Journal of Pharmaceutical and Biomedical Analysis 48 (2008) 1037–1041

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Short communication

Assuring specificity for a multivariate near-infrared (NIR) calibration: The example of the Chambersburg Shoot-out 2002 data set

Karl H. Norris, Gary E. Ritchie*

11204 Montgomery Road, Beltsville, MD 206705, USA

ARTICLE INFO

Article history:
Received 1 April 2008
Received in revised form 16 July 2008
Accepted 19 July 2008
Available online 3 August 2008

Keywords:
Specificity
Pharmaceutical tablets
Transmittance spectra
Second derivatives
Multivariate calibrations
NIR
PLS
MSC
EMSC
PAT

ABSTRACT

A method for demonstrating specificity has been developed for NIR calibrations involving the use of partial least squares (PLS) regressions. The method is demonstrated with near-infrared transmittance data on pharmaceutical tablets. A regression was performed with PLS and the first principal component spectrum is shown to match the spectrum of the active ingredient as well as a spectrum extracted from the original file. A good match demonstrates that the calibration is specific for the active component. The effects on the match for different pretreatments were evaluated.

1. Introduction

Multivariate NIR calibrations are most often used for the measurement of the active ingredient (API) in pharmaceutical tablets [1,2], but such calibrations, alone, do not provide adequate specificity. Typically as the result of a calibration, several figures of merit are assessed for the quality of a calibration: correlation, root mean square error of the calibration (RMSEC), root mean square error of the prediction (RMSEP), and bias [3]. Correlation (r) addresses what the strength of the (usually linear) relationship is between two variables. Thus, correlation cannot sufficiently explain what the "cause" of a relationship is; therefore says nothing about the specificity of a NIR calibration.

Additionally, while the predictions by a calibration may give results that are in perfect agreement with the laboratory reference method, it is only a measure of the degree to which a NIR measurement predicts. Another way to say this is that one attempts to calibrate X on Y variables and then determines if there is a correlation. The correlation is then validated by assessing the degree to which the regression model can predict representative samples from the same population from which they were calibrated on.

The strength of the prediction can be evaluated for all future samples by comparing the RMSEC plot to the RMSEP. Again, as with correlation, predictions cannot sufficiently explain what the "cause" of a relationship is; therefore says nothing about the specificity of a NIR calibration. "Relationships" means that there is some structured association (linear, quadratic, etc.) between X and Y. Note, however, that even though causality implies association, association does not imply causality.

Causality is not proved by association. Correlation is a measure of the strength of the (usually linear) relationship between two variables. The validity of a prediction is the degree to which a measure predicts the future behavior or results it is designed to predict.

Finally, bias is a measure of the difference between the average or expected value of a distribution (i.e. NIR predictions) and the true value (i.e. assay). As with correlation and prediction, bias cannot sufficiently explain what the "cause" of a relationship is; therefore it too says nothing about the specificity of a NIR calibration.

If correlation, prediction or bias are not sufficient metrics for demonstrating NIR specificity, what is?

Specificity of an analytical method is defined in several places. The AOAC INTERNATIONAL defines specificity as

* Corresponding author at: Division of Standard Development, U.S. Pharmacopeia, 12601 Twinbrook Parkway, Rockville, MD 20852, USA.
E-mail addresses: knnirs@comcast.net (K.H. Norris), GER@usp.org (G.E. Ritchie).

0731-7085/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2008.07.021

"The ability of a method to respond exclusively to the target analyte and not to any degradant, impurity, or other component of the matrix. Very few methods are absolutely specific, so the term "selectivity" is often used for this property. This parameter shows that the method can be used to quantitate the analyte without interference." (Official Methods of Analysis of AOAC INTERNATIONAL, 18th Edition, Appendix E)

As defined by the International Conference on Harmonization (ICH) guidance:

"Ability to assess unequivocally the analyte in the presence of components which may be expected to be present." (ICH Validation of Analytical Procedures: Text and Methodology, Q2R).

The United States Pharmacopeia (USP) chapter <1119> Near-Infrared Spectroscopy defines specificity:

"The extent of specificity testing depends on the intended application. Demonstration of specificity in NIR methods is typically accomplished by using the following approaches: Qualitative—Identification testing is a common application of qualitative NIR spectroscopy. Identification is achieved by comparing a sample spectrum to a reference spectrum or a library of reference spectra. The specificity of the NIR identification method is demonstrated by obtaining positive identification from samples coupled with negative results from materials that should not meet criteria for positive identification. Materials to demonstrate specificity should be based on sound scientific judgment and can include materials similar in visual appearance, chemical structure, or name. Quantitative—Quantitative applications of NIR spectroscopy typically involve establishing a mathematical relationship between NIR spectral response and a physical or chemical property of interest. Demonstrating specificity against a physical or chemical property of interest is based on interpreting both NIR spectral attributes and chemometric parameters in terms of the intended application.

For the purposes of this paper, specificity of a NIR method as described by AOAC, ICH and USP shall have as its core characteristics:

(a) the analyte of interest in the presence of components which may be expected to be present in the sample,
(b) a physical or chemical property of the analyte and or components, and
(c) independent and dependent (X/Y) paired variables

We propose to demonstrate specificity by showing that when the first principal component spectrum is shown to match the spectrum of the active ingredient as well as a spectrum extracted from the original file, then the criterion for specificity has been met.

The USP chapter <1119> Near-Infrared Spectroscopy gives as a criterion for specificity of a quantitative method: wavelengths used by regression analysis for the calibration (e.g., for multiple linear regression (MLR) models) or the loading vector for each factor (e.g., for partial least squares (PLS) or principal component regression (PCR) models), can be examined to verify relevant spectroscopic information that is used for the mathematical model, but does not provide a method for examining these loading vectors or factors. We explore the use of an approach developed by one of the authors, Karl Norris, using (PLS) regression data on a calibration set developed by Gary Ritchie.

2. Experimental

2.1. Materials used

The data set used in the present study is available on the web [4] and is known as the *Chambersburg Shoot-out 2002* data set. This set contains transmission spectra of 655 pharmaceutical tablets recorded on two similar instruments (Foss/NIRSystems Multitab Spectrometers) [5] for a total of 1310 spectra over the spectral region from 600 to 1898 nm with 2 nm increments on the wavelength scale. The data were organized into six different subsets, Calibrate1 (155), Calibrate2 (155), Test1 (460), Test2 (460), Validate1 (40), and Validate2 (40). More details about these spectra are available [6,7].

The weight of the API and the weight of each tablet were provided, but the identity of the API was not made available for the original use of the spectra. The identity and the digital spectrum for each of the four ingredients were made available later. These ingredients are the API tramadol, plus talc, ethyl cellulose, and stearyl alcohol, with the API varying from 151.6 to 239.1 mg with tablet weights from 363.9 to 400 mg. NIR absorbance spectra are more linear with concentration than with amount on samples with total weight variations, so the API concentration for each tablet was computed by dividing the API weight of each tablet by the corresponding tablet weight multiplied by 100 to provide values in percent. The word absorbance will be used to refer to the spectra which were measured as transmission (T) values and converted to $\log(1/T)$ by the spectrometer, although it is recognized that the spectra are not true absorbance spectra.

The software program, Unscrambler 9.7 [8], was used for the multivariate regressions and other spectral manipulations. Since this study is emphasizing a method to assure specificity, all of the spectra except for the validation samples were included in the regression. It was recognized from other studies that the total set of 1230 spectra included a significant number of outliers, but they were not removed in the initial regression to provide a more robust test of the specificity procedure. The 1230 spectra are shown in Fig. 1. Regressions were performed with no pretreatment as well as with multiplicative scatter correction (MSC), extended multiplicative scatter correction (EMSC), and Savitsky–Golay [5] second-derivative conversion using 11 points and order 4 (SG1104).

A relatively low number of points are recommended for the derivative conversion, because the API has relatively narrow absorption bands, and the instrument noise is low enough to permit little smoothing.

3. Results and discussion

This study began with the discovery that it was possible to extract a spectrum from the data set that matched the spectrum of the API. This extracted spectrum was computed by subtracting the average of six spectra with low API levels from the average of six spectra with high API levels. The six low-level spectra had an average API of 41.29% with a standard deviation (S.D.) of 0.0318%, and the high-level spectra had an average API of 63.64% with an S.D. of 0.069%. Spectra from both instruments were included in the high and low value sets, and averaging the spectra reduced the effect of instrument noise to provide a better match to that of the API spectrum. The API spectrum and the extracted spectrum are compared in Fig. 2 after applying SG1104 conversion to both spectra. The spectral region from 900 to 1630 nm was chosen for this plot and for the regressions, because the API has many clear absorption bands in this region, and the instrument noise precludes the use of the data at both short and long wavelengths. The second-derivative conversion enhances the features of the spectra, and produces negative peaks

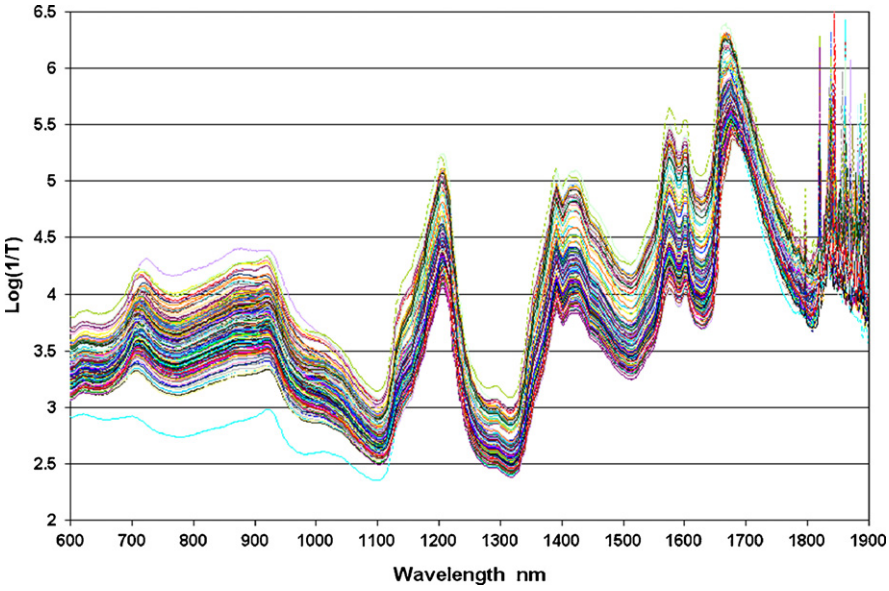

Fig. 1. Spectra of the total data set of 1230 spectra.

at the wavelengths of the absorption bands to more clearly show the degree of match. The positions of the negative peaks are the best evidence that the spectra match.

Fifteen significant absorption bands at the following wavelengths: 920, 1130, 1160, 1200, 1294, 1350, 1368, 1386, 1408, 1426, 1458, 1504, 1534, 1574, and 1602 nm can be clearly identified in the API derivative spectrum. The extracted spectrum duplicates these wavelengths except at 1130 nm with a shift of 2 nm at this peak. The actual difference is less than 2 nm, but with the data point spacing of 2 nm an interpolation is required to determine the true shift. A different set of spectra can be subtracted to obtain an extracted spectrum similar to the API. A test was done (not shown) using the average of spectra that differed in API by only 3.5% with very little change other than lower signal to noise.

The PLS regression was performed on the total data set of 1230 spectra with no pretreatment using the concentration values of the API to obtain a calibration for API.

The regression program indicated 32 possible outlier samples, and was not a particularly good calibration, because of the uncorrected scatter effects, and the presence of the outliers. However, the spectra of the first principal component, the API, and the extracted spectrum described above matched very well after converting each to the second-derivative format. The wavelengths (API and PC_1A) are listed in Table 1, and the spectra are shown in Fig. 3, with the vertical axis of the spectra scaled for clarity.

Please note that all 15 of the negative peaks of the API and the extracted spectrum match on the wavelength scale within 1 data point to that of the first principal component (PC_1). This provides assurance that this calibration is specific for the API.

This data set may be judged as an ideal set for this type of analysis, because it includes a wide range of API content, the API has many relatively narrow absorption bands, the other ingredients do not vary, and the instrument noise is not a major factor. On the other hand, the data do include two instruments, and the

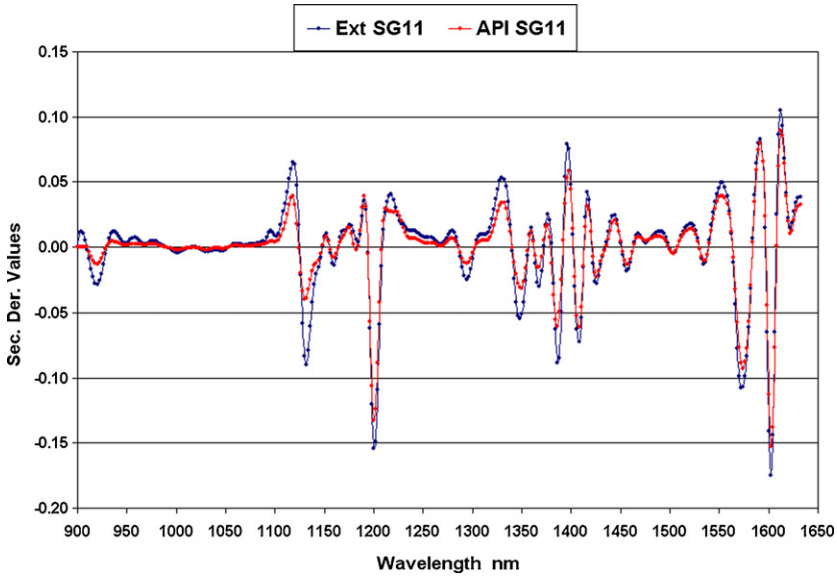

Fig. 2. Second-derivative spectra of the API and the spectrum extracted from the data set.

Table 1
Wavelengths (nm) of negative peaks for different spectra

API	Extracted	PC_1A	PC_1B	PC_1C	PC_1D	PC_1E	PC_1F	PC_1G
920	920	920	918	920	918	918	920	918
1130	1132	1132	1132	1132	1132	1132	1130	1130
1160	1160	1158	1158	1158	1158	1158	1158	1158
1200	1200	1200	1200	1200	1200	1200	1200	1200
1294	1294	1294	1294	1294	1294	1294	1294	1294
1350	1348	1350	1348	1348	1348	1348	1348	1348
1368	1368	1368	1368	1368	1368	1368	1368	1368
1386	1386	1386	1386	1386	1386	1386	1386	1386
1408	1408	1406	1406	1406	1406	1406	1406	1406
1426	1426	1426	1424	1424	1424	1424	1424	1424
1458	1456	1456	1456	1456	1456	1456	1456	1456
1504	1504	1504	1504	1504	1504	1504	1504	1504
1534	1534	1534	1534	1534	1534	1534	1534	1534
1574	1574	1574	1574	1574	1574	1574	1574	1574
1602	1602	1602	1602	1602	1602	1602	1602	1602

PC_1A 1230 spectra, no pre-treatment; PC_1B 1230 Spectra, MSC & SG1104; PC_1C 1198 spectra, omit 32 outliers, SG1104; PC_1D 1198 spectra, omit 32 outliers, MSC & SG11; PC_1E 1198 spectra, omit 32 outliers, EMSC & SG11; PC_1F 150 spectra, omit 5 outliers, MSC & SG11; PC_1G 150 spectra, omit 5 outliers, EMSC & SG11.

tablets are optically very dense, with absorbance values from 2.5 to 6.2 absorbance units. The random noise on the spectra in this data set varies from about 300 to 500 micro-absorbance units for the region from 900 to 1630 nm. These spectra show large effects from scattering particles, and the normal procedure is to apply MSC, EMSC, and/or derivative conversion before the PLS regression to minimize the scatter effects. Therefore, these options were explored to evaluate possible effects on the specificity test.

Regressions were done with the following options:

1. 1230 spectra, with no pretreatment (PC_1 A).
2. 1230 spectra, with MSC and SG1104 pretreatment (PC_1 B).
3. 1198 spectra, omitting 32 samples identified as outliers, with SG1104 (PC_1 C).
4. 4.1198 spectra, omitting 32 samples identified as outliers, with MSC and SG1104 pretreatment (PC_1 D).
5. 1198 spectra, omitting 32 samples identified as outliers, with EMSC and SG1104 pretreatment (PC_1 E).
6. 150 spectra from Calibrate1, omitting five samples identified as outliers, with MSC and SG1104 pretreatment (PC_1 F).
7. 150 spectra from Calibrate1, omitting five samples identified as outliers, with EMSC and SG1104 pretreatment (PC_1 G).

If the derivative conversion is included in the pretreatment, it is not necessary to convert the PC_1, because it reflects the derivative used in the PLS regression. Therefore the negative peaks of the PC_1's from the different treatments can be compared directly with the derivative conversion of the API. These data are compiled into Table 1, and the consistency is amazing. With the seven different data treatments no absorption peaks shifted more than 1 data point, which is not a significant shift, and this occurred in only 4 of the 15 peaks. The MSC and EMSC affect the magnitude of the different absorption bands in the PC_1's, but does not shift the wavelength location of the negative peaks.

Therefore, these wavelength peaks provide an excellent proof of specificity with each of these different treatments. The treatments which included a derivative as well as omission of the outliers provided good calibrations, with low prediction errors on the two validation sets. The calibration which included the use of the Savitsky–Golay, 11 point smoothing and 4th order polynomial derivative on 1198 spectra as the only additional treatment (set 3 above) provided an RMSEP of 1.64 mg, an SEP of 1.52 mg, and a bias of 0.60 mg for the 40 spectra in the Validate1 file using five factors.

The same calibration provided an RMSEP of 1.70 mg, an SEP of 1.70 mg, and a bias of 0.28 mg for the Validate2 file. Hopkins [6]

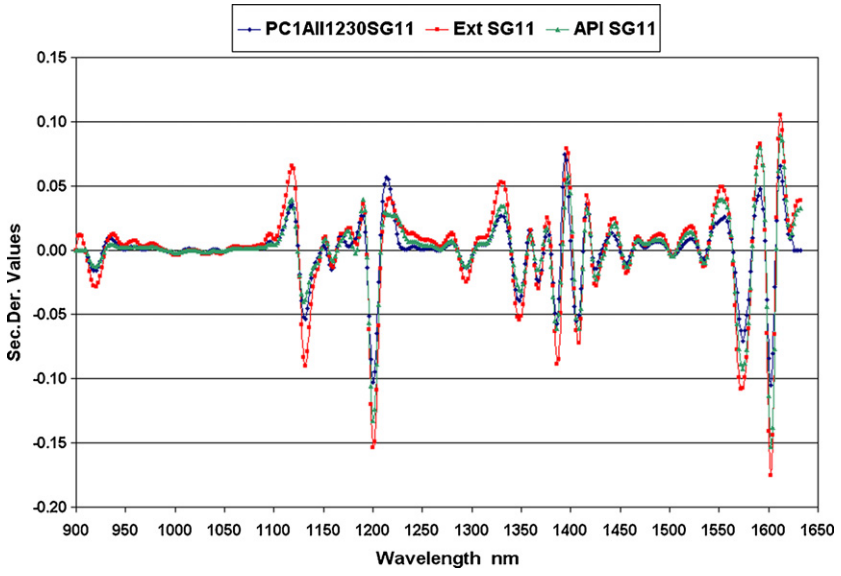

Fig. 3. Second-derivative spectra from the PC_1, API, and extracted spectrum.

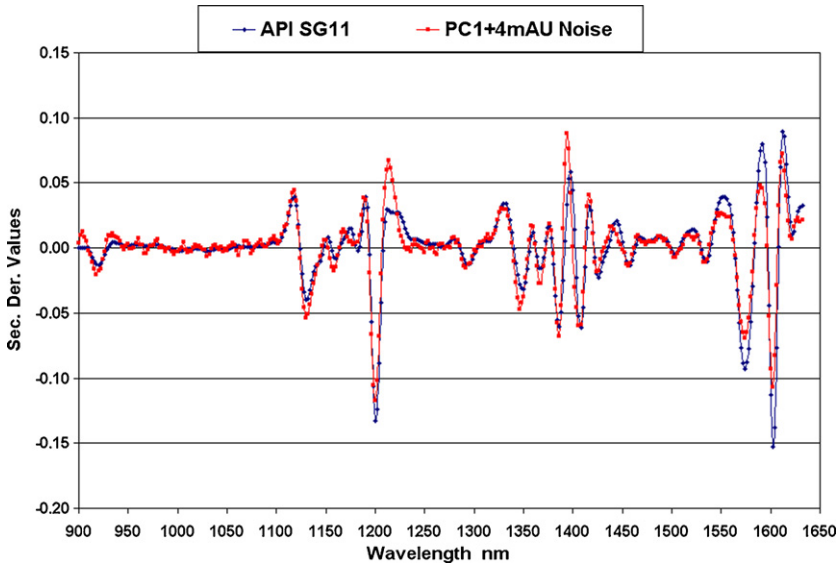

Fig. 4. Effect on PC_1 of adding random noise to the data set before doing the regression.

provided a calibration using 150 spectra from the Calibrate1 file. His prediction results were an RMSEP of 2.1 mg and a bias of 0.5 mg on the Validate1 file, and RMSEP of 1.8 mg with a bias of 0.43 mg on the Validate2 file using Savitsky–Golay, 11 point smoothing and 4th order polynomial and MSC with four factors.

Unscrambler 9.7 offers three different types of derivative calculations, so all were evaluated for possible effects on the specificity. Regressions were performed using the Gap-Segment option with a gap of 3 points and a segment of 1 point, and the Norris Gap derivative with a gap of 4 points. The same 15 peaks were found in the PC_1 spectra, with no peaks differing by more than 1 data point from the other spectra. The effect of random noise was also evaluated by adding different levels of random noise to each spectrum before doing the regression. Added noise of 1000, 2000, and 4000 micro-absorbance levels caused no peak shifts of more than 1 data point, although the noise was clearly evident on the PC_1 spectra. Fig. 4 shows the PC_1 spectrum with no added noise and with the highest added noise. These evaluations demonstrate the robustness of this method of determining specificity.

This procedure for specificity is best applied to transmission spectra of tablets and capsules containing only one API, and that ingredient must have distinguishable absorption spectra in the wavelength region of the measurement. It has not been tested on other pharmaceutical preparations, but the concept does have merit on other samples. The procedure has not been tested on other types of spectrometers, but if the instrument noise is low enough and the resolution is high enough with a stable wavelength scale it should perform.

4. Conclusions

A method of providing specificity for NIR calibrations on pharmaceutical tablets is described. The method compares the absorption bands of the first principal component of a PLS regression with the absorption bands of the API, using second-derivative treatments to clearly show the bands. In the case of the ideal sample used in this study, where the calibration covers a wide range of API content, the API has many relatively narrow absorption bands, the other ingredients do not vary, and the instrument noise is not a major factor, it is concluded that all of the bands in the API must match within 1 data point on the wavelength scale to guarantee specificity. This differs with other tests proposed for specificity [9], in that they require the use of active substances or excipients close to the API tested and also forced degraded samples. Such challenged samples should be rejected in the qualification step of the sample. It is hypothesized that this [qualification] step serves as an identity or outlier test for specificity of the calibration model. The algorithms used for these tests apply several approaches, such as spectral matching, and residual analysis to name a few, using Malahnobis or Euclidian distance calculations, for comparing the unknown sample spectrum to a previously known and calibrated sample set to try to determine their similarity in either wavelength space or principal component space [10]. These analogs or degraded samples should not exhibit specificity for the calibrated samples due to discrete wavelengths differences in their NIR ($\log 1/R$) absorption or derivative spectrum. We propose the novel use of the high–low extracted spectra and compare their difference to the first principal component spectrum of the model and the pure API spectrum for verifying specificity of the calibration model and for showing its suitability of its intended use for assaying API concentration.

Results are presented for a set of 1230 tablet spectra, with regressions done with many different pretreatments of the data and with addition of random noise up to 10 times the normal instrument noise to demonstrate the robustness of the procedure.

References

[1] United States Pharmacopeia, Supplement No. 2, 31st ed., United States Pharmacopeial Convention, Rockville, 2008, pp. 3735–3739.
[2] N. Broad, P. Graham, P. Hailey, A. Hardy, S. Holland, S. Hughes, D. Lee, K. Prebble, N. Salton, P. Warren, Guidelines for the Development and Validation of Near-infrared Spectroscopic Methods in the Pharmaceutical Industry, John Wiley & Sons, Chichester, West Sussex, UK, 2002, pp. 21–42.
[3] A.C. Olivieri, K.M. Faber, J. Ferré, R. Boqué3, J.H. Kalivas, H. Mark, Pure Appl. Chem. 78 (2006) 633–661.
[4] http://www.idrc-chambersburg.org/shootout2002.html.
[5] FOSS NIRSystems, Inc., 7703 Montpelier Road, Suite 1, Laurel, MD 20723.
[6] D.W. Hopkins, NIR News 14 (2003) 10–13.
[7] R. Marbach, J. Near Infrared Spectrosc. 13 (2005) 241–254.
[8] Unscrambler Software from CAMO, Inc., Oslo, Norway.
[9] CPMP/QWP/3309/01 (EMEA/CVMP/961/01) Note for Guidance on the Use of Near Infrared Spectroscopy by the Pharmaceutical Industry and the Data to be forwarded in the Part II of the Dossier for a Marketing Authorisation (Adopted by CPMP/CVMP February 2003).
[10] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, Chemometr. Intell. Lab. 50 (2000) 1–18.