Available online at www.sciencedirect.com
ScienceDirect
Journal of Pharmaceutical and Biomedical Analysis 47 (2008) 238–247
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

The use of net analyte signal orthogonalization in the separation of multi-component diffraction patterns obtained from X-ray powder diffraction of intact compacts

Michael D. Moore, Robert P. Cogdill, Steven M. Short, Colleen R. Hair, Peter L.D. Wildfong *

Duquesne University, Mylan School of Pharmacy, Graduate School of Pharmaceutical Sciences, 600 Forbes Avenue, Pittsburgh, PA 15282, USA

Received 2 October 2007; received in revised form 19 December 2007; accepted 21 December 2007
Available online 8 January 2008

Abstract

X-ray powder diffraction (XRPD) analysis of intact multi-component consolidated mixtures has significant potential owing to the ability to non-destructively quantify and discriminate between solid phases in composite bodies with minimal sample preparation. There are, however, limitations to the quantitative power using traditional univariate methods on diffraction data containing features from all components in the system. The ability to separate multi-component diffraction data into patterns representing single constituents allows both composition as well as physical phenomena associated with the individual components of complex systems to be probed. Intact, four-component compacts, consisting of two crystalline and two amorphous constituents were analyzed using XRPD configured in both traditional Bragg–Brentano reflectance geometry and parallel-beam transmission geometry. Two empirical, model-based methods consisting of a multiple step net analyte signal (NAS) orthogonalization are presented as ways to separate multi-component XRPD patterns into single constituent patterns. Multivariate figures of merit (FOM) were calculated for each of the isolated constituents to compare method-specific parameters such as sensitivity, selectivity, and signal-to-noise, enabling quantitative comparisons between the two modes of XRPD analysis.

Keywords: XRPD; Chemometrics; Net analyte signal; Whole-compact analysis

1. Introduction

X-ray powder diffraction (XRPD) is a non-destructive analytical tool having numerous applications in pharmaceutical materials science [1–5]. Recent advances in XRPD techniques have enabled quantitative analyses of APIs in their final product form. Additionally, whole-compact *in situ* analysis has been demonstrated as a viable technique used to characterize solid-state structural phenomena and physical interactions of consolidated powders without further sample preparation [1,5]. Instrument and sample-specific signal aberrations such as diffraction feature attenuation resulting from the effects of consolidation, or resolving overlapping diffraction peaks due to geometrically associated low angular resolution complicate this method of analysis. Despite these obstacles, the analytical applications of XRPD on intact compacts have been successfully used for phase quantification [1,2], active pharmaceutical ingredient (API) identification in multi-component tablets [3], and the assessment of consolidation effects on various crystal habits [4]. Traditionally, instrument and sample related issues confounding quantitative and qualitative capabilities of XRPD analyses were handled via rigorous mathematical corrections based upon first principles of idealized diffraction theory [6]. As an alternative, chemometrics provides a multivariate approach for troubleshooting many quantitative difficulties associated with linear methods.

Chemometrics is defined by the International Chemometrics Society (ICS) as "... the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods.

* 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.2007.12.042

" In quantitative chemometrics, a multivariate relationship is sought between input data (e.g., diffraction patterns) and output data (e.g., composition, various physical phenomena). Therefore, quantitative chemometric tools are typically related to linear regression and function estimation. Some good reviews of chemometrics methods can be found elsewhere in literature [7–11].

Multivariate methods such as whole-pattern refinement have been in practice for many years [12]. Recently, some examples employing chemometric treatments of XRPD data have appeared in the literature. Applications of various preprocessing methods such as Fourier transforms, orthogonal signal correction, and wavelet transforms in improving regression model performance in quantitative XRPD have been demonstrated [13]. Principal component analysis (PCA) of XRPD data has been shown to be an effective tool for the screening of solid-state information such as degree of disorder, modeling of dehydration, and detection of anomalous data sets [14–17]. Many of the traditional univariate methods are now being displaced by more efficient and accurate multivariate techniques.

Phadnis et al. used a pattern subtraction technique in separating multi-component diffraction data for identification of active pharmaceutical ingredients in dosage forms [3]. One difficulty associated with this method is that specific knowledge of the crystalline component weight fraction is required. More importantly, pattern subtraction ignores the covariance structure of angular variables in diffraction patterns. The implication of the latter manifests in intensity reduction of angular variables correlated to both the subtracted component and the component of interest. Therefore, the separation may result in a loss of information correlated to the isolated component of interest, or "leakage" of spurious/incorrect diffraction features across pure component patterns.

Limitations exist concerning the types of analyses that can be performed on diffraction patterns containing features from all components in the system. It is imperative that the loss of information is minimized through all mathematical operations performed on the diffraction data. As more accurate methods for properly isolating features of diffraction patterns due to single constituents are developed, the amount of information retained for experimental analyses will also increase. This ultimately enables better quantitative modeling of solid-state related physical phenomena (i.e., the effects of consolidation on the structure of individual components). In the present work, two methods of isolating the diffraction features attributable to a single constituent are proposed. XRPD patterns of compacts were successfully separated into individual constituent patterns lacking the characteristic negative intensities associated with the inefficiencies of methods such as simple pattern subtraction. A complex quaternary model formulation was used for the study to provide a realistic chemometric challenge to the proposed methods. Materials were compressed into compacts consisting of two crystalline and two amorphous materials having overlapping diffraction signals. Further, multivariate figures of merit (FOM) were calculated for both traditional Bragg–Brentano reflectance geometry and transmission geometry allowing quantifiable differences in sensitivity, analytical sensitivity, selectivity and signal-to-noise to be examined.

Table 1
Concentration design matrix

Tablet #	Theophylline (w/w)	Lactose (w/w)	MCC (w/w)	Starch (w/w)
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

2. Materials and methods

Four-component mixtures comprising of anhydrous theophylline (Knoll AG, Ludwigshafen, Germany), Lactose 316 Fast Flo NF Monohydrate (Hansen Labs, New Berlin, WI), microcrystalline cellulose (Avicel PH 200, FMC BioPolymer, Mechanicsburgh, PA), and soluble starch GR (EMD Chemicals, Inc., Gibbstown, NJ) were prepared and compressed. The compact concentration design matrix was fully balanced with values ranging 0–60% (w/w) (Table 1). Approximately 800 mg was compressed at 67.0, 117.3, 167.6, 217.8, and 268.1 MPa using a single station Carver Press (Carver, Inc., Wabash, Indiana) equipped with a 13 mm flat-faced punch and a right-cylindrical die. Further information concerning sample preparation and supplemental quality tests can be found in Short et al. [18]. Compacts of the individual pure components, compressed at 67.0 MPa, were also made.

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

All XRPD experiments were performed using a X'Pert Pro MPD system (PANalytical B.V., Alemlo, The Netherlands)

equipped with a copper anode ($\lambda_{K\alpha}$ = 1.5406 Å), programmable divergence slit, and X'Celerator™ RTMS detector. The operational voltage and amperage were set to 45.0 kV and 40.0 mA, respectively. Diffraction patterns were acquired using an angular step size of 0.02° 2θ over a range of 2–60° 2θ. Analyses were performed using both Bragg–Brentano reflectance geometry (equipped with a horizontal spinning sample stage rotating at 16 rpm) and parallel-beam transmission geometry (equipped with a vertical spinner sample stage with the samples fixed between two films of Kapton® film). For transmission experiments, quasi-parallel X-rays were generated using an auxiliary elliptical mirror.

Chemometric analyses were performed in the Matlab programming environment (v7.1, MathWorks, Natick, MA) using the PLS_Toolbox (v3.0, Eigenvector Research, Manson, WA), together with many analysis routines developed in-house at Duquesne Center for Pharmaceutical Technology (DCPT, Pittsburgh, PA).

2.2. Data preprocessing

Prior to modeling, sample diffraction patterns were first corrected for continuous axis shift using an iterative program which tests for correlation between the sample pattern and a reference pattern as a function of incremental axis shifts. Ultimately, the corrected shift having the highest correlation with the reference pattern was selected. Although this effectively corrects continuous axis shifts, many anisotropic shifts are observed in XRPD, specifically those accompanying traditional Bragg–Brentano reflectance geometry [19,20]. Corrections for anisotropic peak aberrations were not applied. Following continuous axis shift correction, diffraction patterns were smoothed by applying a Savitsky-Golay filter (21 point, 2nd order polynomial) [21].

2.3. Classical least squares (CLS) regression

Classical least squares regression gained popularity in spectroscopy as a method for quantitative modeling due to its agreement with Beer's law. The fact that XRPD intensity is a linear function of the number of specific diffracting planes at a given diffraction angle (2θ), warrants the use of CLS regression for modeling of XRPD data. CLS assumes a linear combination of pure component sensitivities, where each component is weighted by concentration. A unique feature of CLS is that the regression vectors serve as estimates of the pure component diffraction patterns [22,23]. CLS regression models can be generated using XRPD data provided that the concentrations of all reference constituents are known. The standard CLS model is represented by:

$$\mathbf{X} = \mathbf{KY} + \mathbf{E}_c \tag{1}$$

where $\mathbf{X}$ is a matrix of diffraction intensities, $\mathbf{K}$ is a matrix of the regression vectors, $\mathbf{Y}$ is a matrix of concentration values for all constituents, and $\mathbf{E}_c$ is the error matrix. Model error is assumed, in this case, to be due to errors in diffracted intensities. The calculation of the regression vector (estimated pure components) is as follows:

$$\mathbf{K} = \mathbf{Y}^{+} \cdot \mathbf{X} \tag{2}$$

where superscript "+" indicates the Moore–Penrose pseudoinverse. The use of CLS regression in multivariate modeling is limited when the concentrations of all constituents that diffract are not known or when non-linearities are present.

2.4. Net analyte signal (NAS)

Net analyte signal was introduced in 1986 as a method for determining the portion of an instrument signal attributed to a specific component, and is estimated to be that which is orthogonal to the remaining interfering factors present within the data (e.g., diffraction from other components, noise, etc.). The net analyte signal for the kth constituent of a multi-component system, can be computed as the part of its diffraction pattern orthogonal to the contributions of the other constituents if it is assumed that the linearity assumptions of the CLS model hold [24]. The part of the vector $\mathbf{u}$ that is orthogonal to the matrix $\mathbf{X}$ is calculated by:

$$\mathbf{v} = (\mathbf{I} - \mathbf{X}\mathbf{X}^{+})\mathbf{u} \tag{3}$$

where $\mathbf{v}$ is the orthogonal vector, $\mathbf{I}$ is an identity matrix, and the superscript "+" indicates the Moore–Penrose pseudoinverse. By projecting $\mathbf{u}$ into the null space of $\mathbf{X}$, which includes the space spanned by all interference factors (other constituents, noise, etc.), the portions of the vector $\mathbf{u}$ which are covariate with $\mathbf{X}$ are suppressed. For applications of NAS in calculations which require a scalar value (e.g., signal-to-noise), the scalar form can be achieved by calculating the Euclidean Norm:

$$\hat{\mathrm{NAS}}_k = ||\mathbf{v}_k|| \tag{4}$$

2.5. Figures of merit from net analyte signal

Multivariate FOM are used to determine various metrics associated with method suitability using the calculated NAS [18,25–28]. Four FOM were calculated for comparing XRPD modes of analysis.

2.5.1. Sensitivity

Sensitivity is a measure of the extent of intensity variation due to changes in analyte concentration. In XRPD, this response is ultimately due to an increased number of diffracting planes at a given angle 2θ. For the kth component and ith sample, sensitivity is calculated by:

$$\hat{\mathbf{SEN}}_i = \frac{\mathbf{v}_k}{y_i} \tag{5}$$

where $\hat{\mathbf{SEN}}_i$ is the vector of sensitivities for each angular variable, $\mathbf{v}_k$ is the net analyte signal vector for the kth component, and y_i is the measured concentration for the ith sample. Sensitivity can also be expressed in scalar form with units of X-ray intensity (counts) per concentration unit.

2.5.2. Analytical sensitivity

Intensity unit differences between different instrumental platforms prohibit direct comparison of sensitivity measurements. Analytical sensitivity ($\boldsymbol{\gamma}$), expressed in units of inverse concentration, affords the direct comparison of sensitivity estimates for different analytical methodologies through normalization by instrumental noise. Analytical sensitivity is calculated using the equation:

$$\boldsymbol{\gamma} = \frac{\mathbf{S\hat{E}N}}{\delta_x} \tag{6}$$

where δ_x represents an estimate for the standard deviation of instrumental noise. For the purpose of this calculation, the standard deviation of the calculated NAS for diffraction patterns from (i) repeat scans with repositioning and (ii) repeat scans without repositioning was used as the estimated measurement error. These scans were performed in both reflectance and transmission geometries, as noise associated with each was expected to be different. Analytical sensitivity is in units of inverse concentration. The inverse of analytical sensitivity is an effective way of comparing the relative resolution of methods in concentration units.

2.5.3. Selectivity

Selectivity explains the fractional amount of the total signal retained from component signal overlap. Selectivity can be calculated by:

$$\mathrm{S\hat{E}L}_k = \frac{\mathrm{N\hat{A}S}_k}{\|\mathbf{x}_i\|} \tag{7}$$

where $\mathbf{x}_i$ is the sample diffraction pattern for the ith sample. Selectivity is sample dependent in that analyte concentration relative to other constituents will affect the outcome.

2.5.4. Signal-to-noise

Multivariate signal-to-noise (S/N) is a dimensionless ratio of the magnitude of useful signal to an estimate of noise. When δ_x is the estimate for the standard deviation of the measurement error (as calculated for δ_x), the (S/N) may be directly calculated as:

$$\frac{\mathrm{S}}{\mathrm{N}_k} = \frac{\mathrm{N\hat{A}S}_k}{\delta_x} \tag{8}$$

2.6. Separation of multi-component diffraction patterns

The mathematical separation of multi-component diffraction patterns into patterns having features solely related to a single constituent was accomplished through a multi-step orthogonalization procedure similar to the NAS calculation. The first step in the separation involved establishing an interference matrix including variance due to other components and instrumental noise. Diffraction patterns from samples having zero concentration of the constituent of interest (j) and background eliminated [29] pure component patterns (from either CLS regression vectors or XRPD scans) were modeled via singular value decomposition (SVD). Loadings explaining high variance in the interference matrix (while exhibiting low correlation to the component of interest (k)) were retained for the interference basis set $\mathbf{X}_{-j}$. Let $\mathbf{X}$ represent a matrix containing raw diffraction patterns collected for all the multi-component samples. The first orthogonalization can then be calculated by:

$$\mathbf{X}_k = (\mathbf{I} - \mathbf{X}_{-k}\mathbf{X}_{-k}^{+})\mathbf{X} \tag{9}$$

The second step of the single pattern isolation involved the pure component diffraction pattern (via CLS regression vector or XRPD scan) of the constituent of interest. A basis set is calculated by orthogonalizing the information retained from the first step to the pure component diffraction pattern. The result of this orthogonalization is a matrix containing variance not attributable to the pure component of interest. This basis set $\boldsymbol{N}$ is calculated by:

$$\mathbf{N} = (\mathbf{I} - \mathbf{B}_k\mathbf{B}_k^{+})\mathbf{X}_k \tag{10}$$

where $\mathbf{B}_k$ is the pure component diffraction pattern for the constituent of interest. Although this may seem counterintuitive at first, if it is assumed that everything remaining after suppressing the features associated with the pure component of interest is a result of the inefficiencies and non-linearities of the first net analyte orthogonalization (which is assumed to be unrelated to the physical nature of the constituent of interest), then this "signal" can be eliminated. The final step in the isolation of the single constituent diffraction pattern is to suppress all of the residual features after the first net analyte signal $\mathbf{X}_j$ correlated to the calculated basis set $\mathbf{N}$. The final net analyte signal orthogonalization is represented by:

$$\mathbf{X}_{\mathrm{fin}} = (\mathbf{I} - \mathbf{N}\mathbf{N}^{+})\mathbf{X}_j \tag{11}$$

where $\mathbf{X}_{\mathrm{fin}}$ is a matrix of the sample diffraction patterns of the single constituent of interest. It is important to note that smoothing was applied at each step using an optimized window width (2nd order polynomial used throughout) to mitigate the effects of uncertainty in calculations. Window width was optimized based upon the elimination of spurious features of calculated interference basis sets thereby preventing those errors from propagating throughout the net analyte signal orthogonalization. The two proposed techniques were implemented using the same manipulation scheme; the only difference being the use of pure component diffraction patterns either acquired directly by XRPD or calculated pure component diffraction patterns obtained from CLS regression.

3. Results

3.1. Diffraction pattern separation

The application of CLS regression to XRPD data provides a method for estimating pure component diffraction patterns of multi-component systems. There are, however, inefficiencies (E_c) associated with this technique. Fig. 1 shows the estimated regression vectors for each of the four constituents within the compacts, in addition to the diffractograms of the pure components as analyzed by transmission XRPD. The two regression

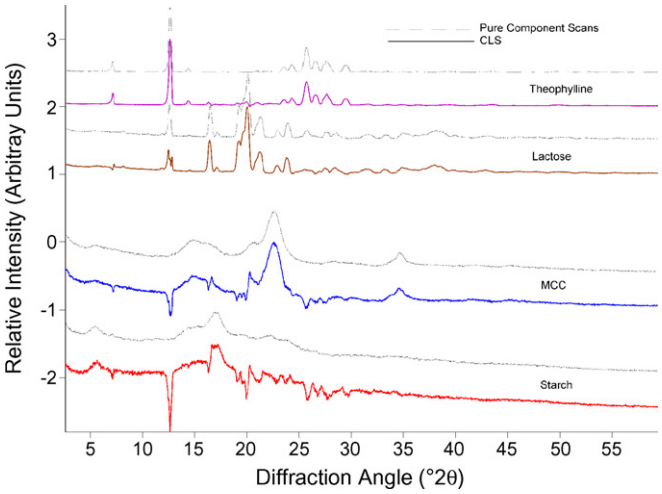

Fig. 1. PXRD pure component scans (- - -) and CLS calculated regression vectors (–).

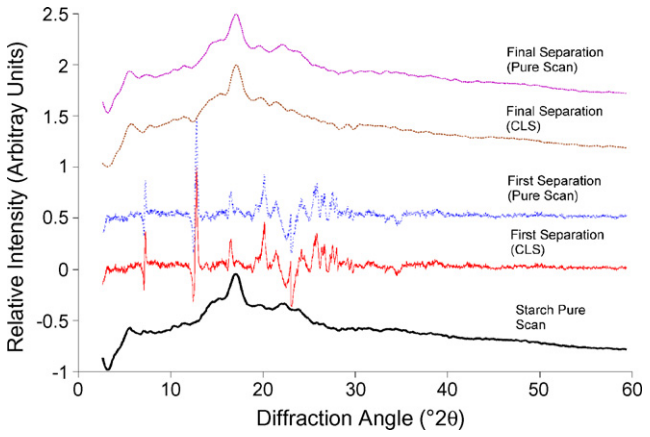

Fig. 3. Starch from Compact 20 following the first net analyte signal suppression using CLS regression vector (–) and pure scan (- - -). Starch final product using CLS regression vector (– –) and pure scan (- –). Starch pure component diffraction pattern (–).

vectors for the two crystalline components appear to be highly correlated with their respective XRPD patterns. Furthermore, both anhydrous theophylline and lactose monohydrate diffract at approximately 12° 2θ. Classical least squares regression was able to distinguish between the smaller lactose peak and the more intense theophylline peak, which overlap in the composite diffraction pattern. At first glance, the two disordered components appear to be highly correlated to their pure component patterns. The problem is that the regression vectors exhibit negative intensities located at positions corresponding to areas in the pattern at which the crystalline materials exhibit high intensity diffraction. Therefore, a program was created to correct for this based on *a priori* knowledge of the crystalline peak positions. Fig. 2 compares the corrected regression vectors with the original regression vectors. It can be seen in the comparison of the original and corrected regression vectors of the amorphous components that the negative intensities originally observed in the regression vectors have been removed. It is imperative to the separation technique that the regression vectors are true

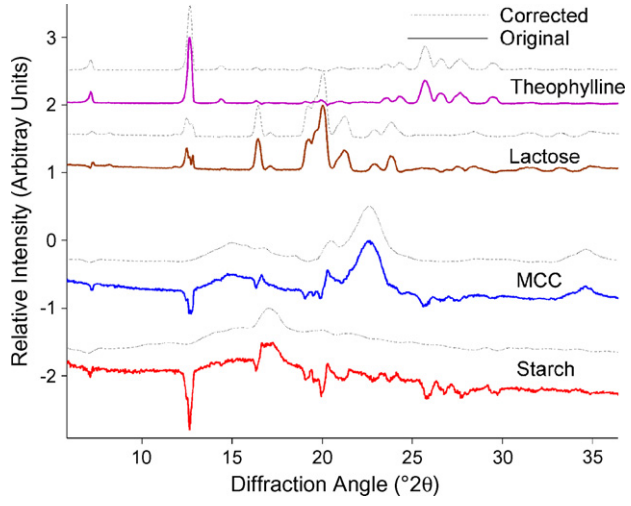

Fig. 2. Corrected CLS regression vectors (- - -) and original CLS regression vectors (–).

representations of the pure components; therefore, they should primarily contain only those diffraction characteristics related to the specified constituents.

The first interference matrix contains diffraction patterns from samples containing zero concentration of the constituent of interest and background eliminated [29] pure component patterns. The reason for eliminating the background should become apparent when considering the disordered components. The characteristic amorphous diffraction halos that are obtained for each of the disordered constituents are nearly superimposable. Therefore, when suppressing the features of one component, the features of the other would also be suppressed. Thus, by eliminating the shared background, only non-halo features are suppressed.

As previously mentioned, the two methods for diffraction pattern separation differ only by the use of the CLS-derived pure components or XRPD scans of the pure components. Fig. 3 shows the results for starch (analyzed via transmission mode) of the first net analyte suppression and the final net analyte suppression compared to the XRPD diffraction pattern of starch for both separation techniques. Starch, possessing the overall lowest diffraction intensity, was the most difficult to extract. After the first orthogonalization, the distinct shape corresponding to the diffraction of starch can be seen in the remaining pattern. There are, however, features not directly related to starch that dominate the overall pattern. These spurious positive and negative features align with significantly diffracting regions of the other materials. Since the magnitude of diffraction intensity is lower relative to the other components, these spurious features are much greater than those observed in the other separations. This is best illustrated in the separation of theophylline in Fig. 4. The first separation for theophylline, a crystalline material possessing high diffraction intensity, is not dominated by large spurious features, but clearly resembles the pure component scan. There are, as in the case for starch, negative intensities attributable to the inefficiencies of the first orthogonalization.

The final vectors, which were isolated using both methods, are illustrated in Fig. 5. Table 2 shows the effective rank of the

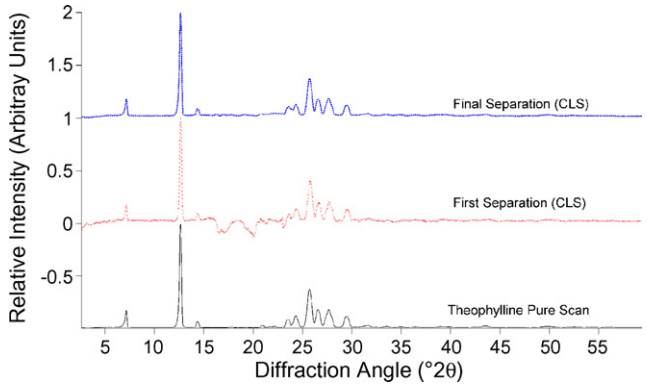

Fig. 4. Theohpylline separation via CLS regression vector method first orthogonalization (- -), final orthogonalization (– –), and theophylline pure scan (–).

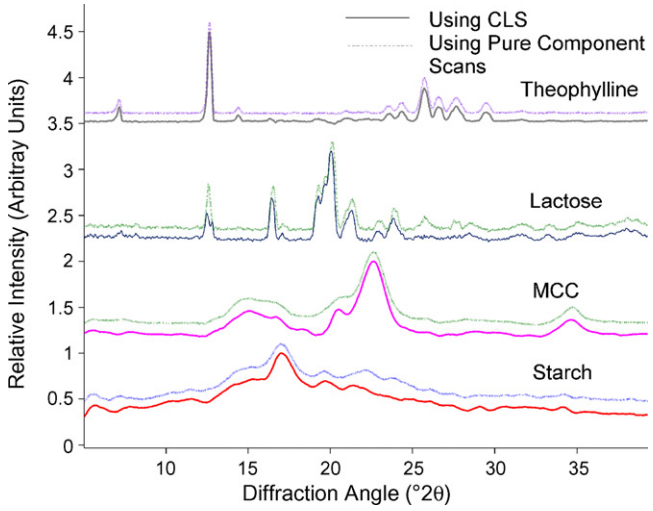

Fig. 5. Isolated individual constituent patterns as obtained using the pure component scan method (dashed line) and the CLS method (solid line).

data matrices as a function of the separation steps. It is important to note that the effective rank of the matrices never reaches one. This would be equivalent to multiplying the pure component patterns by the reference concentration value. By virtue of the fact that the final isolated patterns possess a rank greater than one, it can be stated that the matrices contain variation other

Table 2
Effective rank of data sets throughout the separation process

	Theophylline	Lactose	MCC	Starch
Reflectance, original		28		
Transmission, original		28		
CLS method				
Reflectance, first separation	22	22	22	22
Transmission, first separation	22	22	22	22
Reflectance, final separation	16	16	16	16
Transmission, final separation	16	16	16	16
Pure scan method				
Reflectance, first separation	22	22	22	22
Transmission, first separation	22	22	22	22
Reflectance, final separation	16	16	16	16
Transmission, final separation	16	16	16	16

Table 3
Angular variable correlation to reference concentration data for transmission mode XRPD

Variable correlation to reference concentration values			
Multi-component	Average	Maximum	Minimum
Anhydrous theophylline	0.5122	0.9866	−0.7502
Lactose Fast-Flo	−0.3353	0.9874	−0.9169
Microcrystalline cellulose	−0.1302	0.9228	−0.6852
Starch soluble	−0.0465	0.9356	−0.8187
Separated (CLS b-Vectors)			
Anhydrous theophylline	0.9661	0.9752	0.8571
Lactose Fast-Flo	0.8976	0.9631	0.4750
Microcrystalline cellulose	0.5838	0.6185	0.5485
Starch soluble	0.5361	0.5772	0.4494
Separated (pure scan)			
Anhydrous theophylline	0.9540	0.9760	0.6723
Lactose Fast-Flo	0.9051	0.9628	0.6117
Microcrystalline cellulose	0.5494	0.5816	0.5194
Starch soluble	0.5743	0.6152	0.4937

than that which can be explained by linear scaling of the pure components. Upon close examination, subtle peak position differences are observed between the two methods. Specifically, the lactose monohydrate peaks in the patterns separated via the pure scan method are shifted to higher scattering angles. When using the CLS regression vectors separation method, the respective pure component scans not only contain variation attributed to concentration difference, but possibly variation as a consequence of powder consolidation. Conversely, this variation is lacking when the pure components from XRPD scans are implemented.

The effectiveness of separation is demonstrated by the enhanced correlation of the angular variables to the reference concentration data. Due to the lack of diffraction features below 6° 2θ, correlation values were calculated between 6° and 60° 2θ. Table 3 contains the average, minimum, and maximum correlation values between angular variables and reference concentration data for the multi-component diffraction patterns and separated patterns as obtained by each method. It should be apparent that since the multi-component diffraction patterns are mainly dominated by the high intensity diffraction peaks of the crystalline components, the variable correlation of the disordered components to their respective reference concentrations is largely negative resulting from the loss of the crystalline materials. It is important to note that the separation eliminated all negatively correlated variables; this was observed for all constituents. Further, the average correlation for all components increased due to the positive correlation between each diffraction variable and the reference concentration.

3.2. Consolidation effects

Wildfong et al. have demonstrated that consolidated samples analyzed via transmission XRPD are vulnerable to signal attenuation due to both the solid fraction and thicknesses of the samples, thereby requiring a mathematical correction for accurate quantification [1]. In general, these effects were observed in

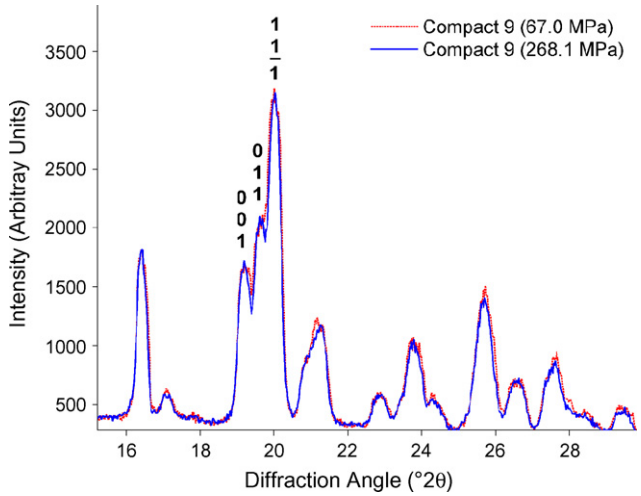

Fig. 6. XRPD scan of Compact 9 compressed at 67.0 MPa (dashed line) and 268.1 MPa (solid line).

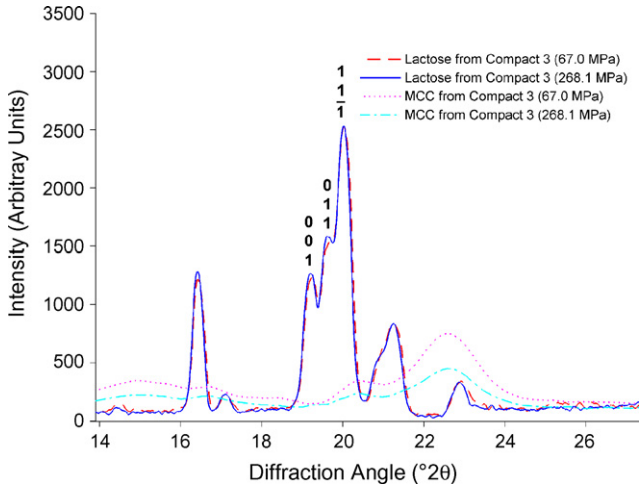

Fig. 8. Separated XRPD scans from Compact 3 of lactose 67.0 MPa (–) and 268.1 MPa (- - -), and microcrystalline cellulose 67.0 MPa (– –) and 268.1 MPa (- –).

the present data. Fig. 6 shows the multi-component diffraction patterns for a samples having a fixed composition (20% (w/w) anhydrous theophylline, 60% (w/w) lactose monohydrate, and 20% (w/w) soluble starch) compressed at 67.0 and 268.1 MPa, while Fig. 7 shows the diffractograms for samples having a different fixed composition (20% (w/w) anhydrous theophylline, 60% (w/w) lactose monohydrate and 20% (w/w) microcrystalline cellulose) compressed at the identical forces. The (0 0 1), (0 1 1), and (1 1 $\bar{1}$) reflections for lactose monohydrate (Fig. 7) are attenuated as a function of compaction pressure, whereas the same lactose monohydrate peaks in Fig. 6 are not. Although microcrystalline cellulose is disordered, its diffraction is significant in the region occupied by these three lactose monohydrate peaks (Fig. 7). The portion of signal attributable to lactose monohydrate was extracted from the scans in Fig. 7, and is presented in Fig. 8. The signal in the original composite pattern is no longer observed. In contrast, the microcrystalline cellulose signal was attenuated as a function of compaction pressure, particularly within the region of the three principally diffracting planes for the lactose monohydrate (Fig. 8). Therefore, the signal attenuation observed for the lactose monohydrate peaks in the multi-component diffraction pattern was a residual effect correlated to the microcrystalline cellulose attenuation. If simple pattern subtraction had been used without considering the covariance structure, the false sense of signal attenuation in lactose monohydrate would have propagated through the separation.

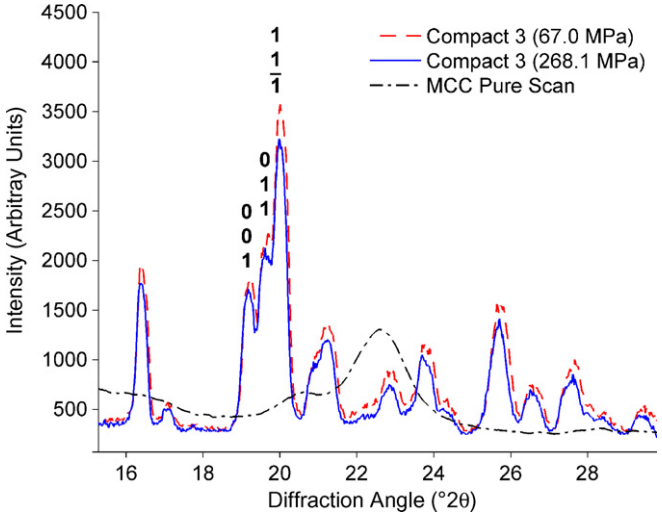

Fig. 7. XRPD scan of microcrystalline cellulose (– –), Compact 3 compressed at 67.0 MPa (- - -) and 268.1 MPa (–).

3.3. Figures of merit

Multivariate figures of merit were calculated using the separated diffraction patterns, which were obtained via net analyte signal theory. Table 4 compares the FOM between reflectance geometry and transmission geometry using the CLS regression vector separation. Overall, the FOM for transmission geometry are superior to that of the reflectance. One value of particular interest occurs for the lactose monohydrate selectivity, which was greater for the reflectance data than for the corresponding transmission data. This is likely due to the increased angular resolution associated with reflectance geometry. When considering how selectivity is calculated, high resolution will be observed for deconvolved peaks versus overlapping peaks, provided that the overall intensity is approximately equal. Relative to the multi-component diffraction patterns, the three largest lactose monohydrate peaks in the separated patterns obtained from transmission data are increasingly convolved, thereby decreasing the resolution.

When comparing the FOM for the method of separation that utilizes the pure scans (Table 5), an extreme reduction in analytical sensitivity and S/N is observed for the reflectance mode analysis, whereas the transmission mode analysis is relatively unaffected. Even though both the sample scans and the pure component scans were corrected for a continuous axis shift, there are many errors associated with Bragg–Brentano reflectance geometry that causes anisotropic peak shifts/distortions [19,20]. Since

Table 4
FOM for reflectance and transmission geometries of compacts compressed at 67.0 MPa separated via the CLS method

	FOM for CLS separation method (67.0 MPa)			
	Theophylline	Lactose	MCC	Starch
Reflectance				
Sensitivity (counts/(w/w))	189.06	200.35	109.29	115.38
Analytical sensitivity $(w/w)^{-1}$	3.25	1.00	1.48	1.38
Selectivity	0.41	0.44	0.23	0.24
Signal-to-noise	101.86	30.55	42.96	39.54
Transmission				
Sensitivity (counts/(w/w))	829.64	365.84	471.94	541.03
Analytical sensitivity $(w/w)^{-1}$	7.16	2.82	3.63	4.41
Selectivity	0.65	0.32	0.43	0.49
Signal-to-noise	226.54	83.44	107.87	132.50

Table 5
FOM for reflectance and transmission geometries of compacts compressed at 67.0 MPa separated via the pure scan method

	FOM for pure scan separation method (67.0 MPa)			
	Theophylline	Lactose	MCC	Starch
Reflectance				
Sensitivity (counts/(w/w))	174.60	138.56	95.76	109.43
Analytical sensitivity $(w/w)^{-1}$	0.47	0.65	0.72	1.24
Selectivity	0.38	0.30	0.20	0.23
Signal-to-noise	14.97	19.25	21.18	35.34
Transmission				
Sensitivity (counts/(w/w))	740.25	375.70	475.58	517.39
Analytical sensitivity $(w/w)^{-1}$	17.21	4.12	4.03	4.55
Selectivity	0.58	0.33	0.44	0.47
Signal-to-noise	553.28	120.66	121.06	135.13

the multivariate techniques employed only account for variation in the direction of intensity, it is hypothesized that this reduction is a result of a combination of these errors. It is also known that stress/strain from material consolidation can result in anisotropic peak distortions. The use of the CLS regression vectors retain more of this information, whereas the pure scans alter the intensities to conform to the positions as determined by XRPD. This supports the use of CLS regression calculated vectors in the separation method.

Table 6 compares the FOM between transmission and reflectance geometry using the CLS regression vector separation of compacts compressed at 268.1 MPa. As anticipated, a decrease in performance was observed with an increase in compaction pressure. Interestingly, only minor differences were observed in transmission mode data for lactose monohydrate as a result of increasing compaction pressure. Recall that the lack of signal attenuation as a function of increasing solid fraction and compact thickness was observed for this component. Additionally, a dramatic decrease in the FOM for the starch pattern separated from reflectance data was observed with increasing compaction force relative to the other constituents. This suggests that because starch has the lowest overall diffraction intensity

Table 6
FOM for reflectance and transmission geometries of compacts compressed at 268.1 MPa separated via the CLS method

	FOM for CLS separation method (268.1 MPa)			
	Theophylline	Lactose	MCC	Starch
Reflectance				
Sensitivity (counts/(w/w))	171.29	261.61	174.96	22.40
Analytical sensitivity $(w/w)^{-1}$	2.94	1.27	2.36	0.27
Selectivity	0.38	0.55	0.37	0.05
Signal-to-noise	90.92	36.43	68.52	8.16
Transmission				
Sensitivity (counts/(w/w))	706.22	420.73	290.43	416.92
Analytical sensitivity $(w/w)^{-1}$	6.10	3.24	2.23	3.40
Selectivity	0.57	0.38	0.30	0.38
Signal-to-noise	192.47	94.33	71.53	97.98

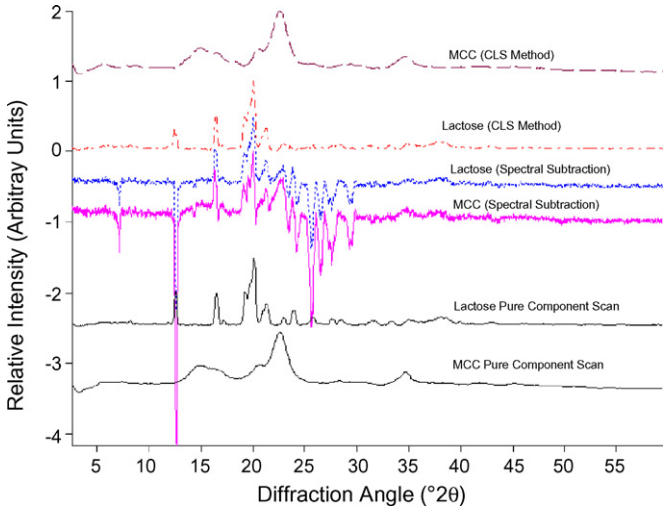

Fig. 9. Pure constituent diffraction patterns (solid lines) relative to spectral subtraction results for lactose (– –) and MCC (–·–) versus CLS regression separation method for lactose (—) and MCC (– –).

of the four components, any signal attenuation corresponds to a severe decrease in available detectable variation.

3.4. Pattern subtraction comparison

One of the greatest inaccuracies associated with pattern subtraction is the failure to address covariation of angular variables. As a result, intensity reduction of angular variables correlated to both the subtracted component and the component of interest occurs. To illustrate this point, pattern subtraction isolation of lactose and microcrystalline cellulose versus the CLS method outlined in this work are illustrated in Fig. 9. The first feature of the pattern subtraction isolated components is the strong negative intensities; specifically located at scattering angles of diffraction peaks belonging to theophylline. The magnitude of scattering at angular variables in this region is drastically different for theophylline, lactose, and microcrystalline cellulose. Furthermore, pattern subtraction does not account for any errors present in the diffraction data. Therefore, the magnitude of the difference in intensities results in negative intensities at scattering angles consistent with diffraction features of the other components. The second aberrant feature that should be apparent is that the separated lactose and microcrystalline cellulose patterns obtained via pattern subtraction are nearly identical. This indicates that the lack of sensitivity of this method for components possessing low intensity diffraction (i.e., the two disordered components). In contrast, the methods presented in this work indirectly compensate for the errors through the final orthogonalization step.

4. Conclusions

The use of XRPD in the analysis of intact compacts remains an important mainstay in the characterization of solid-state phenomena. Performing analyses using multivariate methods affords the ability to improve parameters such as signal-to-noise, sensitivity, and selectivity. Accurate separation of multi-component diffraction patterns with respect to the covariance structure enables direct observation of changes to material structure without compromising the intensity of correlated angular variables.

As formulation design moves towards utilization of increasingly disordered materials, the importance of accurate detection of scattering in XRPD from these constituents becomes imperative. Furthermore, when attempting separation of multi-component patterns, the errors and assumptions associated with given techniques must be considered. It has been shown that the implications of performing pattern subtraction for the isolation of amorphous materials in the presence of highly crystalline materials can have detrimental effects on the retained features.

The proposed methods were successful in separating diffraction patterns obtained from XRPD of intact quaternary consolidated mixtures. Improved correlation to reference concentration data substantiates the effectiveness of this technique. The ability to separate compound effects of signal attenuation between multiple components within the system demonstrates the usefulness of these methods. Multivariate FOM revealed differences between the modes of analysis as well as the separation methods. The work herein supports that the CLS regression vector method retains an increased amount of structural information due to the pure components directly being calculated from the samples. This does, however, deserve further investigation as to the correlation of physical information retained via this technique. Ongoing work will attempt to establish metrics to determine sufficient feature suppression while maximizing the retention of material-specific information.

Acknowledgement

The authors thank support provided through the National Science Foundation grant #NSF-DUE-0511444.

References

[1] P.L.D. Wildfong, N.A. Morley, M.D. Moore, K.R. Morris, J. Pharm. Biomed. Anal. 39 (2005) 1–7.
[2] X. Chen, S. Bates, K.R. Morris, J. Pharm. Biomed. Anal. 26 (2001) 63–72.
[3] N.V. Phadnis, R.K. Cavatur, R. Suryanarayanan, J. Pharm. Biomed. Anal. 15 (1997) 929–943.
[4] R. Bandyopadhyay, J. Selbo, G.E. Amidon, M. Hawley, J. Pharm. Sci. 94 (2005) 2520–2530.
[5] W. Cao, S. Bates, G.E. Peck, P.L.D. Wildfong, Z. Qiu, K.R. Morris, J. Pharm. Biomed. Anal. 30 (2002) 1111–1119.
[6] L.S. Zevin, G. Kimmel, Quantitative X-ray Diffractometry, Springer, New York, 1995.
[7] P.J. Brown, J. R. Stat. Soc. B 44 (1982) 287–308.
[8] S.D. Brown, S.T. Sum, F. Despagne, B.K. Lavine, Anal. Chem. 68 (1996) 21–61.
[9] H. Martens, T. Næs, Multivariate Calibration, John Wiley and Sons, New York, NY, USA, 1989.
[10] H. Martens, T. Næs, Multivariate calibration by data compression, in: P.C. Williams, K.H. Norris (Eds.), Near-Infrared Technology in the Agricultural and Food Industries, AACC, St. Paul, MN, 2001, pp. 59–100.
[11] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbooks of Chemometrics and Qualimetrics, Elsevier, New York, 1998.

[12] H.M. Rietveld, J. Appl. Crystallogr. 2 (1969) 65–71.
[13] T. Artursson, A. Hagman, S. Bjork, J. Trygg, S. Wold, S.P. Jacobsson, Appl. Spectrosc. 54 (2000) 1222–1230.
[14] A.C. Jorgensen, I. Miroshnyk, M. Karjalainen, K. Jouppila, S. Siiria, O. Antikainen, J. Rantanen, J. Pharm. Sci. 95 (2006) 906–916.
[15] M. Karjalainen, S. Airaksinen, J. Rantanen, J. Aaltonen, J. Yliruusi, J. Pharm. Biomed. Anal. 39 (2005) 27–32.
[16] P.R. Nassab, R. Rajko, P. Szabo-Revesz, J. Pharm. Biomed. Anal. 41 (2006) 1191–1197.
[17] M. Savolainen, A. Heinz, C. Strachan, K.C. Gordon, J. Yliruusi, T. Rades, N. Sandler, Eur. J. Pharm. Sci. 30 (2007) 113–123.
[18] S.M. Short, R.P. Cogdill, C.A. Anderson, AAPS PharmSciTech 8 (2007), Article 96.
[19] B.D. Cullity, S.R. Stock, Elements of X-ray Diffraction, third ed., Prentice Hall, Upper Saddle River, 2001.
[20] R.L.S. Ron Jenkins, Introduction to X-ray Powder Diffractometry, Wiley-Interscience, New York, 1996.
[21] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627–1639.
[22] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1988) 1193–1202.
[23] E.V. Thomas, D.M. Haaland, Anal. Chem. 62 (1990) 1091–1099.
[24] A. Lorber, Anal. Chem. 58 (1986) 1167–1172.
[25] A. Lorber, K. Faber, B.R. Kowalski, Anal. Chem. 69 (1997) 1620–1626.
[26] R. Bro, C.M. Andersen, J. Chemom. 17 (2003) 646–652.
[27] J. Ferre, N.K.M. Faber, Chemom. Intell. Lab. Syst. 69 (2003) 123–136.
[28] J.W.B. Bragga, R.J. Poppi, J. Pharm. Sci. 93 (2004) 2124–2134.
[29] R.P. Goehner, Anal. Chem. 50 (1978) 1223–1225.