Journal of Pharmaceutical and Biomedical Analysis 49 (2009) 619–626

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Evaluation of chemometric algorithms in quantitative X-ray powder diffraction (XRPD) of intact multi-component consolidated samples

Michael D. Moore[a], Robert P. Cogdill[b], Peter L.D. Wildfong[a,*]

[a] Duquesne University, Mylan School of Pharmacy, Graduate School of Pharmaceutical Sciences, 600 Forbes Av., Pittsburgh, PA 15282, USA
[b] University of Nebraska, College of Engineering, 114k Othmer Hall, P.O. Box 880642, Lincoln, NE 68588-0642, USA

ARTICLE INFO

Article history:
Received 12 September 2008
Received in revised form 2 December 2008
Accepted 5 December 2008
Available online 13 December 2008

Keywords:
XRPD
Chemometrics
Compacts
Multi-variate analysis

ABSTRACT

Quantitative X-ray powder diffraction (XRPD) data obtained from intact, consolidated samples affords the opportunity to analyze mixtures that simulate pharmaceutical drug products without the need for reversion back to powders; an analytical preparation step that destroys the contextual solid-state information intrinsic to intact consolidated samples. Traditional, standardless quantitative methods generally involve sophisticated pattern refinement procedures (e.g., Rietveld refinement) and are limited to crystalline materials. Methods that incorporate an internal standard are not optimal for compact analysis, and may often be susceptible to prediction errors associated with intensity attenuation. Chemometric-based XRPD utilizes full-pattern methods that combine analyses of both Bragg diffraction and diffuse scatter, thereby allowing for enhancement of signal-to-noise, sensitivity, and selectivity. Classical least-squares (CLS) regression, principal components regression (PCR) and partial least squares (PLS) regression are three chemometric algorithms commonly employed in spectroscopy. In the present work, quantification of a consolidated four-component system, composed of two crystalline materials and two disordered materials was analyzed intact, using two different XRPD optics geometries. Calibrations constructed for the prediction of individual constituent concentrations using the aforementioned three multi-variate algorithms were statistically compared with traditional diffraction–absorption univariate calibration. PLS regression modeling of data collected in transmission geometry provided the best statistical results for the quantification of constituent concentration. Further, this calibration was minimally affected by diffraction pattern anomalies traditionally corrected prior to phase quantification.

1. Introduction

X-ray powder diffraction (XRPD) of intact, consolidated samples is an important technique for pharmaceutical materials and drug product characterization. Quantitative XRPD methods are the most universal of the structure-sensitive methods for phase analysis of multi-phase systems [1,2]. The foundation of quantitative XRPD techniques for analyzing mixtures was originally developed by Alexander and Klug [3], and some examples applied to pharmaceutical mixture analyses are reported in the literature [4–9].

Traditional quantitative XRPD methods can be subcategorized into those which incorporate a standard and those which are standardless. Standardless techniques, such as whole-pattern fitting and Rietveld refinement, are particularly useful because they permit quantification without the use of specific calibration standards. These methods have the ability to assess and account for physical phenomena (such as preferred orientation); however, selecting an appropriate function, is not trivial [10]. Standardless quantification has also been limited to uses with crystalline materials. Further, they require a known crystal structure, which is often obtained from the literature/online databases.

Numerous quantitative XRPD methods exist to incorporate a standard material into mixtures for analysis [11], the use of which permits correction for matrix absorption. The most commonly employed of these is the internal standard technique. Incorporation of a standard within a sample, however, prevents analysis of intact, marketable drug products, for which the inclusion of an analytical dopant would be prohibited. Moreover, finding a standard that is stable, has approximately the same absorption characteristics as the sample, does not exhibit preferred orientation, and possesses isolated peaks which do not overlap with sample peaks are all potential barriers to use this method [11].

Diffraction–absorption calibrations are created through empirical linear regression modeling of component concentration against single diffraction peak intensity, a few diffraction peak intensities, or the integrated area under diffraction peak(s). These techniques assume constant mass attenuation across the sample, and therefore, are susceptible to errors resulting from diffraction intensity attenuation. Anisotropic intensity reduction due to microabsorption and extinction often result in non-linearities observed in

* Corresponding author. Tel.: +1 412 396 1543; fax: +1 412 396 4660.
E-mail address: wildfongp@duq.edu (P.L.D. Wildfong).

0731-7085/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2008.12.007

quantitative XRPD calibrations [3,10]. The largest contributor to intensity attenuation, however, is preferred orientation, which can be particularly problematic in consolidated samples [10]. Problems with preferred orientation in powdered samples are traditionally resolved by optimizing the particle size used for XRPD analysis, which may add significantly to method development and execution [8], as well as the potential for induction of experimental artifact owing to conversions elicited by triturative particle sizing [2].

By using only Bragg diffraction intensities, the traditional methods are limited to quantification of crystalline materials. Many solid oral dosage forms are developed using excipient materials having no long-range order. The physicochemical benefits afforded by incorporating amorphous active pharmaceutical ingredients in formulations of solid oral dosage forms have resulted in increased interest in the use of these systems [12]; however, for XRPD to be sufficiently sensitive to quantify disordered materials, a method that models both diffuse scatter and Bragg diffraction is required.

Many of the aforementioned issues are addressed through the use of multi-variate calibrations [13]. In quantitative chemometrics, a multi-variate relationship is sought between input data (e.g., diffraction patterns) and output data (e.g., composition, various physical phenomena). Applications of chemometric-based analyses in XRPD are starting to surface in the literature. Examples of applications of partial least squares (PLS) regression [14], principal components analysis (PCA) [15,16], and advanced techniques such as artificial neural networks (ANN) [17,18] to XRPD data have all been reported. The possible benefits of incorporating entire diffraction patterns into empirical models include enhanced signal-to-noise, analyte sensitivity and selectivity [19]. The objective of this work was to investigate applications of different multi-variate calibrations used to quantify both crystalline and amorphous components in consolidated samples. Intact compact analysis using XRPD was used to determine which algorithm was most suitable for intact quantification of multi-phase consolidated pharmaceutical systems.

Quaternary mixtures composed of two crystalline materials and two disordered materials were compressed at multiple compaction pressures and subjected to intact XRPD analysis using two different instrumental optics setups. Quantitative calibrations were created using the traditional (univariate) diffraction–absorption technique, and three multi-variate algorithms commonly employed in spectroscopy. Calibration linearity, precision, and prediction error were calculated for assessing model suitability.

2. Materials

Four-component mixtures comprised of anhydrous theophylline (Lot No. 92577, Knoll AG, Ludwigshafen, Germany), Lactose 316 Fast Flo NF Monohydrate (Lot No. 8502113061, Hansen Labs, New Berlin, WI), microcrystalline cellulose (Avicel PH 200, Lot No. M427C, FMC BioPolymer, Mechanicsburgh, PA), and soluble starch GR (Lot No. 39362, EMD Chemicals, Inc., Gibbstown, NJ) were prepared. The design matrix was fully balanced for compact concentration, having values ranging from 0 to 60% (w/w). Approximately 800 mg of each mixture was compressed at 67.0 MPa, 117.3 MPa, 167.6 MPa, 217.8 MPa, and 268.1 MPa using a single station Carver Press (Carver, Inc., Model 3887.1SDOA00, Wabash, IN) equipped with a 13-mm flat-faced punch.

3. Methods

3.1. X-ray powder diffraction analysis of the four-component compacts

XRPD data were collected using an X'Pert Pro MPD system (PANalytical B.V., Alemlo, the Netherlands) equipped with a copper anode ($\lambda = 1.5406$ Å), programmable divergence slit, and X'Celerator™ detector. The operational voltage and amperage were set to 45.0 kV and 40.0 mA, respectively, and diffraction patterns were acquired using an angular step size of 0.02° 2θ over a range of 2–60° 2θ. Data were collected with the instrument set in both Bragg–Brentano reflectance geometry (equipped with a spinning sample stage) and transmission geometry (equipped with a vertical spinner sample stage with the sample sandwiched between X-ray transparent kapton film), optically fitted with an auxiliary elliptical mirror used to expose the sample to quasi-parallel beam radiation.

All chemometric routines were performed in the Matlab programming environment (v7.1, MathWorks, Natick, MA) using the PLS_Toolbox (v3.0, Eigenvector Research, Manson, WA), together with several analysis routines developed in-house.

3.2. Data preprocessing

Prior to the application of chemometric algorithms, sample diffraction patterns were corrected for anisotropic peak (axis) shift using an iterative program that tested for correlations between a reference pattern and the sample pattern as a function of incremental calculated shifts. Corrected shifts that resulted in the highest correlation were ultimately selected. Additionally, the diffraction pattern of an empty sample holder (geometry dependent) was used to correct for scatter from that sample holder by pattern subtraction. Sample absorption correction was also performed according to published equations specific to the geometric setup of the instrument [20], and finally, a modified version of the Rachinger method [21] was used to mathematically correct for $K_{\alpha 2}$ radiation.

3.3. Inverse least squares (ILS) regression

Inverse least squares regression assumes that component concentration is a function of diffraction intensity. The standard ILS univariate model for a single component is calculated by:

$$y = xb + e \tag{1}$$

where y is the reference concentration value, b is the regression coefficient, x is the diffraction value at a specific peak position (i.e., a single peak intensity, a calculated peak area, ratio of intensities, etc.), and e is the error, which is assumed to be attributed to the concentration values. One of the features of ILS regression is that quantitative analysis can be performed even when the concentration of only a single component is known in the calibration mixture. In multiple linear regression (MLR), several independent variables are used, and the regression vector is calculated as follows (in matrix notation):

$$\mathbf{B} = \frac{\mathbf{X}'\,\mathbf{Y}}{\mathbf{X}'\,\mathbf{X}} \tag{2}$$

where $\mathbf{Y}$ is a matrix of concentration values, $\mathbf{X}$ is a matrix of intensity values, $\mathbf{B}$ is an matrix of regression coefficients. MLR is limited, in that the number of variables selected cannot exceed or equal the number of samples in the calibration set. For example, a calibration set having $n_{vars} = n_{samples}$ only possesses enough statistically independent information to estimate the mean; a larger sample set provides additional degrees of freedom to estimate other statistical parameters. Although statistical methods are available to assist with variable selection (i.e., stepwise regression), calibrations created with suboptimal selections can introduce modeling errors such as noise inflation from collinearity and over-fitting [22].

3.4. Classical least squares (CLS) regression

Classical least squares regression is often used in spectroscopic quantitative modeling owing to its agreement with Beer's law.

Under some conditions, XRPD intensity will be a linear function of the number of diffracting planes present, which qualifies the application of CLS regression in XRPD data modeling. CLS assumes a linear combination of pure component sensitivities, where each component is weighted by concentration. Unique to this technique is the feature that the regression vectors can be used as estimates of the pure component diffraction patterns. CLS regression models can be generated using XRPD data provided all reference constituent concentrations are known (Eq. (3)):

$$\mathbf{X} = \mathbf{KY} + \mathbf{E}_c \tag{3}$$

where **X** is a matrix of diffraction intensities, **K** is a matrix containing the regression vectors (pure component estimates), **Y** is a matrix of concentration values for all constituents, and $\mathbf{E}_c$ is the error matrix. Model error is attributed to diffracted intensities (i.e., microabsorption, extinction, preferred orientation, etc.). The calculation of the regression vector (estimated pure components) is as follows:

$$\mathbf{K} = \mathbf{Y}^{+}\bullet\mathbf{X} \tag{4}$$

where the superscript "+" indicates the Moore–Penrose pseudoinverse. The effectiveness of CLS regression in multi-variate modeling is limited when the concentrations of all constituents contributing to signal (in the case of XRPD, diffract or cause diffuse scatter) are not known, or when non-linearities are present [22].

3.5. Principal components regression (PCR)

Principal components regression is the ILS regression of "scores" calculated from principal components analysis against a dependent variable(s). The objective of PCA is to maximize the variation contained in a data matrix with as few underlying "factors" as possible [23]. The PCA model has the form of

$$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E} \tag{5}$$

where **X** is a data matrix of k sample-independent variables (diffraction intensities), for n samples, **T** is a matrix of component "scores" in vector form for each component, **P** is a matrix of component "loadings," and **E** is a matrix containing the residuals (the prime indicates the transpose of the matrix). Principal component scores and loadings are calculated using singular value decomposition [24]. The algorithm proceeds through a least squares fit of a straight line through the data in a k-dimensional hyperspace [25]. New components are fitted to the data until an adequate amount of the variance has been explained. In PCR, the dependent variable(s) (i.e., reference concentration values) are linearly regressed against the scores. The regression coefficients are calculated as follows:

$$\mathbf{B} = \frac{\mathbf{T}'\,\mathbf{Y}}{\mathbf{T}'\mathbf{T}}\bullet\mathbf{P} \tag{6}$$

where **Y** is a matrix of concentration values. A full-length regression vector is obtained by projecting the regression coefficients onto the loading vector(s). Calibrations may be created with fewer samples than variables. Reviews of PCA and PCR can be found elsewhere in the literature [22,25].

3.6. Partial least squares regression

The PLS and PCR algorithms, being factor-based analysis methods, have similar goals. The objective of PLS is to maximize the amount of co-variation explained between dependent variables and independent variables using the fewest number of factors. Contrary to PCA where only the independent variables are used to calculate the basis set, PLS incorporates both dependent and independent variables in the basis set calculation. It may be envisioned that this method is particularly powerful when the dependent variable set contains low noise. Generally speaking, PLS explains the total variance in fewer factors relative to PCR. In this study, the *SIMPLS* algorithm was used to directly calculate factors as linear combinations of the original variables, while constrained to orthogonality and normalization restrictions [26]. The PLS algorithm used herein [27], and representative reviews can be found elsewhere [22,25].

4. Results

The fully balanced concentration design matrix used for this study is shown in Table 1 and contains five concentration points (0, 20, 25, 40, 60%, w/w) for each of the four constituents. The concentration points were added to the design matrix in a one-by-one fashion, followed by a calculation of the matrix covariance; each point was adjusted to minimize this covariance. It should be noted that separate experimentation concerning instrument sensitivity, selectivity, and signal-to-noise has been previously reported in Moore et al., [19]. At each concentration point, mixtures were compacted using five different pressures (67.0 MPa, 117.3 MPa, 167.6 MPa, 217.8 MPa, and 268.1 MPa), resulting in a calibration sample size of 145 compacts. Recognizing that consolidation may induce changes to diffraction intensity [9,28] and diffuse scatter, samples compacted at different pressures were included in the calibration set to incorporate pattern variability into the model approximation. An additional sample at each concentration point, compacted using a randomly assigned pressure, was used to test the calibrations.

Univariate calibrations were created by regressing a single intensity value for each individual component (i.e., the largest Bragg peak for crystalline materials and the largest overall intensity for the disordered materials) against constituent concentration (nominal value from the design). The PCR and PLS regression vectors for each constituent were estimated individually from a component-specific orthogonal basis set. Selection of the ideal number of principal components/latent variables (shown in Table 2) was performed

Table 1
Sample composition design matrix.

Tablet (#)	Theophylline (wt/wt)	Lactose (wt/wt)	MCC (wt/wt)	Starch (wt/wt)
1	0.600	0.200	0.200	0.000
2	0.400	0.400	0.200	0.000
3	0.200	0.600	0.200	0.000
4	0.400	0.200	0.400	0.000
5	0.200	0.400	0.400	0.000
6	0.200	0.200	0.600	0.000
7	0.600	0.200	0.000	0.200
8	0.400	0.400	0.000	0.200
9	0.200	0.600	0.000	0.200
10	0.600	0.000	0.200	0.200
11	0.400	0.200	0.200	0.200
12	0.200	0.400	0.200	0.200
13	0.000	0.600	0.200	0.200
14	0.400	0.000	0.401	0.200
15	0.200	0.200	0.400	0.200
16	0.000	0.400	0.400	0.200
17	0.200	0.000	0.600	0.200
18	0.000	0.200	0.600	0.200
19	0.400	0.200	0.000	0.400
20	0.200	0.400	0.000	0.400
21	0.400	0.000	0.200	0.400
22	0.200	0.200	0.200	0.399
23	0.000	0.400	0.200	0.400
24	0.200	0.000	0.400	0.400
25	0.000	0.200	0.400	0.400
26	0.200	0.200	0.000	0.600
27	0.200	0.000	0.200	0.600
28	0.000	0.200	0.200	0.600
29	0.250	0.250	0.250	0.250

Table 2
Selected statistical values for the different quantitative PXRD calibrations relative to the collection geometry.

Method	Component	LV	Transmission			Reflectance		
			R^2	SEP (%)	Precision (%)	R^2	SEP (%)	Precision (%)
Univariate	Theophylline	N/A	0.961	3.79	0.00024	0.898	5.09	0.09491
	Lactose		0.927	5.70	0.00081	0.911	4.95	0.03195
	MCC		0.809	9.70	0.00324	0.940	5.81	0.04044
	Starch		0.292	15.30	0.00935	0.006	18.18	0.07608
CLS	Theophylline	N/A	0.972	2.90	0.00027	0.942	3.57	0.16771
	Lactose		0.963	3.48	0.00076	0.939	3.63	0.12726
	MCC		0.955	3.69	0.00237	0.958	5.47	0.07263
	Starch		0.946	5.24	0.00413	0.866	4.91	0.01797
PCR	Theophylline	1	0.972	2.92	0.00027	0.926	4.81	0.07780
	Lactose	3	0.968	3.16	0.00062	0.919	4.82	0.02580
	MCC	4	0.751	9.77	0.00346	0.875	7.49	0.03176
	Starch	5	0.941	4.98	0.00133	0.868	5.00	0.10102
PLS	Theophylline	1	0.972	2.91	0.00028	0.933	4.49	0.08007
	Lactose	3	0.978	2.48	0.00110	0.943	3.71	0.02915
	MCC	3	0.955	3.84	0.00210	0.937	6.23	0.03226
	Starch	3	0.950	4.88	0.00256	0.804	7.15	0.07829

through minimization of both the root mean square error of calibration (RMSEC) and the root mean square error of cross validation (RMSECV).

Traditionally, it is not common to use a univariate XRPD calibration to predict the concentration of a disordered component in a mixture; however, it was performed in this study as a means to illustrate the practicality of different multi-variate methods. In general, data collected in transmission geometry provided better linearity, precision and lower concentration prediction error relative to reflectance geometry for both the crystalline and disordered components. In transmission geometry, the entire sample volume is irradiated, whereas only a fraction of the sample is interrogated in reflectance analysis. Therefore, data collected in transmission mode is less susceptible to errors derived from analyzing inhomogeneous "regions" within a sample matrix.

The data in Table 2 indicate that the standard error of prediction (SEP) for theophylline in most instances is greater when using data collected in reflectance geometry relative to that in transmission. This indicates a correlation between instrument geometry and prediction error; more than likely a result of the diffraction pattern anomalies related to specific analytical optics setup modes. Barring microabsorption, extinction, and other anomalies, regression vectors should mimic pure component diffraction patterns. In Fig. 1, the pure component pattern and the calculated regression vectors for all three multi-variate theophylline prediction models are shown for data collected in reflectance geometry. There is good agreement between all regression vectors and the pure component pattern, as would be expected. For PCR and PLS, negative correlations are observed at ~18° 2θ, which specifically corresponds with the location of Bragg peaks resulting from lactose monohydrate diffraction. Therefore, the PCR and PLS models are sensitive to the changes in both theophylline and lactose concentration at this angle. In Fig. 2, the theophylline regression vectors calculated from data collected using transmission geometry are identical to the pure component scan. Further, the negative correlations observed at ~18° 2θ in the reflectance geometry regression vectors are nearly absent from transmission data. The calculated regression vectors from the transmission geometry are, therefore, more highly correlated to only changes in theophylline concentration, thereby resulting in enhanced theophylline sensitivity and decreased prediction errors.

When considering disordered materials, the diffuse scatter that produces the characteristic "amorphous halo" may not be linearly related to constituent concentration. Non-linear relationships may manifest as negative correlations in regression vectors as a result of modeling the decrease in concentration of other components. In Fig. 3, the calculated regression vectors for the three multi-variate starch prediction models and the corresponding pure component scan are shown for data collected in reflectance geometry. The CLS regression vector bears the highest correlation with the pure

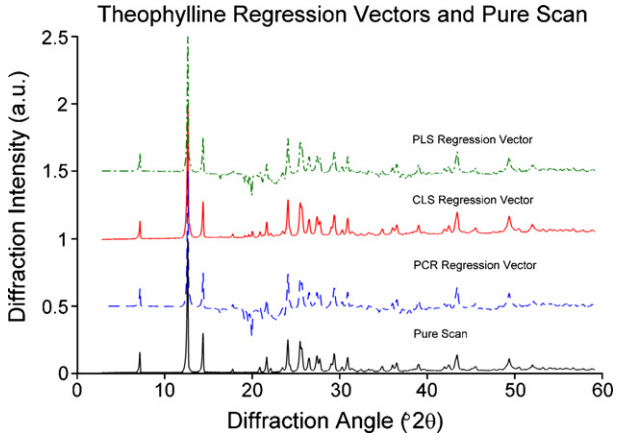

Fig. 1. Calculated regression vectors for each multi-variate calibration used in the prediction of anhydrous theophylline from data collected in reflectance geometry.

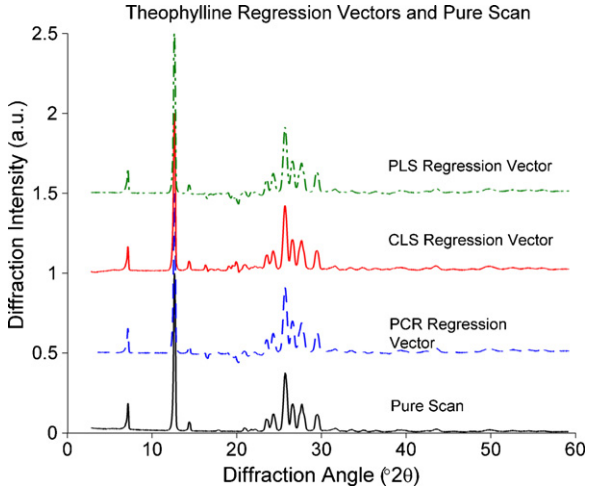

Fig. 2. Calculated regression vectors for each multi-variate calibration used in the prediction of anhydrous theophylline from data collected in transmission geometry.

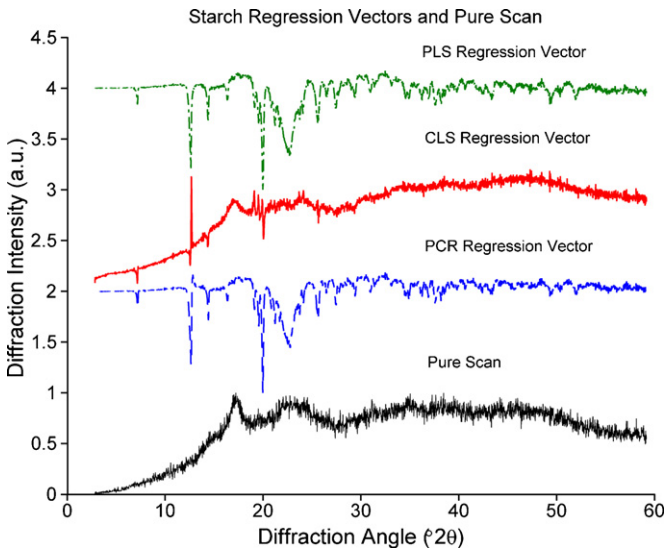

Fig. 3. Calculated regression vectors for each multi-variate calibration used in the prediction of starch from data collected in reflectance geometry.

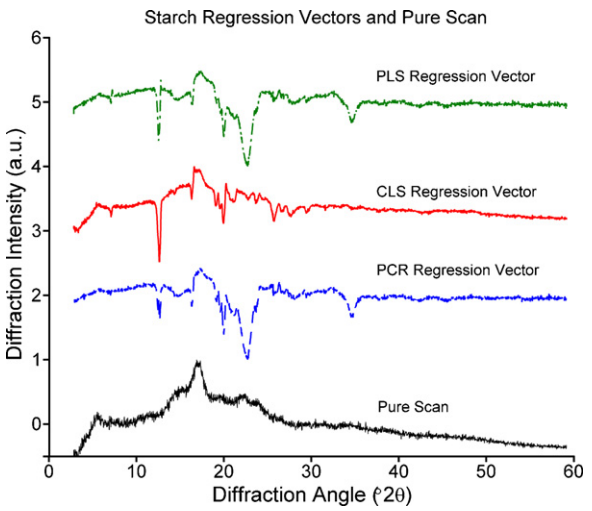

Fig. 4. Calculated regression vectors for each multi-variate calibration used in the prediction of starch from data collected in transmission geometry.

component diffraction pattern, thereby affording enhanced linearity and decreased prediction error relative to PCR and PLS (from Table 2). Although they possess the same positive correlations as the starch pure component scan, the regression vectors from both PCR and PLS once again exhibit negative correlations attributable to changes in the concentrations of other mixture components at angles corresponding to their respective peaks of principal diffraction: theophylline (~12° 2θ), lactose (~18° 2θ) and MCC (~22° 2θ). This same trend is also observed for the transmission data, as shown in Fig. 4, where both the PLS and PCR regression vectors exhibit the same negative correlations seen in reflectance. Additionally, the CLS regression vector of transmission data representing diffraction by the starch component possesses negative correlations at ~12° 2θ and ~18° 2θ. These regression vector similarities resulted in near-equal linearity and prediction error for the PCR, PLS, and CLS starch calibrations created from data collected in transmission. The negative correlations observed in the disordered component regression vectors, therefore, could quite possibly be a result of the non-linear relationship between constituent concentration and diffuse scatter intensity.

In Table 3, linearity and SEP are reported as a function of applied pattern corrections using data collected in reflectance geometry. Statistics are reported as an average of the two crystalline components and an average of the two disordered components. The importance of correcting for axis shift prior to creating a univariate calibration in the prediction of crystalline components is supported by the observed increased linearity and decreased prediction error. When modeling a single intensity, or the area of a single peak, anisotropic peak distortions may build errors into the calibration and result in inaccurate predictions. Interestingly, in multi-variate calibrations (such as PLS) the linearity and predictability are not significantly affected by corrections to the powder patterns relative to univariate modeling. By modeling multiple intensities having correlation with constituent concentration, the anisotropy of error-related variance is compensated for through increased correlation at unaffected variables. Further, through maximization of the explained covariance between diffraction intensity and constituent concentration, variability due to anisotropic peak aberrations is modeled in the regression vectors.

Table 4 shows the effects of XRPD pattern corrections on linearity and prediction error using data collected in transmission geometry. Comparison with Table 3 (reflectance geometry collection) indicates that performance statistics are superior for transmission data, particularly for disordered (weakly diffracting) materials. This is not unexpected, given that transmission experiments interrogate the entire compact sample volume. Although

Table 3
The effects of various powder pattern corrections on linearity and the standard error of prediction of the univariate and PLS calibrations for prediction of both crystalline and disordered components using data collected by reflectance geometry.

	No corrections	Axis shift	$K_{\alpha 2}$ subtraction	Absorption correction	Background subtraction	All
R^2						
Univariate						
Crystalline	0.723	0.898	0.759	0.729	0.723	0.904
Disordered	0.493	0.489	0.484	0.486	0.493	0.473
PLS						
Crystalline	0.906	0.936	0.906	0.908	0.906	0.938
Disordered	0.864	0.890	0.866	0.849	0.864	0.871
SEP (%)						
Univariate						
Crystalline	7.61	5.59	6.75	7.34	7.61	5.02
Disordered	11.41	11.19	11.54	11.76	11.41	11.99
PLS						
Crystalline	4.39	4.45	4.40	4.11	4.39	4.10
Disordered	5.97	6.40	5.98	6.42	5.97	6.69

Table 4
The effects of various powder pattern corrections on linearity and the standard error of prediction of the univariate and PLS calibrations for prediction of both crystalline and disordered components using data collected by transmission geometry.

	No corrections	Axis shift	$K_{\alpha 2}$ subtraction	Absorption correction	Background subtraction	All
R^2						
Univariate						
Crystalline	0.940	0.947	0.937	0.939	0.941	0.944
Disordered	0.260	0.590	0.260	0.186	0.260	0.551
PLS						
Crystalline	0.968	0.972	0.968	0.971	0.968	0.975
Disordered	0.770	0.948	0.772	0.771	0.770	0.952
SEP (%)						
Univariate						
Crystalline	4.24	4.10	4.42	4.47	4.15	4.75
Disordered	16.30	10.72	16.30	16.49	16.30	12.50
PLS						
Crystalline	3.07	2.89	3.07	2.96	3.07	2.70
Disordered	8.07	4.17	8.10	8.17	8.07	4.36

the low mass attenuation coefficients of pharmaceutical materials permit some sample penetration (on the order of mm), reflectance experiments interrogate a much smaller volume compared to transmission experiments. For the crystalline components, both models are invariant to all of the applied corrections. However, the SEP for the disordered components using PLS shows significant dependence on axis shift correction. Fig. 5 illustrates the effects of applied pattern corrections on the calculated PLS regression vectors. The calculated regression vectors of raw diffraction patterns, patterns corrected for absorption only, patterns corrected for $K_{\alpha 2}$ only, or background subtraction only are highly correlated to one another. Further, these regression vectors exhibit differences from the regression vector calculated from patterns corrected for axis shift. The dashed boxes highlight the main differences observed between the aforementioned group of regression vectors and those for which axis shift was corrected. These differences represent uncertainties correlated to peak aberrations and result in modeling concentration changes of other components. Ultimately, the prediction error was larger due to the error in modeling the concentration changes in multiple components.

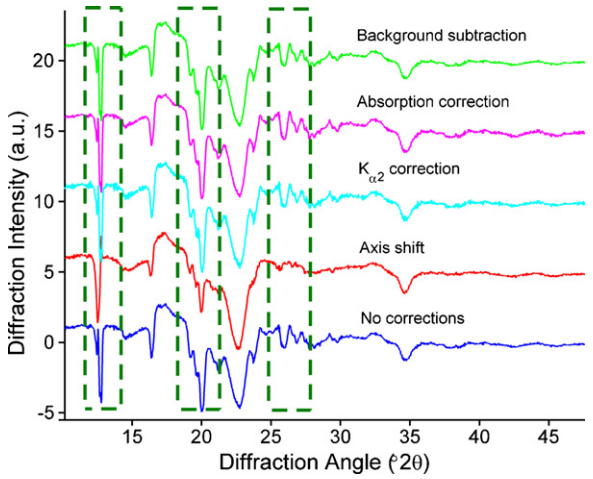

Fig. 5. The effects of various corrections on the PLS calibration regression vectors as applied to data collected in transmission geometry.

For many spectroscopic methods, building quantitative models that include variability associated with compression force is imperative. However, the variability in diffraction intensity observed to have resulted from consolidation was minimal because the materials used in this study have low mass-attenuation coefficients and require few applied absorption corrections. Admittedly, creating a calibration and test sample matrix the size of the set presented in this work ($n = 174$) may be impractical in a multi-product industrial development group. Given the results of the present data, however, efficient calibrations could have been created using fewer samples. This was tested by compressing a single sample from each

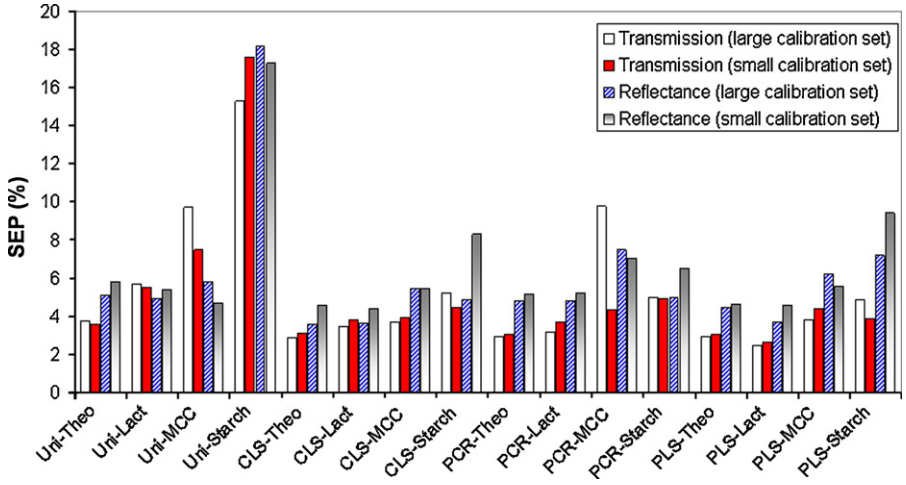

Fig. 6. The effects of the number of samples in the calibration set on the standard error of prediction for both collection geometries.

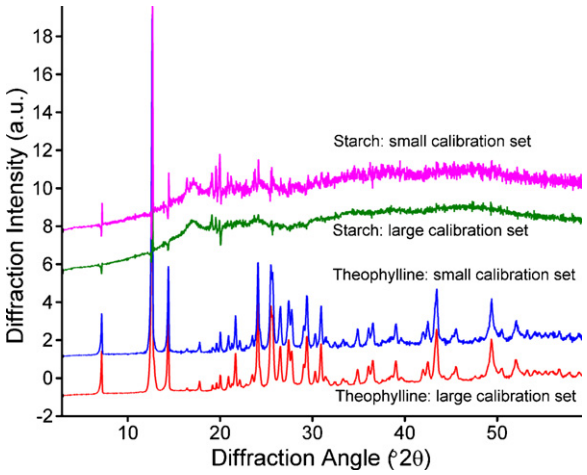

Fig. 7. The effects of the number of samples in the calibration set on the starch and theophylline CLS regression vectors for data collected from reflectance geometry.

concentration point (Table 1) at 268.1 MPa in order to re-create the calibrations (29 samples total), while the remaining samples (n = 145) were used to test the models generated from this reduced data set. The graph in Fig. 6 shows prediction error as a function of component, as modeled by each calibration method for differently sized calibration sets relative to the experimental optics utilized. Linearity statistics (not shown) were not affected by the reduction of the dataset used for calibration. The data in Fig. 6 illustrate that prediction error associated with transmission geometry is relatively unaffected by the size of the calibration set employed. Reflectance geometry, however, indicates that a slight increase in prediction error occurs when using the smaller calibration set, possibly due to the decreased signal-to-noise.

The starch and theophylline calculated CLS regression vectors for reflectance and transmission geometry are shown in Figs. 7 and 8, respectively. When comparing the starch regression vectors as calculated from data collected in reflectance geometry, the smaller calibration set vector exhibits increased noise (rougher) and larger peak/trough correlations ultimately attributable to increased uncertainty. Though the starch regression vector (as calculated from data collected in transmission geometry) for the smaller calibration set is "rougher" relative to the vector from the larger set, the overall shape is the same for both. The regression vectors calculated from the smaller calibration set using data collected in transmission mode demonstrated increased correlation to the regression vectors calculated from the larger calibration set relative to the reflectance data. Therefore, this enhanced correlation results in nearly identical error statistics independent of calibration sample size (seen in Fig. 6).

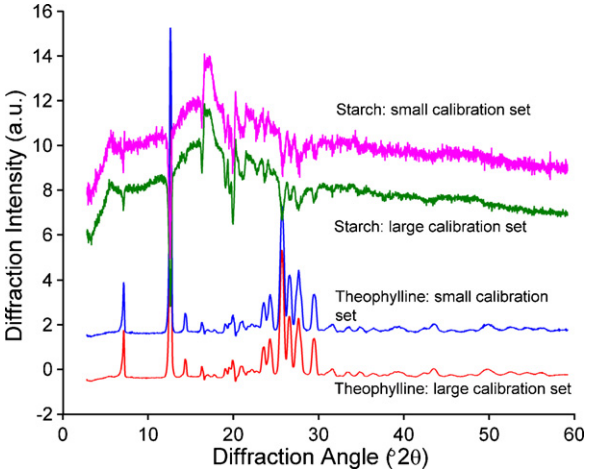

Fig. 8. The effects of the number of samples in the calibration set on the starch and theophylline CLS regression vectors for data collected from transmission geometry.

Based on the results presented herein, calibrations via PLS, created from the data collected in transmission geometry can be recommended as optimal for quantitative XRPD of similar systems. The calibrations created using this algorithm exhibited optimal linearity values of 0.972, 0.978, 0.955, and 0.950 for theophylline, lactose, MCC, and starch, respectively. The calibrations created in the present work stem from a quaternary design using the design matrix concentration values. Given that each nominal design value is likely to differ slightly from the actual concentration, each component reference concentration value has some random error. Overall, a cumulative 2–3% error in content uniformity of all the constituents may be anticipated; linearity values approaching 0.98 in this design are, therefore, statistically acceptable. Additionally, minimum prediction error and enhanced precision for both crystalline and disordered components in complex, intact compacts was achieved. Furthermore, PLS models were observed to be less susceptible to errors associated with diffraction pattern anomalies and effects related to the size of the calibration set.

5. Conclusions

The ability to analyze intact compacts makes XRPD an important analytical tool for non-destructive pharmaceutical characterization. The structure-sensitivity afforded by XRPD enables quantitative applications having the ability to discriminate between different chemical components, polymorphs, and other phase mixtures. It has been shown in this work that the traditional univariate calibrations are affected by peak distortion, variable selection, and applied powder pattern corrections. Multi-variate calibrations, however, provided enhanced linearity, decreased prediction errors, and exhibited less susceptibility to errors attributable to peak distortions relative to single-point calibrations. Further, calibration errors related to pattern anomalies were minimized through empirical modeling of the entire diffraction pattern (i.e., both Bragg diffraction and diffuse scatter intensities). As an increased amount of mixed amorphous/crystalline systems are formulated into dosage forms, the need for discriminative and sensitive quantitative analytical tools for intact analysis will become more prevalent.

References

[1] B.D. Cullity, S.R. Stock, Elements of X-ray Diffraction, third ed., Prentice Hall, Upper Saddle River, 2001.
[2] R.L.S. Ron Jenkins, in: J.D. Winefordner (Ed.), Chemical Analysis, vol. 138, Wiley-Interscience, New York, 1996.
[3] L. Alexander, H.P. Klug, Anal. Chem. 20 (1948) 886–889.
[4] A.K. Dash, K. Khin-Khin, R. Suryanarayanan, J. Pharm. Sci. 91 (2002) 983–990.
[5] I. Fix, K.J. Steffens, Drug Dev. Ind. Pharm. 30 (2004) 513–523.
[6] R. Suryanarayanan, C.S. Herman, Pharm. Res. 8 (1991) 393–399.
[7] R. Suryanarayanan, C.S. Herman, Int. J. Pharm. 77 (1991) 287–295.
[8] M. Tiwari, G. Chawla, A.K. Bansal, J. Pharm. Biomed. Anal. 43 (2007) 865–872.
[9] P.L.D. Wildfong, et al., J. Pharm. Biomed. Anal. 39 (2005) 1–7.
[10] S. Yamamura, Y. Momose, Int. J. Pharm. 212 (2001) 203–212.
[11] L.S. Zevin, G. Kimmel, in: I. Mureinik (Ed.), Quantitative X-ray Diffractometry, Springer, New York, 1995.
[12] B.C. Hancock, G. Zografi, J. Pharm. Sci. 86 (1997) 1–12.
[13] H. Martens, T. Næs, Multivariate Calibration, John Wiley and Sons, New York, NY, USA, 1989.
[14] M. Suda, K. Takayama, M. Otsuka, Anal. Sci. 24 (2008) 451–457.
[15] A.C. Jorgensen, et al., J. Pharm. Sci. 95 (2006) 906–916.
[16] P.R. Nassab, R. Rajko, P. Szabo-Revesz, J. Pharm. Biomed. Anal. 41 (2006) 1191–1197.
[17] S. Agatonovic-Kustrin, et al., J. Pharm. Biomed. Anal. (2000) 985–992.
[18] T. Okumura, et al., Colloids Surf. B. Biointerfaces 49 (2006) 153–157.
[19] M.D. Moore, et al., J. Pharm. Biomed. Anal. 47 (2008) 238–247.

[20] T. Egami, S.J.L. Billinge (Eds.), Pergamon Materials Series, vol. 7, Pergamon, Elsevier, Oxford, 2003.
[21] R. Delhez, E.J. Mittemeijer, J. Appl. Crystallogr. 8 (1975) 609–611.
[22] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1988) 1193–1202.
[23] H. Hotelling, J. Educ. Psychol. 24 (1933), pp. 417–441, 498–520.
[24] E. Anderson, et al., LAPACK User's Guide, 1999.
[25] J. Gabrielsson, N.-O. Lindberg, T. Lundstedt, J. Chemometr. 16 (2002) 141–160.
[26] I.S. Helland, Chemometr. Intell. Lab. Syst. 58 (2001) 97–107.
[27] S. de Jong, Chemometr. Intell. Lab. Syst. 18 (1993) 251–263.
[28] W. Cao, et al., J. Pharm. Biomed. Anal. 30 (2002) 1111–1119.