Talanta 85 (2011) 2218–2225

Contents lists available at ScienceDirect

Talanta

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

NIR analysis of pharmaceutical samples without reference data: Improving the calibration

M. Blanco, R. Cueva-Mestanza, A. Peguero

Department of Chemistry, Faculty of Sciences, Universitat Autònoma de Barcelona, E-01893 Bellaterra, Barcelona, Spain

ARTICLE INFO

Article history:
Received 14 April 2011
Received in revised form 13 July 2011
Accepted 23 July 2011
Available online 29 July 2011

Keywords:
Calibration set
Process variability matrix
Near infrared spectroscopy
Calibration without reference data
PLS calibration

ABSTRACT

Using an appropriate set of samples to construct the calibration set is crucial with a view to ensuring accurate multivariate calibration of NIR spectroscopic data. In this work, we developed and optimized a new methodology for incorporating physical variability in pharmaceutical production based on the NIR spectrum for the process. Such a spectrum contains the spectral changes caused by each treatment applied to the component mixture during the production process. The proposed methodology involves adding a set of process spectra (*viz.* difference spectra between those for production tablets and a laboratory mixture of identical nominal composition) to the set of laboratory samples, which span the wanted concentration range, in order to construct a calibration set incorporating all physical changes undergone by the samples in each step of the production process. The best calibration model among those tested was selected by establishing the influence of spectral pretreatments used to obtain the process spectrum and construct the calibration models, and also by determining the multiplying factor *m* to be applied to the process spectra in order to ensure incorporation of all variability sources into the calibration model. The specific samples to be included in the calibration set were selected by principal component analysis (PCA). To this end, the new methodology for constructing calibration sets for determining the Active Principle Ingredients (API) and excipients was applied to Irbesartan tablets and validated by application to the API and excipients of paracetamol tablets. The proposed methodology provides simple, robust calibration models for determining the different components of a pharmaceutical formulation.

1. Introduction

Near infrared spectroscopy has become a widely used tool in the pharmaceutical industry for a variety of purposes ranging from the determination of the API in end-products to the monitoring of the different steps of a pharmaceutical process. During the pharmaceutical process, a mixture of the different components of the formulation is subjected to operations including granulation, drying, pressing and coating, all of which introduce some variability in the spectra for production samples in relation to those for a sample obtained simply by mixing the formulation ingredients. Such variability precludes the use of the same model to determine any parameter such as the API concentration, an excipient concentration or some physical property (*e.g.* particle size, compaction pressure) in the pharmaceutical preparation. As a result, accurately determining each target parameter requires constructing an appropriate calibration model, which is a slow, difficult task [1].

Introducing process analytical technology (PAT) in a pharmaceutical process entails determining some parameters in order to establish the critical quality attributes for the process or its product. This is usually accomplished by using a fast-response technique such as NIR or Raman spectroscopy. Constructing accurate models for pharmaceutical processes requires using a simple, fast, flexible methodology for this purpose [2].

The greatest difficulty in developing a robust analytical model is constructing a representative calibration set. In fact, the samples to be included in the calibration set should span a wide enough concentration range to enable the prediction of samples with abnormal concentrations and meet the ICH guidelines as regards maximum ranges; also, they should encompass the whole physical and chemical variabilities of the production process [3]. In order to meet the former requirement, the ICH guidelines [4] recommend using a range ±20% around the nominal value to determine the API; also, correlation between components in the calibration samples should ideally be as low as possible.

Physical variability in production samples arises from differences in density, particle size, granulation and compaction, among other factors, which result in differences in light trajectory and scatter that in turn lead to spectral dissimilarities [5]. Because even small differences can have a strong impact, incorporating them into the calibration model is crucial with a view to obtaining acceptable predictions.

The previous requirements are essential in order to construct effective calibration sets [6]; also, the more similar the calibration samples and those to be predicted (*viz.* production samples) are, the

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2011.07.082

Table 1
Concentration correlation coefficients of the different components in Irbesartan laboratory blends.

Correlation coefficients	Irbesartan	Lactose	Avicel pH 102	PVP 25	Croscarmellose sodium	Aerosil 200	Sodium stearyl fumarate
Irbesartan	1	–	–	–	–	–	–
Lactose	0.32	1	–	–	–	–	–
Avicel pH 102	0.23	0.20	1	–	–	–	–
PVP 25	−0.99	−0.39	−0.29	1	–	–	–
Croscarmellose sodium	−0.99	−0.40	−0.28	0.99	1	–	–
Aerosil 200	−0.99	−0.38	−0.29	1.00	0.99	1	–
Sodium stearyl fumarate	−0.99	−0.40	−0.28	0.99	1.00	0.99	1

better will be the predictions. Fulfillment of these requirements can be ensured by using a number of methods [7–14]. Some tablet production variables including grain size [1,8,9], compaction pressure [1,11], tablet shape [5] and coating [17] influence the NIR spectrum for the product and should therefore be considered in constructing a multivariate calibration model to determine the API content. Recently, Blanco et al. [15] proposed a new method for incorporating physical variability in a pharmaceutical production process into the set of calibration spectra via a calculated process spectrum. The process spectrum is the difference spectrum between that for the sample at an intermediate stage of the process (mixture, granulate, core) or its final stage (tablet) and a powder mixture of identical composition. The differences calculated for several production samples are expressed as a set of mathematical vectors defining variability in the production process. This is the so-called "process variability matrix", which is added to the matrix containing the set of NIR spectra for several powder mixtures spanning the desired range of API concentrations in order to obtain the spectral matrix for the calibration set. The spectrum for a calibration sample is the sum of the weighted contributions of the different sample components (laboratory-made samples in our case) and that of the production process (*i.e.* the process spectrum).

Usually, the number of available process spectra is small and may not contain the whole variability in the production samples. Under such conditions, the variability included in the process variability matrix can be increased by multiplying the process spectra by an empirical coefficient m to obtain an expanded set: the extended process variability matrix. The sum of the spectral matrix for the laboratory powder samples – which contains chemical variability – and the extended process variability matrix – which contains physical variability introduced by the different steps of the production process – provides the extended total variability matrix – which contains both physical and chemical variabilities in the process, and is used to construct the calibration model. A poor selection of samples for the calculation of process spectra can introduce systematic errors in the prediction of production samples [15].

In this work, we developed a methodology for selecting the most suitable empirical coefficient m to be used in calculating the extended variability matrix and assessed the effect of spectral treatments used to obtain the process spectrum on the simplicity of the ensuing calibration model and its predictive ability.

2. Experimental

2.1. Production samples

The pharmaceutical products studied were 100 tablets of Irbesartan 300 mg and 52 tablets of Paracetamol 1 g, both from Laboratorios KernPharma. Irbesartan 300 mg is commercially available as uncoated white cylindrical tablets containing Irbesartan (59.29 wt%) as API; lactose (19.76 wt%) and Avicel pH 102 (10.97 wt%) as major excipients; and PVP 25 (2.96 wt%), Croscarmellose sodium (2.67 wt%), sodium stearyl fumarate (2.57 wt%) and Aerosil 200© (1.78 wt%) as minor excipients. The paracetamol 1 g tablets were also white and uncoated, and contained Paracetamol (89.90 wt%) as API; pre-gelatinized starch (8.70 wt%) as major excipient; and stearic acid (0.50 wt%) and Povidone (1.00 wt%) as minor excipients.

2.2. Laboratory samples

The pure components of Irbesartan 300 mg tablets were used to prepare a total of 30 laboratory samples consisting of accurately weighed amounts of the powdered ingredients spanning a concentration range ±20% around the nominal API content as per the ICH guidelines [4]. The sample set was established by using a D-optimal design in order to minimize correlation between concentrations (see Table 1).

Similarly, the pure components of Paracetamol 1 g tablets were used to prepare a total of 26 mixtures spanning a concentration range from 80 to 111% the nominal API content (898 mg/g) in the commercial tablets, the latter value corresponding to pure paracetamol. Such a high API concentration and low excipient concentrations precluded using an experimental design to prepare a sample set with low correlated concentrations.

All samples were made by weighing the required amounts of each component on an analytical balance. The mixtures were homogenized in a solid mixer and assumed to be homogeneous when they gave two identical consecutive NIR spectra. All concentrations were expressed as percentages of the total weight of each pharmaceutical preparation.

2.3. Instrumentation and software

Laboratory samples were homogenized in a Turbula T2C WAB shaker mixer. The NIR spectra for the Irbesartan and Paracetamol laboratory samples and tablets were recorded on a Bruker MPA Fourier Transform spectrophotometer equipped with an integrating sphere and a rotary module furnished with a cast of the same shape and size as the tablets. The instrument was governed via the software Opus v. 6.5.

The D-optimal design was developed with the software Modde v.6.0 from Umetrics.

All spectra were processed and multivariate models were constructed by using The Unscrambler v. 9.8 from Camo.

2.4. Recording of NIR spectra

The NIR reflectance spectra for the samples were recorded over the range 12 500–3750 cm^{-1} (800–2666 nm) with a maximum spectral resolution of 2 cm^{-1} and an accuracy of 0.3 nm. Each spectrum was the average of 32 scans.

Powder samples were held in vials for direct placement on the autosampler window. Each sample was used to record two spectra, with turnover between the two, for averaging.

The spectra for Irbesartan tablets were recorded in cylindrical casts, and those for Paracetamol tablets in oblong casts, which were directly placed on the autosampler. Each sample was used to record

two spectra and their average used to predict the component concentrations.

2.5. Preparation of the calibration–validation set

The process spectrum (S_p, which contained the contributions of granulation and pressing), was calculated as

$$S_p = S_T - S_{ref} \quad (1)$$

where S_T is the spectrum for the production sample (tablet) and S_{ref} that for a laboratory powder sample containing the same API and excipient concentrations as the production sample. Calculations were based on the data for 5 tablets from as many production batches of Irbesartan 300 mg and 10 of Paracetamol 1 g. The 5 Irbesartan and 10 Paracetamol process spectra thus obtained contained the spectral variability introduced by the production process. These spectra constituted the reduced process variability matrix, $\mathbf{S}_{p.red}$, for the production steps, which was added to the spectral matrix for laboratory samples of known concentration in order to obtain the reduced total variability matrix, $\mathbf{S}_{t.red}$.

$$\mathbf{S}_{t.red} = \mathbf{S}_{conc} + \mathbf{S}_{p.red}$$

The number of process spectra used was small; also, they might not contain the whole variability of the production process. In order to increase the variability in $\mathbf{S}_p$, we used a multiplying factor ranging from 1.5 to 0.5; specifically, we used the values 0.5;1.5, 0.75;1.25 and 0.91;1.1. This not only increased the number of process spectra, but also the variability spanned by the new set of process spectra, which constituted the extended process variability matrix, $\mathbf{S}_{p.ext}$. This new spectral set was randomly added to the spectra for laboratory powder samples spanning an API concentration range ±20% around the nominal value for the Irbesartan formulation, and 80–111% for the Paracetamol formulation:

$$\mathbf{S}_{t.ext} = \mathbf{S}_{conc} + \mathbf{S}_{p.ext} \quad (2)$$

where $\mathbf{S}_{t.ext}$ is the spectral matrix with the desired process variability and concentration range for calibration, $\mathbf{S}_{conc}$ that for the laboratory powder samples and $\mathbf{S}_{p.ext}$ the extended process variability matrix. The spectral set $\mathbf{S}_{t.ext}$ constituted the extended total variability matrix, which was subjected to various treatments such as SNV and derivation prior to construction of the calibration model [16].

2.6. Reference values

The reference values for the laboratory-made samples were obtained from the weights of the formulation components.

The API content of the Irbesartan 300 mg production samples used to validate the models was determined by HPLC–UV/Vis under the following conditions: a steel column 25 cm long × 4.6 mm i.d. packed with Kromasil 100 C_{18} in 5 μm particle size; a mobile phase consisting of 65:35 (v/v) buffer at pH = 2.5/acetonitrile (A) and 30:70 (v/v) buffer at pH = 2.5/acetonitrile (B), and used in a gradient regime at a flow rate of 0.8 mL min^{-1}; an injected volume of 20 μL; a chromatographic run time of 35 min and a detection wavelength of 275 nm.

The API content of the Paracetamol 1 g production tablets used to validate the model was also determined by HPLC–UV/Vis, using a steel column 25 cm long × 4 mm i.d. packed with Spherisorb ODS-2 in 5 μm particle size; a 70:30 (v/v) methanol/(water/acetic acid) at a flow rate of 1.0 mL min^{-1} as mobile phase; an injected volume of 25 μL; a chromatographic run time never exceeding 10 min and a wavelength of 280 nm for detection.

2.7. Construction of calibration models

The spectra constituting the extended total variability matrix were used to develop PLS calibration models. Previously, the spectra were subjected to various treatments including SNV, and first and second derivatives, using the Savitzky–Golay algorithm with a 11-point window and fitting to a second-order polynomial in all cases.

The calibration model was constructed by using the PLS model with the number of factors leading to the lowest RMSE, it was calculated from:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}(Y_i^{NIR} - Y_i^{REF})^2}{n}}$$

As determined from a plot of RMSE vs number of factors, Y_i^{NIR} being the API content predicted by the model, Y_i^{REF} that provided by the reference method (weighing or HPLC) and n the number of samples.

3. Results and discussion

Constructing an effective calibration set is crucial with a view to obtaining accurate predictions with multivariate models based on NIR spectroscopy; developing a new method involving calculation of the process spectrum requires a thorough study of the variables influencing its performance. To this end, we used various methods to obtain the process spectrum with a twofold purpose, namely: (a) to establish the effect of spectral treatments in constructing calibration models and obtaining the process spectrum and (b) to incorporate as much variability in the production samples as possible in order to ensure goodness of the model, which entailed determining the multiplying factor m to be applied to the process spectrum in order to ensure the inclusion of production samples in the calibration set.

A good calibration set should include chemical variability in the production process and also variability derived from each sample processing step. This can be accomplished by using a series of laboratory mixtures prepared in accordance with a D-optimal design and spanning the desired concentration range for each component with minimal collinearity between concentrations. Table 1 shows the concentration correlation in the laboratory-made Irbesartan samples. As can be seen, the correlation coefficients between the concentrations of the major components (Lactose, Avicel pH 102 and the API) were all less than 0.50; this precluded reducing collinearity between the minor components, which spanned the concentration range 1.5–3%.

Physical variability in the process was incorporated into the calibration model by using the process spectrum methodology, which involved combining the process spectra ($\mathbf{S}_p$) with those for laboratory samples prepared as described above. As can be seen from Fig. 1, which compares the spectrum for a production sample with that of a powder mixture of identical concentration in each ingredient, the former exhibited a shift and slightly different band intensities as a result of physical changes during the production process (granulation and compaction); the differences, however, were reduced by adding the spectrum for a laboratory sample to $\mathbf{S}_p$ [5]. The process spectrum is not unique; rather, it is a function of the samples used in its calculation since each production process has a different contribution of the factors introducing variability in it. Therefore, the matrix of process spectra is obtained from several process spectra for different production samples in order to ensure that the varying magnitudes of sample variability are included. In any case, it is rather difficult to ensure that the set of samples used to record the process spectra will be representative of the overall variability in the production process.

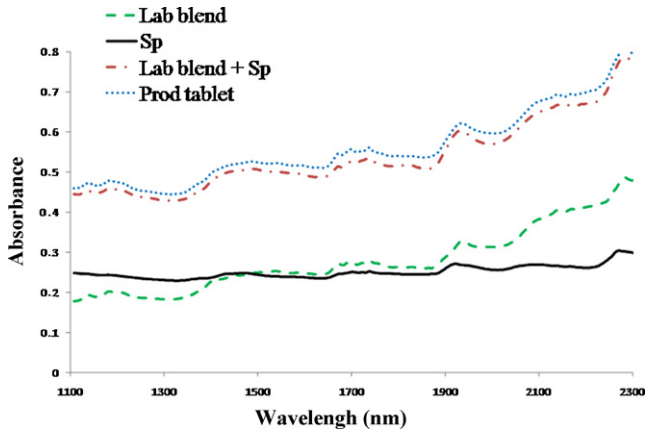

Fig. 1. Absorbance spectra of a laboratory blend, Irbesartan 300 mg production tablet, spectrum process (S_p) and laboratory blend + S_p.

A scatter plot of the scores for the first two PCs (Fig. 2) revealed that the spectra for laboratory samples combined with $\mathbf{S}_p$ (matrix $\mathbf{S}_{t.red}$) were more similar to the production samples. As can be seen from Fig. 2A, the laboratory mixtures departed from the production samples; by contrast, the combination of the spectra for the laboratory samples and the process spectra ($\mathbf{S}_p + \mathbf{S}_{conc} = \mathbf{S}_{t.red}$) clustered with the production tablets (Fig. 2B), which suggests that adding process spectra to laboratory-made samples ensures the incorporation of a sizeable fraction of variability in the process.

3.1. Influence of spectral treatments

The best spectral treatment for obtaining the calibration models was identified from a scatter plot of the PCA scores for $\mathbf{S}_{t.red}$ and production tablets subjected to the SNV, first-derivative and second-derivative treatments over the wavelength range 1100–2300 nm. The clusters of laboratory samples for all spectral treatments included virtually all production samples (Fig. 3A shows this sample distribution for first derivative spectra and that are similar for different pre-treatments); therefore, the samples in $\mathbf{S}_{t.red}$ incorporated most of the variability in the production process and any of the above spectral treatments could in theory be used to construct the calibration model. This was confirmed by constructing PLS models from spectra subjected to each treatment. The model constructed from laboratory mixtures alone ($\mathbf{S}_{conc}$) (*i.e.* including no process spectra, $\mathbf{S}_p$) resulted in large prediction errors for the production samples despite the low intensity of the $\mathbf{S}_p$ bands (see Table 2). A comparison of the RMSEC/P values for the models including $\mathbf{S}_{t.red}$ revealed that all models provided reasonably accurate predictions; however, the simplest, best predictive model was that obtained with the first-derivative spectral treatment, which was thus adopted to determine the optimum multiplying factor, *m*.

Using the different spectral treatments to obtain $\mathbf{S}_p$ resulted in no difference in the ensuing models irrespective of the time of application. This was not the case with the process spectra subjected to SNV; in fact, the $\mathbf{S}_p$ models constructed from SNV-treated absorbance spectra provided increased bands by effect of the autoscaling involved in the SNV treatment.

With the other spectral treatments, altering the application sequence made no difference, which suggests that the point of application of the treatment has no effect on the goodness of the results.

3.2. Increasing the variability of calibration samples

As can be seen from Figs. 2 and 3A – which corresponds to first-derivative spectra – the production samples were not completely included in the $\mathbf{S}_{t.red}$ cluster; therefore, variability in the production samples was not completely represented in these samples. This problem was solved by increasing the initial variability (specifically,

Table 2
Relevant parameters of the PLS models constructed from laboratory blends and laboratory blends + S_p for quantifying the Irbesartan (API). Effect of different pre-treatments over the wavelength range 1100–2300 nm.

Spectral pre-treatment	Calibration samples	PLS factors	Explained variance (%)	Calibration	Prediction
				RMSEC	RMSEP
1st derivative	Lab blends	4	99.0	0.70	8.07
Absorbance	Lab blends + S_p	4	97.3	1.14	3.07
SNV	Lab blends + S_p	4	98.0	0.97	2.88
1st derivative	Lab blends + S_p	3	97.7	1.17	1.53
2nd derivative	Lab blends + S_p	3	97.8	1.10	1.75

Calibration samples: 20 laboratory blends (concentration range of 46.4–69.6%, w/w), prediction samples: 10 batches (10 tablets are analyzed in each batch).

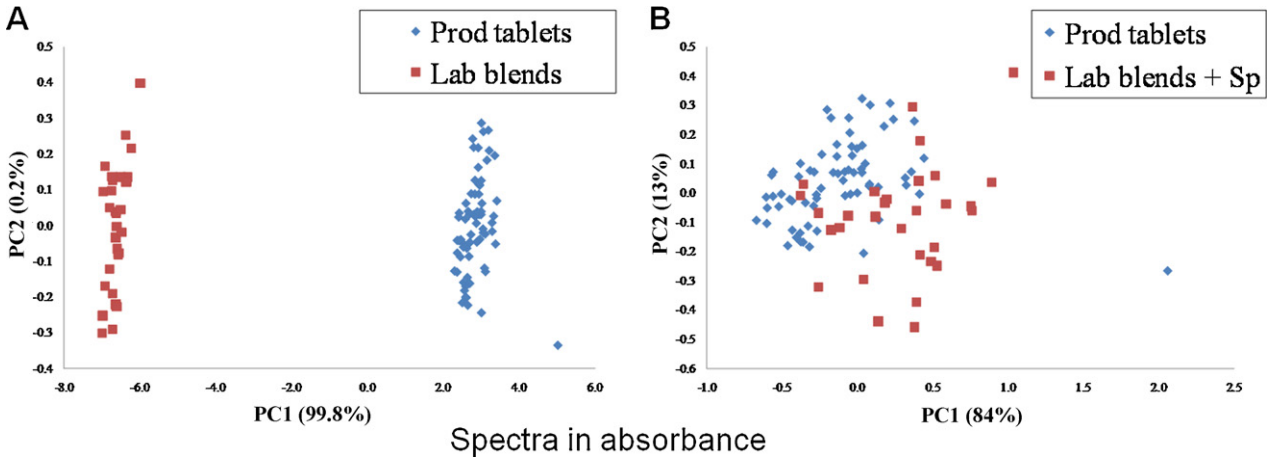

Fig. 2. Scatter plot of PCA scores from absorbance spectra over the wavelength range 1100-2300 nm for: (A) Laboratory blends and production tablets and (B) Laboratory blends + S_p and production tablets.

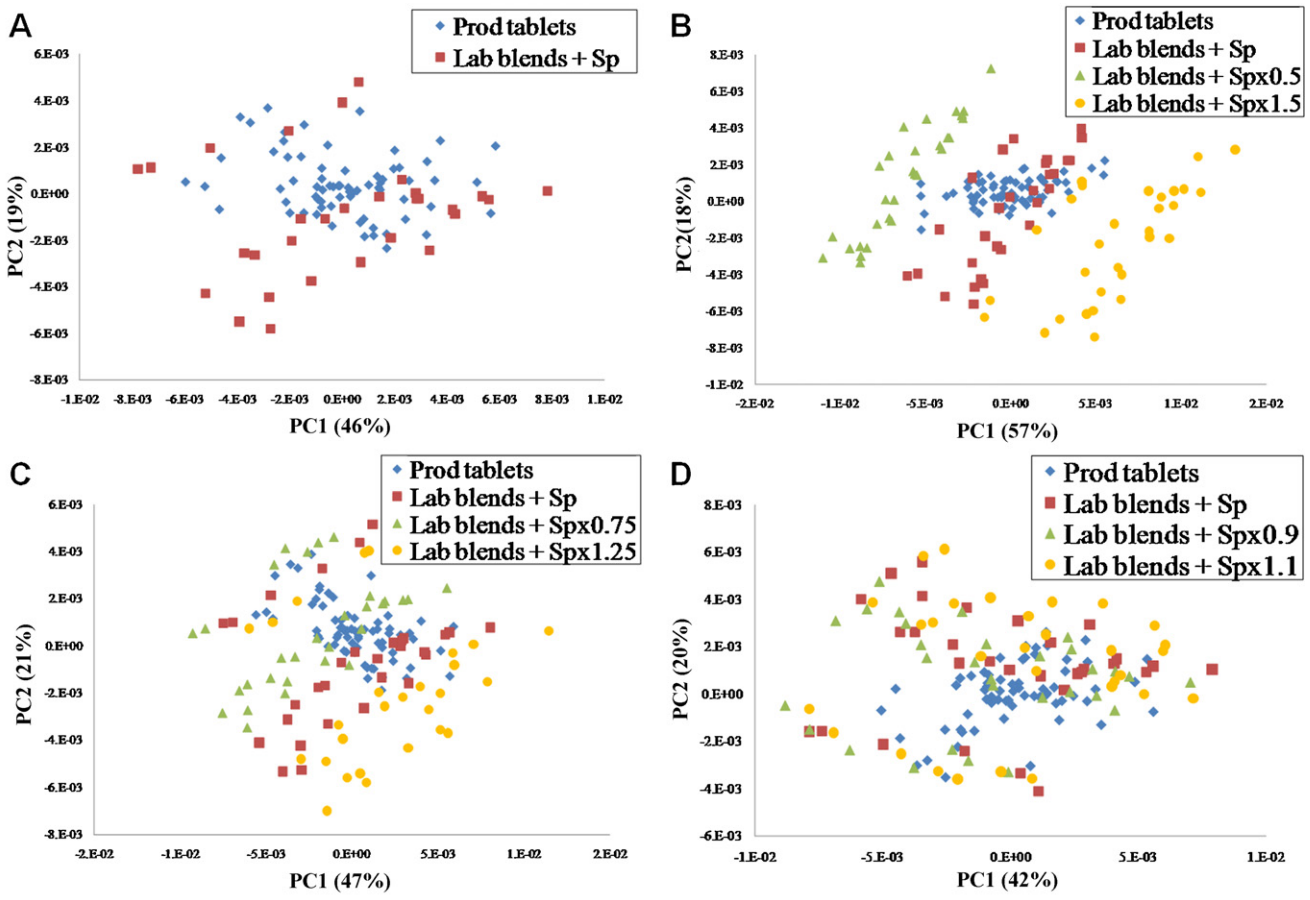

Fig. 3. Scatter plot of PCA scores for production tablets and laboratory blends added reduced or extended process variability matrix. Value of Multiplicative factors (m): (A) 1, (B) 0.5; 1.5, (C) 0.75; 1.25, (D) 0.9; 1.1.

by multiplying the process spectrum by a near-unity factor m). We used principal component analysis (PCA) to determine the specific m value ensuring inclusion of the spectra for all production samples in the cluster of $\mathbf{S}_{conc}$ combined with the products of the process spectra by m.

We used a scatter plot of PCA scores to explore the spectra for laboratory samples modified with the extended process variability matrix ($\mathbf{S}_{conc} + m\mathbf{S}_p = \mathbf{S}_{t.ext}$) and production tablets in order to select the most suitable m values with a view to constructing a set of spectra $\mathbf{S}_{conc} + m\mathbf{S}_p$ incorporating the whole variability in the production samples. Fig. 3 shows the scatter plots for four PCAs on combinations of spectra for laboratory samples with process spectra multiplied by the m values 0.5;1.5, 0.75;1.25 and 0.91;1.1; the spectra were subjected to the first-derivative treatment over the wavelength range 1100–2300 nm. As can be seen, $\mathbf{S}_{t.ext}$ failed to include all production samples (Fig. 3A); also, the laboratory samples whose spectra were applied an m value of 0.5;1.5 departed considerably from the production samples (Fig. 3B) and those with an m value less than 0.75;1.25 fell closer to them (Fig. 3C), but those whose spectra were multiplied by $m = 1.25$ fell below and distant from the production samples. Finally, using an m factor of 0.9;1.1 led to the laboratory samples falling closer to the production samples (Fig. 3D). These results were confirmed by the figures of merit of the ensuing PLS models (Table 3). As can be seen, the PLS model obtained with $m = 0.9;1.1$ was the simplest (*i.e.* that using the smallest number of factors) and also that resulting in the smallest RMSEP.

The goodness of the proposed methodology was further confirmed by using a variability matrix containing a variable number of process spectra (5 and 15). Fig. 4 shows three scatter plots for the PC scores. As can be seen, the cluster $\mathbf{S}_{conc} + \mathbf{S}_p$ ($\#S_p = 5$) failed to include all production samples (Fig. 4A), that constructed from

Table 3
Figures of merit for PLS models constructed from reduced and extended total variability matrix for Irbesartan (API) determination.

Samples for calibration	Multiplicative factors (m)	PLS factors	Explained variance (%)	Calibration		Prediction
				No. samples	RMSEC	RMSEP
Reduced total variability matrix	1	4	98.0	20	0.97	2.88
Extended total variability matrix	0.5; 1.5	5	97.9	60	1.01	1.67
	0.75; 1.25	4	97.7	60	1.07	1.58
	0.9; 1.1	3	97.7	60	1.17	1.53

Calibration samples with concentration range 46.4–69.6% (w/w). Prediction samples: 10 batches (10 tablets are analyzed in each batch) using 1st derivative mode over the wavelength range 1100–2300 nm.

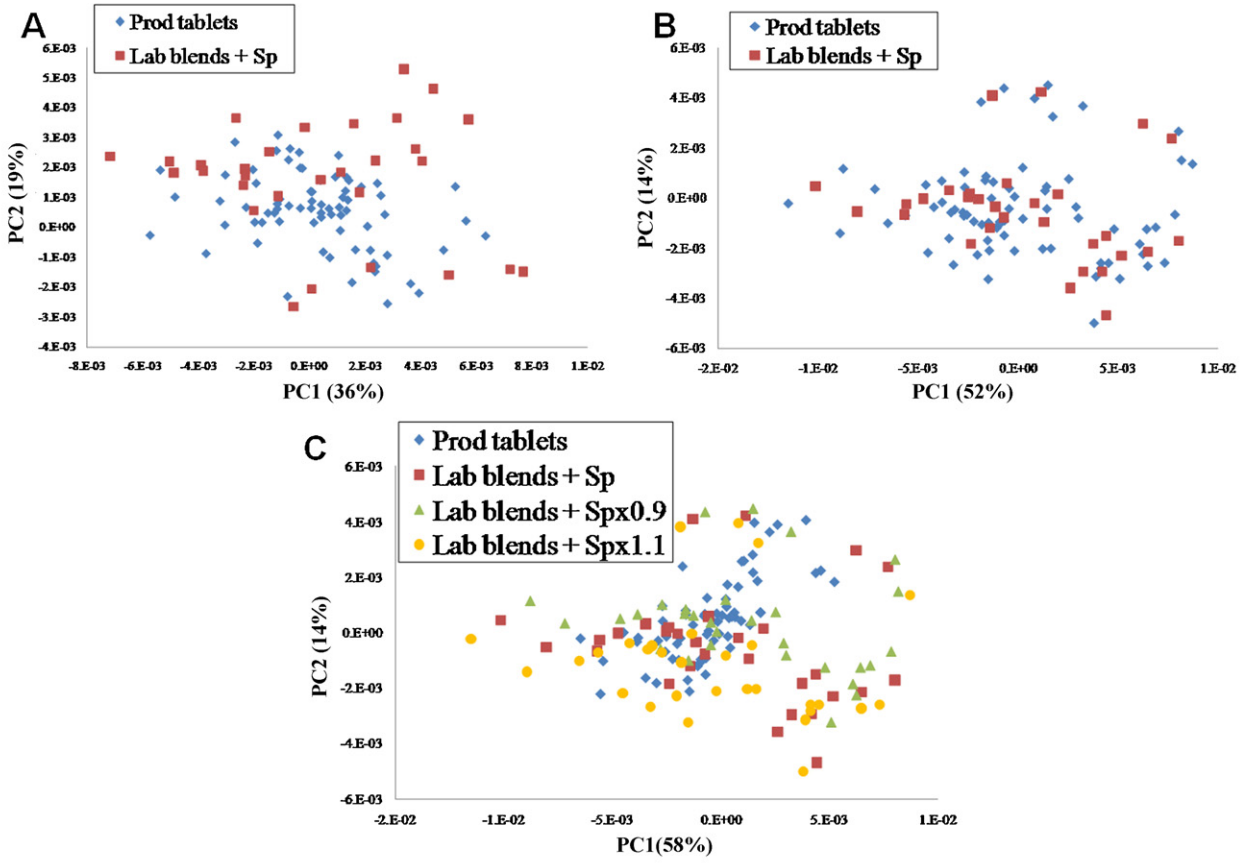

Fig. 4. Scatter plot of PCA scores for production tablets and laboratory blends added reduced process variability matrix: (A) 5 process spectra, (B) 15 process spectra (C) 5 process spectra added extended process variability matrix (m = 0.9:1.1).

Table 4
Figures of merit for PLS models constructed with different quantity of spectra process (S_p) for quantifying Irbesartan tablets.

Samples for calibration	No. S_p	Multiplicative factor (m)	PLS factors	Explained variance (%)	Calibration		Prediction
					No. samples	RMSEC	RMSEP
Reduced process variability matrix	5	1	3	98.2	20	0.92	2.09
	15	1	3	97.5	20	1.10	0.94
Extended process variability matrix	5	0.9; 1.1	3	98.4	60	1.17	1.53

Calibration samples with concentration range 46.4–69.6% (w/w). Prediction samples: 10 batches (10 tablets are analyzed in each batch) using 1st derivative mode over the wavelength range 1100–2300 nm.

15 $\mathbf{S}_p$ included more (Fig. 4B), and that obtained with an m factor of 0.9;1.1 included virtually all (Fig. 4C). Table 4 shows the figures of merit of the three models; as can be seen, the results confirmed the conclusions drawn from the scores plots. Thus, the best models were those encompassing the greatest variability without significantly increasing the complexity of the treatment. Also, using a multiplying factor and an increased number of $\mathbf{S}_p$ to expand the variability of the calibration samples led to better predictions of the production samples. However, using an increased number of process spectra entailed recording more spectra and detracted

Table 5
PLS models for excipients determination in Irbesartan tablets using the extended total variability matrix.

Components	Manufacturer's specification value (% w/w)	PLS factors	Pre-treatment	Explained variance (%)	Prediction[a]		
					Mean (% w/w)	Conc. range (% w/w)	S.D.
Lactose	19.76	10	2nd derivative	99.8	19.3	18.1–21.2	0.9
Avicel pH 102	10.97	9	2nd derivative	97.6	10.9	9.6–12.5	0.9
PVP 25	2.96	5	1st derivative	97.9	3.5	2.4–4.0	0.5
Croscarmellose sodium	2.67	5	1st derivative	99.1	3.2	2.2–3.6	0.4
Sodium stearyl fumarate	2.57	5	1st derivative	99.1	2.8	2.4–3.2	0.3
Aerosil 200	1.78	5	1st derivative	97.4	2.1	1.4–2.4	0.3

[a] Prediction samples: 10 batches (10 tablets are analyzed in each batch). Wavelength range 1100–2300 nm.

Table 6
Figures of merit of PLS models for paracetamol (API) and excipients determination using extended total variability matrix.

Components	PLS factors	Explained variance (%)	Calibration		Prediction			
			Conc. range (% w/w)	RMSEC	Conc. range (% w/w)	RMSEP	Mean (%w/w)	S.D.
Paracetamol	3	99.4	72–100	0.67	86.6–88.3	0.91	87.5	0.9
Corn starch	3	98.8	0–25	0.70	10.3–12.0	–	11.1	0.4
Stearic acid	7	97.4	0–2	0.11	0.1–0.9	–	0.4	0.2
Povidone	5	98.0	0–4	0.19	0.2–1.0	–	0.6	0.2

Prediction set consisted in 52 individual tablets; models in 1st derivative mode over a wavelength range 1100–2500 nm.

from expeditiousness, whereas using the multiplying factor led to similarly good results in a shorter time.

3.3. Determination of the excipients

The high predictive ability of the proposed methodology was confirmed by constructing PLS models for the other formulation components. Such models were established from the same spectra used for the API (Irbesartan) and excipient concentrations in the laboratory mixtures. Table 5 shows the figures of merit of the different models. Worth special note here is the large number of PLS factors required by the models for Lactose (10) and Avicel (9), even at relatively high concentrations. This can be ascribed to the strong correlation between their spectra. The models required different spectral treatments for optimal performance. The results of Table 5 are quite consistent with the nominal concentrations of these two excipients, the actual contents of which in the formulation are unknown because their analysis is usually not required. This testifies to the capabilities of the proposed methodology. Not knowing the actual concentrations of these excipients precluded calculation of their RMSEPs; however, their standard deviations are suggestive of the quality of their quantitation.

The primary aim of constructing these models was not to determine the concentrations of the excipients as target parameters, but rather to show that the proposed methodology affords their quantitation, if desired, without the need to prepare a new set of samples.

3.4. Analysis of paracetamol tablets

The proposed methodology was used to analyze paracetamol production tablets containing the API (89.8 wt%) and maize starch (8.7 wt%) as major components. The high concentration of both precluded the use of an experimental design based on laboratory samples to reduce correlation between concentrations. Even so, the PLS models obtained were simple (only three factors) and exhibited good predictive ability (see Table 6). On the other hand, the PLS models for Povidone and stearic acid, which are present in relatively low proportions, were more complex (5 and 7 factors, respectively) owing to the similarity of their spectra (Fig. 5), the low concentration of the two excipients and the high correlation between their spectra ($r = 0.97$). In any case, both models had standard deviations less than 0.25 (Table 6).

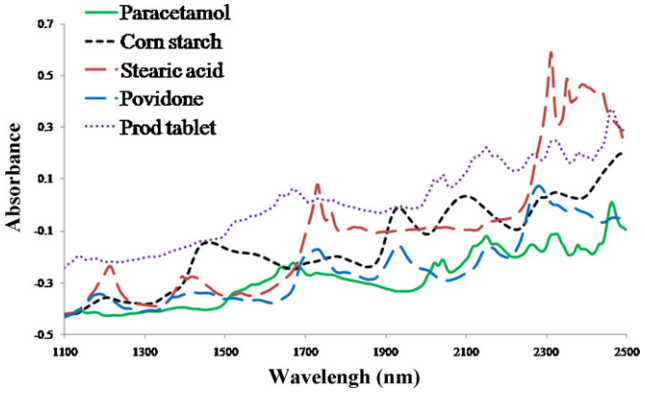

Fig. 5. NIR spectra for production tablet, Paracetamol (API) and excipients (Corn starch, Povidone and Stearic acid).

4. Conclusions

The study conducted in this work demonstrates the usefulness of the proposed methodology, based on the obtainment of a process spectrum that is used to construct a calibration set incorporating the variability in a pharmaceutical production process in the form of a process variability matrix. As shown here, the proposed methodology is compatible with the application of an SNV or derivative spectral treatment, but the most suitable treatment for each specific model cannot be established beforehand. Using a multiplying factor on process spectra expands the variability of the calibration set, and hence the quality of the ensuing PLS models, without additional experimental work. Also, a PCA scores plot provides a highly useful tool for determining the optimum value of the multiplying factor. Using an increased number of process spectra leads to more robust calibration models; this, together with an appropriate choice of the multiplying factor, provides simple models giving small prediction errors. This paper provides the keys and tools needed to use the proposed methodology as an alternative choice for constructing a calibration set incorporating variability in a production process with a view to quantifying any API and excipients with acceptable predictive ability in any type of pharmaceutical formulation.

Acknowledgements

The authors are grateful to Spain's MCyT for funding this research within the framework of Project CTQ2009-08312 and to KernPharma (Terrassa, Spain) for kindly supplying the samples. R.C. wishes to acknowledge additional support from Universitat Autònoma de Barcelona in the form of a scholarship.

References

[1] M. Blanco, R. Cueva-Mestanza, A. Peguero, J. Pharm. Biomed. Anal. 51 (4) (2010) 797–804.
[2] M. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J.P. Remon, C. Vervaet, Int. J. Pharm. (2011), doi:10.1016/j.ijpharm.2010.12.012.
[3] European Medicines Agency. Guideline on Specification and Control Test on the Finished Product, June 1992. 3 AQ11a. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003368.pdf.
[4] ICH Q2(R1) Validation of Analytical Procedures: Methodology: Text and Methodology, 2005. http://bioforum.org.il/Uploads/Editor/karen/q2_r1_step4.pdf.
[5] M. Blanco, A. Peguero, J. Pharm. Biomed. Anal. 52 (1) (2010) 59–65.
[6] M. Blanco, M. Alcalà, in: K. Bakeev (Ed.), Process Analytical Technology, 2nd ed., 2010, pp. 463–491.
[7] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, N. Pou, Fresenius J. Anal. Chem. 368 (5) (2000) 534–539.
[8] A.J. O'Neil, J. Andrew, D. Roger, A.C. Moffat, Analyst 123 (11) (1998) 2297–2302.
[9] A.J. O'Neil, R.D. Jee, A.C. Moffat, Analyst 124 (1) (1999) 33–36.

[10] M. Alcalà, M. Blanco, M. Bautista, J.M. González, J. Pharm. Sci. 99 (1) (2010) 336–345.
[11] M. Blanco, M. Alcalà, J. Pharm. Sci. 95 (10) (2006) 2137–2144.
[12] J.J. Moes, M.M. Ruijken, E. Gout, Int. J. Pharm. 357 (1–2) (2008) 108–118.
[13] J.D. Kirsch, J.K. Drennen, J. Pharm. Biomed. Anal. 13 (10) (1995) 1273–1281.
[14] S.H. Tabasi, R. Fahmy, D. Bensley, C. O'Brien, S.W. Hoag, J. Pharm. Sci. 97 (9) (2008) 4052–4066.
[15] M. Blanco, A. Peguero, Trends Anal. Chem. 29 (10) (2010) 1127–1136.
[16] A. Candolfi, R. De Maesschalck, D. Jouan-Rimbaud, P.A. Hailey, D.L. Massart, J. Pharm. Biomed. Anal. 21 (1) (1999) 115–132.
[17] M. Römer, J. Heinämäki, C. Strachan, N. Sandler, J. Yliruusi, AAPS PharmSciTech 9 (4) (2008) 1047–1053.