International Journal of Pharmaceutics 380 (2009) 49–54

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

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

Applications of NIR in early stage formulation development. Part II. Content uniformity evaluation of low dose tablets by principal component analysis

Weiyong Li[a,*], Laurent Bagnol[b], Maxine Berman[a], Renato A. Chiarella[a], Michael Gerber[b]

[a] Chemical & Pharmaceutical Development, Johnson & Johnson Pharmaceutical Research & Development, LLC, Welsh & McKean Roads, Spring House, PA 19477, United States
[b] Chemical & Pharmaceutical Development, Johnson & Johnson Pharmaceutical Research & Development, LLC, 1000 US Highway 202 South, Raritan, NJ 08869, United States

ARTICLE INFO

Article history:
Received 8 April 2009
Received in revised form 12 June 2009
Accepted 24 June 2009
Available online 2 July 2009

Keywords:
Near infrared spectroscopy
Principal component analysis
Partial least squares
Content uniformity
Tablet

ABSTRACT

A near infrared method based on principal component analysis (PCA) was developed for predicting content uniformity of low dose tablets manufactured by a direct compression process. The work was conducted in early stage formulation development. NIR spectra of one hundred and eighty tablets from three feasibility batches were used as the pseudo-calibration set. A correlation was established between PCA scores and a set of reference values obtained by HPLC analysis. The reference values were also used to define a concentration range for the active pharmaceutical ingredient to facilitate content uniformity prediction by PCA. Analyses of unknown samples were conducted by forming a prediction set that included the calibration and unknown samples, followed by PCA. Samples from two development batches were predicted using the PCA model and the results were consistent with the reference HPLC values. Remarkably, the model was able to predict CU for tablets that were prepared using different grades of lactose (anhydrous versus monohydrate). Additionally, during this study, the impact of spectrum pretreatments on PCA is demonstrated. A brief discussion is given to highlight the advantages of PCA over partial least squares (PLS) regression for analysis of samples generated in early stage formulation development.

1. Introduction

The development of low dose tablets poses a series of challenges. In particular, content uniformity (CU) becomes a critical product attribute, creating challenging processing and scale up scenarios compared to conventional tablets. Near infrared (NIR) spectroscopy is a widely accepted analytical tool in pharmaceutical analysis with applications in the study of content uniformity (Ritchie et al., 2002), tablet hardness (Blanco and Alcala, 2006), dissolution (Freitas et al., 2005), and polymorph conversion (Blanco and Villar, 2003; Li et al., 2005), etc. When developing NIR calibration models for analysis of low dose tablets, special care is needed to ensure specificity, accuracy, and robustness. Norris and Ritchie (2008) proposed to demonstrate specificity by showing a good match between the first principal component spectrum and the spectrum of the active pharmaceutical ingredient (API). Alcala et al. (2008) reported an approach to improve quantitation and detection limits by reducing the upper concentration level of the partial least squares (PLS) models. It has been demonstrated that NIR calibration models can be developed for formulations with as low as 0.5% (w/w) API. Xiang et al. (2009) discussed robustness of NIR calibration models. They pointed out that the strong correlations observed in the PLS models were not solely based on the API information in some cases.

As opposed to most NIR applications for quantitative pharmaceutical analysis that use PLS regression (Roggo et al., 2007), this article explores simpler and more efficient alternatives to support early stage formulation development. In Part I of this series, the determination of blend uniformity and content uniformity by NIR without calibration models was described (Li et al., 2007). A semi-quantitative approach was developed by using the assumption that homogeneity of powder blends and tablets can be evaluated based on relative intensity changes of properly selected NIR signals. The method was applied to an early stage formulation development project and was able to distinguish between batches that had satisfactory and unsatisfactory content uniformity and potency. In this article, we explore the use of principal component analysis (PCA) in early stage formulation development. A recent project required the development of a low dose tablet formulation manufactured by direct compression. In this case, the method in Part I could not be used because the API signals were not strong enough to be separated from the background. For this development project, a PCA model was developed instead. The model is very different from PLS models due to its simplicity, which takes advantage of the PCA algorithm to separate chemical variations of the API from those of excipients and physical changes. Specificity and robustness of the PCA model are demonstrated. The model was successfully applied to monitoring CU of tablets from formulation development batches.

* Corresponding author. Tel.: +1 215 628 5122; fax: +1 215 628 5897.
E-mail address: wli1@its.jnj.com (W. Li).

0378-5173/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2009.06.032

Reliable results were obtained even when two batches of tablets were prepared using two different grades of lactose (anhydrous versus monohydrate).

2. Experimental

2.1. Materials

The excipients used were lactose anhydrous (direct compression grade, purchased from Kerry Bio-Science, Norwich, NY, USA), lactose monohydrate (Fast Flo, purchased from Foremost, Baraboo, WI, USA), and microcrystalline cellulose (MCC, Avicel PH102, purchased from FMC Biopolymer, Philadelphia, PA, USA). The API is a proprietary compound under development by Johnson & Johnson Pharmaceutical Research & Development, LLC (JJPRD).

2.2. NIR instrument

A FOSS XDS near infrared rapid content analyzer (RCA) was used for the sample analysis (FOSS NIR Systems, Laurel, MD, USA). The NIR settings were as follows:

Sampling module—MasterLab.
Detection/detector—Transmission/InGaAs.
Number of scans—32.
Resolution—0.5 nm.
Wavelength range—800–1650 nm.
Spectrum pretreatment—1st derivative (Gap-Segment; Gap size = 5, Segment size = 15) unless otherwise specified.

The tablets were scanned directly without any preparation.

2.3. PCA

PCA was conducted using the chemometrics software Unscrambler (Camo Process AS, Nedre Vollgate, Norway; version 9.7).

2.4. Formulation

The tablet formulation consisted of 5 mg API, lactose, MCC and other excipients. Tablets were round concave with 7/32-inch diameter for a total weight of 80 mg (6.25% API).

2.4.1. Formulation feasibility batches

Blends were prepared by passing the API and a premixing excipient (lactose or MCC) through a mill and then blended in a bin blender with the remaining excipients. Tablets were compressed using a SMI Piccola 10-station press (SMI Incorporated, Lebanon, NJ, USA) at 50 rpm with three compression forces (140, 160, and 180 MPa).

2.4.2. Process development batches

Blends were prepared by passing the API and lactose (anhydrous or monohydrate) through a mill and then blended in a bin blender. Tablets were compressed using a Manesty Betapress (OYSTAR Manesley, Merseyside, England) at 160 MPa. Samples were retrieved at the beginning, middle and end of tablet compression.

2.5. Tablet content uniformity test by HPLC

Sample solutions were prepared at a concentration of 50 μg/mL in a 1:1 acetonitrile/water (v/v) sample solvent. An Agilent (Wilmington, DE, USA) 1100 HPLC system equipped with a photodiode array detector was used for sample analysis. Reversed-phase chromatography was carried out on an Ascentis Express C18 column, 150 mm × 4.6 mm, 2.7 μm particles (Supelco, Bellefonte, PA, USA), with an isocratic mobile phase consisting of 0.05% trifluoroacetic acid in water/acetonitrile (53:47, v/v). The flow rate was set at 1.0 mL/min, and the column temperature was maintained at 35 °C. The injection volume was 20 μL. Waters (Milford, MA, USA) Empower 2 software was used for data acquisition and processing.

3. Results and discussion

The content uniformity evaluations were performed on two manufacturing campaigns of a direct compression formulation. In the first campaign, the feasibility studies compared two premixing excipients, lactose and MCC. In the second campaign, the process development studies compared anhydrous and monohydrate lactose for effectiveness in improving content uniformity. Table 1 shows the detailed campaign, batch and sub-batch information for easy identification of the samples. The focus of this article is to demonstrate the feasibility of using PCA for evaluation of content uniformity of low dose tablets.

Table 1
Formulation, sample, and processing conditions.

Campaign/batch	Sub-batch	Process cond.	Premixing excipient	No. of tablets
Feasibility 1	1	CP01	Lactose anhydrous	20
	2	CP02		20
	3	CP03		20
Feasibility 2	4	CP01	Lactose anhydrous	20
	5	CP02		20
	6	CP03		20
Feasibility 3	7	CP01	MCC	20
	8	CP02		20
	9	CP03		20
Development 1	10	Beginning	Lactose anhydrous	20
	11	Middle		20
	12	End		20
Development 2	13	Beginning	Lactose monohydrate	20
	14	Middle		20
	15	End		20
HPLC ref.	16			30[a]
Validation				20[b]

CP, compression force.
[a] 10 from each feasibility batch.
[b] 10 from each development batch.

Table 2
PCA results for feasibility and development batches.

Sample	Tablet (*n*)	PC01		PC02		PC03		PC04		PC05	
		ID	*X*%	ID	*X*%	ID	*X*%	ID	*X*%	ID	*X*%
Feasibility 1	60	Hd	86	Unk	13	API	1				
Feasibility 1 + 2	120	Hd	85	Unk	14	API	1				
Feasibility 1 + 2 + 3	180	Hd	85	Unk	13	API	2				
Feasibility + development	300	Hd	65	Unk	26	Unk	5	Unk	4	API	<1

Hd, hardness; Unk, unknown; *X*, variance % expressed.

3.1. Principal component analysis of the feasibility batches

One hundred and eighty tablets were randomly selected from the three feasibility batches according to the sampling plan in Table 1 and scanned by NIR in transmission mode. Advantages and disadvantages of transmission versus reflectance NIR have been discussed in the literature (Merckle and Kovar, 1998). After pretreatment, principal component analyses were performed on three sets of spectra obtained from 60 (feasibility batch 1), 120 (feasibility batches 1 and 2) and 180 (feasibility batches 1, 2 and 3) tablets. The PCA results are summarized in Table 2. In all cases, three principal components (PCs) were used to account for the chemical and physical variations within each sample set. Even though a different number of tablets was used in each data set, the same number of PCs was identified (3 PCs), and the percent variance expressed (PVE) by each PC was very similar. Examination of the *X*-loading plots indicated that the first PC (PC01) was related to hardness variations (details not discussed here). Similar observations have been reported in the literature (Alcala et al., 2009). The chemical or physical meaning of the second PC (PC02) remains unknown whereas the third PC (PC03) was identified as the one related to API content variations in tablets. Efforts were made to explore the possibility of increasing PVE by PC03 using different pretreatment algorithms. Raw spectra of the 180 tablet set were first pretreated with standard normal variate (SNV), multiplicative scatter correction (MSC), or extended multiplicative scatter correction (EMSC), followed by the 1st derivative calculation and PCA (Table 3). Interestingly, the added pretreatment step caused a decrease in PVE by PC03 from 2% to 1% in all cases. A re-distribution of PVE between PC01 and PC02 was also observed. In addition, the *X*-loading plots of PC01 and PC02 were significantly altered because of the added pretreatment step whereas those of PC03 were not (Fig. 1). The fact that *X*-loading plots of PC03 were consistently present supported the notion that there may be a relationship between PC03 and API concentration variations.

The relationship between the third PC and the API content was further established based on two observations. First, there is certain resemblance between the third *X*-loading plot and 1st derivative spectrum of the pure drug substance (Fig. 2a). Good spectral matches are observed at 867, 885, 980, and 1100 nm. On the other hand, the spectral match in the 1140–1400 nm range is not obvious. The presence of overwhelming amounts of lactose and MCC may have caused the lack of good match in this region. Both lactose and MCC have a strong and overlapping absorption band in the 1300–1400 nm range, which may mask the API absorption in this region. The absorption band(s) of lactose and MCC in the 1100–1200 nm range is strong but slightly shifted from each other and from that of the API, which may cause the shape of the API band to be significantly changed (Fig. 2b). The most important evi-

Table 3
PCA results for the calibration set after different spectral pretreatments.

Pretreatment	PC01		PC02		PC03	
	ID	*X*%	ID	*X*%	ID	*X*%
1st derivative	Hd	85	Unk	13	API	2
SNV–1st derivative	Hd	94	Unk	5	API	1
MSC–1st derivative	Hd	95	Unk	4	API	1
EMSC–1st derivative	Hd	96	Unk	3	API	1

Hd, hardness; Unk, unknown; *X*, variance % expressed; SNV, standard normal variate; MSC, multiplicative scattering correction; EMSC, extended multiplicative scattering correction.

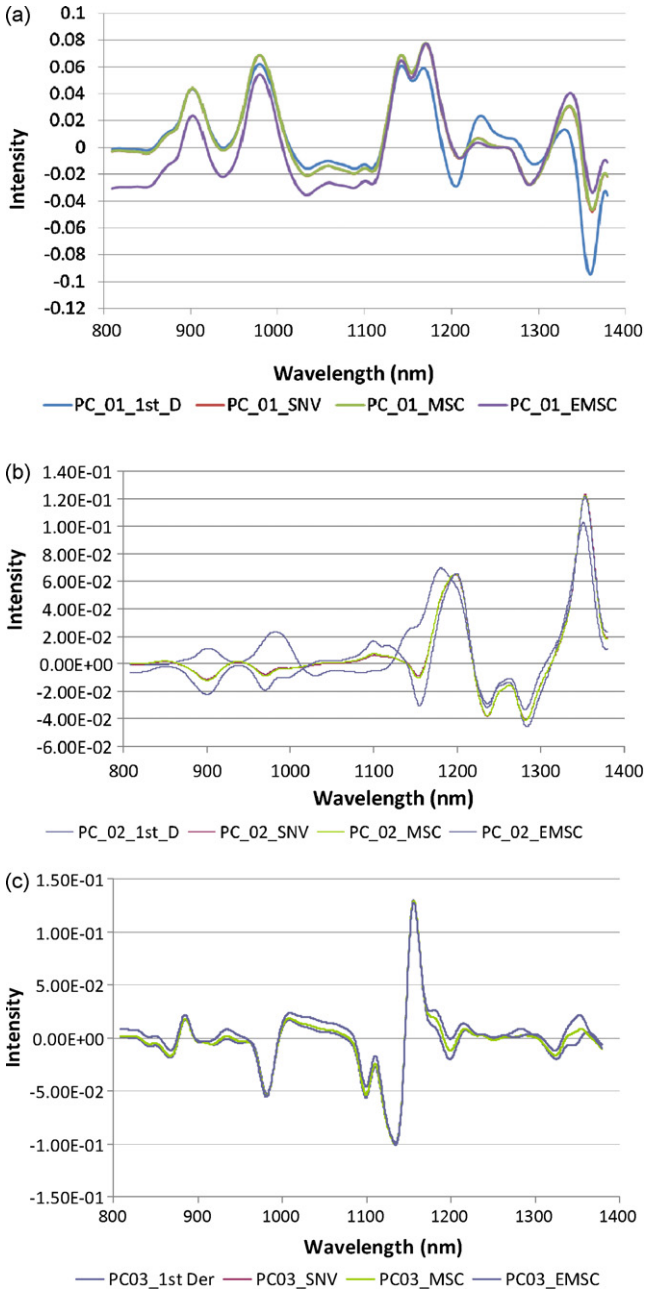

Fig. 1. *X*-loading plots of three PCs after multiple-step pretreatments.

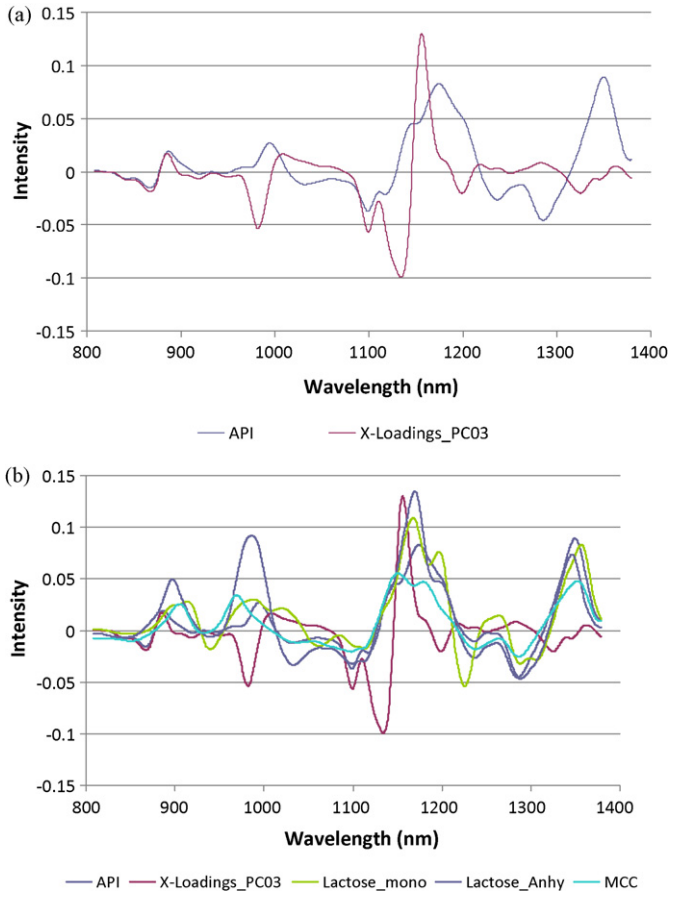

Fig. 2. Comparison of the third *X*-loading plot (red) from the PCA model with the 1st derivative spectra of (a) the API and (b) MCC and lactose. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dence was obtained by identifying the relationship between PCA scores and HPLC CU results. Thirty tablets (10 from each feasibility batch) were randomly selected and analyzed by HPLC. Content uniformity results from these reference tablets were plotted against the third set of scores (corresponding to PC03) of the same tablets. The PC03 scores displayed a good correlation with the CU results, which confirms the relationship between the third PC and API content (Fig. 3). Scores of PC01 and PC02, however, did not correlate with the CU results.

Based on the above discussion, a PCA model was "built" by using the spectra of all 180 tablets. Performance of the model is schematically presented in Fig. 4. The third set of scores of all 180 tablets is plotted according to their sub-batch numbers, with the red triangles indicating the locations of the 30 reference samples. Scores of the reference samples are also plotted using a pseudo-sub-batch number 16. A content uniformity range can then be defined by using the scores range of the reference samples. For example, a score range of −0.002 to +0.002 corresponds to 90–110% of the API. Without further calculation, it can be seen that all tablets from sub-batches 2 and 5 had concentration values within 90–110%, which is in agreement with the API content measured by HPLC for the selected samples from sub-batches 2 (96.8–104.1%) and 5 (99.4–102.7%). For other sub-batches, the concentration values lay outside the range for some of the tablets. It should be pointed out that the concentration range could be increased by testing more tablets that have extreme PCA scores using the reference method.

The above-discussed evidences demonstrate that the PCA model is specific in separating variations related to API concentration from other variances. Robustness of the model is hinted by the third *X*-loading plots from the principal component analyses, which were very consistent (Fig. 5). This also indicates that the PCA model may be suitable for content uniformity prediction of unknown tablets.

3.2. CU of development batches by the PCA model

In early stage formulation development, it is desirable to obtain analytical data quickly so that the effectiveness of formulation or processing modifications can be evaluated rapidly. To predict the CU of tablets from the development batches, the 180 spectra data set was used as a calibration set. Also, 120 tablets (60 from each development batch) were randomly selected (Table 1), scanned and the spectra pretreated as specified in Section 2.2. In the next step, the data set was combined with the calibration set to form a collection of 300 spectra as the prediction set. PCA was performed for the prediction set, which produced PCA scores and loadings corresponding to five PCs (Table 2). During manufacturing of the development batches, processing conditions were varied and two

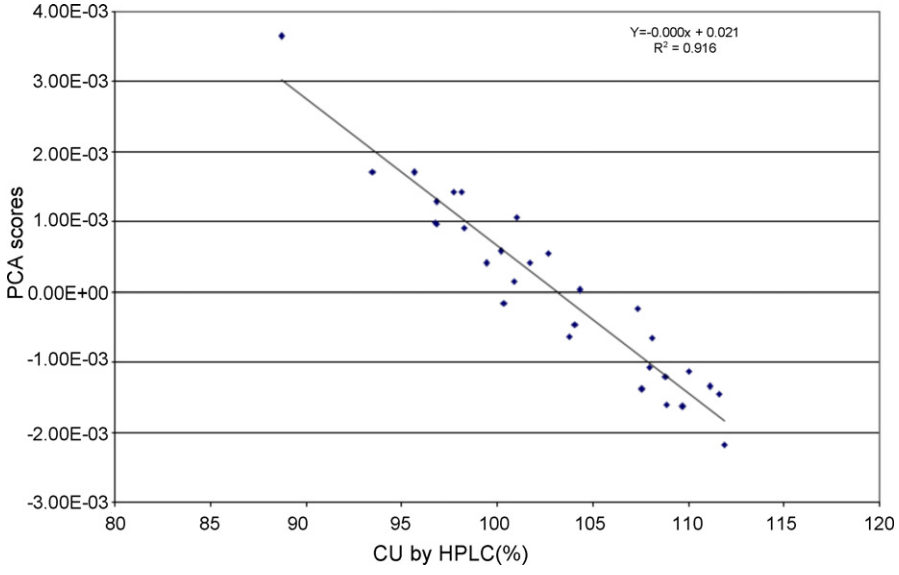

Fig. 3. Correlation between the third PCA scores of the reference samples and HPLC content uniformity results.

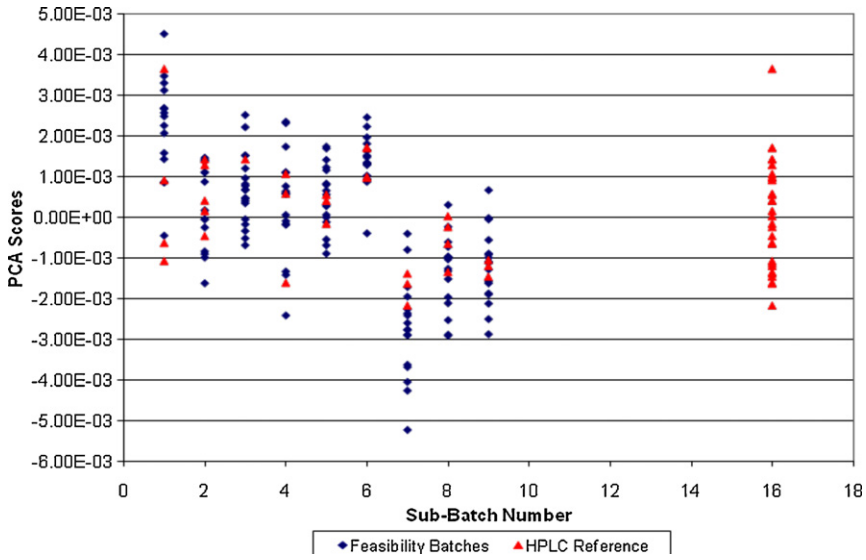

Fig. 4. The third PCA scores plotted against sub-batch numbers for samples from the three feasibility batches (red triangles showing scores of reference samples). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

different grades of lactose were used. These variations caused the PCA to pick up two additional PCs. Based on comparison of the *X*-loading plots, the fifth PC (PC05) was identified as describing API concentration variations in the whole data set. To confirm this relationship further, the third set of scores from the calibration set was plotted against the fifth set of scores from the prediction set (Fig. 6). A good correlation between them confirmed the identity of PC05. It is particularly interesting that the PCA model worked for both development batches even though they were prepared using two different grades of lactose.

The reason for performing PCA on the prediction set rather than the unknown samples directly is to take advantage of the available reference values and for easier identification of the correct PCs as demonstrated in Fig. 6. It may be difficult to identify the correct PC for CU calculation for an unknown sample set. In addition, the calibration set should span a wide range of API concentrations whereas the later development batches are expected to have less variation.

The predicted results for the development batches were plotted using Fig. 4 as a template and presented in Fig. 7 (sub-batch numbers 10–15). Again, without further calculations, it can be seen that the composition of all tablets from the development batches are within the 90–110% range. It is also evident that the development batch 1 has better content uniformity compared with batch 2 (sub-batches 10–12 and 13–15 in Fig. 7, respectively), indicating that anhydrous lactose is a better premixing excipient in terms of content uniformity. In an attempt to verify the PCA results, 20 tablets (10 from each development batch) were analyzed using the reference method. Batch 1 (anhydrous lactose) had CU values of 97.2–100.1% (mean = 98.8; RSD = 1.0) compared with a range of 91.2–98.5% (mean = 95.8; RSD = 2.1) for batch 2. These results were consistent with the PCA predictions.

3.3. Principal component analysis versus partial least squares regression

It has been demonstrated that a PCA model can be used for CU analysis of low dose tablets. However, what are the advantages of using a PCA model instead of a partial least squares (PLS) model? To answer this question, a PLS model was developed using samples from the feasibility batches (details about the model are not discussed here). The model failed to give good predictions for samples from the development batches. One interesting observation was that PC01 and PC02 of the model described hardness variation

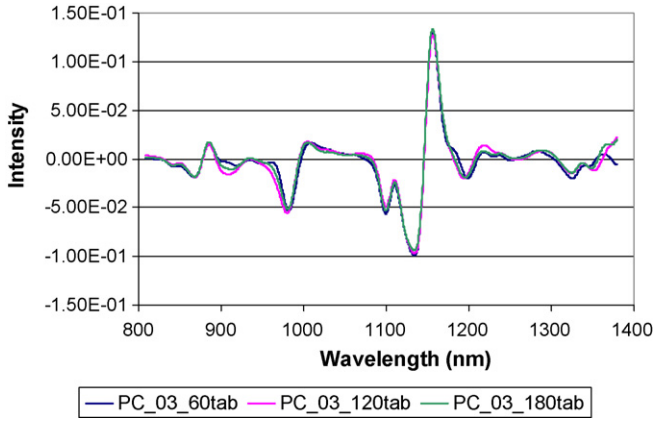

Fig. 5. The third *X*-loading plots from three separate principal component analyses with 60, 120, and 180 tablets.

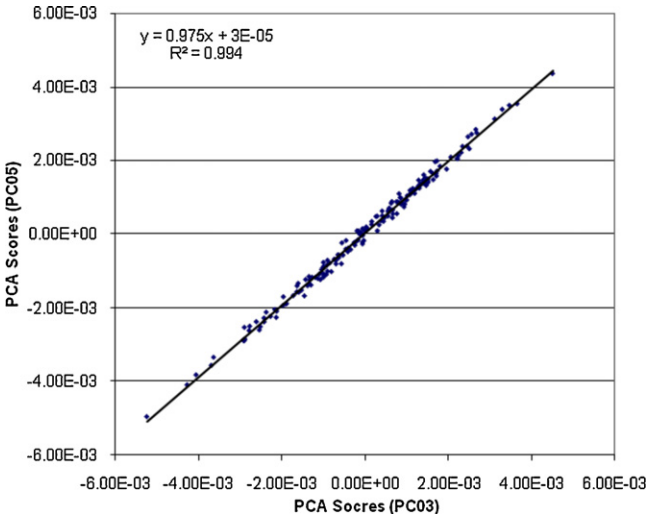

Fig. 6. Correlation between the third and fifth sets of PCA scores from the calibration and prediction set.

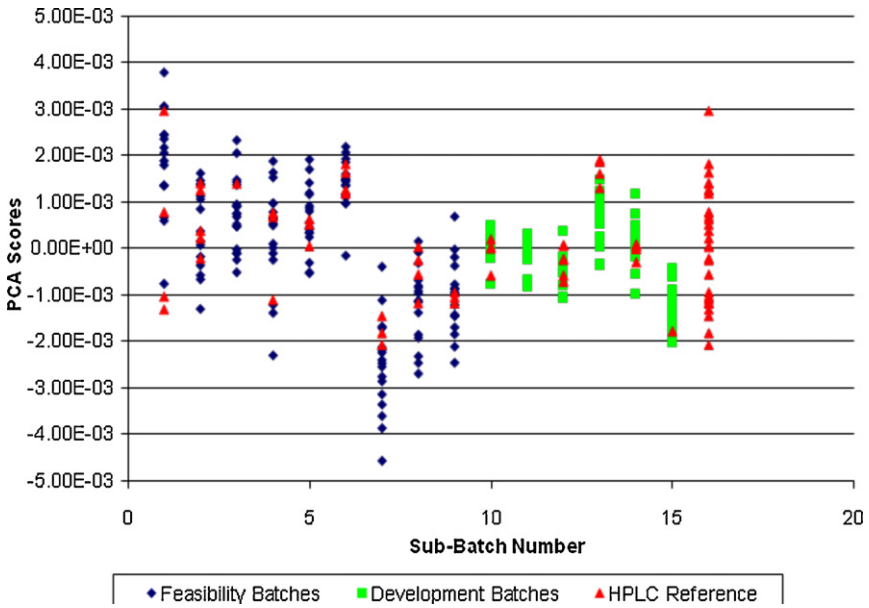

Fig. 7. PCA prediction of content uniformity for tablets from the development batches (in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and other excipient-related variances (e.g. lactose anhydrous versus monohydrate) rather than API concentration in the tablets. In early formulation development, the physical changes are not well controlled, which hinders the development of robust PLS models. On the other hand, the PCA model separates chemical changes from the physical ones and therefore it is more robust. Detailed discussions with regard to physical property changes and PLS regression modeling for low dose tablets are beyond the scope of this article and will be given in a separate publication.

4. Conclusions

This article demonstrates that a simple PCA algorithm can be used for content uniformity determination of solid formulations. The model appears particularly advantageous for analysis of low dose tablets during early stage formulation development. A PCA model can be generated by selecting a set of spectra that contain the suitable types (e.g. variances related to tablet hardness and API concentration) and amounts (e.g. a suitable API concentration range) of variances. The quantitative prediction by a PCA model is based on correct identification of the *X*-loadings and the establishment of a PCA scores versus API concentration relationship. Limited reference testing is needed to define the quantitative relationship and a concentration range. A PCA model may be less accurate compared with a PLS model but it appears to be more robust. Because NIR is a fast and non-destructive method, a large number of samples can be analyzed in a timely and cost-effective manner, which should result in better characterization of formulation development batches.

Acknowledgements

The authors would like to thank Smeet Deshmukh and Wayne Wang for preliminary formulation development work, as well as John Troisi for assisting with analyses of reference samples.

References

Alcala, M., Leon, J., Ropero, J., Blanco, M., Romanach, R., 2008. Analysis of low content drug tablets by transmission near infrared spectroscopy: selection of calibration rangers according to multivariate detection and quantitation limits of PLS models. J. Pharm. Sci. 97, 5318–5327.

Alcala, M., Ropero, J., Vazquez, R., Romanach, R., 2009. Deconvolution of chemical and physical information from intact tablets NIR spectra: two- and three-way multivariate calibration strategies for drug quantitation. J. Pharm. Sci. 98, 2747–2758.

Blanco, M., Villar, A., 2003. Development and validation of a method for the polymorphic analysis pharmaceutical preparations using near infrared spectroscopy. J. Pharm. Sci. 92, 823–830.

Blanco, M., Alcala, M., 2006. Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy. A contribution to process analytical technologies. Anal. Chim. Acta 557, 353–359.

Freitas, M.P., Sabadin, A., Silva, L.M., Giannotti, F.M., do Couto, D.A., Tonhi, E., Medeiros, R.S., Coco, G.L., Russo, V.F., Martins, J.A., 2005. Prediction of drug dissolution profiles from tablets using NIR diffuse reflectance spectroscopy: a rapid and nondestructive method. J. Pharm. Biomed. Anal. 39, 17–21.

Li, W., Worosila, G., Wang, W., Mascaro, T., 2005. Determination of polymorph conversion of an active pharmaceutical ingredient in wet granulation using NIR calibration models generated from the premix blends. J. Pharm. Sci. 94, 2800–2806.

Li, W., Bashai-Woldu, A., Ballard, J., Johnson, M., Agresta, M., Rasmussen, H., Hu, S., Cunningham, J., Winstead, D., 2007. Applications of NIR in early stage formulation development. Part I. Semi-quantitative blend uniformity analyses by reflectance NIR without calibration models. Int. J. Pharm. 340, 97–103.

Merckle, P., Kovar, K.A., 1998. Assay of effervescent tablets by near-infrared spectroscopy in transmittance and reflectance mode: acetylsalicylic acid in mono and combination formulations. J. Pharm. Biomed. Anal. 17, 365–374.

Norris, K.H., Ritchie, G.E., 2008. Assuring specificity for a multivariate near-infrared (NIR) calibration: the example of the Chambersburg Shoot-out 2002 data set. J. Pharm. Biomed. Anal. 48, 1037–1041.

Ritchie, G.E., Roller, R.W., Ciurczak, E.W., Mark, H., Tso, C., MacDonald, S.A., 2002. Validation of a near-infrared transmission spectroscopic procedure. Part B. Application to alternate content uniformity and release assay methods for pharmaceutical solid dosage forms. J. Pharm. Biomed. Anal. 29, 159–171.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., Jent, N., 2007. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J. Pharm. Biomed. Anal. 44, 683–700.

Xiang, D., Berry, J., Buntz, S., Gargiulo, P., Cheney, J., Joshi, Y., Wabuyele, B., Wu, H., Hamed, M., Hussain, A., Khan, M.A., 2009. Robust calibration design in the pharmaceutical quantitative measurements with near-infrared (NIR) spectroscopy: avoiding the chemometric pitfalls. J. Pharm. Sci. 98, 1155–1166.