ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
21 (1999) 115–132

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra

A. Candolfi [a], R. De Maesschalck [a], D. Jouan-Rimbaud [a], P.A. Hailey [b], D.L. Massart [a,*]

[a] *ChemoAC, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium*
[b] *Analytical Research and Development, Pfizer Central Research, Ramsgate Road, Sandwich, Kent CT13 9NJ, UK*

Received 26 August 1998; received in revised form 26 February 1999; accepted 28 February 1999

Abstract

The effect of data pre-processing (no pre-processing, offset correction, de-trending, standard normal variate transformation (SNV), SNV + de-trending, multiplicative scatter correction, first and second derivative transformation after smoothing) on the identification of ten pharmaceutical excipients is investigated. Four pattern recognition methods are tested in the study, namely the Mahalanobis distance method, the SIMCA residual variance method, the wavelength distance method and a method based on triangular potential functions. The performance of the 32 method combinations is evaluated on the basis of two NIR data sets. The first one, measured in 1994, is used to build the classification models, the second, measured from 1994–1997, is used to assess the quality of the models. The best approach for the given data sets is the wavelength distance method combined with de-trending, a simple baseline correction method. More general recommendations for pre-processing excipient NIR data and for choosing an appropriate classification method are given. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Pharmaceutical excipients; NIR spectroscopy; Pre-processing; Pattern recognition; Positive identification

1. Introduction

In the pharmaceutical industry, excipients are required to be identified prior to release for use in the manufacture of dosage forms. This is one of the most common applications of near-infrared (NIR) spectroscopy combined with pattern recognition methods [1–5]. The incoming material is either scanned in its original container by means of an optical fibre connected to the NIR spectrophotometer, which can be placed directly in the warehouse, or by withdrawing samples from the material container and performing the measurement in the analytical laboratory. After acquiring the NIR spectrum of the excipient, its identity is determined by means of a pattern recognition method. If an accurate classification model is already established, this is a simple task. To develop such a model, however, requires time, effort and experience. Initially one must define the clas-

* Corresponding author. Tel.: + 32-2-4774737; fax: + 32-2-4774735.
E-mail address: fabi@vub.vub.ac.be (D.L. Massart)

PII: S0731-7085(99)00125-9

sification goal of the application, i.e. the type of result, which has to be provided by the pattern recognition method. Then, an appropriate classification method can be selected. A wide range of classification methods is available, which requires a careful selection and evaluation of the methods. When working with NIR data, an important decision is whether a pre-processing method is necessary. The selection of the suitable pre-processing method is therefore another important step in the method development. NIR spectra are subject to large baseline shifts due to the reflectance mode in which they are usually recorded [6–9]. This problem is especially important in the case of solid powdered samples, which contain materials of varying particle size distributions. By appropriate application of suitable pre-processing methods it is possible to minimise the contribution of physical effects to the NIR spectra.

The aim of this work is to propose a strategy for developing a chemometrics method for identifying excipient samples based on their NIR spectra. As a case study, the strategy is applied to a data set, containing ten solid powdered excipients, which are commonly used in the pharmaceutical industry. The classification models will be developed on the basis of a data set, which was measured in 1994. Spectra of new incoming excipient batches obtained over the following 3 years will then be predicted with the models, to evaluate whether the identification system is successful. For the method development we focus on two subjects: (i) the selection of an appropriate pre-processing method to correct for physical effects and (ii) the choice of a suitable classification technique which leads to accurate and acceptable identification of the excipients. The goal of the classification is to obtain the smallest possible α-error while eliminating any β-error. An α-error is the incorrect rejection of an object from its own class; a β-error is the false acceptance of a foreign object into a class. The selected method combination must discriminate sufficiently well to eliminate β-errors; it should not discriminate too much (i.e. be robust) so that samples of new batches that have slightly different spectra are still considered as belonging to its class.

2. Theory

A strategy for developing a classification method, consisting of data investigation, pre-processing, applying diagnostics to reveal inhomogeneities in the data, data set division, feature selection and modelling was already described in a previous work [10]. In the present study parts of this strategy is refined. The flow-chart in Fig. 1 explains the procedure for the steps pre-processing and classification optimisation. The scheme concerns the initial method development.

After having stated the classification goal and obtained the NIR spectra, the raw $\log(1/R)$ data are carefully investigated. One needs to examine, whether the variance within a class is large and, if so, why (using additional data and measurement related information). For separating several classes, it is necessary that their between-class variance is larger than their within-class variance. There exist a number of data pre-processing methods, which are able to reduce the within-class variance. Suitable pre-processing methods are selected, which correct for specific problems, in order to avoid blind data transformation. Classifiers should be selected based on the classification problem, i.e. the type of answer, which is expected from the pattern recognition method and the data structure. For a simple classification problem (e.g. ten very different classes) a univariate classifier might perform satisfactorily, while for a difficult classification problem, more powerful multivariate methods might be required. In the modelling phase, all proposed pre-processing and classification method combinations could be tested in order to obtain acceptable classification results.

The following theory sections describe the selection of suitable pre-processing and classification methods for the hereby studied application and data sets.

2.1. Pre-processing

2.1.1. Preliminary selection of the pre-processing methods

Three issues occur in diffuse reflectance NIR spectroscopy for solid samples: the multicollinearity among variables, light scattering and particle size [6–9]. The multicollinearity of the variables is typical for spectroscopic data since the data consist of continuous signals. Some transformation methods have the ability to reduce the correlation between variables. Scattering occurs on the surface of a material and depends therefore on the physical nature of the material and the particle size. Interaction of the incident light and the medium occurs within the material, e.g. within the particles of the powder. Therefore, the particle size defines the spectral pathlength and varying particle sizes result in a baseline shift in the spectra. Additional factors influence NIR spectra, for instance the particle size distribution, the density of a powder and consequently the packing of the

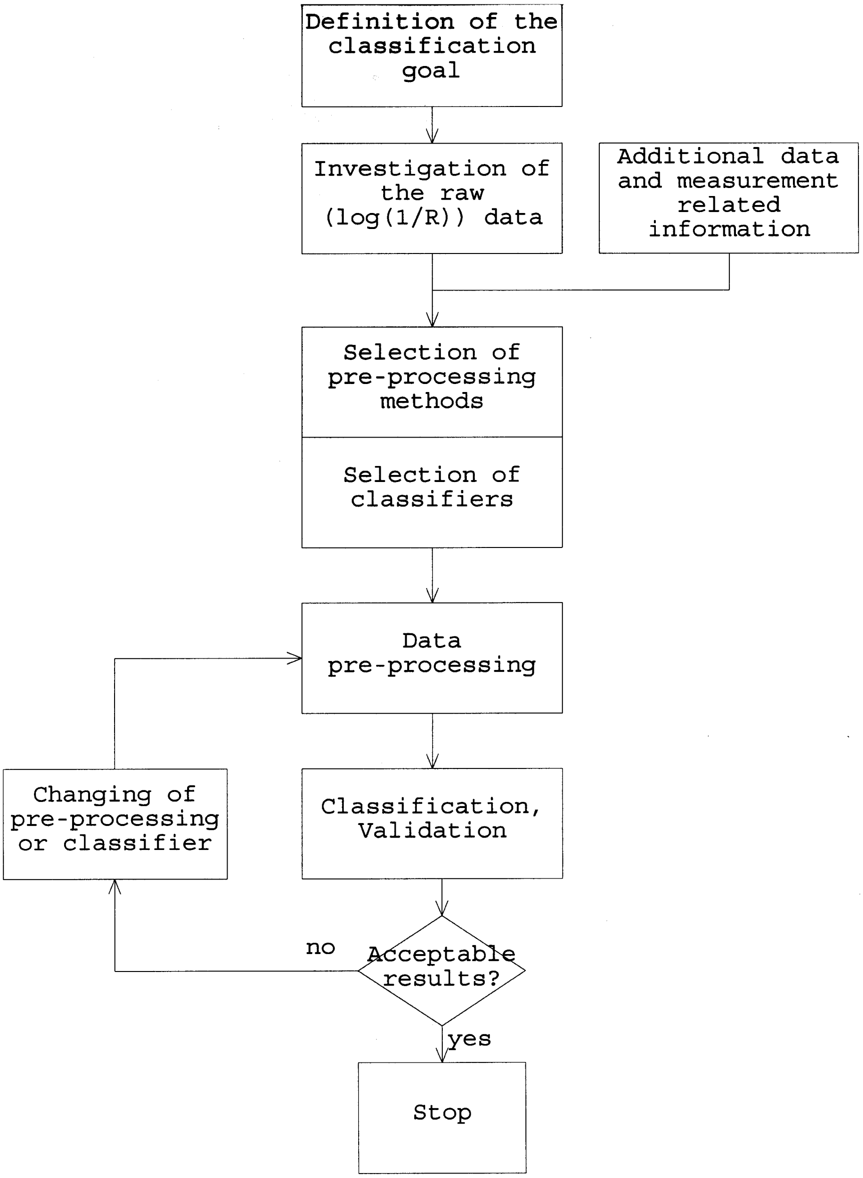
Definition of the classification goal
Investigation of the raw (log(1/R)) data
Additional data and measurement related information
Selection of pre-processing methods
Selection of classifiers
Data pre-processing
Changing of pre-processing or classifier
Classification, Validation
no
Acceptable results?
yes
Stop

Fig. 1. Flow-chart for pre-processing and classification method optimisation.

material inside a measuring cup, the moisture content of a material, the instrument itself and temperature. Multiplicative interference of these effects are responsible for baseline shifts, slope changes and curvilinearity in the spectra. As a result NIR spectra contain not only chemical, but also physical information about the sample and the measuring conditions, which may be irrelevant to the problem under study. If such information is indeed undesired for the data analysis, it is possible to pre-process the spectra, in order to extract the more relevant chemical information. Certain transformation techniques are able to remove baseline shifts, slope changes and curvilinearity of spectra, i.e. they reduce the influence of particle size, scattering and other influencing factors.

The present study describes the identification of ten powdered excipients. The spectra show the above discussed powder related problems, such as baseline shift and slope changes. A detailed discussion of the basic data set, revealing different sources of variance, is described in [11]. As a result the data of the individual classes contain large within-class variances. In this application the identification of the excipients is restricted to chemical differences between substances. Therefore, the spectra can be pre-processed without loosing this information. A number of pre-processing (signal processing) methods, which correct for the observed problems, are selected to transform the spectra. The methods and their rational for the selection are described below. The list does not contain explicit matrix transformation techniques (such as auto-scaling, logarithmic transformation etc.). These transformations do not correct for typically NIR related problems. Nor are particular noise reduction methods considered, as for instance Fourier transform (FT) or wavelets. Such methods should be regarded as feature reduction methods, comparable to principal component analysis (PCA).

The following pre-processing methods are selected:

1. Offset correction
2. De-trending
3. SNV transformation
4. SNV transformation + de-trending
5. Multiplicative scatter correction (MSC)
6. First derivative after smoothing
7. Second derivative after smoothing

An initial classification is performed with the original data, in order to determine, whether any spectral pre-processing is necessary at all. Aside from SNV + de-trending, no other combinations of pre-processing methods were chosen. This combination is considered to be a standard approach for solid materials and described throughout the literature [12]. In general it is not advisable to combine signal processing methods, unless there is a specific reason.

2.1.2. Pre-processing methods

2.1.2.1. Offset correction. Offset correction is applied to correct for a parallel baseline shift. An arbitrary chosen value is subtracted from each spectrum independently. In this application the mean absorbance of the first five variables of each spectrum is used for the correction, in order to obtain positive values for the whole spectrum and a zero baseline at the beginning of the spectrum.

$$x_{ij,0} = x_{ij} - \bar{x}_{i,1-5} \tag{1}$$

where $x_{ij,0}$ is the transformed element, x_{ij} the original element and $\bar{x}_{i,1-5}$ the mean absorbance of the first five variables of each spectrum.

2.1.2.2. De-trending. De-trending is another baseline correction method. It removes offset and curvilinearity, which occurs in the case of powdered, densely packed samples. The baseline is modelled as a function of wavelength, with a second-degree polynomial, and subtracted from the spectrum.

$$x_{ij,\mathrm{d}} = x_{ij} - \mathrm{bl}_{ij} \tag{2}$$

where $x_{ij,\mathrm{d}}$ is the transformed element, x_{ij} the original element and bl_{ij} the baseline value at wavelength j of spectrum i.

Normally, de-trending is carried out in combination with SNV transformation. If one is interested in the different shapes of the spectra, de-trending is applied alone [12].

2.1.2.3. SNV transformation. SNV removes the multiplicative interferences of scatter and particle size. To remove slope variations on individual spectrum basis, each object is transformed independently using the following equation:

$$x_{ij,\mathrm{SNV}} = (x_{ij} - \bar{x}_i)/\sqrt{\frac{\sum(x_{ij} - \bar{x}_i)^2}{p-1}} \tag{3}$$

where $x_{ij,\mathrm{SNV}}$ is the transformed element, x_{ij} the original element, $\bar{x}_i$ the mean of spectrum i and p the number of variables in the spectrum.

SNV can be combined with de-trending in order to remove the curvilinearity of spectra [12].

2.1.2.4. Multiplicative scatter correction. Another method compensating for different scatter and particle sizes is MSC. Here the correction is carried out based on the assumption, that all samples have the same scatter coefficient at all NIR wavelengths. An ideal spectrum, usually the average spectrum of a representative data set, is used to estimate the scatter of the spectra. All other spectra are corrected to have the same scatter level as the selected one. Each individual spectrum is shifted and rotated so that it fits as closely as possible to the chosen mean spectrum. The fit for the individual and the mean spectrum is achieved by least squares.

$$\mathbf{x}_i = a_i + b_i\bar{\mathbf{x}}_j + \mathbf{e}_i \tag{4}$$

where $\mathbf{x}_i$ is an individual spectrum i, $\bar{\mathbf{x}}_j$ the mean spectrum of the data set, and $\mathbf{e}_i$ the residual spectrum, which ideally represents the chemical information in the data. The fitted constants a_i (offset, intercept) and b_i (slope) are used to correct each value of the spectrum i.

$$\mathbf{x}_{i,\mathrm{MSC}} = (\mathbf{x}_i - a_i)/b_i \tag{5}$$

Since scatter and particle size are independent of chemical information, the user normally defines a sub-region of the spectrum, which represents explicitly the baseline and no chemical information. This sub-region is then used to obtain the parameters a_i and b_i, which are then applied to correct the entire spectrum [13,14]. In this application the first 100 variables are selected as sub-region, which correspond to the spectral range of 1100–1300 nm, where only little chemical information was found.

In pattern recognition, MSC is typically applied to each class separately. This includes the determination of the corresponding ideal spectrum, i.e. the mean spectrum, and the definition of the correction terms for each class.

2.1.2.5. Derivatives after smoothing. The goal of differentiation is to remove background and to increase spectral resolution. A constant background is removed by transforming the original spectra into first derivative spectra, a linear background by transforming them into the second derivative spectra. In general, the second derivative is more often used, because the data interpretation is considered to be easier. This transformation is largely historical too.

The drawback of differentiation is that it amplifies noise. Therefore, it is necessary to smooth the data beforehand. The most often used smoothing method is the one proposed by Savitzky and Golay [15], which is a moving window averaging method. Similar to the approach for the computation of derivatives, a window is selected, where the data are fitted by a polynomial of a certain degree. In this application the differentiation and smoothing is carried out according to [16]. A window width of 17 variables is selected.

2.2. Classification

2.2.1. Preliminary selection of the classification methods

In order to develop classification rules one uses supervised pattern recognition techniques. One can distinguish between discriminating and class-modelling methods [17]. In discrimination analysis one tries to find boundaries between given classes. New objects are assigned to one class depending on the classification rule. With class-modelling techniques an individual model is established for each given class separately, based on similarities between the objects within this class [18]. Class borders are defined around the samples of the class. The model therefore consists in fact of a hypervolume, described by the samples of the class. This approach enables a positive identification. A new object is assigned to the class only, if it falls within the hypervolume. If it falls outside, it is considered as an outlier. For this reason class-modelling techniques can be regarded as outlier detection methods.

The classification of excipients requires a method, which leads to a positive identification. A sample should be identified as member of a class, only if it is similar enough to the class under investigation. Class-modelling techniques are appropriate for dealing with such a situation. With these methods the identification as well as the quality of an excipient can be determined. For this study we selected the following four pattern recognition methods, which lead to a positive identification.

1. Mahalanobis distance method (Hotelling's T^2 test)
2. SIMCA residual variance method
3. Wavelength distance method
4. Method based on triangular potential functions

The Mahalanobis distance, SIMCA residual variance and the wavelength distance method are commercially available. In the Mahalanobis distance and SIMCA residual variance method it is however important to investigate, which variant is used, because they exist in several versions. The method based on potential functions applied here was originally developed to study the representativity of prediction samples towards a multivariate calibration model [19]. It is selected as a nonparametric alternative to the three parametric methods.

2.2.2. Classification methods

2.2.2.1. Mahalanobis distance and SIMCA residual variance method. As used here, the Mahalanobis distance method and SIMCA residual variance method are class-modelling techniques based on PCA [20–23]. The two methods are complementary, the Mahalanobis distance method covers the space defined by the significant principal components (PCs), the SIMCA residual variance method the residual space. For each class separately a PCA is carried out yielding a class PC model. The model for each class, with a certain number of significant PCs, is obtained by the following equation:

$$(\mathbf{X}-\bar{\mathbf{X}}) = \mathbf{UWV}^{\mathrm{T}} + \mathbf{E} \tag{6}$$

where $(\mathbf{X}-\bar{\mathbf{X}})$ is the mean centred data matrix, **U** the normed score matrix obtained for n objects and r selected principal components, **W** a diagonal matrix with the singular values, $\mathbf{V}^{\mathrm{T}}$ the loading matrix obtained for r selected principal components and p variables and **E** the residual matrix.

The Mahalanobis distance method, as used here, is equivalent to the Hotelling's T^2 test, which is known from statistical process control [24]. In the literature there is some confusion about this classification method, since several variants are known, working in the original and in the PC space [3,25–27]. The method always involves the computation of the Mahalanobis distance, T_i^2, of an object $\mathbf{x}_i$ to the mean spectrum (centroid) of the class, $\bar{\mathbf{x}}_j$, which is then compared to a critical value. This critical value can be obtained either from an F, χ^2 or a β distribution, and can be found in corresponding tables or it is defined by the user.

In the original space, the Mahalanobis distance is defined as followed:

$$T_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}}_j)\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}_j)' \tag{7}$$

with

$$\mathbf{S} = \frac{1}{(n-1)}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}}_j)'(\mathbf{x}_i - \bar{\mathbf{x}}_j) \tag{8}$$

where $\bar{\mathbf{x}}_j$ is an estimate of the mean vector and **S** an estimate of the variance–covariance matrix of the class.

In the PC space the computation of the Mahalanobis distance is simpler, because the PCs are orthogonal. Therefore the covariance of the variance–covariance matrix of the PC scores vanishes. The **S** then becomes a diagonal matrix. Moreover the centroid of the class is zero for mean centred data. Eq. (7) can therefore be simplified to:

$$T_i^2 = (n-1)\mathbf{u}_i\mathbf{u}_i' \tag{9}$$

where $\mathbf{u}_i$ is the vector of the normed scores of object i.

In this application two critical values are needed for the computation of the Hotelling's T^2

test. For the training set, where the data are only fitted, the critical value must be defined differently compared to the test set, which is based on prediction. According to [24] the critical T^2 value for the training set is obtained in the following way:

$$T^2_{\text{crit}} = \frac{(n-1)^2}{n} B(\alpha, r/2, (n-r-1)/2) \qquad (10)$$

where B refers to the β-distribution. This distribution is proposed for a situation, where the training set samples are used to obtain the mean and the variance–covariance matrix needed for the computation of their Mahalanobis distances. B is the tabulated value for the confidence level α, and for $r/2$ and $(n-r-1)/2$ degrees of freedom.

When predicting a new object, the mean and the variance–covariance matrix of the training class are used to calculate its T^2 value. The F-distribution is appropriate to establish the critical value in this situation [24]. Thus, the critical value is obtained in the following way:

$$T^2_{\text{crit}} = \frac{r(n-1)(n+1)}{n(n-r)} F(\alpha, r, n-r) \qquad (11)$$

where F is the tabulated value for the confidence level α and r and $n-r$ degrees of freedom.

In the SIMCA residual variance method, class boundaries are constructed around the modelled PCs, based on the distances (Euclidean distances) between the objects and the origin in the space of the residual PCs.

$$s_0 = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{p} \frac{e_{ij}^2}{(p-r)(n-r-1)}} \qquad (12)$$

e_{ij}^2 is the squared residual of object i on the latent variable j and s_0 is the mean distance between all objects belonging to the class and the class model.

An unknown object is classified by projecting it into the PC space defined for the class. Then its distance towards the class model (s_i) is computed.

$$s_i = \sqrt{\sum_{j=1}^{p} \frac{e_{ij}^2}{p-r}} \qquad (13)$$

With the help of an F-test at a given level of confidence the obtained value s_i is compared to a critical value, s_{crit}.

$$s_{\text{crit}} = \sqrt{F_{\text{crit}} s_0^2} \qquad (14)$$

If $s_i < s_{\text{crit}}$ the object is considered to be a member of the class, otherwise it is regarded to be an outlier. s_0 obtained from the fitted scores of the training class and s_i obtained from the predicted scores of a new object should not be directly compared [11,23,28]. Therefore, it was proposed to predict all spectra of the training set first with leave-one-out cross-validation (LOOCV) in order to obtain the predicted scores [23]. These scores are then used to establish the value s_0.

In this work the Mahalanobis distance and the SIMCA residual variance method are applied at two levels of confidence, $\alpha = 0.05$ and 0.01.

2.2.2.2. Wavelength distance method. The wavelength distance method is a univariate classification method, applied in the spectral domain [29]. Thanks to its mathematical simplicity and the ease of interpretation of the results, it found widespread use by practitioners of NIR-spectroscopy [4,27,30]. As it is a class-modelling technique, a model is established for each class separately. To identify a new spectrum, its residual spectrum, z_i, is computed by using the mean spectrum $\bar{\mathbf{x}}_j$ and the standard deviation spectrum $\mathbf{s}_j$ of the training class under investigation.

$$\mathbf{z}_i = (\mathbf{x}_i - \bar{\mathbf{x}}_j)/\mathbf{s}_j \qquad (15)$$

$$\mathbf{t}_i = \mathbf{z}_i[n/(n+1)]^{1/2} \qquad (16)$$

For each variable (wavelength) a t-test is carried out at a given level of significance (α). A spectrum is considered to belong to the class under investigation if for all variables $t_i \leq t_{\max}$ (tabulated) ($H_0, z_i = 0$), otherwise the sample is considered to be an outlier ($H_1, z_i > 0$).

This test is correct only for one single variable. Here multiple t-tests are carried out, one for each variable, in order to examine whether a new sample belongs to the class. The probability of p (with p being the number of variables) successful outcomes, pr(p), from such multiple tests is much lower.

For multiple comparisons it is therefore necessary to increase $t_{\max}$ in order to achieve correct results. Gemperline et al. enhanced the original wavelength distance method by including parametric statistical tests and probability thresholds, which depend on the number of training samples (n) and the number of variables (p) per spectrum [4]. Probabilities for selected critical values for t ($t_{\max}$) taking into account different values of objects and variables, are reported. This modification of the method is in agreement with the $t_{\max}$ value of 6 which is proposed in the commercial software for NIR spectra, where $p \approx 700$. A problem of the method is that the correlation between variables is neglected. As a consequence the class borders are larger compared to multivariate approaches, which leads to lower α-errors, but on the other hand to a higher risk for β-errors. Nevertheless, it was found that the method performs well in practice [4,27].

In this study the wavelength distance method is applied using a $t_{\max} = 6$, as proposed in the commercial software.

2.2.2.3. Method based on triangular potential functions. Different methods based on potential functions are known in pattern recognition [31–33]. Recently, Jouan-Rimbaud [19] studied the applicability of triangular potential functions to detect prediction outliers and inliers in multivariate calibration. Since we can formulate our classification problem as an outlier problem, the method used in the present study is derived from this work.

In potential methods one simulates a potential field in space around each object of a given class. The value of such a potential field is maximum at the location of the object and decreases with distance from the object. The individual potentials from the whole class can then be averaged in each point of the considered space, leading to a global potential for the class, which can be considered to be a probability density function. The shape of the individual potential fields depends on the choice of the potential function and a smoothing factor. In our approach triangular potential functions are selected. It is a simple function, which avoids that the user needs to evaluate critical parameters, such as the selection of a correct cut-off value and confidence level, needed for instance with Gaussian functions [19]. In the case of triangular functions the cut-off value is simply zero. In the univariate space triangular potential functions are defined as follows:

$$\phi(x_i,x_k) = 0, \quad \text{for}\left|\frac{x_i - x_k}{\text{sm}}\right| > 1$$

$$\phi(x_i,x_k) = 1 - \left|\frac{x_i - x_k}{\text{sm}}\right|, \quad \text{for}\left|\frac{x_i - x_k}{\text{sm}}\right| \leq 1 \qquad (17)$$

where $\phi(x_i,x_k)$ is the potential induced by object x_i on object x_k of the class. The width of the function depends on the smoothing factor sm, which has to be optimised for each class separately.

The global potential field f, an estimate of the probability density, is obtained as:

$$f = \frac{1}{n}\sum_{i=1}^{n} \phi(x_i,x_k) \qquad (18)$$

with n being the number of objects in the class.

In the definition of triangular potential functions, applied to the multivariate space, the absolute value of $|x_i - x_k|$ is replaced by the Euclidean norm of the vector $\mathbf{x}_i - \mathbf{x}_k$. The global potential, creating so-called potential hypersurfaces, is then defined:

$$f = \frac{1}{n\,\text{sm}^p}\sum_{i=1}^{n} \phi(\mathbf{x}_i,\mathbf{x}_k) \qquad (19)$$

where p is the number of variables. In this application the smoothing (sm) is kept constant within one class. The optimisation of the parameter sm for each class is a critical step. If the smoothing is too small, local potential fields are obtained around isolated objects or small groups of objects. Within the class there can be still regions with a zero potential. Therefore no continuous potential function is obtained. On the other hand, if the smoothing is too large, the global potential field is becoming flat and too large on the border of the class. Therefore, the smoothing must be chosen carefully as a compromise between an acceptable α- and β-error. In order to define the smoothing parameter (sm) we work with the K nearest neighbour's distance. The median Euclidean distance of all objects and its K neighbour is used. The median distance is chosen, because it is not influenced by extreme objects. The smoothing is then optimised by optimising K according to [19] with the centroid method and by LOOCV. In the centroid method pairs of objects are selected and the potential of their centroid is determined. If this is non-zero for all or most possible pairs of objects, the smoothing is considered to be adequate. In the LOOCV procedure one object is left out and the global potential induced by the other objects of the class on the left out object is calculated. Again if the potential is positive for most objects the smoothing is adequate. The smoothing is optimal for the smallest K with positive results. The obtained model can then be used to predict new samples. If the potential of a new sample, projected in the space of the class under investigation is positive, the sample is classified to the class, if its potential is zero it is considered to be an outlier.

3. Experimental

Two data sets, called ‘basic’ and ‘historical’ data set were investigated. Both of them contain spectra of ten white, powdered excipients, commonly used in the pharmaceutical industry. They were collected from different excipient batches, delivered from various suppliers. In the case of the basic data set each class contains between 15 and 22 samples, the total number of spectra in the data set is 175. These data were collected in 1994, over a 9-month period. This data set was already investigated in a previous study [11]. The historical data set includes samples of new incoming excipient batches, obtained from 1994 to 1997. Here there are a total of 259 spectra available. The composition of the two data sets is given in Table 1.

The samples were also analysed by conventional pharmacopoeial tests, which were all passed. Therefore all the excipient batches were released for production.

The spectra were measured with a NIRSystem 6500 spectrophotometer (NIRSystem, Silver Spring, MD, USA), with the standard sample cup (NIRSystem, Silver Spring, MD, USA). Every spectrum, which is the average of 32 scans, was ratioed against a Spectralon standard (99% reflective, SRS-99-010, Labsphere, North Sutton, NH, USA). The spectra were obtained from 1100 to 2468 nm, in 2-nm steps, leading into 685 variables. The conventional system suitability tests were performed prior to any data acquisition.

Table 1
Composition of the basic and historical data set (types of excipients and number of samples in each lass)

Excipient	Basic data set (1994) No. of samples	Historical data set (1994–1997) No. of samples
Class 1: anhydrous dicalcium phosphate	17	13
Class 2: anhydrous lactose	16	15
Class 3: explotab	19	11
Class 4: lactose	22	43
Class 5: magnesium stearate	15	10
Class 6: methocel	18	24
Class 7: povidone	15	5
Class 8: sodium lauryl sulphate	17	0
Class 9: starch	19	43
Class 10: avicel	17	95

The programs for the data analysis are written in MATLAB code (V.4.0, Mathworks, Natick, USA). The spectral acquisition was carried out with NSAS (V.3.50, NIRSystems, Silver Spring, MD, USA).

4. Results and discussion

In order to find the best identification system, each pre-processing method is combined with each selected pattern recognition technique. The performance of the method combinations is evaluated on the respective α- and β-errors obtained in classification.

The basic data set is used as training set to construct the classification models. One individual model is built for each excipient class. This data set is considered to be representative for the materials. It contains excipient samples from different batches, obtained from several suppliers. As the data were collected over some months, also instrument dependent sources of variance, such as the instrument instability over time, are included. The α-error is determined for each class. As described in the theory, in the Mahalanobis distance method the α-error for the training set is obtained using the fitted data and the adapted critical values. In the SIMCA residual variance method the predicted scores are used to establish the confidence limits. For the wavelength distance and the potential-function method, LOOCV and the centroid method, respectively, are applied within each class to obtain the α-errors of the basic data set. Then all samples of the training sets of the other classes are predicted with the established models for determining the β-errors. We use the historical data set, to evaluate the performance of the classification models. Again the α- and β-errors are determined by predicting those data with the established pattern recognition models. The spectra of the historical data set are obtained from new excipient batches received within the following 3 years after the data of the basic data set. Since they are real data, they show realistic variations, which can be used to evaluate the potential for correct recognition of future samples.

In order to obtain an idea about the pattern recognition problem, the mean spectra of the basic data set are displayed together. Fig. 2 shows the mean spectra for the original data.

It is evident, that most of the mean spectra are very different in shape, except for two cases. The spectrum of magnesium stearate is similar to the one of sodium lauryl sulphate, since both excipients are fatty acids, and the spectrum of explotab resembles the one of starch, as explotab is a modified starch.

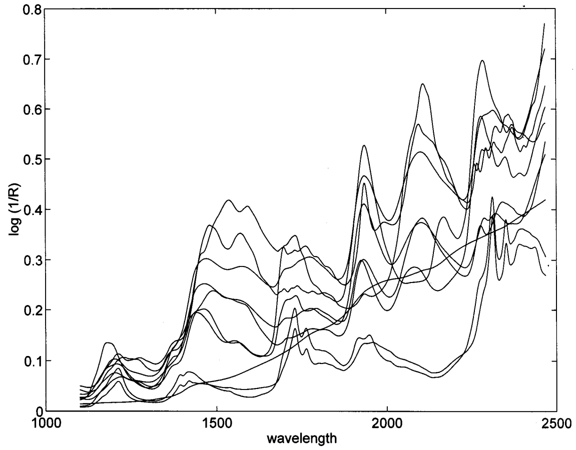

Fig. 2. Mean spectra for the ten excipient classes (original data).

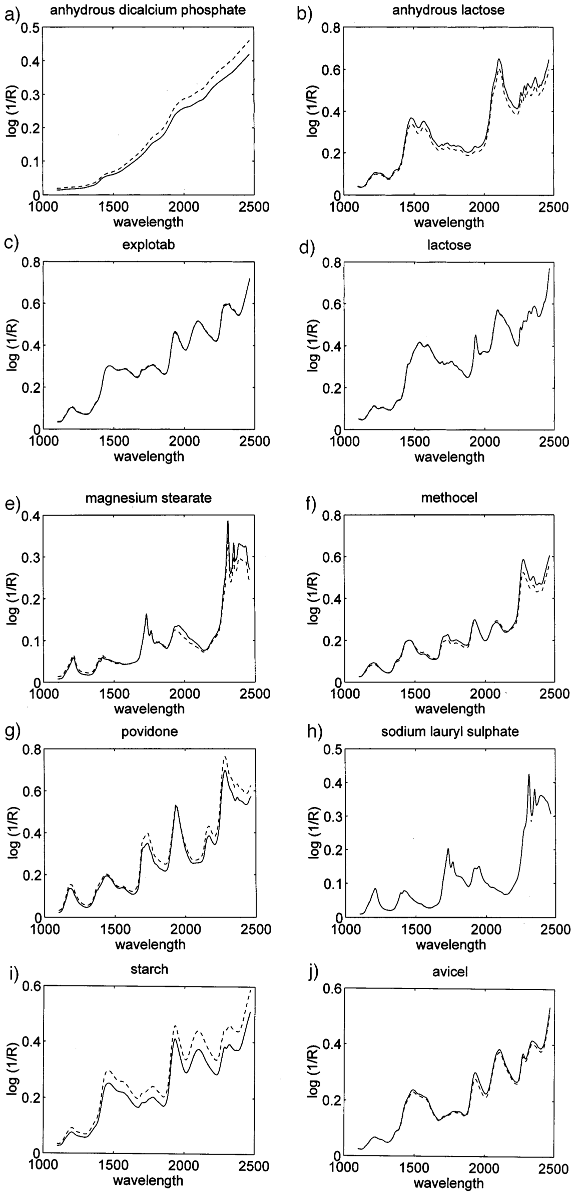

Fig. 3. Mean spectra of the basic (—) and the historical (--) data set for: (a) anhydrous dicalcium phosphate; (b) anhydrous lactose; (c) explotab; (d) lactose; (e) magnesium stearate; (f) methocel; (g) povidone; (h) sodium lauryl sulphate; (i) starch and (j) avicel.

The mean spectra obtained from the original spectra of the basic and the historical data set for each class are presented in Fig. 3a–j.

The mean spectrum of the basic data set is represented by a solid line, the one of the historical data set by a dashed line. For some classes, anhydrous dicalcium phosphate, anhydrous lactose, starch and avicel (Fig. 3a, b, i and j), there is a baseline shift between the two mean spectra, which increases with increasing wavelengths. For three classes, explotab, lactose and methocel (Fig. 3c, d and f) both spectra almost overlap. There occur several crossings of the two mean spectra, one spectrum is somewhat flatter than the other is. In two other cases, povidone and magnesium stearate (Fig. 3g and e), both mean spectra also cross, but isolated spectral bands have also slightly different shapes, due to increased or decreased intensities. In the case of the last class, sodium lauryl sulphate (Fig. 3h), there are no spectra in the historical data set. No general trend is found for the differences between the two data sets. Prior to data acquisition, instrumental system suitability tests were performed. This confirms that all spectra were correctly obtained, and hence require no preliminary standardisation to be performed [34,35].

The first classifiers discussed here are the complementary techniques, the Mahalanobis distance- and the SIMCA residual variance method. Both methods are based on PCA. For the construction of the classification models the number of significant PCs has to be determined for each class separately. To obtain these values, we used the method of the reduced eigenvalue, a test that was proposed by Malinowski [36]. Between one and nine factors were selected for the ten excipient classes, pre-treated with the different pre-processing methods. The results for the two pattern recognition methods obtained with the basic data set is presented in the Table 2a and b.

The class models were constructed at two levels of significance, 95 and 99%. The overall α- and β-errors, summarised for the ten excipient classes, are given.

As can be seen in Table 2a for the Mahalanobis distance method the overall α-error on the 95% level of confidence is between 9 and 12% for all types of pre-processing. By increasing the confidence interval to 99% the overall α-error can be reduced to around 5%. The rejection rate is somewhat too high probably due to the fact that the data of the training sets are heterogeneous, obtained from different excipient batches provided occasionally by various suppliers. The α-error decreases only slightly with data pre-processing. In general also more or less the same objects are classified as outliers. This appears to demonstrate that pre-processing does not have a large influence on the α-error in the case of the Mahalanobis distance method. The method is based on the distribution of the data in the space of the modelled PCs. The Mahalanobis distance takes the variance–covariance structure of the data after different types of pre-processing into account.

The situation is different for the β-error. A β-error occurs, when spectra are wrongly classified into a class, i.e. when classes overlap. This happens when classes are similar and their between-class variance is smaller than their within-class variance. For the original data two samples are wrongly classified at the 99% level of confidence. Larger β-errors occur in the case of the first and second derivative data. 13 and 17 samples, respectively, are wrongly classified at the 99% level of confidence and five and 11 samples, respectively at the 95% level of confidence. These misclassifications are sodium lauryl sulphate samples, which are classified as members of magnesium stearate, and explotab samples, which are classified as starch samples. The spectra of the corresponding two classes are indeed very similar, as already explained before. However, the spectra within one class are slightly different in shape. This may happen, if there are several suppliers to provide the same excipient. By the differentiation the small spectral differences in the data (within one class) are emphasised, and consequently the ratio of the between-class variance/within-class variance (between the two similar classes) is decreased, which leads then to the β-error. For the other types of pre-processing there is no β-error at both levels of confidence. Therefore, for these other types of pre-processing, one is able to work at the 99% significance level, where better α-errors are obtained.

Table 2b gives the corresponding results for the SIMCA residual variance method. It can be seen, that the α-errors are around 5 and 3% at the respective levels of confidence, which is expected at these significance levels. Pre-processing again has no strong influence on the α-error. The results are better than for the Mahalanobis distance method. Almost all samples, which are rejected as outliers here, are also outlying in the Mahalanobis distance method. The extreme samples, classified as outliers, represent normal variations, which can occur for excipients. The only β-error again appears for the second derivative data.

As mentioned earlier, the Mahalanobis distance and the SIMCA residual distance method are complementary. Therefore, the individual results from both methods are combined. All objects, which are outlying in one or both methods, are considered to be outliers, so that the full data space is covered. The same is valid for the wrongly classified samples, i.e. the β-errors. The summarised results are given in Table 2c.

As we require that no β-error and the smallest possible α-error must be obtained, we conclude from the results in Table 2c, that the original data, first and second derivative data cannot be used, since β-errors occur. There are only small differences between the five remaining pre-processing methods.

To investigate, whether the models can correctly classify new samples, received from new incoming excipient batches, the data of the historical data set are predicted with the models ob-

Table 2
Classification results (α- and β-errors) obtained with the Mahalanobis distance method, SIMCA residual variance method and the combined Mahalanobis distance and SIMCA residual variance methods carried out at two significance levels for the basic data set (175 samples)

	$\alpha = 0.05$		$\alpha = 0.01$	
	α-Error (%)	β-Error (%)	α-Error (%)	β-Error (%)
(a) Mahalanobis distance method				
Original	12.6	0	6.3	0.13
Offset	12.0	0	5.7	0
De-trending	9.7	0	4.6	0
SNV	9.1	0	5.1	0
SNV+de-trending	9.7	0	5.7	0
MSC	12.0	0	5.1	0
First derivative	10.9	0.32	5.7	0.83
Second derivative	11.4	0.70	5.1	1.08
(b) SIMCA residual variance method				
Original	5.1	0	4.0	0
Offset	4.0	0	2.3	0
De-trending	5.1	0	2.3	0
SNV	4.0	0	3.4	0
SNV+de-trending	4.0	0	2.9	0
MSC	3.4	0	2.3	0
First derivative	4.6	0	4.0	0
Second derivative	4.6	0	2.3	0.13
(c) Combined Mahalanobis distance and SIMCA residual variance methods				
Original	13.1	0	6.3	0.13
Offset	12.0	0	5.7	0
De-trending	10.9	0	4.6	0
SNV	9.7	0	5.7	0
SNV+de-trending	10.3	0	5.7	0
MSC	12.0	0	5.1	0
First derivative	11.4	0.32	5.7	0.83
Second derivative	12.6	0.70	5.7	1.21

Table 3
Classification results (α- and β-errors) obtained with the Mahalanobis distance method, the SIMCA residual variance method and the combined Mahalanobis distance and SIMCA residual variance methods carried out at two significance levels for the historical data set (259 samples)

	$\alpha = 0.05$		$\alpha = 0.01$	
	α-Error (%)	β-Error (%)	α-Error (%)	β-Error (%)
(a) Mahalanobis distance method				
Original	44.8	0.04	27.8	0.04
Offset	49.4	0	36.7	0
De-trending	29.0	0	18.9	0
SNV	47.5	0	28.2	0
SNV + de-trending	48.7	0	34.8	0
MSC	51.7	0	31.3	0
First derivative	37.8	0	25.5	0.09
Second derivative	22.8	0.21	12.4	0.26
(b) SIMCA residual variance method				
Original	54.8	0	52.9	0
Offset	58.7	0	55.6	0
De-trending	59.5	0	54.4	0
SNV	53.7	0	47.9	0
SNV + de-trending	64.1	0	55.2	0
MSC	55.6	0	49.8	0
First derivative	57.5	0	49.4	0
Second derivative	45.2	0	33.6	0.47
(c) Combined Mahalanobis distance and SIMCA residual variance methods				
Original	60.6	0.04	56.4	0.04
Offset	64.1	0	59.1	0
De-trending	64.1	0	57.1	0
SNV	60.6	0	50.6	0
SNV + de-trending	70.3	0	61.0	0
MSC	64.5	0	54.8	0
First derivative	60.6	0	50.6	0
Second derivative	47.1	0.21	34.0	0.73

tained for the basic data set. The results are given in the Table 3a and b, and in Table 3c for the combined methods.

High α-errors are obtained for both methods. The α-errors are mainly due to four classes, where many objects are rejected as outliers. They are anhydrous dicalcium phosphate, magnesium stearate, starch and avicel (class 1, 5, 9 and 10). For those classes the basic data set is not representative for most of the new samples. Fig. 4 shows the PC1 versus PC2 score plot of the SNV transformed avicel data (class 10).

The PC space is defined by the training set samples, the scores of which are represented by stars. The objects of the historical data set are projected into the established PC space and their scores are characterised by points. Most of the predicted samples are clearly lying outside the space spanned by the training set samples, and are therefore rejected as outliers by the classification model. A pattern recognition method can only recognise samples for which it was trained. Therefore the training set plays a central role in any identification system. In the case of the other six excipient classes (class 2, 3, 4, 6, 7, 8), the basic data set is more representative for the historical data set, and as a result most of the data are correctly classified. When a classification model is not able to identify further samples of the class, because the training set is not representative, model updating is required. Model updating consists of incorporating new sources of variance, which occur in practice, in the classification model in order to make it more robust. Fewer objects are rejected with the Mahalanobis distance method compared to the SIMCA residual variance method. The class volume, defined during the modelling stage, is considerably large in the case of the Mahalanobis distance method. For the SIMCA residual variance method very tight class borders are obtained. It is interesting, that with the second derivative data a smaller α-error is achieved for both methods. This is again an indication that for this type of pre-processing the within-class variance is enlarged, as described before. The class-volumes are then becoming large also, and consequently smaller α-errors occur, but also β-errors. From Table 3c, which shows the combined results from the Mahalanobis distance and the SIMCA residual variance method, it can be seen, that the final classification results are unsatisfactory, since the results are summarised for the ten classes. For most types of pre-processing no β-error occurs at both levels of significance, but α-errors up to 61.0% are found. SNV seems to be the 'best' pre-processing method, at the 99% level of confidence the α-error is 50.6%. These summarised results are not acceptable in an industrial context.

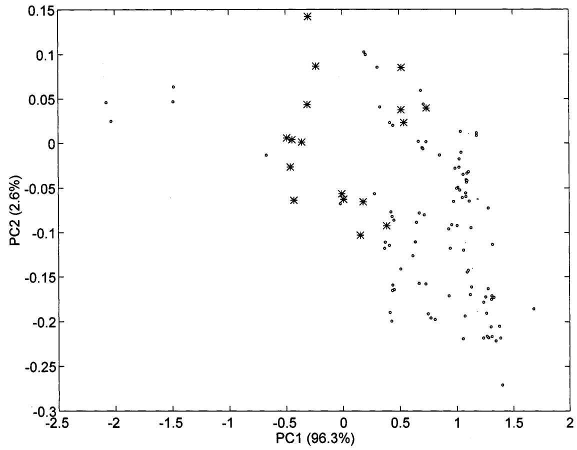

Fig. 4. PC1 versus PC2 scores plot for the SNV pre-treated Avicel data with the training set data (✱) and the predicted test set data (●).

Table 4
Classification results (α- and β-errors) obtained with the wavelength distance method (samples) and carried out with $t_{max} = 6$ for the basic (175 samples) and the historical (259 samples) data sets

	$T_{max} = 6$ α-Error (%)	β-Error (%)
(a) Basic data set		
Original	0.6	32.0
Offset	1.1	2.10
De-trending	2.3	0
SNV	1.7	0
SNV + de-trending	1.1	0
MSC	1.7	0
First derivative	2.3	0
Second derivative	2.9	0
(b) Historical data set		
Original	0	29.34
Offset	6.6	4.59
De-trending	6.2	0
SNV	19.3	0
SNV + de-trending	21.2	0
MSC	7.0	0
First derivative	7.3	0
Second derivative	28.6	0

The wavelength distance method is a simple univariate classification method. The basic statistical test in this method is the t-test. As explained in the theory section multiple t-tests are necessary for testing NIR spectra, one for each variable, and therefore the critical t-value has to be increased in order to maintain the overall α-error. As shown in [4] a t_{max} value of 6, for $n = 15$ and $p = 700$, leads to good results. This critical value for t is also proposed by the commercial software. Therefore we constructed the classification models with a t_{max} value of 6. The results are included in Table 4a.

It can be seen, that the overall α-errors are satisfactorily small for all types of pre-processing. However, large β-errors occurred for the original and offset data, which means that these method combinations are not acceptable. Concerning the remaining method combinations it looks indeed reasonable to apply the wavelength distance method with a $t_{max} = 6$. The method performance is again tested with the historical data set. The results are given in Table 4b.

As was to be expected β-errors occur again for the original and offset corrected data. Concerning the α-errors, one observes two groups of results. For de-trending, MSC and first derivative, recognition rates up to 93% are achieved. These pre-processing methods correct especially for the baseline. This is also true for MSC as applied here, where the correction parameters are defined by the first 200 nm of the spectra. This spectral band contains mostly baseline information. Due to the small spectral corrections, the standard deviations for the spectra at most variables are still rather large and so are the class volumes. Therefore, good recognition rates are obtained in the prediction of new samples. For SNV, SNV combined with de-trending and second derivative the results are less good, recognition rates around 70–80% are obtained. The reason in the case of SNV is the following. SNV reduces the original variance, which is due to particle size and scattering. The remaining variance is spread over the entire spectrum. As a result the standard deviation for most variables is small and tighter class volumes are obtained. Possible problems, which are connected to SNV, are described in [37]. In the case of the second derivative data the spectra cross each other many times due to the differentiation. At the location of the crossing the standard deviation is becoming very small, which is then problematic in the wavelength distance method, and is responsible for the large α-error. Simple baseline correction methods, such as de-trending or first derivative, combined with the wavelength distance method lead to acceptable results for this data set, namely to correct identification of around 93% of the excipient samples.

Classification methods based on potential functions have been used in the field of pattern recognition, but they were not yet applied in the present application. In potential functions the most important step in the training phase is to find a suitable smoothing parameter. In order to optimise it, a compromise between an acceptable α- and β-error is often necessary. We require, that no β-error occurs. This means, that one can theoretically increase the smoothing as long as this obligation is fulfilled. Two methods are applied to define the smoothing factor, the centroid method and LOOCV. The smoothing factor is determined for each class separately during modelling. It is optimised by determining the α-errors obtained with the triangular potential function method performed with a stepwise increased K nearest neighbour distance. The smoothing is considered to be appropriate if the α-error is acceptably small and no β-error occurs.

For the centroid method, where in total 1464 centroids are possible for the ten classes, a smoothing from 2 to 17 nearest neighbours median distance was found optimal for the individual classes. Generally smaller smoothing factors are determined by LOOCV. There a value between 2 and 11 is usually optimal, except for one case (class 9, second derivative data) where a $K = 15$ is needed. The results, obtained with the corresponding smoothing factors for the basic data set are presented in Table 5a.

The table shows the α- and β-errors obtained for the centroid method and by LOOCV. As the smoothing parameter was optimised such that no wrong acceptance of samples would occur, the β-error is zero. The α-error in the centroid method varies from 0 to 6.3% and represents the amount of centroids, which have a zero potential for the given smoothing factor and are therefore outlying. It is evident that better validation results are obtained when models are established with large smoothing parameters. However, a high smoothing factor leads to a flat global potential field and to large positive zones on the border of the class. This situation can be dangerous for possible β-errors when predicting new samples. A small smoothing on the other hand can create local potential fields around isolated objects or groups of objects within single classes, leading to possible α-errors in the prediction step. With LOOCV α-errors between 2.9 and 8.0% are obtained. The small α-errors for the derivative data are due to the large smoothing factor for class 9.

Data pre-processing mainly influences the selection of the smoothing parameter. For homogeneous data smaller K-values are found optimal compared to heterogeneous data. Consequently, if spectral pre-processing removes data inhomogeneities (e.g. originating from different particle sizes of the powders within a class), smaller smoothing parameters are found optimal after data transformation. The data of the historical data set are analysed. The α- and β-errors are presented in Table 5b.

No β-errors are observed, even with large smoothing factors. The individual α-errors for the different types of pre-processed data vary between 15.4 and 42.5% for the centroid methods, and between 25.1 and 45.6% for LOOCV. For both methods the best results are achieved with offset correction, the worst results with SNV + de-trending. As it was to be expected, the prediction results depend on the smoothing parameter. For offset correction rather large smoothing factors were necessary to obtain acceptable results for the training set. In the case of SNV + de-trending small smoothing factors are defined during modelling. The classification results are not very satisfying. The reason for that is again the data itself: as explained before, the training sets of several classes are not representative for the prediction of the new excipient samples. Based on the results presented in Table 5b, we conclude that the best approach for the potential function method in this application is to work with offset correction and to use the smoothing factors defined by the centroid method. For this combination an α-error of 6.3% is obtained for the training set, and of 15.4% for the prediction of the historical data set.

It is shown that the influence of data pre-processing depends on the data and the pattern recognition method. The important feature of pre-processing is found in reducing possible β-errors. Transforming NIR spectra mostly decreases the within-class variance, so that possible β-errors might be eliminated. An exception to that was found in the case of using derivatives. This data transformation emphasis small spectral differences. This effect is often desired, but it appeared here to be problematic, since the spectra within certain classes were slightly different in shape, which can happen in practice if excipient samples from different batches are measured. Therefore this pre-processing method should be applied with special care. In this study de-trending, SNV and MSC never lead to any β-error with any pattern recognition method, i.e. the classes are always well separated. This situation is particularly fa-

Table 5
Classification results (α- and β-errors) obtained with the triangular potential function approach carried out with the centroid method and with LOOCV for the basic data set (175 samples) and the historical data set (259 samples)

	Centroid		LOOCV	
	α-Error (%)	β-Error (%)	α-Error (%)	β-Error (%)
(a) Basic data set				
Original	5.9	0	8.0	0
Offset	6.3	0	5.7	0
De-trending	2.9	0	5.1	0
SNV	1.4	0	5.1	0
SNV + de-trending	0	0	4.0	0
MSC	0.4	0	4.6	0
First derivative	3.0	0	2.9	0
Second derivative	4.2	0	2.9	0
(b) Historical data set				
Original	21.6	0	42.5	0
Offset	15.4	0	25.1	0
De-trending	25.9	0	38.2	0
SNV	42.1	0	45.2	0
SNV + de-trending	42.5	0	45.6	0
MSC	21.2	0	38.2	0
First derivative	28.6	0	36.7	0
Second derivative	23.6	0	28.6	0

vourable for including further excipient classes in the system at a later time. Data pre-processing does not influence the α-error obtained for the training set in the case of the Mahalanobis distance and the SIMCA residual variance method. The wavelength distance method is sensitive to small standard deviations for some individual variables. Therefore for this method, pre-processing may influence the α-error, namely, if the SD for certain wavelengths are very small after the transformation (e.g. multiple spectra-crossing in the case of the second derivative data). In the potential function method the data pre-processing influences the choice of the smoothing factor.

From the four investigated classifiers, the wavelength distance method performed best for the studied excipient data. Thanks to its univariate character large class volumes are constructed (neglecting the correlation between variables) and most test set samples are correctly identified. It is clear, that this pattern recognition approach is only successful for situations, where the excipient classes are very different and proper data pretreatment is applied in order to optimise the ratio of the between-class variance over the within-class variance. The Mahalanobis distance and SIMCA residual variance method are complementary methods and should be used together. The results obtained for the historical data set with these multivariate methods appear to be unacceptable. This is partly due to the parametric way the class borders are determined. A version of SIMCA, based on robust statistics, might possibly improve the results. However, even more important is the fact, that the data of the training sets for four classes are not representative for the future samples to be analysed. This situation may occur, if excipient suppliers subtly change the manufacturing process of a certain material. The multivariate approach based on potential functions suffers from the same data problem. The results are however better compared to the SIMCA method. This can be explained by the non-parametric characteristic of the method, and by the larger flexibility of the method, i.e. the possibility of choosing large smoothing values (as long as the β-error remains zero).

More than 93% of the samples from the historical data set are correctly classified with the wavelength distance method applied to de-trending, MSC and first derivative data. This means that 93% of the time-consuming conventional pharmacopoeial identification methods can be replaced by fast NIR identification. Moreover no β-error occurred. From the point of view of application MSC is more complex compared to the other transformations. This transformation method depends of a defined ideal spectrum (here the mean spectrum) of the training set. De-trending and first derivative are similar methods. Between these two, de-trending can be preferred, since the shape of the spectra remains unchanged.

5. Conclusions

It was shown, that transforming excipient NIR spectra obtained from powdered materials with an appropriate pre-processing method is generally advised. Possible β-errors can be eliminated. Concerning pre-processing, we suggest applying SNV first. It may happen, that using this data transformation not the best α-error is obtained (as it is the case here with the wavelength distance method), but the within-class variance due to particle size effects is considerably reduced, thereby avoiding β-errors. Other methods, which are sometimes useful, are: MSC and de-trending. As described, special care should be taken, when working with derivatives.

Concerning the pattern recognition method, we propose to perform a first trial with the wavelength distance method. This method is indeed successful in the case of an easy data set, i.e. for a classification situation consisting of a few classes only (here ten classes), where the between-class variance is larger than the within-class variance for all variables. However, it is not evident that this method will be always best. When the number of substances studied and the number of batches for each substance grows, it may be that better discriminating power will be required and that methods such as potential function methods or a robust variant of SIMCA will be the better solution. In that case, for the studied data set, model updating procedures will be needed, such as including new incoming samples, which could not be correctly classified, into the training set of the class to which they belong, and building a new model. This subject is now under study.

References

[1] J. Griffin, W. Kohn, NIR Spectroscopy as a Key Element in Total Quality Management in the Pharmaceutical Industry, Ciba-Geigy AG, Analytical Promotion K-127.2.30, CH-4002 Basel.

[2] P.J. Gemperline, L.D. Webber, F.O. Cox, Anal. Chem. 61 (1989) 138–144.

[3] N.K. Shah, P.J. Gemperline, Anal. Chem. 62 (1990) 465–470.

[4] P.J. Gemperline, N.R. Boyer, Anal. Chem. 67 (1995) 160–166.

[5] P. Corti, L. Savini, E. Dreassi, G. Ceramelli, L. Montecchi, S. Lonardi, Pharm. Acta Helv. 67 (1992) 57–61.

[6] G. Downey, Analyst 119 (1994) 2367–2375.

[7] B.G. Osborne, T. Fearn, P.H. Hindle, Practical NIR spectroscopy, 2nd edn., Longman, Harlow, UK, 1993.

[8] D.L. Wetzel, Anal. Chem. A 55 (1983) 1165A–1176.

[9] T. Naes, T. Isaksson, NIR News 1,5,15 (1994) 2.

[10] A. Candolfi, W. Wu, D.L. Massart, S. Heuerding, J. Pharm. Biomed. Anal. 16 (1998) 1329–1347.

[11] A. Candolfi, R. De Maesschalck, P.A. Hailey, A.C.E. Harrington, D.L. Massart, Identification of pharmaceutical excipients using NIR spectroscopy and SIMCA, J. Pharm. Biomed. Anal. 19 (1999) 923–935.

[12] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Appl. Spectrosc. 43 (1989) 772–777.

[13] P. Geladi, D. MacDougall, H. Martens, Appl. Spectrosc. 39 (1985) 491–500.

[14] T. Isaksson, T. Naes, Appl. Spectrosc. 42 (1988) 1273–1284.

[15] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1994) 1627–1639.

[16] P.A. Gorry, Anal. Chem. 62 (1990) 570–573.

[17] M.P. Derde, D.L. Massart, Anal. Chim. Acta 191 (1986) 1–16.

[18] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: A Textbook, Elsevier, Amsterdam, 1988, p. 385.

[19] D. Jouan-Rimbaud, E. Bouveresse, D.L. Massart, O.E. de Noord, Detection of prediction outliers and inliers in multivariate calibration, Anal. Chim. Acta 388 (1999) 283–301.

[20] S. Wold, M. Sjöström, in: B.R. Kowalski (Ed.), Chemometrics: Theory and Application, American Chemical Society, Washington, DC, 1977, pp. 243–281.

[21] M.P. Derde, D.L. Massart, Chemometr. Intell. Lab. Syst. 4 (1988) 65–93.

[22] B. Mertens, M. Thompson, T. Fearn, Analyst 119 (1996) 2777–2784.

[23] R. De Maesschalck, A. Candolfi, S. Heuerding, D.L. Massart, Decision Criteria for SIMCA Applied to Near Infrared Data, Chemometr. Intell. Lab. Syst. 52 (1999) 63–75.

[24] N.D. Tracy, J.C. Young, R.L. Mason, J. Quality Technol. 24 (1992) 88–95.

[25] H.L. Mark, D. Tunnell, Anal. Chem. 57 (1985) 1449–1456.

[26] H.L. Mark, Anal. Chem. 58 (1986) 379–384.

[27] M.A. Dempster, B.F. MacDonald, P.J. Gemperline, N.R. Boyer, Anal. Chim. Acta 310 (1995) 43–51.

[28] S. Wold, M. Sjoestroem, J. Chemometr. 1 (1987) 243–245.

[29] IQ2 V.1.02, NIRSystems, Inc., Silver Spring, MD (1990).

[30] C.I. Gerhaeuser, K.-A. Kovar, Appl. Spectrosc. 51 (1997) 1504–1510.

[31] D. Coomans, D.L. Massart, I. Broeckaert, A. Tassin, Anal. Chim. Acta 133 (1981) 215–224.

[32] D. Coomans, D.L. Massart, Anal. Chim. Acta 133 (1981) 225–239.

[33] D. Coomans, I. Broeckaert, Potential Pattern Recognition in Chemical and Medical Decision Making, Research Studies, UK, 1986.

[34] O.E. De Noord, Chemometr. Intell. Lab. Syst. 25 (1994) 85–97.

[35] E. Bouveresse, D.L. Massart, Vibrational Spec. 11 (1996) 3–15.

[36] E.R. Malinowski, Factor Analysis in Chemistry, 2nd. edn, Wiley, New York, 1991.

[37] W. Wu, Q. Guo, D.L. Massart, The robust normal variate transform for pattern recognition with near-infrared data, Anal. Chim. Acta 382 (1999) 87–103.