Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis 37 (2005) 109–114
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Implementation of a simple semi-quantitative near-infrared method for the classification of clinical trial tablets

R. De Maesschalck*, T. Van den Kerkhof

Janssen Pharmaceutica NV, Pharmaceutical Research and Development, Global Analytical Development, Turnhoutseweg 30, B-2340 Beerse, Belgium

Received 15 July 2004; received in revised form 30 September 2004; accepted 10 October 2004
Available online 8 December 2004

Abstract

Near infrared transmission spectroscopy combined with chemometrical methods can be applied for identity confirmation of double-blind clinical trial tablets. Samples of two clinical studies, investigating the dose and placebo effect of an experimental drug, were studied. The identity of the blistered tablets was checked using partial least squares beta classification (PLSBC) applied to their NIR transmission spectra. PLSBC is a new supervised classification approach based on partial least squares (PLS) regression combined with β-error driven class boundaries. It has the ability to limit the probability for misclassification to a known number and therefore providing the method developer a tool for deciding whether the NIR spectra of the different strengths of tablets are specific enough to obtain a robust classification model. The presented approach has the advantage to be applicable on most commercial available near infrared spectroscopy (NIRS) instrumentation software and it can be applied in a GMP environment since validation according to the ICH Q2A and Q2B guidelines on analytical method validation is fast and relatively easy.

Keywords: Double-blind clinical studies; NIR; PLS; Beta classification; Validation

1. Introduction

During the development of a new drug product, several clinical studies are carried out to evaluate the effect of different dosage strengths compared to the effect of a placebo or a clinical registered comparator. To assure objective results, most clinical studies are double-blind. The correctness of the packaging and labelling of the blister packs needs to be checked before shipping the samples. It is important to limit the waiting time between packaging and the shipping as much as possible. Identification of the samples is usually performed with high performance liquid chromatography (HPLC), thin layer chromatography (TLC), or UV spectroscopy. Near infrared spectroscopy (NIRS) has been shown to be a less time-consuming alternative. Near infrared spectroscopy has gained acceptance by the regulatory agencies which was reflected in the publication of a draft USP chapter on the technique [7]. In a GMP environment NIRS can be considered as a true alternative for the identity confirmation if it complies with the following basic conditions: (a) the method must be 100% reliable, i.e. no dosage forms attributed to the wrong class; (b) the method development and the routine application must be fast; (c) the model must be robust and applicable for several studies and batches of drug product although only a limited number of tablets initially is available for building the model, (d) the chemometrical pattern recognition technique must be available or applicable on the instrument software; (e) the method must be validatable according to the ICH guidelines for analytical method development Q2A and Q2B [5,6] and (f) the instrumentation must be qualified and 21 CFR part 11 compliant. This article will focus on points (a), (b), (d) and (e).

NIR measurements on solid dosage forms can be performed in diffuse reflectance mode [2] or transmittance mode [11–13]. The obtained NIR spectra need to be pretreated using mathematical algorithms such as, e.g. standard normal variate (SNV) [14], first derivative or second derivative [15]

* Corresponding author. Tel.: +32 14605576; fax: +32 14605034.
E-mail address: rdmaessc@prdbe.jnj.com (R. De Maesschalck).

0731-7085/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2004.10.016

in order to remove the physical information in the spectra such as, e.g. scattering. An overview of typical pretreatment methods for NIRS can be found in literature [16,17]. After pretreatment, different chemometrical classification techniques can be applied to the measured NIR spectra. They can be divided into so-called supervised and unsupervised techniques [18]. The supervised classification techniques are suitable since one starts with a set of spectra with a known identity, the so-called training set. Each type of spectra or tablets is called a class. The training set should contain a representative set of spectra of each class of tablets. A classification model is built using the training set which allows to predict whether a newly measured spectrum belongs to a certain class. There are several supervised classification techniques described in the literature [18] and some are available in the commercial software packages of the NIR instrumentation or in stand-alone chemometrical software packages. Although very powerful, stand-alone software packages often are a burden towards application in a GMP environment since they are difficult to qualify due to the large number of available techniques and due to the necessary data format transformations.

A well known supervised classification method is SIMCA [3,19,21]. The fact that the technique develops a separate model for each class [3,4] has as a consequence that the most discriminating spectral features between the classes do not necessary get the highest importance. The linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) techniques [1,20] model all classes together and determine the features which are the most discriminatory between the different classes. The techniques have the disadvantage that every spectrum must be attributed to one of the modelled classes including spectra which are near to the borders of two or more classes. This can lead to misclassifications. Discriminant PLS (DPLS) [8–10] is easy to perform using the PLS algorithm [22] available in most commercial instrument software packages. DPLS has been mostly used to check whether a sample belongs to a certain class or not. Usually the *y*-value is set to 1 (belonging to the class) and 0 (not belonging to the class) [8,9]. In case of three classes, three DPLS models are needed, or alternatively, only one model can be applied when, e.g. giving the *y*-variable the values 1, 2 or 3 for the different classes. The class boarders are set in the middle of the *y*-values predicted by the model. This however means again that each spectrum must be attributed to a certain class and that again no direct indication is available on the probability for misclassification. In a GMP environment in the pharmaceutical industry it is unacceptable to have misclassifications and therefore, a simple semi-quantitative approach partial least squares beta classification (PLSBC) is proposed. The nominal concentrations of the tablets are used as y-variables and the limits are based on the probability for misclassifications (β-error) instead of the α-error as is the case for most classification techniques. It is shown that using the proposed approach one can correctly identify clinical trial tablets and also easily comply with the ICH guidelines on analytical method validation.

2. Theory

2.1. PLSBC

In a first step a PLS regression model is built using the nominal concentration of the calibration samples. Compared to DPLS this has the advantage that optimal correlation determined between *X* and *y* is more likely to be concentration related and more meaningful than when, e.g. 0 and 1 are used as *y*-variables. The obtained model has to be chemometrically validated by leave-one-out or leave-a few-out cross validation (LOOCV or LFOCV) [22]. The final PLS model is described by a selected spectral region, a certain spectra pretreatment and a number of PLS factors. The chemometrical method validation yields predicted *y*-values of the training set spectra (internal validation). Provided that the training set samples are representative for future samples, the deviation between the predicted and nominal concentration is a good indication for the prediction error which can be expected for new tablet spectra.

Secondly, boundaries have to be set to assign a sample to a certain class. We consider the total population of spectra (tablets) of a class to be normally distributed around the nominal concentration. It is assumed that the nominal value is a good estimate of the real value when considering multiple tablet batches due to the strict content uniformity limits imposed for the release of the clinical tablets. The boundaries of the classification model are set by using the average and standard deviation of the values predicted by internal validation. The boundaries of neighbouring classes may not exceed each other. The predictions by internal validation are often an underestimation of the real variation. If enough samples are available, it may be better to set the boundaries based on the predicted average and standard deviation of the independent test set. In our case the internal validation is used as the independent test will be applied for the regulatory validation of the model specificity. Upper boundary of class 1

$$(\mathrm{UB}_1) \leq \text{lower boundary of class 2 } (\mathrm{LB}_2). \tag{1}$$

According to statistics, when the number of samples for a class is high enough ($\geq$30), it can be assumed that the sample average, $\langle y \rangle$, and standard deviation, s, are equal to the average, μ, and standard deviation, σ, of the total normal distributed population. In this case the boundaries for a certain class are set by:

$$\mathrm{LB} = \mu - z\sigma \qquad \mathrm{UB} = \mu + z\sigma \tag{2}$$

For a certain *z*-value, the corresponding confidence interval and α-error can be found in a two sided normal distributed table. When two neighbouring classes are considered, there is not only a probability to obtain false outliers (α-errors), but also a chance to wrongly classify a sample (β-error). The probability for such a misclassification can be determined by calculating the α-error in case the upper boundary of the first

class coincides with the lower boundary of the second class.

$$\mathrm{UB}_1 = \mathrm{LB}_2 \qquad \Leftrightarrow \mu_1 + z_1\sigma_1 = \mu_2 - z_2\sigma_2 \tag{3}$$

Now, let's consider that the boundaries of the two classes are set in such a way that they cover the same confidence interval

$$z_1 = z_2 = z_{\max} \qquad \Leftrightarrow z_{\max} = \frac{\mu_2 - \mu_1}{\sigma_2 + \sigma_1}. \tag{4}$$

When the boundaries are constructed with the $z_{\max}$-value a discriminating classification model is obtained. When the classification boundaries are constructed with a z-value lower than $z_{\max}$, the chance for a β-error will decrease, while the probability for an α-error will increase. When the number of spectra belonging to a certain class is lower than 30, the population average, μ, and standard deviation, σ, have to be considered as unknown. In this case the z-value has to be replaced by the Student t-value, where the number of spectra, n, equals the degrees of freedom. The boundaries of a class are then set by:

$$\mathrm{LB} = \langle y \rangle - t_n s \qquad \mathrm{UB} = \langle y \rangle + t_n s. \tag{5}$$

As the Student t-distribution is a bit broader compared to the normal distribution, the chance for obtaining α- and β-errors becomes a bit larger. This increase obviously depends on the number of spectra. In order to obtain a robust classification model, we recommend to always measure at least 18 spectra per class.

The construction of the classification boundaries is based upon the β-error, which must always be lower than 0.005%. Furthermore, a class-modelling approach is preferred above a discriminating one, because it has the ability to detect non-representativeness of the sample or of the chemometrical model. At the same time, the classification model should try to assign every sample to a class, i.e. a minimal probability for an α-error. Therefore, we propose to first calculate, depending on the number of spectra, the $z_{\max}$- or $t_{n,\max}$-values between every two neighbouring classes. When a minimum number of spectra per class (18) is used and a $t_{18,\max}$ value of 5 is obtained, the β-error shall always be lower than 0.005%. If a value of 5 or higher is calculated, a robust classification model can be achieved. Depending wether the achieved $z_{\max}$-, or $t_{\max}$- value is respectively lower than 6; lower than 7 and; equal to or higher than 7, the boundaries are constructed with a z- or t_n-, value of respectively, 4, 5 and 6. This approach is clarified using the examples shown in the experimental part.

2.2. Regulatory method validation

The objective of the regulatory validation of an analytical method is to demonstrate that it is suitable for its intended purpose. Our intended application is the identification of clinical trial supplies. According to the ICH and pharmacopeial guidelines only the characteristic specificity needs to be evaluated. Specificity for a NIR method for the identification of clinical trial samples means that unknown clinical trial supplies are unambiguously assigned to the correct class. In case of the PLSBC technique, the specificity of the method is tested by predicting the identity of a set of tablet NIR spectra with a known class identity. The criterion for the validation parameter specificity is that none of the tested tablets are misclassified (no β-error).

3. Experimental

3.1. Samples

The samples of two clinical studies, for investigating oral immediate release tablets containing haloperidol (Haldol) and risperidone (Risperdal), were studied. The Haldol tablets used in this work are coated circular tablets with a total weight of about 91 mg and were manufactured to contain 1, 2, 4 and 5 mg of haloperidol, or to be a matching placebo. Each of the five classes consists of 40 tablets, of which 20 are randomly selected for inclusion in the training set, and the remaining 20 tablets are reserved for the regulatory validation of the developed classification model.

The Risperdal tablets are coated circular tablets with a total weight of 100 mg, and were manufactured to contain 0.25, 0.5, 1, 2, 3 and 4 mg of risperidone, or to be a matching placebo. Again, the spectra of each class were randomly divided for calibration and regulatory validation.

3.2. Instrumentation and software

The FT-NIR spectra were recorded in transmission on a Bruker Vector 22/N-T spectrometer, using a Tungsten Halogen source in combination with a Quartz beamsplitter and an InGaAs detector. The interferograms were recorded with a resolution of 8 cm^{-1}, averaged over 32 scans, Blackman–Harris 3-term apodized and Fourier transformed with a zero filling factor of 2.

Calibration models were built using the Quant 2 software package version 4.2, an add-on to the general OPUS spectroscopy software version 4.2. Quant 2 contains the PLS regression technique needed for the PLSBC approach. Setting the class boundaries was performed by using a validated Excel sheet (Microsoft).

4. Results and discussion

The NIR spectra of Haldol tablets belonging to the different classes are shown in Fig. 1, for visibility reasons these spectra are presented in a shifted form. A visual inspection of these spectra, shows that a weak broad drug specific band in the 9000–8700 cm^{-1} region can be observed for tablets with a relatively high amount of active compound. The second derivative spectra (Fig. 2) clearly show that all the active tablets possess this active-dependent spectral information,

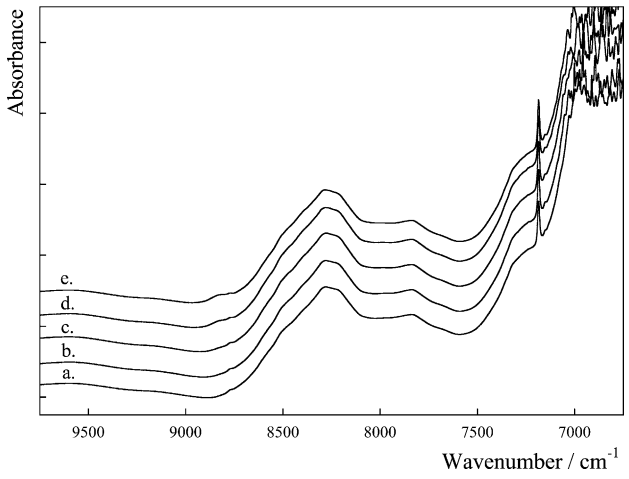

Fig. 1. Near infrared spectra of tablets used in Haldol clinical trial studies. (a) Placebo; (b) 1.1 % w/w API; (c) 2.2 % w/w API; (d) 4.4 % w/w API and (e) 5.5 % w/w API. The spectra are presented in a shifted form.

and that all the different classes can be distinguished from one another. A calibration model was built for the spectral region 10000–7500 cm^{-1} with 20 tablets per class. The best calibration model was obtained after taken the SNV of each spectrum and using 7 PLS components. The performance of a leave-5-out cross validation on the calibration model resulted in a prediction error (RMSECV) of 0.0361 and a regression coefficient (R^2) of 0.9996. The $t_{20,max}$-values between neighbouring classes are all higher than 7, so a reliable classification model can be built. Here, the classification boundaries are constructed according to equation (5), with a t_{20}-value of 6. This implies that borders are set in such a way that they cover a 99.99928% confidence interval. Here, the biggest chance for a β-error will be the assignment of a 4 mg tablet to the 5 mg class. The probability for such a misclassification amounts to approximately 6.2×10^{-14}, which is clearly negligible. The cross validation results and classification model are summarized in Table 1.

Table 1
Calibration results and classification model for the Haldol clinical study

Data pre-treatment			SNV
Spectral region			10000–7500 cm^{-1}
No. of PLS factors			7
Training set			20 spectra/class
Test set			20 spectra/class
Cross-validation results (leave 5-out)			
RMSECV	0.0361		
Class	$\langle y \rangle$	s	$t_{20,max}$
1	0.002	0.032	18.3
2	0.990	0.022	18.0
3	2.016	0.035	24.3
4	3.984	0.046	12.5
5	5.009	0.036	
Classification model			
−0.190 mg < placebo < 0.194 mg			
0.858 mg < 1 mg < 1.122 mg			
1.806 mg < 2 mg < 2.226 mg			
3.708 mg < 4 mg < 4.260 mg			
4.793 mg < 5 mg < 5.225 mg			

SNV pretreated NIR transmission spectra of a 1 mg and 4 mg Risperdal tablet and a matching placebo are shown in Fig. 3. Even though the raw NIR spectra and their pretreated spectra, by using different mathematical algorithms, of the different strengths of Risperdal tablets reveal visually no clear drug dependent spectral information, a calibration model could be built. Here, the best calibration model was obtained after taken the standard normal variate of each spectrum, selecting a spectral regio of 10000–8750 cm^{-1}, and using 6 PLS components. A prediction error (RMSECV) of 0.0359 and a regression coefficient (R^2) of 0.9993 were ob-

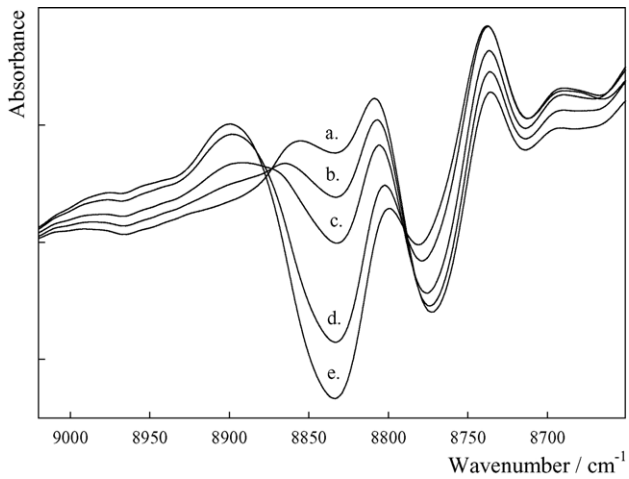

Fig. 2. Second derivative of the near infrared spectra of Haldol tablets. (a) Placebo; (b) 1.1 % w/w API; (c) 2.2 % w/w API; (d) 4.4 % w/w API and (e) 5.5 % w/w API.

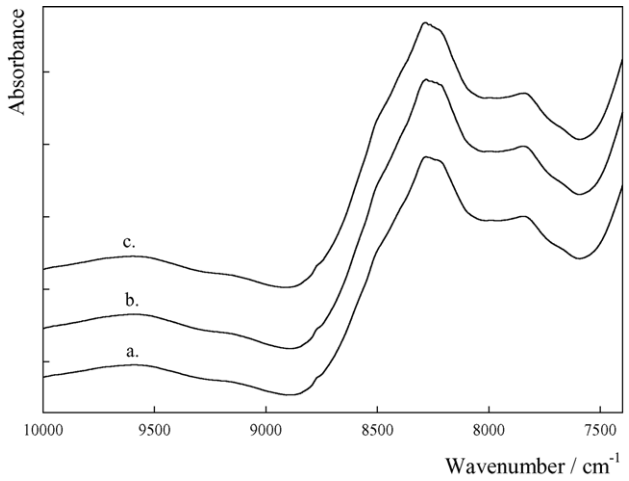

Fig. 3. Standard normal variate pretreated near infrared spectra of tablets used in Risperdal clinical studies. (a) Placebo; (b) 1 % w/w API and (c) 4% w/w API. The spectra are presented in a shifted form.

Table 2
Calibration results and classification model for the Risperdal clinical study

Data pretreatment	SNV
Spectral region	10000–8750 cm^{-1}
No. of PLS factors	6
Calibration set	20 spectra/class
Validation set (regulatory)	20 spectra/class

Cross-validation results (leave 5-out)

RMSECV	0.0408		
Class	$\langle y \rangle$	s	$t_{20,max}$
1	−0.009	0.037	4.07
2	0.268	0.031	3.67
3	0.503	0.033	7.19
4	0.992	0.035	14.51
5	1.993	0.034	13.18
6	2.995	0.042	12.72
7	4.000	0.037	

Classification model

Placebo or 0.25 mg or 0.5 mg < 0.782 mg
0.782 mg < 1 mg < 1.202 mg
1.789 mg < 2 mg < 2.197 mg
2.743 mg < 3 mg < 3.247 mg
3.778 mg < 4 mg < 4.222 mg

Table 3
Regulatory validation results for the Haldol classification model

	Class				
	Placebo	1 mg	2 mg	4 mg	5 mg
No. of test set tablets	20	20	20	20	20
No. of correctly identified	20	20	20	20	20
No. of wrongly identified	0	0	0	0	0
Average prediction value	0.00	0.99	2.03	4.01	5.03
Standard deviation	0.02	0.02	0.03	0.04	0.06

tained. The resulting $t_{20,max}$-values between the neighbouring classes are shown in Table 2. As several $t_{20,max}$-values are lower than 5, according to the above described PLSBC technique no robust classification model can be constructed for the identity confirmation of all the different strengths of Risperdal tablets. According to Table 2, the PLSBC technique allows to discriminate between neighbouring classes with a content of active higher than 0.5 mg. If only the classes with a content of active lower than 1 mg are used to built a calibration model, still no discrimination can be made between these classes. Table 2 shows that classification boundaries for the high content classes can be set in a highly accurate and reliable way, i.e. a t_{20} of 6. So, a partial PLSBC classification model can be built for which 1, 2, 3 and 4 mg Risperdal tablets can be identified, while the other tablets are left unassigned. Because of the fast and non-destructive nature of NIR measurements, routine analysis for the identity confirmation of these clinical studies should always start with the above described NIR–PLSBC technique. Subsequently, all remaining unidentified samples have to be analysed by a more traditional and time-consuming technique, such as UV spectroscopy.

The regulatory validation of the two described analytical methods is performed by confirming the known class identity of 20 tablets per class. The validation is very fast and easily performable, as only the NIR spectra have to be measured and their predicted concentration has to be evaluated with respect to the developed classification model. The regulatory validation results of the Haldol and Risperdal classification models are summarized in Tables 3 and 4. For the Haldol classification model all test set tablets are correctly identified. In fact, none of the test set spectra have a predicted concentration that's close to one of the class boundaries. For the validation of the Risperdal NIR–PLSBC method the spectra of the 1, 2, 3 and 4 mg test set tablets are positively identified, while all the test set tablets of the remaining classes have a predicted concentration lower than the lower boundary of the 1 mg class and are consequently left unassigned.

Clinical trial samples of the two studies, Haldol and Risperdal, previously were solely identified with UV spectroscopy. Due to time-consuming sample preparation, about 8 h were needed to confirm the identity of 100 clinical samples. With the above described NIR–PLSBC technique, these samples were identified in less than 2.5 h and still could be shipped the same working day. For UV spectroscopy no additional time has to be added for method development and validation, as the content uniformity method can be adopted for identification purposes. As the development and validation of a NIR–PLSBC classification model takes about 16 h, the technique yields a gain in time when the clinical study demands the identity confirmation of more than 290 samples. Concerning waiting time before shipment, there

Table 4
Regulatory validation results for the Risperdal classification model

	Class						
	Placebo	0.25 mg	0.5 mg	1 mg	2 mg	3 mg	4 mg
No. of test set tablets	20	20	20	20	20	20	20
No. of correctly identified	0	0	0	20	20	20	20
No. of wrongly identified	0	0	0	0	0	0	0
No. of unassigned	20	20	20	0	0	0	0
Average prediction value	0.00	0.25	0.50	0.99	1.99	3.00	4.01
Standard deviation	0.03	0.03	0.03	0.04	0.05	0.04	0.04

is a clear advantage to develop the NIR–PLSBC classification model before the arrival of the clinical samples for the identification.

5. Conclusion

Near infrared spectroscopy in combination with the PLSBC technique was successfully implemented in a GMP environment. The PLSBC technique combines PLS modeling with the construction of class boundaries designed to minimize the risk for misclassification (β-error), which is very important for this type of analysis. The approach was demonstrated for two examples out of daily practice. The developed model was shown to be able to correctly identify clinical study tablets with an amount of active of respectively 1, 2 and 3% g/g using their NIR transmittance spectra. The examined 0.5 and 0.25% g/g tablets are still distinguishable but the risk for misclassifications is no longer negligible. In this case a combination of the classical identification technique (UV spectroscopy) and NIR spectroscopy is possible because of the non-destructive nature of NIR spectroscopy.

Applying the NIR transmittance spectroscopy and PLSBC approach was shown to be much faster than the classical approach with UV spectroscopy, leading to a significant reduction of the waiting time between packaging and shipping of the clinical trial blistered tablets. The gain in time and resources increases with the number of samples to be identified.

The presented approach is applicable on most commercial available NIRS instrumentation software since these all contain the PLS modeling algorithm.

The classification model can be easily validated according to ICH guidelines which determines specificity testing has to be performed. Specificity is demonstrated by the correct classification of a number of newly measured tablets for each class.

References

[1] A. Candolfi, W. Wu, D.L. Massart, S. Heuerding, J. Pharm. Biomed. Anal. 16 (1998) 1329–1347.
[2] M.A. Dempster, J.A. Jones, I.R. Last, B.F. MacDonald, K.A. Prebble, J. Pharm. Biomed. Anal. 11 (1993) 1087–1092.
[3] M.A. Dempster, B.F. MacDonald, P.J. Gemperline, N.R. Boyer, Anal. Chim. Acta 310 (1995) 43–51.
[4] P.K. Aldridge, R.F. Mushinsky, M.M. Andino, C.L. Evans, Appl. Spectrosc. 48 (1994) 1272–1276.
[5] International Conference on Harmonisation, ICH-Q2A, Text on validation of analytical procedures, Federal Register 60 (1995) 11260–11262.
[6] International Conference on Harmonisation, ICH-Q2B, Validation of analytical procedures, Federal Register 62 (1997) 27464–27466.
[7] The United States Pharmacopeial Convention, United States Pharmacopeia 26, Webcom Limited, Toronto, Ontario, 2002, pp. 2388–2391.
[8] J. McElhinney, G. Downey, T. Fearn, J. Near Infrared Spectrosc. 7 (1999) 145–154.
[9] G. Downey, J. McElhinney, T. Fearn, Appl. Spectrosc. 54 (2000) 894–899.
[10] H. Yoshida, R. Leardi, K. Funatsu, K. Varmuza, Anal. Chim. Acta 446 (2001) 485–494.
[11] M. Scheiwe, D. Shilling, P. Aebi, Die Pharmazeutische Industrie 61 (1999) 179–183.
[12] A. Eustaquio, P. Graham, R.D. Jee, A.C. Moffatt, A.D. Trafford, Analyst 123 (1998) 2303–2306.
[13] J. Gottfries, H. Depui, M. Jongeneelen, M. Josefson, F.W. Langkilde, D.T. Witte, J. Pharm. Biomed. Anal. 14 (1996) 1495–1503.
[14] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Appl. Spectrosc. 43 (1989) 772–777.
[15] P.A. Gorry, Anal. Chem. 62 (1990) 570–573.
[16] J. Sun, J. Chemom. 11 (1997) 525–532.
[17] T. Fearn, A look at some standard pre-treatments for spectra, NIR News 10 (3) (1999) 10–11.
[18] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: a Textbook. Data Handling in Science and Technology 2, Elsevier, Amsterdam, 1988, pp. 328–412.
[19] S. Wold, M. Sjostrom, ACS Symp. Ser 52 (1977).
[20] W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D.L. Massart, S. Heuerding, F. Erni, Anal. Chim. Acta 329 (1996) 257–265.
[21] R. De Maesschalck, A. Candolfi, D.L. Massart, S. Heuerding, Chemom. Intell. Lab. Syst. 47 (1999) 63–75.
[22] P. Geladi, B.R. Kowalski, Anal. Chim. Acta 185 (1986) 1–17.