International Journal of Pharmaceutics 511 (2016) 488–504

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

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

Considerations on the quantitative analysis of apparent amorphicity of milled lactose by Raman spectroscopy

Samaneh Pazesh[a,*], Lucia Lazorova[a], Jonas Berggren[b], Göran Alderborn[a], Johan Gråsjö[a]

[a] Department of Pharmacy, Uppsala University, Uppsala, Sweden
[b] Recipharm Pharmaceutical Development AB, Solna, Sweden

ARTICLE INFO

Article history:
Received 17 May 2016
Received in revised form 30 June 2016
Accepted 1 July 2016
Available online 7 July 2016

Keywords:
Raman spectroscopy
Lactose
Amorphous content
Spectral data analysis
Principal component analysis (PCA)
Milling induced disorder

ABSTRACT

The main purpose of the study was to evaluate various pre-processing and quantification approaches of Raman spectrum to quantify low level of amorphous content in milled lactose powder. To improve the quantification analysis, several spectral pre-processing methods were used to adjust background effects. The effects of spectral noise on the variation of determined amorphous content were also investigated theoretically by propagation of error analysis and were compared to the experimentally obtained values. Additionally, the applicability of calibration method with crystalline or amorphous domains in the estimation of amorphous content in milled lactose powder was discussed.

Two straight baseline pre-processing methods gave the best and almost equal performance. By the succeeding quantification methods, PCA performed best, although the classical least square analysis (CLS) gave comparable results, while peak parameter analysis displayed to be inferior.

The standard deviations of experimental determined percentage amorphous content were 0.94% and 0.25% for pure crystalline and pure amorphous samples respectively, which was very close to the standard deviation values from propagated spectral noise.

The reasonable conformity between the milled samples spectra and synthesized spectra indicated representativeness of physical mixtures with crystalline or amorphous domains in the estimation of apparent amorphous content in milled lactose.

1. Introduction

The interest in the quantitative analysis of amorphous content of pharmaceutical solids has increased considerably the last years. This stems from the facts that firstly, amorphous solids may be used to solve certain formulation problems, such as low solubility of a drug, and, secondly, the processing of crystalline particles may result in, often undesired, formation of an amorphous phase (sometimes referred to as process induced disordering) that may still have an impact on physical and chemical properties of the materials and thus the final pharmaceutical product performance. Processing operations such as size reduction (Caron et al., 2011; Chamarthy and Pinal, 2008; Otte et al., 2012), compression (Kaneniwa et al., 1985) and dry mixing (Pazesh et al., 2013) has been shown to induce the formation of thermodynamically unstable amorphous regions of predominately crystalline particles. Therefore, the quantification of low levels of amorphous materials in powders has become an important part of the development of pharmaceutical preparations and this paper has been written in the context of this aspect.

Many analytical techniques may be used to quantify modest to high levels of amorphous content in a solid, including X-ray powder diffraction (XRPD), differential scanning calorimetry (DSC), solution calorimetry, isothermal microcalorimetry and dynamic vapour sorption. It is reported that the quantification limit of XRPD and DSC are greater than 5% (Shah et al., 2006), whereas solution calorimetry (Hogan and Buckton, 2000), isothermal calorimetry (Buckton et al., 1995) and dynamic vapour sorption (Mackin et al., 2002; Sheokand et al., 2014; Young et al., 2007) may enable the quantification of amorphous content of less than 1%.

Vibrational spectroscopy techniques, such as near- infrared spectroscopy (NIR) (Fix and Steffens, 2004; Hogan and Buckton, 2001; Savolainen et al., 2007), mid-infrared spectroscopy (MIR) (Agatonovic-Kustrin et al., 2001; Bartolomei et al., 1997) and Raman spectroscopy (Niemelä et al., 2005; Susi and Ard, 1974; Taylor and Zografi, 1998), have gained increasing interest in past

* Corresponding author at: Department of Pharmacy, Uppsala University, Uppsala Biomedical Center, P.O. Box 580, SE-751 23 Uppsala, Sweden.
E-mail address: samaneh.pazesh@farmaci.uu.se (S. Pazesh).

http://dx.doi.org/10.1016/j.ijpharm.2016.07.001

decades for the analysis of chemical and physical properties of pharmaceutical solids. In a review, Strachan et al. (Strachan et al., 2007) give an overview of a number of such applications of the Raman technique, including the analysis of crystallinity of a solid. The limited water sensitivity of the Raman signal is a major advantage of this technique in the study of amorphous solids since water is frequently present as moisture in the amorphous phase.

A number of papers on the quantification of amorphous content of pharmaceutical substances by Raman spectroscopy have been reported in the literature. An early paper on the subject was authored by Taylor and Zografi (Taylor and Zografi, 1998) which was followed by other papers dealing with both quantification of amorphous content of drugs (e.g. Heinz et al. (Heinz et al., 2009), Mah et al. (Mah et al., 2015)) and excipients (Savolainen et al. (Savolainen et al., 2007) and Fix et al. (Fix and Steffens, 2004)).

According to literature, there are two ways of tackling the quantitative analyses of a Raman spectrum. The first group of methods is based on peak analysis, using peak variables such as amplitudes and areas or combinations thereof. In the second group of methods, the spectral values are used as they are, irrespectively of what spectral feature they may belong to, and subjected to multivariate analyses, such as principal component analysis (PCA) and classical least square analysis (CLS). Before the quantitative analysis, a raw Raman spectrum is often pre-processed in order to remove unwanted baseline and intensity variations in the Raman spectrum due to fluorescence and macroscopic variations in sample structure. The choice of pre-processing strategy may be crucial in order to obtain robust and accurate quantitative information from Raman spectra. Nevertheless, there are only few papers discussing the quantification of amorphous content of a pharmaceutical substance that also reports on the effect of the pre-processing procedure of raw Raman spectra for the out-come of the analysis. Hence, this is an aspect of the spectrum analysis that deserves more attention.

The standard procedure used in the literature to transform a Raman spectrum of a sample of milled particles into amount disordered content, is to use a standard curve obtained from spectra of reference samples. These reference samples are typically physical mixtures of completely crystalline and completely amorphous particles of known proportions, i.e. two-state systems with domains that are either crystalline or amorphous. The representativeness of such a two-state system for the physical structure of milled particles that are considered partially amorphous may however be questioned. Two conceptions are used in the literature (Chamarthy and Pinal, 2008; Luisi et al., 2012) to describe the physical nature of partially amorphous particles. Firstly, particles can be considered as a one-state system where the degree of disorder depends on the concentration of defects or the size and orientation of crystallites forming the particle. Secondly, particles can be considered as a two-state system where the degree of disorder depends on the proportion of amorphous and crystalline domains. The physical form of milled particles is thus an intricate issue as the exact physical nature of the disorder in milled samples is not obvious. Thus, a careful interpretation of the determined percentage amorphous content must be seen as an apparent measure and will at least partly be a relative measure of the degree of disorder rather than an absolute measure of amorphicity. The term apparent amorphicity or apparent amorphous content is used in this paper to acknowledge the fact that the quantitative analysis of a disordered solid results in a single percentage value although the exact nature of the solid is unknown and possibly complex.

The overall aim of this study is to address the question of the possibility to derive physically sound values of the apparent amorphous content for milled powders with a special reference to the importance of the combination of pre-processing and quantification methods for the determination of amorphous content in a powder. Lactose, one of the most common pharmaceutical excipients, is used as model material in the study. The overall aim was sub-divided into three sub-objectives. The first objective was to evaluate the effect of a comprehensive series of pre-processing methods combined with a series of quantification methods for the determination of fraction of apparent amorphous solid in a powder. The quantification methods could be divided into two main approaches; either based on peak parameter analysis or multivariate analysis (as illustrated in Figs. 1 and 2). In this part of the study, a series of physical mixtures of crystalline and amorphous lactose was used. The second objective of the study was to evaluate how spectral noise influenced the variation in determined fraction of amorphous content for the physical mixtures under investigation. This can be done by simulations, but, if possible, a derivation of an analytical expression will facilitate this evaluation. In this paper, an expression for propagation of errors for the combination of linear background and PCA analysis was derived. The third objective was to determine and compare the apparent amorphous content of milled powders of lactose, using the same pre-processing and quantification methods as were used in the assessment of amorphous content (Figs. 1 and 2), and to evaluate the representativeness of the calibration method in the estimation of apparent amorphous content.

2. Materials and methods

2.1. Materials and sample preparation

Crystalline α-lactose monohydrate Pharmatose® 200 M, (DFE Pharma, the Netherlands) was used as received. Amorphous lactose was prepared from a 10% (*w/w*) aqueous solution of α-lactose monohydrate, using a Büchi Mini Spray Dryer B-290 Advance (Büchi Labortechnik AG, Flawil, Switzerland). The aspirator rate was set to 100%, the spray gas flow to 40 L/min and the feed pump rate to 4.0 mL/min. The inlet and outlet temperatures were maintained at 155 °C and 98 °C respectively.

Ball milling of crystalline lactose was performed in a planetary ball mill (PM 100, Retsch, Germany) at 25 °C and 30 ± 3% relative humidity (RH) with rotation speed of 400 rpm. 1 g of lactose was milled in a stainless steel milling jar of a volume of 12 cm^3 with 50 stainless steel balls with a diameter of 5 mm corresponding to a ball to powder mass ratio of 25:1. In order to prepare lactose samples that could represent low, intermediate and high degree of apparent amorphous content of lactose by the ball mill used in this study, a series of pre-trials was conducted. Based on these pre-trials, milling times of 10, 300 and 1200 min were used for sample preparation. For the two longer milling times, a combination of milling periods and pause periods was applied to allow cooling of the jar and thus minimize heating of the sample, i.e. after each milling period of 20 min a pause period of 5 min was used.

For standard curves, physical mixtures with 0%, 5%, 10%, 15%, 30%, 50%, 70%, 85%, 90%, 95% and 100% (*w/w*) amorphous content with total weight of 2 g were prepared by combining spray dried lactose powder, representing 100% amorphous lactose, and Pharmatose powder, representing 100% crystalline lactose. Constituents of each mixture were added by geometrical dilution and mixed in a mortar with a pestle at 25 °C and 30 ± 3% RH.

2.2. Experimental methods

2.2.1. X-ray powder diffraction

Powder X-ray analysis were performed using a Bruker D8 Advance diffractometer equipped with a position sensitive detector (PSD), LynxEye (Bruker AXS, Inc., Madison, WI, USA)

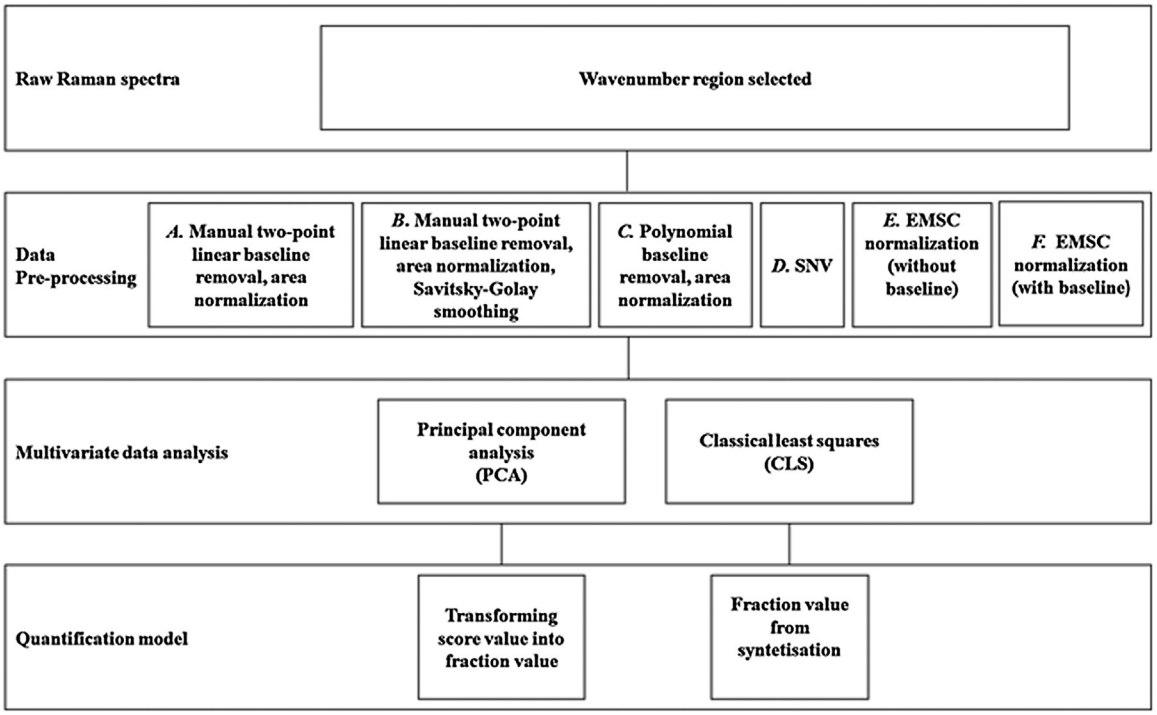

Fig. 1. Flow chart of different pre-processing and quantification approaches using multivariate analysis of Raman spectra.

and CuKα anode; λ = 1.5406 Å. The sample was placed in a sample holder and scanned at 40 kV, 40 mA from 10° to 60° 2θ using a scanning speed of 0.2 s and a step size of 0.016°. A motorized primary and secondary divergence slit with 0.4° and 2.48° were used. The diffractogram were collected using DIFRACT.SUIT software. The experimental diffraction patterns were compared with theoretical patterns which were based on the data from the Cambridge Crystallographic Data Centre (CCDC).

2.2.2. Raman spectroscopy

Raman spectra was collected using Raman SLSR-ProTT analyser (Enwave Optronics., Irvine, CA, USA), equipped with TE cooled CCD

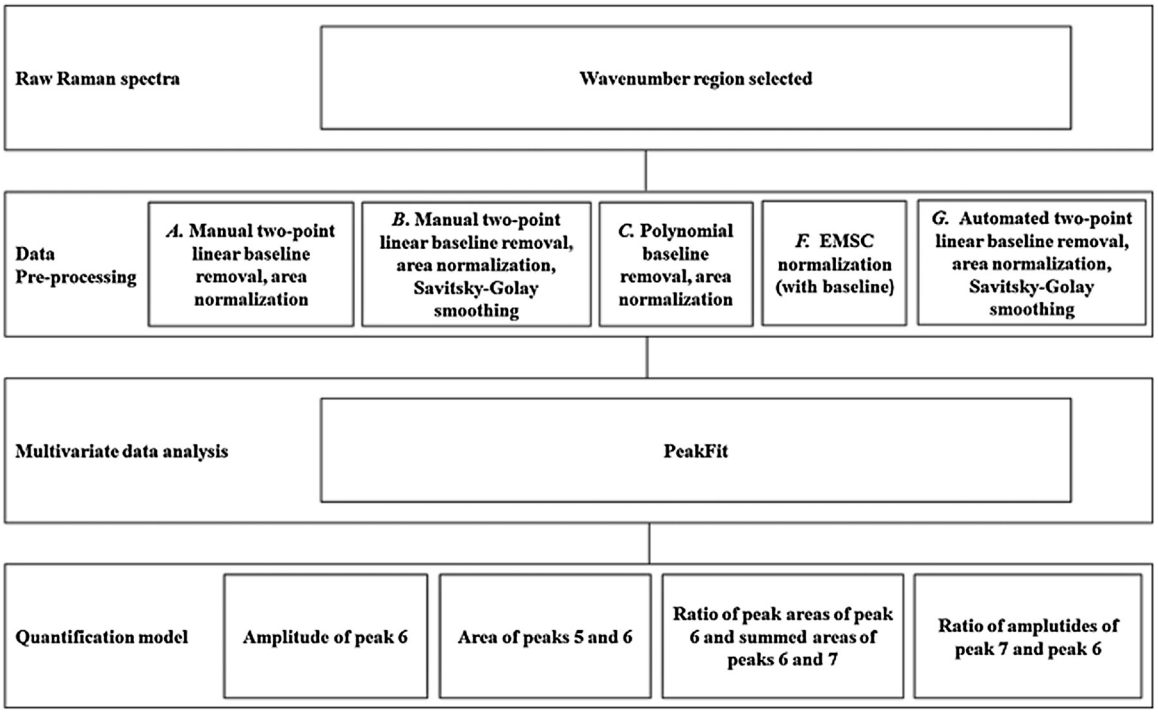

Fig. 2. Flow chart of different pre-processing and quantification approaches using peak parameter analysis of Raman spectra.

detector and a laser source with an excitation wavelength of 785 nm. A reference standard (ECS-I, provided by the manufacturer together with the Raman equipment) was used to calibrate the spectrometer each time the instrument was switched on. The measurements were performed at room temperature in a dark room to minimize the influence of external light. In order to identify potential procedure factors that could be of importance for quantification of amorphous fraction of lactose by Raman, a series of pre-trials was conducted. These included sample preparation (compact or powder), rotating or stationary sample holder, laser power, measurement time and sample to laser probe distance. Based on the pre-trial outcome, the following procedures were decided to be used: Compacts containing 400 ± 2 mg of lactose samples were prepared by powder compression in a single punch press (Korsch EK0, Korsch, Germany) at an applied pressure of 100 MPa, using circular flat-faced punches with a diameter of 11.3 mm. The compressed sample (3 ± 0.5 mm thick) was then placed in a rotating sample holder which was positioned under the laser probe with a fixed distance of 7 mm. The rotation speed was ~30 rpm and the radius of the circle swept by the laser spot was around 3 mm. Laser power and integration time was tuned in order to avoid excessive heating of the sample. The laser output power was set to 30 mW and the integration time was set to 3 s over the Raman shift range of 100 cm^{-1} to 2200 cm^{-1}. Each spectrum consisted of 100 scans, resulting in average measurement time of 5 min per measurement. The diameter of the laser beam at focus (the illuminated spot) was approximately 0.3 mm.

2.3. Pre-processing methods

2.3.1. Baseline correction and normalization methods

The measured spectra may be regarded as consisting of the Raman shifted signal superimposed on a background offset, mainly originating from fluorescence. Further slight signal intensity variations are noted due to variations in the preparation of the samples. Thus to make the different Raman spectra comparable, the spectra have firstly, to be corrected by subtracting a baseline that ideally should be equal to the fluorescence signal profile or another offset reflecting the baseline, and secondly, to be normalized by the intensity. Basically, normalization is performed by dividing the intensity at each wavenumber of a spectrum with a constant, specific for the spectrum and normalization method.

Regarding the background, one is, in practice, more or less referred to qualified guesses to at least come close to the exact shape of the real fluorescence profile. The problem is well known in most types of spectroscopy, e.g. NMR spectroscopy, mass spectroscopy, UV spectroscopy and different types of X-ray based spectroscopies. On the other hand, if the systematic error to a large degree is the same in all different sample spectra that are to be compared, the errors will in some degree cancel out each other if spectra are used for quantification.

Regarding the normalization, typical strategies has been to normalize to a certain peak amplitude or the area covered by the signal in a certain wavenumber section. As shown in the succeeding sections, also other strategies to handle the background and normalization problem have been used.

In this investigation, low wavenumber region of spectrum (319 cm^{-1} to 491 cm^{-1}) was chosen for quantification analysis. The wavenumber region used in Raman spectrum analysis may affect the sensitivity of measurement dependent on the type of quantitative analysis to be performed e.g. (Mah et al., 2015). Within the scope of the presented study, the exact choice of wavenumber region is less important and a region that has earlier been used in Raman spectrum analysis of lactose sample was arbitrarily chosen (Niemelä et al., 2005).

2.3.2. Straight baseline methods and higher polynomial baseline method approaches

In the straight baseline method, hereafter denoted method *A*, the simplest baseline approach was employed, where a straight line between the spectrum intensity values at 319 cm^{-1} and 491 cm^{-1} was used as baseline. The corrected spectrum was calculated by subtracting the spectra intensity values by the straight line values at the respective wavenumber. The baseline approach using a straight line to correct the spectrum was also applied in straight baseline with Savitsky-Golay filtered endpoints, hereafter donated method *B*. Instead of using the recorded intensities at 319 cm^{-1} and 491 cm^{-1} when calculating the linear baseline, the corresponding Savitsky-Golay filtered (3rd degree, n = 9) intensity values at 319 cm^{-1} and 491 cm^{-1} was used. A possible advantage with the method *B* could be the diminishing of noise effects on the baseline that subsequently could propagate into the determined amorphous fraction.

In polynomial method, here denoted method *C*, a modification of the Lieber et al. (Lieber and Mahadevan-Jansen, 2003) method was done. Lieber proposes an iterative algorithm (Polyfit) that suggests a separation of regions of the spectrum that could be regarded as pure background (due to fluorescence) from regions that is due to vibrational effects. A polynomial is fitted to the measured intensity in the background sections of the spectrum. The polynomial (in this work a 3rd degree polynomial) was fitted to the wavenumber region of the spectra that was to be investigated (in this work from 250 cm^{-1} to 540 cm^{-1}). Portions of the spectra that are positive relative the fitted line and above the noise level are interim regarded as vibrational peaks and removed in a new polynomial fit. This procedure is repeated until convergence is reached and the then obtained polynomial is used as the baseline. The calculations were carried out on an EXCEL spread sheet linked to an in house developed visual basic macro program. The original method suggested by Lieber et al. does not consider the noise level when determine vibrational regions of the spectra.

In methods *A*, *B* and *C*, the spectra, after baseline subtraction, were normalized by dividing the spectrum with the area under the curve, numerically determined by the trapezoidal method.

2.3.3. Standard normal variate (SNV)

In the standard normal variate method (SNV), here denoted method *D*, the spectrum is transformed according to Eq. (1): (Barnes et al., 1989)

$$y_{SNV,i} = \frac{z_i - < z >}{s_z} \quad (1)$$

where z_i is the measured intensity at wavenumber i, $< z >= \frac{1}{n}\sum z_i$, i.e. the mean in the interval, and $s_z = \sqrt{\sum (z_i - < z >)^2/(n-1)}$, i.e. the standard deviation in the interval. The summations are carried out over all wavenumber points in the interval and n is number of wavenumber points.

By the subtraction of the mean in the numerator, consideration for the various backgrounds of the spectra of the samples was carried out, however not necessarily resulting in a proper background removal as in the methods described above. Therefore, in this way pre-processed spectrum is not usable for Peak parameter based quantification methods (see below). Further, the normalization is carried out by the division by the standard deviation. In this work the method was applied to the wavenumber region from 319 cm^{-1} to 491 cm^{-1}.

2.3.4. Extended multiple scattering correction (EMSC) methods

In extended multiple scattering correction (EMSC), hereafter denoted method *E*, the spectrum is regarded being described

according to the following model (Eq. (2)): (Martens et al., 2003)

$$\bar{z}_j = a_j + b_j \cdot \left(\frac{\bar{z}_{am} + \bar{z}_{cryst}}{2}\right) + h_j \cdot (\bar{z}_{am} - \bar{z}_{cryst}) + d_j \cdot \bar{x} + e_j \cdot \overline{x^2} \qquad (2)$$

where $\bar{z}_j$ is the vector representation of the spectrum for sample j, $\bar{z}_{cryst}$ and $\bar{z}_{am}$ are the vector representation of the average spectrum for the pure crystalline and amorphous samples respectively and $\bar{x}$ and $\overline{x^2}$ are the vectors of wavenumbers and squared wavenumbers respectively. Further a_j, d_j and e_j is the coefficients of the wavenumber dependent offset polynomial of degree 2, b_j is the scaling (for normalization) and h_j is linked to the amorphicity. The coefficients a_j, b_j, d_j and e_j was determined by multiple linear regressions and used to determine an offset and normalized corrected spectrum, $\bar{y}_{EMSC,j}$ according to Eq. (3):

$$\bar{y}_{EMSC,j} = \frac{\bar{z}_j - (a_j + d_j \cdot \bar{x} + e_j \cdot \overline{x^2})}{b_j} \qquad (3)$$

In this work, the method was applied to the wavenumber region from 319 cm^{-1} to 491 cm^{-1}. By the b_j coefficient, normalization is performed and by the a_j, d_j and e_j coefficients, this method compensates for offset effects such as fluorescent effects, but, in similarity with the SNV method, this is not necessarily resulting in a proper baseline removal. Therefore, in this way pre-processed spectra is not either usable for peak parameter based quantification methods (see below). To obtain a proper baseline removal, and with that pre-processed spectra appropriate for peak parameter based quantification methods, a second approach of EMSC was applied, here denoted method *F*. In Eq. (2) the spectra vectors $\bar{z}_{cryst}$ and $\bar{z}_{am}$, were replaced by corresponding vectors of background corrected and normalized spectra (according to the method *A*). Thereafter a normalized corrected spectrum $\bar{y}_{EMSC,j}$ was obtained according to Eq. (3).

2.3.5. Pre-processing method provided by PeakFit software

A baseline algorithm offered within the PeakFit software was also used, here denoted method *G*. Sections of the spectrum in the range 315–496 cm^{-1} were detected, where the second derivative of the data is both constant and zero, which were regarded to be sections where the baseline is not superposed by spectral features. A linear baseline was fit to these sections and subtracted from the spectrum, where after the data were normalized to unit area.

2.4. Quantification methods

2.4.1. Principal component analysis (PCA)

Principal component analysis (Simca software, Umetrics, Sweden) was conducted on data pre-processed in the different ways described above (method *A-F*). The wavenumbers were treated as variables and the different physical mixtures of lactose as observations. Before the extraction of principal component, the data were centered scaled (Ctr). The other standard scaling methods such as unit variance (UV) and non-scaling were briefly investigated, but showed a slightly worse performance regarding quantification and were therefore suspended in advantage for the centered scaling. Regarding the observations included in the PCA two main approaches were tested, where either the PCA was applied on the spectra of all physical mixtures or applied just on the spectra of pure crystalline and pure amorphous samples. The results of the analyses did show only minute differences in the loading-values and therefore the latter, simpler, approach was chosen.

Using the PCA loadings, the score (t) for the 1 st principal component (PC1) of each physical mixture and milled sample spectra (after centering) was calculated. The fraction of amorphous content (f_{am}) could then be obtained from the score according to Eq. (4):

$$f_{am} = \frac{t_c - t}{(C-1)t - (C \cdot t_a - t_c)} \qquad (4)$$

Where t_c and t_a is the averages of the score for PC1 of pure crystalline sample, and pure amorphous sample respectively and C is the estimated integrated intensity ratio between pure amorphous and pure crystalline spectra. The derivation of Eq. (4) and the procedure to determine C is given in Appendix A. The scores and quantification calculations were carried out on an EXCEL spread sheet where in the determination of C, the with the software provided solver function was used.

2.4.2. Analysis by classical least squares (CLS)

The measured spectra (in the selected wavenumber region of Raman spectra pre-processed as in *A-F*) were expected to be the sum of amorphous and crystalline spectra weighted by the proportions, which gave the following relation (Eq. (5)):

$$\bar{y}_{synth} = \frac{(1 - f_{am}) \cdot \bar{y}_c + f_{am} \cdot C \cdot \bar{y}_a}{(1 - f_{am}) + f_{am} \cdot C} \qquad (5)$$

where f_{am} is the proportion of amorphous content of the measured sample, $\bar{y}_c$ and $\bar{y}_a$ are the vector representations of the normalized crystalline spectrum and the normalized amorphous spectrum respectively and $\bar{y}_{synth}$is the vector representation of the resulting synthesized spectrum, mimicking the normalized measured spectrum. The derivation of Eq. (5) and the procedure to determine C is given in Appendix A. The fraction of amorphous content (f_{am}) for a sample of unknown composition could be obtained by an ordinary curve-fitting procedure, i.e. by minimizing the sum of squares (Eq. (6)):

$$\sum_{i=1}^{n} \left(y_i - y_{synth,i}\right)^2 \qquad (6)$$

where y_i and $y_{synth,i}$ are the i-th components of the sample spectrum and $\bar{y}_{synth}$ respectively, according to the equation above (Eq. (5)). The calculations and curve fitting (by the Levenberg-Marquardt method) were carried out on am EXCEL spreadsheet linked to a visual basic macro.

2.4.3. Peak parameter analysis by PeakFit

The software PeakFit (version 4.12, Systat Software, Inc., San Jose, USA) was used to extract peak parameters from the raw Raman spectra (Fig. 3). Prior to the peak fitting, several pre-processing methods were applied (*A*, *B*, *C*, *F*, *G* described in the pre-processing section), where after the pre-processed data were smoothed by Savitzky-Golay algorithm for noise reduction.

The number of peaks to be examined in selected wavenumber region (315–496 cm^{-1}) was set to seven (Fig. 4), based on the outlook of the spectra in the region. An initial manual fit of the seven peaks (assumed having Voigt profiles) was done, followed by an automated fit using the Levenberg-Marquardt algorithm, where the peak amplitudes, peak centers, full peak width at half-maxima (FWHM) and full base width for each peak were obtained. The factors used as indicators for the goodness of the fit were coefficient of the determination and the F-statistic for the fit.

In order to sift out the most potent parameters for quantification, a pre-investigation of the correlation or anti-correlation between the obtained peak parameters and fraction amorphous lactose was performed. Based on calculated correlation coefficients between fraction amorphous lactose and the different peak parameters and by visual inspection of the corresponding graphs, it was concluded that peak areas and peak amplitudes showed the best correlation or anti-correlation to fraction amorphous lactose.

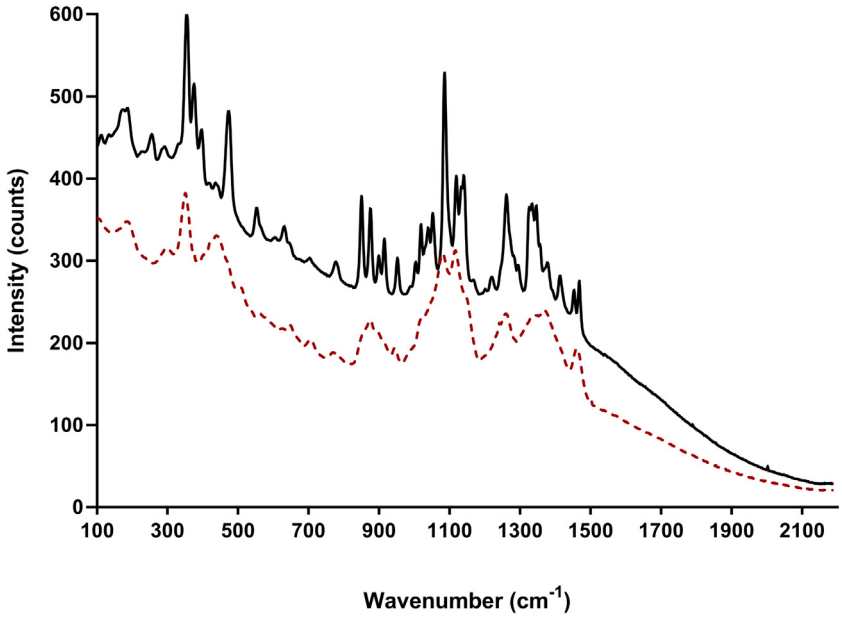

Fig. 3. Raw Raman spectra of pure crystalline α-lactose monohydrate (the black line) and pure amorphous lactose (red dashed line) before baseline correction.

Different ways using the peak data to obtain measures for quantification were further investigated by single peak amplitudes (I), summed peak areas (A_S), the ratios of peak areas (A_R) and ratios of peak amplitudes (I_R).

From such measures, the fraction of amorphous content (f_{am}) could then be obtained according to Eq. (7):

$$f_{am} = \frac{r_c - r}{(C-1)r - (C \cdot r_a - r_c)} \quad (7)$$

where r_c is the average of the I, A_S, A_R or I_R measure of pure crystalline sample, r_a is the average of the I, A_S, A_R or I_R measure of pure amorphous sample, r is the I, A_S, A_R or I_R measure of the sample under investigation and C is the estimated integrated intensity ratio between pure amorphous and pure crystalline spectra. The derivation of Eq. (7) and the procedure to determine C is given in Appendix A. It should here be mentioned that applying Eq. (7) on peak area ratios and peak intensity ratios, the obtained quantification methods become almost identical to the quantification method suggested by Taylor and Zografi (Taylor and Zografi, 1998) (Kontoyannis et al. (Kontoyannis et al., 1997), and Roberts et al. (Campbell Roberts et al., 2002)).

2.5. Quantification of milled samples

Milled lactose samples were analyzed by all above described quantification methods and the amount of apparent amorphous fractions (w/w%) after 10 min, 300 min and 1200 min milling were determined.

2.6. Calculation of the percentage error due to propagated spectral noise

To estimate the spectral noise level by the variance, ten repeated measurements was performed on the same rotating sample, with duration of 30 s of, i.e. 1/10 of the time of the standard measure time in this work. The variance was then calculated at each wavenumber and pooled, where after the variance of a 300 s measurement was estimated by dividing the pooled variance with

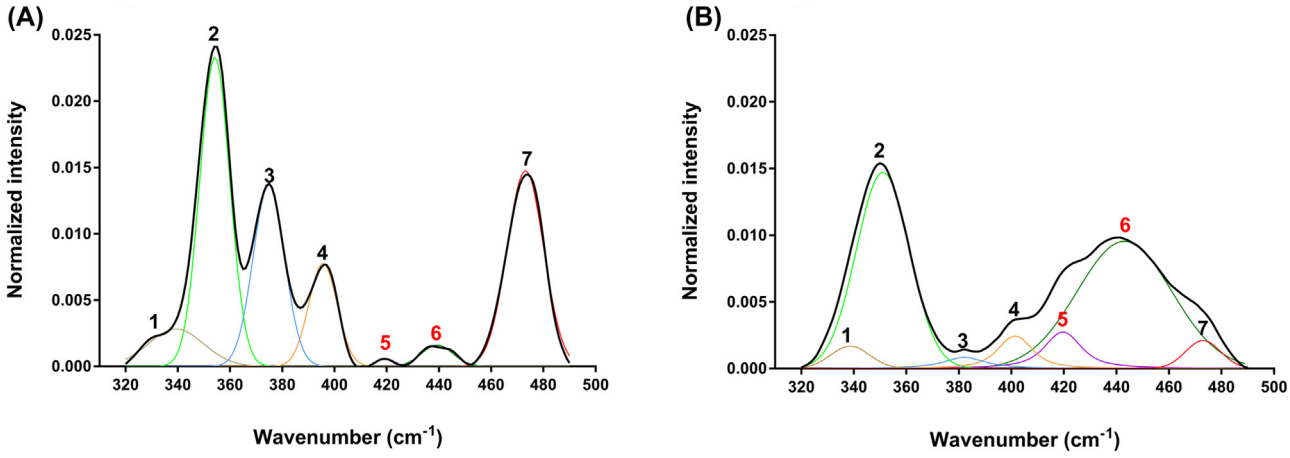

Fig. 4. Spectra of crystalline lactose (A) and 100 w/w% amorphous lactose (B) in the selected wavenumber region (black lines). The spectra are pre-processed by method *A* and Savitsky-Golay filtered (quadratic, window size = 9). The single peaks (coloured lines) are extracted with aid of the PeakFit software. Peaks co-variating with increasing amount of amorphous lactose are labelled with red numbers and peaks co-variating with increasing amount of crystalline lactose are labelled with black numbers.

the factor 10. To simplify the succeeding analysis, the variance of the noise was assumed being constant (σ_z^2) within the wavenumber region used in the analysis. The effect of the spectral noise propagated into the standard deviation (σ_f), of determined amorphous content was then calculated according to Eq. (8) (see Appendix B for derivation):

$$\sigma_f = \frac{C \cdot (t_a - t_c)}{[C(t - t_a) - (t - t_c)]^2} \sqrt{\frac{\sigma_{num}^2}{A^2} + N^2 \frac{\sigma_{area}^2}{A^4}} \tag{8}$$

where

$$\sigma_{num}^2 = \sigma_z^2 \cdot \sum_{i=1}^{n-1} p_i^2 + \frac{2\sigma_z^2}{\Delta x_{range}^2} \left(\sum_{i=1}^{n-1} p_i \cdot x_i \right)^2 \tag{9}$$

$$\sigma_{area}^2 = (n-1) \cdot \sigma_z^2 \cdot (\Delta x)^2 + \frac{\sigma_z^2}{2} \cdot (n-1)^2 \cdot (\Delta x)^2 \tag{10}$$

$$N = \sum_{i=1}^{n-1} p_i \cdot z_i - k \cdot \sum_{i=1}^{n-1} p_i \cdot x_i \tag{11}$$

and were x_i is the wavenumber difference to the center-wavenumber of the spectral portion under analysis, $\Delta x = x_i - x_{i-1}$, A is the norming area under the baseline corrected spectrum and p_i is the loading for PC1 for wavenumber i.

3. Results

3.1. Characterization by XRPD

The X-ray diffraction pattern of α-lactose monohydrate as received corresponded to its theoretical diffraction pattern and confirmed its crystalline nature. Spray dried lactose was completely amorphous, indicated by a "halo" pattern and the absence of distinct diffraction reflections. Fig. 5 shows the experimental X-ray diffraction pattern of the crystalline α-lactose monohydrate, amorphous lactose and milled lactoses of milling times of 10 min, 300 min and 1200 min. For the milled samples, the intensity of peaks became lower and the peaks became slightly broader and merged with increasing milling time.

3.2. Qualitative observations of relations between Raman peaks and amorphous content

The selected region of spectrum appeared to significantly differentiate between mixtures of various composition of crystalline and amorphous lactose (Fig. 6). By the seven peaks found in the selected region, the amplitudes and areas of peaks 5 and 6 were co-variating with increasing amount of amorphous lactose in physical mixtures and thus were referred to as "amorphous peaks" and on the contrary, amplitudes and areas of peaks 1, 2, 3, 4, and 7 were co-variating with increasing amount of crystalline lactose and thus were referred to as "crystalline peaks" (Fig. 4).

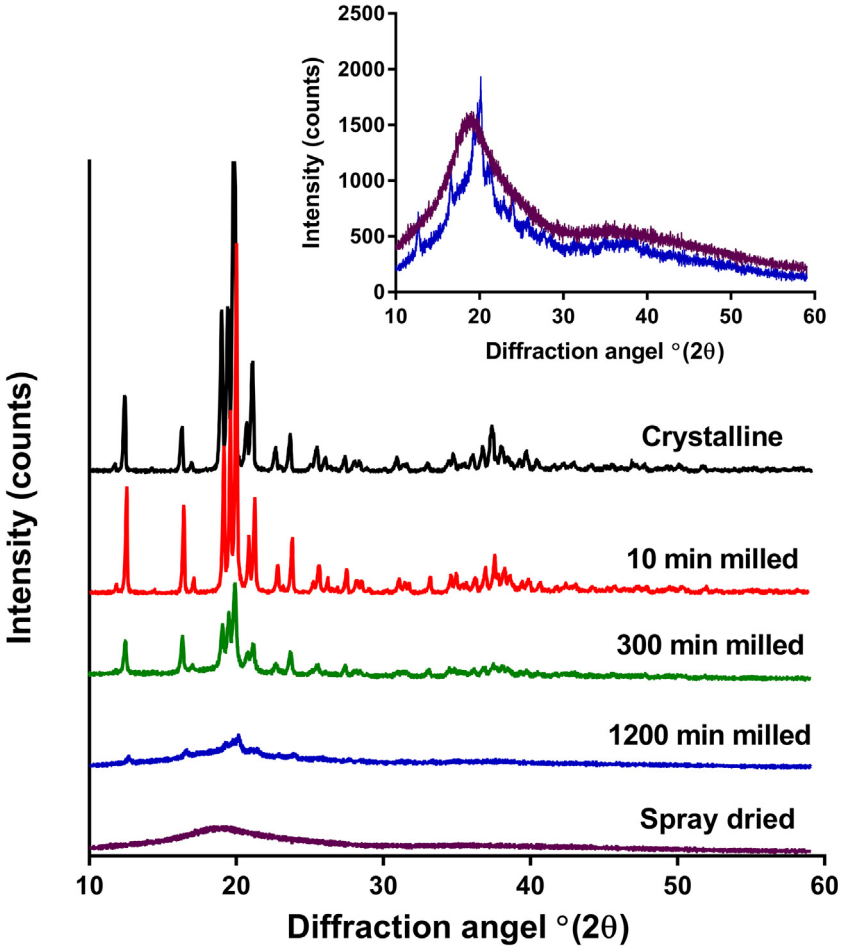

Fig. 5. X-ray powder patterns of crystalline α-lactose monohydrate, amorphous lactose (spray dried) and milled lactose. The insert shows an expanded view of the weaker intensity diffractograms of 1200 min milled lactose (blue) and spray dried lactose (purple).

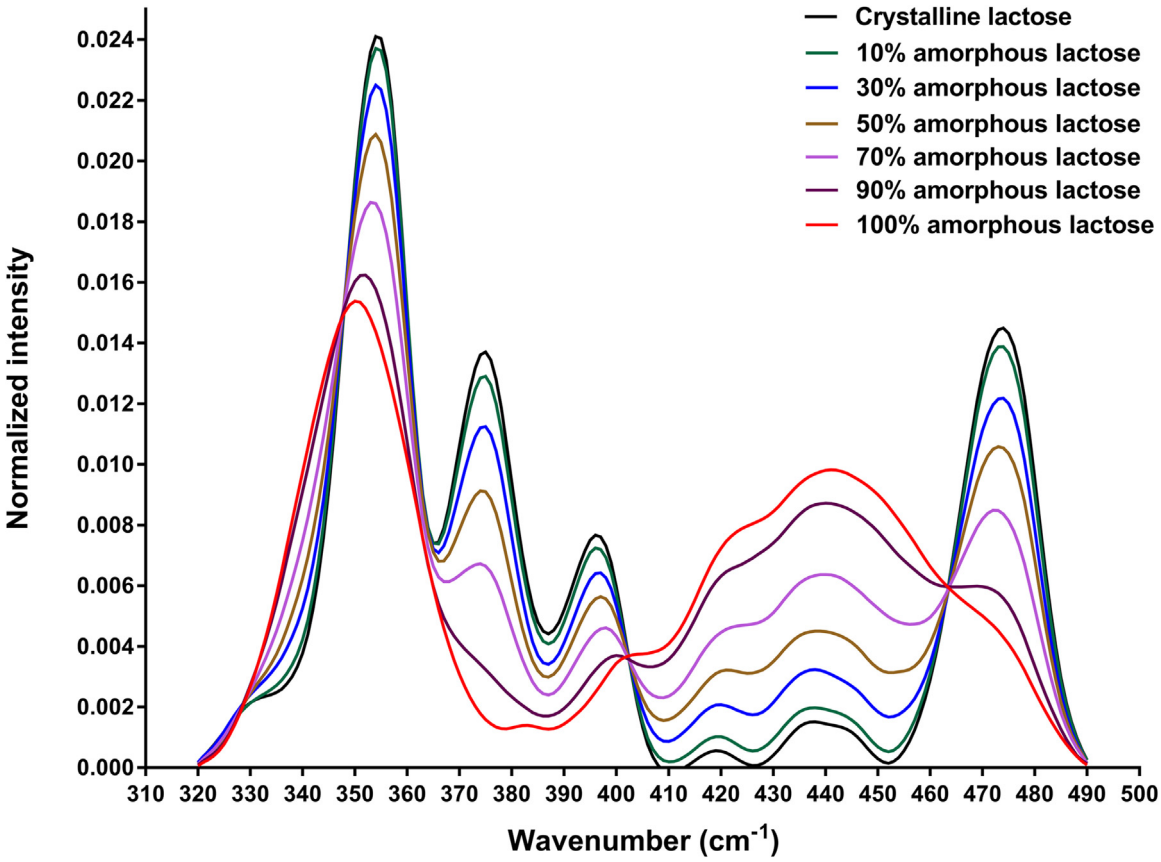

Fig. 6. Raman spectra of physical mixtures containing merely 0, 10, 30, 50, 70, 90, and 100% amorphous lactose (w/w%). The spectra are pre-processed by method *A* and Savitsky-Golay filtered (quadratic, window size = 9). Black line indicates pure crystalline samples; red line indicates pure amorphous samples.

With increasing content of amorphous lactose, FWHM was broadening in peaks with amorphous character (e.g. peak 6), but was generally quite stable in crystalline peaks (e.g. peak 7). In addition to peak broadening, peak overlapping and merging was observed with increasing content of amorphous lactose (Figs. 4 and 6). Peak centra were also shifting to some extent, mostly in peak 3 (from 375 to 383 cm^{-1}), peak 6 (from 439 to 444 cm^{-1}) and peak 4 (from 396 to 402 cm^{-1}) and to least extent in peak 1 (from 339 to 336 cm^{-1}), peak 2 (from 354 to 351 cm^{-1}), peak 5 (from 419 to 420 cm^{-1}) and peak 7 (from 473 to 474 cm^{-1})

To be noted in the crystalline spectrum (Fig. 6), there are small humps on the high wavenumber flank of peak 6 and the low wavenumber flank of peak 1, which probably are indications of smaller peaks merged with a stronger neighboring peak. These probable peaks have not been considered in the peak fitting procedure by the limiting of the number of peaks in order to reduce

Table 1
Summary of quantification analysis of amorphous fraction by different pre-processing methods of physical mixtures using principal component analysis (PCA).

Pre-processing method	*A*	*B*	*C*	*D*	*E*	*F*
Nominal amorphous content (w/w%)	Calculated amorphous content (w/w%)					
0	0.00 ± 0.94	0.00 ± 0.97	−0.01 ± 1.57	−0.09 ± 4.20	0.00 ± 1.91	−0.01 ± 1.91
5	4.40 ± 1.43	4.31 ± 1.44	2.33 ± 2.00	7.30 ± 1.67	5.02 ± 1.94	5.02 ± 1.94
10	9.58 ± 1.06	9.56 ± 1.15	7.27 ± 1.10	9.10 ± 1.65	10.50 ± 1.22	10.50 ± 1.22
15	12.93 ± 1.50	12.86 ± 1.57	11.85 ± 1.56	14.18 ± 2.70	14.25 ± 1.77	14.25 ± 1.77
30	30.42 ± 0.43	30.33 ± 0.36	28.16 ± 0.80	30.37 ± 0.13	30.88 ± 0.33	30.88 ± 0.33
50	49.84 ± 2.05	50.19 ± 1.97	49.54 ± 1.98	47.86 ± 2.28	49.91 ± 2.05	49.91 ± 2.05
70	70.22 ± 1.95	70.21 ± 1.99	71.99 ± 2.23	67.54 ± 2.85	69.10 ± 2.11	69.10 ± 2.11
85	85.85 ± 1.37	85.57 ± 1.37	88.87 ± 0.92	84.67 ± 1.38	84.40 ± 1.37	84.40 ± 1.37
90	91.64 ± 0.75	91.35 ± 0.72	94.02 ± 0.58	90.77 ± 0.55	90.41 ± 0.67	90.41 ± 0.67
95	95.95 ± 0.57	95.77 ± 0.51	97.24 ± 0.26	95.52 ± 0.51	95.18 ± 0.47	95.18 ± 0.47
100	100.00 ± 0.25	100.00 ± 0.25	100.00 ± 0.21	100.00 ± 0.39	100.00 ± 0.28	100.00 ± 0.28
Average SD	1.12	1.12	1.20	1.66	1.28	1.28
Max SD	2.05	1.99	2.23	4.20	2.11	2.11
RMSD	1.02	0.95	2.75	1.43	0.57	0.57
MD	2.07	2.14	4.02	2.46	0.90	0.90

Raw spectra were pre-processed using methods *A-F* as described in method section. The performance of results was summarized in average standard deviation (average SD), max standard deviation (max SD), root mean square deviation (RMSD) and max deviation (MD).

Table 2
Summary of quantification analysis of amorphous fraction by different pre-processing methods of physical mixtures using the classical least squares method (CLS).

Pre-processing method	*A*	*B*	*C*	*D*	*E*	*F*
Nominal amorphous content (w/w%)	Calculated amorphous content (w/w%)					
0	−0.01 ± 1.22	−0.01 ± 1.25	−0.01 ± 1.16	−0.08 ± 3.91	0.00 ± 1.92	−0.02 ± 2.47
5	4.55 ± 1.30	4.48 ± 1.32	2.95 ± 1.77	5.66 ± 1.64	5.07 ± 1.95	5.59 ± 1.87
10	9.73 ± 1.40	9.75 ± 1.51	7.52 ± 1.16	9.59 ± 2.18	10.58 ± 1.23	11.13 ± 1.59
15	13.69 ± 1.78	13.66 ± 1.87	12.80 ± 2.15	14.83 ± 2.81	14.37 ± 1.78	15.63 ± 2.17
30	30.13 ± 0.50	30.26 ± 0.37	27.06 ± 0.84	29.95 ± 0.14	31.06 ± 0.33	31.85 ± 0.22
50	49.81 ± 1.96	50.53 ± 1.89	48.54 ± 2.09	48.34 ± 2.03	50.17 ± 2.03	51.30 ± 1.98
70	69.72 ± 1.89	69.84 ± 1.93	70.20 ± 2.35	67.81 ± 2.25	69.30 ± 2.10	69.56 ± 2.07
85	85.53 ± 1.47	84.02 ± 1.59	86.45 ± 1.08	83.09 ± 1.46	84.49 ± 1.38	84.06 ± 1.46
90	91.40 ± 0.81	90.04 ± 0.74	92.35 ± 0.79	89.36 ± 0.58	90.45 ± 0.66	90.03 ± 0.67
95	95.50 ± 0.67	94.28 ± 0.52	95.75 ± 0.42	94.00 ± 0.57	95.15 ± 0.46	94.52 ± 0.49
100	100.00 ± 0.25	100.00 ± 0.24	100.00 ± 0.32	100.00 ± 0.50	100.00 ± 0.27	100.00 ± 0.30
Average SD	1.22	1.20	1.28	1.64	1.28	1.39
Max SD	1.96	1.93	2.35	3.91	2.10	2.47
RMSD	0.72	0.67	1.95	1.21	0.57	0.97
MD	1.40	1.34	2.94	2.19	1.06	1.85

Raw spectra were pre-processed using methods *A-F* as described in method section. The performance of results was summarized in average standard deviation (average SD), max standard deviation (max SD), root mean square deviation (RMSD) and max deviation (MD).

the risk of over-fitting and un-robustness of the peak parameter determination.

3.3. Comparison of performance between the different pre-processing and analysis methods

The results by the determined percentages of amorphous content of lactose are shown in the Tables 1–6, where each table reports values from a specific quantification analysis. The determined amorphous percentage values and their standard deviations are given for all physical mixtures (row wise) and for the different pre-processing methods applicable (column wise). Summarized values for each pre-processing method are also given by average and maximum standard deviation of the determinations as well as root mean square deviation (RMSD) and maximum deviation (MD) from nominal percentage of the physical mixtures.

For the peak parameter quantification analyses, in order to get a better overview, the results for the four best performing peak parameter measures for quantification are shown: i) single peak amplitudes by peak 6, ii) summed peak areas of peaks 5 and 6, iii) the ratios obtained by dividing the area of peak 6 by the sum of the areas of peak 6 and 7 and iv) the ratios obtained by dividing the amplitudes of peak 7 by the amplitudes of peak 6. In order to determine the accuracy of the quantification methods, validation curves were constructed by plotting the nominal vs. determined amount of amorphous lactose (w/w%) in the physical mixtures for all acquired calibration curves. Fig. 7 illustrates validation curve for the pre-processing and quantification method here graded being the best performing, i.e. PCA preceded by the method *A* as pre-processing method.

3.4. The results from milled samples

The determined amounts of apparent amorphicity of milled lactose powder are shown in Table 7. The mean of apparent amorphicity values and their standard deviation are given for the three milling times (row wise) and for the different pre-processing methods applicable (column wise). Summarizing values for each pre-processing method are also given by average and maximum standard deviation of the determinations.

The obtained degrees of disordered lactose were inserted into Eq. (5) synthesising corresponding theoretical spectrum of

Table 3
Summary of quantification analysis of amorphous fraction by different pre-processing methods of physical mixtures using analysis based on single peak amplitude of peak 6.

Pre-processing method	*A*	*B*	*C*	*F*	*G*
Nominal amorphous content (w/w%)	Calculated amorphous content (w/w%)				
0	−0.03 ± 2.42	−0.02 ± 2.40	−0.06 ± 3.22	−0.04 ± 3.17	−0.03 ± 2.74
5	4.05 ± 2.56	3.93 ± 2.56	1.87 ± 3.44	1.72 ± 3.17	4.54 ± 2.52
10	9.39 ± 0.92	9.34 ± 0.93	6.45 ± 2.37	7.84 ± 0.95	10.86 ± 1.23
15	11.80 ± 1.13	11.66 ± 1.12	9.06 ± 1.66	10.67 ± 1.38	11.33 ± 1.99
30	29.79 ± 0.60	29.87 ± 0.69	26.38 ± 2.09	30.99 ± 0.56	28.86 ± 0.76
50	48.88 ± 2.11	48.68 ± 2.06	48.77 ± 2.93	49.87 ± 1.86	48.23 ± 2.24
70	71.22 ± 2.76	71.94 ± 1.81	75.13 ± 2.13	71.27 ± 0.45	72.26 ± 1.94
85	87.51 ± 1.46	88.26 ± 1.45	91.36 ± 0.70	86.98 ± 1.43	88.73 ± 1.66
90	92.85 ± 0.97	93.46 ± 1.06	96.79 ± 1.11	91.00 ± 2.20	94.48 ± 1.86
95	96.73 ± 0.45	97.50 ± 0.45	99.44 ± 0.77	97.34 ± 0.38	99.02 ± 0.51
100	99.94 ± 0.79	99.94 ± 0.78	100.08 ± 0.98	99.97 ± 0.76	99.99 ± 1.24
Average SD	1.44	1.35	1.91	1.38	1.52
Max SD	2.76	2.56	3.44	3.17	2.52
RMSD	1.88	2.29	4.77	2.29	2.60
MD	3.20	3.46	6.79	4.33	4.48

Raw spectra were pre-processed using methods *A-F* as described in method section. The performance of results was summarized in average standard deviation (average SD), max standard deviation (max SD), root mean square deviation (RMSD) and max deviation (MD).

Table 4
Summary of quantification analysis of amorphous fraction by different pre-processing methods of physical mixtures using analysis based on summed integrated peak areas of peak 5 and 6.

Pre-processing method	*A*	*B*	*C*	*F*	*G*
Nominal amorphous content (w/w%)	Calculated amorphous content (w/w%)				
0	0.00 ± 0.96	0.00 ± 0.95	−0.05 ± 2.65	−0.02 ± 1.86	−0.01 ± 1.57
5	3.73 ± 1.75	3.60 ± 1.71	0.37 ± 2.92	1.29 ± 2.09	4.83 ± 1.65
10	9.64 ± 0.92	9.59 ± 0.97	6.18 ± 2.94	7.81 ± 1.44	12.15 ± 1.20
15	13.79 ± 1.54	13.63 ± 1.59	11.57 ± 2.28	12.25 ± 2.61	14.19 ± 2.16
30	31.12 ± 1.23	31.96 ± 0.85	27.68 ± 1.81	33.32 ± 0.33	27.97 ± 1.11
50	49.69 ± 4.11	49.50 ± 4.09	50.13 ± 4.41	50.80 ± 4.05	48.18 ± 4.01
70	69.51 ± 2.43	70.03 ± 1.64	71.50 ± 1.69	69.38 ± 0.26	71.55 ± 1.46
85	85.92 ± 1.81	85.59 ± 1.77	89.36 ± 0.97	84.97 ± 1.97	87.10 ± 1.89
90	91.69 ± 0.94	91.43 ± 0.80	95.01 ± 0.94	89.43 ± 2.66	93.32 ± 1.00
95	95.88 ± 0.68	95.76 ± 0.83	97.49 ± 0.75	96.02 ± 0.92	97.82 ± 0.81
100	99.98 ± 1.38	99.98 ± 1.36	99.99 ± 0.65	99.98 ± 1.53	99.98 ± 1.43
Average SD	1.71	1.58	2.08	1.82	1.72
Max SD	4.11	4.09	4.41	4.05	4.01
RMSD	1.02	1.11	3.43	2.10	2.00
MD	1.69	1.96	5.01	3.71	3.32

Raw spectra were pre-processed using methods *A*-*F* as described in method section. The performance of results was summarized in average standard deviation (average SD), max standard deviation (max SD), root mean square deviation (RMSD) and max deviation (MD).

Table 5
Summary of quantification analysis of amorphous fraction by different pre-processing methods of physical mixtures using analysis based on ratios of integrated peak areas of peak 6 and summed areas of peaks 6 and 7.

Pre-processing method	*A*	*B*	*C*	*F*	*G*
Nominal amorphous content (w/w%)	Calculated amorphous content (w/w%)				
0	0.00 ± 0.73	0.00 ± 0.71	−0.04 ± 2.64	0.00 ± 1.16	0.00 ± 1.43
5	2.90 ± 1.62	2.77 ± 1.56	−0.65 ± 2.44	0.95 ± 1.72	4.45 ± 1.61
10	9.05 ± 0.91	8.99 ± 0.94	5.54 ± 3.22	7.50 ± 1.34	11.74 ± 1.17
15	12.46 ± 1.27	12.32 ± 1.32	8.78 ± 1.53	11.31 ± 2.07	13.52 ± 2.23
30	30.15 ± 1.78	31.17 ± 0.60	27.36 ± 1.16	32.24 ± 0.34	27.00 ± 1.13
50	49.59 ± 4.63	49.31 ± 4.59	49.94 ± 4.38	50.39 ± 4.97	49.74 ± 4.71
70	70.50 ± 2.64	71.04 ± 1.79	72.71 ± 1.81	70.26 ± 0.03	71.04 ± 2.21
85	87.29 ± 1.60	86.95 ± 1.51	91.22 ± 1.40	86.10 ± 1.93	87.53 ± 1.66
90	92.55 ± 1.22	92.13 ± 1.14	95.78 ± 0.88	90.64 ± 2.77	92.10 ± 1.39
95	97.17 ± 0.82	96.69 ± 0.81	98.47 ± 2.02	97.52 ± 0.93	97.35 ± 0.69
100	99.99 ± 2.48	99.99 ± 2.40	99.99 ± 0.93	99.99 ± 2.51	99.99 ± 2.32
Average SD	1.83	1.58	2.09	1.79	1.87
Max SD	4.63	4.59	4.38	4.97	4.71
RMSD	1.79	1.74	4.58	2.34	1.89
MD	2.55	2.68	6.22	4.05	3.00

Raw spectra were pre-processed using methods *A*-*F* as described in method section. The performance of results was summarized in average standard deviation (average SD), max standard deviation (max SD), root mean square deviation (RMSD) and max deviation (MD).

Table 6
Summary of quantification analysis of amorphous fraction by different pre-processing methods of physical mixtures using analysis based on ratio of peak amplitudes of peak 7 and peak 6.

Pre-processing method	*A*	*B*	*C*	*F*	*G*
Nominal amorphous content (w/w%)	Calculated amorphous content (w/w%)				
0	0.08 ± 2.05	0.09 ± 2.07	0.07 ± 3.25	0.17 ± 2.88	0.08 ± 2.40
5	4.32 ± 2.37	4.24 ± 2.40	1.93 ± 3.51	2.15 ± 3.08	4.30 ± 2.43
10	9.60 ± 1.04	9.63 ± 1.06	7.03 ± 2.77	8.39 ± 1.09	10.86 ± 1.38
15	11.97 ± 1.06	11.95 ± 1.09	9.56 ± 1.32	11.23 ± 1.56	11.46 ± 1.77
30	30.39 ± 0.84	30.76 ± 0.41	28.88 ± 1.25	31.64 ± 0.49	29.31 ± 0.80
50	49.05 ± 2.33	48.95 ± 2.30	49.19 ± 1.97	50.08 ± 2.41	48.94 ± 2.27
70	70.72 ± 2.56	71.38 ± 1.75	72.61 ± 1.93	70.60 ± 0.21	71.80 ± 1.86
85	86.79 ± 1.24	86.66 ± 1.21	89.12 ± 1.09	86.33 ± 1.33	87.07 ± 1.27
90	92.36 ± 0.74	92.16 ± 0.71	94.25 ± 0.45	90.50 ± 2.51	92.34 ± 0.99
95	97.14 ± 0.74	96.88 ± 0.73	97.78 ± 1.09	97.40 ± 0.82	97.26 ± 0.64
100	100.01 ± 1.22	100.01 ± 1.18	100.00 ± 0.34	100.01 ± 1.23	100.01 ± 1.13
Average SD	1.43	1.29	1.71	1.50	1.49
Max SD	2.56	2.40	3.51	3.08	2.43
RMSD	1.66	1.65	3.32	1.99	1.93
MD	3.03	3.05	5.44	3.77	3.54

Raw spectra were pre-processed using methods *A*-*F* as described in method section. The performance of results was summarized in average standard deviation (average SD), max standard deviation (max SD), root mean square deviation (RMSD and max deviation (MD).

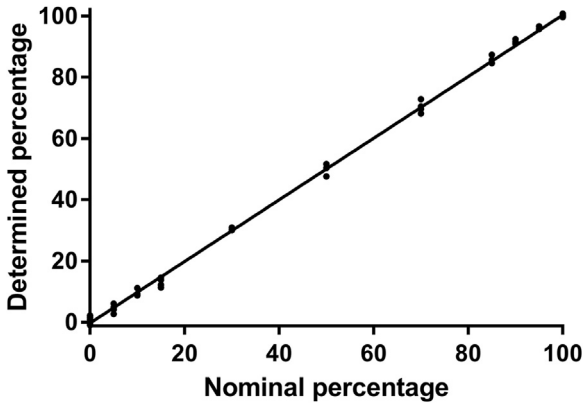

Fig. 7. Validation plot of the best performing quantification analysis, i.e. using pre-processing method *A* combined with PCA. A linear regression with an R^2 value > 0.999 was obtained.

amorphous crystalline mixture. The spectra of milled lactose were similar to synthesized spectra with the corresponding amorphous content (Fig. 8) although some deviations however were obtained for 300 min and 1200 min milled samples relative the corresponding synthesized spectra.

3.5. Theoretical propagation of spectral noise analysis

For the crystalline sample spectrum, the estimated variance was $\sigma_z^2 = 0.20$ ($\sigma_z = 0.45$) in the 320 cm^{-1} to 490 cm^{-1} range. The variances did not show any significant correlation to either wavenumber or intensity spectral intensity and could therefore be regarded as constant in the interval and the approximation of a variance being constant over all wavenumbers in the wavenumber region was regarded appropriate. Putting the above determined σ_z^2-value into Eqs. (8)–(11) gave a $\sigma_f = 0.74\%$ at 0% amorphous content and a $\sigma_f = 0.21\%$ at 100% amorphous content. The standard deviations of amorphous content in the amorphous and crystalline and amorphous samples was experimentally determined to $\sigma_f = 0.94\%$ and $\sigma_f = 0.25\%$ respectively ($n_{cryst} = n_{amorph} = 20$). It could thus be concluded that the theoretically determined standard deviations values are in the same magnitude as, however slightly smaller than, the corresponding experimentally determined values.

4. Discussion

4.1. Physical mixtures

Amongst the four wavenumber regions identified showing differences between amorphous and crystalline lactose, the lowest wavenumber regions (319 cm^{-1}–491 cm^{-1}) were chosen. This region might contain features emanating from lattice vibration (Carteret et al., 2009; Roy et al., 2013), which are expected to disappear in the amorphous state and therefore being a good discriminator between crystalline and amorphous lactose.

The general broadening of peaks, especially peaks 5 and 6, observed in the amorphous lactose spectrum, could be attributed to an ensemble of slightly inequivalent sites. In crystalline lactose, the lattice arrangement of molecules gives that for each vibration mode, the frequency is expected to be the same in molecules in equivalent lattice points. In amorphous lactose, on the contrary, the molecules are not regularly arranged, i.e. the surrounding of every molecule differs and with that a variation in frequencies of each vibration mode are expected leading to a broadening of the Raman peaks.

Table 7
Determined apparent amorphous fraction in milled lactose samples using different pre-processing methods and different quantification approaches.

Pre-processing method	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Milling time (min)	Calculated apparent amorphous fraction (w/w%)						
			PCA				
10	0.84 ± 0.43	0.86 ± 0.51	3.57 ± 0.71	1.64 ± 0.54	2.37 ± 0.15	2.36 ± 0.15	–
300	49.20 ± 1.86	48.96 ± 1.83	51.64 ± 0.67	52.76 ± 2.11	51.54 ± 1.88	51.54 ± 1.88	–
1200	82.36 ± 0.81	82.22 ± 0.82	83.79 ± 1.98	77.94 ± 1.23	81.83 ± 0.79	81.83 ± 0.79	–
			CLS				
10	1.21 ± 0.63	1.24 ± 0.71	4.12 ± 0.78	2.68 ± 0.50	2.40 ± 0.15	2.89 ± 0.30	–
300	51.90 ± 2.35	51.77 ± 2.31	55.78 ± 2.00	54.44 ± 2.40	51.88 ± 1.86	55.45 ± 2.49	–
1200	83.88 ± 0.97	83.81 ± 0.97	85.33 ± 2.57	78.01 ± 0.98	82.08 ± 0.78	84.46 ± 0.96	–
			Single peak amplitude analyses of peak 6				
10	−0.04 ± 0.84	−0.04 ± 0.89	2.26 ± 1.00	–	–	1.81 ± 0.65	0.17 ± 1.14
300	56.18 ± 1.57	55.63 ± 1.41	56.66 ± 0.49	–	–	59.70 ± 1.60	56.56 ± 1.04
1200	92.38 ± 1.04	92.40 ± 1.11	85.54 ± 2.77	–	–	90.95 ± 0.63	93.06 ± 1.01
			Integrated peak areas analyses of peaks 5 and 6				
10	0.14 ± 0.42	0.16 ± 0.51	2.68 ± 0.78	–	–	1.88 ± 0.15	−0.11 ± 1.05
300	47.19 ± 1.78	47.01 ± 1.63	42.68 ± 0.67	–	–	55.14 ± 1.71	48.52 ± 0.98
1200	76.07 ± 1.12	75.95 ± 1.14	72.53 ± 4.73	–	–	77.33 ± 1.04	77.25 ± 1.36
			Ratios of integrated peak areas analyses of peak 6 and summed areas of peaks 6 and 7				
10	0.01 ± 0.23	0.04 ± 0.29	1.41 ± 0.61	–	–	1.34 ± 0.11	−0.29 ± 0.82
300	46.42 ± 1.52	45.98 ± 1.49	36.51 ± 0.71	–	–	52.68 ± 1.45	47.71 ± 1.21
1200	75.40 ± 0.73	75.36 ± 0.76	76.67 ± 5.01	–	–	74.85 ± 0.93	75.19 ± 0.66
			Ratio of peak amplitudes analyses of peak 7 and peak 6				
10	0.28 ± 0.72	0.29 ± 0.78	2.71 ± 0.85	–	–	2.15 ± 0.56	0.43 ± 1.10
300	52.81 ± 1.70	52.63 ± 1.57	53.29 ± 0.63	–	–	56.46 ± 1.62	53.34 ± 1.53
1200	81.15 ± 0.64	81.16 ± 0.69	84.48 ± 1.31	–	–	81.47 ± 0.65	81.46 ± 0.58

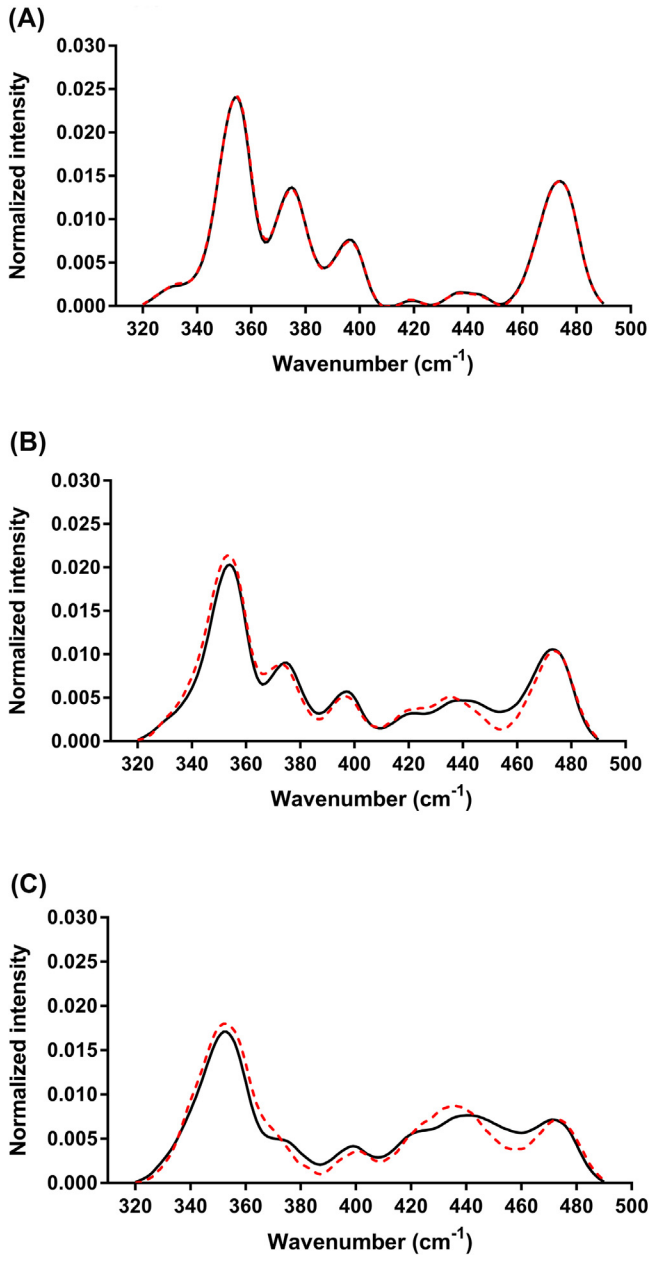

Fig. 8. Pre-processed Raman spectra of milled lactose (red dashed line) and the synthesized spectra (the black line) with the corresponding amorphous content. (A) indicates spectra of 10 min milled lactose; (B) indicates spectra of 300 min milled lactose and (C) indicates the spectra of 1200 min milled lactose. Raman spectra of milled sample are pre-processed using pre-processing method *A* and quantified by PCA. The synthesized spectra are derived from the obtained amorphous content of lactose using PCA.

In this work measure of maximum deviation (MD) and RMSD have been used as measures of accuracy and maximum and average standard deviation as measure of precision for the overall performance. As discussed below, the precision of the amorphicity determination in mixed samples may be unrepresentatively high due to sample heterogeneity. Therefore the precision measures (standard deviation) for the pure crystalline (0% amorphous) and pure amorphous (100% amorphous) samples i.e. where the samples necessarily must be totally homogeneous with regard to composition, is more representative.

Assessing on the basis of these measures, the straight baseline approaches (Method *A* and *B*) combined with quantification by multivariate analysis were graded to be the best performing and most reliable combination of pre-processing and quantification methods. EMSC (Method *E* and method *F*) combined with quantification by PCA actually performs even better than *A* and *B* regarding RMSE and maximum deviation, however the standard deviation in the pure crystalline sample is twice the standard deviation when Method *A* and *B* have been applied. To be noted, the pre-processing methods *A* and *B* and *E* combined with the CLS (synthetization) performs almost as well as when combined with PCA.

The methods *A* and *B* differed only by the Savitsky-Golay filtering of the baseline determining end points and it could be concluded that the filtering had just a minor impact on the baseline and on the quantification. The worse performance using the pre-processing method *C* (polyfit) may be explained by as well too much flexibility of the curvature of the baseline polynomial, as too much flexibility in the choice of data points used for the determination of the baseline, while the assumed positive effect of finding the true curved baseline is subordinated. By the same reasoning the worse performance using pre-processing method *G* might be explained.

The peak parameter based methods showed to be inferior compared to the multivariate methods most probably due to the fact that only a part of the wavenumber region was used in these cases, using less of the available spectral information in the quantification procedure. Also when dealing with overlapping peaks, the peak decomposition procedure may induce uncertainties in the determinations in peak parameters such as peak amplitudes and peak areas. However, the method of choice will depend on the compound in question and unit operation under investigation, thus in certain cases can peak parameter analysis contribute to sufficient information and reliable predictably (Rantanen et al., 2005).

In Appendix B expressions for spectral noise propagating into variation of the determined amorphicity were developed when applying the method combination that was graded performing best in this work, i.e. pre-processing method *A* combined with quantification by PCA. By using these expressions in a theoretical analysis, it was observed that due to spectral noise, the standard deviation in determined amorphicity by the chosen pre-processing and quantification methods and the choice of wavenumber region is expected to be around 0.74% for samples with low amorphous content. The standard deviation successively decreased for samples with higher amorphous content down to 0.2% at pure amorphous sample. The observed difference in variances between samples containing high crystalline content and samples containing high amorphous content is to a large extent a reflection of the PCA score (*t*-value) to fraction value transfer function (Eq. (A.6)) and its derivative (Eq. (B.10)). The derivative is about three times higher at 0% amorphicity compared to 100% amorphicity, which nearly agrees with the observed standard deviation ratios.

By putting spectral data from measurements of pure amorphous and pure crystalline samples in Eqs. B.6-8 in Appendix B, it is shown that the spectral noise contribution to the total variation in determined fraction amorphous is dominated by the impact of the baseline. The impact of the baseline by the endpoints of the spectral interval was a factor 3 (amorphous) to a factor 6 (crystalline) higher compared to the impact of spectral points within the spectral interval.

The experimentally determined standard deviations of amorphicity were around a factor 0.25 larger than the above theoretically obtained standard deviations both for the pure amorphous and pure crystalline samples. This discrepancy of standard deviation is due to unknown sources of error, but is most probable partly due to the fact that the baseline to some degree is approximate, leading to additional variations in the normed spectra wherefrom the quantifications are made.

Knowing the sources of error and the magnitude of influence on the variation in determined amorphicity gives a hint of where to put the effort if improved precision in the determinations are required. According to the relations derived in Appendix B, spectral noise induced standard deviations of the determinations can be decreased by increasing the intensity of the measured spectra by the exposure time and/or exposure intensity in the measurements and in the limit of infinite exposure in principle be reduced to zero. However considerations must be made that increasing the exposure might lead to changes of the sample and sample structure e.g. by heating. The remaining error, due to the above mentioned unknown sources, is then close to the theoretical minimum error, applying the chosen pre-processing and quantification methods and the choice of wavenumber interval. One way of reducing them would, if the speculation above is correct, to further investigate and refine the preprocessing approaches. Investigating further other quantification methods or wavenumber interval might also be a strategy of such issue. The relations derived in Appendix B could also be applied on any wavenumber region and any binary system applying the chosen preprocessing and quantification methods.

The magnitude of the variation of determined amorphicity for the pure crystalline and amorphous samples is thus regarded as being due to the measuring and analysis procedure. Any larger observed variation, such as clearly observed for the physical mixtures, are here regarded being indications of additional sources of variation, such as heterogeneity in the samples. Indications of larger variations, however smaller than for the physical mixtures, of determined amorphicity are also seen in the milled samples indicating a small degree of heterogeneity also in these samples. By the small statistical basis, a more extensive investigation would be required to determine the role of heterogeneity for the final variation of determined amorphous content

The effect of the sample heterogeneity is clearly illustrated by a pre-trial series of Raman measurements made on physical mixtures with the sample fixed, which resulted in remarkably higher variances of the determined fractions amorphous substance (data not shown) than on the series where the samples were rotated. In the rotating set up, the Raman signal sampling area became considerable larger, smearing out the heterogeneity by the averaging effect. This observation also supports the use of the more favorable and easily gained measuring set up by rotating the sample during the measurements.

4.2. Milled powders

Determination of the amount of apparent amorphous content of each milled powder were made using all the pre-processing and quantification methods that are described in the method section (Table 7). Regarding the average of the determined percentage of apparent amorphous content, a dependency of pre-processing and quantification methods used was notable and the effect of the choice of pre-processing and quantification seemed more pronounced for the milled powders than for the physical mixtures. This discrepancy in obtained percentage values could be explained by, that the relative intensity of different peaks obviously differs between milled samples and physical mixtures together with the quantification methods are considering different parts of the spectra differently for the percentage determination, see Fig. 8. The latter is most obvious in the peak parameter methods where one or two peaks are used while spectral values in the wavenumber regions covering the remainder of the peaks are not at all considered. Also in the PCA quantification this is the case, since the loadings are giving different weight at different wavenumbers. Nevertheless, a careful selection of the data handling procedure in the spectrum analysis is critical, especially in the determination of low levels of apparent amorphous content. As a generalisation, the Raman analysis gave indications of the apparent amorphicity of the samples milled for 10, 300 and 1200 min of around 1%, 49% and 82% respectively.

The validation of the obtained apparent amorphous content for the milled powders is problematic since the actual contents of amorphous material of the milled particles are not known. However, the X-ray diffraction measurements gave supporting indications of the micro-structure of the bulk of the particles. The diffraction pattern for powders milled for 10 min was similar to the original powder but the intensity of the peaks was lower and it is concluded that the shortest milling time gave a limited reduction in crystallinity of the particles. For powders milled for 300 min, a further reduction in the intensity of the peaks together with slight broadening was observed. This indicates that milling for 300 min gave a considerable increased degree of disorder in the particles. The diffraction pattern for powders milled for 1200 min finally was characterised by a halo with less distinct and further broadened peaks of low intensity. It is concluded that milling for 1200 min gave particles of high degree of disorder and the halo pattern indicates presumable amorphous domains. Supported by the qualitative results of X-ray diffraction measurements, it thus could be concluded that the values of apparent amorphous content obtained by Raman spectroscopy were reasonable.

Although amorphous particles are described in the literature as either defective or partially amorphous there is in practice difficult to unambiguously differentiate between these two physical natures of amorphous material. It is argued (Chamarthy and Pinal, 2008) that the processes of forming defects or fragmenting crystallites precedes the formation of amorphous regions, i.e. a sequential amorphisation process. Mechanistically this is described as an accumulation of defects or a reduction in size of crystallites and when the concentration of defects within a particle region becomes high it will subsequently lead to the formation of an amorphous phase. If a spatial distribution of amorphous domains exists within a particle, a two-state model is applicable to describe the physical nature of the particle. In addition, it has also been argued (Pazesh et al., 2013) that an amorphous phase can be formed by a vitrification process during handing of powders due to particle-particle collisions involving sliding and friction. The amorphous phase thus formed will be peripherally located in the particle. The exact physical nature of milled particles is thus an intricate issue and the representativeness of the calibration procedure used in Raman spectrum analysis involving reference samples of physical mixtures can be questioned. In Fig. 8, Raman spectra of milled samples were compared to synthesized theoretical spectra with the corresponding amorphous content as the milled samples. The synthetized spectra generally captured well the overall profiles of the measured spectra for the milled samples. The agreement was dependent on the wavenumber region and the apparent amorphous content of the samples. For the lowest apparent amorphous content, the two spectra coincided as expected, since the sample is almost entirely crystalline. For the higher apparent amorphous content, the peaks numbered 2, 3, 4 and 7, i.e. peaks with high relative intensity in the crystalline sample, more or less coincided with the corresponding physical mixture, while the peaks numbered 5 and 6, i.e. peaks with high relative intensity in the amorphous sample were merged and deviated relative the corresponding physical mixture. It is however concluded that by synthesising a spectrum using the crystalline and amorphous spectra, it is possible to obtain a reasonable description of the spectroscopic pattern of the milled material. It seems natural to think that milling thus results in the formation of amorphous regions and that the apparent amorphous content of the milled particles predominately is explained by the formation of amorphous regions, located at the surface as a film of amorphous

material or dispersed within the particles as amorphous domains. However, there were observable differences between measured and synthesized Raman spectra which may indicate that the nature of the milled particles cannot be described solely by a two-state system with amorphous and crystalline domains. The apparent amorphous content determined from Raman spectra of the milled particles may thus represent an indication of amorphous domains, other detecting elements such as dislocation and crystallite boundaries and transition boundaries between amorphous and crystalline phases. The peak broadenings observed in the X-ray diffractograms may support that defects and fragmented crystallites existed in the milled particles. However, further amorphous response could occur if the crystallites or domains of order are so small that the diffraction features becomes too broad and weak to be observed and/or not supporting phonon modes. Nevertheless, to better understand the disordering mechanisms of a crystalline solid during milling, the use of low frequency Raman regions that offer a more sensitive probe of the lattice vibration of a crystalline structure (Hédoux et al., 2011; Mah et al., 2015) could be of potential value.

To give a general advice regarding pre-processing and quantification methods would be questionable and most probable highly dependent on the systems that are to be studied. The intensity of fluorescence, the spectral outlook and the difference in spectra between the different constituents of the system may be parameters that are important for what methods actually will be optimal for the quantification. The conclusions based on the experiments done here are, to be strict, just applicable on amorphous/crystalline α-lactose system, however the methods and design of investigation could be applied on other systems, any frequency region and with other spectroscopies, such as IR- and Microwave spectroscopies. To be noted the here suggested methods relies on that the spectra of samples containing different fractions of the two species could be described as linear combinations of the spectra of the pure species. Therefore, care must be taken how well this actually is applicable on the studied system for the spectral method used and in the spectral range used for the quantification. If departures from linear combinations occur in the spectra of the mixed system, considerations if the here suggested method approximate the mixed system well enough or if modifications of the method has to be done. The conclusions from the theoretically derived equations on the other hand could be regarded as more general.

5. Conclusions

The result of the here done pre-processing and quantification method comparison indicated that the two straight baseline pre-processing methods gave the best performance for lactose. Multivariate analysis methods generally performed superior to the peak parameter methods.

By optimizing the experimental set-up, a precision in the estimate of amorphous content below 1% could be achieved. The experimental precision is slightly worse than could be expected from theoretical calculations of propagated spectral noise. The development of theoretical propagation of error expressions made it possible to estimate the importance of difference noise sources, and this could be an aid if improvement of the precision of the measurements are to be made in the here studied or in similar systems. From the expressions it could also be read out the importance of the baseline approaches for the variation of determined amorphicity.

This study demonstrates that Raman spectroscopy proved to be an appropriate and effective technique in the quantification of apparent amorphous content of milled lactose powder. This was indicated by the achieved reasonable conformity between spectra of milled sample and synthesized spectra with corresponding amorphous content.

Acknowledgments

Recipharm Pharmaceutical Development AB, Solna, Sweden is gratefully acknowledged for funding. The authors are grateful to Prof. Peter Lazor, Department of Earth Sciences, Mineralogy Petrology and Tectonics, Uppsala University, Sweden, for providing the Raman instruments used in this research and for his helpful comments on an earlier draft. The authors wish to thank Joel Hellrup, Department of Pharmacy, Uppsala University, Sweden for help with X-ray measurement and Göran Frenning, Department of Pharmacy, Uppsala University, Sweden for helpful discussion of the calculation equations in appendixes.

Appendix A.

Derivation of equations used in the modified CLS analysis and derivation of a transformation equation from a scalar (e.g. PCA score values, peak values or peak ratios) to fraction amorphous lactose

Due to the classical least squares method CLS spectrum, the spectrum vector $\bar{y}^+$ of a mixed sample is assumed to be the sum of pure amorphous and pure crystalline spectrum vectors (y_a^{-+} and y_c^{-+} respectively) weighted by the fraction amorphous (f_{am}) and fraction crystalline lactose (1-f_{am}) respectively (Haaland and Thomas, 1988). The relation is expressed in Eq. (A.1):

$$\bar{y}^+ = (1 - f_{am}) \cdot \bar{y}^+{}_c + f_{am} \cdot \bar{y}^+{}_a \tag{A.1}$$

However, due to the normalization procedure in the pre-processing, every spectrum was given the same integrated intensity, or corresponding norming quantity, in the wavenumber region used for quantization. Dealing with normalized spectra, the crystalline and amorphous spectra have to be multiplied by their norming quantities, C_c and C_a, respectively, if to be used in a relation like Eq. (A.1). To obtain a resulting spectrum that also is normalized, the weighted sum has to be divided by the sum of the weights. The following expression is then obtained:

$$\begin{aligned}\bar{y} &= \frac{(1 - f_{am}) \cdot C_c \cdot \bar{y}_c + f_{am} \cdot C_a \cdot \bar{y}_a}{(1 - f_{am}) \cdot C_c + f_{am} \cdot C_a} \\ &= \frac{(1 - f_{am}) \cdot \bar{y}_c + f_{am} \cdot C \cdot \bar{y}_a}{(1 - f_{am}) + f_{am} \cdot C}\end{aligned} \tag{A.2}$$

where C=C_a/C_c and $\bar{y}_a$, $\bar{y}_c$ and $\bar{y}$ are the vector representation of the normed amorphous and crystalline and sample spectra.

Having a sample set of known fractions of amorphous lactose, the ratio factor C could be determined by a non-linear fit, by minimizing the following expression:

$$\sum_{j=1}^{n_j} \sum_{i=1}^{n_i} \left(y_{j,i} - \frac{(1 - f_{am,j}) \cdot y_{c,i} + f_{am,j} \cdot C \cdot y_{a,i}}{(1 - f_{am,j}) + f_{am,j} \cdot C} \right)^2 \tag{A.3}$$

where i runs over the wavenumbers in the spectral portion, j runs over the samples of different amorphous content, $f_{am,j}$ is the fractions of amorphous lactose in sample j, $y_{j,i}$, $y_{c,i}$ and $y_{a,i}$ are the normed spectral intensity at wavenumber i for sample j, for the crystalline sample and the amorphous sample respectively.

Knowing C, a determination of f_{am} could be obtained for any sample by a non-linear fit by minimizing the following expression:

$$\sum_{i=1}^{n_i} \left(y_i - \frac{(1 - f_{am}) \cdot y_{c,i} + f_{am} \cdot C \cdot y_{a,i}}{(1 - f_{am}) + f_{am} \cdot C} \right)^2 \tag{A.4}$$

where y_i is the normed spectral intensity at wavenumber i for the sample.

Applying an analogous reasoning as for the spectrum vectors above, any scalar quantity (r) varying with f_{am}, such as a PCA score (t), peak parameters or from peak parameters derived quantities, could be expressed by fraction weighted sum of the scalar quantity determined in a pure crystalline and amorphous samples respectively. Those relations could thus be expressed as:

$$r = \frac{(1-f_{am})\cdot r_c + f_{am}\cdot C_r \cdot r_a}{(1-f_{am}) + f_{am}\cdot C_r} \tag{A.5}$$

from which f_{am} could be resolved as:

$$f_{am} = \frac{r_c - r}{(C_r - 1)r - (C_r \cdot r_a - r_c)} \tag{A.6}$$

where r_c and r_a are the scalar quantities for the pure crystalline and pure amorphous samples respectively and C_r is the scalar ratio between the crystalline and amorphous spectra.

Having a sample set of known fractions of amorphous lactose, the ratio factor C_r could be determined by a non-linear fit, by minimizing the following expression:

$$\sum_{i=1}^{n}\left(f_{am} - \frac{r_c - r}{(C-1)r - (C\cdot r_a - r_c)}\right)^2 \tag{A.7}$$

Appendix B.

Error analysis: Derivation of equation for propagating error of measured spectral intensity into determined t- and percentage values in the case of linear baseline pre-processing (method A) followed by quantification by PCA

Denoting the measured spectral intensity by z_i, where i is wavenumber index, the recalculated intensity after baseline correction and normalization of the spectra, y_i is determined according to:

$$y_i = \frac{z_i - z_{b,i}}{A} \tag{B.1}$$

where $z_{b,i}$ is the baseline z-coordinate and A is the area under the baseline corrected spectrum.

With a linear background, $z_{b,i}$ could be expressed as $z_{b,i} = k\cdot x_i + z_m$, where x_i is the wavenumber relative the center-wavenumber (405 cm^{-1}) of the spectral portion under analysis, k is the slope of the baseline and z_m is the offset of the baseline at the center wavenumber. In turn, k and z_m are given by:

$$z_m = (z_n + z_0)/2$$

and

$$k = (z_n - z_0)/(x_n - x_0) = (z_n - z_0)/\Delta x_{range} \tag{B.2}$$

where z_n and z_0 are the z-values at the upper (x_n) and lower (x_0) limits respectively of the spectral portion under analysis and $\Delta x_{range} = x_n - x_0$.

Further using the trapezoidal rule, A is given by:

$$\begin{aligned} A &= \sum_{i=1}^{n-1}(z_i - z_{b,i})\cdot\Delta x = \sum_{i=1}^{n-1}(z_i - k\cdot x_i - z_m)\\ &\cdot\Delta x = \sum_{i=1}^{n-1} z_i\cdot\Delta x - k\cdot\Delta x\sum_{i=1}^{n-1} x_i - \frac{z_0+z_n}{2}\cdot(n-1)\\ &\cdot\Delta x = \sum_{i=1}^{n-1} z_i\cdot\Delta x - \frac{z_0+z_n}{2}\cdot(n-1)\cdot\Delta x \end{aligned} \tag{B.3}$$

where $\Delta x = x_i - x_{i-1}$. By the definition of the baseline $z_0 - z_{b,0} = z_n - z_{b,n} = 0$, which has been used in the first equality in Eq. (B.3). Further, by symmetry reasons $\sum x_i = 0$, which has been used in the last equality in Eq. (B.3). Putting these expressions for area and baseline into Eq. (B.1) then gives:

$$y_i = \frac{z_i - kx_i - z_m}{\sum_{i=1}^{n-1} z_i\cdot\Delta x - \frac{z_0+z_n}{2}\cdot(n-1)\cdot\Delta x} \tag{B.4}$$

In PCA analysis, centered scaled values are used, i.e. by subtracting the central spectrum value calculated as $y_{cent,i} = (y_{c,i} + y_{a,i})/2$, where $y_{c,i}$ and $y_{a,i}$ are the spectral value of the pure crystalline and pure amorphous sample respectively. The score (t) for the 1st principal component (PC1) has then been calculated by weighting each centered values by corresponding loading, p_i, for PC1, and then summing these values, i.e.:

$$\begin{aligned} t &= \sum_{i=1}^{n-1} p_i\cdot(y_i - y_{centr,i}) = \sum_{i=1}^{n-1}\frac{p_i\cdot(z_i - z_m - kx_i)}{A} - \sum_{i=1}^{n-1} p_i\cdot y_{centr,i}\\ &= \frac{\sum_{i=1}^{n-1} p_i\cdot z_i - z_m\cdot\sum_{i=1}^{n-1} p_i - k\cdot\sum_{i=1}^{n-1} p_i\cdot x_i}{A} - \sum_{i=1}^{n-1} p_i\cdot y_{centr,i}\\ &= \frac{\sum_{i=1}^{n-1} p_i\cdot z_i - k\cdot\sum_{i=1}^{n-1} p_i\cdot x_i}{A} - \sum_{i=1}^{n-1} p_i\cdot y_{centr,i} \end{aligned} \tag{B.5}$$

The PCA is carried out on normalized and centered data, which leads to that $\sum_{i=1}^{n-1} p_i = 0$, (see Appendix C) giving $z_m\cdot\sum_{i=1}^{n-1} p_i = 0$, which in turn is used in the third equality in Eq. (B.5).

The variance of the score (t), σ_t^2

Since terms with the products of loadings and spectral center values in the last sum in Eq. (B.5) are not changing between the measurements, that sum will not contribute to the variance of t. Thus only the term $\frac{\sum_{i=1}^{n-1} p_i\cdot z_i - k\cdot\sum_{i=1}^{n-1} p_i\cdot x_i}{A}$ will give a contribution. Denoting the numerator $\sum_{i=1}^{n-1} p_i\cdot z_i - k\cdot\sum_{i=1}^{n-1} p_i\cdot x_i = N$, the variance of the score σ_t^2, using Gauss approximation (Bevington and Robinson, 1992), can be expressed as:

$$\begin{aligned} &\frac{\sigma_t^2}{t^2} = \frac{\sigma_t^2}{N^2/A^2} = \frac{\sigma_{num}^2}{N^2} + \frac{\sigma_{area}^2}{A^2} - 2\frac{Cov(N,A)}{N\cdot A} \Rightarrow \sigma_t^2 =\\ &\frac{N^2}{A^2}\left(\frac{\sigma_{num}^2}{N^2} + \frac{\sigma_{area}^2}{A^2} - 2\frac{Cov(N,A)}{N\cdot A}\right) = \frac{\sigma_{num}^2}{A^2} + N^2\frac{\sigma_{area}^2}{A^4} - 2N\frac{Cov(N,A)}{A^3} \end{aligned} \tag{B.6}$$

where σ_{num}^2 is the variance of numerator, σ_{area}^2 is the variance of the area, and $Cov(N,A)$ is the covariance between the numerator and area.

The variance of the numerator

Assuming that the variances, $\sigma_{z,i}^2$, of the measured intensities are equal for each spectral point we set $\sigma_{z,i}^2 = \sigma_z^2 =$ constant at all wavenumbers. The measured intensities could also be assumed being statistically independent. The relation $k = (z_n - z_0)/\Delta x_{range}$ is a weighted sum of z_i-values which then gives:

$$\sigma_k^2 = \frac{\sigma_n^2 + \sigma_0^2}{\Delta x_{Range}^2} = \frac{2\sigma_z^2}{\Delta x_{Range}^2}$$

Using that also the whole numerator can be seen as a weighted sum of z_i-values the following is obtained:

$$\sigma_{num}^2 = \sum_{i=1}^{n-1} p_i^2 \cdot \sigma_{z,i}^2 + \sigma_k^2 \left(\sum_{i=1}^{n-1} p_i \cdot x_i \right)^2 = \sigma_z^2 \cdot \sum_{i=1}^{n-1} p_i^2 + \frac{2\sigma_z^2}{\Delta x_{range}^2} \left(\sum_{i=1}^{n-1} p_i \cdot x_i \right)^2 \quad \text{(B.7)}$$

The variance of the area (denominator)

In a similar way, using that the expression for the area also can be seen as a weighted sum of z_i-values, (see Eq. (B.4)), the variance of the denominator σ_{area}^2 is given by

$$\sigma_{area}^2 = \sum_{i=1}^{n-1} \sigma_{z,i}^2 \cdot (\Delta x)^2 + \left(\frac{\sigma_{z,0}^2}{2^2} + \frac{\sigma_{z,n}^2}{2^2} \right) \cdot (n-1)^2 \cdot (\Delta x)^2 = (n-1) \cdot \sigma_z^2 \cdot (\Delta x)^2 + \frac{\sigma_z^2}{2} \cdot (n-1)^2 \cdot (\Delta x)^2 \quad \text{(B.8)}$$

The covariance between numerator and area

Assuming that the intensity observations at different wavenumbers are statistically independent, the following must be valid regarding the covariances: $Cov(z_i,z_j) = \delta_{i,j}\sigma_z^2$ and $Cov(z_0 + z_n, z_n - z_0) = 0$

This gives:

$$\begin{aligned} Cov(N,A) &= Cov\left(\left[\left(\sum_{i=1}^{n-1} p_i \cdot z_i \right) - \left(\frac{z_n - z_0}{\Delta x_{range}} \cdot \sum_{i=1}^{n-1} p_i \cdot x_i \right) \right], \left[\left(\sum_{i=1}^{n-1} z_j \cdot \Delta x \right) - \left(\frac{z_0 + z_n}{2} \cdot (n-1)\Delta x \right) \right] \right) \\ &= Cov\left(\left(\sum_{i=1}^{n-1} p_i \cdot z_i \right), \left(\sum_{i=1}^{n-1} z_j \cdot \Delta x \right) \right) \\ &+ Cov\left(\left(\sum_{i=1}^{n-1} p_i \cdot z_i \right), \left(\frac{z_0 + z_n}{2} \cdot (n-1)\Delta x \right) \right) \\ &+ Cov\left(\left(\frac{z_n - z_0}{\Delta x_{range}} \cdot \sum_{i=1}^{n-1} p_i \cdot x_i \right), \left(\sum_{i=1}^{n-1} z_j \cdot \Delta x \right) \right) \\ &+ Cov\left(\left(\frac{z_n - z_0}{\Delta x_{range}} \cdot \sum_{i=1}^{n-1} p_i \cdot x_i \right), \left(\frac{z_0 + z_n}{2} \cdot (n-1)\Delta x \right) \right) \\ &= Cov\left(\left(\sum_{i=1}^{n-1} p_i \cdot z_i \right), \left(\sum_{i=1}^{n-1} z_j \cdot \Delta x \right) \right) + [0 + 0 + 0] \\ &= \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} Cov(p_i \cdot z_i, z_j \Delta x) = \Delta x \sum_{i=1}^{n-1} p_i \cdot \sigma_z^2 = 0 \end{aligned} \quad \text{(B.9)}$$

Recalling that $\sum_{i=1}^{n-1} p_i = 0$, the last equality is obtained.

The variance of fraction amorphous lactose determined

According to Appendix A, the fraction amorphous f_{am} is obtained from the score t as:

$$f_{am} = \frac{t_c - t}{(C-1)t - (C \cdot t_a - t_c)},$$

where r has been replaced by t in Eq. (A.6). This gives:

$$\frac{df_{am}}{dt} = \frac{C \cdot (t_a - t_c)}{[C(t - t_a) - (t - t_c)]^2} \quad \text{(B.10)}$$

and Gauss approximation gives in turn:

$$\sigma_f = \frac{df_{am}}{dt}\sigma_t = \frac{C \cdot (t_a - t_c)}{[C(t - t_a) - (t - t_c)]^2}\sigma_t \quad \text{(B.11)}$$

where σ_f^2 is the variance of the determined fraction amorphous lactose.

Appendix C.

Showing that the sum of scores equals zero when PCA is performed on centered values of normalized spectra

$$\bar{y}_{c,centr} = \bar{y}_c - \frac{\bar{y}_c + \bar{y}_a}{2} = \frac{\bar{y}_c - \bar{y}_a}{2} \quad \text{(C.1)}$$

$$\bar{y}_{a,centr} = \bar{y}_a - \frac{\bar{y}_c + \bar{y}_a}{2} = -\frac{\bar{y}_c - \bar{y}_a}{2} \quad \text{(C.2)}$$

Summing the individual centered spectral intensity values gives:

$$\sum_{i=1}^{n-1} y_{c,centr,i} = \frac{1}{2}\sum_{i=1}^{n-1} y_{c,i} - y_{a,i} = \frac{1}{2}\left(\sum_{i=1}^{n-1} y_{c,i} - \sum_{i=1}^{n-1} y_{a,i} \right) = \frac{1}{2} \cdot (1-1) = 0 \quad \text{(C.3)}$$

since by norming both $\sum_{i=1}^{n-1} y_{c,i} = 1$ and $\sum_{i=1}^{n-1} y_{a,i} = 1$

The scores, p_i, are the direction cosines of PC1 relative the wavenumber axes, $\hat{y}_i$. Since the PCA is only carried out on spectra of the pure amorphous and pure crystalline samples, PC1 is the axis crossing the averages of pure amorphous spectra and pure crystalline spectra respectively. Thus:

$$p_i = \frac{y_{c,centr,i}}{\sqrt{\sum_{i=1}^{n-1} y_{c,centr,i}^2}} \Rightarrow \sum_{i=1}^{n-1} p_i = \frac{1}{\sqrt{\sum_{i=1}^{n-1} y_{c,centr,i}^2}} \sum_{i=1}^{n-1} y_{c,centr,i} = \frac{1}{\sqrt{\sum_{i=1}^{n-1} y_{c,centr,i}^2}} \cdot 0 = 0 \quad \text{(C.4)}$$

References

Agatonovic-Kustrin, S., Rades, T., Wu, V., Saville, D., Tucker, I.G., 2001. Determination of polymorphic forms of ranitidine–HCl by DRIFTS and XRPD. J. Pharm. Biomed. Anal. 25, 741–750.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43, 772–777.

Bartolomei, M., Ramusino, M.C., Ghetti, P., 1997. Solid-state investigation of fluocinolone acetonide. J. Pharm. Biomed. Anal. 15, 1813–1820.

Bevington, P.R., Robinson, D.K., 1992. Data Reduction and Error Analysis for the Physical Sciences, 2nd ed. Mc Graw-Hill, New York, U.S.A, pp. 38–47.

Buckton, G., Darcy, P., Mackellar, A.J., 1995. The use of isothermal microcalorimetry in the study of small degrees of amorphous content of powders. Int. J. Pharm. 117, 253–256.

Campbell Roberts, S.N., Williams, A.C., Grimsey, I.M., Booth, S.W., 2002. Quantitative analysis of mannitol polymorphs. X-ray powder diffractometry–exploring preferred orientation effects. J. Pharm. Biomed. Anal. 28, 1149–1159.

Caron, V., Willart, J.F., Lefort, R., Derollez, P., Danede, F., Descamps, M., 2011. Solid state amorphization kinetic of alpha lactose upon mechanical milling. Carbohydr. Res. 346, 2622–2628.

Carteret, C., Dandeu, A., Moussaoui, S., Muhr, H., Humbert, B., Plasari, E., 2009. Polymorphism studied by lattice phonon Raman spectroscopy and statistical mixture analysis method. Application to calcium carbonate polymorphs during batch crystallization. Cryst. Growth Design 9, 807–812.

Chamarthy, S.P., Pinal, R., 2008. The nature of crystal disorder in milled pharmaceutical materials. Coll. Surf. A 331, 68–75.
Fix, I., Steffens, K.J., 2004. Quantifying low amorphous or crystalline amounts of alpha-lactose-monohydrate using X-ray powder diffraction, near-infrared spectroscopy, and differential scanning calorimetry. Drug Dev. Ind. Pharm. 30, 513–523.
Hédoux, A., Decroix, A.-A., Guinet, Y., Paccou, L., Derollez, P., Descamps, M., 2011. Low- and high-frequency Raman investigations on caffeine: polymorphism, disorder and phase transformation. J. Phys. Chem. B 115, 5746–5753.
Haaland, D.M., Thomas, E.V., 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Anal. Chem. 60, 1193–1202.
Heinz, A., Strachan, C.J., Gordon, K.C., Rades, T., 2009. Analysis of solid-state transformations of pharmaceutical compounds using vibrational spectroscopy. J. Pharm. Pharmacol. 61, 971–988.
Hogan, S.E., Buckton, G., 2000. The quantification of small degrees of disorder in lactose using solution calorimetry. Int. J. Pharm. 207, 57–64.
Hogan, S.E., Buckton, G., 2001. The application of near infrared spectroscopy and dynamic vapor sorption to quantify low amorphous contents of crystalline lactose. Pharm. Res. 18, 112–116.
Kaneniwa, N., Imagawa, K., Otsuka, M., 1985. Effect of tabletting on the degree of crystallinity and on the dehydration and decomposition points of cephalexin crystalline powder. Chem. Pharm. Bull. 33, 802–809.
Kontoyannis, C.G., Bouropoulos, N.C., Koutsoukos, P.G., 1997. Use of Raman spectroscopy for the quantitative analysis of calcium oxalate hydrates: application for the analysis of urinary stones. Appl. Spectrosc. 51, 64–67.
Lieber, C.A., Mahadevan-Jansen, A., 2003. Automated method for subtraction of fluorescence from biological Raman spectra. Appl. Spectrosc. 57, 1363–1367.
Luisi, B.S., Medek, A., Liu, Z., Mudunuri, P., Moulton, B., 2012. Milling-induced disorder of pharmaceuticals: one-phase or two-phase system. J. Pharm. Sci. 101, 1475–1485.
Mackin, L., Zanon, R., Park, J.M., Foster, K., Opalenik, H., Demonte, M., 2002. Quantification of low levels (<10%) of amorphous content in micronised active batches using dynamic vapour sorption and isothermal microcalorimetry. Int. J. Pharm. 231, 227–236.
Mah, P.T., Fraser, S.J., Reish, M.E., Rades, T., Gordon, K.C., Strachan, C.J., 2015. Use of low-frequency Raman spectroscopy and chemometrics for the quantification of crystallinity in amorphous griseofulvin tablets. Vib. Spectrosc. 77, 10–16.
Martens, H., Nielsen, J.P., Engelsen, S.B., 2003. Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. Anal. Chem. 75, 394–404.
Niemelä, P., Päällysaho, M., Harjunen, P., Koivisto, M., Lehto, V.P., Suhonen, J., Järvinen, K., 2005. Quantitative analysis of amorphous content of lactose using CCD-Raman spectroscopy. J. Pharm. Biomed. Anal. 37, 907–911.
Otte, A., Zhang, Y., Carvajal, M.T., Pinal, R., 2012. Milling induces disorder in crystalline griseofulvin and order in its amorphous counterpart. CrystEngComm 14, 2560–2570.
Pazesh, S., Heidarian Höckerfelt, M., Bramer, T., Berggren, J., Alderborn, G., 2013. Mechanism of amorphisation of micro-particles of griseofulvin during powder flow in a mixer. J. Pharm. Sci. 102, 4036–4045.
Rantanen, J., Wikström, H., Rhea, F.E., Taylor, L.S., 2005. Improved understanding of factors contributing to quantification of anhydrate/hydrate powder mixtures. Appl. Spectrosc. 59, 942–951.
Roy, S., Chamberlin, B., Matzger, A.J., 2013. Polymorph discrimination using low wavenumber Raman spectroscopy. Org. Process Res. Dev. 17, 976–980.
Savolainen, M., Jouppila, K., Pajamo, O., Christiansen, L., Strachan, C., Karjalainen, M., Rantanen, J., 2007. Determination of amorphous content in the pharmaceutical process environment. J. Pharm. Pharmacol. 59, 161–170.
Shah, B., Kakumanu, V.K., Bansal, A.K., 2006. Analytical techniques for quantification of amorphous/crystalline phases in pharmaceutical solids. J. Pharm. Sci. 95, 1641–1665.
Sheokand, S., Modi, S.R., Bansal, A.K., 2014. Dynamic vapor sorption as a tool for characterization and quantification of amorphous content in predominantly crystalline materials. J. Pharm. Sci. 103, 3364–3376.
Strachan, C.J., Rades, T., Gordon, K.C., Rantanen, J., 2007. Raman spectroscopy for quantitative analysis of pharmaceutical solids. J. Pharm. Pharmacol. 59, 179–192.
Susi, H., Ard, J.S., 1974. Laser-raman spectra of lactose. Carbohydr. Res. 37, 351–354.
Taylor, L.S., Zografi, G., 1998. The quantitative analysis of crystallinity using FT-Raman spectroscopy. Pharm. Res. 15, 755–761.
Young, P.M., Chiou, H., Tee, T., Traini, D., Chan, H.K., Thielmann, F., Burnett, D., 2007. The use of organic vapor sorption to determine low levels of amorphous content in processed pharmaceutical powders. Drug Dev. Ind. Pharm. 33, 91–97.