The comparison of the results, when the contrast method and independent test set are used, for UVE-PLS are also listed in Table 3. It shows that UVE-PLS gives the same good results (CCR = 1.0000) for both training and test sets. However, UVE-PLS only selects four contrast variables, while PLS uses all variables. Most wavelengths corresponding to the four selected contrast variables are in the characteristic peak region of wavelength 430 (Fig. 8b). Since the differences between classes are difficult to see in Fig. 8a,b, Fig. 8c displays the difference between the two mean spectra for classes 1 and 2 versus the index of the wavelengths. It demonstrates that most of the selected wavelengths with UVE-PLS are situated in the re-

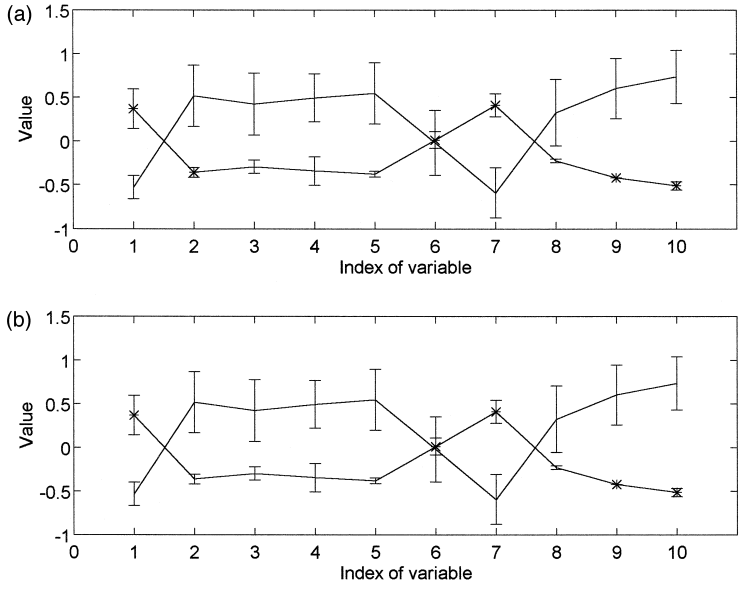

Fig. 11. The mean value ± standard errors of the two classes for data set 4 versus the index of the variables; ∗ Denotes the selected variables by (a) PLS with univariate feature selection and (b) UVE-PLS.

gion where there are relatively large absorbance differences between classes. Moreover, UVE-PLS uses two components to build the model, while PLS needs seven components. The results suggests that UVE-PLS eliminates the un-informative variables and simplifies the model complexity.

4.4. Data set 4

This data set is used as an example when contrasts are applied to other kinds of data than NIR. For this data set, only the contrast method is applied. No other pretreatment methods are compared, since they are specific for NIR. Fig. 10 shows that the Fisher ratios are higher for the contrast-transformed data than the original data. The results in Table 8 show that the contrast method requires a lower number of variables for RDA. When PLS is used as the classifier, the contrast method always gives better results (CCRs) than the original data (Table 9). In Table 3, the results of the independent test set show that the CCR of the test set is lower than that of the training set. However, after UVE-PLS is performed, the CCRs of both data sets are similar. The chance correlation is reduced by eliminating uninformative variables.

Fig. 11a,b show the mean value $\pm$ standard errors of the two kinds of sludges versus the index of the variables after the data are normalised to eliminate the scaling differences between the variables.

The selected variables with PLS and UVE-PLS are displayed in Fig. 11a,b, respectively. The second variable (Cd) is selected by PLS, but eliminated by UVE-PLS. The relative amounts of Cd (variable 2) and Ni (variable 10) are correlated in nature due to geochemical mobility. For UVE-PLS, the selected variables correspond to the following six paired original variables: (7–6), (10–6), (10–1), (10–0), (10–9) and (7–9), where (10–0) denotes variable 10 of the original data. This variable (Ni) is frequently used. The concentration of Ni plays an important role in the discrimination, because the difference between the two kinds of sludges is large.

5. Conclusion

Spiegelman's contrast method is modified to pretreat multivariate data for classification. Three NIR data sets and one pollution data set are used as examples to study the power of the method. Our results show that the contrast method greatly improves the ratios of between- to within-class variance. It is more powerful than offset correction, SNV, first and second derivative methods or using the original data. This conclusion does not depend on the classifiers (RDA and PLS2) with univariate feature selection based on Fisher's ratio. Our results indicate that the selection of data pretreatment methods is more crucial than selection of classifiers. The results of the independent test set show that there is a risk of chance correlation after the contrast pretreatment. The chance correlation decreases after eliminating the un-informative variables by the modified UVE-PLS method. However, one needs more data to critically compare different pretreatment methods.

Acknowledgements

We thank Prof. S.C. Rutan for her helpful comments, and Prof. C. Spiegelman for kindly providing his manuscript of the presentation at Compana '95, Germany. Thanks are also given to the Fonds van Wetenschappelijk Onderzoek, the DWTC and the Standards, Measurement and Testing program of the EU for financial assistance. A. Baldovin is acknowledged for providing the pollution data set.

References

[1] W. Wu, B. Walczak, D.L. Massart, K.A. Prebble, I.R. Last, Spectral transformation and wavelength selection in NIR spectra classification, Analytica Chimica Acta 315 (1995) 243–255.

[2] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra, Applied Spectroscopy 43 (1989) 772–777.

[3] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Correction to the description of standard normal variate (SNV) and de-trend (DT) transformations in practical spectroscopy with applications in food and beverage analysis (2nd edition), Journal of Near Infrared Spectroscopy 1 (1993) 185–186.

[4] L.S. Aucott, P.H. Garthwaite, S.T. Buckland, Transformations to reduce the effect of particle size in near-infrared spectra, Analyst 113 (1988) 1849–1854.

[5] M.S. Dhanoa, S.J. Lister, R. Sanderson, R.J. Barnes, The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra, Journal of Near Infrared Spectroscopy 2 (1994) 43–47.

[6] W. Wu, Q. Guo, D.L. Massart, The robust normal variate transform for pattern recognition with near-infrared data, Analytica Chimica Acta, submitted.
[7] C.H. Spiegelman, Calibration: a look at the mix of theory, methods and experimental data, presented at Compana '95, Wuerzburg, Germany.
[8] J.H. Friedman, Regularized discriminant analysis, Journal of the American Statistical Association 84 (1989) 165–175.
[9] W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D.L. Massart, S. Heuerding, F. Erni, Comparison of regularized discriminant analysis, linear discriminant analysis and quadratic discriminant analysis, applied to NIR data, Analytica Chimica Acta 329 (1996) 257–265.
[10] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Analytica Chimica Acta 185 (1986) 1–17.
[11] W. Wu, B. Walczak, D.L. Massart, S. Heuerding, F. Erni, I.R. Last, K.A. Prebble, Artificial neural networks in classification of NIR spectral data: design of the training set, Chemometrics and Intelligent Laboratory Systems 33 (1996) 35–46.
[12] J. Zupan, J. Gasteiger, Neural Networks for Chemists: An Introduction, Weinheim, New York, 1993.
[13] V. Centner, D.L. Massart, O.E. de Noord, S. de Jong, B.M. Vandeginste, C. Sterna, Elimination of uninformative variables for multivariate calibration, Analytial Chemistry 68 (1996) 3851–3858.
[14] W. Wu, S.C. Rutan, A. Baldovin, D.L. Massart, Feature selection using the Kalman filter for classification of multivariate data, Analytica Chimica Acta 335 (1996) 11–22.
[15] A. Baldovin, W. Wu, V. Centner, D. Jouan-Rimbaud, D.L. Massart, L. Favretto, A. Turello, Feature selection for the discrimination between pollution types with partial least squares modelling, Analyst 121 (1996) 1603–1608.