Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy

Marcus H. Ferreira [a,b], Jez W.B. Braga [c,e], Marcelo M. Sena [a,d,e],*

[a] Mestrado em Ciências Moleculares, UnUCET, Universidade Estadual de Goiás, P.O. Box 459, 75001–970, Anápolis, GO, Brazil
[b] Indústria Química do Estado de Goiás SA, IQUEGO, 74450–0101, Goiânia, GO, Brazil
[c] Instituto de Química, Universidade de Brasília, 70904–970, Brasília, DF, Brazil
[d] Departamento de Química, ICEx, Universidade Federal de Minas Gerais, 31270–901, Belo Horizonte, MG, Brazil
[e] Instituto Nacional de Ciência e Tecnologia em Bioanalítica, 13083–970, Campinas, SP, Brazil

ARTICLE INFO

ABSTRACT

This work developed and validated a new multivariate diffuse reflectance near infrared method for direct determination of hydrochlorothiazide in powder pharmaceutical samples. The best partial least squares (PLS) model was obtained in the spectral region from 1640 to 1780 nm, with mean centered data preprocessed by first derivative and Savitzky–Golay smoothing followed by vector normalization. This model was built with 4 latent variables and provided a root mean square error of prediction of 1.7%. The method was validated according to the appropriate regulations in the range from 21.25 to 29.00 mg of hydrochlorothiazide per 150 mg of powder (average mass tablet), by the estimate of figures of merit, such as accuracy, precision, linearity, analytical sensitivity, capability of detection, bias and residual prediction deviation (RPD). The concept of net analyte signal (NAS) was used to estimate some figures of merit and to plot a pseudo-univariate calibration curve. The results for determinations in powdered manufactured tablets were in agreement with those of the official high performance liquid chromatographic method (HPLC). Finally, the method was extrapolated for determinations in intact tablets, providing prediction errors smaller than ±9%. The developed method presented the advantage of being about fifteen times faster than the reference HPLC method.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The combination of near infrared spectroscopy (NIRS) and multivariate calibration has emerged in the last decade as a promising alternative for the quality control of active pharmaceutical ingredients (API) [1–8], providing methods that are simple, rapid, non-destructive and of low cost. In addition, methods based on NIRS are environmentally friendly and solvent free, generate no chemical waste, may not require any sample pre-treatment, and provide sufficient accuracy and sensitivity with less human intervention. The development of these methods requires a robust calibration design that incorporates all the possible sources of variation, thus improving the quality control of the final product. However, in practice few quantitative NIRS methods are used for the determination of APIs in the quality control laboratories of the pharmaceutical industry. The main challenge is developing of NIRS methods that meet the stringent requirements of this highly regulated industry [3]. In the last years, Brazilian and US Pharmacopoeias [9,10]

have published general monographs about NIRS methods, but they have no monographs about multivariate methods for the quantification of specific APIs in pharmaceutical formulations. The most of the present regulation has been established based on chromatographic methods in a univariate way, such as the guidelines of ANVISA (National Health Surveillance Agency) [11], in Brazil, and ICH (International Conference on Harmonisation) [12,13]. Thus, these regulations should be harmonised in order to encompass the specific aspects of NIRS technology related to multivariate methods.

The main aspect of this harmonisation is the incorporation of the multivariate thinking in the traditional analytical validation. The requirement of total selectivity/specificity [11,12] should be eliminated for multivariate methods, since they are only useful when a selective variable/wavelength does not exist. The traditional calibration curves (signal as a function of analyte concentration) should also not be employed with multivariate methods. Some of these aspects have been discussed since about ten years ago [14], but the main concept that emerged in the last years is the net analyte signal (NAS). The NAS concept is an advance in the multivariate calibration theory that allows separating the information specific of the analyte from the whole signal, and can be used for estimating important figures of merit (FOM) in pharmaceutical applications. In addition, NAS

* Corresponding author at: Departamento de Química, ICEx, Universidade Federal de Minas Gerais, 31270–901, Belo Horizonte, MG, Brazil. Tel.: +55 31 34096389; fax: +55 31 34095700.
E-mail address: marcsen@ufmg.br (M.M. Sena).

values of each sample can be used to represent multivariate models as pseudo-univariate curves, an easier and simpler manner to interpret them in routine analyses. Thus, this work aimed to incorporate the state of the art in the multivariate validation [15] for the development of a NIRS method for quality control of a hydrochlorothiazide (HCTZ) formulation by using diffuse reflectance measurements and partial least squares (PLS). In the context of process analytical technology (PAT), this method was developed for quantification in powder samples, but it was also applied on intact tablets.

Moreover, this method development adopted a robust strategy, including an experimental design, variable selection, data preprocessing and outlier detection.

HCTZ, 6-chloro-3,4-dihydro-2H-1,2,4-benzothiadiazine-7-sulphonamide-1,2-dioxide [16], is a thiazide drug used for the treatment of hypertension, congestive heart failure, hepatic cirrhosis and some kidney diseases. It is a diuretic, which inhibits the ability of the kidney to retain water by increasing the excretion of sodium, chloride and, to a lesser extension, potassium ions. The official methods for HCTZ determination in pharmaceutical formulations are based on HPLC [9,10], but many other methods have been published, based on UV/visible spectrophotometry [17,18], chemiluminescence [19], capillary electrophoresis [20] and electroanalytical techniques [21,22]. Nevertheless, the majority of these methods have not been able to determine directly HCTZ in the presence of interferences, such as excipients, impurities and other active principles, demanding steps of separation and the use of reagents or solvents. The method proposed in this work provides a direct analysis of powder pharmaceutical samples and uses no solvent or reagent.

2. Multivariate analytical validation

2.1. NAS and pseudo-univariate calibration curves

The concept of NAS is useful in the development of NIRS methods. It can be defined as the part of the analytical signal uniquely related to the analyte or property of interest, which is orthogonal to the space of the interferences. This concept was first proposed by Lorber in 1986 for direct multivariate calibration methods (classical least squares) [23], but has found few applications until some years ago, when it was improved and extended to inverse multivariate calibration methods, such PLS and PCR (principal component regression) [24]. A NAS vector, $\hat{\mathbf{x}}_{A,i} nas$, is estimated for each sample i from the regression vector of a PLS/PCR model with A latent variables (LV), $\mathbf{b}$, according to Eq. (1).

$$\hat{\mathbf{x}}_{A,i}^{nas} = \mathbf{b}\left(\mathbf{b}^{\mathrm{T}}\mathbf{b}\right)^{-1}\mathbf{b}^{\mathrm{T}}\mathbf{x}_i \qquad (1)$$

The NAS vectors can be used for qualitative analysis and the norm of each one provides a scalar $nâs_i$, which is equivalent to a selective univariate analyte signal for each sample. A more complete theoretical description of NAS model can be found elsewhere [8,15,23,24].

The use of NAS is important in the analytical validation for estimating FOM, such as sensitivity, selectivity, and limits of detection and quantitation [15,25]. Moreover, it also allows presenting multivariate calibration models in a more interpretable way, through the pseudo-univariate calibration curves [26]. In the first step, $\hat{\mathbf{x}}_{A,i} nas$ vectors are calculated for the calibration samples, and then a regression coefficient, $\hat{b}_{nas}$, is estimated by a linear regression between a vector containing the scalar nâs values, $\mathbf{nâs}$, and the vector of the analyte concentrations ($\mathbf{y}$).

$$\hat{b}_{nas} = \left(\mathbf{nâs}^{\mathrm{T}}\mathbf{nâs}\right)^{-1}\mathbf{nâs}^{\mathrm{T}}\mathbf{y} \qquad (2)$$

In the last step, the regression model is expressed as:

$$\hat{\mathbf{y}} = \hat{b}_{nas}\mathbf{nâs} + \mathbf{e} \qquad (3)$$

where $\mathbf{e}$ is a vector containing the residuals of the model.

2.2. Figures of merit (FOM)

The traditional regulation [11–13] prescribes that the validation of analytical methods for API content determination requires the estimation of the following FOM: specificity/selectivity (SEL), linearity, accuracy, precision and range. Limit of detection is not required for this type of method, but only for the quantification of impurities. Aiming at a complete multivariate analytical validation, this work also estimated sensitivity (SEN), analytical sensitivity (γ), bias, capability of detection (CCβ) and residual prediction deviation (RPD). The SEL for multivariate methods is calculated as the ratio between the norm of the NAS vector and the norm of each spectrum. Since each sample presents a different SEL, an average value is used to characterized the method [15,25]. As mentioned in the Section 1, NIRS multivariate methods do not require signal resolution. Therefore, there is no sense in establishing a limiting SEL value. An alternative for evaluating specificity is the demonstration that the method is able to distinguish other components, such as impurities, degradation products or other active principles.

The precision and accuracy can be estimated in a similar manner for univariate and multivariate methods. In this work, precision was assessed at two levels, repeatability and intermediate precision. The accuracy can be evaluated through the relative prediction errors of the individual samples. However, the accuracy for multivariate methods uses to be also evaluated through parameters, such as RMSEC (root mean square error of calibration), RMSECV (root mean square error of cross validation) and RMSEP (root mean square error of prediction). Among these, RMSEP is the most robust parameter, because it is estimated from external validation samples. The linearity can be evaluated from the verification of the random behaviour of the fit residuals of a plot of reference *versus* predicted values. Once this random behaviour has been assured, the linearity can be expressed through the correlation coefficient (r) of this plot. Alternatively, the parameters of the pseudo-univariate calibration curve could be used to express linearity. However, it should be stressed that the value of r should be the same for both of these plots. The range is established by confirming that the method provides an acceptable degree of linearity, accuracy and precision when applied to samples in the range from 80 to 120% of the test content.

The SEN of multivariate methods is estimated as the NAS at unit concentration, which is equivalent to the following equation.

$$\mathrm{SEN} = 1/||\mathbf{b}|| \qquad (4)$$

where "|| ||" indicates the Euclidian norm of a vector. A more useful FOM is the analytical sensitivity (γ), which is defined, by analogy with univariate calibration [27], as the ratio between SEN and the instrumental noise (ε). The instrumental noise can be estimated through the pooled standard deviation of a vector containing fifteen replicate spectra of the blank [15].

$$\gamma = \mathrm{SEN}/\varepsilon \qquad (5)$$

The inverse of γ (γ^{-1}) provides an estimation of the minimum concentration difference that is discernible by the analytical method considering the random experimental noise as the only source of error, regardless of the specific technique employed.

The term bias evaluates the presence of systematic errors, and is defined as the difference between the limiting mean and the true value. It is calculated only from the validation set, according to ASTM [28]:

$$\boldsymbol{bias} = \sum_{i=1}^{n_v} \frac{\left(\boldsymbol{y^{ref}}_i - \hat{\boldsymbol{y}}_i\right)}{n_v} \qquad (6)$$

where $\mathbf{y}^{\mathrm{ref}}_i$ and $\hat{\mathbf{y}}$ are the analyte concentration values of reference and predicted by PLS model, respectively, and n_v is the number of samples in the validation set. Standard deviation of validation errors (SDV) is

also calculated and used in a *t* test to determine if the validation estimates show a statistically significant bias.

$$SDV = \sqrt{\frac{\sum\left[\left(\mathbf{y}_i^{ref} - \hat{\mathbf{y}}_i\right) - bias\right]^2}{n_v - 1}} \qquad (7)$$

The calculated t value (t = |bias| $\sqrt{n_v}$/ SDV) is then compared to the critical t value with the adequate number of degrees of freedom, which is equal to n_v.

The concept of CCβ, present in the ISO 11843–2 [29], has been extended for multivariate calibration by Ortiz et al. [30], according to Eq. (8), considering a calibration set without replicated samples.

$$\boldsymbol{CC\beta} = \delta_{\alpha,\beta,\nu} \frac{\mathbf{s(e)}}{\mathbf{b_{cal}}} \sqrt{\frac{1}{\mathbf{n_c}} + 1 + \frac{\mathbf{y}^2}{\sum_{\mathbf{i=1}}^{\mathbf{n_c}} (\mathbf{y_i} - \mathbf{y})^2}} \qquad (8)$$

In this equation s(e) and b_{cal} are, respectively, the standard deviation of the residuals and the slope of the reference *versus* predicted concentration values curve, n_c is the number of samples in the calibration set, and $\delta_{\alpha,\beta,\nu}$ is the parameter of the non-central t-distribution with probabilities α and β and the number of degrees of freedom (ν) equal to n_c-2; in this work, δ = 3.328 was used, corresponding to α and β probabilities of 0.05 (95% confidence level) [31].

The residual prediction deviation or relative predictive determinant (RPD) [32,33] is the ratio of natural variation in the calibration or validation samples to the size of probable errors occurring during the prediction. The RPD represents how well the calibration model predicts a specific set and it is more appropriate for evaluating the performance of a model in absolute terms. Specifically, the RPD is calculated for the calibration and validation sets according to the following equations:

$$RPD_{cal} = SD_{cal}/RMSECV \qquad (9)$$

$$RPD_{val} = SD_{val}/RMSEP \qquad (10)$$

where SD is the standard deviation of the reference values for the calibration (SD_{cal}) and validation (SD_{val}) sets.

2.3. Detection of outliers

The detection of outliers is fundamental for developing good NIRS models. This work employed a robust procedure for outlier detection based on the identification of samples with extreme leverages, large residuals in the spectral data or large residuals in the analytical concentration values [28,34]. The leverage, h_i, is a measure of the influence of each sample on the model. Samples with h_i larger than a limit value (three times the number of LVs plus one, divided by the number of calibration samples) should be deleted. Outliers with high spectral data residuals are detected by comparison of the total standard deviation, s(e), with the standard deviation of each sample, s(e_i), related to the values of absorbance or log(1/R) measured and predicted with A LVs [8,34]. If a sample has s(e_i) > 2 s(e), it should be deleted, at about 95% confidence level. Outliers with high residuals in the concentration values can be detected by comparing the RMSEC of the model with absolute errors of individual samples. If a sample has a difference between its reference value and its estimate larger than three times the RMSEC, it should be deleted.

3. Materials and methods

3.1. Apparatus and software

Spectra were recorded on a Foss NIRSystems 4500 Smart Probe Analyzer spectrophotometer (Silver Spring, USA), equipped with a diffuse reflectance accessory. The equipment was controlled and data were acquired using the Foss Vision 3.3.0.0 software package. Data were handled using MATLAB software, version 7.13 (The Math-Works, Natick, USA). The PLS routine came from PLS Toolbox, version 6.5 (Eigenvector Technologies, Manson, USA), and a homemade routine was also used for the detection of outliers. This last routine is compatible with both Matlab and Octave, and can be found in the supplementary material.

3.2. Reagents and samples

The target pharmaceutical formulation has 25 mg of HCTZ and the following excipient composition: microcrystalline cellulose, sodium croscarmellose, colloidal silicon dioxide, and magnesium stearate. All the chemical reagents were of analytical grade, purchased from certified suppliers and used without further purification. Powder samples were prepared by weighing with an analytical balance (±0.0001 g), according to an experimental design.

3.3. Methodology

3.3.1. Experimental design

Ninety three powder samples were prepared according to an experimental design with three factors: HCTZ, cellulose, which is the main excipient, and other excipients (croscarmellose, silicon dioxide, and stearate), as shown in Fig. 1. The range of HCTZ content was varied from 21.25 to 29.00 mg per 150 mg of powder (the average tablet mass) in approximately equally spaced intervals, corresponding from 85.0 to 116.0% of the target content (25 mg). This range was chosen in order to cover from 90.0 to 110.0% (from 22.50 to 27.50 mg per tablet) of the HCTZ target content, which are the acceptable limits established by the Brazilian Pharmacopoeia [10] for this API content in this type of formulation. The mass of 25 mg corresponds to about 17% w/w of each produced HCTZ tablet. The other two factors were also varied according to the experimental design depicted in Fig. 1. In order to reduce the errors in the samples preparation, the weighted total mass of each sample was fixed in 30.0000 g.

3.3.2. Procedure

The powder samples were prepared in amber glass flasks, manually homogenized and directly measured. The spectra were recorded from 1100 to 2500 nm (step 2 nm) as the average of 32 scans. During all the measurements, the laboratory temperature and the relative air humidity were controlled at 22–28 °C and 35–65%, respectively. Triplicates of samples at three HCTZ content levels (90.4, 100.3 and
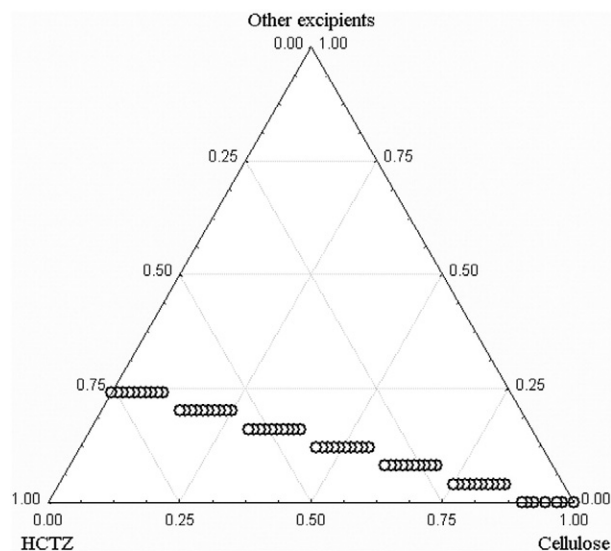

Fig. 1. Experimental design used in this work.

110.2%) were also obtained for evaluating repeatability. These replicates were compared with other triplicates obtained at these same three levels on another day by a different analyst for estimating intermediate precision. Fifteen spectra of the empty cell were recorded for estimating the instrumental noise.

3.3.3. Analysis of real samples

The developed method was applied to real samples in three different ways. Firstly, triplicates of powdered tablet samples from three different batches were analyzed. Each sample was prepared by finely powdering twenty tablets in an agate mortar with a pestle. Secondly, ten intact tablets from the same batch were randomly sampled and directly analyzed with the diffuse reflectance accessory. All these previous results were compared with the reference HPLC method. Finally, triplicates of powder samples under control were collected in the production line after the mixture step and before the compression to tablets. These samples were obtained from six different batches, during a period of three months, and analyzed.

3.4. Chromatographic analysis

The analysis with HPLC was based on the official method [9] and was carried out with a Perkin Elmer liquid chromatograph, series 200, with UV/Vis detection. An analytical C-18 column (250×4.6 mm, 5 µm) was used. The mobile phase was water/acetonitrile (9:1, v/v), adjusted with phosphate buffer at pH 3.0 ± 0.1. A flow rate of 2.0 mL min^{-1} and detection at 254 nm were used. The powder samples were dissolved in acetonitrile, sonicated for 15 min, and passed through a 0.45 µm membrane filter. All the injections were repeated three times.

4. Results and discussion

4.1. Pure HCTZ spectrum

A diffuse reflectance spectrum of pure HCTZ powder is shown in Fig. 2. A qualitative analysis of this figure helps in highlighting the spectral regions that may most contribute for predictive models. The region below 1400 nm, associated with second overtone vibrations, do not present intense bands. First overtones of N–H stretchings were observed in the region between 1480 and 1620 nm, where the bands centered at 1518, 1552 and 1590 nm can be attributed to the three N–H bonds of the amino groups of HCTZ. First overtones of C–H stretchings were observed in the region from 1640 to 1780 nm, where the bands centered at 1664 and 1738 nm can be attributed to the aromatic C–H and aliphatic C–H (CH$_2$ of the non-aromatic ring) bonds, respectively. The most intense band at 2034 nm is associated with the primary amine N–H combination of stretching and bending, and another
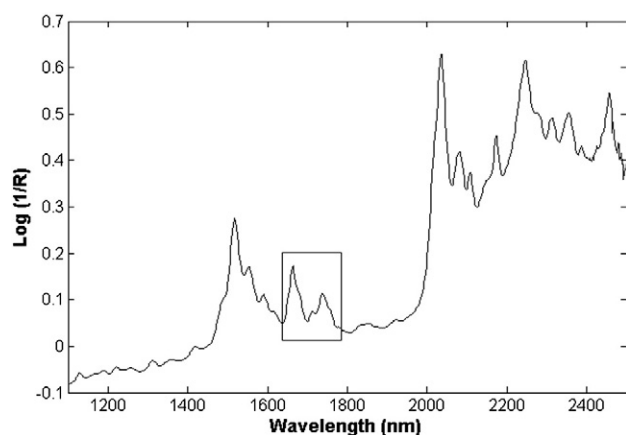
combination N–H band is present at 2173 nm. Combination bands of carbon bonds are observed above 2200 nm, with the intense band at 2246 nm attributed to the combination of stretchings of =CH and C=C [35].

4.2. PLS model

The spectra of all 93 prepared samples are shown in Fig. 3. They were divided into 56 for the calibration set and 37 for the validation set. The calibration samples were selected in order to ensure the representative and homogenous distribution of them in the HCTZ content range, according to the experimental design (Fig. 1). The very noisy spectral region above 2400 nm (Fig. 3) was deleted. NIR diffuse reflectance spectra of powder samples use to present non-linear baseline deviations, due to the multiplicative light scattering caused by the lack of homogeneity of the particle size distribution. These instrumental deviations are not related to the chemical sample composition and require the use of preprocessing methods. In this work, the most common preprocessing techniques used for NIR spectra [36] were tested: multiplicative scatter correction (MSC), standard normal variate (SNV), and first derivative with Savitsky-Golay smoothing followed by vector normalization. The number of LVs of the PLS models was chosen by contiguous block (with 7 splits) cross validation and the combination of first derivative, smoothing (11 points in filter and second order polynomial fit) and vector normalization provided the best model, with 5 LVs and a RMSECV of 3.0%. Models using MSC and SNV provided RMSEC values a bit higher with more LVs, seven in both cases. The chosen preprocessing aimed at compensating any change in experimental conditions and increasing the signal to noise ratio, stabilizing the model.

The results were optimized by testing local models on different regions, which were chosen based on the qualitative analysis of the spectrum of pure HCTZ, discussed on Section 4.1. According to this discussion, four spectral regions were tested, 1100–1400 nm (A), 1480–1620 nm (B), 1640–1780 nm (C) and 2000–2400 nm (D). Local PLS models were compared with the PLS model using the whole spectra and the best results were obtained for region C: 4 LVs and a RMSECV of 2.6%. Other spectral regions provided similar RMSECV values and higher number of LVs, such as 7 (B) and 8 (D), which is an indication of overfitting. Thus, the best PLS model was built with mean centered spectra preprocessed by first derivative/Savitsky-Golay smoothing/vector normalization in the region from 1640 to 1780 nm (Fig. 4), which is associated to the first overtones of the C-H stretchings of the analyte. This model accounted for 99.68% of the total variance in the X block and 97.31% in the Y block. Considering the chosen number of LVs (4), it was assured that the minimum necessary number of calibration and validation samples was used, in accordance with ASTM guidelines
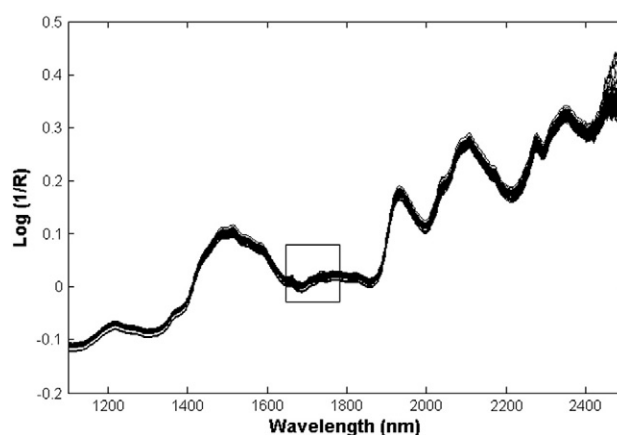

Fig. 2. NIR diffuse reflectance spectrum of a pure HCTZ sample. The spectral region selected for PLS model is marked in a rectangle.

Fig. 3. Spectra of 93 HCTZ samples, corresponding to the calibration and validation sets. The spectral region selected for PLS model is marked in a rectangle.
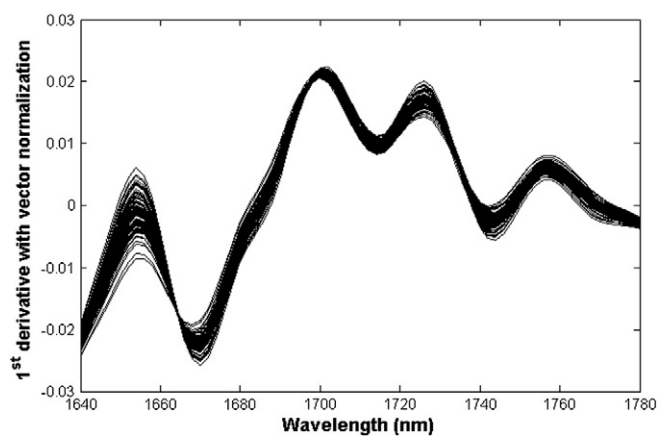
Fig. 4. First derivative normalized spectra in the range from 1640 to 1780 nm.
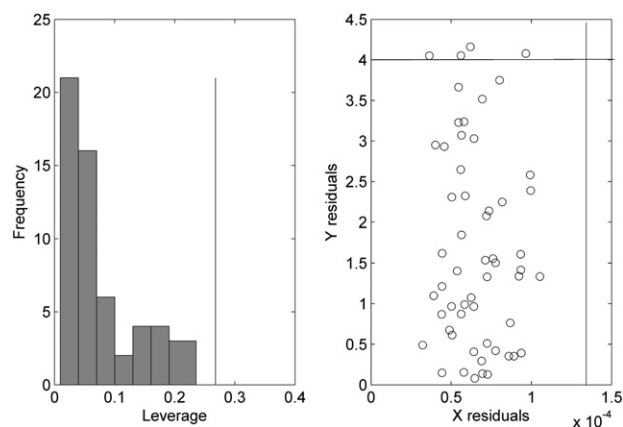

Fig. 5. Visualization of the outlier detection in the calibration set (first model). (a) Histogram of leverage values. (b) Plot of spectral residuals (X) *versus* concentration residuals (Y). The solid lines indicate the acceptance limits for outlier detection.

[28], which prescribes 6x(number of LVs + 1) and 4x(number of LVs) samples for the calibration and validation sets, respectively.

4.3. Detection of outliers

Following the selection of the spectral region, the PLS model was optimized using a procedure for detecting outliers based on the methodology described in Section 2.3. In this optimization, seven outliers were detected in the calibration set, and other seven in the validation set. The whole procedure is detailed in Table 1, which shows the changes of the RMSEC and RMSEP values. All of these models were built with four LVs. Fig. 5 represents the detection of outliers for the first model through the histogram of leverages and the plot of spectral *versus* concentration residuals, including the acceptance limits. It can be noted that for this first model four outliers were detected, based on their large residuals in the concentration values (Y). Considering the whole process (Table 1), ten more outliers (three in the calibration set and seven in the validation set) were detected, all of them based on their high residuals in Y. Therefore, the optimized PLS model used 49 calibration and 30 validation samples.

4.4. Analytical validation

Table 2 summarizes the parameters estimated for evaluating the main FOM of the developed method. The average accuracy can be evaluated through the parameters RMSECV, RMSEC and, mainly, RMSEP. The obtained RMSEP of 1.7% is below ±2%, the acceptable limits of accuracy commonly adopted for analytical methods in the pharmaceutical industry [37]. The relative prediction errors for the validation samples were between −2.8 and +2.7%. The precision was evaluated through the relative standard deviation (RSD) obtained for triplicates at three levels of HCTZ content. RSD values of 1.7 and 2.2% were obtained for repeatability and intermediate precision, respectively. Both results are in accordance with the Brazilian regulations [11], which prescribes a maximum RSD of 5%.

The linearity of the method was evaluated by the residuals of the PLS model, which are shown in Fig. 6. It is possible to verify the absence of systematic trends in the residuals distribution, evidencing their random behaviour. Once the residual plot indicates that the linear model is a valid assumption, the fitting of a straight line to the reference *versus* predicted values can be used to estimate a correlation coefficient that can express the average agreement between the estimated and reference values. Nevertheless, this correlation coefficient cannot be assumed as a quantitative measurement of the linearity. The results of this fit for the calibration samples are presented in Table 2, and the obtained value, 0.9865, is in accordance to the Brazilian regulations [11]. Considering the linearity, precision and accuracy studies, the range of the method was established from 85.0 to 116.0%, corresponding from 22.5 to 27.5 mg of HCTZ per 150.0 mg of powder (one tablet).

The estimation of the method SEL was only possible with the application of the NAS concept, providing an estimative of the amount of the instrumental signal that was used by the calibration model for the determination of the analyte. Although SEL is an essential parameter for determinations with HPLC, in multivariate NIRS models it has no practical meaning for quality control, since low values of SEL can be obtained even with accurate results. The average SEL estimated for the method corresponds to 33%, indicating a reasonable overlapping of the interferences (excipients) in the HCTZ signal.

Table 2
Parameters estimated for evaluating the main FOM of the developed method.

Figures of merit	Parameter	Value
Accuracy	RMSECV	1.6%
	RMSEC	1.4%
	RMSEP	1.7%
Precision	RSD repeatability	1.7%
	RSD intermediate precision	2.2%
Linearity	Slope	1.02 ± 0.04[a]
	Intercept	-2.01 ± 3.6[a]
	Correlation coefficient	0.9865[a]
Range		85.0 to 116.0%
Selectivity		0.33
Sensitivity		3.3×10^{-4}[b]
Analytical Sensitivity (γ)		0.9%$^{-1}$
γ^{-1}		1.1%
CCβ		12.3%
Bias		-0.267 ± 1.729
RPD	RPD calibration	5.6
	RPD validation	5.1

[a] Values for the line fitted to the calibration samples.
[b] Value expressed as the ratio between log (1/R) and %.

Table 1
Results for the optimization of the PLS model by detection of outliers.

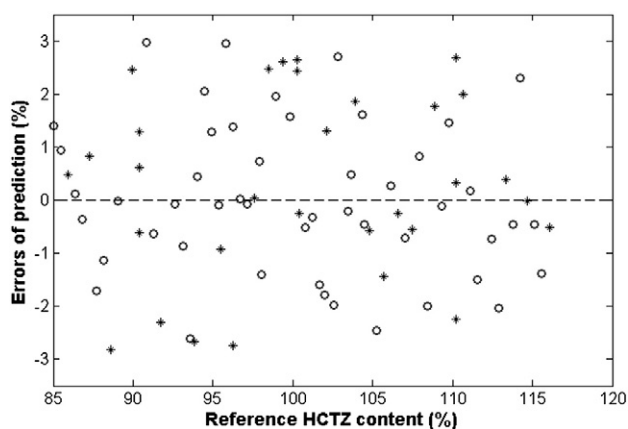| Model	1st	2nd	3rd	final
Number of calibration samples	56	52	49	49
Number of validation samples	37	37	37	30
RMSEC (%)	2.0	1.9	1.4	1.4
RMSEP (%)	2.5	2.3	2.3	1.7

Fig. 6. PLS residuals for the calibration (circles) and validation (asterisks) samples.

The SEN of the method was estimated as 0.00033. This value is not appropriate for comparison with other methods, since it depends on the analytical technique employed and the analysed matrix. Thus, γ was also calculated as $0.9\%^{-1}$, based on the estimated instrumental noise, 0.00038. The inverse of γ, 1.1%, indicates the minimum concentration difference that the method can discern considering the random instrumental noise as the only source of error. The CCβ was estimated as 12.3%, indicating that the method is not able to estimate contents below this value. However, this method is adequate for its purpose, since the lower limit of the analytical range is about seven times higher than this CCβ. The bias was estimated as -0.267 ± 1.729 and a t test with 30 degrees of freedom and at 95% confidence level demonstrated that there are no systematic errors in the model ($t_{calc} = 0.860 < t_{crit} = 2.042$). The estimates of RPD for calibration and validation were 5.6 and 5.1, respectively. These results are satisfactory, since models with RPD values larger than 5 are considered good for quality control [33].

Finally, a pseudo-univariate calibration curve was obtained (Fig. 7), demonstrating a simple manner to express this NIRS multivariate model. This representation is particularly useful in routine analysis, where the analyst can visualize the PLS model in a univariate way. The values of norm of NAS are equivalent to selective signals obtained from each sample spectrum vector. The fit of this curve is presented in the following equation:

$$[HCTZ] = 3055.4 \, ||\mathbf{n\hat{a}}s|| + 2.6943 \qquad (11)$$
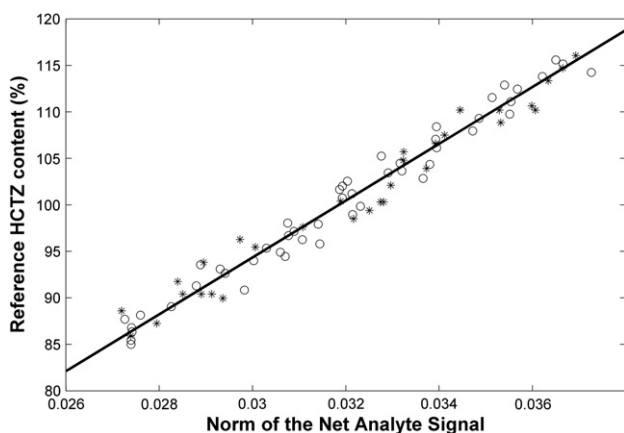

Fig. 7. Pseudo-univariate calibration curve. Plot of the NAS norms *versus* reference values for the calibration (circles) and validation (asterisks) samples.

Table 3
Determination of HCTZ in powdered tablets by the proposed NIRS method and the official HPLC method.

Batch	Official method (%)[a]	Proposed method (%)[a]
#A	97.7 ± 2.3	99.0 ± 0.4
#B	99.3 ± 1.3	96.8 ± 1.6
#C	100.4 ± 2.3	98.9 ± 2.5

[a] Mean values and standard deviations of three determinations.

4.5. Analysis of real samples

Firstly, the NIRS method was applied to the determination of HCTZ in powder samples obtained from manufactured tablets of three different production batches. The results were compared with the reference HPLC method (Table 3). Non paired Student's t tests with four degrees of freedom showed that there are no significant differences between the methods for all the three batches, at 95% confidence level (estimated t values below the critical t value, 2.776).

Although the NIRS model was developed for determining HCTZ in powder samples, it was also tested on ten intact tablets from a single batch. These analyses were carried out without replicates and the results were compared with the HPLC method (Table 4). A paired t test with 10 degrees of freedom showed that there are no significant differences between these results at 95% confidence level ($t_{calc} = 2.094 < t_{crit} = 2.228$). The individual prediction errors varied between -6.1 and $+8.6\%$, and a RMSEP of 4.6% was calculated. These results are worse than those obtained for powder samples, but taking into account that they were obtained for samples in another physical form, they were considered reasonable.

Finally, the NIRS method was applied to powder samples obtained before the compression to tablets. These samples were considered under control, what means HCTZ contents between 95.0 and 105.0% (23.75 and 26.25 mg per 150 mg of powder), the limits adopted for this specific release in the production quality control. This control is carried out by HPLC, but we did not have access to these results for comparison purposes. Thus, the developed method was applied to triplicates of samples collected from six different batches during a period of three months. The results were plotted in way similar to a control chart (Fig. 8) and, as can be observed, all are considered in agreement with the expected values.

5. Conclusions

This paper developed and validated a diffuse reflectance NIRS method for direct determination of HCTZ in powder pharmaceutical formulations. This method has many advantages over the official HPLC and other alternative methods, such as low cost, simplified procedure, no need for reagents or solvents and no generation of chemical waste. However, its main advantage is the rapidity of the analysis. The time interval estimated for the whole chromatographic

Table 4
Determination of HCTZ in ten intact tablets by the proposed NIRS method and the official HPLC method.

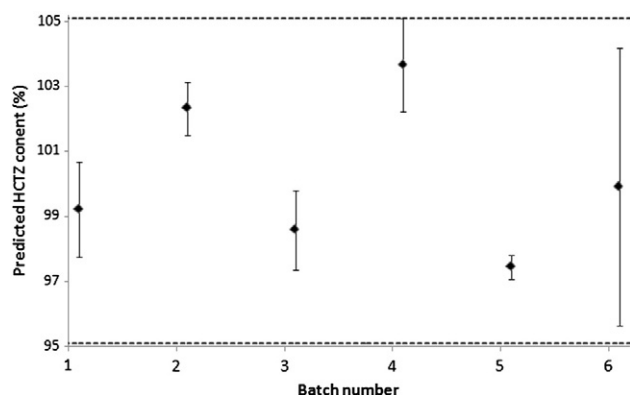| Tablet	Official method (%)	Proposed method (%)	Prediction errors (%)
#1	95.1	103.6	8.6
#2	98.0	104.2	6.1
#3	99.2	103.5	4.3
#4	100.1	94.0	-6.1
#5	100.3	101.2	0.8
#6	98.2	98.0	-0.1
#7	99.9	105.2	5.3
#8	95.0	97.6	2.6
#9	100.8	103.1	2.3
#10	99.3	102.0	2.7

Fig. 8. Results (mean and standard deviation of triplicates) for the determinations of HCTZ content in powder formulations collected from six batches before the compression. The dashed lines indicate the acceptance limits for release.

analysis of real samples (accounting tablet powdering, solvents extractions, filtering, chromatographic runs and data treatment) was 88 min, compared with this same estimative for NIRS analysis (accounting tablet powdering, average spectra acquisition, data transfer and chemometric calculations), which was 6 min. Thus, the NIRS method was considered 15 times faster than the HPLC one.

The spectral region between 1640 and 1780 nm provided the best PLS model, corresponding to the first overtone of the stretchings of C–H bonds of HCTZ. The NIRS method was validated in accordance with Brazilian and international guidelines [11–13,28] and was considered accurate, precise, linear and sensitive in the range from 85.0 to 116.0% of the target HCTZ content per tablet (25 mg). This model was successfully applied to powdered manufactured tablets and to powder samples collected from the production line before compression. In addition, it was also extrapolated for determination in intact tablets, providing prediction errors that were considered reasonable, despite of their higher values in relation to those obtained for powder samples. Finally, this paper presented a robust multivariate methodology, which employed the concept of NAS, and may be applied in the development and validation of other analytical methods based on NIRS and multivariate calibration.

Appendix A. Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.microc.2012.03.008.

References

[1] P. Chalus, Y. Roggo, S. Walter, M. Ulmschneider, Near-infrared determination of active substance content in intact low-dosage tablets, Talanta 66 (2005) 1294–1302.
[2] Z. Wang, B. Xiang, Application of artificial neural network to determination of active principle ingredient in pharmaceutical quality control based on near infrared spectroscopy, Microchem. J. 89 (2008) 52–57.
[3] D. Xiang, M. Konigsberger, B. Wabuyele, K. Hornung, J. Cheney, Development of robust quantitative methods by near-infrared spectroscopy for rapid pharmaceutical determination of content uniformity in complex tablet matrix, Analyst 134 (2009) 1405–1415.
[4] M.C. Sarraguça, J.A. Lopes, The use of net analyte signal (NAS) in near infrared spectroscopy pharmaceutical applications: Interpretability and figures of merit, Anal. Chim. Acta 642 (2009) 179–185.
[5] M.H. Ferreira, J.F.F. Gomes, M.M. Sena, Development and validation of a multivariate calibration model for determination of dipyrone in oral solutions by near infrared spectroscopy, J. Braz. Chem. Soc. 20 (2009) 1680–1686.
[6] W.F.C. Rocha, A.L. Rosa, J.A. Martins, R.J. Poppi, Determination and validation of nimesulide in pharmaceutical formulation by near infrared spectroscopy, J. Braz. Chem. Soc. 21 (2010) 1929–1936.
[7] M.C. Sarraguça, S.O. Soares, J.A. Lopes, A near-infrared spectroscopy method to determine aminoglycosides in pharmaceutical formulations, Vib. Spectrosc. 56 (2011) 184–192.
[8] M.A.M. Silva, M.H. Ferreira, J.W.B. Braga, M.M. Sena, Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy, Talanta 89 (2012) 342–351.
[9] The United States Pharmacopoeia, 33rd rev., U.S.P. Convention, Rockville, , 2011.
[10] Farmacopéia Brasileira, 5ª ed Ateneu Editora, São Paulo, Brazil, 2010.
[11] Agência Nacional de Vigilância Sanitária (ANVISA), Guia para Validação de Métodos Analíticos e Bioanalíticos, Resolução – RE nº 899, Brazil, , 2003.
[12] International Conference on Harmonization, Tripartite Guideline – Q2A Text on Validation of Analytical Procedures, Fed. Regist., 60 FR 11260, USA, 1995.
[13] International Conference on Harmonization, Tripartite Guideline – Q2B Validation of Analytical Procedures: Methodology, Fed. Regist., 62 FR 27464, USA, 1997.
[14] A.C. Moffat, A.D. Trafford, R.D. Jee, P. Graham, Meeting the International Conference on Harmonisation's Guidelines on Validation of Analytical Procedures: Quantification as exemplified by a near-infrared reflectance assay of paracetamol in intact tablets, Analyst 125 (2000) 1341–1351.
[15] P. Valderrama, J.W.B. Braga, R.J. Poppi, State of the art of figures of merit in multivariate calibration, Quim. Nova 32 (2009) 1278–1287.
[16] H.P. Deppeler, Hydrochlorothiazide, in: K. Florey (Ed.), Analytical Profiles of Drug Substances, 10, 1981, pp. 405–441.
[17] E. Satana, S. Altinay, N.G. Goger, S.A. Ozkan, Z. Senturk, Simultaneous determination of valsartan and hydrochlorothiazide in tablets by first-derivative ultraviolet spectrophotometry and LC, J. Pharm. Biomed. Anal. 25 (2001) 1009–1013.
[18] M.A. Gotardo, L. Pezza, H.R. Pezza, Determination of hydrochlorothiazide in pharmaceutical formulations by diffuse reflectance spectroscopy, Eclet. Quim. 30 (2005) 17–24.
[19] J. Xi, X. Ji, S. Zhang, X. Ai, Z. He, Investigation of RuBPS-Ce(IV) chemiluminescence reaction and its application in determination of two diuretics, Anal. Chim. Acta 541 (2005) 191–196.
[20] M.G. Quaglia, E. Donati, G. Carlucci, P. Mazzeo, S. Fanali, Determination of losartan and hydrochlorothiazide in tablets by CE and CEC, J. Pharm. Biomed. Anal. 29 (2002) 981–987.
[21] B. Rezaei, S. Damiri, Multiwalled carbon nanotubes modified electrode as a sensor for adsorptive stripping voltammetric determination of hydrochlorothiazide, IEEE Sens. J. 8 (2008) 1523–1529.
[22] H. Karimi-Maleh, A.A. Ensafi, H.R. Ensafi, Ferrocenedicarboxylic acid modified carbon paste electrode: A sensor for electrocatalytic determination of hydrochlorothiazide, J. Braz. Chem. Soc. 20 (2009) 880–887.
[23] A. Lorber, Error propagation and figures of merit for quantification by solving matrix equations, Anal. Chem. 58 (1986) 1167–1172.
[24] R. Bro, C.M. Andersen, Theory of net analyte signal vectors in inverse regression, J. Chemom. 17 (2003) 646–652.
[25] A.C. Olivieri, N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas, H. Mark, Uncertainty estimation and figures of merit for multivariate calibration, Pure Appl. Chem. 78 (2006) 633–661.
[26] N.M. Faber, Exact presentation of multivariate calibration model as univariate calibration graph, Chemom. Intell. Lab. Syst. 50 (2000) 107–114.
[27] L.C. Rodriguez, A.M.G. Campaña, C. Jimenez-Linares, M. Román-Ceba, Estimation of performance characteristics of an analytical method using the data set of the calibration experiment, Anal. Lett. 26 (1993) 1243–1258.
[28] Annual Book of ASTM Standards, Standard Practices for Infrared Multivariate Quantitative Analysis - E1655-05, ASTM International, West Conshohocken, Pennsylvania, USA, 2005.
[29] ISO 11843–2, Capability of detection, International Standards Organization, Geneva, Switzerland, 2000.
[30] M.C. Ortiz, L.A. Sarabia, A. Herrero, M.S. Sánchez, M.B. Sanz, M.E. Rueda, D. Giménez, M.E. Meléndez, Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares, Chemom. Intell. Lab. Syst. 69 (2003) 21–33.
[31] H.S.P. Carneiro, A.R.B. Medeiros, F.C.C. Oliveira, G.H.M. Aguiar, J.C. Rubim, P.A.Z. Suarez, Determination of ethanol fuel adulteration by methanol using partial least-squares models based on Fourier transform techniques, Energy Fuel 22 (2008) 2767–2770.
[32] M. Forouzangohar, R.S. Kookana, S.T. Forrester, R.J. Smernik, D.J. Chittleborough, Midinfrared spectroscopy and chemometrics to predict diuron sorption coefficients in soils, Environ. Sci. Technol. 42 (2008) 3283–3288.
[33] B. Igne, C.R. Hurburgh Jr., Local chemometrics for samples and variables: optimizing calibration and standardization processes, J. Chemom. 24 (2010) 75–86.
[34] P. Valderrama, J.W.B. Braga, R.J. Poppi, Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy, J. Agric. Food Chem. 55 (2007) 8331–8338.
[35] J.J. Workman Jr., L. Weyer, Practical Guide to Interpretive Near-Infrared Spectroscopy, CRC – Press, Boca Raton, USA, 2008.
[36] A. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC, Trends Anal. Chem. 28 (2009) 1201–1222.
[37] L. Huber, Validation and Qualification in Analytical Laboratories, 2nd ed. LabCompliance, New York, USA, 2007.