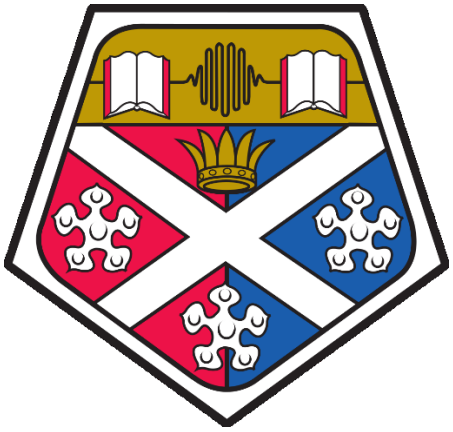
Spatially and angularly resolved diffuse reflectance measurement for in-line analysis of particle suspensions - A multi-sensor approach



Carla Ferreira

Department of Chemical and Process Engineering

University of Strathclyde

This dissertation is submitted for the degree of Doctor of Philosophy

2019

Spatially and angularly resolved diffuse reflectance measurement for in-line analysis of particle suspensions - A multi-sensor approach / by Carla Ferreira. Glasgow: University of Strathclyde, 2019.

Supervisors:

Dr. Yi-Chieh Chen

Prof. Jan Sefcik

This project was based at the Future Manufacturing Research Hub of Continuous Manufacturing and Crystallisation (CMAC), and was associated to the Measurements and Analytics Team (MAT), at the Department of Chemical and Process Engineering.

The work was funded by the EPSRC and the Doctoral Training Centre in Continuous Manufacturing and Crystallisation (CMAC). The contents of the publication reflect only the author view.

Declaration of author's rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Carla Ferreira

September 2019

List of publications

C. Ferreira, J. Cardona, O.S. Agimelen, C. Tachtatzis, J. Sefcik, Y-C. Chen. 'Quantification of particle size and concentration using in-line techniques and multivariate analysis'. Submitted to Powder Technology, 2020.

J. Cardona, C. Ferreira, J. McGinty, A. Hamilton, O.S. Agimelen, A. Cleary, R. Atkinson, C. Michie, S. Marshall, Y-C. Chen, J. Sefcik, I. Andonovic, C. Tachtatzis. 'Image analysis framework with focus evaluation for in situ characterisation of particle size and shape attributes'. *Chemical Engineering Science.* 2018. 191: 208-231.

Y-C. Chen, S. Tiernan-Vandermotten, L. Lue, C. Ferreira, J. Sefcik & S. Thennadil. 'Effect of particle size distribution on spatially and angularly resolved diffuse reflectance measurement'. *European Pharmaceutical Review* 2018. 23(1): 34-37.

J. Cardona, C. Ferreira, J. McGinty, A. Hamilton, O.S. Agimelen, A. Cleary, R. Atkinson, C. Michie, S. Marshall, Y-C. Chen, J. Sefcik, I. Andonovic, C. Tachtatzis. 'The role of in-line image analysis in the transition to continuous manufacturing in the pharmaceutical industry'. *Advances in Transdisciplinary Engineering* 2017. 8: 27-32.

Preface

This thesis presents the research performed during my PhD project at the University of Strathclyde. Most of the work was carried out at the CMAC National Facility supported by UK Research Partnership Fund (UKRPIF), awarded from the Higher Education Funding Council for England (HEFCE) (Grant ref HH13054), and was funded by the EPSRC and the Doctoral Training Centre in Continuous Manufacturing and Crystallisation (Grant Ref: EP/K503289/1). The results presented in this thesis were developed in close collaboration with the Measurements and Analytics Team (MAT).

Acknowledgements

The completion of this thesis would not be possible without the contribution of the following persons or institutions, to which I'm most happy to express my sincere thankfulness:

- to my supervisors, Dr. Yi-chieh Chen and Prof. Jan Sefcik, for the indispensable guidance and support throughout the execution of this work, patience, commitment and for enhancing my experience as a researcher and a person;

- To CMAC and the Chemical and Process Engineering Department that have a great team who are always ready to help.

- To Javier Cardona for all for the guidance, work discussions, contributions, and for being such a kind friend.

- I am truly thankful to Martin Prostredny for being sharing all the moments throughout these 3 years, good or bad, for always listening to me, for his patience and always being there for me.

- To all my friends that contributed for these marvellous years in Glasgow: Carlota, Georgia, Ruben, Eleonora, Frederik, Michael, Rab, David, Dasha, Suse, Bruce, Bilal, Scott, Paul, Alexandra, Catia, Andrew.

- To Tiago for believing in me, for being understanding, patience, and his unconditional support through the good and bad times.

- To Charles for all the support, patience, to help to make it easier these last steps and for all his love.

- To my parents, for believing in my dreams. My kindest thanks for them for the support and understanding through the good and less good times;

-To my sisters, Inês and Beatriz, for their patience, affection and encouragement provided throughout this journey.

Abstract

Particle size and shape are critical quality attributes for active pharmaceutical ingredients (API) as they have direct impact on downstream processing, as well as on the performance behaviour of the finished product. None of the current process analytical technologies (PAT), although capable of providing an indication on these key attributes, measures particle size directly. Obtaining a reliable and robust quantitative information of these particle attributes in real-time remains a great challenge across the multiple manufacturing steps and cannot be achieved by just using a single sensor.

A Spatially and Angularly-Resolved Diffuse Reflectance Measurement (SAR-DRM) technology is for the first time applied to monitor micron size particles. SAR-DRM relies on multiple scattering of particles and collects multi-wavelength (UV-visible-NIR) diffuse reflectance spectra from optical fibres of multi-angle multi-space arrangements. Each SAR-DRM spectrum yields differences which correspond to the light travelling differently in the sample and to being differently affected by scattering and absorption effects. This technique is used alongside with Focus Beam Reflectance Measurement (FBRM) and Particle Vision and Measurement (PVM) which have been widely applied for in-line monitoring of particle attributes in crystallisation processes, in order to examine the SAR-DRM performance and investigate the possibility to broadening the size and concentration ranges of the applications.

The investigation was carried out on two model systems: polystyrene beads suspensions in water and alpha lactose monohydrate suspensions in acetone, for size (<38 to 800 μm) and concentration (0.5 to 25 wt.%) ranges, relevant to many pharmaceutical processes, e.g., during crystallisation and granulation. As particles scatter light differently depending on the size, shape and solid concentration, the spectral changes in SAR-DRM can be related to the sample's attributes. These properties can be either obtained by applying light propagation theory, which involves intensive computational calculations and are challenging to invert the information in real-time, or by multivariate regression analysis, an alternative and faster method. Characterisation of the particles attributes was performed by both in-line and off-line

commercial technologies, and served as an input to validate SAR-DRM sensitivity, accuracy and capability to track the differences in size and solid loading in the model system. Robust calibration models were established to predict particle size and particle concentration for the individual technologies and for the combined measurement data, by applying multivariate regression analysis. The results suggest that the accuracy of multivariate calibration models can be improved by combining key SAR-DRM configurations with FBRM data. This is due to the spectra captured by SAR-DRM containing more complete information about the light scattering properties of samples, which can be related to the particle size and concentration. As SAR-DRM prefers high turbidity samples, i.e., samples with higher solid content, the study shows that SAR-DRM technology can be a potential complementary to other PAT tools such as FBRM and PVM, which tend to perform better for low turbidity samples. Currently, solid concentration is not measured during pharmaceutical processes and particle size is not directly obtained. Our result suggests that combining the strengths of each technique can help to obtain reliable and quantitative information about particle attributes, allowing to achieve robust process monitoring and enable improved control and optimisation of manufacturing processes.

Table of Contents

List of Figures

List of Tables

Abbreviations & Symbols

ATR-FTIR – Fourier Transform Infra-red Spectroscopy

ATR-UV/vis - Ultra Violet/visible Spectroscopy

BSE – Backscattered Electrons

CCD – Charge-coupled Device

CL – Chord Length

CLD – Chord Length Distribution

D1 – 0.3 mm

D2 – 0.6 mm

D3 – 0.9 mm

D4 – 1.2 mm

ECD – Equivalent Circle Diameter

EISC – Extended Inverse Signal Correction

EMA – European Medicines Agency

EMSC – Extended Multiplicative Scatter Correction

ESD – Equivalent Sphere Diameter

FBRM – Focus Beam Reflectance Measurement

FDA – Food and Drug Administration

FTIR – Fourier-transform infrared spectroscopy

ISC – Inverse Signal Correction

L1 – 0°

L2 – 30°

L3 – 45°

LD – Laser Diffraction

LED – Light-emitting Diode

LV - Latent Variables

MB-PLS - Multiblock PLS

MSC - Multiplicative Scatter Correction

NIR – Near Infrared

NIR – Near infrared Spectroscopy

PAT – Process Analytical Tools

PC – Principal Components

PCA – Principal Component Analysis

PCM – Pixact Crystallization Monitoring

PDF – Probability Density Functions

PLSR – Partial Lest Squares Regression

PS – Polystyrene

PSD – Particle Size Distribution

PVM – Particle Vision and Measurement

Rd – Diffuse Reflectance

RMSECV – Root Mean Square Error of Cross-Validation

RMSEP – Root Mean Square Error of Prediction

SAR-DRM – Spatially and Angularly Resolved Diffuse Reflectance Measurement

SE – Secondary Electrons

SEM – Scanning Electron Microscope

SFV – Spatial Filtering Velocimetry

SG – Savitzky-Golay

SNV – Standard Normal Variate

Td – Diffuse Transmittance

Notation & Units

The notation used throughout this thesis is described in this section. Matrices are denoted by capital letters while vectors by denoted by lowercase letters. The matrix inverse is represented as X^{-1}. The transpose of a matrix or a column vector is indicated as, e.g., X^T or p^T, respectively.

B	Regression coefficients
COV(X)	Covariance matrix
E	Residuals matrix of independent variables
e	Residuals vector of independent variables
F	Residuals matrix of dependent variables
p	Loading vector/ eigenvector of independent variables
P	Loading matrix of independent variables
Q	Loading matrix of dependent variables
q	Loading vector of dependent variables
R	Loading weights matrix of independent variables
t	Scores vector of independent variables
u	Score vector of dependent variables
U	Scores matrix of dependent variables
X	Data matrix of independent variables
Y	Data matrix of dependent variables

a	Intercept
b	Slope
i	Sample number
J	Number of LVs
J	Total number of LVs
K	total wavelength
k	Number of columns (or observations)
m_i	Mean spectra of all samples
n	Number of rows (or samples)
R^2	Correlation coefficient
SD_i	Sample's standard deviation
SS_{res}	Sum of squared residuals
SS_{tot}	Total sum of squares
Y	Measured reference value
$Y_{i,cv}$	Estimated reference value from cross-validation
$Y_{i,p}$	Predicted variable
λ	Eigenvalue
Bon, Boff	Baseline light on and baseline light off
Ion, Ioff	Sample light on and sample light off
$V_{particle}$	Volume of particle

N	Number of particles
$m_{solution}$	Mass of solution
n_d	Number density
$m_{particles}$	Total mass of particles added to the solution
ρ_{PS}	Density of polystyrene
m_1	Mass of one particle
L_i	Length
W_i	Width
V_1	Characteristic volume of one particle
V_{1i}	Characteristic volume of each particle
A	Absorption
l	Thickness
I	Intensity of light passing through the sample
I_0	Intensity of the incident light
c_i	Concentration of chemical species
R	Diffused reflected light
σ_{ext}	Extinction cross section
σ_a	Absorption cross section
σ_S	Scattering cross section
Δt	Length of backscattering time
v_s	Scanning speed

Chapter 1 - Overview

1.1. Introduction

With the pharmaceutical industry moving away from traditional batch manufacturing processes to continuous drug manufacturing, there is an increased demand for developing and implementing tools for process understanding, monitoring and control. The objective is to allow a continuous and fast response of process to factor changes to achieve a reduction of the quality overhead costs, increase the production quality, lower production costs and to obtain self-adjusting production processes [2]. The systems based on timely measurements developed for analysing, designing, improving and controlling processes, have been defined as Process Analytical Technologies (PAT).

The European Medicines Agency (EMA) set a PAT team in 2003 to support applications of PAT in the European Union to enhance process and product understanding by quality-by-design. The aim was to identify and manage critical sources of variability to ensure an effective and efficient manufacturing process to accomplish a 'right first time' performance. This would enable a step change from a compliance manufacturing paradigm. In 2004, the United States Food and Drug Administration (FDA) introduced and encouraged the use of PAT driven by the need of a feedback/feedforward process-control strategies at real-time [3]. PAT was described in the FDA guideline document as "*A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*". The in-line monitoring of drugs using PAT would promote the identification of critical-to-quality attributes and their relationship to the quality of the product, the design of a robust process to control the critical-to-quality attributes, and a systems approach to use and correlate all significant process information [4].

The idea is that the more information gathered from a process, the better the causes of process deviations can be identified and the better strategies can be applied to ensure they do not occur, ultimately leading to reduced downtimes and improved process control thereby ensuring product quality. Application of the right PAT tools and analysis methodology can assure the desired product quality to be consistently manufactured and that any non-conforming material detected is eliminated by an appropriated investigation process.

The ability to monitor chemical attributes in chemical process reactions in real-time has been the key motivation to drive the development of PAT. As many processes and reactions are multiphase and contain particulate substance, such as particles and droplets, physical attributes of these particulates are sometimes overlooked by most of the PAT tools. For example, particle size and shape are critical physical attributes for pharmaceutical ingredients as they have direct impact on downstream processing (e.g. isolation, formulation) as well on the performance of the finished product (e.g., uniformity, dissolution, bioavailability, stability) [5-8]. Controlling and monitoring these properties in real-time during a continuous crystallisation process remains a rather challenging area of research. It impacts the active pharmaceutical ingredient (API), semi-solids (suspensions), sterile liquid products (injectable) and properties such as the homogeneity, flow and processing of solid oral drug products (tablets and capsules). Various optical-based measurements (such as imaging, backscattering, laser diffraction) or ultrasonic sensing technologies, although capable of providing an indication on these key attributes, show limitation on real-time quantitative information due to the speed of data processing and analysis, as well as an ill-posed nature of the inverse problems involved.

An innovative spatially and angularly-resolved diffuse reflectance measurement (SAR-DRM) system was developed for in-line monitoring in a variety of chemical manufacturing applications [9]. The system collects multi-wavelength (UV-visible-Near infrared) diffuse reflectance spectra from optical fibres of multi-angle multi-space arrangements, making use of the optical properties, namely scattering and absorption coefficients that are dependent on wavelength to assess chemical composition and particle attributes information in the suspensions. In this thesis, SAR-

DRM is for the first time applied to monitor particles in the micrometre size, using a polystyrene system, and a crystal system by using α-lactose monohydrate crystal suspensions. This size range is particularly important for the process monitoring and control of API used for the production of oral dosage forms, such as tablets and capsules. Assessing it in real-time can determine whether the desired size range was achieved or whether size reduction is needed.

The SAR-DRM technology can be complementary to other PAT tools measuring particles, such as Focus Beam Reflectance Measurement (FBRM) and Particle Vision and Measurement (PVM), both commonly applied to monitor crystallisation processes. These particle sizing technologies measure the optical response of single particle (single scattering) at a time using a light source of a single wavelength light source. Currently, these are used to provide an indication of the particle size and shape rather than a quantitative measure and do not measure the solid concentration in a suspension. Besides, the increase of solid concentration leads to complexity in the signal acquired or images resolution. In a highly turbid medium, when photons encounter several particles, they undergo multiple scattering, changing the direction during each scattering event. This leads to nonlinear scattering effects and absorption competition. Techniques relying on single scattering may result in a deteriorated and misleading captured signal. SAR-DRM relies on particles multiple scattering, hence, the higher the turbidity of the medium the better the signal acquired. Moreover, SAR-DRM ability to capture information from multi-wavelengths at different points and angles results in more detailed information about the particles attributes and system behaviour and can provide a more comprehensive assessment.

Although SAR-DRM spectra can be inverted to obtain bulk optical properties [10], and subsequently to estimate particle size distribution (PSD) [11], the reported inversion methods were developed for nanometre particles estimation. Moreover, these methods involve complex mathematical operations and intensive computational calculations, not being currently suitable for real-time applications.

An alternative approach for modelling the particulate systems using SAR-DRM consists on the development of multivariate calibration models. In this thesis, spectral datasets collected from different spatial and angular configurations from multiple

wavelength in visible to near infrared (NIR) regions are investigated to develop multivariate calibration models to estimate particle size and concentration. The different fibre measurement configurations are combined using two types of data approaches to utilise the information for enhancing the model performance. Raw spectra as well as scatter corrected spectra are analysed to establish understanding of how the information is utilised.

Additionally, a fusion of FBRM data with SAR-DRM spectra is investigated to examine the utilisation of complementary information by the calibration model.

The overarching goal of the present thesis is to propose a fusion of multi-sensors data for in-situ monitoring of processes, and the use of a spatially and angularly resolved diffuse reflectance spectroscopy as a PAT tool. In order to fulfil this goal, the objectives of his study are:

- Critically analysing SAR-DRM, FBRM and PVM and their performance using a model system;
- Establish a reliable setup for optimum data acquisition and data processing from the different in-line tools;
- Develop multivariate regression models to obtain reliable and robust models for estimating particle size and concentration for individual techniques/single detectors;
- Determine whether multiple detector-source configurations in SAR-DRM can lead to a better performance of the calibration models developed;
- Implement a data fusion approach to combine different information, CLD, vis-NIR and NIR to assess the impact on the calibration and prediction models;
- Determine whether SAR-DRM can be applied to monitor crystal systems;

In addition, the understanding developed in this study also contributes to the improvement of the PVM and FBRM algorithm apps to convert the information captured in particle size and aspect ratio. The work is published in [12]. As this is outside the scope of this study, the thesis will focus on the research for the aforementioned main objectives.

1.2. Thesis Structure

This thesis is divided into eight chapters. The first chapter introduces the thesis aims and contributions, and explains the importance and the need of implementing PAT tools to monitor manufacturing processes, more specifically, in the pharmaceutical industry. Chapter 2 comprises the theoretical background on common off-line and in-line techniques used in the pharmaceutical industry for measuring/monitoring particles size and introduces the newly developed SAR-DRM technique. Chapter 3 describes the data pre-processing commonly used and the theoretical background required to understand the process of building quantitative multivariate calibration models.

Chapter 4 describes the methodology and materials studied, the equipment and experiments used to collect and process the data, and the methodology used to build and test calibration models. It also provides initial assessment of the experimental setup and measurement reliability.

In Chapter 5 and 6 the results are based on experiments using a polystyrene system. Chapter 5 focus on understanding the impact of particle size and concentration on multiple techniques. It reports and compares the performance from off-line and in-line techniques and assesses the SAR-DRM response in comparison to the existing technologies. In Chapter 6, multivariate regression models are built to estimate particle size and concentration. The models first evaluate the information contained in single techniques or from single measurement configurations. For SAR-DRM, spectra from different spatial and angular configurations are evaluated. The understanding developed are subsequently applied to examine and identify approaches for best utilizing the measurement from different techniques.

In Chapter 7, SAR-DRM is, for the first time, applied to study crystalline suspensions. The analysis model approaches investigated in Chapter 6 are implemented to predict particle size and concentration of α-lactose monohydrate crystals.

Chapter 8 presents the overall conclusions gathered throughout this thesis and suggestions for future work.

Chapter 2 – Introduction to Particle Sizing Techniques

This chapter reviews the theoretical background of various off-line and in-line particle measurement techniques, their measurement principles, the type of information collected and presented, their applicability as well as their advantages and drawbacks.

2.1. Introduction

Traditionally, particle size measurements have been conducted using off-line techniques such as sieving, laser diffraction and microscopy [13], being laser diffraction the most broadly technique used in the pharmaceutical industry [14]. Although the interpretation of the results from the off-line analysis is usually more straightforward, due to the nature of the underlying principles of the analyser, errors can also be introduced during sampling, sample preparation or transport. The representativeness of the sample, the difficulty of sampling, and the sampling frequency are the common concerns associated with off-line analysis. The consequences of delayed detection by the off-line analyser for problems occurring in the process can be costly, sometimes could even lead to the discard of the product.

Among many particle analysis techniques developed, only a few are capable of in situ/in-line analysis for real-time process control and optimisation. The in-line analysis methods offer fast, non-destructive and non-invasive measurements, presenting an attractive alternative to the off-line options. The in-line analysers normally are configured with reduced measurement flexibility and tighter control of the measurement parameters. These constraints are often set based on the optimal measurement conditions found during the instrument development stage and aim to reduce the complexity in interpreting the results by the end-user of the technology. For complex processes where particulate substances evolve significantly with the process and/or vary with process conditions, the selection of suitable in-line particle analysers for the entire process is not always straightforward. The key benefit of obtaining real-time information from in-line particle analyser is to enable real-time feedback/feedforward process-control, a major contribution to optimal process operation and process/product consistency. The most common techniques for in-line particle measurements include particle chord length measurement, microscopic imaging, and spectroscopy.

Depending on the measurement principles, each method presents its own challenges in data interpretation. Section 2.2 will review the common in-line and off-line particle sizing techniques, their theoretical and measurement principles, and key advantages and constraints.

2.2. Standard Particle Sizing and Measurement Techniques

2.2.1. Sieve analysis

One of the simplest and most widely used sizing analysis techniques is sieve analysis. This easy and cheap yet precise and reproducible performance in separating a broad particle size range in relative short time make it an attractive reference method. Each sieve is made of a wire mesh with specific aperture. The smallest lateral dimension of the particle will pass through a given sieve opening. Multiple sieve plates can be stacked on top of each other, from the sieve with the smallest aperture to the sieve with the large. However, sieving is not suitable for all materials. Sprays or emulsions, dry powders with sizes below 38 μm (electrostatic charges may cause loss of material) and cohesive or agglomerated materials are very difficult to measure [15]. Other disadvantages are associated with attrition during the sieving process, destruction of a granularity of tough and fragile particles can be caused, reducing their size dramatically due to abrasion and collisions. Measurement times (particles need sufficient time to orient themselves to fall through the sieve, aided by vibration, yet the increase of the shaking time increases the risk of particle erosion) and operating methods (amplitude, frequency of the shaking) need to be optimized since they can lead to inaccurate results that do not reflect the real size distribution of particles in the measured sample. Sieve analysis assumes that all particles will be round or nearly round whereby when measuring the particle size of needle-like or rod-like particles peculiar results are generated. Sieve analysis requires a relatively large amount of sample and therefore is not adequate for use pricey materials and must be carefully cleaned as vigorous brushing may distort sieve openings [16].

2.2.2. Microscope-based imaging methods & analysis (Off-line)

Microscopy is perhaps one of the most accurate methods to determine particle size and shape. In the field of particles characterisation through imaging analysis, researchers typically rely on optical and electron microscopes. Both principles are described in more detail in this section.

2.2.2.1. Optical microscopy

In the optical microscopes, a visible light is transmitted through or reflected from the sample through a single or multiple lenses to produce magnified visual or photographic images of objects that are too small to be seen with the naked eye. Although there are optical microscopes with different levels of complexity, the underlying principle is the same. The 2D image of the particles is captured by normal light-sensitive cameras, usually complementary metal-oxide-semiconductor and charge-coupled device (CCD) cameras. The latter ones are used in digital microscopes to show the image directly on a computer screen without the need for eyepieces.

For the purpose of this thesis, an optical microscopy analysis was chosen. Among the different equipment available, an advanced optical microscope, Morphologi G3 [17]was chosen for off-line analysis of particle suspensions due to its ability to automatically scan through samples, providing information of the particle size and shape distributions of the samples in study. As it is shown in Figure 2.1, a 3D particle is captured in a 2D image. The area of the projected particle is then converted as an equivalent circular area to 2D image [18]. The diameter of this circular area is then expressed as the Equivalent Circle Diameter (ECD), see Equation 2.1, which describes the diameter of a circle with the same area as the 2D image of the particle captured, regardless its shape.

$$ECD = \sqrt{\frac{4Area}{\pi}}$$ **Equation 2.1**

The volume of a sphere with the same ECD, is calculated as:

$$\text{Volume} = \frac{\pi \times \text{ECD}^3}{6}$$ **Equation 2.2**

Morphologi G3 provides a simple and objective measure to a particle automatically. This high quality information could not be provided with a conventional microscope.

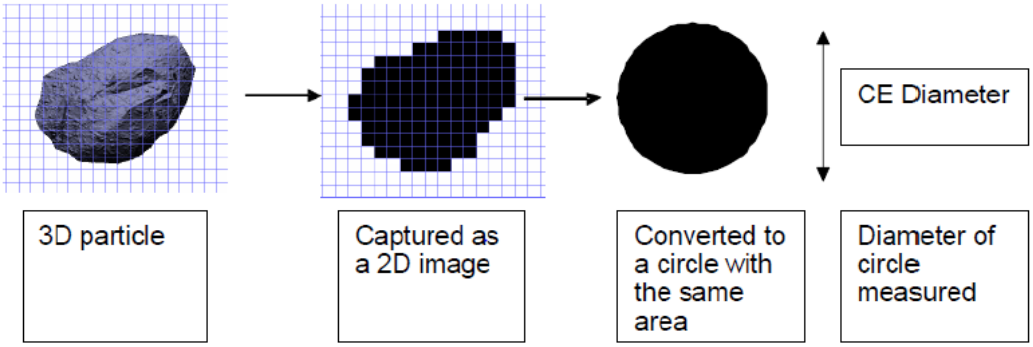


Figure 2.1: Determination of Equivalent Circle Diameter [18].

The main disadvantage of this method is that particles of very different shape could be characterized as identical, as shown in Figure 2.2, where all the particles present an ECD of 2.257 units. In process monitoring, the shape of particles often plays an important role, impacting the flowability, abrasive efficiency, bio-availability. Hence, the shape of the particles needs to be characterised, either by analysing the images (qualitative analysis) or by analysing the aspect ratio distribution provided by the software. The aspect ratio values are between 0 and 1 and consists on ratio of width to length of each 2D particle.

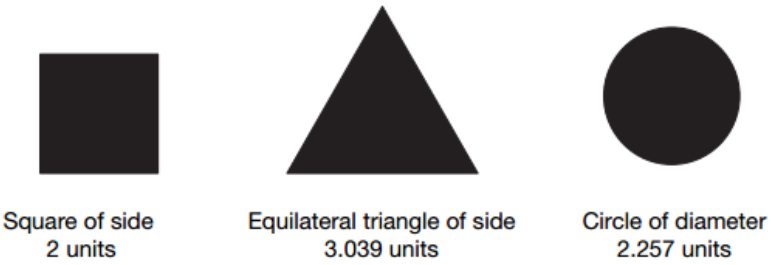


Figure 2.2: Three different particle shapes of identical ECD [18].

Other disadvantages are related to the fact the instrument only is capable for off-line measurement. Human error can be introduced since it involves sampling or sample preparation and transport. No instantaneous feedback is get online in order to control and optimize the process immediately. A relatively long time is also required to complete scanning of the sample, which is time-consuming and costly although the selection of a suitable scanning area can help to reduce it [19].

2.2.2.2. Electron microscopy

The Scanning Electron Microscope (SEM) uses a high energy beam of primary electrons to focus on the sample and generate surface signals [20]. The interaction of electrons with the sample results in the emission various types of radiation and electrons, such as: backscattered electrons (BSE), high-energy electrons that are ejected by an elastic collision of an incident electron; secondary electrons (SE), which result from inelastic collisions between the electrons present in electron clouds of sample atoms; diffracted backscattered electrons (EBSD), used to determine crystal structures and orientations; photons (X-rays), used for elemental analysis; and visible light (cathodoluminescence). The SE are the responsible to produce the SEM images and the BSE to obtain the topographic contrast, being the most valuable for showing morphology and topography of the samples [21] and to extract the particle distribution profiles.

In the case of Transmission Electron Microscopy (TEM), the image of the analysed structure is formed using the electrons that penetrated through the sample, providing information of the atomic structure and defects in the material, as well as particle size and shape.

At the moment there are also devices using a series of 2D image captured to reconstruct the 3D volume of electron microscopy data, e.g., Serial Block Face Scanning Electron microscopy (SBFSEM), however they have been mainly used for biologic applications [22, 23].

While optical microscopy is limited to visible light, using a beam wavelength of 400-700 nm, SEM and TEM offer energies up to a thousand times greater than visible light, offering a greater depth of focus and resolving power. However, the information provided is more visual and descriptive. The quantification of particle size distribution can be misleading as only few particles per sample are usually analysed. Moreover, these analyses are more expensive, more challenging to use and more time consuming.

2.2.3. Microscope-based imaging methods & analysis (In-line)

In in-line measurements, particles are in constant movement and rotation relatively to the field of view. This contrasts with off-line microscope-based imaging measurements as particles are static during the image acquisition and can be placed close or at the focal point the instrument. The lens to object distance can be adjusted and do not represent an issue. In in-line measurements, even when adjusting the focus for optimal measurements, capturing particles in focus is challenging, often, only a small fraction can be considered to be in focus. The different particle projections can lead to an increased number of particles out-of-focus and consequently, to bias results of particle size and shape measurement. It is crucial to exclude these out-of-focus particles from the analysis in order to correctly characterise particle size distribution of the samples, avoiding the under or overestimation of the respective size.

In the recent years many in-line microscopes-based imaging having been developed and optimised to monitor a variety of processes. Two types of microscopes are presented in this section: Machine Vision Based on 2D Imaging and Photometric Stereo Imaging (3D imaging).

2.2.3.1. Machine Vision Based on 2D Imaging

The Machine Vision Based on 2D imaging technologies usually use a white light source and the images are acquired using a monochrome camera. These technologies utilise the reflected light from the particles to capture a 2D grey image. Few tools such as Particle Vision and Measurement (PVM), Parsum IPP 70S using Spatial Filtering Velocimetry (SFV) and Pixact Crystallization Monitoring (PCM), have been tailored to provide solution for in-line process monitoring. These probe-based video in-situ imaging systems provide high resolution images of particle size and morphology in a stirred medium. Parsum probe can be also utilised to measure granules.

PVM has been widely used for in-line monitoring of crystallisation processes, to indicate the particles shape and size during nucleation, crystal grow and dissolution and to assess the extent of agglomeration [24-26]. The PVM is a simple in-process

optical video microscope which collects the backscattered light from the samples. It consists of six to eight lasers arranged at angles different angles around the objective tubus, to illuminate the suspensions. The lens focuses the light through a sapphire window on an adjustable focal plane. The camera is able to capture 10 images per second.

Although the obtained 2D projection only gives a qualitative analysis, quantitative information in form of distribution of characteristics sizes and aspect ratio can be extracted by applying an image processing algorithm. For example, Cardona et al., reported the development of an imaging processing algorithm coupled with filtering and screening functions to examine the images acquired and eliminate the incomplete or out-of-focus particle images [12]. This imaging analysis algorithm has been applied to suspensions of polystyrene and elongated silicon particles [12], and to monitor a wet milling process [6]. The output was compared to the manufacture mean size and to offline optical imaging analysis, respectively. The polystyrene suspensions showed a good agreement for particle size ranges from 150 to 500 μm, and consistent among the solid loadings 1-5 wt.% [12]. In the milling experiment, PVM showed to be able to quantitative extract particle size and shape, although some inconsistencies were observed when compared to Morphologi G3.

According to Cardona et al., a minimum number of 500 particles are necessary to be statistically representative [12]. The procedure followed by the image processing and analysis algorithm is explained in Figure 2.3. The greyscale PVM images are passed through a median filter to remove speck noise from the image background to smooth the filtered image. A Laplacian filter is then applied to increase the resolution of the particle edges, which shows where the pixel intensity varies. After a threshold is set to separate the particles from the background, the particles/objects are closed and filled using a disk of the same specific size as the object in analysis. Then, particles on the boundary of the image frame or too small to be adequately resolved are removed. Lastly, the focus of the particles that passed through the imaging process is evaluated and each individual particle is characterised in terms of area, perimeter, centroid, eccentricity, convexity. The particle size is achieved by fitting each particle in an

ellipse and then in a bounding box to determine the particle length and width. The results are then converted to ECD, similarly to Morphologi G3.

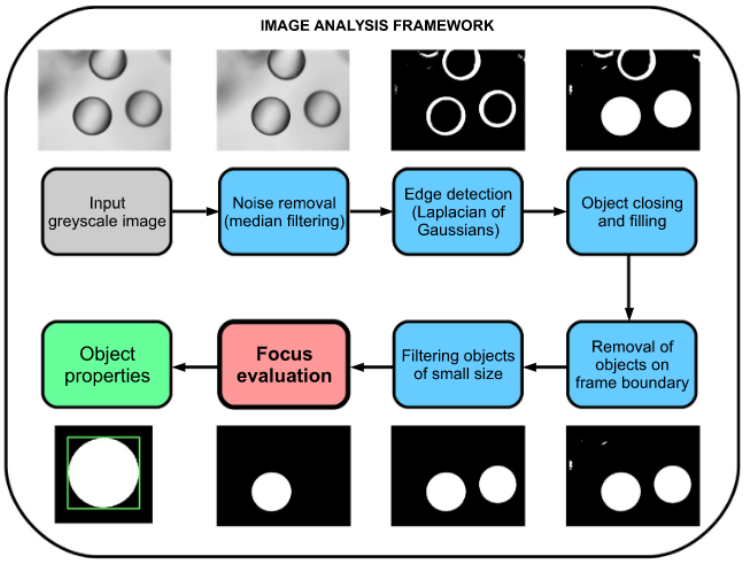


Figure 2.3: Diagram of the key steps of the image processing method used in our previous study [12].

PVM is easy to use, it is a low maintenance probe and the images are quickly acquired. The probe can be used to monitor solid suspensions under harsh conditions with respect to temperature and pressure during manufacturing processes. However, it may lose resolution of the images when dealing with highly turbid mediums, which will consequently affect the results from the algorithm. Inconsistencies with other techniques can also be related to the dispersion of particle agglomerates when in a suspension (while in off-line optical microscopes particles are static and the agglomerates kept), camera resolution limits to detect particles of a certain minimum sizes, the impact of movement/orientation during the process measurements [6].

2.2.3.2. Photometric Stereo Imaging

Photometric stereo imaging allows the reconstruction of three-dimensional images of objects by varying the direction of the incident light sources, keeping the viewing

direction constant. The aim is to generate images with a decent contrast to reliably discern defects, symbols, objects or other features of interest from the background. The technique can be advantageous to investigate samples that are moist or cohesive, when dispersion of particles is challenging or impossible.

An example of a particular equipment of interest using this principle is the Eyecon technology. The Eyecon technology has been developed not only for visual analysis of shape and surface morphology of particles, but also for quantitative purposes. The technique allows the in-line quantification of particle size and shape distribution and it has been investigated as a PAT to monitor different processes in the pharmaceutical industry. The Eyecon particle sizing technology has been applied to monitor hot-melt extruded pellets [27], a placebo of Avicel and Lactose in fluid bed granulations [28] and a twin screw granulation [29]. The technology is based on photometric stereo imaging, illuminating the sample with three different Light-emitting Diode (LED) colours, green, blue and red from different angles, to capture the 3D features of the particles in addition to a regular 2D image. The lights are arranged circularly around the lens as observed in Figure 2.4(a). It uses a flash imaging technique to capture moving particles. The camera is able to capture a continuous image sequence every 600 µs, using powerful short pulses of illumination (length between 1 and 5 µs) to avoid motion blurring of particles and capture a sharp image. It enables to capture images of moving particles up to 10m/s.

The colour distribution on the particle surface is captured in an image and for each individual pixel, a map of the surface height is captured. The colour change allows the distinction of particle edges, overlapping or touching particles, avoiding two particles being detected as one. The particles partially cut by the field of view are excluded from the analysis algorithm. The 3D-projected image built for each particle fully captured, as shown in Figure 2.4(b), is fitted with an ellipse, with the same major and minor axis information as the particle in analysis. The particle size is then expressed as ECD and the distribution of sizes is built by combining all the measured particles. The results are presented as the mean of the particles diameter, such as, D10, D50 and D90 from the volume-based distribution. The mean aspect ratio is also provided. The particle size range evaluated can vary from 50 to 3000 µm.

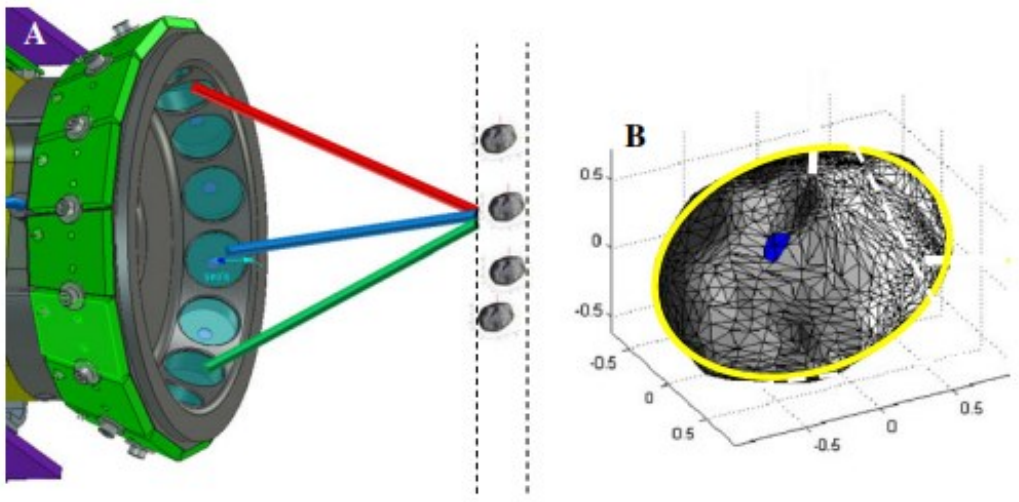


Figure 2.4: a) Working principle of the Eyecon equipment and **b)** particle size determination from a 3-D projected image [29].

A study conducted by Treffer *et al*. has shown, in general, a good agreement of the D50 and D90 values obtained from Eyecon with the respective results from an off-line image analysis tool (QICPIC), with a small deviations ((deviations <5%) which could be related with the non-spherical shape of the pellets [27]. Similar PSD curves were also obtained. Although the results were considered accurate, the in-line analysis measured only 0.3–0.6% of the total number of pellets to calculate the PSD.

Eyecon is not in direct contact with the sample, hence, there is no risk of contamination, and provides directly the sample measurements. However, it is mainly useful for the analysis of powders and granules. As the technique is based on diffuse reflected light, black, strongly reflecting or transparent particles cannot be correctly detected. Other challenges arise from the difficulty to properly separate individual particles for the size analysis, which can lead to erroneous measurements.

2.2.4. Light scattering based method

2.2.4.1. Laser Diffraction (LD) measurement (Off-line)

Laser diffraction measurement has become a standard method used in many industries for particle characterisation and quality control. The relatively quick, reliable and high reproducible measurement operates in a range of 0.1 – 3000 μm, which makes it an attractive technique.

The method relies on the fact that diffraction angle is inversely proportional to particle size [15]. The laser beam that irradiates the dispersed sample and the scattered light from the samples is then collected by a multi-element detector. The angular variation in intensity of light scattered is largely particle size dependent: large particles scatter light at small angles relative to the direction of the laser beam and small particles scatter light at large angles. Therefore, particle size information can be calculated from the collected scattering pattern. Older instruments mainly rely on the Fraunhofer approximation to derive particle size information from the scattering pattern. The approximation assumes that the particle is much larger than the wavelength of light employed with a refractive index significantly different from the surrounding medium, and most of the light is scattered in the forward direction at small angles relative to the propagation direction of the incident beam [30]. It also assumes that all sizes of particle scatter with equal efficiencies and particles are opaque. These assumptions are not appropriate for many materials which present a relative small contrast in the refractive index or the particles are small, below 50 μm. The main advantage is that there is no need to know the exact refractive index of the particles and the medium. Nowadays, LD method is based on Mie theory. Mie theory makes use of the complex refractive index (including the imaginary part, which corresponds the reduction of intensity of a light due to absorption) of both the sample and the medium, to provide appropriate size analysis. In this case, it is assumed that all particles are spherical and homogenous in the suspension, and the dispersion is dilute, so that light is scattered once. It requires to know the refractive index and the particles can be either opaque or transparent. However, satisfying all requirements is rarely possible for most applications and is difficult to achieve when applying in-line for samples with high particle content.

Despite the constraint and challenge to achieve an ideal LD measurement and analysis, this is a well established technique of industrial standard by ISO13320:2009 [31]. LD results are generally presented as a volume-weighted particle size distribution.

2.2.4.2. Chord length measurement (In-line)

In-line devices based on light scattering have been developed through the years for measuring particles attributes while they are moving, however they measure particle chord length (CL) rather than actual particle size.

Spatial Filtering Velocimetry (SFV) [32, 33] and Focused Beam Reflectance Measurements (FBRM) [34] are example of devices that measure particle chord length (CL), which is the distance between the two edges of a particle. The measurement is straightforward; it collects signals reflected by particles passing in front of a laser illumination beam and coverts the duration of the reflected signal into CL information. However, while SFV calculates the CL from the shadow cast by particles that cross the laser beam, FBRM calculates CL from the reflected light from the particles back to the probe. In SFV, particles pass through a spatial filter in front of a receiver and a shadow image of the moving particle is collected. It is assumed the shadow is the same size as the particle.

FBRM has been extensively investigated to monitor changes in the CLD related to crystals morphology, to establish a relationship between the chord length measured and the corresponding particle size and to qualitatively detect nucleation, growth and dissolution of crystals [24, 25, 35-37]. Hence, FBRM was the selected device to monitor particulate suspensions of different particle size ranges in this study and a more detailed description of its mechanism, measurement and analysis principle is described here.

FBRM uses a circularly moving laser beam focused near the sapphire window to measure particles. A schematic of the FBRM optical design is shown in Figure 2.5.

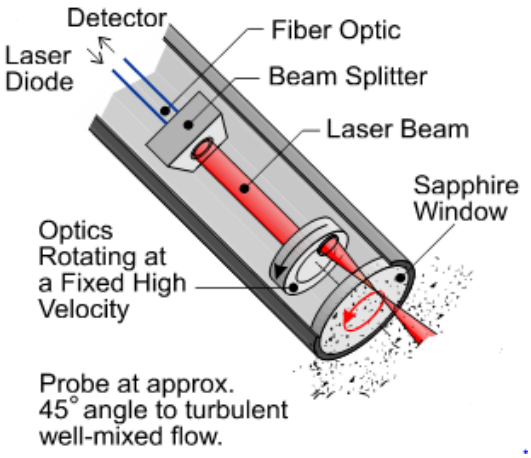


Figure 2.5: Schematic illustration of measurement mechanisms of FBRM [38].

The focused beam intersects at one edge of a particle, resulting in a backscatter light to be detected by FBRM probe until the focused beam reaches the opposite edge of the particle [39]. The beam moves at high velocity (2 to 6 m/s), therefore, it is assumed the laser beam scans across the particles near the window on a straight line. The scattered light collected by FBRM optics is converted into an electronic signal. Chord length (CL) is defined as the duration of each reflection multiplied by the velocity of the scanning beam:

$$Chord\ length = \Delta t\ v_s \qquad \textbf{Equation 2.3}$$

where Δt is the length of back scattering time (s) and v_s is the scanning speed of the laser (m/s) as shown in Figure 2.6. The results are presented in the form of chord length distribution (CLD), which is the statistical representation of the CL from a population of particles.

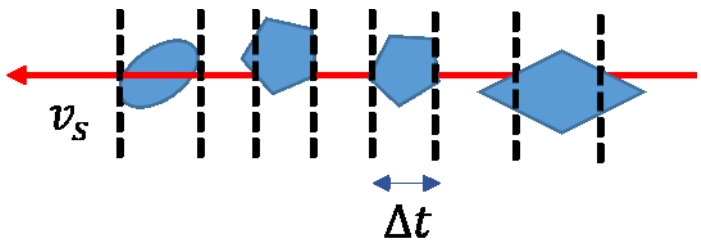


Figure 2.6: Definition of CL for particles of different shapes. Adapted from [1].

As CL measurement does not assume particle shape, the technique became well suited for many in-line application where particles are non-spherical. Moreover, FBRM can be easily operated and it requires little maintenance or calibration, making it an attractive tool for in-line monitoring of a variety of processes. However, the FBRM measurements are not easy to interpret since results are presented as CLD rather than PSD.

The relationship between CLD and PSD is given by a Fredholm integral equation (first kind) [40],

$$Cq_{CLD,0}(s) = \int_0^\infty q_{P,0}(s,L)q_{PSD,1}(L)dL \qquad \textbf{Equation 2.4}$$

where the measured probability density function of a CLD (s), $q_{CLD,0}(s)$, is proportional to the length-weighted PSD, $q_{PSD,1}(L)$. The size and shape information is expressed using the kernel density function, $q_{P,0}(s,L)$, which defines the CLD of a single particle of a characteristic size of L.

The calculation of $q_{CLD,0}(s)$ from the corresponding PSD and kernel is straightforward and it has been described by LT and Wilkinson, and Agimelen *et al.* for spherical and non-spherical particles [41, 42].

Different methods to transform the measured CLD into a PSD have been developed [41, 43, 44]. However, to invert PSD from a CLD is challenging as there are potentially multiple solutions in terms of particle size distributions and shape which give essentially the same CLD within the accuracy of experimental data [42]. Another limitation is that the developed approaches require an assumption about a fixed particle shape. In most of the cases particles are assumed to be spherical, however, providing the right geometry of the crystals plays a critical role for achieving more accurate PSD [45]. It adds extra difficulty to the ill-posed inversion problem, hence the technique is usually applied alongside with PVM which offers critical information to assess particle shape and give an indication of particle size and morphology [42, 46].

2.2.5. Spectroscopy methods

Spectroscopy is the most prominent method for in-line process analytics. When an electromagnetic wave of a given intensity propagates through a particulate system, the amplitude of the wave decreases due to the absorption and scattering of the particle [47]. Two different frequency/wavelength ranges, acoustic and optical, have been explored for particle size analysis. Acoustic and optical spectroscopy analyse scattering and/or absorption effects to relate the changes of magnitude in certain frequency/wavelength to the properties of the system, such as size, concentration, composition and structure.

2.2.5.1. Acoustic spectroscopy

Acoustic spectroscopy is one of the emerging technologies to characterise concentrated solutions. The technique utilises sound/ultrasonic waves of various frequencies to interact with the dispersant and disperse phase for assessing particle size and concentration of particles in a suspension. The sound waves propagate with certain velocity, intensity, and its amplitude decreases when passing through the particulate system, mainly due to absorption and scattering effects [47]. This consequence is commonly called attenuation. The total attenuation is related with the system properties, such as size, concentration, composition and structure. Therefore, particle size and concentration of particles in a suspension can be assessed by exploiting the sound velocity and attenuation. Many theories have been developed in order to establish the relationship between the materials acoustic properties and their physical and chemical properties [48-51]. Although ultrasonic devices have been mainly applied to investigate and characterize nanoparticles, it is suitable for particle size ranges from 10 nm to 3mm and the measurements can be performed in concentrations from 0.1 by volume up to 50 % or more by volume, of the material dispersed in liquid [52]. Some of the acoustic devices have been developed for in-line and on-line particle size analysis [53, 54]. A major advantage of the technique is the ability to study highly concentrated and optically opaque systems. In addition, it allows fast, non-destructive and non-invasive measurements. However, prior to any acoustic measurement,

information about the materials and dispersing phase characteristics, such as density, viscosity and heat capacity must be collected to serve as an input to set the correct parameters for a reliable and more accurate analysis [55]. The success of the measurement technique is highly dependent on the quality of the analysis algorithm. Moreover, there is the need to design, develop and calibrate instruments according to the target information, operating conditions, materials characteristics and know which theories to use to interpret measurements.

2.2.5.2. Optical spectroscopy

In optical spectroscopy, it is measured the interaction of the light with a variety of materials. When light is emitted to a system, it can be absorbed, transmitted, reflected or scattered (Figure 2.7).

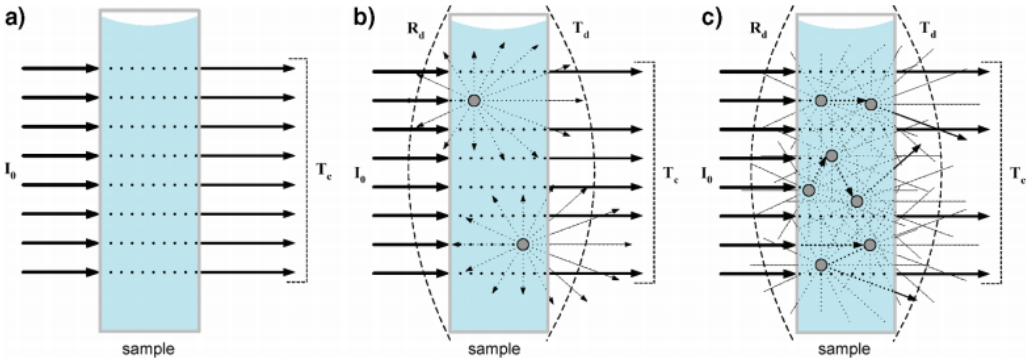


Figure 2.7: a) Homogeneous sample (e.g., liquid mixture), absorption only. **b)** Turbid sample with very low concentration of scatterers (particles): absorption and single scattering. **c)** Turbid sample with a high concentration of scatterers: absorption and multiple scattering [56].

Depending on the wavelength chosen, from 100nm to 1mm, as illustrated in Figure 2.8, different modes of interaction can be observed and therefore, different information can be obtained from the system in analysis. PAT tools such as attenuated total reflectance ultra violet/visible spectroscopy (ATR-UV/vis), attenuated total reflectance Fourier transform infra-red spectroscopy (ATR-FTIR), near infrared (NIR) spectroscopy, Raman are the most commonly used to monitor batch crystallizers to

obtain information about, for example, solution concentration and polymorph or pseudopolymorph conversion [57-60].

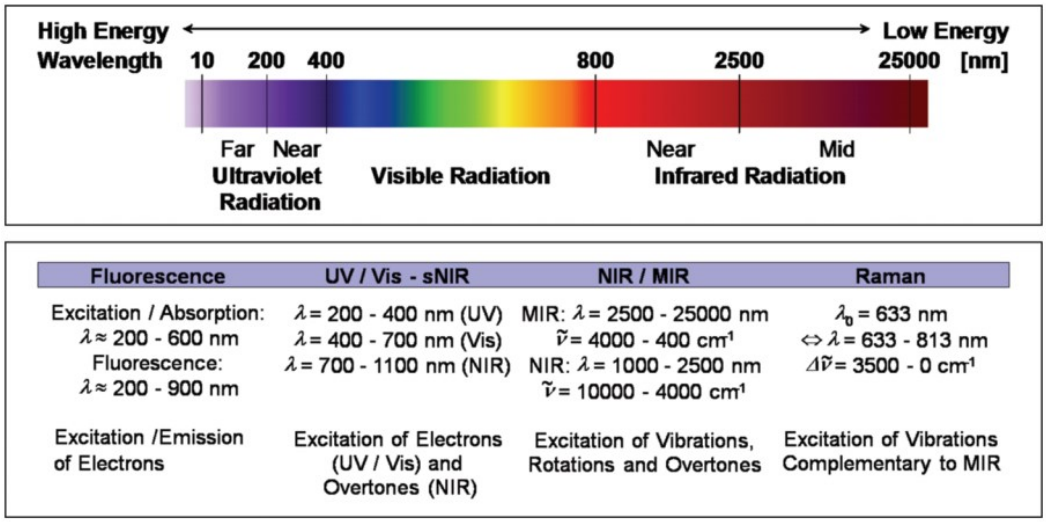


Figure 2.8: Electromagnetic spectrum and range of useful wavelength for spectroscopic analysis [2].

The Ultraviolet and visible (UV-vis) wavelength range spans from 200 to 700 nm and it is routinely used in analytical chemistry to determine the concentration of the chemical species by applying Lambert-Beer's law. The optical spectroscopies based on this principle are typically used to obtain chemical information of the sample. Classical absorption spectroscopy is based on Lambert-Beer's law which relates the concentration of the absorbing species to the sample absorbance at a given wavelength. The absorption bands are obtained from the electronic transition and are typically very strong. The electronic transition occurs when the energy of the absorbed light is equal to the energy that separates the electron from the ground state to the excited state. In these cases, the electron absorbs energy and jumps to a higher energy level. The energy of the absorbed photon is related to its frequency or wavelength and consequently, related to the structural groups in the molecule. As the absorption bands are usually strong, sample preparation is generally required for transmission or reflection measurements. If the sample is not a liquid, the powder needs to be dissolved in order to measure its absorbance. A common PAT used to monitor crystallisation processes is ATR-UV/vis. In ATR-UV/vis measurements, the light penetrates a short distance

into the sample. The technique is not affected by particles scattering in the medium. The short light path length allows light to be absorbed in a way that the transmitted beam carries information on the spectral absorbance and consequently, the sample concentration.

The mid-infrared (MIR) spans from 2500 to 25 000 nm and its spectral features are related fundamental vibrations and rotations of the molecules. Differently from NIR, the molecular vibration mode is excited from ground-state (n=0) the first vibration excited state (n=1). It is generally translated in sharper, strong and well-defined peaks compared to NIR, which makes the spectra usually more straightforward to interpret. However, NIR radiation typically penetrates much further into a sample and perhaps more information can be obtained using data analysis.

With the increase of the wavelength, it is decreased the frequency and the energy of photons. The NIR wavelength region spans from 700 to 2500 nm. Despite being neighbour of UV-vis, the energy difference is enough to cause different interaction with the sample. The NIR is related to the combination and overtone modes of the fundamental vibrations of the molecule. The absorbed light is a result of the molecular vibration excitation from the ground state (n=0), to higher vibrational energy levels (n=2,3,4...) or the combinations of several vibrational states. Only vibrations resulting in a change of the dipole moment of the molecule can absorb NIR light, making NIR a suitable technique to detect groups in the molecule with high dipole moment, such as C-H, N-H, S-H or O-H bonds. The NIR techniques are usually relatively fast, non-destructive and non-invasive but the absorption bands are typically broad and overlapping, resulting in a baseline shift of the spectra. Multivariate data analyses have been applied to help the identification of the bands, as well as extracting quantitative information from it.

The sample's absorbance is obtained by measuring the light that passes through the sample, the transmitted light (collimated transmittance, Tc), resulting in a transmission spectrum. Hence, to measure the chemical concentration it is required a homogenous and non-scattering sample Figure 2.7(a). At sufficiently low particle concentration in the liquid, photons passing through the sample will encounter at most one particle that will scatter the light, single scattering, and absorb light (Figure 2.7(b)). In such cases,

the path length contribution of the light is extended equally for all wavelength, therefore the Lambert-Beer's law is still considered valid. Further increase of the number of particles in the system introduces complexity in the analysis. Increasing the turbidity, photons will encounter several particles and undergo multiple scattering before leaving the sample (Figure 2.7(c)). As a result, the total path length travelled by the photon before escaping the sample will be greater than the sample thickness and the Beer's law is no longer completely satisfied. The direction of the photons is changed by each scattering event and eventually becomes random. Some of the light will be returned to the front of the sample where the incident light illuminates and are called diffuse reflectance (Rd). The pure absorption spectrum from chemical species is affected by non-linear spectral variations which causes the broadening of the absorption peak and a shift of the baseline. In reflectance systems, the Lambert-Beer's law needs to be redefined as:

$$A(\lambda) = -\ln(R) = \ln\left(\frac{1}{R}\right) = -\ln\left(\frac{I(\lambda)}{I_0(\lambda)}\right) = l\sum_{i=1}^{n}\sigma_{ext,i}(\lambda)c_i \qquad \textbf{Equation 2.5}$$

where A is the absorbance, R is the diffused reflected light detected, l is the thickness, I_0 is the intensity of the incident light, I is the intensity of light passing through the sample, n is the number of chemical species, c_i is its concentration of the chemical specie i, and λ is the wavelength of light. $\sigma_{ext,i} = \sigma_{a,i} + \sigma_{S,i}$, is the extinction cross section, $\sigma_{a,i}$ is the absorption cross section and $\sigma_{S,i}$ is the scattering cross section. The $\sigma_{ext,i}$ is nonzero for those species which are particles. As the scattering cross section is a highly nonlinear function of particle size and shape, even though Beer's law applies, the situation is complicated by the presence of nonlinear scattering effects. For the same concentration of the scattering species, two different particle sizes would lead to changes in the absorbance.

In high turbid mediums, not all light is absorbed or transmitted and the scattering events are dominant. In these cases, visible and NIR are regions dominated by particle scattering as the light penetrates deeper into the sample. The scattering events are caused by differences in the refractive index due to the particle size, shape, concentration (number of particles in the suspension) in analysis. These characteristics

make the visible and NIR ranges suitable for particle size analysis and solid concentration. The impact of particle attributes normally manifests a shift and offset of the spectra baseline.

Several authors have proposed to decouple the absorption and scattering effects in order to extract particle size and particle size distribution from metallic and semiconductor nanoparticles [61], latex emulsions [62], powders or granular powders [63], using different technologies and mathematical approaches. In case of suspensions of high solid content, either diffuse reflectance (Rd) or diffuse transmittance (Td) measurements are taken. As more light is reflected back from the samples than transmitted for samples with mediums with a high solid content, diffuse reflectance measurements (DRM) are usually preferable. The DRM is developed for in-situ and online applications, and usually exploit the use of optical probes configured to emit UV-vis-NIR light from one or more sources (with the same incident light angle), and to collect the reflected light from several detection fibres.

Recent studies on biomedical diagnosis on cancer [64, 65], diabetics [66] and biological tissues [67, 68] introduce the use of a spatially resolved diffuse reflectance measurement (SR-DRM) for determination of the optical scattering and absorption coefficients. In SR-DRM, the detecting fibres are set at specific distances from the light source, and the reflectance spectra collected from each one of them corresponds to a different traveling path of the light. Usually, a decrease in the intensity of the reflected light is observed as a function of increasing the source-to-detector distances. As the source-to-detector distance increases, the photons will travel a longer path and undergo more scattering and absorption events, leading to weaker reflectance signal. The diffuse reflectance spectra collected from the samples encloses information of both bulk absorption and bulk scattering properties. Those optical properties are dependent on chemical composition, concentration and particle size, and can be determined by applying light propagation theory to invert the spectra [69]. However, as the inversion involves intensive computational calculations, multivariate regression models can be used as an alternative to develop calibration models to acquire real-time information of particle size and concentration.

Recently, a Spatially and Angularly-Resolved Diffuse Reflectance Measurement (SAR-DRM) system was developed for in-line monitoring in a variety of chemical manufacturing applications [9]. In addition to the spatially resolved fibre arrangement, an angular feature was introduced. The SAR-DRM collects multi-wavelength (UV-visible-Near infrared) diffuse reflectance spectra from optical fibres of multi-angle and multi-space arrangements. According to Chen *et al*., SAR-DRM outperforms common reflectance measurement for highly turbid particle suspensions [9]. Depending on the angle used to irradiate the sample, light will travel a different path length inside the sample and the spectra will contain additional information of the overall system, which can be explored to improve the estimation of particle size and concentration.

In this study, SAR-DRM system will be, for the first time, applied to monitor micron size particles in particulate suspensions. Calibration models to estimate particle size and particle concentration of polystyrene in water and α-lactose monohydrate in acetone are built by applying partial least square regression analysis. Three different in-line technologies were used, PVM, FBRM and SAR-DRM. A sensor data fusion strategy based on data augmentation and co-adding is performed to assess whether there is an improvement of the model performance. The particle size used as reference was obtained from laser diffraction measurements and the particle concentration from gravimetric methodologies.

2.3. Representation of particle size distribution (PSD)

Due to the nature of the measurement principles used, each particle size instrument has a preferred mode of expressing the particle size measurement. Comparison of PSD from different measurement techniques is not straightforward. Typically, there are two aspects to consider; the expression of particle size, and the expression of PSD. The following section briefly introduces the different choices available, and what steps to take to allow direct comparison of PSDs from different measurement techniques.

2.3.1. Expression of size of non-spherical particles

The expression of size for spherical particles is straightforward. It can use either radius or diameter, and since this dimension is the same in 3D, the expression is always consistent. The most common expressions are equivalent spherical diameter (ESD) and equivalent circle diameter (ECD). Both describe the particle characteristics using a single parameter, the diameter. The difference between the ESD and ECD resides on the measured size-dependent property obtained by each technique. The ESD of a particle is defined as the diameter of a sphere with the same volume as the particle. Expressing particle size as ESD is common for techniques such as laser diffraction, which contemplate the 3D nature of the particles by obtaining ESD distributions from the diffraction pattern of a corresponding distribution of spheres. The ECD is defined as the diameter of a circle with the same area as the particle measured, and is commonly used in imaging-based analysis. The FBRM analysis algorithm also estimates particle size as ECD. This is due to the inversion of CLD into PSD considering the 2D projection of an ellipsoid with its major and minor axes parallel to the projection plane. The ellipsoid area is then converted into a circle with an equivalent diameter, a similar principle as used by the imaging techniques.

ESD and ECD measures are comparable as they both use a single length, the diameter, of a particle with an aspect ratio of 1. The disadvantage is that the characteristics associated with non-spherical particles can mislead their interpretation, resulting in invalid analysis of the measurement. Nevertheless, ESD and ECD remain an attractive means to infer particle size information in reasonable accuracy for many applications while enabling comparison of results obtained from the different equipment.

2.3.2. Expression of the particle size distribution

PSD is commonly used to express the overall size characteristics of a population of particles since particles often have a non-identical size or uneven shape. Depending on the main size range of interest and the technique used, PSDs can be expressed as either number-weighted or volume-weighted distributions. The number-weighted PSD

typically emphasizes small size as each particle has equal weight and is often used for microscopes and image-based analysers. The volume-weighted PSD emphasizes the presence of large particles due to their larger volume and is normally chosen for techniques such as laser diffraction. However, the choice and interpretation of the type of distribution are not always straightforward. In the pharmaceutical industry, for example, it has been reported that the preference is to present and discuss the results on a volume basis for most applications [70]. Furthermore, the size interval of a PSD, also known as bin size, is often unequal. This is designed to improve the efficiency of the analysis algorithm by reducing the total numbers of bins for calculation. However, each instrument has its own bin size setting, which makes a direct comparison of PSDs from different measurement techniques impossible without further treatment.

A more appropriate approach is to convert PSD into a probability density functions (PDF) to allow a direct comparison for all measurement results. The PDF stipulates the probability of a random variable falling within a particular range of values which can be calculated from the area under the PDF curve [71]. Using PDF makes the distribution grid-independent. In this representation, the only differentiating factor between various techniques is the resolution of the original grid. The smaller the bin size is, the more exact the discretised PDF will be. A PDF is calculated by dividing the fraction of a volume/number-weighted distribution (H_i/H_{total}) to the width of the bin, as shown in the following equation:

$$PDF(bin_i) = \frac{H_i}{H_{total}} \cdot \frac{1}{(bin_{i+1} - bin_i)} \qquad \textbf{Equation 2.6}$$

where i is the bin number and bin_i and bin_{i+1} are the left and right limits of bin i, respectively. H_i corresponds to the count or percentage of particles whose particle size falls within the ranges defined for a bin, and $H_{total} = \sum_i H_i$. As the population is normalised to the bin width, it eliminates the issue of comparing different PSD when different size intervals are used. In addition, the total population of a PDF is one. This brings the different PSD measurement into the same scale and ease the comparison between results from different techniques.

Few studies were reported to compare the performance and reliability of different particle analysis techniques [16, 72, 73]. As previously mention, each technique, relies on different physical principles and have different methods of converting the signal collected into PSD. Moreover, each equipment also has its own number and size (wide or narrow) of the bins set differently, which is related with the resolution of the instrument. A comparison of a performance matrix of in-line and off-line particle analysis techniques is essential for a reliable and consistent PSD analysis and can be valuable to guide users on the selection of appropriate particle analyser, pitfalls to avoid, and widen the new applications of the analyser.

Chapter 3 – Introduction to Empirical Scatter/Signal Correction Methods and Multivariate Analysis

This chapter covers some pre-conditioning methods used to solve signal related issues, such as mean centring, scaling and smoothing, and the empirical based models commonly used to reduce differences in sample thickness and eliminate the scatter effect in spectra obtained through diffuse spectroscopy, more specifically, the Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), Extended Multiplicative Scatter Correction (EMSC), Inverse Signal Correction (ISC) and the derivative of the signal. The chapter also attempts to give some elementary mathematical background required to understand the process of building quantitative multivariate calibration models for estimation of particle size and particle concentration using Partial Least Squares Regression (PLSR) analysis. The models developed need to be validated. Cross-validation, and test set validation are explored.

3.1. Introduction

On a daily basis, most of the product analysis and their respective findings are based on univariate analysis, on a single measurement with a single response. However, most of the problems are dependent on multiple variables, multiple parameters that have an impact on the final outcomes. The aim of multivariate analysis is to assess the relationship among those variables and extrapolate qualitative or quantitative information.

In pharmaceutical industry, there has been an increasing demand for high and consistent product quality by regulatory control bodies such as the Food and Drug Administration (FDA). This has been the driving force for the development and application of PAT, and quite often, multiple PAT can be used simultaneously. The implementation of PAT aims to improve and enhance product efficiency, safety and traceability, as well as to contribute to a growth in profit margins of the manufacturing industries. Spectroscopic sensors, for example, play an important role in numerous manufacturing processes since they offer a rapid, precise, non-destructive analysis and are flexible to be adopted in-line. Depending on the target information, techniques such as UV-vis, NIR, Raman and MIR/FTIR spectroscopy, can be used to predict concentration of the chemical species, quantify physical parameters, fingerprint raw materials and process streams. Research based on the development and implementation of these in-line and on-line analytical techniques often involves the combination of multivariate analysis (chemometrics) with the experimental data collected. Multivariate analysis incorporates mathematical and statistical methods required to look for relationships between different variables and to build-up information from the system [74, 75]. The method is used to resolve the multidimensionality of the data set, reduce the multicollinearity, extrapolate information even with missing data and reduce or eliminate the variation introduced by particles scatter effect, path length effect, experimental error and noise. The multivariate analysis can be divided into two types: exploratory and decisional. The exploratory analysis is often used to classify the samples into different groups and it helps to find outliers in the system. The decisional analysis is used to explain a set of variables, find quantitative relationships among them and to predict the output. For

each type of analysis, the most popular are the Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR), respectively. PLSR is the approach used in this work to estimate the particle size and concentration in suspensions of polystyrene and α-lactose monohydrate.

Most of the spectroscopic sensors are commonly used for chemical purposes rather than particle size. However, there is a widely known technique that has been explored for particle size analysis, the NIR spectroscopy. The on-line/in-line NIR spectroscopy is one of the most popular technologies applied for process understanding as its spectrum is affected not only by the chemical species but also by the sample physical properties and it is normally combined with chemometrics. NIR has been used to extract physical information related to the particle size from granules during wet granulation [76] and high shear wet granulation (wet massing phase) [77], wheat flour analysis for potential on-line/at-line quality control in a flour mill [78], commercial grade lactose mixed with microcrystalline cellulose and chlorpheniramine to enhance the control of tumble mixing of high and low dose formulations [79], blended powders of paracetamol (three excipients used: microcrystalline cellulose, talc and magnesium stearate) [80], dried granules during continuous granulation and drying [81] and polystyrene in suspension polymerization experiments [82]. UV-Vis spectroscopy coupled with multivariate analysis has also shown to be effective at classifying particles of different sizes from polymers [83] and silver nanoparticles [84]. Diffuse reflectance UV-vis-NIR spectroscopy has also been explored to estimate particle size and particle concentration in suspension polymerisation reactions, in micrometre size [85], and polystyrene emulsions in nanometre size [9]. Although Raman spectroscopy only gives information about vibrational and rotational spectra, few studies have been reported linking the effect of particle size distributions with the Raman spectra intensities obtained [86-89]. In general, the trend observed is an increase of the Raman intensity with a decrease of particle size. It has also been stated that the diffuse reflectance signal acquired depends on the optical system for excitation and collection of the Raman spectra (probe-to-sample distance) [88]. However, no quantification or prediction models for particle size estimation by Raman using multivariate methods was found in the literature. Nevertheless, a Raman coupled with optical imaging has shown to be a promising method for characterizing particle shape, size and distribution

of corticosteroid in aqueous nasal spray suspension, as a result of a validation study on polystyrene microsphere size standards [90].

In this work, we propose a multiple in-line sensor approach to assess whether it is possible to extract more information about particles attributes from PVM, FBRM and SAR-DRM. Calibration and prediction models to estimate particle size and particle concentration of polystyrene and α-lactose monohydrate are built using the multivariate regression analysis. The building of the quantitative models involved (1) spectra/CLD acquisition from different sample conditions, (2) collecting of the reference measurement, (3) selection of a representative dataset, (4) organisation of the dataset, (5) selection of pre-processing method, (6) building multivariate calibration model and (7) validation of the model by cross-validation and test set validation. The background of the last three steps are described in more detail in this section.

3.2. Data pre-processing and pre-conditioning methods

3.2.1. Pre-conditioning methods

3.2.1.1. Mean centring

The mean centring comprises the subtraction of the average data from the dataset. In this case, the values change, but not the scale, resulting only in a shift of the original plot. The averaged measurement is set to zero allowing the modelling to assess the deviation of each measurement from the mean of all measurements. The standard deviation, skewness and shape of the data are not affected. This pre-condition can enhance the interpretability of the data and reduce the multicollinearity, i.e., the high correlation among independent variables, when used to characterise individual regression coefficients. A high degree of multicollinearity can lead to bias and unreliable results as it increases the sensitiveness of the coefficients to small changes in the model, reduces the precision of their estimate, and weakens the significance tests power of the regression model, F-test. However, when the aim is to assess the overall fit of the model, i.e., R^2, without taking into consideration the variability across the

predictors and its main effects, or to assess the determinant of the covariance matrix, mean centring will not affect or improve multicollinearity and it may be bypassed [91].

3.2.1.2. Scaling

There are several ways to scale the data, however, the most common is the unit variance scaling. Mathematically, the variables from each block are divided by their standard deviation so the variables have unit variance. This technique is especially important when analysing datasets containing variables of different scales in order to give them equal weights so that all the variables have the same contribution to the model. However, in spectroscopy data, each variable contains important information (peaks), and the intensity variation between wavelengths should be captured, thus other type of pre-conditioning or pre-processing methods would be more appropriate than scaling.

3.2.1.3. Smoothing

Smoothing is typically applied to remove random noise from the spectra and consists of taking the average of adjacent data points (a window) to calculate the new value. Savitzky-Golay (SG) algorithm is the most method applied and it is based on fitting a low degree polynomial function, typically of quadratic or cubic degree, of adjacent data points by linear least squares [92-94]. For that, SG requires the number of data points (window size) and the order of the polynomial in the function to be specified. A balance between these parameters needs to be found as a wider window will lead to higher smoothing but a high polynomial order will better fit spectral features and accentuate the noise.

3.2.2. Empirical scatter correction methods

Spectra are often affected by the noise from the instrument error and by physical effects, such as light scattering, during sample measurement. In diffuse reflectance spectroscopy, the scattered light from the particles present in a suspension is collected. The light scattering is highly correlated with the number of light and surface interactions (which consequently, depends on the particle size, shape and concentration) and with the refractive indices differences between the particles and the medium/surroundings [95]. The light scattering will be higher when analysing small particles or/and when there is a high difference in the refractive indices. These non-chemical effects are regarded as unwanted and introduce non-linear spectral variations, causing the broadening of the absorption peak, non-linear changes of the absorption intensity and a shift of the baseline. Empirical-based signal pre-processing methods are commonly used to eliminate undesired interference and to enhance the chemical spectral features. Several pre-treatments methods have been reported to remove the non-consistent data information, to ascertain common metrics and to linearize the chemical related information, Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), Extended Multiplicative Scatter Correction (EMSC), Inverse Signal Correction (ISC), first or second derivative of the signal. By selecting a suitable pre-processing method, the resultant data can contain information with improved linearity to be captured by a regression modelling method [96].

3.2.2.1. Standard Normal Variate (SNV)

Standard Normal Variate (SNV) is one of the most common methods applied to remove or minimize the multiplicative interferences of scatter and particle size on the spectra. The method is used to mitigate the problem on the spectral intensity variation by centring and scaling each individual measurement. The corrected measurement, $x_{ik}^{*}(SNV)$, is individually obtained for each sample by subtracting each measurement at the k^{th} wavelength for the i^{th} sample, x_{ik}, from the mean of the measurement, m_i, and divided by the standard deviation of the measurement, SD_i, as shown in the following equation [97]:

$$x_i^*(SNV) = \frac{(x_i - m_i)}{SD_i}$$ **Equation 3.1**

where $m_i = \frac{1}{k}\sum_{k=1}^{K} x_{ik}$ and n is the number of samples spectra, and $SD_i = \sqrt{\frac{\sum_{k=1}^{K}(x_{ik}-m_{ik})^2}{K-1}}$ and K is the total wavelength number used.

After applying the SNV pre-treatment, there is always a zero mean value and a variance equal to one, and therefore, the final spectra are independent of the original intensity/counts values. It is also free from the influence of the data in the rest of the dataset, since m_i and SD_i are obtained from within the measurement.

When dealing with solids or densely packed samples, it can be observed a baseline shift and a curvilinear trend of the spectrum. To correct these variations, a de-trending normalisation can be applied jointly with the SNV, by using a second-degree polynomial regression [97].

3.2.2.2. Multiplicative Scatter Correction (MSC)

The Multiplicative Scatter Correction (MSC) was initially developed for NIR data to reduce or remove the signal noise [98-100]. Similar to SNV, it is used to reduce the spectral variability caused by path length effects such as particle size and light scattering among samples. The main difference between the two methods is that SNV standardises each spectrum using only data from that spectrum while MSC uses the entire dataset. The MSC is performed in two steps. The first step consists of fitting each spectrum to a reference spectrum, which is usually the mean spectra of the calibration dataset, through the following equation:

$$x_i(MSC) = a_i + b_i\, m + \, e_i$$ **Equation 3.2**

where the intercept *a* represents the additive effect*,* the slope *b* represents the multiplicative effect, *m* is the mean spectra of entire dataset, and *e* represents the

residuals, i.e., all the other effects in the spectrum that cannot be modelled by an additive or multiplicative constant.

Both a and b are estimated using ordinary least-squares regression of spectrum x_i versus m_i over all variables/wavelength k. The additive effect is usually caused by path length differences from sample-to-sample, due to positioning or irregularities on the sample surface, and can produce a baseline shift of the spectrum along the y axis. The multiplicative effect can be caused by differences in particle size and concentration, as light will be scattered at different angles and with lower or higher intensity, leading to differences in the local slope of the spectrum.

In the second step, the a and b values estimated are used to correct the raw spectrum, through the following equation:

$$x_{i,corr}(MSC) = \frac{x_i - a_i}{b_i} = m + \frac{e_i}{b_i} \qquad \textbf{Equation 3.3}$$

where $x_{i,corr}$ is the corrected spectra. The MSC is especially recommended when there is a significant impact of scatter effect in the spectrum in order to remove most of the spectral variation caused by additive and multiplicative effects. The deviations from the regression lines are then inferred as the chemical information enclosed in the spectra. If little or no scatter effect is present in the spectra, the chemical information will be recognised as the residuals/error in the Equation 3.2, and will have a significant impact on the slope and intercept of the regression line. If the MSC is applied in these cases, some chemical information may be removed from the spectra.

The parameters a and b estimated during MSC are strongly influenced by particle size, concentration and refractive index of the medium, and Chen and Thennadil, suggested that these parameters can be used for particle size and concentration prediction [101].

Several variations from the MSC have been developed and are commonly known as Extended Multiplicative Scatter Correction (EMSC). These versions can be based on a second-order polynomial fitting to the reference spectrum, on a polynomial baseline correction depending on the wavelength and can allow the incorporation of prior information from the spectral components [100] or physics-based models, such as

incorporation of transport theory-based parameters [102] and correction of Mie scattering artefacts in biological samples [103].

In this work, two EMSC forms wavelength dependent are considered. The first form, EMSCW, is described as follows [104]:

$$x_i = a_i + b_i\, m + \sum c_{comp} x_{pure,comp} + \; d_i\lambda + f_i\,\lambda^2 + e_i \qquad \textbf{Equation 3.4}$$

where c_{comp} is the concentration of each component (species) in the system, $x_{pure,comp}$ is the spectrum of the pure component, and both $d_i\lambda$ and $f_i\lambda^2$ are the wavelength dependent terms.

The second form uses the logarithm values of the wavelengths, referred as EMSCL, and has been suggested by Thennadil and Martin [105]:

$$x_i = a_i + b_i\, m + d_i log\lambda + \; e_i \qquad \textbf{Equation 3.5}$$

According to Rayleigh theory, the scattered intensity is strongly dependent on the wavelength, $\sim\lambda^{-4}$, and therefore, shorter wavelengths are scattered more strongly than longer wavelength. Taking the log term of the wavelength will emphasize the scatter effect on the spectra to be removed. However, this method should only be efficient when particles have a smaller size than the wavelength used, which is not the case of this work, thus it is not expected to improve the prediction models.

It has been also reported that the difference between the first polynomial and the logarithm transformation is minimal and the approaches are identical for all practical purposes [106].

3.2.2.3. Inverse Signal Correction (ISC)

Inverse Signal Correction (ISC) is the inverse version of MSC. In this method, the estimation of the correction parameters *a* and *b* is found by fitting the reference

spectrum m to each individual spectrum x_i, using ordinary least squares regression, through a similar equation as Equation 3.6:

$$m = a_i + b_i\, x_i \; + \; e_i \qquad \textbf{Equation 3.6}$$

Similar to MSC, different variations of ISC have been derived to include additional terms and/or reference signals, and are commonly called Extended Inverse Signal Correction (EISC). According to Helland *et al*. [107], the ISC/EISC is applied to model the noise on the reference/mean spectrum while the MSC method models the noise on individual spectra. In ISC, it is assumed that the error of the spectrum to be corrected is smaller than the error for the reference/mean spectrum. However, the expected noise level for this reference is of magnitude $\sqrt{n}$ smaller than the individual spectra and does not take into account the scatter effect differences in the dataset [106]. This can be disadvantageous as the ISC pre-processing will have a greater influence than the original MSC.

3.2.2.4. Derivatives

Another common pre-processing method to reduce scatter effects is the application of derivatives [95]. The first derivative is commonly used to remove the additive baseline, i. e., the baseline offset, and to improve the spectral resolution of the peaks. The peak maximum of the corrected spectrum will correspond to the slope of the peak from the original spectra, and will cross the zero value when the original peak exhibits its maximum value. The second derivative corresponds to the slope of the first derivative spectrum and it is used to eliminate the slope differences between spectra, removing a linear baseline and the physical properties of the sample. It is more similar to the original data but vertically inverted. Typically, the Savitzky-Golay (SG) and Norris-Williams algorithms are used in spectroscopy applications.

The application of derivatives may reduce the signal and enhance the noise in the spectra hence, smoothing needs to be applied prior the calculation of the first and

second derivative. Another disadvantage of the derivatives is the difficulty of data interpretation in the corrected spectra.

3.3. Multivariate Analysis

3.3.1 Principal Components Analysis (PCA)

Although Principal components analysis (PCA) is not applied in this study and it can be also considered as one of the common pre-processing techniques done prior a quantitative analysis, it is here reviewed as it provides a better understanding of the Partial Least Square Regression (PLSR) analysis since both are used for dimension reduction in regression analysis when some independent variables are correlated.

Principal components analysis (PCA) is often used for cluster analysis, to identify patterns in data highlighting their similarities and differences without reference to prior knowledge about the sample conditions. In a dataset that contains a large number of variables, PCA is applied to reduce the number of dimensions while keeping as much as possible the variation in the dataset. This data compression is achieved by modelling a set of observed spectra into linear combinations of a smaller set of uncorrelated and ordered variables, called principal components (PC) or latent variables (LVs). The reason behind the data compression is that smaller sets are much easier to explore, analyse, visualise and detect outliers. The first few LVs hold most of the variation present in the original variables, and the succeeding LVs retain the remaining variability present in the data [108]. Hence, two or three LVs are usually enough to plot to identify trends and patterns, but for modelling purposes the total number of LVs capturing most of the variance in the dataset should be properly determined, for example, by cross-validation or test set validation [109].

The first step of PCA is to create a covariance matrix $COV(X)$ as an input dataset to assess how the variables are correlated with each other and describe the variance of the data by using the Equation 3.7:

$$COV(X) = \frac{X^T X}{n-1}$$ **Equation 3.7**

in which the matrix X has n rows and k columns.

The matrix X is then decomposed as the sum of the outer product of vectors score *t* and loading *p* plus a residual matrix e as shown in Equation 3.8:

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_j p_j^T + e = TP^T + E$$ **Equation 3.8**

where $j = 1,2 \ldots J$ represents the number of LVs used by the model which must be less than or equal to the smaller dimension of X, i.e. $j \leq \min\{n, k\}$ [110, 111]. The residual matrix E represents the noise and contains the unexplained variability of X. The scores for each sample and LV are stored in matrix T and are multiplied by a matrix P that contains the all the loadings, in order to approximate the systematic variation in X. Figure 3.1 shows a schematic of the decomposition of X matrix by PCA described in Equation 3.8.

For each sample (each row in the matrix X), there is a t_i score value that describes their distance from the origin as well as their relationship to each other. The loadings, also called eigenvectors of the covariance matrix, describe the direction of the variables in the hyperplane. Each sample has an eigenvalue λ_i associated with the eigenvector p_i:

$$COV(X)p_i = \lambda_i p_i$$ **Equation 3.9**

This linear relationship allows the fitting of a straight line that passes through the origin, a LV, that describes the variance of the data. Therefore, it possible to interpret p_i as the LVs that model X. PCA tries to find the best fitting line, i.e., LV, by rotating the line until it fits the data as well as possible.

The score vector t_i , orthogonal set, is the linear combination of the original X data, which is defined by p_i , orthonormal set, and it is obtained by the following equation:

$$t_i = Xp_i \qquad \textbf{Equation 3.10}$$

In this case, p_i can be interpreted as the weights, defining the contribution of each variable to the respective LV.

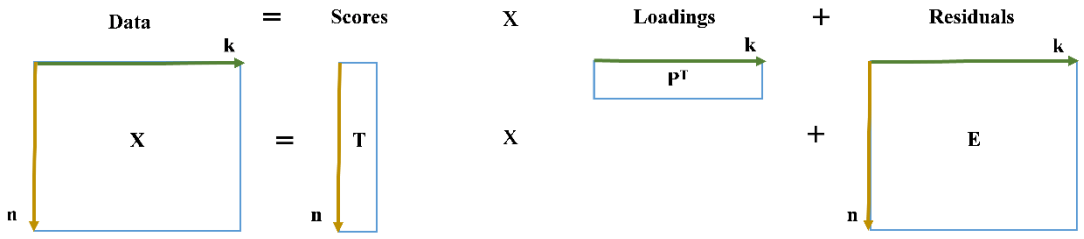


Figure 3.1: Schematic of the decomposition of X matrix by PCA, and its relationship with scores T, loadings P and residuals E, where n represents the sample number and k the wavelength/bin/Chord length.

The amount of variance/information in $t_i p_i$ explained by each LV is ranked from the highest to the lowest order of the eigenvalue λ_i. The first LV captures most of the variance in the dataset. The next LV repeats the process to capture most of the variance remaining in the data, and the same process repeats until a maximum number of LVs is reached. The maximum number of LVs is normally determined by considering the true variation in the dataset. For instance, a simple two-component spherical particle suspension could have a maximum number of 3 LVs for explaining the NIR data (medium concentration, particle concentration, particle number density). This process reduces the original amount of data to describe a system/process to much fewer variables. The following LVs capture the remaining variance.

For interpretation, the scores and loadings can be plotted individually. For example, the scores can be plotted against each other to evaluate the samples similarities or differences, find clusters and outliers. The loadings plot, can be used to evaluate which of the variables have a bigger contribution in each of the LVs. The loadings can also be plotted versus wavelength number, to find the spectral pattern of each LV.

3.3.2. Partial Least Squares Regression (PLSR) analysis

Partial Least Square Regression (PLSR) analysis is an extension of PCA. While PCA maximises the covariance of only matrix X, PLSR maximises the covariance between all the linear functions of X and its reference Y values. In modern instrumentation, such as UV-vis-NIR spectroscopy, the X variables are often affected by multicollinearity. This happens because the number of X-variables are often much larger than the number of available samples, resulting in linear relationships among the variables in matrix X. PLSR was developed to handle complex problems, to analyse strongly collinear, noisy, numerous and incomplete X variables, and model their response variables Y [112]. The technique is essentially used for quantitative analysis, to predict the Y values from the X-variables.

Before the analysis, the X and Y variables are often transformed by pre-conditioning or pre-processing in order to eliminate unwanted features or interference in the spectra and have variables with a fairly similar contribution to the model.

The original data, matrix X and Y are decomposed into T-scores and U-scores, respectively.

The algorithm selects the initial score vectors u, usually $u_1 = y$, to describe the maximum covariance in Y matrix, and determines the initial X variables loading weights, r_1, which are normalise to unity length, by regressing u on X:

$$u_1 = Xr_1 \qquad \textbf{Equation 3.11}$$

The loading weights describe the direction of the first LV.

The first approximation of the score vector t_1 is obtained by multiplying the normalised weights (r_1) to X:

$$t_1 = Xr_1 \qquad \textbf{Equation 3.12}$$

The regression coefficient q is obtained by regressing y onto t_1:

$$q = yt_1$$ **Equation 3.13**

The new u score can then be obtained by multiplying the weight:

$$u_2 = yq_1$$ **Equation 3.14**

Hence, the updated Y matrix can be described by the multiplication of U scores by the Q loadings with the associated residuals G:

$$Y_{ik} = \sum_J u_{ij}\, q_j + g_i \qquad Y = UQ^T + G$$ **Equation 3.15**

where $j = 1,2 \ldots J$ represents the number of LVs used by the model which must be less than or equal to the smaller dimension of X, i.e. $j \leq \min\{n, k\}$.

The cycle is repeated until the score vector converges until the residuals are smaller or the maximum number of LVs is reached. When a convergence is reached, the loadings **p** are calculated based on **X** and the **t** scores:

$$p_j = \frac{X^T t}{t^T t}$$ **Equation 3.16**

The X matrix can then be described by the multiplication of scores T with the loadings P so the residuals E are small, through the following formula:

$$X_{ik} = \sum_J t_{ij}\, p_{jk} + e_{ik} \qquad X = TP^T + E$$ **Equation 3.17**

The X-scores are also good predictors of Y:

$$Y_i = \sum_j q_j \, t_{ij} + f_i \qquad Y = TQ^T + F$$ **Equation 3.18**

where F are the residuals. Figure 3.2 shows the components obtained from matrices X and Y decomposition, presenting the basic equations of multivariate latent variable analysis.

In the presence of multiple response variables and a predictor, the Equation 3.12 and Equation 3.18 can be rewritten as the following Equation 3.19 in order to obtain a multiple regression model:

$$Y = XRQ^T + G = XB + G$$ **Equation 3.19**

The regression coefficients B can be written as:

$$B = RQ^T = R(P^TR)^{-1}Q^T$$ **Equation 3.20**

The residual matrices E and F are then obtained by subtracting the loading and scores from X, and Y loadings, scores and weights from Y, respectively:

$$E = X - TP^T$$ **Equation 3.21**

$$F = Y - UQ^T$$ **Equation 3.22**

The subsequent LVs are calculated by replacing X and Y for E and F to start the new regression models. The number of LVs needed to describe a model is usually determined empirically through cross-validation, which is described in more detail in the next section 3.2, until there is no more significant information in X about Y to be captured.

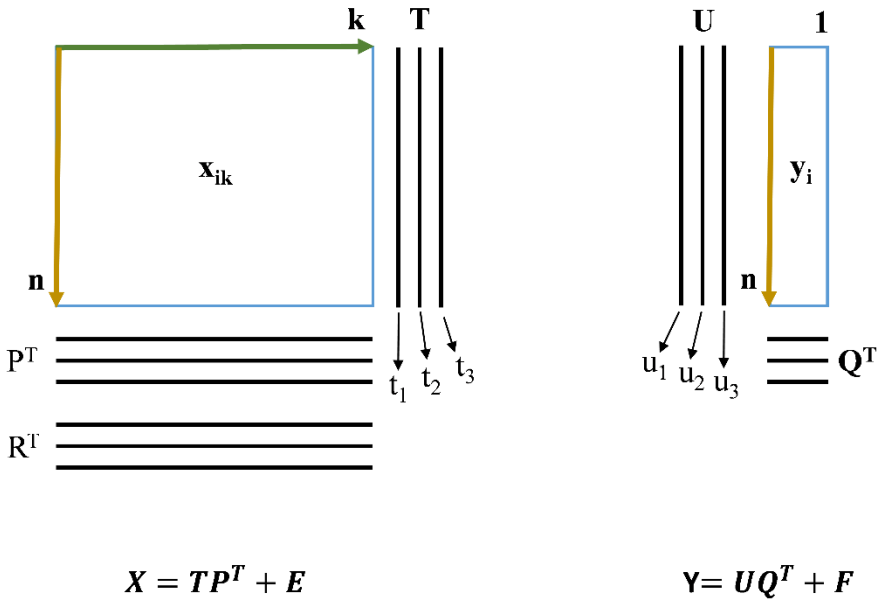


Figure 3.2: PLSR matrix decomposition of the spectra/CLD in scores T, loadings P^T, weights, R^T, and matrix decomposition of the reference values in scores U^T and loadings Q^T.

Few techniques that are PLSR based, such as PLS2 and Multiblock PLS (MB-PLS, have for example, have been also used to extract particle size information from NIR spectroscopy. PLS2 is an extension of PLSR which includes several dependent variables and it has been applied to investigate PSD of aggregates (a natural product composed basically by silicon dioxid) and powdered microcrystalline cellulose in [113, 114]. In the case of the Multiblock PLS, while PLSR combines all the variables in a single block for prediction purposes, Multiblock PLS (MB-PLS) separates the predictors into subsets or blocks. The advantage of this strategy is that by focusing on separate subset of data, it is possible to learn more of some local phenomena and hence, improve the interpretability of the results [115]. MB-PLS approach has been applied for modelling and monitoring a variety of systems, e.g., to assess the influence of each stage (inoculum production and fermentation) in the overall production process of an API [116], for determination of corn and tobacco samples by using near-infrared diffuse reflection spectroscopy [117], to detect and quantify the adulteration of diesel/biodiesel blends by vegetal oil using MIR and NIR spectroscopy [118], for calibration development of soybean flour quality properties (crude protein and moisture using MIR and NIR spectroscopy [119] and to predict PSD in pharmaceutical

powders [80]. In general, even if the prediction model does not improve by using MB-PLS, the method provides deeper understanding of the contributions of the data blocks.

3.4. Validation

Multivariate data analysis relies on the application of empirical models to linearize and summarise the information from the dataset, to allow a better prediction of the properties of interest. Since the model is based on the best fit for the calibration data, it is crucial to assess whether it captures the true information and is robust and reliable for predicting future measurements.

The models performance can be validated by using only the training data, denominated cross-validation (CV), or by adding an external test set, test set validation. According to Kos et al., cross-validation is suitable for medium to small dataset (<50), while test set validation can be applied for larger datasets (>50) [120].

Cross-validation is especially valuable when the number of experiments/samples analysed is limited and to understand the dependence of the models on the number of latent variables used, i.e., to define the number of components required to reconstruct the data matrix with the smallest error [121].

The cross-validation is given by the root mean square error of cross-validation (RMSECV), according to the Equation 3.23:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_{i,cv} - y_i)^2}{n}} \qquad \textbf{Equation 3.23}$$

where n is the number of calibration samples, y_i is the measured reference y value for sample i and $y_{i,cv}$ is the estimated y_i obtained from the calibration equation.

The simplest version of cross-validation is the leave-one-out cross-validation. This means that the PLS model is built using $n - 1$ samples, i.e., one of the samples is removed from the calibration set and it is used as a validation sample for the model.

After the $y_{i,cv}$ from the removed sample is estimated and compared to its respective y_i, the sample is re-introduced in the calibration set, and another sample is removed. The process is repeated until each one of the calibration samples i is used once as a validation sample.

The advantage of leave-one-out cross-validation is that if used to validate a large enough dataset, all samples are used and the bias will be low. However, the fact that the model is validated against only one sample, the $y_{i,cv}$ estimated is highly dependent on the quality of the measured sample, and if the sample is an outlier, it will lead to a high variation in the model and reduce its effectiveness.

Alternatively, a n fold cross-validation can be implemented. This validation technique consists of removing a group of samples from the calibration set, usually between 5 to 10 samples, to be used as a validation set. In this case, a large calibration set is required to ensure that any relevant information or trend is removed from the model, otherwise it will lead to a higher bias. The n fold cross-validation can be particularly useful when analysing replicates of samples, in which the replicates can be used as a validation set and it is warranted that the calibration set contains this information to build an all-inclusive PLS model.

To truly test the model performance and stability, a test set validation is desired. The test set comprises a set of independent samples which needs to be representative of the sample conditions contained in dataset as a whole. The test samples are not included in the calibration set and should only be used after the model is trained, i.e., after the calibration set is appropriately validated with CV, in order to ensure an optimum number of components is used to describe the data variance. The model performance can then be evaluated by calculating the root mean square error of prediction (RMSEP), described in Equation 3.24:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_{i,p} - y_i)^2}{n_p}} \qquad \textbf{Equation 3.24}$$

where n_p is the number of samples in the test set, $y_{i,\mathrm{p}}$ is the predicted variable and y_i is the measured reference y value for sample i.

The accuracy of the model can also be expressed using the correlation coefficient R^2, shown in Equation 3.25, which indicates the degree of agreement between the predicted and the measured reference values.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \qquad \textbf{Equation 3.25}$$

where SS_{res} is the sum of squared residuals:

$$SS_{res} = \sum_{i=1}^{n} (y_i - y_{i,cv})^2 \qquad \textbf{Equation 3.26}$$

and SS_{tot} is the total sum of squares:

$$SS_{tot} = \sum_{i=1}^{n} (y_i - y_{mean})^2 \qquad \textbf{Equation 3.27}$$

A R^2value of 0 indicates no correlation was found while a value closer to 1, indicates a linear relationship and a good agreement between the reference and predicted values.

Chapter 4 – Experimental Procedure

This chapter describes the materials, the off-line and in-line techniques used for sample characterisation and the respective experimental procedures for evaluating the data collected. The chapter also comprises the strategy to extract and process the data generated by each technique. The methodology used to build and test the multivariate calibration models developed is also described.

Since SAR-DRM is for the first time applied to study micron-sized particle suspension, the last section (section 4.7) explores the optical interference from other optical probes used simultaneously during the experiments (FBRM and PVM), and evaluates SAR-DRM's signal reproducibility and consistency.

4.1. Materials

In this study, two types of samples were investigated: polystyrene (PS) and α-lactose monohydrate. As SAR-DRM was for the first time applied to study particles in the micrometre range, the first challenge was to find a simple, affordable and robust model system, containing only spherical particles and available in relatively large quantities, to evaluate the performance of the optical probes. Polystyrene particles synthetized in-house from suspension polymerisation reactions [122] were selected to be a suitable candidate for the purpose. The PS particles were sieved into 6 size groups from < 90 to 800 μm in diameter. The particles from each particle size group were suspended in DI water in order to prepare suspensions of different solid loadings, from 1 to 10 wt.%. The knowledge acquired from these controlled experiments may lead to a better understanding of the impact of the particle attributes on the optical measurements. The model system allows a comparison between the performance of different commercial techniques used and enables to assess the feasibility of applying SAR-DRM for measuring particle size and concentration. Next, α-lactose monohydrate crystals, obtained from DFE pharma, were sieved into 9 size groups and suspended in acetone (99.5 % purity) to test whether the same methodology can be applied to characterise crystals. Suspensions varying from 0.5 to 25 wt.% of solid loading were prepared. Table 4.1 summarises the size ranges of the materials and their respective nomenclature used in this study. The gaps found in the sieved sizes ranges are due to the insufficient sample to prepare the required solid loadings through a gravimetric analysis. The corresponding properties are described in Table 4.2.

Table 4.1: Range of sizes analysed per each sample.

Material	Size range (μm)	Nomenclature
Polystyrene (PS)	<90	PS0-90
	125-180	PS125-180
	180-250	PS180-250
	250-350	PS250-350
	300-500	PS300-500
	630-800	PS630-800
α-lactose monohydrate	<38	L0-38
	38-75	L38-75
	90-125	L90-125
	125-150	L125-150
	150-180	L150-180
	180-212	L180-212
	250-300	L250-300
	300-355	L300-355
	355-400	L355-400

Table 4.2: Materials and solvent used and corresponding properties.

Material	Product Information	Density	Refractive Index
Polystyrene (PS)	Produced in-house and described in [122]	1.04 g/cm^3	1.59 (λ = 587.6 nm) [123]
α-lactose monohydrate	DFE pharma	1.54 g/cm^3 (at 20ºC) [124]	1.65 [125]
Water		1.0 g/cm^3	1.33 [126]
Acetone		0.7845 g/cm^3	1.359 (λ = 587.6 nm) [127]

4.2. Measurement and data analysis procedures

In this work, calibration models are built for particle size and concentration of polystyrene and lactose samples. The objective is to be able to monitor this changes in real-time. To build these models, reference values are needed. Although particle concentration can only be determined by gravimetric analysis, PSD can be obtained through a variety of off-line techniques. As discussed in Chapter 2, each technique relies on different physical principles and has a preferred mode of expressing PSD, which makes their direct comparison not straightforward. Moreover, most of the in-line techniques do not directly provide the PSD of the sample but a signal that can be related to it, e.g., CLD, baseline shifts in the spectra. In order to compare the performance among off-line and in-line technologies, prior to the model building, it is important to know how the data is collected, how the PSD distribution is represented and how the signal can be processed or inverted to obtain the PSD.

Hence, this section presents not only the measurement conditions but also details on the particle representation and processing algorithms used for posteriori analysis. Most of the methods presented are widely used in the pharmaceutical industries or/and in current research to study crystallisation processes, which makes them attractive to validate SAR-DRM.

4.2.1. Off-line methods

Imaging

The static microscope-imaging analysis was performed using Morphologi G3 (Malvern Instruments Limited, UK) to measure particles size and shape. The PS and α-lactose monohydrate particles were dispersed using an integrated dry powder disperser and individual particles were scanned under the microscope with images captured using a high resolution digital camera. The setup typically measures particles of size between 0.5 and 1000 μm.

The instrument utilises an advanced graphing and data classification software that provides analysis on single particles as well as particle distribution statistics. To exclude partially imaged or overlapped particles, as well as some contaminants such as dust or fibres, from the analysis, morphological filtering of the raw data was applied based on the particles area. Exclusion of particles was double-checked manually. At least 500 particles per each sample group are analysed to form a statistical representation of the population. Results of the analysis can be displayed as Equivalent Circle Diameter (ECD), aspect ratio, elongation, circularity, convexity and solidity. In this study, ECD and aspect ratio are used for further analysis.

Laser diffraction

Laser diffraction measurements were conducted using a Mastersizer 3000 (Malvern Instruments Limited, UK). PS samples were dispersed in DI water using a Hydro MV cell (Malvern Instruments Limited, UK). The α-lactose monohydrate crystals were dispersed in acetone. Particles were added to the cell until a laser obscuration (internal parameter of the instrument used to avoid multiple scattering effects) of 15% was reached. Five measurements were taken for each sample. A Hydro Sight (Malvern Instruments Limited, UK) was connected to the dispersion unit to capture images during the laser diffraction measurement. The results of the measurements are expressed as the volume weighted distribution of Equivalent Sphere Diameter (ESD).

The D50 from the PSDs obtained, i.e., the diameter of the particles at which 50% of the particles in the sample are smaller, was used as the reference particle size to build calibration models. This technique was chosen as it is widely accepted as a standard particle sizing across many industry sectors for a wide range of particle sizes.

Scanning electron microscope (SEM)

SEM images of polystyrene particles were recorded in backscattered mode at a beam voltage of 5kV using a Hitachi SU6600 Field Emission Scanning Electron Microscope

(FE-SEM). Sample preparation involved mounting the samples on aluminium stubs using a carbon tape, and a carbon coating was applied to avoid the charging effect.

Optical microscopy

Microscopic images of polystyrene samples were obtained using an optical microscope (LEICA QEDAF), to assess particles morphology and surface roughness.

4.2.2. In-line methods

In-situ imaging

The images of the particles in suspension were captured using a PVM V819 probe (Mettler Toledo, UK). The PVM probe provides eight near-infrared (808 nm) laser beams, six for forward illumination and two for back illumination. Each of the laser beams can be individually turned on/off for different illumination requirements. Greyscale images were collected from the scattered laser illumination using a CCD with a field of view at 1075 μm × 825 μm. Each image frame consists of 1360×1024 pixels with a pixel size of 0.8 μm. Diffuse back illumination is achieved using a reflector Teflon cap to diffuse and direct the illumination into the suspensions, resulting in a transmission image rather than a backscattered image. Ten images per second were collected for all experiments ran.

The processing of the particle images follows the procedure described in Ref. [12]. The image analysis algorithm [128] only accepts particles in focus and larger than 24 μm, and the results are expressed in the form of Equivalent Circle Diameter (ECD). As the image of a particle is a 2D projection of the 3D objects, the correct description of the PSD requires analysis of a large number of 2D particle views, typically >500 valid particles. The algorithm also provides aspect ratio distributions, where the aspect ratio is defined as the ratio of width to length of the bounding box that best encloses the particle, as described in [12].

Chord length measurement

CLD measurement was obtained using an FBRM G400 system (Mettler-Toledo, UK). The instrument can acquire up to thousands of chord lengths per second, and the counts of the chords are organized into 100 CL groups to express as a CLD. The FBRM software offers two modes to express CL measurements, namely, primary and macro. The primary mode has enhanced sensitivity to fine particles or sharper edges of particles, while the macro mode provides a slightly larger focal zone and a digital filter to lower sensitivity to the edges when particles are in proximity, favouring the measurement of large particles. Furthermore, the shape of the CLD is affected by the weighting method employed, namely unweighted or square weighted. Previous studies suggest the square weighted CLD provides a good qualitative approximation to the volume weighted PSDs obtained by other particle size measurement techniques [44]. However, the CLD can only be used to qualitatively infer the particle size. PSD is more effective as it is directly related to product quality and process productivity. To allow more direct comparison to other techniques, the volume-weighted PSD can be inverted from the unweighted CLD using the Fredholm integral equation (first kind) [40].

Considering the broad range of particle sizes analysed in this work and the requirement for quantitative comparison with other particle measurement techniques, the macro mode is chosen for all samples investigated. The unweighted CLD is inverted to estimate PSD using the algorithm developed by Agimelen *et al*. [42, 129, 130]. No *a prior* information about particle size range or particle aspect ratio are required for the inversion, and particles in each size group are assumed to have the same aspect ratio. The final results are presented as an ECD distribution.

Spatially and angularly resolved diffuse reflectance measurement (SAR-DRM)

SAR-DRM is a multi-wavelength fibre-optic system which collects diffuse reflectance at various spatial and angular configurations between the source and detecting fibres [9]. The system consists of a light source, an optical multiplexer, two spectrometers and an optical fibre probe. Figure 4.1 shows a schematic drawing of SAR-DRM system.

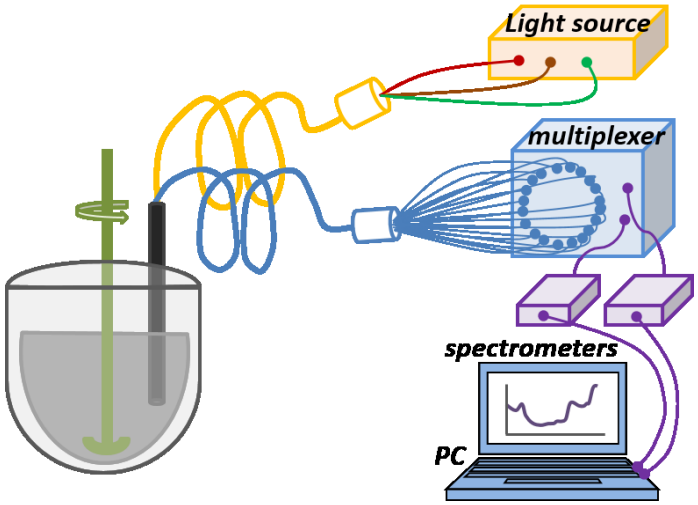


Figure 4.1: Schematic drawing of SAR-DRM system [9].

SAR-DRM uses a 50 W halogen lamp to emit radiation. The lamp is located inside of a custom-made lamp house and has an input voltage of 12V. The light is directed to source fibres through a brass ferrule, which can be manually connected to the lamp house. In order to block any incident light entering the optical fibre, a shutter is included in the lamp house.

The emitted light propagates from the lamp house and couples with the source to deliver the incident light at the probe end, until it reaches the sample. The light reflected from the sample is collected by detecting fibres and guided into a multiplexer. In this work, a 16 x 2 fibre-optic multiplexer (FOM-UVIR200-2x8, Avantes) is employed to sequentially couple two output channels with 16 input channels. The multiplexer uses a controlled stepper motor and a rotary block to switch the channels

according to the pre-defined sequence and time interval set by the user. Two optical fibres are coupled to the multiplexer to direct the light to 2 spectrometers, a UV-vis spectrometer (USB4000-UV-VIS, Ocean Optics) and a NIR spectrometer (NIRQuest 2.2-512, Ocean Optics). The multiplexer and both spectrometers are connected to a computer via USB ports, enabling the recording of the data as a function of wavelength over the given spectral range and the collection of a sample spectrum. In this work, the UV-vis spectrometer uses 3648 pixels to cover a spectral range between 350 and 1000 nm with a 0.2 nm spectral interval. The NIR spectrometer uses 512 pixels to provide a spectral range between 900 and 2100 nm with a 5 nm spectral interval. As the optical multiplexer enables the controlled switching of the detector fibres, a UV-vis-NIR spectrum is acquired per each detector channel used.

The SAR-DRM probe is a custom built probe with 5 sources fibres to irradiate incident light to the sample and 16 detecting fibres to collect the diffuse reflectance signal. Figure 4.2(a) shows the SAR-DRM's source and detecting fibres arrangements viewed from the probe end.

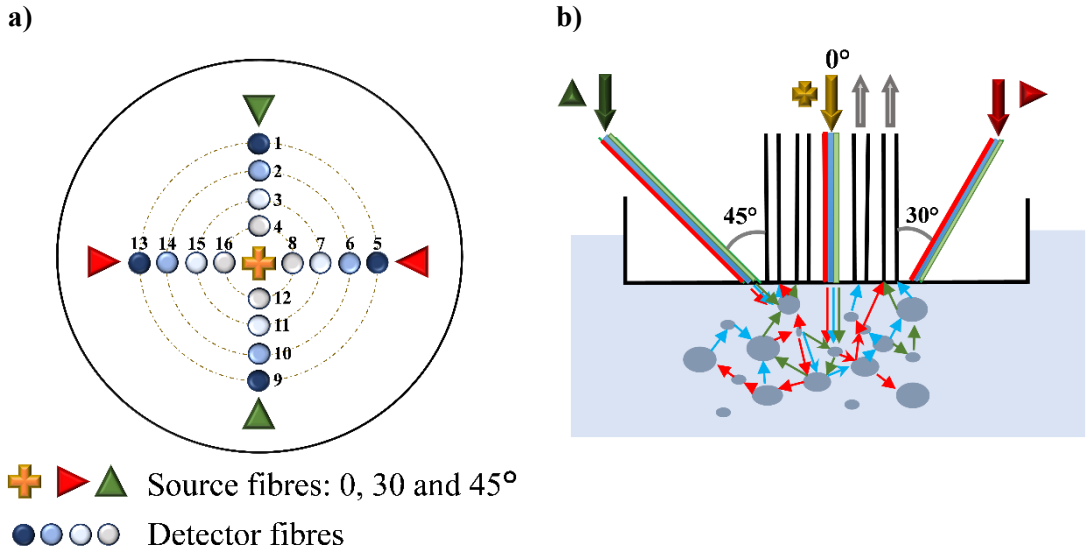


Figure 4.2: a) Design of SAR-DRM probe. **b)** Illustration of light scattering when sample is irradiated using the incident light source at 0, 30 and 45°.

The probe offers 3 angular light sources, 0, 30 and 45°, respective to the probe axis, denominated L1, L2 and L3, respectively. The 16 detecting fibres are arranged at 4 different spatial distances, 0.3, 0.6, 0.9 and 1.2 mm, core-to-core, to the source fibre in the branch and are named in this work as D1, D2, D3 and D4, respectively. As illustrated in Figure 4.2(b), the different spatial and angular arrangement between the source and detecting fibre results in differences in the light travelling pattern in a particle suspension. A 50 wt.% suspension of 25 nm silica nanoparticles in water was used to demonstrate the propagation of normal and angular incident light in a low turbid medium, as seen in Figure 4.3.

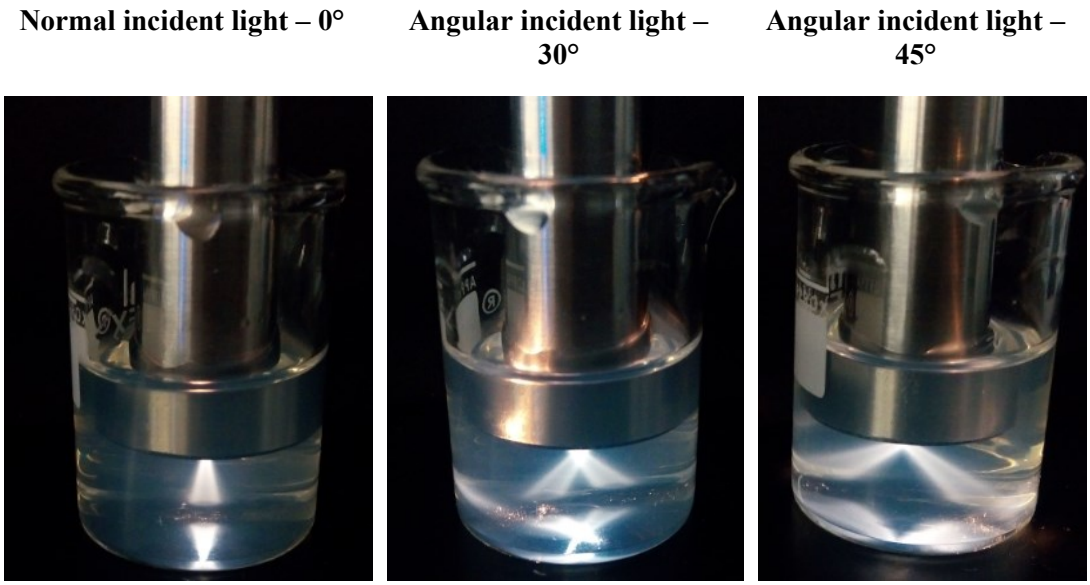


Figure 4.3: Light propagation in a low turbid medium of a 50 wt.% suspension of silica in water, with a particle size of ~25 nm, using the SAR-DRM angular sources 0, 30, 45°.

A maximum of 48 combinations (3 incident angles x 16 collecting fibres) of source and collecting fibres are offered. In total, SAR-DRM system produces 96 spectra per sample analysed, 48 in the UV-vis-short NIR wavelength range and 48 in the NIR range. However, taking into consideration the fibre arrangement at the probe end, four branches of detecting fibres are at equivalent distance from the normal incident light (Fibres 1-4, fibres 5-8, fibres 9-12 and fibres 13-16) and should result in an identical signal. Regarding the position of the angular incident sources, two branches of

detectors are at equivalent distance. For example, when light is emitted by the 30° incident source, the fibres 13-16 (at 0.3, 0.6, 0.9 and 1.2 mm, respectively) are a repetition of the fibres 5-8. This identical fibre arrangement can be used to assess the measurement precision as it is expected identical fibre arrangements to have identical spectral responses from the suspension. The exclusion of the measurement repetitions from the analysis (after assessing its signal consistency) reduces the data from 48 spectra per sample and per region analysed to 12 meaningful spectra (3 incident angles × 4 spatial distances), simplifying the data analysis. The selection of the branches is described in more detail in section 4.7.3.

The different configurations used are described in relation to the source and detector distance used, for example, if the detector at 0.3 mm (D1) from the normal incident light source (L1) is used, the configuration is named as D1 – L1. If all the detectors at different distances from the normal incident light are used, it is described as D1:D4 – L1.

During the measurements, only one angular light source is used at a time to avoid signal cross-talking, e.g. if the incident angle at 0° to the probe axis is illuminating the sample, the angular sources 30 and 45°, will not be used at the same time.

SAR-DRM uses low-OH optical fibres with 200 μm in diameter and with a numerical aperture of 0.22. The fibre length is 2.75 m. The 16 detector leg ends have a SMA 905 strain relief connector to couple with the optical multiplexer.

The optical probe geometry, angle of incident light and source detector distance can affect the optimal illumination, the penetration depth of light and the light collection in spectroscopic applications [131-133]. This project studies whether emitting light from different angles and collecting it from different source-to-detector distances offers more qualitative and quantitative information of the system in analysis.

- SAR-DRM spectra acquisition parameters

An ideal spectrum should not be either saturated or too weak in order to preserve the wavelength information and to display the spectral features and bands well resolved. Parameters such as integration time and averaging have a significant impact on the

quality of the spectra taken. The integration time defines the length of time over which the light detected. The longer the integration time, the higher the amount of light detected and the larger the peaks will appear in the spectra. However, if too long, it can lead to the acquisition of unnecessary noise, saturation of the detectors and the lag of the software. If the integration time is too short, less photons will reach the detector and they might not be in a sufficient number to overcome the signal associated with the noise, resulting in a poor signal-to-noise ratio spectrum. Therefore, the integration time needs to be adjusted to maximise the signal without endangering the reliability of the spectrum. As the signal measured varies from sample to sample, with sample conditions, such as particle concentration and size, and with the spatial distance of the detector regarding the light source in use, several tests were performed to ensure suitable integrating times for the suspensions studied, for each spectral region. The conditions used in this study are summarised in Table 4.3.

Table 4.3: Integration times and spectra averaging used to study polystyrene beads and α-lactose monohydrate crystals by SAR-DRM.

Sample	Spectrometer	Integration time (ms)	Spectra Averaging	Spectra Repetitions
Polystyrene	UV-vis	30	5	5
	NIR	2000	1	5
α-lactose monohydrate	UV-vis	40	5	5
	NIR	10000	1	5

The noise component in NIR is in general much stronger than in UV-vis, therefore, the NIR region measurements required a much higher integration time. As the signal-to-noise ratio can be improved by averaging the signal, the measurements were signal-averaged from five measurements for the UV-vis region. In all the experiments performed in this work, 5 spectra repetitions were analysed to assess the precision of the spectra measured by a fibre and its equidistant counterpart fibre as regards to the same light source.

Since scattering is the prominent effect manifested in the visible to short NIR spectral region, only signal in wavelength between 500-900 nm is analyzed for both Polystyrene and α-lactose monohydrate suspensions. The NIR region selected to develop calibration models was from 1000-1850 nm. However, the NIR region was only evaluated for the polystyrene system as a significant number of lactose samples resulted in the saturation of the signal.

- SAR-DRM spectra acquisition

Spectral measurements were performed using a graphical interface developed in MATLAB (R2012b, Mathwork). Both spectrometers and optical multiplexer were purchased with their own software. Nonetheless, the algorithm developed in MATLAB enables the synchronisation of the spectral acquisition with the switching of the multiplexer channels. The program also provides a quick visualisation of the spectrum while it is being recorded.

Before commencing data collection, the SAR-DRM system was switched on for at least half an hour to allow the spectrometers to warm up and the light source to stabilise. Figure 4.4 displays the spectra acquisition process by using this interface and the data analysis after all the spectra are acquired.

After the hardware setup, the graphical interface is loaded and followed by the initialisation of the spectrometers. The spectrometers and multiplexer connection with the computer is checked. Before starting the experiments, it is crucial to define the experiment name, the number of scanning repetitions and the spectral acquisition parameters. These parameters specify the detector fibres to be used, the integration times and the averaging to be applied in each spectral region to be analysed. After all these parameters have been established, the program is ready to run. For each fibre analysed, the acquired signal is updated and displayed in the computer program, with the data being continuously saved in a *.mat* file.

In total, four type spectra are acquired during the experiment: reference spectrum, reference background spectrum, sample spectrum and sample background spectrum.

To improve the measurement accuracy, it is essential to calibrate SAR-DRM by taking a reference and a background spectrum. To obtain the reference spectrum, the probe is fitted into a 3.3" integrating sphere (Newport, New Hampshire, US) via an adaptor for a 1.0" port. The integrating sphere uses a highly diffuse reflective white coating made of Spectralon® which is designed to exhibit a near perfect Lambertian surface. When SAR-DRM's incident light hits the sphere wall, it undergoes multiple diffuse reflections, distributing the light uniformly inside of the sphere. As the integrating sphere preserves the light intensity, the total incident light emitted is collected. The reference spectrum will therefore be a measure of the light intensity emitted to analyse the samples afterwards. This baseline information is collected for each light source used over the specified spectral range. The spectra are taken when the incident illumination is on and off. The shutter included in the lamp house is used to block the incident light. The reason behind taken a spectrum when no light is coming in from the light source, designated as a background spectrum, is to collect any possible bias relative to the electronic noise from the system. This background spectrum is then subtracted from the reference spectrum to eliminate any interference from the desired light source intensity quantification spectrum. Both measurements are taken when the integrating sphere and probe are wrapped with a dark cloth.

The same procedure is followed for the sample measurement. The probe is immersed in the particle suspension in the reactor and a spectrum is taken when the incident illumination is on and off, for all SAR-DRM incident light sources. The sample background spectrum is used to capture any possible contamination from spurious light from the experimental environment, to account for the nonlinear response of the optical components, and variation of coupling efficiency among the channels of the multiplexer. This contamination can then be removed by subtracting the sample background spectrum from the sample spectrum.

After all the spectra are collected, they are subjected to further signal process processing to remove system dark noise. The normalized spectra are obtained by computing the following formula:

$$\text{Processed Signal} = \frac{\text{Sample light on } (I_{on}) \; - \; \text{Sample light off } (I_{off})}{\text{baseline light on } (B_{on}) \; - \; \text{baseline light off } (B_{off})}$$ **Equation 4.1**

The processed spectra are then used to build multivariate regression models to estimate particle size and concentration. All data process was done by using MATLAB (R2016b, Mathwork).

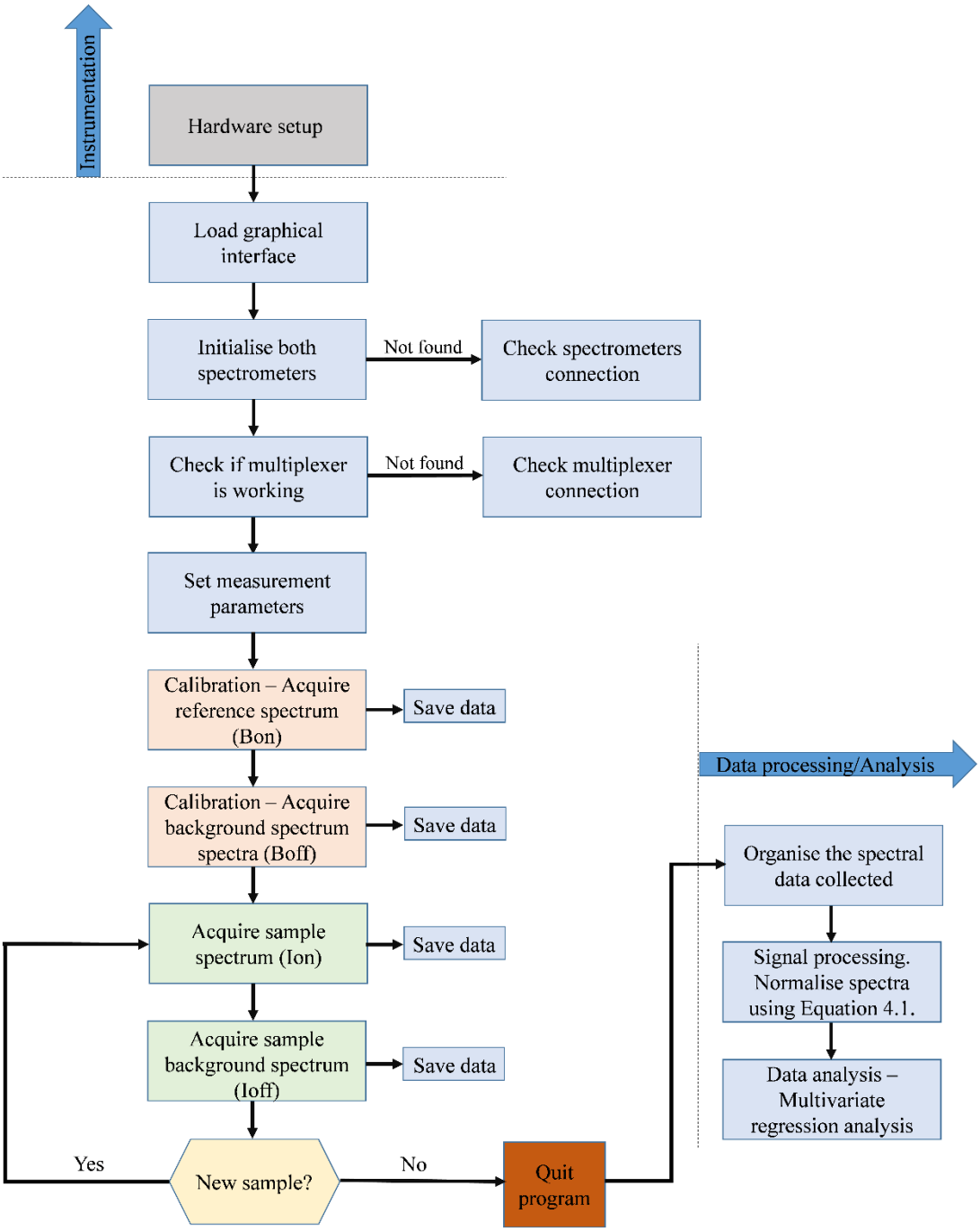


Figure 4.4: SAR-DRM spectra acquisition and data analysis process.

4.3. In-line experimental set-up

4.3.1. Experimental set-up for Polystyrene Measurements

The measurements were performed on polystyrene suspensions placed in a 250 ml Radleys glass jacketed tank reactor with a Polytetrafluoroethylene (PTFE) pitch-blade stirrer. The effect of stirring speed, direction and probe positions on the measurement were assessed to ensure the final setup has minimum interference to the optical measurements. Solid loading and particle size effect were evaluated using SAR-DRM, FBRM and PVM at the optimum setup found, as shown in Figure 4.5. The stirrer operated clockwise at 500 rpm and the FBRM and PVM probes were set at an inclination of 30º. The setup was covered by a dark cloth to free SAR-DRM from interference of ambient light. A total of 30 samples with a solid loading ranging from 1 to 10 wt.% and a particle size range from <90 to 800 µm were analysed. Solid loading was determined by a gravimetric analysis. For that, 300 g of water were initially weighted and added to the reactor. The overhead stirrer was switched one. The mass of polystyrene to add to the reactor to obtain the desired solid loadings (1, 2.5, 5, 7.5 and10 wt.%) was previously calculated, and then the mass wanted was weighted and added to the system through a funnel. The probes were switched on and the particle suspension was left to stir for approximately 10 min prior to start taking the measurements.

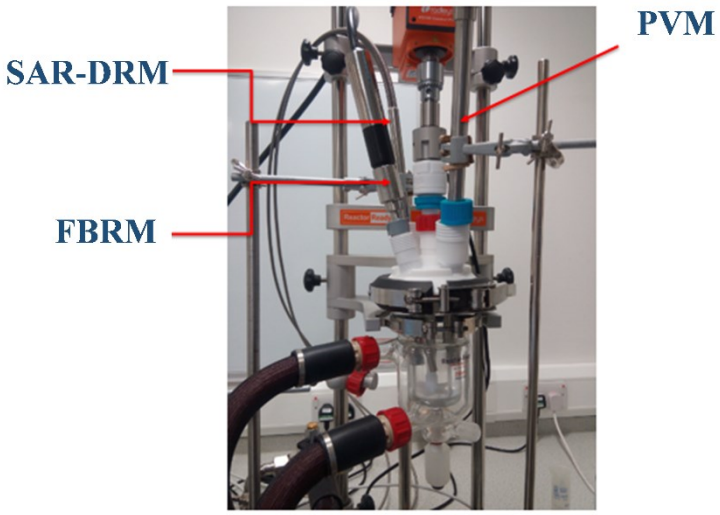


Figure 4.5: Experimental setup for measurements of polystyrene suspensions.

4.3.2. Experimental set-up for α-lactose monohydrate

A similar set-up to the polystyrene was used to prepare α-lactose monohydrate suspensions in acetone. In this case, Mettler Toledo OptiMax™ workstation consisting of a 1 L stirred tank crystalliser equipped with an inline Hastelloy® Pt100 temperature sensor was used for all experiments with the system controlled by the *iControl* v5.4 software. A pitch-blade impeller made of glass was used. FBRM, PVM and SAR-DRM were integrated within the vessel, as shown in Figure 4.6. The experiments were conducted by keeping the suspension temperature at 20 °C and the agitator set at 700 rpm. Similarly, to the polystyrene study, the effect of stirring speed and the probe position on the measurement were evaluated to guarantee minimum interference to the optical measurements. The different solid loadings were achieved through a gravimetric analysis and then added to the reactor through a funnel. The interference of ambient light in SAR-DRM measurements was avoid by covering the reactor with dark plastic and cloth. A total of 125 samples with a solid loading ranging from 0.5 to 25 wt.% and a particle size range from <38 to 400 µm, were analysed using this set-up. Table A.1 in Appendix A, summarises the range of conditions tested for each particle size analysed.

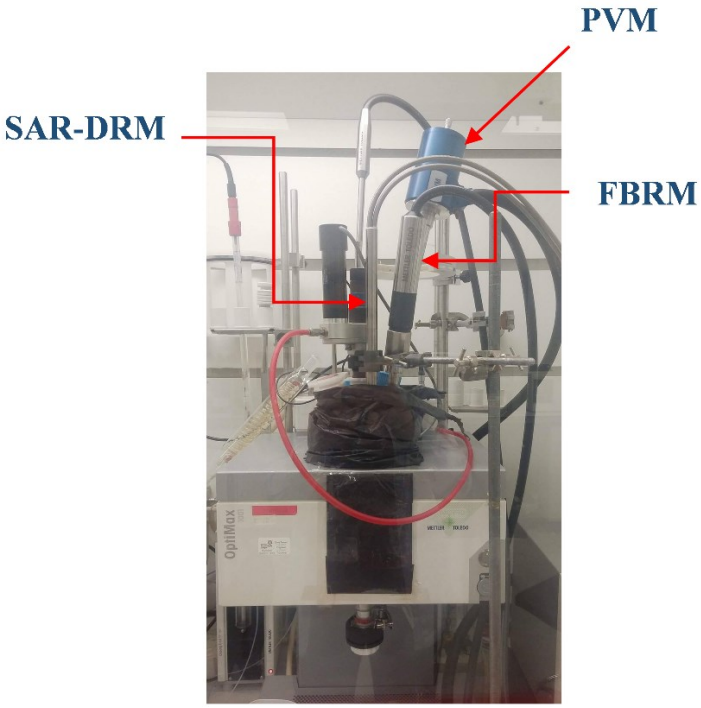


Figure 4.6: Experimental setup for measurements of α-lactose monohydrate suspensions.

4.4. Description of dataset

Polystyrene datasets

Polystyrene samples are the system studied in Chapters 6 and 7. In Chapter 6, the particulate system is used to compare the performance among different off-line and in-line commercial sizing techniques and to qualitative evaluate SAR-DRM response. In total, 30 polystyrene samples were evaluated i.e., it was used the data acquired from five different concentrations (1, 2.5, 5, 7.5 and 10 wt.%) of six different particle size ranges (PS0-90, PS125-180, PS180-250, PS250-355, PS300-500, PS630-800). All the bin sizes/spectral points offered by the different techniques studied are included in the analysis.

In Chapter 7, the polystyrene samples are used to build multivariate regression models. In this study, 25 samples were used as a calibration set. The PS300-500 sample was not included in the in the model as the measurements were acquired from the damaged SAR-DRM probe. To ensure the model is free from this known interference, it was decided to use only data obtained from the new probe to build calibration models. No test set was used. Although the polystyrene dataset is relatively small, which can lead to larger uncertainties, the aim of this work is to validate the viability of exploring further SAR-DRM and whether there is an improvement of the models when fusing data from other techniques.

In section 4.2, it was observed the number of bin size/spectral points used for the CLD, vis-NIR and NIR varies considerably. The vis-NIR data has about 5 times more data points than the NIR data, and about 30 times more data points than the CLD data. Thus, the vis-NIR information could have a much higher contribution when performing the regression analysis. To make the contribution of each technique more even, a reduction or interpolation of the dataset was performed, according to type of data:

- The vis-NIR data is organized into four blocks, where each of them corresponds to a different detector distance from the light source, being the first block related to D1 and the last one to D4. The range of wavelength used was 500-900 nm with a ~0.2 nm step size giving rise to 2112 spectral

points. The vis-NIR data was reduced from 2112 to 352 spectral points by just keeping the first spectral point of a sequence of 6 spectral points.

- The NIR data is also organised into four blocks using the same sequence of detectors. The wavelength evaluated range was 1000-1850 with a ~2.5 nm step that originated 352 spectral points.
- The FBRM data is organised in two blocks containing the unweighted and square weighted CLDs respectively. The 100 CLD points obtained by the FBRM software are interpolated to give rise to 357 data points.

In all cases only the first 350 data points were used to build the models. This matrix reduction or enlargement had no significant effect on the model and only eliminates the numbers of variables on the PLS regression model. This data treatment was performed so each block would have the same weight on the PLS model.

α-lactose monohydrate dataset

α-lactose monohydrate crystals are used in Chapter 8 to evaluate the performance of calibration and prediction multivariate regression models. In this system, a more diverse set of solid concentrations (0.5 to 25 wt.%) and particle size ranges (<38 to 400 μm) were evaluated to ensure a sufficient number of training samples to build more robust models and to average out most of the random error associated with the reference technique. A test set for external validation was used to improve the model performance and stability. In total, 120 samples were evaluated. The dataset was split into 80 samples for the calibration set and 40 samples for the test set. The 40 samples in the test set were randomly selected. Table A**.1** displays the conditions evaluated in the calibration and in the test datasets.

Similarly to the polystyrene data set, a dataset reduction/interpolation was performed:

- The vis-NIR data is organized into four blocks corresponding to the increase of the detector-source spatial distance, from D1 to D4. However, as the interference from FBRM could not be fully removed by the spectra normalisation (Equation 2.1), after the vis-NIR data reduction to 2112

spectral points, the variables corresponding to the wavelength region from 760-800 nm were eliminated. This resulted in a reduction from 2112 to 1894 spectral points. By just keeping the first spectral point of a sequence of 5 spectral points, the data is reduced to 379 spectral points. The final 357 vis-NIR spectral data points are achieved by removing the first 11 and the last 12 data points from the dataset.

- The FBRM data is again organised in two blocks containing the unweighted and square weighted CLDs respectively. The 100 CLD points obtained by the FBRM software are interpolated to give rise to 357 data points.

The calibration models build on the individual techniques/detectors and on different measurement combinations, only use the first 350 data points of each block.

4.5. Multivariate regression analysis

Partial least square regression (PLSR) analysis was performed on SAR-DRM and FBRM measurements for building calibration models on polystyrene and α-lactose monohydrate systems for estimating particle size and concentration. The D50 obtained from laser diffraction measurements was used as the reference particle size for both systems studied. For particle concentration, the concentration from gravimetric measurements used.

Figure 4.7 displays the work flow for the PLSR analysis using the lactose dataset as an example. The split of the datasets was performed prior any data pre-processing steps. Hence, the pre-processing parameters determined by the calibration dataset were applied to the test set. Different pre-processing methods were applied to assess whether they lead to an improved performance of the models.

The datasets were pre-processed using the transformation of reflectance (R) to -ln(1/R), five light scattering and baseline correction methods, SNV, MSC, EMSCL, EMSCW and ISC, smoothing and mean centring. A combination of the log transformed data with the empirical correction models (one at a time), was evaluated.

A Savitzky–Golay filter with a moving smoothing window of 15 points and second order polynomial to the polystyrene vis-NIR and NIR spectral data. No smoothing was required for the α-lactose monohydrate data.

All the models were built using leave-one-out cross-validation in all the datasets. The optimal number of latent variables (LVs) to be used for modelling was chosen by examining the curve of root mean square error of cross-validation (RMSECV) plotted versus the number of LVs. A maximum of 10 LVs was evaluated. In case of a RMSECV relatively flat profile with an appreciable drop in value in the first few LVs, the number of LVs was chosen based on the clear minimum value observed. In these situations, the use of higher LVs gives negligible reduction in RMSECV. When the RMSECV curves decrease monotonically with an apparent flat profile a model using no less than five LV was chosen. The minimum of five LVs is established by the number of varying components in system (particle size, particle size distribution, concentration of particles, shape, and suspension medium). Yet, a visual inspection of the respective LV loading curves was performed to avoid the over fitting of the model. A number of LVs above 5 was selected when the RSMECV curve systematically decreased and after examining the loadings. For example, if the RMESCV indicates 10 LVs are needed to explain the model but the respective loading curves are noisy, the LV number 9 is selected (i.e., if the LV number 9 shows a methodical behaviour).

Comparison of the calibration models for each input data set was chosen based on the number of LVs used and on the lowest RMSECV. In the case of the lactose system, after selecting the best calibration model for each input data, the models comparison was based on both RMESCV and root mean square error of prediction (RMSEP). The latter error is obtained by using the test samples.

The Matlab code used has been written in-house and it was used in previous studies [105, 134-136].

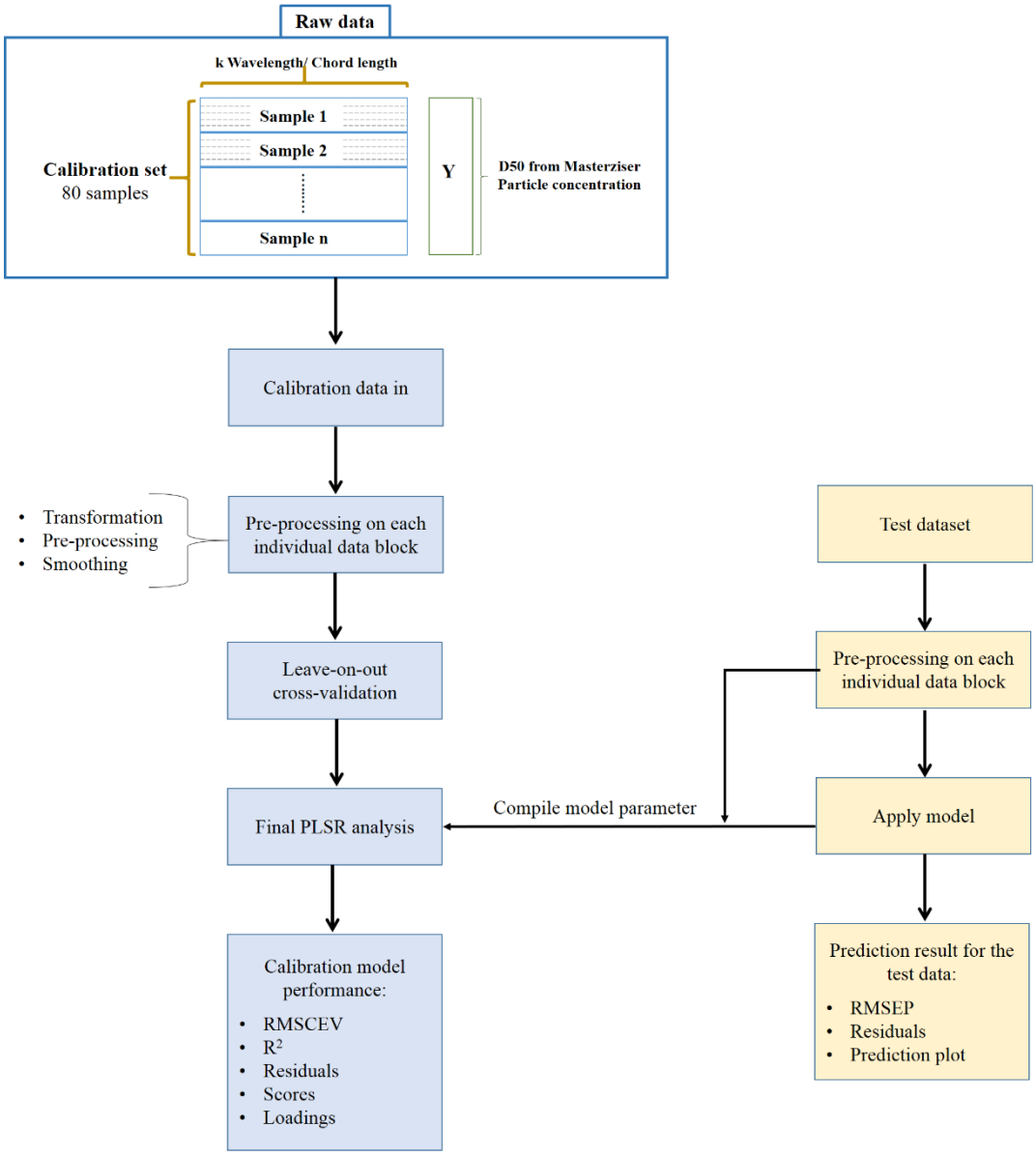


Figure 4.7: Work flow for PLSR analysis.

4.6. Data fusion approaches

Multivariate regression analysis was initially performed on the data collected from each sensor (FBRM and SAR-DRM) separately and on the spectra collected by each single detecting fibres individually. In order to assess the influence of incorporating additional information in the data, and how the PLSR analysis utilises the additional information, two different data handling strategies were evaluated: a) co-addition of

spectra and b) data augmentation. Figure 4.8 illustrates the referred strategies adopted for fusing the data by decomposing a 3D version of the matrix X, in which n represents the number of samples and k represents the wavelength/CLD modes. The same approaches are used in [122] to combine spectral data from different source-detector distances. In the first strategy, the spectra from multiple source-detector distances or/and CLD are combined to form a single spectrum for the sample. This may produce a spectrum with higher signal-to-noise ratio. Nevertheless, the averaging of the spectra is not expected to make a difference in the following regression analysis. The second strategy relies on concatenating the different data blocks into an augmented matrix, i.e., the spectra collected from the different distances will form a large data matrix where the total number of variables equals the sum of the numbers of variables in each individual data blocks. In this approach, it is essential the weight of each one of the different blocks to be concatenated to have equal importance in the model in order to obtain an equal impact in the PLS regression model performance. When combining the data from multiple detectors, sources or instruments through this approach, the pre-processing was performed on the individual data block and the resultant pre-processed data blocks of each detector were augmented to form a final data block.

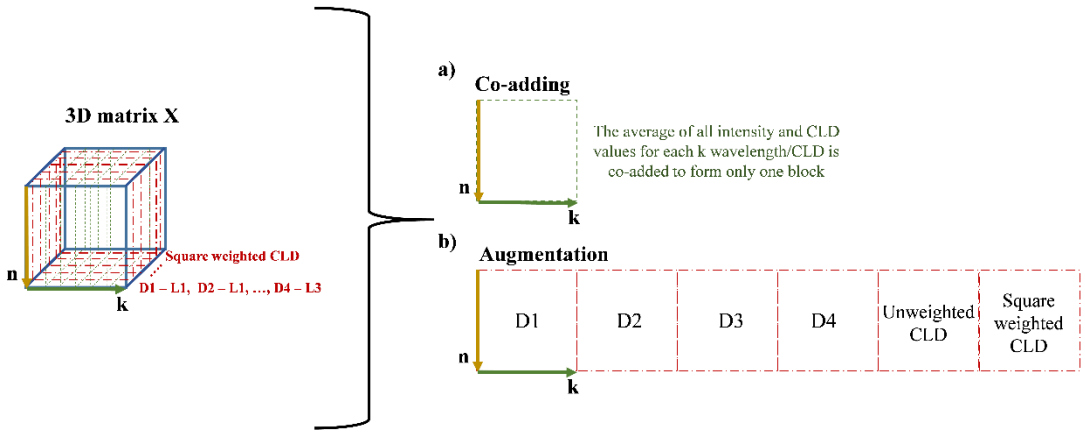


Figure 4.8: a) Co-adding and **b)** augmentation data approaches of matrix X displayed in a 3D system.

4.7. Analysis of SAR-DRM signal quality

As SAR-DRM is, for the first time, applied to study micron size particles, the setup and signal quality need to be examined to determine whether reliable measurements can be obtained. This section explores whether SAR-DRM is able to capture the spectral features of polystyrene beads, its sensitiveness to the other optical probes used, namely, PVM and FBRM, the best set-up conditions to acquire the signal and whether SAR-DRM results are consistent, repeatable and trustable. PS suspensions of different solid loadings are chosen to produce different turbidity conditions to test SAR-DRM in a range of optical conditions. The analysis carried out includes optical interference from other optical probe, signal reproducibility and consistency.

4.7.1. Signal interference from the optical probes

The signal interference of FBRM and PVM in SAR-DRM was evaluated by using the set-up described in section 4.3.1. Figure 4.9(a) exhibits a peak at 785 nm in the SAR-DRM spectra from the sample of 1 wt.% PS0-90. The signal was collected from fibre configuration of smallest spatial distances, which represents the shortest optical path length. Figure 4.9(b) shows the SAR-DRM spectra when no light was emitted from the probe. It was expected that no signal would be detected, except for the background noise. However, a peak at the same wavelength, 785 nm, was detected, suggesting SAR-DRM receives light from other sources. The same peak is observed for all other configurations, and also appears when analysing samples of different particles sizes and solid concentrations. This interference is found to be due to FBRM emitting laser light at 785 nm for the chord length measurement. This interference can be overcome or minimized when subtracting the known interference (Sample light off) from the sample spectra (Sample light on), as shown in Figure 4.9(c). No interference from PVM was detected in SAR-DRM signal in any of the experiments.

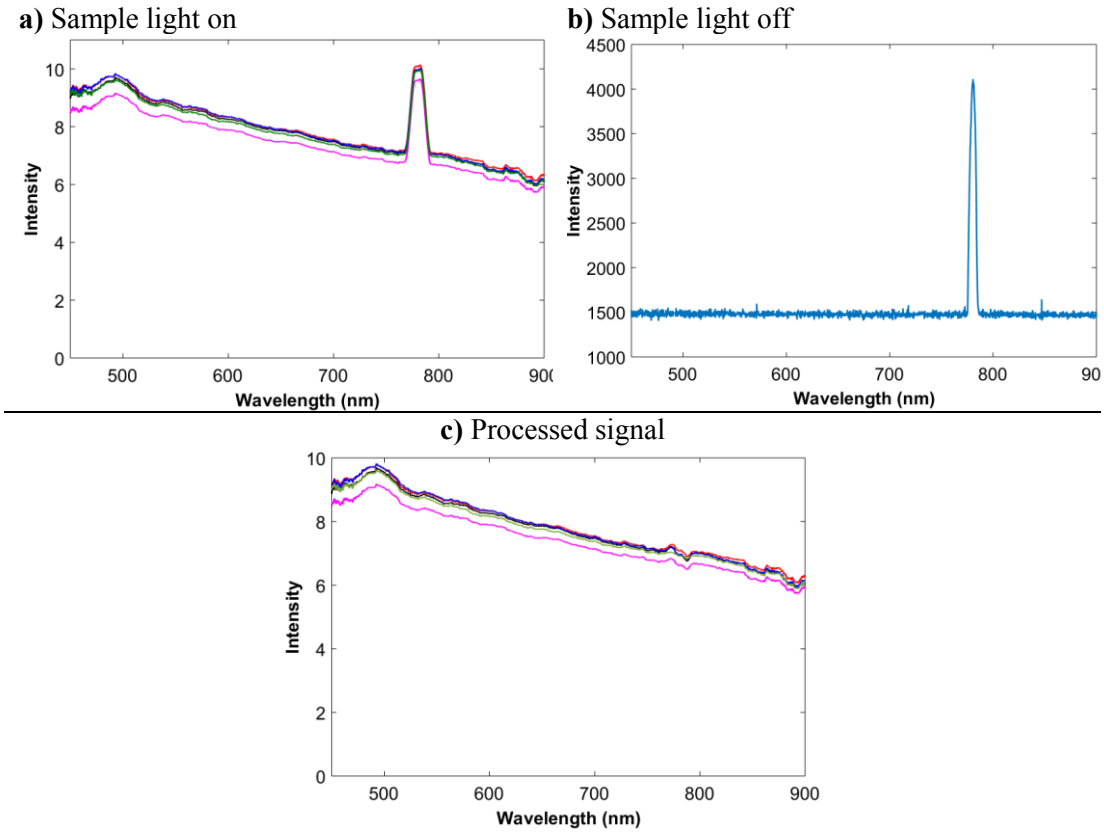


Figure 4.9: FBRM interference detected when SAR-DRM is **a)** analysing 1 wt.% of PS0-90 with the light on, **b)** with the light off and **c)** after subtraction (processed signal).

4.7.2. Stirring speed effect & Repeatability of the method

To determine the optimal stirring speed to achieve uniformity of the suspension for reliable measurements from SAR-DRM, five spectra repetitions at a stirring speed of 200, 300, 400 and 500 rpm were taken. Each stirring speed was studied using 1 and 10 wt.% of PS0-90, PS90-300 and PS630-800 particles. An example of the obtained results is shown in Figure 4.10 for 1 and 10 wt.% of PS0-90 evaluated at 100, 300 and 500 rpm. The spectra displayed were collected from configuration of the smallest spatial distance to the normal incident light source. As shown in Figure 4.10(a1) and (b1), the magnitude of the signal varies around 30 % when using a stirring speed of 100 rpm regardless the solid loading. However, the analysis of lowers solid concentrations demonstrates higher sensitivity of SAR-DRM to FBRM interference. An increase in the stirring speed to 500 rpm leads to a decrease in the signal variation among the repetitions to 1%. This indicates the repeatability of SAR-DRM when a

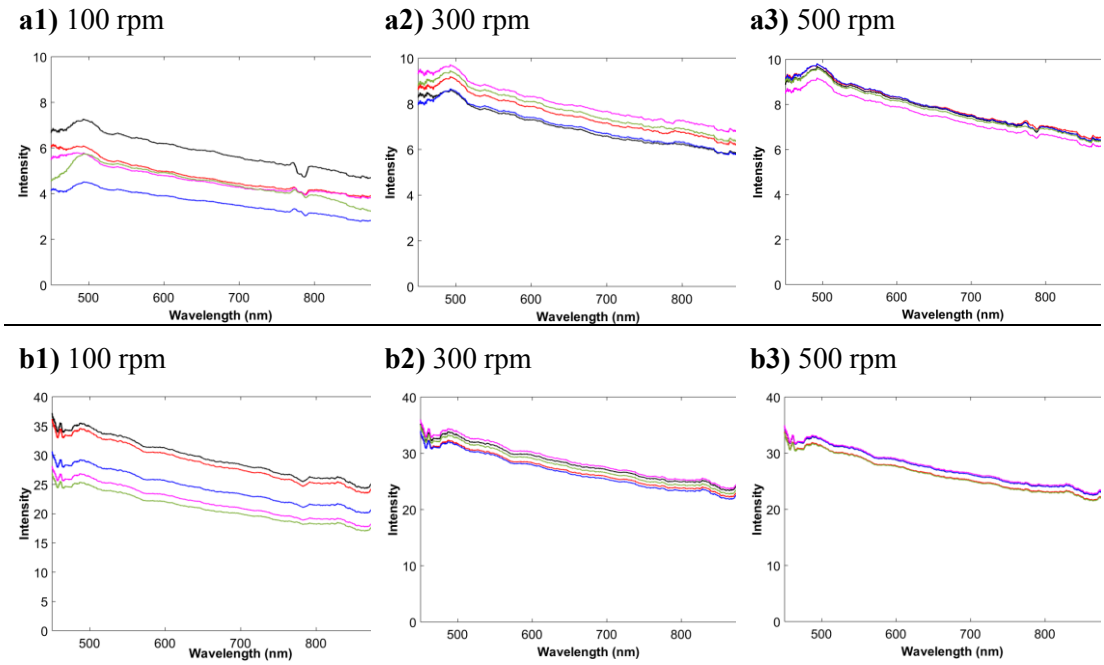


Figure 4.10: Five replicates of SAR-DRM spectra using **a)** 1 and **b)** 10 wt. % of PS0-90 at different stirring speeds **(1)** 100, **(2)** 300 and **(3)** 500 rpm. All spectra were collected from configuration of the smallest spatial distance to the normal incident light source.

proper mixing of the samples is achieved. However, the optimal mixing condition depends on the system itself. In the case of PS suspensions, a stirring speed above 500 rpm would be suitable to result in a uniformly suspended particle mixture.

Similar examination on the SAR-DRM spectral variation was applied to other source-detector fibre arrangement. Figure 4.11 shows the source-to-detector distance of the five replicates taken for all the available fibre arrangements. In general, the angular incident light resulted in larger spectral variation. The largest spectral variation of the replicates, around 7%, was observed when using the incident light at 30° and the smallest source-detector distance. In Figure 4.11 it is also observed the use of angular incident light resulted in higher intensity values. When using an angular incident light travels shallower into the sample hence the returned reflectance is higher than those from the normal incident light as illustrated in Figure 4.2(b).

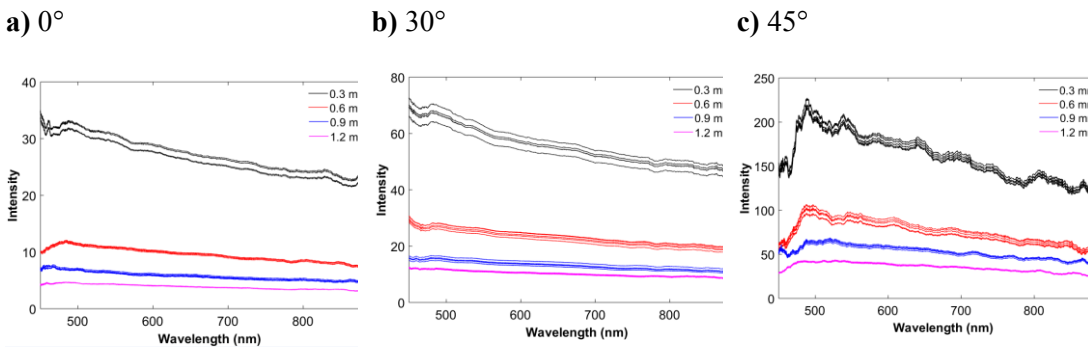


Figure 4.11: Source-to-detector distances and its respective replicates when light is being emitted at a) 0°, b) 30° and c) 45° in a 10 wt.% suspension of PS0-90.

An increase of the particle size leads to a larger spectral variation of SAR-DRM spectra. Although the same concentration was used (10 wt.%), fewer number of particles are in the medium, which results in reduced scattering event and lower intensity as shown in Figure 4.12. This means that SAR-DRM signal more likely to be affected by small number of particles, leading to the larger variation observed. This condition highlights that the higher turbidity of the medium the better the SAR-DRM performance.

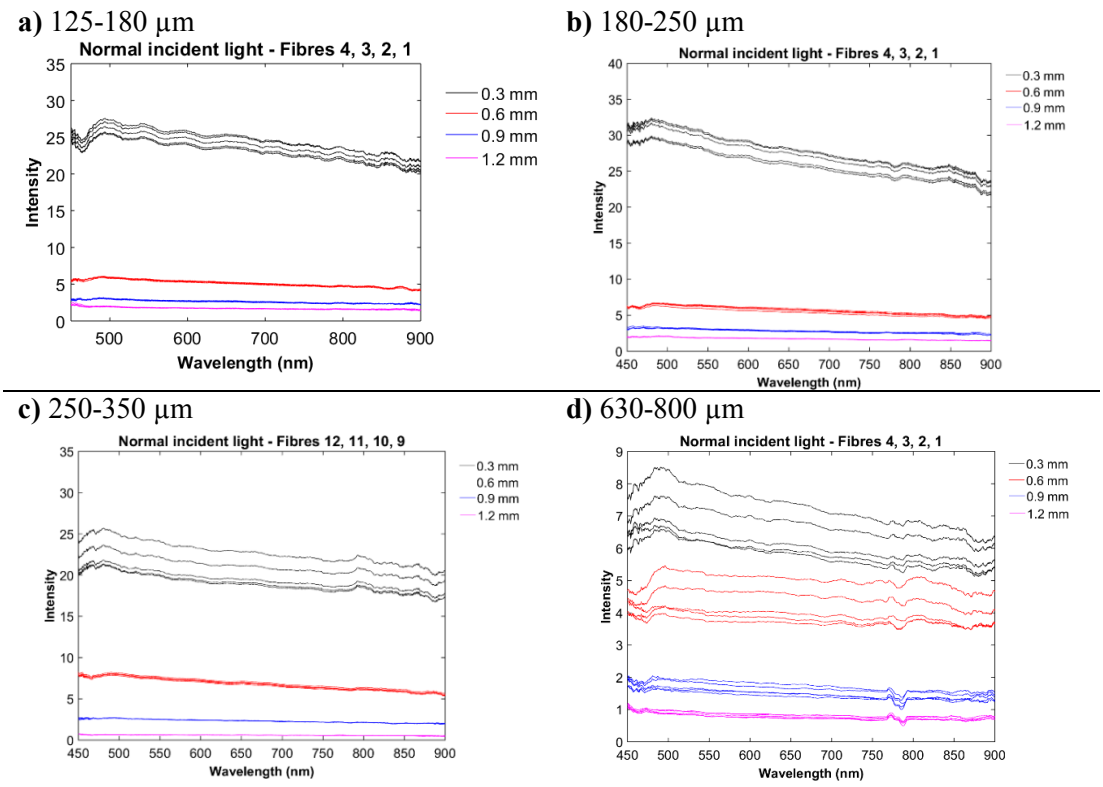


Figure 4.12: SAR-DRM spectral variation for all source-to-detector distances using a normal incident light for 10 wt.% suspension of **a)** PS0-90, **b)** PS180-250, **c)** PS250-350 and **d)** PS630-800.

4.7.3. Confidence check

As described in section 4.2.2., the SAR-DRM probe design contains four sets of fibers located at equivalent distance from the normal incident light, two sets for the light emitted at 30° and two more sets when the angle of incident light is 45°, see Figure 4.2(a). The identical fibre arrangement offered by the probe design allows the check to be made on the consistency of the measurements, or to be used to examine sample uniformity.

An example of the consistency is illustrated in Figure 4.13, using the closest fibres to the normal and angular light sources to analyse 10 wt.% of PS125-180. Five repetitions for each configuration were taken. Fibres 4, 8, 12 and 16 are all located at 0.3 mm of the normal incident light, and Figure 4.13(a) shows that all fibres produce a signal with similar magnitude and variation in the repetition. However, when using the angular

incident source, a gap in intensity is observed in Figure 4.13(b) and (c). Yet, all repetitions are quite consistent in both fibres. This is found to be due to the damage in one of the angular incident fibre, which leads to the samples spectra normalisation (Equation 2.1) to incorrectly account the total intensity from the integrating sphere. Less amount of light being emitted to the system will consequently result in less amount of light being reflected. As such, all the incident light being emitted to the integrating sphere will be collected by SAR-DRM, regardless if one light source is working at 50% or 100 % of their capability; the light collected will always be the overall of the light being emitted to the system. This effect is then translated in a drop of the samples intensity signal, after the signal normalisation.

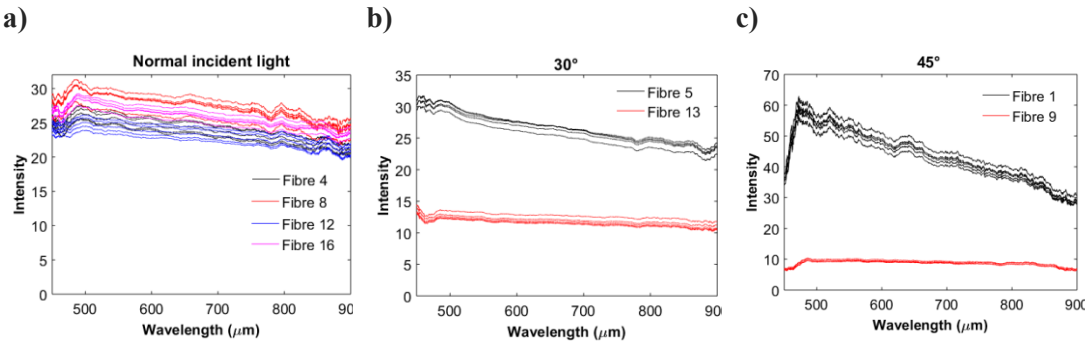


Figure 4.13: Confidence check of the fibres located at 0.3 mm from the incident light at a) 0 °, b) 30 ° and c) 45°.

This issue is confirmed by plotting the intensity versus the source-to-detector distance. Figure 4.14(a) demonstrates that the first set of fibres analysed show the best performance collecting data when compared to the equivalent fibres, exhibited in Figure 4.14(b), located at same distance from the light source at 45°. This trend is shown in all measurements regardless the sample condition. The analysis suggests the damage affects two branches of fibres (Fibres 9-12 and 13-16), while the other two branches (Fibres 1-4 and 5-8), perform most consistently.

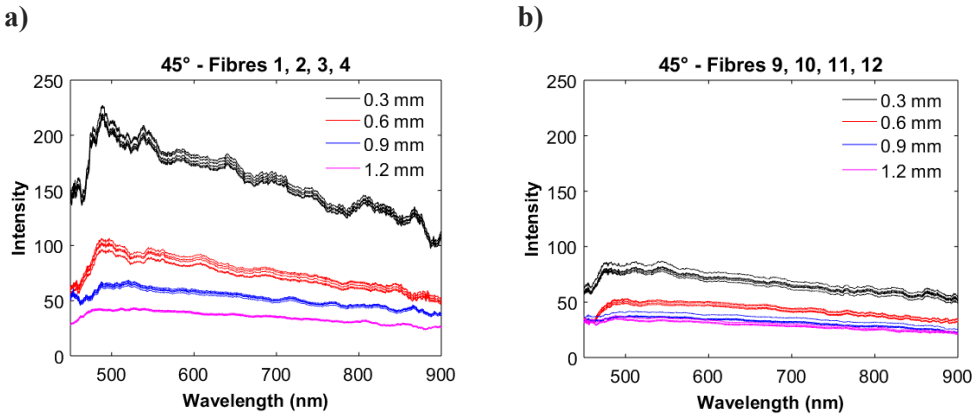


Figure 4.14: Confidence check of the fibres located at different distances from the angular incident light source at 45° for the branch **a)** and **b)**. Results were obtained from 10 wt.% of PS0-90.

Due to this issue, a replacement SAR-DRM probe was ordered. Figure 4.15 exhibits the spectra obtained from a suspension of 10 wt.% of PS125-180 when looking at the fibres at 0.3 and 0.6 mm from the respective light source, similarly to Figure 4.13. With the new optical probe, the signal collected from equidistant fibres was similar and its consistency was improved, as expected, when increasing the collecting fibre distance from the light source. This comparison was performed for each sample analysed. All of the analyses have shown no further damage of the probe. As the comparison suggests that the branch of the optical probe containing the fibres 1, 2, 3, 4 and the branch containing the fibres 5, 6, 7, 8 perform most consistently, since a consistent and valid signal was obtained in the different conditions tested in this work, this set of fibres have been selected for further studies. The impact source-detector distance as well as the effect of particle size and concentration on SAR-DRM spectra is studied in the next chapter.

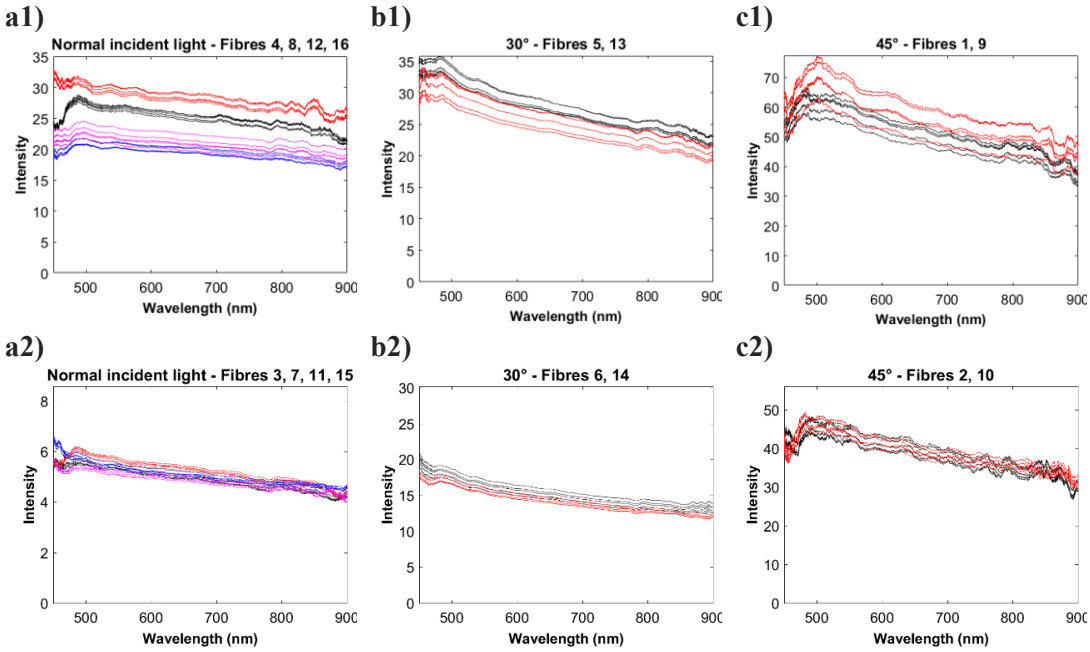


Figure 4.15: Confidence check of the fibres located at 0.3 mm from the incident light at **a)** 0°, **b)** 30° and **c)** 45° of the new SAR-DRM probe.

Chapter 5 - Multi-sensor in situ Measurements of Particle Size, Shape and Concentration - Perspectives of Spatially and Angularly Resolved Diffuse Reflectance Measurement

The previous chapter has shown the materials, experiments and methodologies used to evaluate different off-line and in-line technologies. It also presented the tests performed to establish a reliable experimental setup to collect and process data in-line, in an accurate and repeatable way.

This chapter reports how to combine and compare information from multiple off-line and in-line particle size analysis technologies and demonstrates the potentiality of SAR-DRM to measure particle size and solid concentration. Polystyrene beads suspensions in water are used as model system.

The overall results show SAR-DRM is capable of handling high solid loadings (up to 10 wt.%) and relatively large particle sizes, from <90 to 800 μm, and is sensitivity to both particle size and solid concentrations, performing better than FBRM and PVM for such suspensions.

5.1. Introduction

Real-time monitoring of particulates presents a great challenge for many pharmaceutical, chemical and biological processes. Most particle measurement techniques are developed for off-line analysis while in-line analysis has only become available in recent years. However, all these techniques show limitations when applied in practice, as discussed in more detail in Chapter 2. This study investigates the performance of state-of-art in-line particle measurement techniques, namely, Focused Beam Reflectance Measurements (FBRM) and Particle Vision Microscope (PVM), and Spatially and Angularly-Resolved Diffuse Reflectance Measurement (SAR-DRM). It studies the information contained in both in-line and off-line measurements and the response to a wide range of variations in particle loadings and size. This work represents the first time SAR-DRM is applied to analyse particles in the micron size range. Using polystyrene particle suspensions as a model system, the in-line measurements are performed simultaneously. Mathematical algorithms are applied to FBRM and PVM measurements to extract particle size and aspect ratio distributions and compare these with those obtained from off-line technologies (i.e. laser diffraction and imaging). The performance of SAR-DRM is investigated on its sensitivity, accuracy and capability to track the differences in particle size, shape and concentration. The analysis identifies key challenges in FBRM and PVM analysis and quantitates the upper limits of particle number density for reliable PVM measurement. The capability of SAR-DRM to measure suspensions of high solid loadings and relatively large particle sizes is demonstrated. The results suggest the benefit of utilising the underlying physics from multiple in-line sensors for processes involving significant changes in particle size, shape and concentration.

5.2. Sample characterisation

In the first section of this chapter, polystyrene beads of different particle sizes were characterised using both off-line (Mastersizer, Morphologi G3) and in-line (FBRM, PVM) commercial sizing technologies. The results obtained served as reference to evaluate the SAR-DRM response to the model system. In the second section, it is

evaluated the sensitivity, accuracy and capability of the novel technique to track the differences in size and concentration as observed by the standard methods, and whether SAR-DRM can be a potential complementary tool for in-line analysis.

In this study, all PSDs are converted into probability density functions (PDF) to allow a direct comparison for all measurement results. As show in Table 5.1, each technique offers not only different size measurement ranges but also different number of bins for the PSDs. As mentioned in Chapter 2, the PDF makes the distribution grid-independent and stipulates the probability of a random variable falling within a particular range of values which can be calculated from the area under the curve.

Table 5.1: Particle size range and number of bins for the different techniques used in this work.

Technique	PSD representation	Size range (µm)	Number of bins
Mastersizer	ESD	0.01 – 3283.58	100
Morphologi G3	ECD	0.1 – 2000	1001
FBRM (inverted)	ECD	1.02 – 2940.55	200
PVM	ECD	1 – 1000	100

5.2.1. Sample Characterisation – Particle size effect

5.2.1.1. Off-line analysis

The particle size and shape of the polystyrene beads used in this study were first characterised using a laser diffraction and imaging devices as off-line particle sizing technologies.

Laser diffraction

Figure 5.1 shows the PDF of the volume weighted distribution obtained from the LD measurement for the sieved fractions of polystyrene. The respective volume distributions acquired as raw data are shown in Figure B**.1**. The mode of each curve

displayed good agreement with the sieved size range, resulting in distinctive PSD curves for each size fraction. In the samples PS0-90 and PS180-250, the distribution was skewed towards smaller sizes, which can be attributed to the high number of ellipsoidal particles present in the suspension. Figure 5.2 exhibits an example of the images acquired during the LD analysis. In general, a significant amount of ellipsoidal particles mixed up with spheres in the slurry was detected in the PS beads samples of smaller size ranges (Figure 5.2(a), (b), (c) and (d)). During the sieving process, ellipsoidal particles were probably mostly oriented in the vertical direction, since particles with a width less or equal to the respective mesh diameter will pass through. One of LD limitations is that it assumes all particles are spherical. Particle shape influence the scattering pattern of the light. Non spherical particles will diffract the laser beam depending on the average orientation of the particles in the beam. Thus, in the case of the ellipsoidal particles, the probability of diffracting a beam based on the particles width is much higher than based on the particles length, resulting in the skewed curve towards smaller sizes. Only the samples PS300-500 and PS630-800 seemed to be fully represented by spheres.

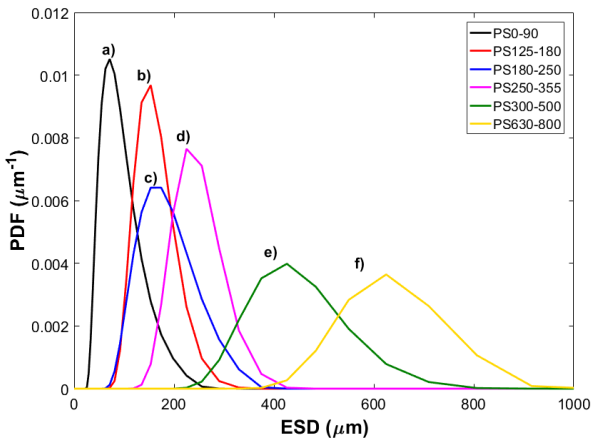


Figure 5.1: Equivalent Sphere Diameter (ESD) obtained from LD analysis of the PS beads of various size ranges.

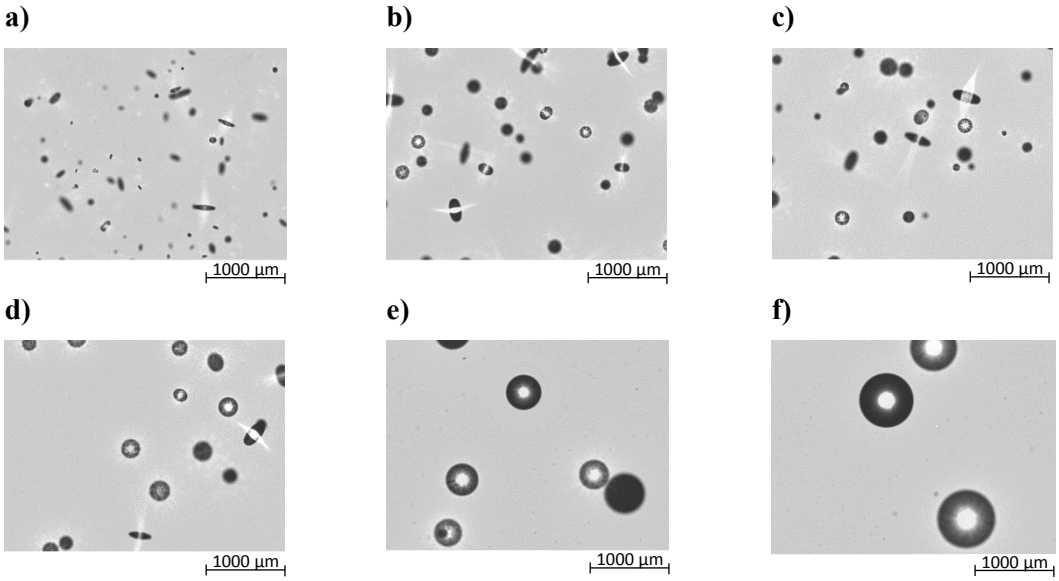


Figure 5.2: Hydro Sight images of **a)** PS0-90, **b)** PS125-180, **c)** PS180-250, **d)** PS250-355, **e)** PS300-500 and **f)** PS630-800.

Imaging

Figure 5.3 shows the results obtained from the Morphologi G3 measurements of the different particle sizes, displayed on a number and volume weighted distribution. Due to the effect of gravity, either the spheres or ellipsoidal particles present in the samples will lie on the measurement plate. As the dispersed particles are perpendicular to the optic axis of the lens during the analysis, it is assumed the Morphologi G3 can provide a relative accurate size. In both distributions (Figure 5.3(a) and (b)), an increase in the beads size leads to an increase in the ECD, although this trend is not so clear when analysing the samples PS125-180 and PS180-250. Moreover, differently from the other samples, PS125-180 and PS180-250 beads exhibit a distinctive bimodal number weighted distribution. This is caused by the presence of large amount of fine particles in the system and by the presence of particles of different shape, spheres and ellipsoids. The difference in shape can be quantified computing the cumulative frequency of the aspect ratio of each sample, as shown in Figure 5.4. It is observed that around 70 and 60% of the samples PS0-90 and PS125-180, respectively, are composed by ellipsoidal particles, which will have a significant impact on the size distribution. The increase in

particle size range shows a decrease in the amount of ellipsoidal particles. The larger samples, PS300-500 and PS630-800, are mainly composed by spherical particles.

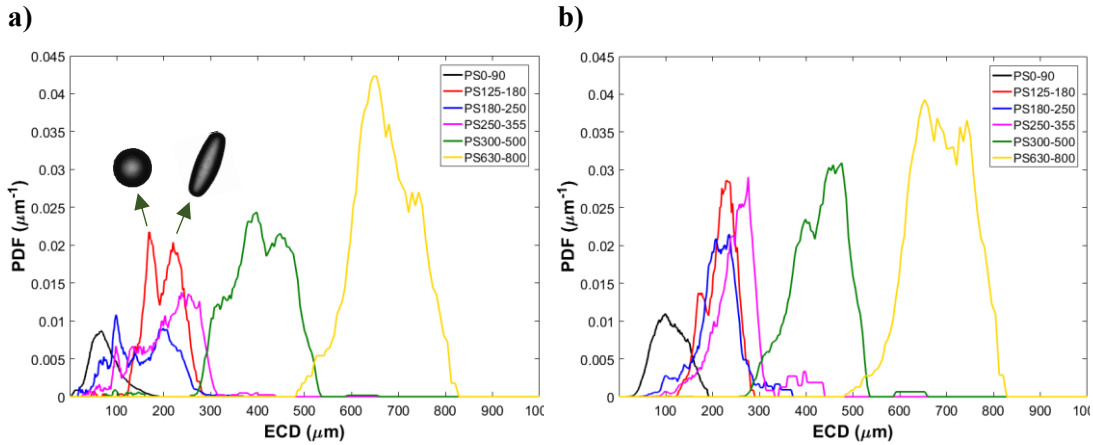


Figure 5.3: Equivalent Circle Diameter (ECD) obtained from Morphologi G3 expressed in **a)** number and **b)** volume weighted distribution.

The number weighted distribution states that each particle analysed has the same contribution to the distribution, i.e. each particle has equal weighting regardless their size, hence, the considerable amount fine particles present in the sample is affecting the distribution. The Morphologi G3 software allows the identification of the particles shape measured at each peak, due to a digital filter that selects the particles of interest regarding to their size. In the case of PS125-180, for example, the software shows that the first peak is mainly related to the spheres and the second to the ellipsoidal particles.

The volume weighted distribution makes the contribution of each particle proportional to its volume and typically resembles a bell-shaped curve. Figure 5.3(b) shows that the conversion of number weighted into volume weighted distribution results in a shift of the ECD towards larger sizes and, in case of PS125-180 and PS180-250, in a single curve. The sensitivity to fines is reduced and the contribution of large particles is dominating the distribution as their volume is much larger than the small particles.

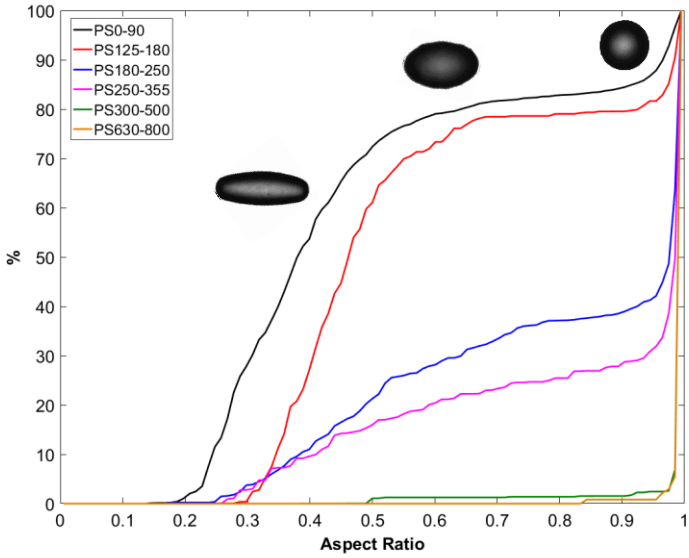


Figure 5.4: Cumulative frequency showing the aspect ratio of each sample analysed by Morphologi G3.

5.2.1.2. In-line analysis

Before comparing the effect of particle size between the different off-line and in-line technologies, we show typical data obtained from the in-line tools, FBRM and PVM. Although the effect of particle size has been studied for all the five solid loadings, the results shown below are for 10 wt.% solid loading as an example of the information that can be collected from both techniques.

FBRM & PVM

Figure 5.5 presents both unweighted and square weighted CLDs obtained from each particle size range tested. The unweighted CLD for PS0-90 shows a broad band with two modes due to the presence of particles with different sizes and shapes in the suspension, which are mostly ellipsoids as it was observed in Figure 5.4. The mode at ~30 µm can be attributed to the laser beam being backscattered predominantly by the ellipsoidal width. The second mode is observed at around 70 µm and it is likely a result from the ellipsoidal particles length and from the spheres diameter. Due to the PS0-90

small particle size, 10 wt.% solid loading signifies a much larger number of particles in the suspension than in the other samples. This is emphasised by the high number of counts detected by FBRM when comparing PS0-90 with samples with larger particle sizes. The number of counts tends to decrease with the increase of particle size (Figure 5.5(a). In both unweighted and square weighted CLDs, it is difficult to observe a clear trend between the different particle sizes, except for the PS0-90. PS300-500 and PS630-800 showed two broad bands in the unweighted CLD which do not correspond to the true particle size. The main peak at 100 µm suggests the particles are outside the expected particle size range of the corresponding sieved fraction. However, PVM pictures acquired during the experiment show that PS300-500 and PS630-800 particles are much larger than the FBRM signal indicates, as observed in Figure 5.6(e) and Figure 5.6(f). As mentioned in Chapter 4, the measurements were performed using a reflector Teflon cap on PVM to obtain a transmission image rather than a backscattered image. PVM images acquired when using the six forward lasers instead of the two back lasers to avoid the sample reflection, reveal that these particles are transparent with a large and smooth surface as shown in Figure 5.7. Transparent particles have been reported to be poorly sized by FBRM in literature.[45, 137-139] These particles produce considerable specular reflection of the light emitted by FBRM or a high degree of internal scattering, which results in chords splitting.

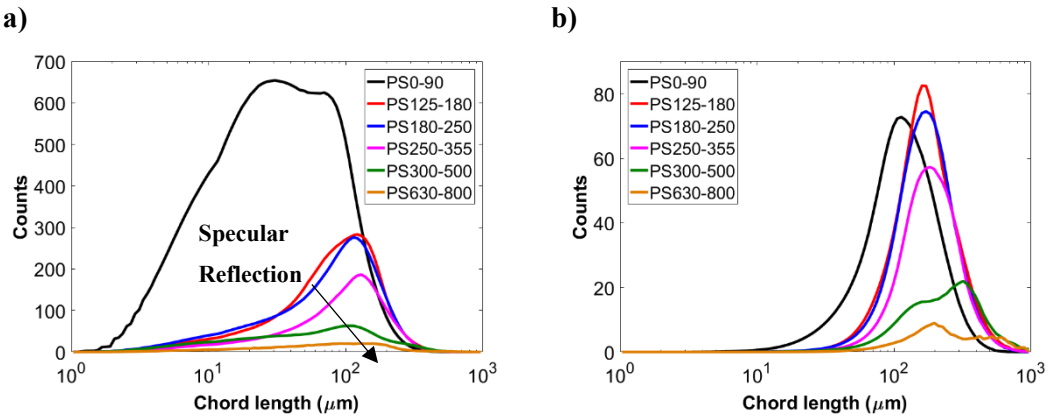


Figure 5.5: a) Unweighted and **b)** Square weighted CLD of 10wt.% of PS samples of various particle size.

However, it is still observed that some chord lengths can be related with the polystyrene size of 300-500 and 630-800 μm when looking at the square weighted CLD in Figure 5.5(b). This happens due to the ability of this representation to emphasise coarse particles. Nevertheless, the amount of chord splitting is so significant that it is difficult to observe a peak in those regions when looking at the FBRM raw signal.

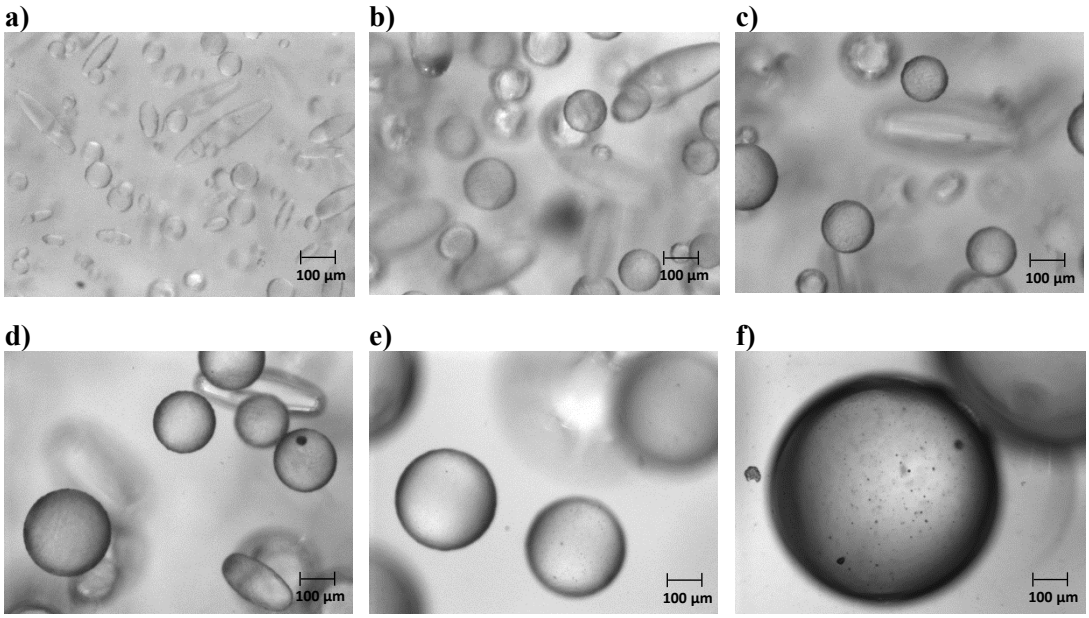


Figure 5.6: PVM images of polystyrene beads at 10 wt.% **a)** PS0-90, **b)** PS125-180, **c)** PS180-250, **d)** PS250-355, **e)** PS300-500 and **f)** PS630-800.

PVM images can be processed in order to estimate the PSD. Figure 5.8 shows the notorious difference between using the number and volume weighted distribution to describe the obtained ECD. In all cases, there is a shift towards larger equivalent diameters when converting the results into volume weighted distributions. This difference is even more significant when analysing large particles such as PS300-500. The PS300-500 peak shown in Figure 5.8(a) at smaller diameters is related with parts of particles that are out-of-focus, as explained in [12], and with small imperfections found on the particles surface (Figure 5.6(e)). Those imperfections, marked by dark spots, may have been caused by particles shocking with each other during the

polymerisation reaction. Images acquired from optical microscopy and SEM show these imperfections in more detail, as seen in Figure 5.9. Some particles show bumps at the surface, which could be due to small particles shocking with large ones during polymerisation reaction, followed by their attachment to the larger particle surface and some particles exhibit a cavity.

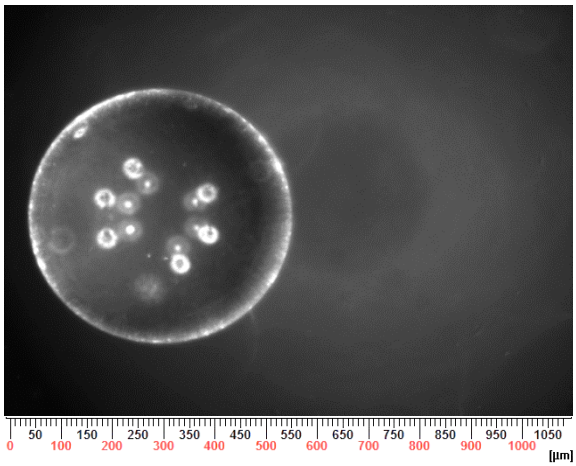


Figure 5.7: PVM image of a larger, transparent particle with a smooth surface.

When computing the volume weighted distribution (Figure 5.8(b)), a peak at around 350 µm is observed, which is in good agreement with the known particle size range. The results from the largest particle size, PS630-800, are not here exhibited since in order to obtain a representative number of particles, millions of pictures would need to be acquired. The probability of getting a particle with a size between 630-800 µm completely in focus without touching the picture frame is very low because the PVM probe has a field of view of 1075 µm × 825 µm.

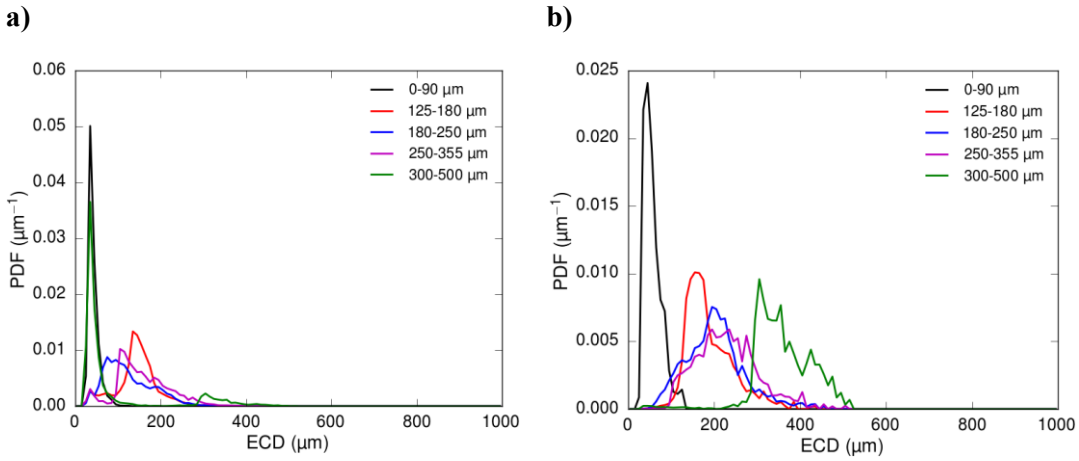


Figure 5.8: a) Number and **b)** volume-weighted ECD obtained from the PVM image processing algorithm when using 10 wt.% solid loading of the different PS samples. The sample PS630-800 is not shown here as it was not possible to capture a representative number of particles for a valid analysis.

Due to the different shape of particles present in the samples, an ECD distribution might be misleading. For a more accurate description of the material in analysis, it is always worth looking at the distributions describing the length and width of the particles. The length describes the longest dimension of a bounding box around the particle and consequently the distribution is shifted towards larger sizes. The width represents the shortest dimension of the bounding box. As during sieving the limiting dimension of ellipsoidal particles to pass through the sieve opening is their width, their width distribution is the most representative comparison with the sieved size range (Figure B.4). However, since this property is only accessible through imaging techniques, it was decided to perform the comparison between different techniques using the ECD.

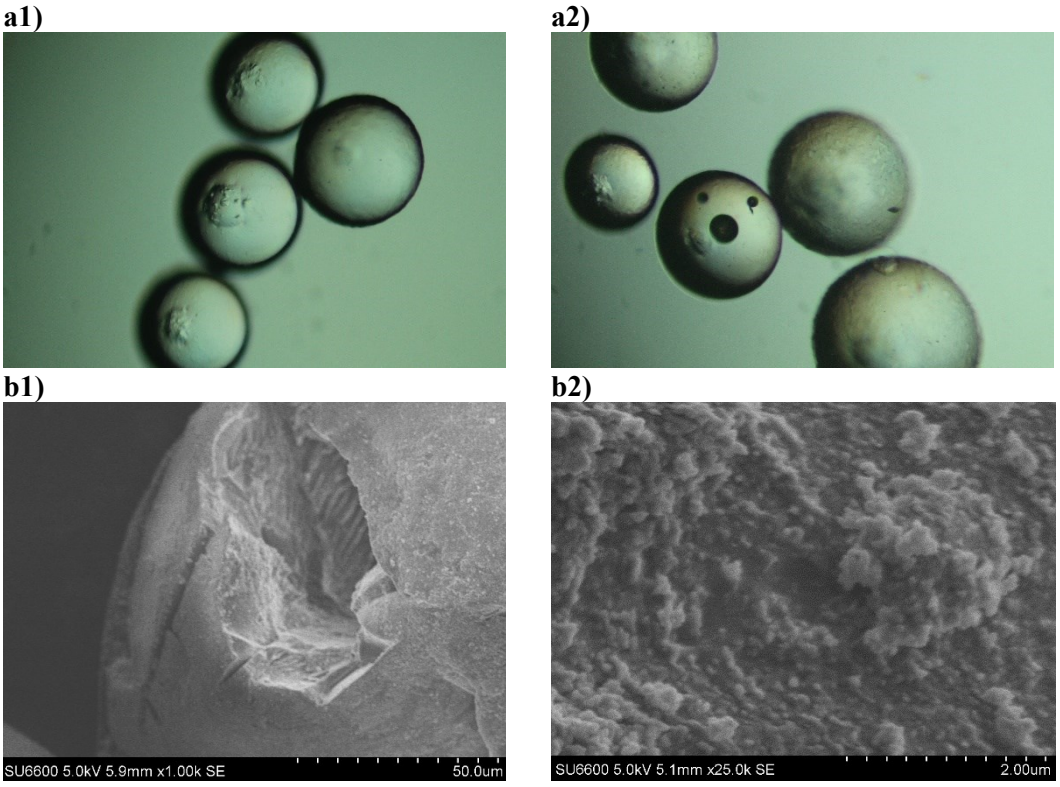


Figure 5.9: PS beads imperfections visualised in **a1), a2)** optical microscopy and **b1), b2)** SEM images.

5.2.1.3. Comparison of off-line & in-line PSD

Effect of particle size in off-line & in-line PSD

A comparison of PDFs calculated based on the volume-weighted ECD/ESD from all particle size measurement techniques is shown in Figure 5.10; for the in-line measurement techniques, the PFDs are estimated using measurements on 10 wt% particle suspensions. Some inconsistencies are observed between the PDFs and the defined sieving range (shaded size range), as well as between the PDF from different methods. To determine the cause of the discrepancies, the unprocessed results from all measurements are investigated.

Particle images from both off-line and in-line imaging techniques reveal the presence of non-spherical particles mixed with spherical ones, as shown in Figure 5.2 and Figure 5.6. Except for the two largest particle size groups, all samples are found to contain spherical and ellipsoidal particles. This finding explains the inconsistency between the

PDFs in Figure 5.10(a)-(d) and the sieved ranged for these particle groups, and demonstrates clearly the limitations of using a sieve to establish a PSD for samples containing particles of non-uniform aspect ratio. A non-spherical particle such as an ellipsoid can pass a mesh with an aperture smaller than the length of its long axis, as long as the length of the other axis is smaller than the mesh aperture. For samples containing particles of diverse aspect ratio, the sieve will not be an effective method in determining the distribution of particle sizes.

Depending on whether in-line or off-line microscopy technique is used, the presence of the non-spherical particles can impact differently the PSD analysis. The shortest axis of particles, for example, the thickness of a platelet particle, is often underrepresented in off-line image analysis due to particles being static on microscopic slides and showing the two preferential longer axes. On the other hand, it is also known that in-line imaging techniques are more susceptible to the presence of out-of-focus particles. The in-line method captures a slice of particles in movement with different orientations, and the presence of elongated particles tilted with respect to the focus plane might be counted as partially detected particles [12]. Although the image analysis algorithm can impose a focus threshold to discard out-of-focus particles, constructing the true shape of the non-spherical particles using in-line imaging is challenging. Instead, the different particle projections observed can be counted as smaller particles, subsequently shifting the distribution towards smaller sizes, as observed in Figure 5.10(a) and (b).

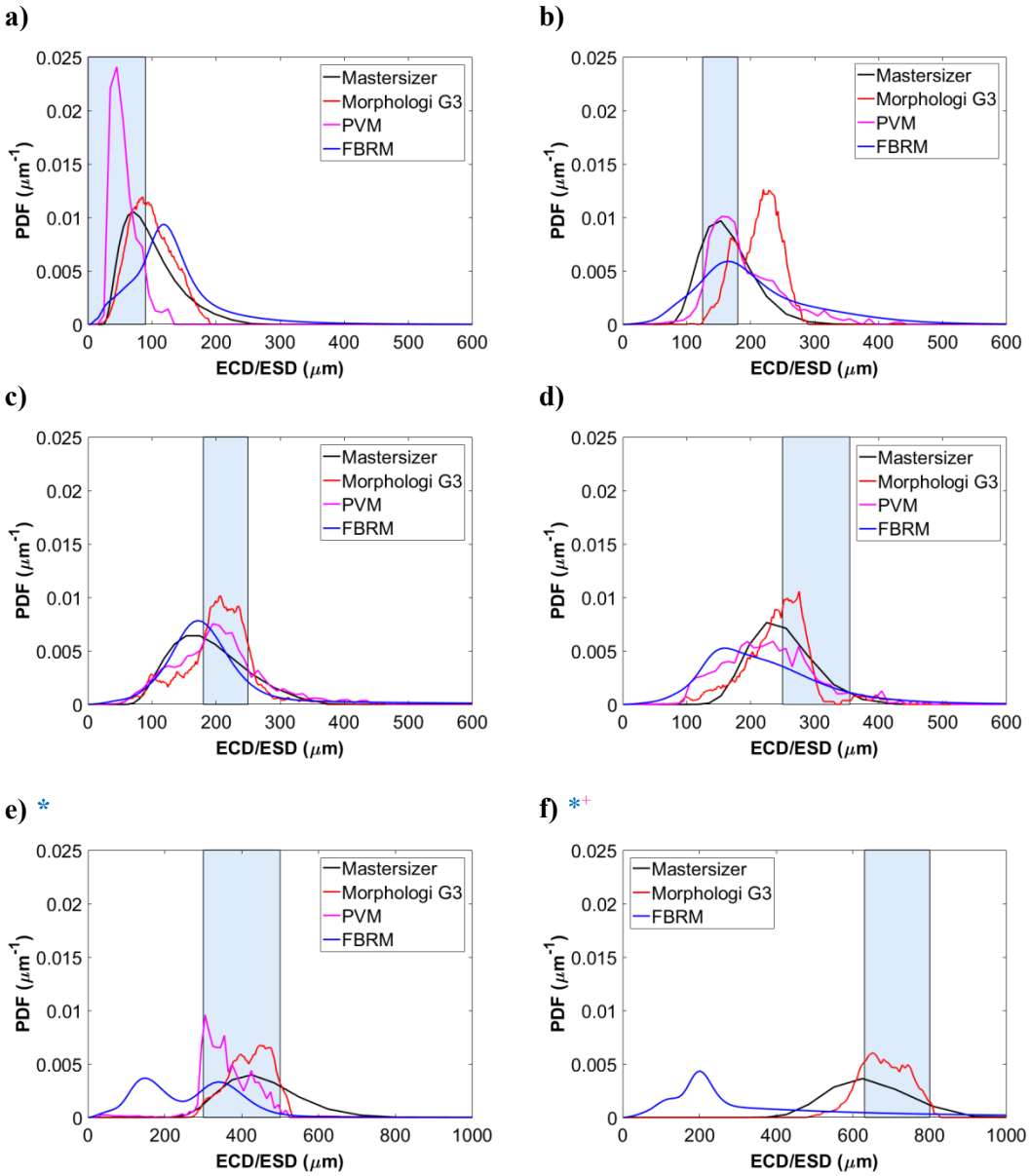


Figure 5.10: PSD obtained or/and estimated from Mastersizer, Morphologi G3, PVM and FBRM for **a)** PS0-90, **b)** PS125-180, **c)** PS180-250, **d)** PS250-355 **e)** PS300-500* and **f)** PS630-800*+. The shaded size ranges correspond to those defined by sieving.

*Inaccurate CLD from FBRM to reconstruct PSD; + A representative number of particles for a valid PVM analysis was not reached.

Figure 5.11 compares the aspect ratio obtained from Morphologi G3 with the aspect ratio extrapolated from PVM analysis. Morphologi G3 reveals two predominant aspect ratios, ~0.5 and 1, for PS0-90 and PS125-180 in Figure 5.11, indicating a significant amount of thin ellipsoids and spherical particles in the sample. It suggests that ~70 and 60% of the particle population of PS0-90 and PS125-180, respectively, have an aspect

ratio of less than 0.5. On the contrary, the PVM processing algorithm indicates that particles with an aspect ratio less than 0.5 are only around 30 and 10% for these samples. This is due to the low probability of observing the longest dimension of the elongated particles given their random orientation with respect to the focal plane. In addition, portions of out-of-focus tilted ellipsoids being treated as in-focus objects by the PVM analysis algorithm can also contribute to larger aspect ratios. Furthermore, off-line imaging analysis seems more sensitive to resolve the particle population of different shapes. Inspecting the particle images from Morphologi G3 reveal that the peaks at 175 and 220 μm in Figure 5.10(b) are mainly related to the spherical and ellipsoidal particles, respectively. On the other hand, the PSD from PVM analysis only exhibits a shoulder which corresponds to the length of the ellipsoids. Better agreement in PSD between Morphologi G3 and PVM can be observed from samples that mostly consist of spherical/nearly spherical particles. For PS180-250 and PS250-355, less than 20% of the particles show an aspect ratio below 0.5 in both analyses. As most of the particles are nearly spherical, the impact of aspect ratio on the PSD is less prominent.

The PDFs from Mastersizer analysis seem to be in better agreement with PVM imaging analysis, as observed in Figure 5.10. Mastersizer measurements are performed in a suspension of particles. Although LD devices assume all particles are spherical, particle shape influences the scattering pattern of the light. Non-spherical particles will diffract the laser beam differently depending on the average orientation of the particles in the beam. Thus, in the case of ellipsoidal particles, the probability of diffracting a beam based on the width of the particles is much higher than based on the length of the particle, resulting in the skewed curve towards smaller sizes. Due to sieving, the width of the ellipsoidal particles should be fairly similar to the diameter of the spheres in the same suspension. In general, a better agreement between imaging-based and LD methods can be obtained for samples with spherical particles, as shown in Figure 5.10(e) and (f). The underestimation of the population for larger particles in PS300-500 from the PVM analysis algorithm can be explained by the fact that larger particles are more likely to be incomplete within the field of view and therefore discarded from the analysis, leaving the smaller particles to dominate and shift the PSD towards smaller sizes. The same issue also leads to insufficient particles being captured from

PVM to form a meaningful PDF for PS630-800. Off-line methods such as Mastersizer and Morphorlogi G3 are less likely to be affected by this issue, hence providing more reliable and consistent PSDs.

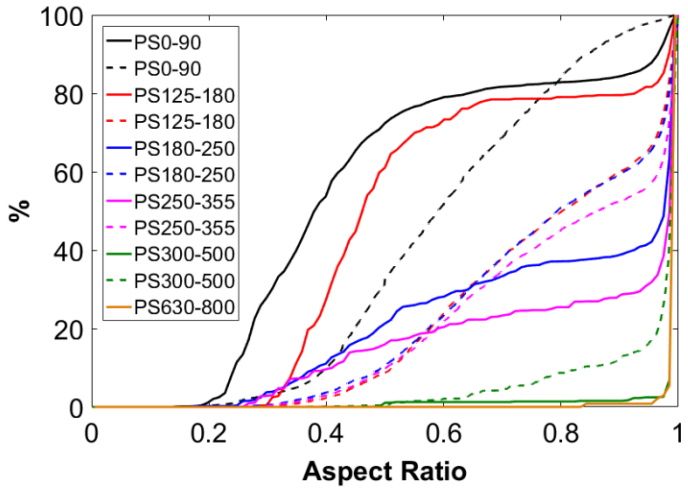


Figure 5.11: Aspect ratio cumulative distributions obtained from Morphologi G3 (solid lines) and PVM (dashed lines). A representative number of particles in PS630-800 for a valid PVM analysis was not reached.

The results in Figure 5.10 also demonstrate several limitations of estimating PSD from CL measurement. Firstly, the overestimated PSD in Figure 5.10(a) can be attributed to the overcrowded particle population in the sample. As seen in Figure 5.6(a) and Figure 5.5(a), images from 10 wt% solid load of the smallest particle group contains a significant amount of overlapping particles, in addition to the large number of ellipsoids. As a result, the CL measurement in macro mode can be misled by counting the overlapping particles as a larger one. Inverting such CLD transfers the error into an overestimation of the particle size. As the particle size increases, while maintaining the same solid loading in the sample, the reduction of particle number density resolves the issue of overlapping particles and more consistent PSDs are observed when compared with the imaging-based method. Another limitation in estimating PSD from CLD is the interesting observation of the peak around 150 µm for the three largest particle size groups. Although other methods such as PVM also indicate the presence of a small number of fine particles in the sample, it is not sufficient to result in the

persistent and prominent peak at 150 μm in Figure 5.10(d)-(f). The significant deviation of CLD-inverted PSD for larger particle group can be attributed to the transparent appearance of the particles. The internal reflection of the PVM's forward laser illumination is clearly observed (as seen in SFigure 2(c)); the CL measurement would suffer similar interference where specular reflection and additional internal scattering splits the length of the chords [45, 137-139]. Such interference can be clearly identified from both unweighted and square-weighted CLDs in Figure 5.5. As the actual particle size range increases from PS250-355 to PS630-800, the impact of chord splitting continues to worsen. Eventually, the CLs that correspond to the actual particle size turn into a second less prominent contribution, as seen in Figure 5.10(e) and (f).

In Figure 5.10(b) and (c), the PSD from the inversion is in relatively good agreement with the PSD from Mastersizer. The probability of FBRM scanning across particles width is higher than across the length. This effect in the CL measurement is then translated upon the inversion to PSD, resulting in a shift towards smaller sizes in comparison to off-line imaging tools. In Figure 5.10(d), the PSD from CLD inversion shows a peak around 150 μm with a shoulder at 300 μm. Also, PVM indicated the presence of fine particles in the suspension, showing a highly distributed size range of particles present in the suspension. This high number of fine particles seems to have a greater impact on the CLD measurement and consequently on the inverted PSD. Another possible source of discrepancies is the issue of chord splitting, which is no longer negligible but is not as significant as for PS300-500 and PS 630-800.

Overall, the comparison of PSD obtained from different analysis methods highlights the advantages and shortcomings of each method. The imaging-based methods provide the most consistent and detailed measurements, although the range of particle size measurable for the in-line method can be constrained by both optical setup and the analysis algorithm. The presence of non-spherical particles poses a significant challenge to the LD method and deteriorate the performance of FBRM. The results also highlight the challenges in analysing samples with high particle density, which will be discussed in details in the following section.

5.2.1.4. Effect of solid loading on in-line measurements

The effect of solid loading is investigated using only in-line methods; the off-line particle measurements used in this study are not capable of quantitating the particle population in suspensions. Furthermore, the PDF is not suitable for this analysis as the particle distribution described is normalised to the total number of particles assessed by the measurement. While the in-line particle measurement techniques are mostly focused on assessing the particle size attribute and not the particle quantity, based on the working principles of the techniques, the increase of solid loading can be manifested as an increase in the numbers of the particle images or chord length captured per unit time. Figure 5.12 shows a typical example of the changes in the unweighted and square weighted CLD with increasing solid loading; an increase in solid loading leads to an increase in the number of counts for all CLs and to a shift towards smaller chord lengths. As increasing the solid loading rapidly increases the particle number density, the presence of larger number of particles may interfere with the flow of the suspension, leading to small particles to crowd the FBRM window and to be measured more often, consequently shifting the distribution towards smaller sizes. The unweighted CLDs for PS125-180 in Figure 5.12(a) show a peak with a shoulder whose relative count to the main peak decreases as the solid loading increases. The peak at ~180 μm corresponds to the diameter of the spheres and the width of ellipsoidal particles. Chords related to the major axis of the ellipse are responsible for the broadening of the peak towards the right. The shoulder at ~70 μm could be related to the presence of fine particles or very thin and elongated particles present in the suspension. As in the unweighted distribution all particles have equal weight, fine particles are emphasized. A similar result can be found for PS0-90 in Figure 5.6(a). The use of the square weighted distribution, Figure 5.12(b), allows the resolution of the shoulder on the right of the peak at ~200 μm, which is related to the length of the ellipsoidal particles. Similar to the unweighted CLD, the PVM analysis algorithm shows a PSD shift towards smaller sizes when increasing the concentration (Figure B.6).

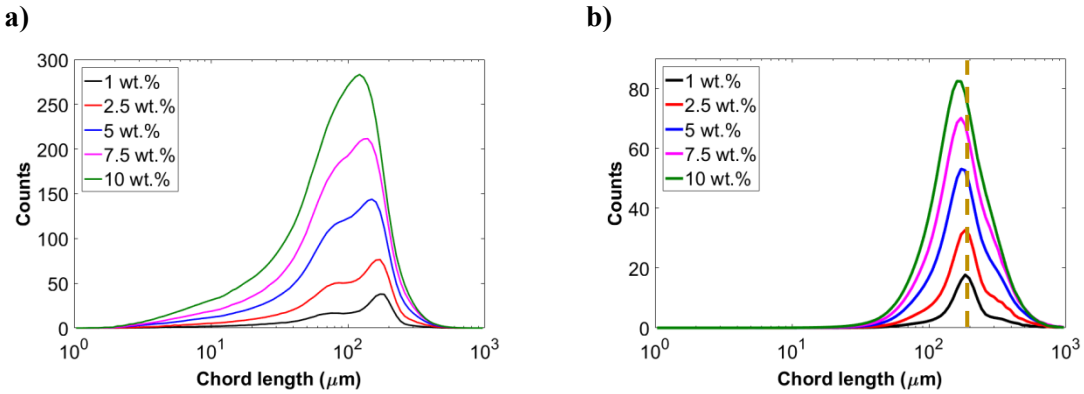


Figure 5.12: Effect of solid loading expressed in **a)** unweighted and **b)** square weighted CLD of PS125-180. The dashed line corresponds to the chord length of maximum count for 1 wt.%

The changes in the total CL counts from the unweighted CLD for all particle size groups is summarised in Figure 5.13(a). As an increased particle number density should increase the number of particles scanned by FBRM, it should also increase the unweighted count. However, although the total counts are number-sensitive, its relationship with the number of particles in the suspension remains unclear. A non-linear increase in the total count with solid loading can be clearly seen for PS0-90, for example. The particle number density, i.e., the number of particles per gram of solution, can affect the resolution of the CL between two particles while localized noise to the reflected intensity can result from the surface roughness of the particles. An increase in solid loading may obscure the laser intensity or even become overcrowded to confuse the reflected signal, as observed in Figure 5.10(a). At low solid loadings, the laser beam may become quite diffuse and lead to chord length overestimation if the particles are highly reflective [140]. The complexity of this relationship increases when dealing with particles of different sizes and shapes, and can result in a nonlinear response of counts to solid loading.

The number density can be a more adequate measure of the concentration of the solution since it considers the size and shape of the particles. The number density n_p represents the number of particles present per gram of solution:

$$n_p = \frac{N}{m_{solution}} \qquad \textbf{Equation 5.1}$$

where $m_{solution}$ is the mass of solution and N is the number of particles in the solution that can be calculated as:

$$N = \frac{m_{particles}}{m_1} \qquad \textbf{Equation 5.2}$$

With $m_{particles}$ being the total mass of particles added to the solution and $m_1 = \rho_{PS} V_1$ being the mass of one particle. The density of polystyrene ρ_{PS} is 1.04 g cm^{-3}. Finally, V_1 is the characteristic volume of one particle and it is calculated as the D50 of the volume distribution of each sample. In order to take into account the shape of the particles present in the suspension, the volume of each particle $V_{1\,i}$ contributing to the distribution was calculated as:

$$V_{1\,i} = \frac{\pi}{6} L_i \, W_i^2 \qquad \textbf{Equation 5.3}$$

, assuming the particles have ellipsoidal shape with equal dimensions for the two minor axes. Length L_i and width W_i were obtained from Morphologi G3 measurements for every individual particle in the sample, to represent each particle size group distribution. Considering the significant number of ellipsoidal particles in some particle size groups, the calculation of particle volume assumes particle of ellipsoidal shape with one major axis (i.e. particle length) and two identical minor axes (i.e. particle width). Figure 5.13(b) shows the relationship between the total number of counts, obtained from the unweighted CLD distribution, and the particle number density. A better linearity between particle number density and the total CL count is observed, with larger particles resulting in higher total count for a given number density. This could be due to the higher probability for a larger particle to be scanned by the FBRM, as suggested by Vaccaro *et al.* [48].

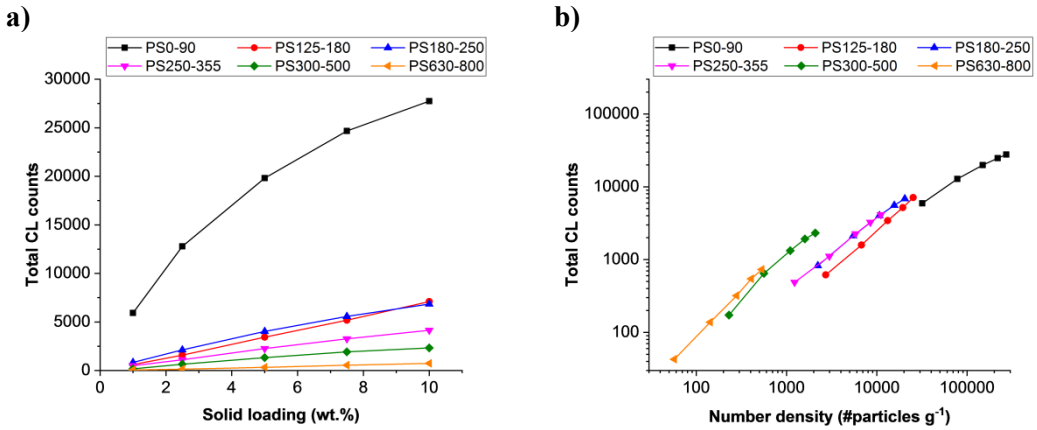


Figure 5.13: Changes in total number of CL counts with **a)** solid loading and **b)** number of particles per g of sample, for all samples.

The effect of solid loading on PVM is analysed using number and volume-weighted distributions of the particles in focus over 10000 frames. An example of the result using PS125-180 is shown in Figure 5.14. As discussed earlier, the two peaks in Figure 5.14(b) are related to the mixture of spherical and ellipsoidal particles in the sample. For PVM analysis, the increase of solid loading should not affect the accuracy of the analysis algorithm to estimate the PSD; the particle number counts should increase with the increase of solid loading as a result of the increase in particle number density. However, it is found that only particles size groups larger than PS180-250 follow the expected increase in the unweighted counts. Smaller particle size group exhibits an unexpected response to the increasing solid loadings, as seen in Figure 5.14 (and Figure B.6(a)-(c)). This can be explained by the overcrowded particle images from these size groups that affect the ability to resolve particles and particle shape. As discussed for Figure 5.6(a), single particles without overlapping are unlikely to be found for smaller particle size group. The PVM analysis algorithm does not directly exclude images of overlapped particles. However, they are more likely to fall on the image frame and be discarded. This, together with the deteriorated contrast of the images for high concentrations, reduces the representativeness of the results from in-line imaging. This can be expressed by comparing Figure 5.15(a) and (b) where, despite the steady increase in total particle (objects) count for all particle size groups, the number of particles (objects) in focus for PS0-90 and 125-180 drops significantly with increasing solid loading. When plotting the particle counts with respect to number

density (Figure 5.15(c) and (d)), a decrease in counts for particles in focus is observed for samples with particle number density larger than 10000 particles g^{-1}. This observation is independent of the particle size and can be considered as the upper limits for quantitative PVM analysis on solid loading.

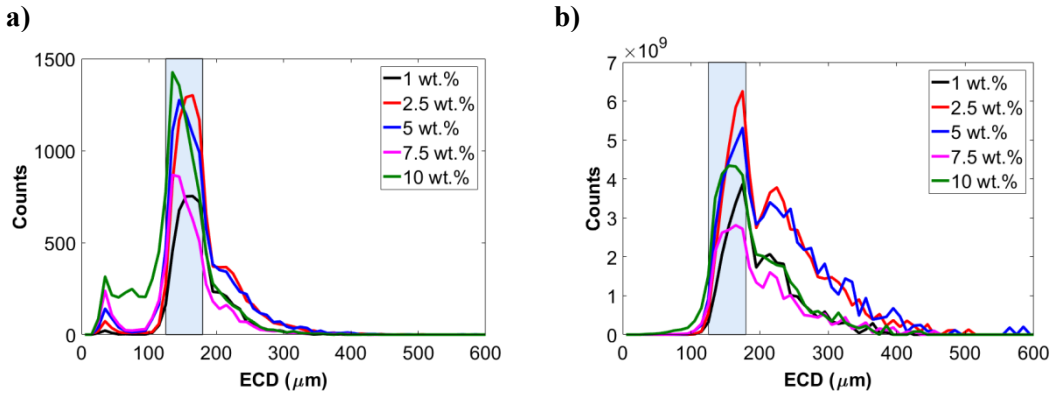


Figure 5.14: Effect of solid loading expressed as **a)** number and **b)** volume-weighted ECD of PS125-180, on the data obtained from the PVM image processing algorithm. The shaded size range corresponds to those defined by sieving.

Another interesting observation in Figure 5.14 (and Figure B**.6**) is the shift of the PSD toward smaller sizes with increasing solid loadings. The concentration at which the shift starts for each particle group coincides with the decrease of the in-focus particle counts in Figure 5.15(b) and (d). Further investigation on the analysed particle images suggests that the larger particles in overcrowded suspensions are more likely to overlap with other particles, and subsequently be discarded. As a result, the smaller particles become the dominant population in the PSD as the solid loading increases. For larger particle size groups, a more representative PSD can be obtained at higher solid loadings.

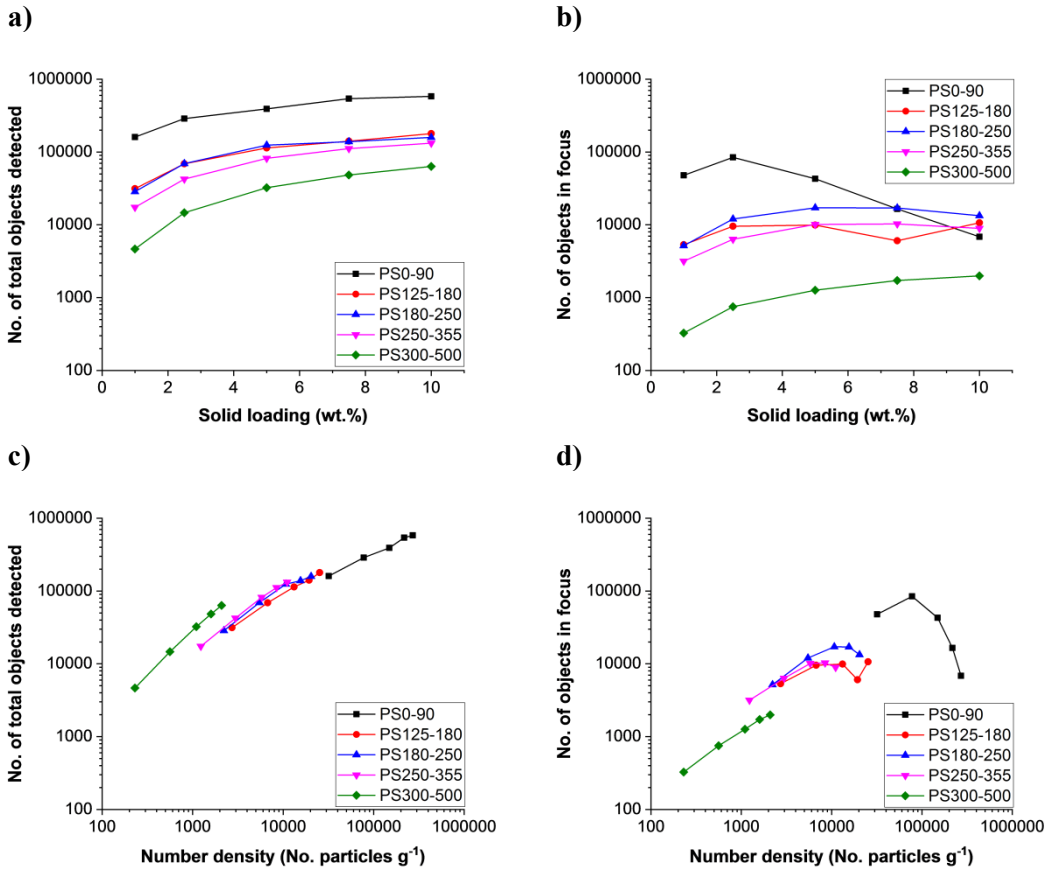


Figure 5.15: a) Total number of objects detected and **b)** number of objects in focus in 10000 frames, by PVM, at different particle sizes and solid loadings. **c)** and **d)** express the total objects in a) and b), respectively, in particle number density.

5.2.4. Performance constraints of in-line measurements

As observed in section 5.2.1.3, the PSD obtained is greatly dependent on the technique used for the measurements, and there is no obvious way to correctly describe the "true" particle size or concentration of a system, especially when dealing with non-spherical particles. Compared to the off-line imaging technique, the in-line imaging-based measurement is limited by the focal depth and the field of view. Particles with a size larger than either of them cannot be correctly measured due to exceeding the dimension of the image, or not being in focus for imaging. This does not only affect the accuracy in describing the PSD but is also more likely to exclude and underrepresent the larger particles. In addition, a clear distinction of particles in-focus from the others is desirable, but not always achievable when measuring highly turbid

mediums. The optical constrain in the field of view and the quality of particle images means that it does not only affect PVM's performance in PSD analysis, but also the concentration estimation. Furthermore, depending on the particle size and concentration, a large number of images may need to be acquired and processing the images for a meaningful PSD can be time-consuming. Nevertheless, the in-line imaging method can be valuable in providing sample information in situ and at real-time to support the interpretation of the data extracted from the other methods, such as the presence of different particle shapes.

Inferring CLDs to PSD is not straightforward. As the chord length is non-specific to the shape of particles, the presence of non-spherical particles complicates the CLD interpretation. For example, it can lead to the appearance of multiple peaks in the CLD. Moreover, the inversion algorithm uses a single aspect ratio for the entire population which can be unrealistic if a sample is a mixture of particles of different shape, adding more uncertainty to the inverted PSD. Using FBRM to assess solid loading may lead to misleading results. Although the total number counts obtained from the unweighted CLD might be used as an indication of solid loading, its relationship with the number of particles in the suspension remains unclear.

Due to the measurement principles used, increasing sample turbidity can significantly affect the performance of current particle measurement techniques. Quantitative in-line methods measure and count the interaction of incident light with single particles. The quality of the measurement can deteriorate rapidly with the increase of turbidity since this condition is unlikely to be met. On the other hand, spectroscopic-based techniques are commonly applied to highly turbid systems which utilise the diffuse reflectance/transmittance. The recently reported SAR-DRM system presents a unique and interesting approach to resolve and utilise different light scattering path caused by the particles [9, 10]. Previous studies using SAR-DRM focus on suspensions with particles below micron size, from 100 to 500 nm in diameter [101]. In this study, for the first time, SAR-DRM will be assessed on micron size particle suspensions, from <90 to 800 μm.

5.3. SAR-DRM – Qualitative analysis

SAR-DRM is applied to the same set of samples to investigate its potential in analysing the suspensions of particles in the range of microns. Although it is reported that SAR-DRM spectra can be inverted to obtain bulk optical properties, [10] and subsequently to estimate PSD, [11] the reported inversion methods focus on systems that consist of particles below one micron. The applicability and performance of the proposed inversion method to micron size suspensions are still unknown. Therefore, this study will focus on providing a qualitative assessment of the effect of size and concentration of particle suspensions on SAR-DRM spectra.

5.3.1. Particle size effect

As described in Methodology, the SAR-DRM probe provides a total of 12 combinations of spatial and angular arrangements between the source and detecting fibres. Figure 5.16 shows an example of SAR-DRM spectra from 10 wt% suspensions for all particle size groups. A decrease in the SAR-DRM intensity with increasing spatial distance from 0.3 to 1.2mm can be clearly observed. This is a result of increased path length, which causes stronger absorption and scattering of the incident light. The spectra from PS0-90 show the strongest intensity for most of the fibre arrangements; the only exception is when using normal incident light with 0.3 mm spatial distance between the source and detecting fibres. The lowest SAR-DRM intensity was obtained from samples of the largest particle size range, PS630-800. This is due to the reduced turbidity as a consequence of lower particle number density. While large particles exhibit stronger scattering, the reduction of particle number density by increasing the particle size at a fixed particle volume concentration supersedes the increased scattering. By changing the angular difference between the source and detecting fibres, different spectral response of the particle size groups can be seen. For example, an increase in the diffusely reflected intensity with the increased angular resolution is observed for all spatial resolutions. This can be explained by the way light travels in a particulate system; an angular incident light travels shallower into the sample with shorter optical path lengths, as illustrated in Figure 4.2(b), and results in stronger

reflectance. By using multiple spatial and angular arrangements, SAR-DRM measures the sample at different depths, offering sensitive and detailed reflectance information which could be utilised further to extract the physical and chemical properties of the samples in the study.

Analysing the change in SAR-DRM spectra with particle size is not straightforward. The SAR-DRM spectra contain the combined effect of multiple scattering events from multiple particles. Furthermore, the magnitude and direction of the scattered light of each particle depend on its characteristics, such as size and shape, and the wavelength of the incident light. The analysis would become further complicated if the absorption of the system is nonnegligible. A previous study on SAR-DRM spectra suggests that the mean particle size and particle concentration can be inverted from the spectra using a physical-based model [10]. However, considering the size range and the mixed shapes of the PS samples used in this study, further development of the model is required to apply the methodology to the micron size particle suspensions in this study. Therefore, an empirical analysis on the SAR-DRM spectra is performed to evaluate its performance at distinguishing samples of different particle size. This is conducted by fitting linearly the SAR-DRM spectra to obtain their slope, a common spectral feature which can be related to the scattering effect.

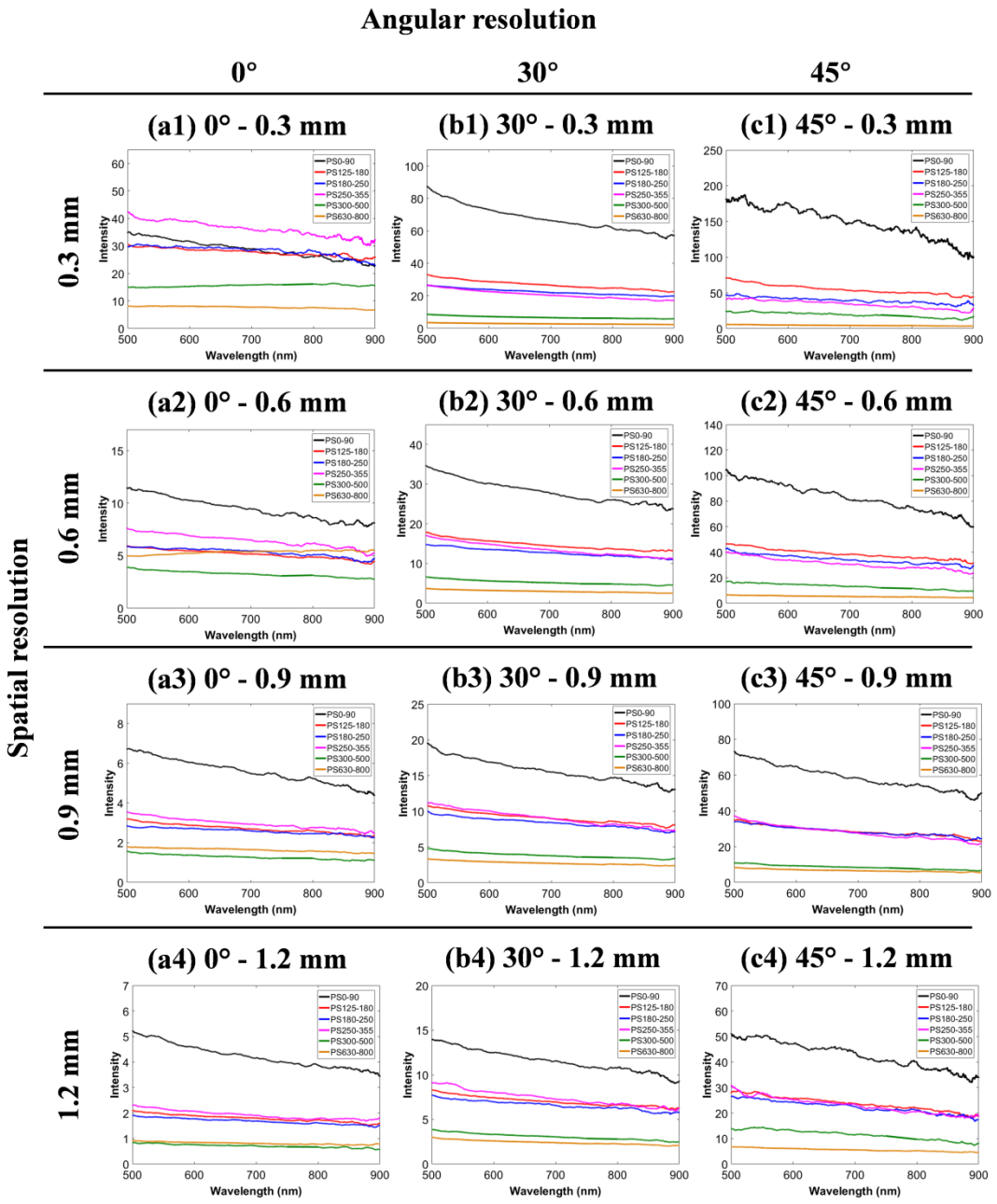


Figure 5.16: Effect of particle size in vis-NIR region, when the light is emitted at a) 0, b) 30 and c) 45° and collected from the 4 different distances from the light source: 1 (0.3 mm), 2 (0.6 mm), 3 (0.9 mm) and 4 (1.2 mm). Solid loading used – 10 wt.%.

Figure 5.17 summarises the slope extracted for the spectra shown Figure 5.16(a1)-(c1) and (a4)-(c4). The slope in general increases with increasing the particle size, except PS250-355. A similar trend is observed when analysing the signal collected from the fibres at 0.6 and 0.9mm from the light source, as observed in Figure B**.7**. Increasing

the incident angle increases the differences in the slope between different particle size groups, with the largest difference obtained from spectra using angular incident light of 45º. Overall, SAR-DRM results in a more prominent difference between the particle size group than either unweighted or volume-weighted CLD from FBRM (shown in Figure B.**2**). The results suggest that the spectral response under different combinations of spatial and angular arrangement can be utilised further to achieve particle size analysis with greater sensitivity.

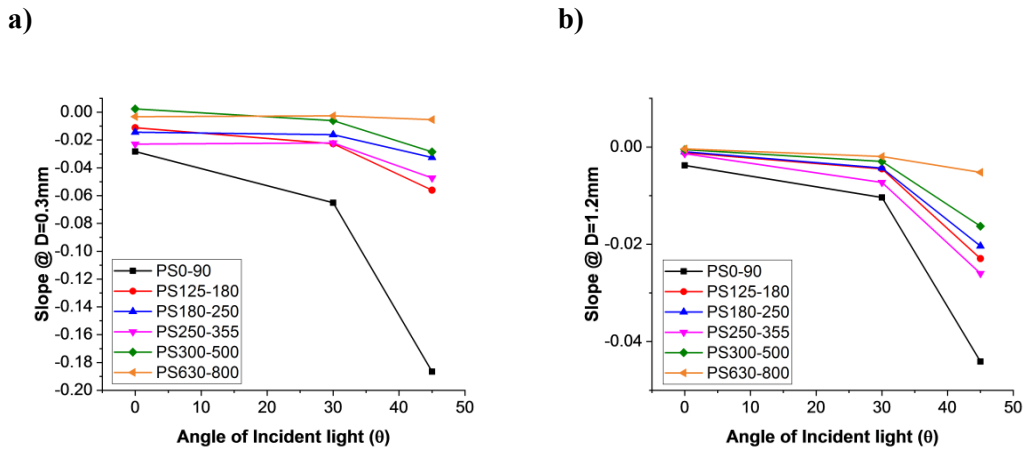


Figure 5.17: Effect of particle size on the vis-NIR slope when using the 0, 30 and 45° light source and collecting the signal from **a)** 0.3 mm and from **b)** 1.2 mm. Solid loading used – 10 wt.%.

5.3.2. Solid loading effect

A similar analysis is performed to investigate the effect of solid loading on SAR-DRM spectra. A typical example of the changes in the SAR-DRM spectra with different particle concentration is shown in Figure 5.18. Overall, the increase of solid loading increases monotonically the intensity of the reflected light. The change in the intensity, however, is non-linear and dependent on the angular and spatial configurations of the measurement. For example, SAR-DRM spectra show closer intensity among different concentration when using an incident angle of 0°. Increasing the incident angle results in a wider spectral separation between different solid loadings, and a slighter decrease in the intensity as the spatial distance increases.

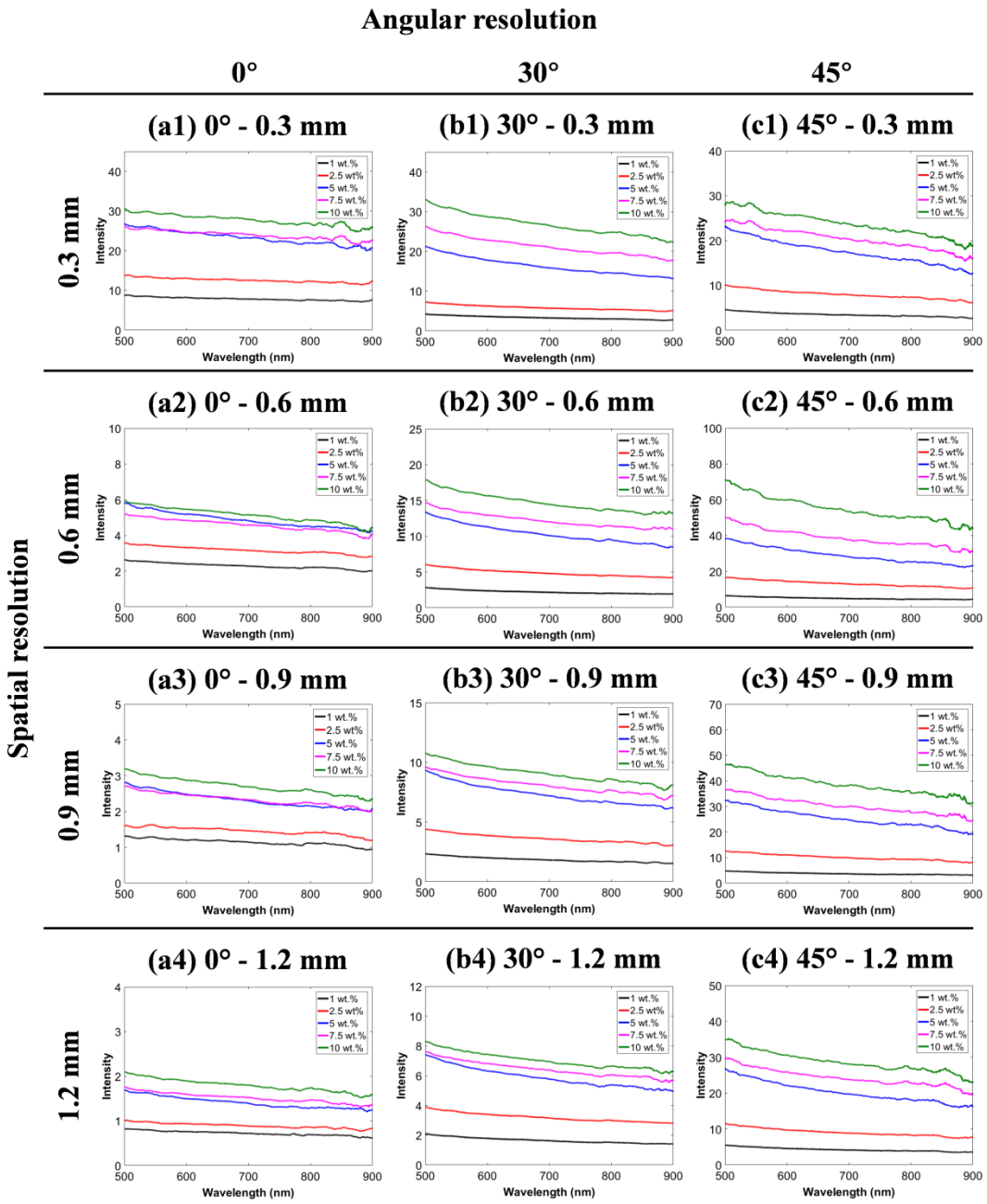


Figure 5.18: Effect of solid concentration of PS125-180 in vis-NIR region, when the light is emitted at **a)** 0, **b)** 30 and **c)** 45° and collected from the 4 different distances from the light source: **1** (0.3 mm), **2** (0.6 mm), **3** (0.9 mm) and **4** (1.2 mm). Particle size group used – PS125-180.

To better quantitate and compare the changes of SAR-DRM intensity with the solid loadings for different particle size groups, the intensity at a wavelength of 700 nm is selected for further analysis. Using the smallest spatial distance as an example, Figure

5.19(a1)-(c1) show that the configuration using angular incident light gives larger intensity difference between different particle size groups, as also discussed for Figure 5.16. Increasing the solid loading not only increases the intensity of the reflected light, but it also provides more consistent intensity differences between the particle groups. For instance, Figure 5.19(b1) and (c1) show clearer separation between PS125-180, PS180-250 and PS250-355 for a solid loading higher than 2.5 wt%. This is in contrary to the imaging and chord length analysis where the techniques start to show limitations for solid loading above 2.5 wt%, as shown in Figure 5.13(a) and Figure 5.15(a)-(b). Figure 5.19(a2)-(c2) shows that the increase in SAR-DRM intensity can be related to the increase of particle number density, similar to the observation on FBRM in Figure 5.13(b). Like FBRM, the relationship between SAR-DRM intensity and particle number density is not straightforward. In this case, the optical path length of light travelling inside the sample is altered by changes in particle size and concentration. To fully describe the phenomena will require the development of a physical model based on light propagation theory to invert the spectra, as done in previous studies on suspensions of particles below micron size [141, 142].

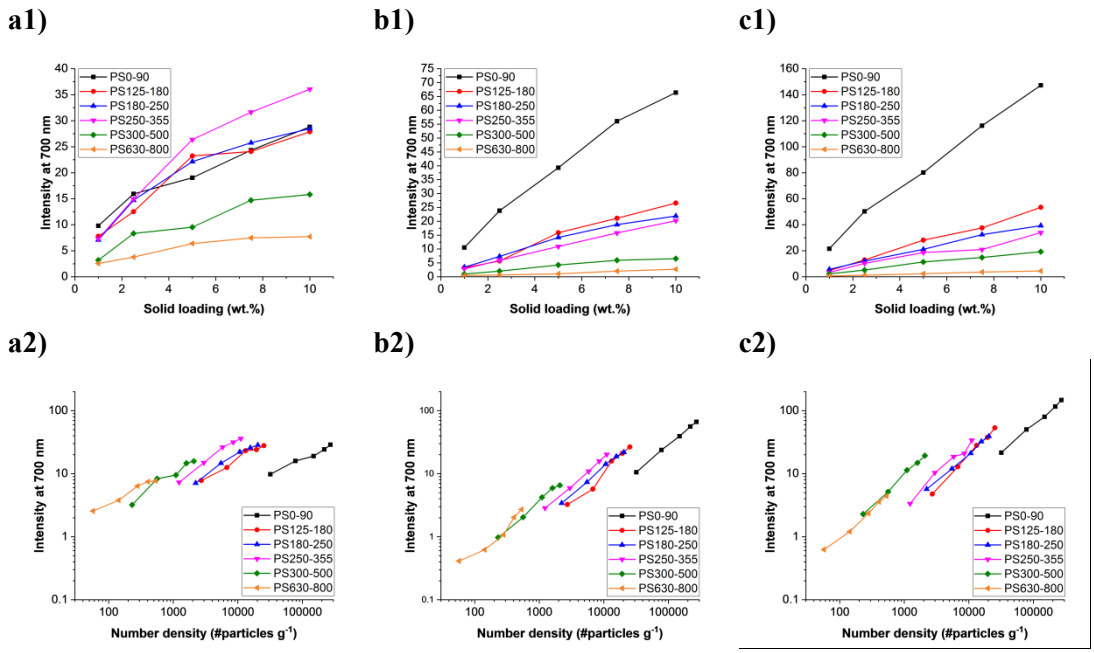


Figure 5.19: Solid loading effect when using **a1)** 0, **b1)** 30 and **c1)** 45° light source and collecting the signal from the closest fibre to the source, at 0.3 mm. (**a2)-(c2)** show the corresponding number density for (a1)-(c1).

Similar analysis of spatial distances of 0.6-1.2 mm can be seen in the Supplementary Information (Figure B.8). However, the increase of spatial distances seems to decrease the sample differentiation, for example, similar intensity for PS125-180, PS180-250 and PS250-355 are observed in Figure B.8(b1)-(b3) and (c1)-(c3). The findings seem to contradict the discussion on Figure 5.16 and Figure 5.17 that the angular fibre arrangement improves the differentiation of particle size group. It is worthy to note that the findings in Figure 5.15 (and Figure B.8) are wavelength specific. Referring to Figure 5.16, the difference in SAR-DRM spectra between PS125-180, PS180-250 and PS250-355 is difficult to resolve, especially for wavelength ranges where the spectra intersect. As a result, determining the particle size group based on only one wavelength could be misleading. It would be more reliable and robust to include multiple, or a range of, wavelengths to build a more accurate analysis. Multivariate analysis (MVA) provides an approach to incorporate information from multiple wavelengths to achieve robust analysis and is commonly applied to model spectroscopic data [79, 83, 143]. Indeed, MVA was applied in the first study reported on SAR-DRM; however, it is suggested that the choice of wavelength and fibre arrangements has a significant impact on the performance of the analysis [9]. Such an analysis is outside the scope of this study and would be of interest for further study.

The results show a systematic response in SAR-DRM spectra with particle size and concentration when analysing micron size particle suspensions. Although the performance at estimating PSD from SAR-DRM is still to be investigated, the technique illustrates its preference to measure high turbidity samples which is complementary to current in-line methods such as PVM and FBRM. Further investigation to select suitable fibre arrangement and wavelength region for MVA will be beneficial in understanding and optimising the robustness of the MVA model. Similar to the study on sub-micron systems, the extension of the physical model to invert SAR-DRM spectra would remove the need to build a calibration model for applying SAR-DRM to in-line PSD measurement.

5.4. Conclusions

In this work, we compare the performance of commercial off-line and in-line particle sizing techniques, and we explore the novel SAR-DRM technique to study the effect of particle size and solid loading in solid suspensions. Overall, the off-line measurements (Mastersizer and Morphologi G3) and PVM demonstrated a better performance at differentiating the samples, even for the three samples which show similar distributions due to the presence of ellipsoidal particles (PS125-180, PS180-250 and PS250-355). In addition, the change in the particle size range was not clearly observable in FBRM measurements, and both FBRM and PVM encountered difficulties in obtaining reliable measurement for larger particles (PS300-500 and PS630-800). The smooth and transparent particle in these size ranges caused chord length split in FBRM and exceeded the field of view for PVM. On the contrary, the results from SAR-DRM offer a different prospect in analysing the effect of particle size. In all cases, the spectra show the strongest intensity for PS0-90, with and exception for the spatial arrangement at 0.3 mm using a normal incident light, and the weakest intensity for PS630-800. Analysis of the slope of the spectra shows that increasing the incident light angle results in better separation between the samples, regardless of the spatial distances.

For the analysis of solid loading, a monotonic increase in total CL count with particle number density is observed while results from PVM analysis show an upper limit in particle number density. However, to determine the amount of particles in the system requires inverting CLD to PSD and cannot be directly inferred. Furthermore, the CLDs show a shift towards smaller sizes with increasing solid loading, which might introduce error in CLD inversion. The upper limit in PVM is identified as a further increase in solid loading deteriorates the quality of the particle images, and causes an increasing number of particles being rejected by the image analysis algorithm. On the other hand, an increase in the turbidity usually leads to an increase in the intensity of SAR-DRM intensity. Better linearity between the solid loading and SAR-DRM intensity is observed using angular incident light. However, the observation is only based on the intensity at one particular wavelength and would require further analysis

on the full wavelength range offered by SAR-DRM to establish the most sensitive spatial and angular arrangements.

In summary, for the first time, the performance of SAR-DRM to measure suspensions of micron size particles is demonstrated. The ability to work with samples with high solid loading over a large range of particle size shows SAR-DRM is a promising technique to complement other in-line analysis methods for process control and optimization. Further analysis of SAR-DRM spectra to extract bulk optical properties using physical-based modelling or to identify optimal spatial and angular arrangements would be valuable for a full assessment of SAR-DRM performance.

Chapter 6 – Estimation of Polystyrene Particle Size and Concentration of Polystyrene using Complementary PAT Tools

The previous chapter confirmed that SAR-DRM is sensitive to changes in particle size and solid concentration, similarly to other PAT tools such as FBRM and PVM.

In this chapter, the use of a multiple in-line sensor approach is investigated to assess the potential to extract quantitative information of particles attributes at real-time. Calibration models to estimate particle size and particle concentration of polystyrene beads in water are built using multivariate regression analysis on data collected from FBRM and SAR-DRM.

The chapter is divided into two parts. The first part presents and discusses the models built for individual techniques. In case of SAR-DRM, the influence of spatial and angular fibre arrangements is investigated for both visible and near-infrared spectroscopies regions. Combinations of data from different configurations are also analysed to assess improvement in the robustness of the calibration models. The impact of different pre-treatments, log-transformation, smoothing, SNV, MSC, EMSC and ISC on the calibration models performance is also evaluated. In the second part, it is examined whether these different technologies can be integrated using a sensor data fusion approach based on data augmentation and data co-adding.

6.1. Part 1 – Individual techniques calibration

6.1.1. FBRM

Particle size

As discussed in Chapter 5, FBRM particle size, concentration and shape induce changes in the CLD patterns. Although the effect is noticeable, further analysis is desired to assess whether such information can be used for quantitative modelling of particle critical attributes.

In this work, it is assumed sufficient averaging is applied to the CLD is constant and each CLD is unique for polystyrene particles of a specific sieved size and at a specific concentration Figure 6.1 shows the unweighted and square weighted CLDs of the 25 samples analysed in the FBRM dataset. PLSR analysis was applied in order to determine the CLDs which are more sensitive to describe particle size and concentration. The median particle size, D50, from laser diffraction, and concentration from gravimetric measurements were used as the reference values for the model.

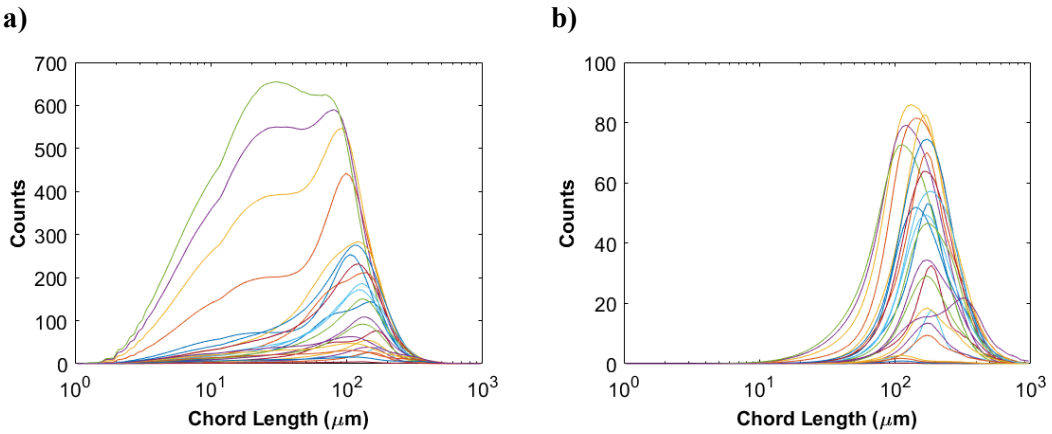


Figure 6.1: a) Unweighted and **b)** square weighted CLD of the 25 samples in the FBRM dataset.

Although empirical results suggest that mean centring reduces the multicollinearity hence, improving the interpretability and accuracy of the model by reducing the

covariances and correlations between the predictor variables and their interaction term, whether mean centring should be always applied or not is unconcluded. Geladi and Kowalski use mean centring to tailor the data in the calibration set to simplify the computing of NIPALS algorithm [144]. However, according to Seasholtz and Kowalski, mean centring should not be applied when there are no offsets or no baseline, which is the case of the measurements taken from FBRM [145]. Figure 6.2 demonstrates the outcome of applying mean centring on the unweighted CLD. The CLD shape variation with the particle sizes is not linear and systematic, subtracting the mean CLD from the dataset would not benefit the PLSR modelling in the next step. Other pre-processing methods that can be more suitable to mean centre the data without changing the shape of the CLD, ensuring the fitted mean of the CLD.

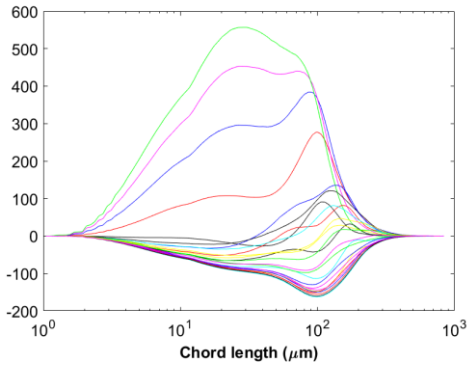


Figure 6.2: Unweighted CLD using mean centring of the 25 samples.

As observed in Figure 6.1(a), the unweighted CLDs present a skewed distribution, with a short tail to the right, a suitable transformation which may improve the modelling results is the log-transformation of the data, in order to decrease the variability of the data, leading to the higher efficiency of the model. In these unweighted CLD measurements, the counts vary from approximately 20 up to 700 counts, depending on the particle size and solid loading used. Converting the data to a comparable y-scale could help to improve the linear relationship between the CLDs and the parameters of interest. Hence, another approach proposed is the use SNV for pre-processing by standardising each CLD by its own average and standard deviation, to remove the

multiplicative interference of the backscattered light (counts) and linearize the measurement more effectively. In these measurements, there is no baseline displacement to be corrected. This method, which centres and scales individual CLD, results in a similar effect to data treated with the multiplicative scatter correction (MSC) with reduced sensitivity to peculiarities in the raw data as it is based on individual observations. It standardises each CLD using only data from that CLD, rather than using the mean of all CLD in the dataset to derive pre-treatment parameters.

A combination of the log transformation with the SNV correction method is also proposed to assess whether it improves the calibration models performance. The same principles are applied to the square weighted distribution.

Table 6.1 summarises the best PLSR models of the pre-processed CLD for the determination of particle size and beads concentration. The criteria followed to select the best models is described in Chapter 4, Section 4.5.

Table 6.1: Summary of the best PLSR models on different pre-processed unweighted and square weighted CLDs for estimating PS mean particle size.

Estimation of particle size						
	Unweighted CLD			Square weighted CLD		
Pre-processing	**No of latent variables**	**RMECV (µm)**	**R^2**	**No of latent variables**	**RMECV (µm)**	**R^2**
None	4	291	0.169	6	207	0.331
Log	6	111	0.697	6	68.1	0.889
SNV	8	49.7	0.945	8	18.3	0.993
Log & SNV	3	97.2	0.733	5	57.9	0.915

Overall, better model performances were obtained for the models built on the square weighted CLDs. This could be due to the square weighted being able to separate better the CLD from different particle size ranges, present a more consistent CLD shape and smaller counts difference among them. In both types of CL distributions, unweighted

and square weighted, the use of log transformation decreased significantly the error of calibration in comparison to model built on the non-treated data, about 62 and 67%, respectively. This is accompanied by a significant improvement of the R^2. The SNV is the pre-treatment leading to the best models performance in both types of distributions, leading to a decrease of the RMSECV of about 83 and 91% from the non-treated data. SNV seems to significantly reduce variance ‘‘within sample’’ in the spectra leading to a separation of samples with different sizes by analysing the samples in a similar/comparable scale. The combination of log transformation and SNV led to lower errors of calibration in comparison to when only using the log transformation of the data, describing the model variance using fewer LVs. However, the loss of the CLD features during the log transformation leads to the poor performance of the model in comparison to when applying only SNV to correct the data.

The best model performance for estimating particle size is built when pre-treating the square weighted CLDs using SNV, presenting an RMSECV of 18.3 μm and an R^2 of 0.993. For a visual overview, Figure 6.3(a) shows the RMSECV curve, Figure 6.3(b) plots estimated particle diameter against reference values and Figure 6.3(c) plots the residuals of the calibration dataset. The accentuated decrease of the RMSECV curve suggests most of the sample variance is captured by the first 3 LVs. After, the RMSECV slowly decreases up to 8 LVs where a minimum is reached. Loading plots show the effect of variables X on scores Y for each LV. Figure 6.4 shows the loading plots of the model using SNV as a pre-processing method to analyse the square weighted CLD. The first 3 LVs are mainly attributed to the differences in particle size, with the 2nd and 3rd LV capturing the variance respective to the larger particle size group PS630-800. LVs 4-8 were added to the model, as no accurate precision was possible with only 3LVs. Their respective loading plots are noisier than the first 3 loading plots, suggesting they contain information about the shift of the peaks with particle size, solid loading differences among samples with the same size and other physical sample variations such as shape (Figure 6.4(b2) Figure 6.4(b3)).

Although a high R^2, samples differences are observed. This is due to the fact the suspensions have different solid loadings within the same particle size group resulting in a shift of the peak position. As mentioned in Chapter 6, the presence of ellipsoidal

particles and their projection also have an impact on the CLD. Figure 6.3(c) shows the residuals, indicating how the model is performing and how far the estimated values are from the reference ones. It is observed that the residuals from PS630-800 are, in general, higher than the other samples of 1, 5, 7.5 and 10 wt.% of PS. This might be due to the different CLD shape obtained for this particle group in comparison to other particle size ranges (See Figure B.3), as it presents a bimodal distribution from the significant impact of specular reflection.

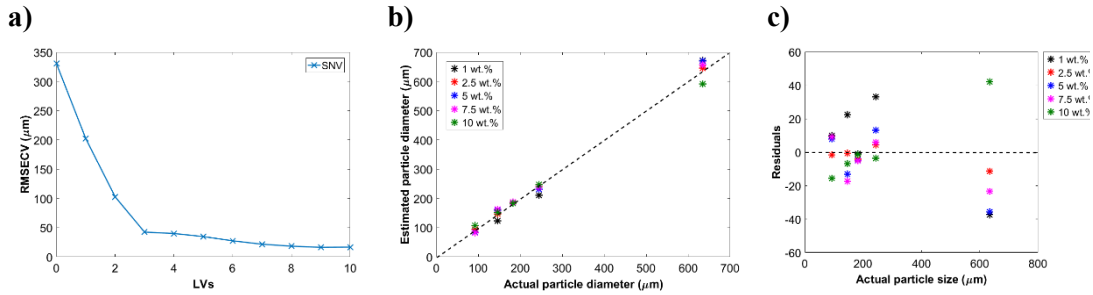


Figure 6.3: a) RMSECV versus LVs obtained from the PLSR analysis of the SNV pre-processed square weighted CLDs. **b)** Estimated versus actual particle mean size. **c)** Residuals in the calibration dataset.

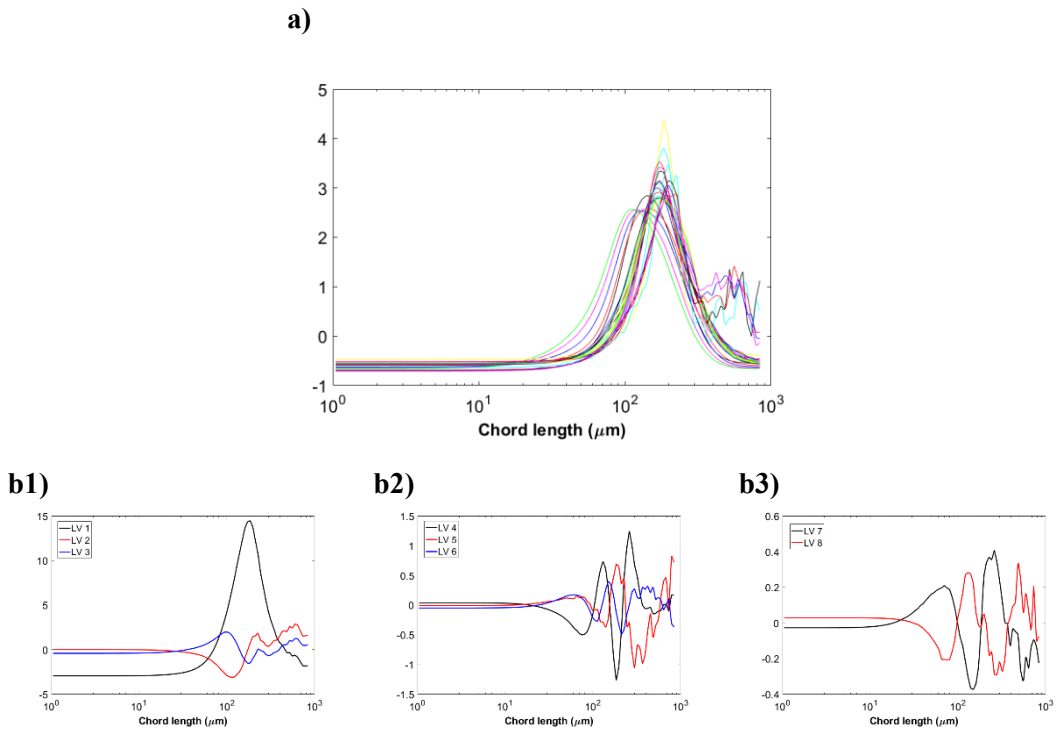


Figure 6.4: a) SNV pre-processed square weighted CLDs. (**b1)-(b3)** Loading plots of the calibration model built for mean particle size estimation for the first 8 LVs.

Particle concentration

The same approach was used to estimate particle concentration from the unweighted and square weighted CLD. As observed in Table 6.2, the PLSR models built showed, in general, a high RMSECV, typically above 2 wt.%. The poor ability to estimate concentration suggests the absence of a linear relationship to the samples. The use of pre-processing methods seems to remove CLD features, the log transformation does not allow an accurate separation of CLDs from different solid loading and the SNV removes the interference of the counts. This suggests the total number of counts contain important information to explain the variability of the model. Similar to the results from particle size, the use of square weighted CLDs led to the best model performance. The fact that this distribution approximates the results to a better well-defined unimodal distribution (except for the largest particle sizes, PS250-355 and PS630-800) not only helps to resolves better the peaks shift related to particle size but also the shift of solid loading within the same particle size group. The deemphasise of the fine particles effect seems to play an important role in model performance. Hence, when building the model using the raw square weighted data, a decrease in the RMSECV from 2 wt.% to 0.45 wt.% is observed. This is translated in a higher number of LVs explain the model as shown in Figure 6.5(a). Figure 6.6 shows the loading plots from using the raw square weighted CLDs. The first 5 LVs seem to be capturing most of the variance from the differences in particle size. As discussed in Chapter 6, the increase in solid loading is not only translated in the increased number of CL counts but also in a shift towards smaller CLs. Although the LVs 7-9 present noisier loadings plot, it suggests they contain information about the peak shift due to solid loading differences among samples with the same size, increase or decrease of the number of counts and shape of the particles.

Table 6.2: Summary of the best PLSR models on different pre-processed unweighted and square weighted CLDs for estimating PS particle concentration.

Estimation of particle concentration						
	Unweighted CLD			Square weighted CLD		
Pre-processing	**No of latent variables**	**RMECV (wt.%)**	**R^2**	**No of latent variables**	**RMECV (wt.%)**	**R^2**
None	5	2.52	0.604	9	0.45	0.982
log	4	2.54	0.448	4	2.36	0.502
SNV	5	2.07	0.618	5	2.26	0.530
Log & SNV	4	2.47	0.447	5	2.40	0.493

Figure 6.5(b) shows the estimated particle concentration versus actual particle concentration, in which an R^2 of 0.982 was obtained. The residuals plot exhibits that the less accurate estimation of particle concentration by the model built using raw square weighted CLD data, is when using a suspension of 10wt.% of PS125-180, resulting in a larger error.

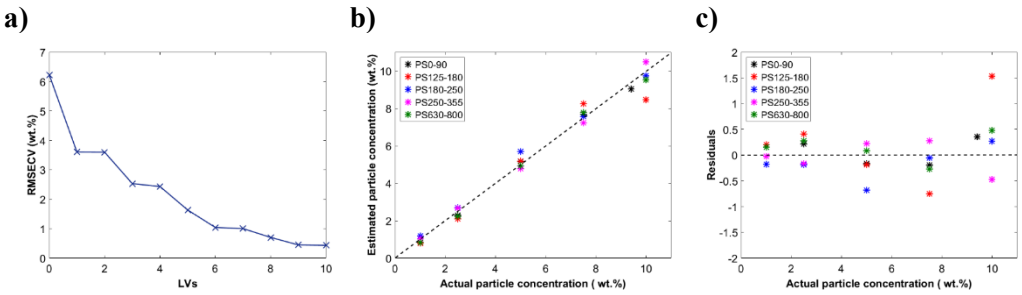


Figure 6.5: a) RMSECV versus LVs obtained from the PLSR analysis raw square weighted data. **b)** Estimated versus actual particle concentration. **c)** Residuals in the calibration dataset.

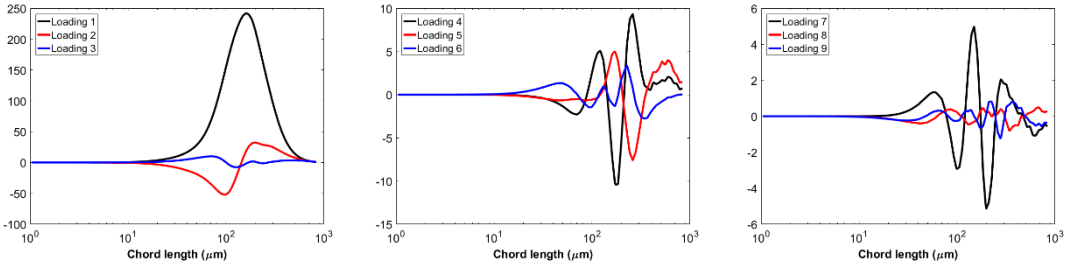


Figure 6.6: Loading plots of the calibration model built for particle concentration estimation using 9 LVs.

The results indicate a potentially promising approach to obtaining quantitative particle information in-line without the need to apply inversion algorithms for CLD analysis. A bigger dataset should be tested in order to increase the robustness and accuracy of the model. A possible application would be to monitor a polymerization reaction.

6.1.2. SAR-DRM

The following sections describe the PLRS analysis on SAR-DRM spectra in the two spectral regions obtained from SAR-DRM, vis-NIR and NIR. The dataset description is described in Chapter 4, Section 4.4. For each one of them, PLSR models are built to estimate particle size and concentration. Due to the spatially and angularly feature of the probe, different measurement configurations are investigated to assess the influence of information in the model performance. The best strategy to combine information from different detectors and sources, whether it is co-adding data or augmentation of the data, is also evaluated. The reference values used to build the models were the median particle size, D50, from laser diffraction, and the concentration from gravimetric measurements. The criteria followed to select the best performing model is described in Chapter 4, Section 4.5.

6.1.2.1. Vis-NIR data fusion – mean particle size and concentration

Particle size

The vis-NIR range is strongly affected by light scattering from the polymer beads. The light emitted to the particulate will change its direction every time it encounters a particle. Depending on the particle size, shape, and optical properties, the light will be scattered in a particular direction and magnitude. In a turbid medium, the scattered light from a particle will encounter other particles, which consequently will also scatter the light. These events lead to the light being scattered multiple times before it is collected. The light will travel differently depending on the nature of the system and depth of light penetration. SAR-DRM can provide a spectrum representative of the entire sample, as both angular sources and the detectors at different distances from the light source give information of different depths of light penetration, from a sample of specific set of scattering and absorption. As observed in Chapter 6, the information of each source-detecting fibre distances is difficult to interpret individually. Hence, we propose the use of combined measurement configurations to estimate particle size and concentration.

Table 6.3 summarises the first measurement configurations between detectors and angles evaluated, where D represents the detector distance at (1) 0.3, (2) 0.6, (3) 0.9 and (4) 1.2 mm from the light source.

In this section, the log transformation of the spectral data is used in order to assess whether it improves the linear relationship among the samples by transforming the spectra into a more comparable scale, leading consequently, to better calibration models.

The first stage of the analysis contains two steps: assessment of the non-processed and log-transformed SAR-DRM spectra impact on the model performance, and selection of two configurations from each light source. This is then followed by the selection of the best approach to combine the information from multiple configurations. Mean centring was applied to all datasets, however, no empirical pre-processing method was applied.

Table 6.3: Initial measurement configurations cases used to build PLSR models. These were constructed using spectra from different detector distances (D1 – 0.3 mm, D2 – 0.6 mm, D3 – 0.9 mm, D4 – 1.2 mm) from the light source L1 (0°), L2 (30°) and L3 (45°).

	Angular source		
Case nº	**L1**	**L2**	**L3**
1	D1		
2	D2		
3	D3		
4	D4		
5		D1	
6		D2	
7		D3	
8		D4	
9			D1
10			D2
11			D3
12			D4

Figure 6.7(a) shows all spectra acquired from the D1 – L1 configuration and its b) log transformation, to illustrate the changes made by the log transformation. While the raw spectra are visually more spread over a large intensity range, the log-transformed spectra show the variables in a more comparable scale.

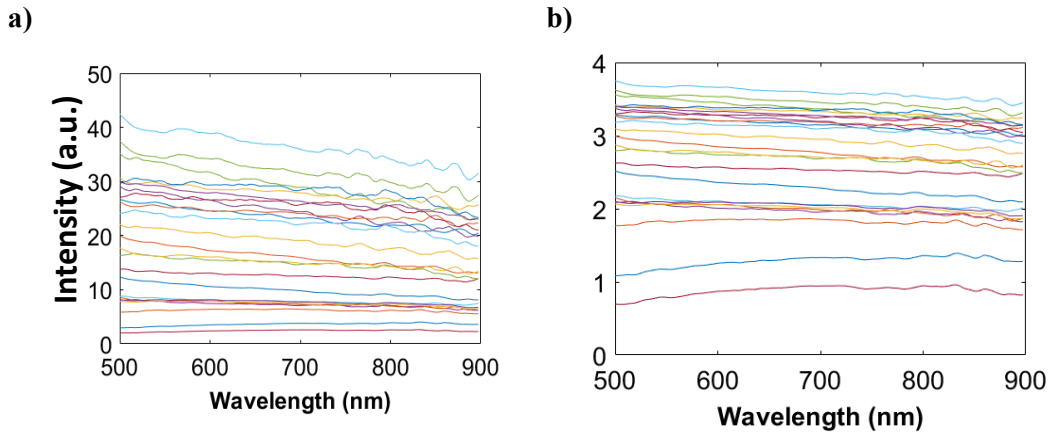


Figure 6.7: a) Raw vis-NIR spectra collect from the detector at 0.3 mm (D1) from the normal incident light and its **b)** log transformation, for the 25 samples analysed.

Figure 6.8 shows the comparison of the RMSECV curves for models built using raw data (solid line) and log-transformed data (dashed line) for each configuration. It is observed that models built on the log-transformed spectra exhibit a decrease in the overall RMSECV. Table 6.4 summarises the best model obtained from each configuration when using both forms of data.

When the models were built using the spectra acquired from the three detectors at 0.3 (D1), 0.6 (D2) and 1.2 (D4) mm from the normal incident light (Cases 1, 2 and 4), the use of log transformation not only lead to a decrease of the RMSECV but it also reduced the number of LVs needed to explain the model. Hence, an increase in the correlation coefficient was observed in these cases. In case 3, the use of log-transformed data led to an increase in the number of LVs, from 3 to 6. However, the RMSECV decreased 70%. Similar results are also observed from models built on log-transformed spectra when using angular incident light. For example, in case 10, the use of an additional LV in the model with log-transformed data resulted in a 27% decrease of the RMSECV and in an increase of the R^2 from 0.331 to 0.638. The cases number 9 and 12 needed more 8 and 6 LVs, respectively, to explain the model after transforming the data. However, the RMSECV decreased 68 and 67%, respectively, in comparison to the models built using the raw spectra.

The best model, with a lower error of calibration, fewer number of optimum LVs to explain the model and higher R^2, was found to be built only using the log-transformed spectra collected from the detecting fibre at 0.6 mm from the normal incident light, case 2.

Due to the significant size gap between the majority of the samples and PS630-800, and the weaker intensity from PS630-800 samples in comparison to the others, as shown in Chapter 5 Section 5.3.1., the model built using raw data might consider these spectra from PS630-800 outliers, which subsequently affect the regression model disproportionately. However, we know the data from these samples are based on a true signal. A log transformation reduces the influence of these points in the regression model, bringing spectra from PS630-800 closer to the centre of the dataset. The method normalises the residuals, consequently leading to a reduced number of LVs, RMSECV and higher R^2. One possibility to mitigate such issue is to include data from

particle size ranges between 355 and 600 µm in the model. However, in this study this was not possible due to the limited amount of sample available to run the experiments.

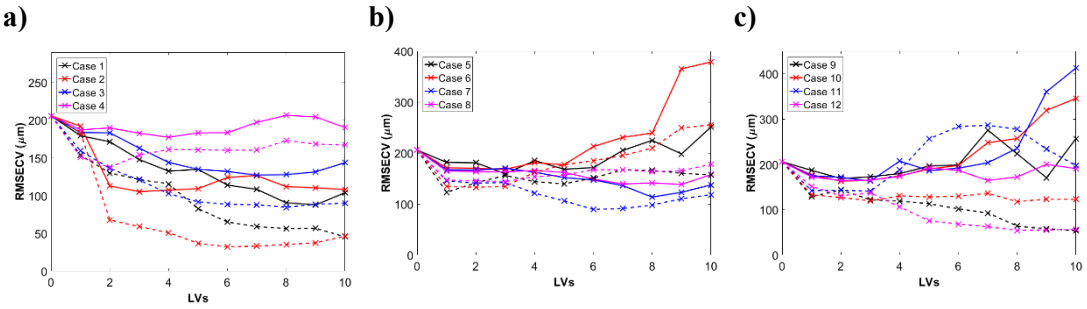


Figure 6.8: RMSECV versus LV obtained from the PLSR analysis using the spectra from each one of the detectors at different distances from the light source at **a)** 0°, **b)** 30° and **c)** 45°. The dashed line represents the RMSECV obtained when transforming the respective spectra (cases) in a logarithm scale.

Table 6.4: Summary of the results obtained from PLSR models when using the vis-NIR spectra and its log transformation, to determine PS mean particle size.

	No transformation			Log transformation		
Case No.	**No of LVs**	**RMESCV (µm)**	**R^2**	**No of LVs**	**RMESCV (µm)**	**R^2**
1	9	87.7	0.814	8	56.5	0.924
2	3	105	0.724	6	32.2	0.974
3	7	127	0.634	6	88.5	0.803
4	4	177	0.301	2	137	0.521
5	3	159	0.361	1	123	0.613
6	3	168	0.305	2	133	0.553
7	8	113	0.710	6	89.4	0.810
8	7	139	0.582	3	144	0.47
9	2	169	0.289	10	53.5	0.9266
10	2	163	0.331	3	119	0.638
11	3	163	0.339	1	142	0.484
12	2	164	0.331	8	54.4	0.926

Analysing only the RMSECV from the models using the log-transformed data, no clear trend between the source-detector distances is observed. It was expected that the shortest detector distances would lead to better model performances since the photons collected are less affected by scattering or absorption due to the shorter light path length taken, and result in higher reflectance intensities collected (see Chapter 5 Section 5.3). An increase in each detector distance increases the probability of absorption and scattering events due to the longer light path length travelled. This leads weaker reflectance intensities and to an increase in the model complexity. However, this also means that the light penetrates deeper in the sample, perhaps enhancing and unveiling chemical or physical information not readily available at the sample surface. Hence, the detecting fibre distances selection/performance can be a compromise between penetration depth and an adequate signal-to-noise ratio.

Combination of the information obtained spatially and angularly may offer a better representation of the sample scattering and absorption behaviour of the particulate system and consequently improve the model performance. For that reason, the measurement configurations using the log-transformation of the spectral data which led to the lowest RMSECV values or fewest LVs are selected to combine the information (see Table 6.4). For example, as the detecting fibres from L2 (cases 5-8), all exhibit a poor model performance, one of the detecting fibres selected to combine information was D1 – L2 (case 5), as it resulted in the fewer number of LVs and lower RMSECV. Two different approaches, data co-adding and data augmentation, are implemented in order to assess the influence of data fusion on the model performance. As the log-transformed data seems to consistently result in better model performance, it is the dataset selected to carry on with PLSR analysis.

Next, the findings so far will be taken to the second stage of this study: selection of the best approach to combine the spectra from multiple detectors. The data co-adding strategy works similar to some commercial particle size analysers, using the average spectrum for analysis. An example is the signal provided by conventional reflectance probes where the spectra from multiple detecting fibres is averaged and returned as a single-block data. The co-adding approach used in this study mimics the spectra normally obtained from these reflectance probes. On the other hand, the augmentation

approach enables the inclusion of the spectra obtained from different configuration in the same dataset, resulting in an increased number of variables and consequently in a long array.

Based on the analysis of each configuration and on the spatially-resolved feature provided by SAR-DRM, combinations of configurations are selected to further utilisation of the information evaluated. A summary of the selected combination is presented in Table 6.5. Cases 13-19, are related to configurations using one or multiple incident angles with all the spatial distance included. Cases 20-26 are selected based on the model performance for all available configuration analysed in Table 6.4. The measurement configurations are chosen by selecting the two detecting fibres from each light source with the lowest RMSECV values or fewest LVs to describe the model, to combine the information.

Table 6.5: Measurement configurations cases used to build PLSR models and compare the data co-adding and augmentation approaches. These were constructed using spectra from different detector distances (D1 – 0.3 mm, D2 – 0.6 mm, D3 – 0.9 mm, D4 – 1.2 mm) from the light source 0°, 30° and 45°.

	Angular source		
Case nº	**L1**	**L2**	**L3**
13	D1:D4		
14		D1:D4	
15			D1:D4
16	D1:D4	D1:D4	
17		D1:D4	D1:D4
18	D1:D4		D1:D4
19	D1:D4	D1:D4	D1:D4
20	D2	D3	D1 and D4
21	D2	D3	D1
22	D2		D4
23	D2	D1 and D3	
24	D2	D3	D4
25	D1:D2	D3	D1 and D4
26	D2	D3	

Table 6.6 summarises the results obtained from PLSR using co-added and augmented datasets, as shown in Figure 4.8. Taking into account the sample conditions analysed, which contain particles of different size, shape and concentration, the models require at least 3 LVs to explain the variance in the dataset. Models using less than 3 LVs are not considered to be robust or representative and therefore, are excluded for further analysis/comparison. Comparing the results in Table 6.6 and Table 6.4, there is not an improvement in the model performance to estimate particle size in combining all the detectors from the same or multiple light sources, more specifically, cases 13-19. The results show that including detecting fibres with poor performance deteriorates of the model performance. For example, case 2 led to lower RMSECV values using fewer LVs to explain the model built on spectra collected by one detecting fibre, D2 – L1, than case 13, which combined all the detectors, including the ones with poor performance, D1:D4 – L1. The use of D1:D4 – L3 should also be discarded as they take up 50 % of the information in cases 17 and 18, leading to poor calibration models. The only small improvement observed was when combining all the detectors from the angular source L2. The combination of information from key configurations is essential to improve the model performance, as observed in cases 19-26. The corresponding RMESCV curves are shown in Figure 6.9. In general, the RMESCV curves related to data augmentation seem to result in smoother LV curves with lower errors of calibration. In case 20, the data augmentation needed 3 more LVs than when using the co-adding approach. However, it resulted in a significant decrease in the RMSECV error, about 74 %. The LV curve of case 26 also seems to be much smoother than the LV curve from case 20 in Figure 6.8. In case 21, both model approaches needed 8 LVs to explain the variance in the model but a significant decrease of RMSECV in the data augmentation approach was observed. The cases 22, 24 and 25, resulted in a higher number of LVs using the data augmentation approach, but in smoother LVs curves and significant decrease of the RMESCV and increase of the correlation coefficient.

When combining multiple configurations, the best model performance was achieved using augmented data of D2 – L1 and D3 – L2, case 26. Although the lowest RMESCV and highest R^2 were achieved by using the data augmentation approach in the case 20, the model includes data from 3 additional configurations and needed 3 more LVs to

explain the model in comparison with the case 26. Considering the decrease of the RMSECV for case 20 from the one for case 26, it is not significant enough to justify the use of 3 extra LVs.

Table 6.6: Summary of the PLSR calibration model performance for estimating particle size using data augmentation and co-adding approaches in the log-transformed vis-NIR spectra.

	Augmented			Co-added		
Case No.	**No of LVs**	**RMESCV (µm)**	**R^2**	**No of LVs**	**RMESCV (µm)**	**R^2**
13	7	40.0	0.959	8	54.5	0.924
14	3	108	0.704	2	129	0.576
15	2	103	0.729	10	40.0	0.959
16	8	66.9	0.889	3	122	0.619
17	3	80.4	0.836	8	50.8	0.935
18	7	61.9	0.903	3	119	0.647
19	5	75.6	0.856	7	98.3	0.764
20	8	26.3	0.983	5	102	0.761
21	8	26.9	0.982	8	63.8	0.906
22	8	27.6	0.98	6	58.7	0.913
23	7	29.7	0.978	8	70.4	0.880
24	8	31.0	0.977	5	98.8	0.758
25	9	29.5	0.980	7	76.9	0.849
26	5	30.2	0.977	8	64.0	0.906

The model performance using the data augmentation approach seems, in general, to improve the model performance. These findings are in agreement with the work developed by Chen *et. al*., which used SAR-DRM in polystyrene suspensions in the nanometre range [9]. The study indicates that extra information can be utilised when using the data augmentation approach and that in some cases, the addition of some angular sources can improve the model performance. This can be due to the significant number of extra variables included in the model provided in relation to the co-adding

approach, where the variables are treated as extra information. Moreover, the distinct differences between the spectra can be preserved. Averaging all the spectra from different detectors distances and light sources can result in a loss of the spectra characteristic features and deteriorate the model performance. In our study, the model performance is dependent not only on the data approach used but also on the measurement combination selected. The model performance can be improved by finding the optimum detectors-sources combinations to utilise the data.

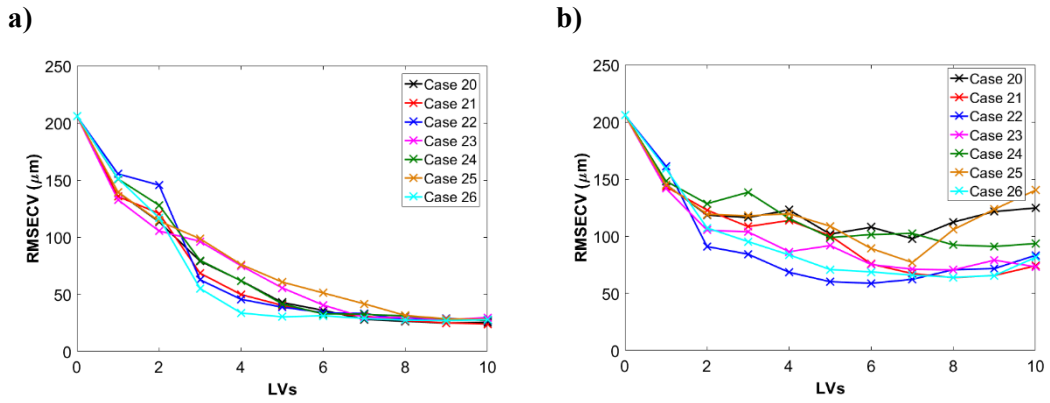


Figure 6.9: Cross-validation curves of PLSR models built for estimation of particle size using **a)** data augmentation and **b)** co-adding approach for cases 20-26.

The third stage of this study is to examine whether the model performance can be improved for cases 1-6 by adding empirical pre-processing to the augmentation and co-adding approaches. The pre-processing treatments used in this study were SNV, MSC and ISC. The SNV is typically used to regularise the scaling of the spectra while MSC is used to remove the spectra slope and its baseline variation. The ISC is simply the inversion of MSC. For the data augmentation approach, the pre-processing was applied to each detector (block) separately to maintain the information collected by each one of them. For the co-added approach, the pre-processing was implemented before the co-adding step. In order to simplify the discussion, only the best model performances achieved for each one of the pre-processing methods are shown and compared. Table 6.7 shows the comparison of models built with and without applying pre-processing methods for both data fusion approaches.

The best model performance was obtained by applying SNV when combining the spectra collected from all the detectors at different distances from the normal incident light and all the detectors collecting information from the angular source at 45°, using the data augmentation approach. This correction resulted in a lower number of LVs to explain the data variance, lower error of calibration and higher R^2. However, this decrease in the RMSECV error does not seem significant compared to when no pre-processing is applied. The cross-validation curves of the respective PLSR models show that both display a relatively smooth curve, as shown in Figure 6.10. Figure 6.11 shows the estimated versus actual particle size and the residuals plots obtained for each model. The model built on none pre-processed data shows similar prediction error for all samples, as shown in Figure 6.11(b1), while the model built on SNV treated data exhibit an outlier (5 wt.% of PS250-355) in Figure 6.11(b2).

Table 6.7: Impact of pre-processing methods (SNV, MSC and ISC) on calibration models built to estimate particle size. The table presents the cases (measurement configurations) in which the pre-processing method resulted in the lowest RMSECV using data augmentation and data co-adding approaches.

	Augmented				Co-added			
Pre-processing method	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**
None	**26**	5	30.2	0.977	**17**	8	50.8	0.935
SNV	**18**	4	27.8	0.980	**26**	4	89.7	0.845
MSC	**26**	7	43.2	0.953	**26**	5	83.6	0.828
ISC	**13**	5	43.1	0.976	**16**	4	40.9	0.958

Usually, MSC and SNV present similar results for all applications [106, 107, 146]. However, in comparison to MSC, the SNV showed a significant improvement, reducing the 35% of RMSECV and using fewer latent variables in the data augmentation approach. In the data augmentation approach, a similar error in the calibration model is observed from ISC and MSC correction method, but the ISC treated data requires fewer LVs to describe the calibration model.

In general, the co-adding approach did not lead to an improvement of the model performance when applying different pre-processing methods. As the best model performance obtained by using SNV and MSC approaches was using the same data combination approach, the impact of these two pre-processing techniques was similar, as it was expected. Interestingly, a significant improvement, >50%, was observed was noticed when using the ISC method. Comparing Table 6.6 where case 16 using co-added approach without pre-processing was considered invalid due to the low number of LVs, the significant improvement in the model performance using the ISC pre-processed data is peculiar. This highlights the importance of taking a systematic approach to appropriately analyse the model performance observed. In this particular case, one may conclude that the model does not contain strong basis for its performance, and will be more likely to struggle to predict any future sample.

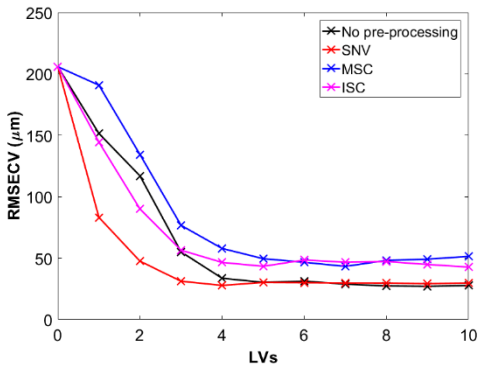


Figure 6.10: Cross-validation curves of PLSR models built for the estimation of particle size using a data augmentation approach on vis-NIR data treated by different pre-processing techniques.

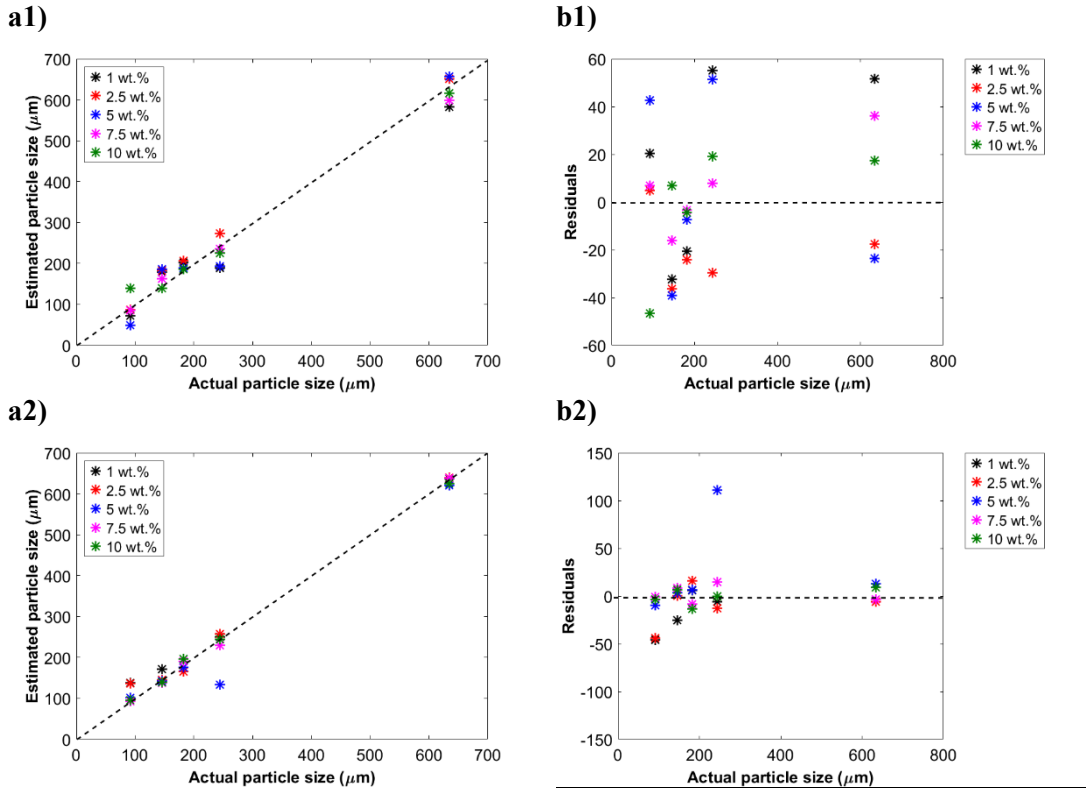


Figure 6.11: a) Estimated versus actual particle size and **b)** residuals in the vis-NIR calibration from **(1)** case 26 (none pre-processed data) and **(2)** case 18 using the SNV correction.

Particle Concentration

The same approach and steps were used to estimate particle concentration from the vis-NIR spectra collected from SAR-DRM. In chapter 6, it was clearly observed that an increase of solid loading resulted in an increase in the light intensity collected from each one of the detectors. In this chapter, it is assessed whether this is a consistent pattern and if modelling strategies are able to capture that information.

In order to select the best configurations to estimate particle concentration, regression models were initially developed for individual configuration (each source-to-detecting fibre combination, see Table 6.3. The results could indicate suitable configurations to use a posteriori. Differently from what was conducted in the particle size estimation study, the log transformation was not applied to datasets used to estimate particle concentration. This is because of the smaller concentration range and similar concentration increment in the sample condition. To be certain, the initial analysis was

conducted and showed that higher RMSECV errors are obtained from models using the log-transformed data. Hence, the data analysed in this section uses the raw spectra. No pre-processing method was initially used. Table 6.8 summarises the PLSR results from each source-to-detecting fibre combination. For the concentration estimation, 2 main independent variables, particle size and shape, are associated with each particle group. Changing the particle concentration for a given particle group only increase particle density in the suspension, but not the particle size or shape distribution. Therefore, in theory a model should require at least 2 LVs. Models using fewer than 2 LVs will be excluded from further analysis. Overall, the estimation of particle concentration using the spectra collected from single configuration resulted in poor calibration models with relatively high errors in comparison with the range of concentrations used to build the model (1, 2.5, 5, 7.5 and 10 wt.%) and low correlation coefficient. Similarly to the results from particle size estimation (Table 6.4), no particular pattern related to the detecting fibre distance and its performance was found. For the normal and 45º incident light, the 2 detecting fibres which presented the lowest calibration errors were selected for further analysis. Due to the poor performance of all detecting fibres at different distances from the 30º angular source, it was selected the detecting fibre with the lowest calibration error and one with the best compromise among RMSECV and number of LVs.

As the angular incident light travels shallower in the samples, as illustrated in Figure 4.2(b), the closest detecting fibres might be collecting the spectra after incident light undergoes single or a few scattering. This effect would have an even higher impact when using larger particles in the suspension, as the number of particles is much lower in comparison to smaller sizes regardless the concentration used.

Table 6.8: Summary of the results obtained from PLSR models when using the original vis-NIR spectra to determine PS beads concentration.

Case No.	No of LVs	RMESCV (wt.%)	R^2
1	6	1.81	0.707
2	3	1.61	0.767
3	2	2.25	0.540
4	3	2.44	0.459
5	5	2.27	0.529
6	3	2.44	0.457
7	3	2.49	0.437
8	5	2.24	0.572
9	5	2.54	0.417
10	4	2.60	0.421
11	8	1.51	0.795
12	7	1.44	0.812

In order to improve the model, different combinations of configurations are analysed using data augmentation and co-adding approaches. Table 6.9 summarises the measurement combinations used to build PLSR models to estimate particle concentration.

Table 6.10 summarises the PLSR analysis results for estimating particle concentration using the cases summarised in Table 6.9 to form the augmented and co-added datasets approaches for modelling. Overall, the data co-adding approach requires fewer LVs to describe the best performing model but poor models performance in comparison with the model on augmented data. Analysing the Cases 13-19, it is observed that for the data augmentation approach, the models with an improvement in the model performance from the single detectors are the cases that include all the detecting fibres from the normal incident light, Cases 13, 16, 18 and 19, even when combined with other configurations. Meanwhile, in the co-adding approach, the addition of angular configurations to the D1:D4 – L1 combination led to poor performance models in all

cases. This suggests the strong influence of the detecting fibres D1 – L1 and D2 – L1 on the PLSR model to estimate particle concentration.

Table 6.9: Measurement configurations cases used to build PLSR models and compare the data co-adding and augmentation approaches for particle concentration estimation. These were constructed using spectra from different detector distances (D1 – 0.3 mm, D2 – 0.6 mm, D3 – 0.9 mm, D4 – 1.2 mm) from the light source 0°, 30° and 45°.

	Angular source		
Case nº	**L1**	**L2**	**L3**
13	D1:D4		
14		D1:D4	
15			D1:D4
16	D1:D4	D1:D4	
17		D1:D4	D1:D4
18	D1:D4		D1:D4
19	D1:D4	D1:D4	D1:D4
20	D2	D2	D3
21	D2		D3:D4
22	D2	D2	D3:D4
23	D2		D3
24	D1:D4	D1:D4	D3
25	D2		D2
26	D2	D2	D3

The cases 20-26 mainly use data from D2 – L1 (case 2) in combination with configurations using angular sources with relatively good model performance. These configurations/cases are based on the summary of the best model performance achieved when analysing all the possible combinations among the single detecting fibres selected previously. In these cases, it is observed an overall improvement of the model performance. Figure 6.12(b2) shows that the profile of the RMSECV curves become smoother in these cases in comparison to the Cases 13-19 for the data augmentation approach, and to all the cases using the co-adding approach (Figure 6.12(a1) and Figure 6.12(b2)-(b2), respectively). However, it is also observed in Table 6.10 that the best models using the augmented data require either an equal number of LV or in a higher number of LVs when compared to the co-adding

approach. This might be due to the competing influence of different configuration in reducing the overall error which subsequently requires additional latent variables to compensate the extra effect in the model.

Table 6.10: Summary of the PLSR calibration model performance for estimating particle concentration using data augmentation and co-adding approaches.

	Augmented			Co-added		
Case No.	**No of LVs**	**RMESCV (wt.%)**	**R^2**	**No of LVs**	**RMESCV (wt.%)**	**R^2**
13	8	1.10	0.889	8	1.37	0.827
14	2	2.88	0.250	6	2.41	0.538
15	4	2.74	0.392	4	2.50	0.536
16	8	0.82	0.938	2	2.19	0.563
17	4	2.76	0.381	4	2.50	0.432
18	10	1.31	0.842	4	2.32	0.512
19	10	1.07	0.895	4	2.39	0.485
20	8	0.73	0.952	6	1.38	0.827
21	8	0.80	0.941	8	1.23	0.867
22	8	0.87	0.931	7	1.54	0.79
23	6	0.87	0.931	6	1.19	0.870
24	10	0.70	0.956	3	2.23	0.550
25	6	0.98	0.911	5	1.68	0.747
26	8	0.89	0.927	4	2.22	0.555

The RMSECV curves for the co-adding data is shown in Figure 6.12(b1)-(b2) and exhibit a jaggy curve with the cases 14-19, 24 and 26 exhibiting an almost flat profile. A closer examination of the loading plots revealed that majority of the cases after the 4^{th} LV are modelling only noise. In general, the first LV captures the major variance of the data as observed in Figure 6.13(d), which shows case 19 as an example. This variance seems to be mainly related to the smaller particle size ranges analysed as

observed in Figure 6.13(e1), and it shows a similar slope to the one observed in Figure 5.16 in Section 5.3.1. In chapter 6 was also observed a change in the spectral slope with the particle size and aspect ratio of the particles, which seem to be captured by the LV 2, 3 and 4 in Figure 6.13(d1). The particle size and shape have an effect on the spectra and therefore are interconnected with particle solid loading. These LVs are mainly capturing the variance from higher particle concentrations of the samples with a particle size ranging from 125 to 355 µm, as displayed in Figure 6.13(e2). The remaining LVs 5, 6, 7 and 8, shown in Figure 6.13(d3) seem to be only capturing noise from the 3 largest particle size groups. The lowest RMSECV value obtained by the co-adding approach uses the data from D2 – L1 and D3 – L3 (Case 23).

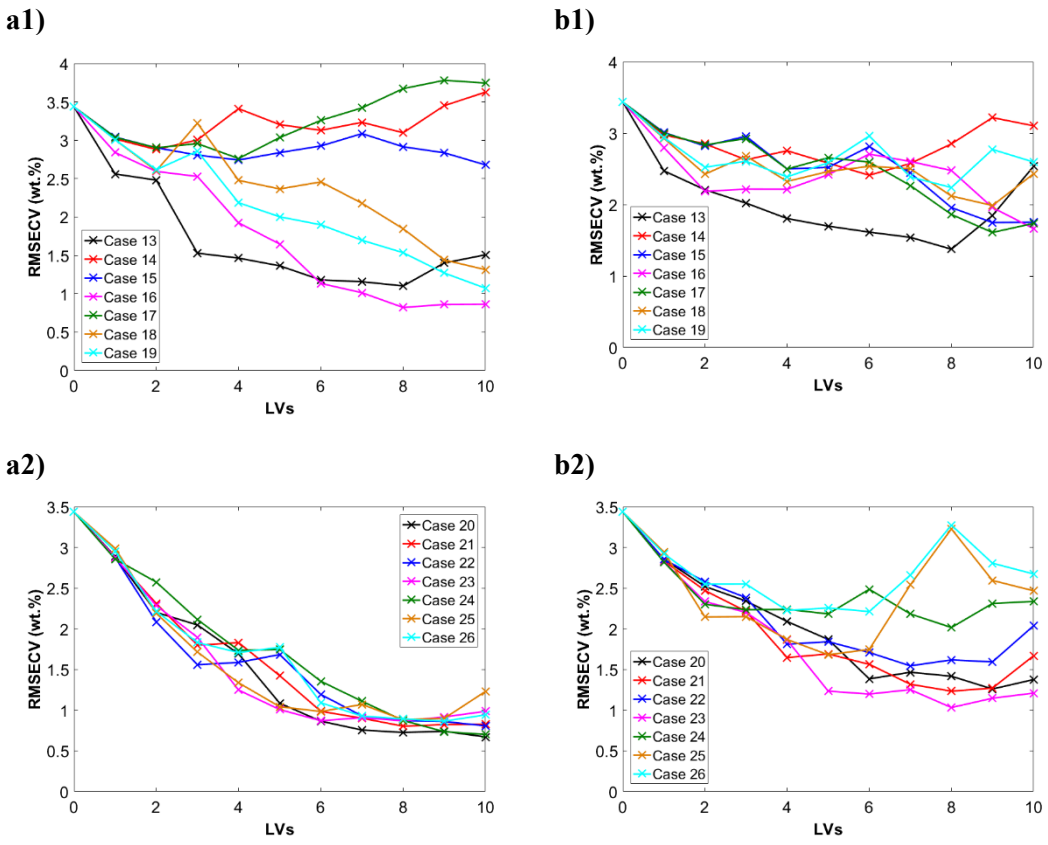


Figure 6.12: Cross-validation curves of PLSR models built for estimation of particle concentration using a**)** data augmentation and **b)** co-adding approaches for **(1)** cases 13-19 and **(2)** cases 20-26.

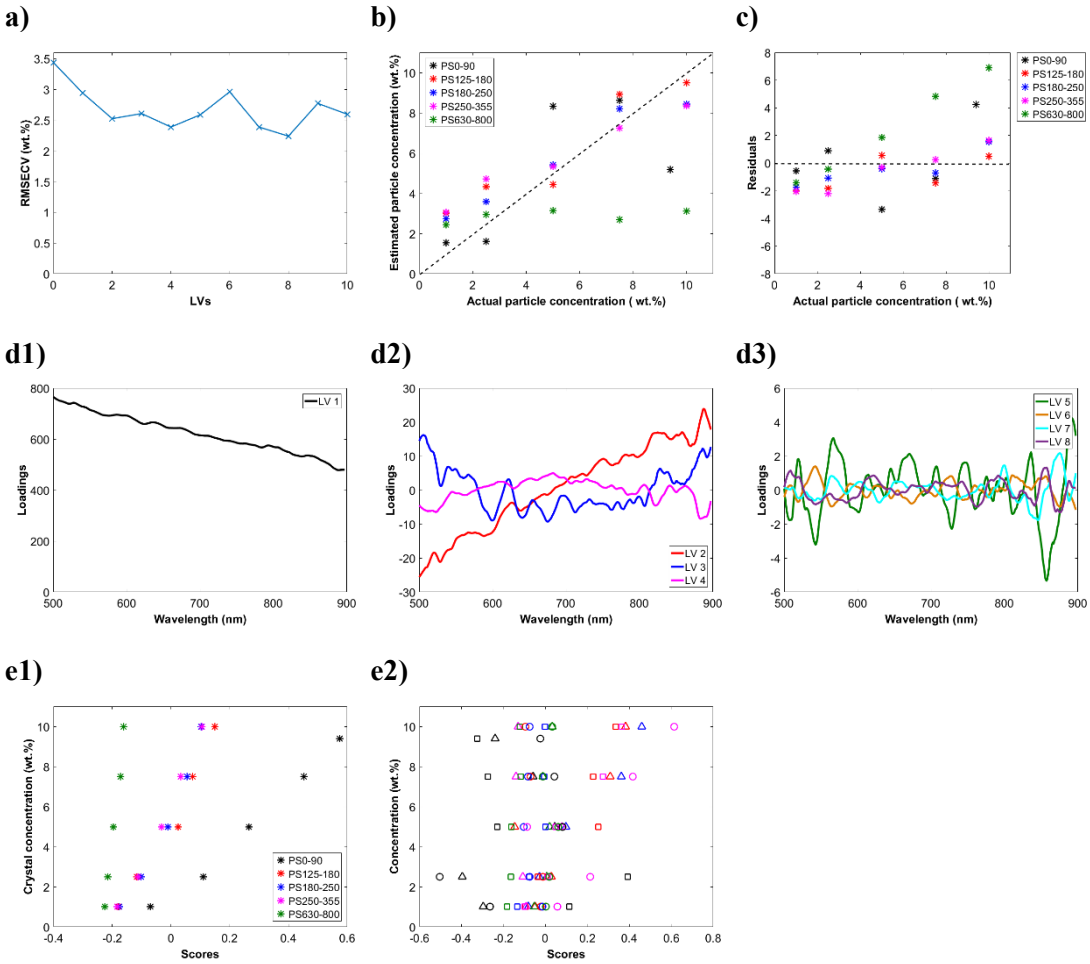


Figure 6.13: a) RMSECV profile versus LVs. **b)** Estimated versus actual particle concentration and **c)** residuals in the calibration of case 19. **c)** Loading plots of case from **(1)** LV 1, **(2)** LV 2, 3, and 4 and **(3)** LV 5, 6, 7 and 8, from the calibration model built for particle concentration estimation using the data co-adding approach. e) Score plot for the respective LV. The open triangle, square and circle in (e2) correspond to LV 4, 3 and 2, respectively

The best model performance was obtained from Case 20, which estimates the particle concentration with an error of 0.73 wt.% and an R^2 of 0.952. Although a lower RMSECV value was achieved in Case 24, there was not a significant improvement. Moreover, Case 24 requires 2 additional LVs. Figure 6.14 presents the loading and score plots for Case 20. The first loading seems to capture the variance from the sample concentration with smallest particle size range analysed, from each one of the detectors used, as shown in the respective scores in Figure 6.14(b1). The second LV seems to capture variance from the largest particle size groups, PS250-355 and PS630-800. For the largest group, it is mainly capturing the variance from high solid loadings, 7.5 and

10 wt.%. The next few LVs continue to capture the remaining features in the data with more complicated LV curves and score plots. The last LV (LV8) seems to adjust the remaining concentration-related information as a negative score is given largely to the smallest and largest concentrations while the medium concentrations were given a positive score. However, it is worthy to note that the small magnitude of LV8 in Figure 6.15(a2), making any correlation found in Figure 6.15(b4) a very small improvement to the overall RMSECV. Figure 6.15 shows the prediction and residuals plots for case 20. The model seems to have difficulty to estimate the higher solid concentrations for PS125-180, which translates into larger residuals. On the contrary, the prediction for PS250-355 samples also shows large residuals for 2.5 and 5 wt.%. The slightly U-shaped residuals distribution suggests the underlying variation could be non-linear and further improvement would be best achieved using a non-linear model.

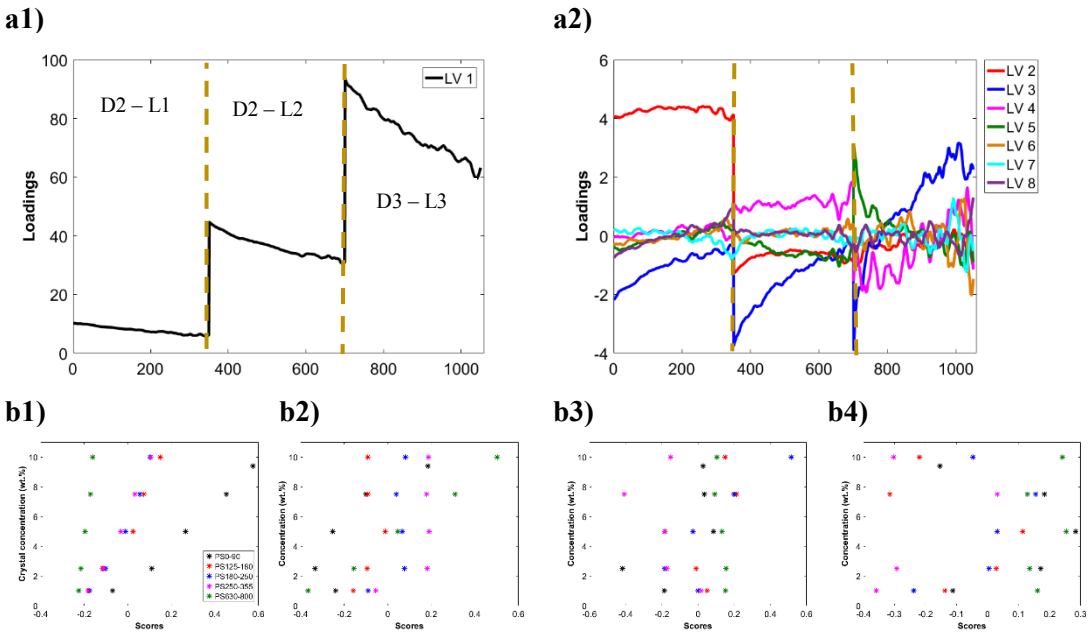


Figure 6.14: Loading plots of case 20 from **a1)** LV 1, **a2)** LV 2-8, from the calibration model built for particle concentration estimation using the data augmentation approach. Scores for **b1)** LV 1, **b2)** LV 2, **b3)** LV 3 and **b4)** LV 8.

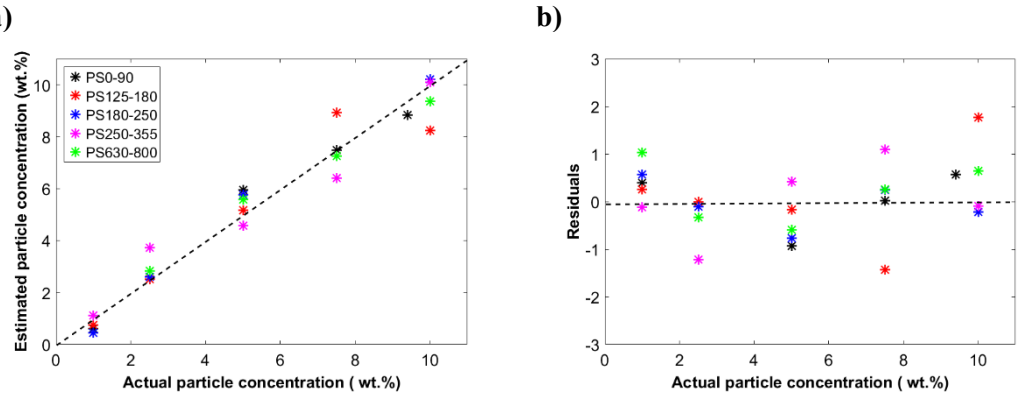


Figure 6.15: a) Estimated versus actual particle concentration and **b)** residuals in the calibration of case 20.

The impact of using different pre-processing methods was also investigated. Table 6.11 summarises the PLSR results on the calibration models for estimating particle concentration on SNV, MSC and ISC pre-processed dataset. Both data augmentation and co-adding approaches were applied to Cases 13-26. Similarly, the observation on analysing particle size estimation, only the best results achieved from the measurement combinations described in Table 6.10 are presented. The results show that the application of pre-processing methods to estimate particle concentration deteriorates the PLSR models. The reason for that is the impact of particle size and shape have in spectra. As observed in Figure 6.16 (a1) and (a2), for the data augmentation approach, the use of SNV as a correction method not only enhances the noise, but also causes alteration in the spectral shape as seen from the sample PS630-800. The latter could be due to the interference of noise in the SNV formula as the correction method uses the own mean value and standard deviation. The correction parameters could be significantly affected by the poor quality of the signal. The impact of these 5 samples is even greater as it results in a sudden change of pattern. More particles within a size range between 350-630 μm should be incorporated in the model, using different concentrations, for a better particle concentration estimation. The use of MSC and ISC pre-treatments in cases 15 and 18, respectively, seem to mainly amplify the noise as observed in Figure 6.16(b2) and (c2). As the vis-NIR spectra collected is featureless, the PLSR algorithm may consider the noise enhanced by the pre-processing as a feature when building the model hence, leading to the unwanted feature to being

captured wrongly in the first few LVs and deteriorate the overall performance. Thus, the application of pre-processing methods to estimate particle concentration for this particular sample in this vis-NIR region is not suitable. Non-linear methods could be implemented in the future to improve the calibration models. For the purpose of this study, only linear methods were assessed.

Table 6.11: Impact of pre-processing methods (SNV, MSC and ISC) on calibration models built to estimate particle concentration. The table presents the cases (measurement configurations) in which the pre-processing method resulted in the lowest RMSECV using data augmentation and data co-adding approaches.

	Augmented				Co-added			
Pre-processing method	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**
None	**20**	8	0.73	0.952	**23**	6	1.19	0.870
SNV	**21**	6	2.54	0.490	**22**	2	3.22	0.123
MSC	**15**	1	3.01	0.175	**17**	1	3.15	0.134
ISC	**18**	3	2.69	0.388	**13**	8	2.33	0.564

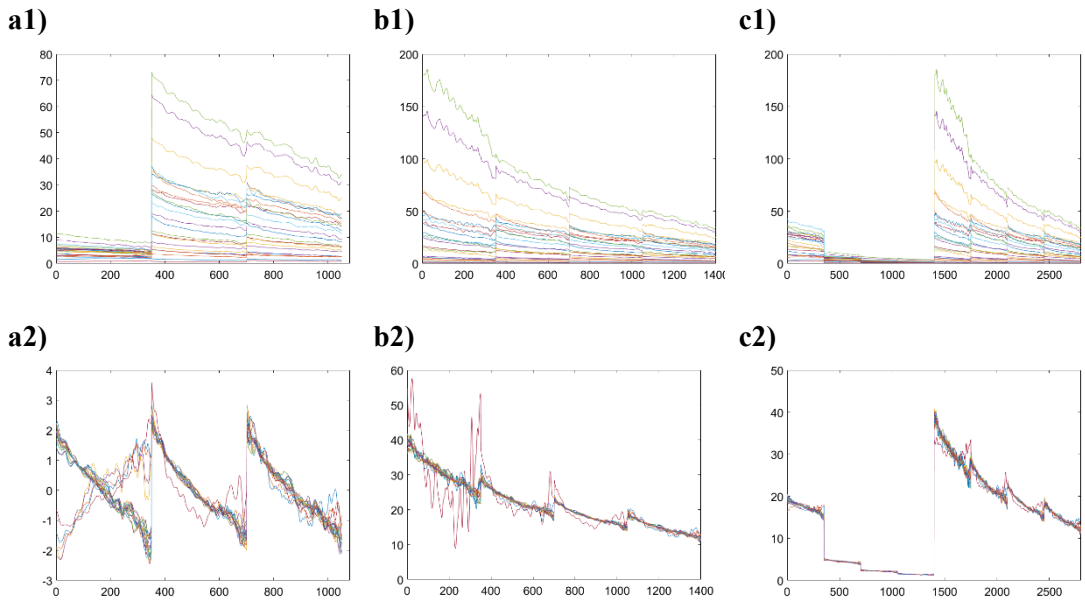


Figure 6.16: Augmented data from **a1)** case 21, **b1)** case 15 and **c1)** case 18. Data corrected with **a2)** SNV, **b2)** MSC and **c2)** ISC.

6.1.2.2. NIR data fusion – mean particle size and concentration

Figure 6.17(a1) and (b1) show the typical NIR reflectance spectra obtained using different particle size ranges with a solid loading of 10 wt.% and the effect of particle concentration of PS0-90, respectively, from the detector at 0.3 mm from the normal incident light. Figure 6.17(b1) and (b2) shows the respective spectra when the reflectance signal is linearized to absorbance using the formula: $A = \log_{10}(1/R)$. The negative values are related to the original intensity values from the measurement. This is due to a constant scaling factor which corresponds to the surface area of the integrating sphere not being accounted in the Equation 4.1. The complete calibration should account for the internal surface since the signal collected from the integrating sphere only corresponds to a small fraction of light which is distributed evenly inside the sphere. However, in practice, this factor is difficult to obtain due to the probe end surface not being the same material as the sphere, the small gap between the probe and the inner wall of the sphere, and the unknown surface area of the internal baffle inside the sphere. As a result, the surface area factor is omitted from Equation 4.1, hence the reflectance can not be scaled to the theoretical range between 0 to 1 which subsequently affects the transformation using the Lambert-Beer law and results in negative values. This approximation to the Lambert-Beer, which states the concentration of a substance is directly proportional to the absorbance of the solution, may lead to better linear regression. In Figure 6.17(b1) and (b2), it is observed a dominating band at 1400-1600 nm from the overtone of water, O-H. However, the water absorption is confined to this region and does not affect the polystyrene fingerprint. The NIR absorption bands related to the presence of polystyrene are mostly from the aromatic C-H overtones. Weak absorption bands are observed at 1640 and 1684 nm, from the first overtone of the aromatic C-H stretch. A second overtone band from the aromatic C-H and C-H_2 stretch is identified at ~1200 nm.

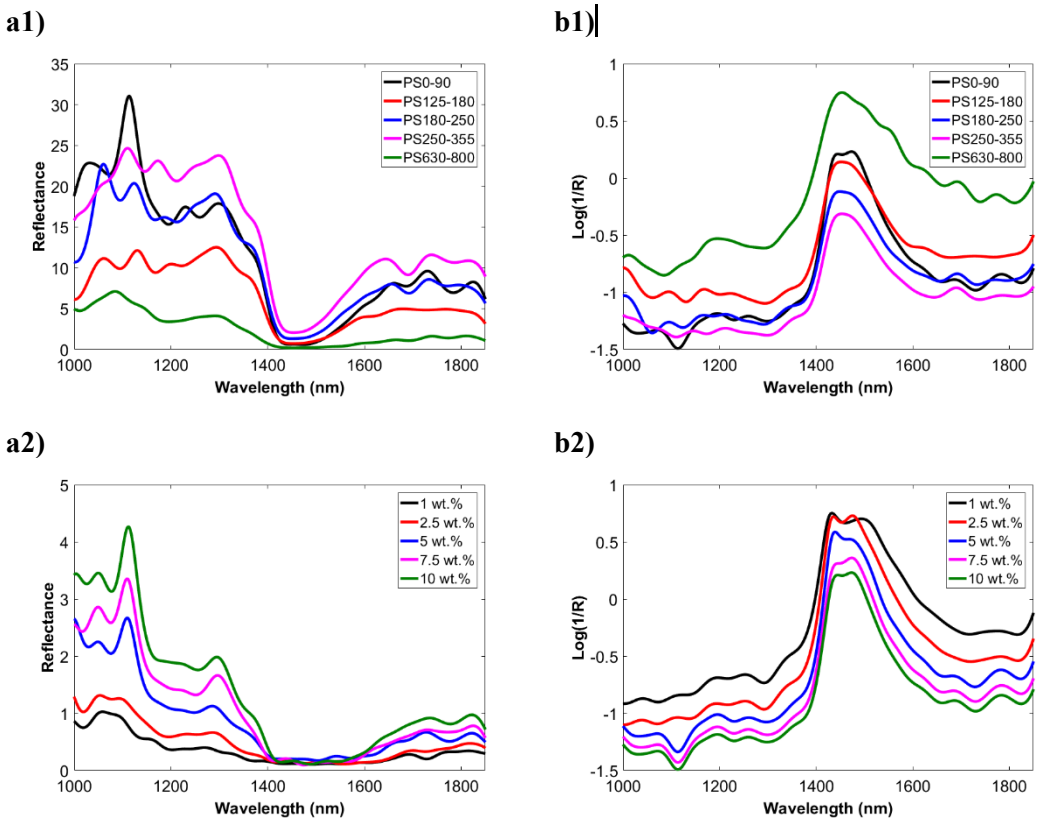


Figure 6.17: NIR **a)** reflectance spectra and respective **b)** absorption spectra for different **(1)** particle size ranges using suspensions of 10wt.% and **(2)** particle concentration using the PS0-90 sample.

Particle size

In Figure 6.17(b1) is observed the effect of particle size results in a non-linear pattern with a baseline shift. The effect of particle concentration in Figure 6.17(b2) leads mainly to a baseline shift, to an increase of the bands related to the polymer and to a decrease of the water band. According to Pasikatan *et al*., the contribution of a size range to the reflectance NIR spectra from powders, is approximately weighted in terms of its volume or mass fraction, hence powder mixtures with higher solid loading tend to lead to better prediction models [63]. Due to the smaller number of samples analysed, the wide distribution of particle size and shape, and the gap for particles with a size range from 355-630 µm, it is expected to obtain relatively high errors of calibration. Table 6.12 summarises the PLSR analysis results from models built on all the available configurations, as described in Table 6.3 Comparing the PLSR results

obtained from the original vis-NIR spectra (Table 6.4) with the results from the absorbance NIR spectra (Table 6.12), it is observed a decrease in the calibration error for all the cases analysed. This suggests that the NIR region may be more suitable to distinguish samples with different conditions and that the results may be improved by combining information from different configurations. When using the normal (cases 1-4) and the 30° incident light (Cases 5-8) the detecting fibres with the best performance for estimating particle size, i.e., lower RMSECV and higher R^2, are the ones at the shortest distance (0.3 and 0.6 mm) from the light source, cases 1 and 2, and cases 5 and 6, respectively. respectively. The use of extra LVs to explain the variance in these models leads to a significant improvement to the calibration error. These results are in agreement with the results using the vis-NIR spectra (Table 6.4: Cases 1-4), in which the closest fibres exhibited better results than the farthest detectors from the light source. This pattern is not found when the light is emitted at 45°. When analysing the model performance from the detectors at different distances from the 45° source, it is observed that the detectors obtaining the lowest calibration errors, 50.0 and 44.4 μm, are found at 0.6 and 1.2 mm from the light source, respectively. These models are also the most complex from the group, using 9 LVs to explain the data variance. To investigate the measurement combination in the next stage, two detectors were selected, one based on the lowest RMSECV acquired and the other based on the model simplicity. Hence, it was selected the detector at 0.3 mm and 1.2 mm from the 45° light source to investigate further. The model developed using the data captured from the detector at 1.2 mm only needed 3 LVs to explain the data variance, resulting in an error of 61.9 μm and in a coefficient of correlation of 0.902, exhibiting better performance than the models built for the closest detectors to the normal incident light. Although this model is not robust, as similar to the vis-NIR results, the models need at least 3 LVs to explain the independent variables, particle size, shape and concentration that affect the sample conditions. However, in this work it was assessed the impact of using such conditions to combine information and improve the model performance.

Overall, the best model performance was obtained when using the 45° angular source and capturing the spectra by the farthest distance, 1.2 mm, which needed 9 LVs to describe the model with an error of 44.4 μm and coefficient of correlation of 0.950, indicating a reasonably good fit.

Figure 6.18 shows the RMSECV versus LVs profile from the cases 1-12 when using the original NIR spectra to estimate mean particle size. Overall, it is exhibited a relatively smooth profile, differently from the models developed using the vis-NIR spectra which displayed an almost flat profile, as shown in Figure 6.8.

Table 6.12: Summary of the results obtained from PLSR models when using the original NIR spectra to estimate mean particle size.

Case No.	No of LVs	RMESCV (µm)	R^2
1	6	68.5	0.883
2	6	85.3	0.816
3	5	102.7	0.773
4	5	100.7	0.756
5	8	48.7	0.949
6	4	71.1	0.873
7	5	72.7	0.866
8	3	87.6	0.806
9	3	61.9	0.902
10	9	50.0	0.945
11	6	76.9	0.849
12	9	44.4	0.950

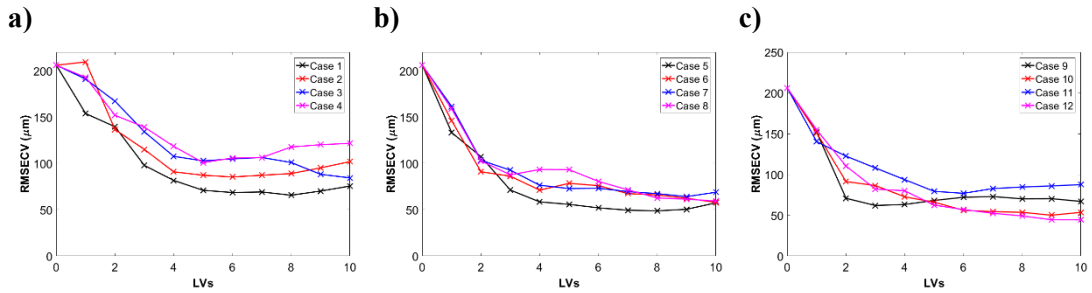


Figure 6.18: RMSECV versus LV obtained from the PLSR analysis using the absorbance NIR spectra from each one of the detectors at different distances from the light source at **a)** 0°, **b)** 30° and **c)** 45°.

In order to identify the influence of combining data from different configurations on the model performance, the same analysis to what was performed on the vis-NIR spectra was carried out. Table 6.13 summarises the measurement combinations selected to investigate the model performance on the augmented and co-added data. Cases 13 to 19 are the same combinations studied in the vis-NIR region. Cases 13-15 combine all the configurations using the normal, 30 and 45° incident light, respectively; Case 16-19 are the different combinations of Case13-15. To simplify the discussion, the cases from 20 to 26 are based on the best performing configurations selected from Table 6.12.

Table 6.13: Measurement configurations cases used to build PLSR models and compare the data co-adding and augmentation approaches for particle size estimation using the absorbance NIR spectra. These were constructed using spectra from different detector distances (D1 – 0.3 mm, D2 – 0.6 mm, D3 – 0.9 mm, D4 – 1.2 mm) from the light source 0°, 30° and 45°.

	Angular source		
Case nº	**L1**	**L2**	**L3**
13	D1:D4		
14		D1:D4	
15			D1:D4
16	D1:D4	D1:D4	
17		D1:D4	D1:D4
18	D1:D4		D1:D4
19	D1:D4	D1:D4	D1:D4
20	D1:D2	D1:D2	D4
21	D1:D2	D1:D2	D1 & D4
22	D1:D2	D1	D1 & D4
23	D1:D2	D1	D4
24	D2	D1:D2	D1 & D4
25	D2	D1	D4
26	D2	D1:D2	D4

Table 6.14 summarises the PLSR models on estimating particle size using augmented and a co-added from the configurations indicated in Table 6.13. Overall, the data augmentation approach resulted in more complex models, i.e., it uses more LVs to explain the data variance, with better performance (lower RMSECV and high

correlation coefficient) than the co-adding approach. These findings are in agreement with the results obtained from the models built using the vis-NIR data in section 6.2.2.1. Figure 6.19 presents the RMSECV curves for PLSR models on cases 13-26. The curves from the data using the augmentation approach exhibit relatively smoother and more stable profile than the curves obtained from data using the co-adding approach.

Table 6.14: Summary of the PLSR calibration model performance for estimating particle concentration using data augmentation and co-adding approaches on absorbance NIR data.

	Augmented			Co-added		
Case No.	**No of LVs**	**RMESCV (µm)**	**R^2**	**No of LVs**	**RMESCV (µm)**	**R^2**
13	8	53.0	0.928	6	81.9	0.829
14	9	24.7	0.985	3	88.7	0.800
15	5	49.2	0.942	10	51.9	0.932
16	9	34.7	0.970	5	59.9	0.910
17	8	30.6	0.978	9	47.42	0.945
18	7	49.1	0.939	8	52.6	0.929
19	9	31.5	0.977	9	48.6	0.940
20	9	16.1	0.994	6	48.7	0.940
21	8	17.7	0.992	8	44.0	0.952
22	8	19.0	0.991	5	52.8	0.929
23	8	19.1	0.991	6	51.2	0.933
24	9	19.9	0.991	9	33.3	0.978
25	9	19.8	0.990	9	39.1	0.962
26	9	21.9	0.989	10	41.3	0.960

By simply combining different configurations, the model performance improves considerably when compared to the single configurations, regardless of the approach used. Combining the best performing single configurations only improved the model

performance if augmentation approach is used. Interestingly, Cases 20-26 show a similar RMSECV curve profile when the same approach is used, as observed in Figure 6.19. This suggests that the model performance is not caused by an improvement of signal-to-noise ratio, as the spectral quality from each detecting fibre is maintained in the augmented approach, and consistent results are obtained regardless the combination of the configurations. Usually, co-adding approach is used to improve the signal-to-noise ratio, as it averages all the spectra, however, the combinations of selected configurations did not improve the general performance for models using the co-added data. In both data augmentation and co-adding approaches, a similar RMSECV profile was observed for the different cases 20-26, as shown in Figure 6.19(a2) and (b2). The similarity could be due to similar configurations being used to build the model, leading to more robust and similar models.

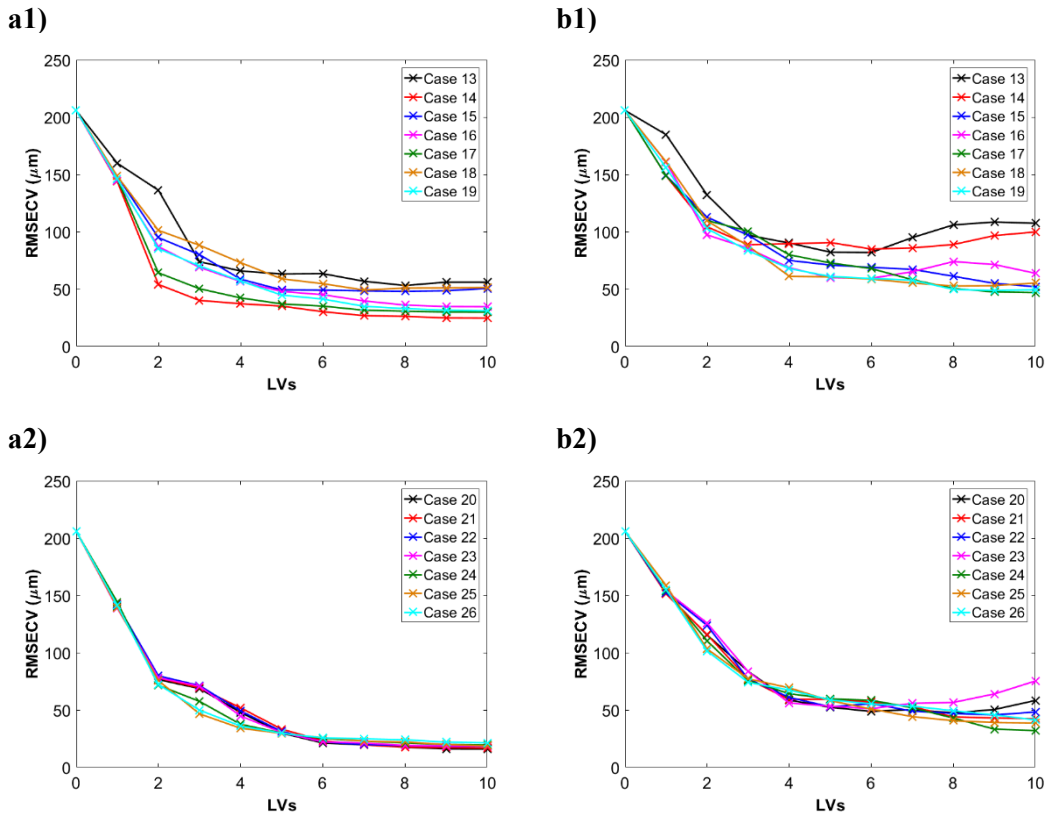


Figure 6.19: Cross-validation curves of PLSR models built for estimation of particle size using **a)** data augmentation and **b)** co-adding approach for **(1)** cases 13-19 and **(2)** cases 20-26.

The best model performance was obtained when combining, through a data augmentation approach, the absorbance NIR spectra from the 2 closest detectors (at 0.3 and 0.6 mm) to the normal incident light with the 2 closest detectors to the 30° source and with the farthest detector (at 1.2 mm) to the 45° source. Figure 6.20 illustrates the respective measurement combination. The PLSR model built using the combination in case 20 resulted in an RMSECV of 16.1 µm and coefficient of correlation of 0.994, indicating a good fit. This calibration error obtained is smaller than the best results obtained using vis-NIR or the CLD data, which were 30.2 (no pre-processed data - Table 6.6, Case 26) and 18.3µm (using SNV as a pre-treatment - Table 6.1). Figure 6.21(a) shows the estimated particle concentration versus actual particle size, in which an R^2 of 0.994 was obtained. The residuals plot in Figure 6.21(b) do not exhibit any particular pattern.

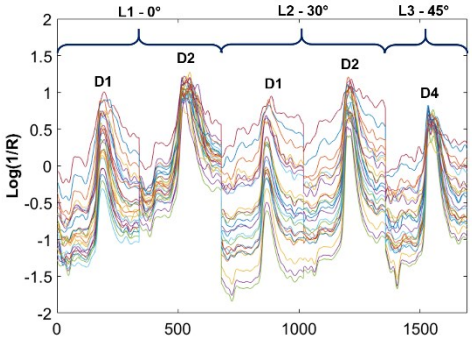


Figure 6.20: Absorbance NIR spectra using the measurement combination described in case 20. D1 – 0.3 mm; D2 – 0.6 mm; D4 – 1.2mm

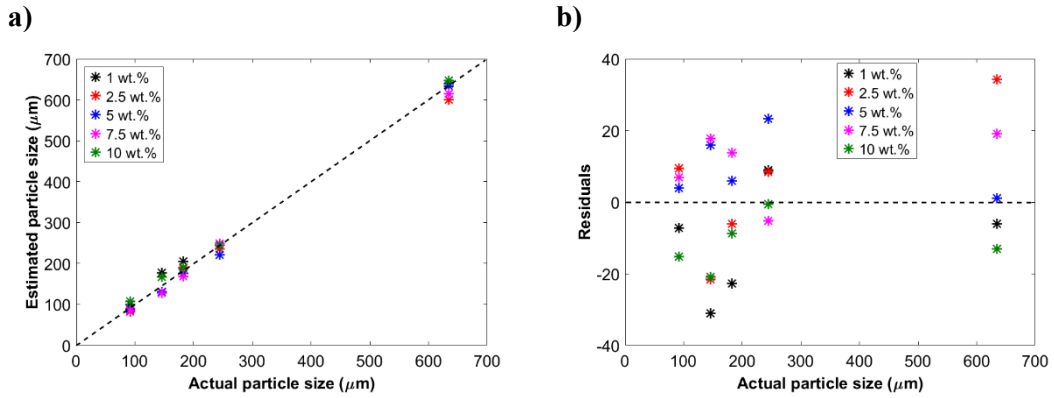


Figure 6.21: a) Estimated versus actual particle size. **b)** Residuals in the NIR calibration.

The impact of pre-processing, using the SNV, MSC, EMSC and ISC on the calibration models of augmented and co-added data is assessed and summarised in Table 6.14. To simplify the discussion, only the best results for each pre-processing technique are shown and compared to the models without using pre-processing, in Table 6.15.

Table 6.15: Impact of pre-processing methods (SNV, MSC and ISC) on calibration models built to estimate particle size from the absorbance NIR spectra. The table presents the cases (measurement configurations) in which the pre-processing method resulted in the lowest RMSECV using data augmentation and data co-adding approaches.

	Augmented				Co-added			
Pre-processing method	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**
None	**20**	9	16.1	0.994	**22**	5	52.8	0.929
SNV	**23**	3	44.0	0.951	**20**	4	53.0	0.928
MSC	**13**	7	40.2	0.964	**20**	4	50.1	0.936
EMSCL	**13**	7	41.5	0.962	**24**	3	54.7	0.925
ISC	**16**	5	46.4	0.949	**23**	4	55.2	0.922

The data augmentation resulted in more complex models, requiring more LVs to explain the data variance, due to the extra information captured by the different detecting fibres, in comparison to the co-adding approach. The use of pre-processing techniques in the co-adding approach resulted in similar RMSECV values among themselves, which is usually expected since the underlying principle of the formula is similar. This indicates that the models have similar accuracy and a simple model can be built using only the unprocessed data. The lowest error of calibration in the co-adding approach was obtained using the MSC correction, resulting in an error of 50.1 µm and correlation coefficient of 0.936.

In the data augmentation approach, the pre-processing methods applied to correct the scattering and baseline effects in the dataset also results in models with similar

accuracy as seen for the co-added data. However, models using SNV and ISC needed fewer LVs to explain the data variance, showing its model simplicity. All the pre-processed data results in models with a significant increase of the error, although fewer LVs are used. While the increased model complexity may reduce the model robustness for future predictions, the higher RMSECV indicates that the pre-processing methods may not be robust enough to discern or correct for the undesired effects in the spectra.

Particle Concentration

The absorbance NIR spectral data was used to build PLSR models to estimate particle concentration. Table 6.16 summarises the model performance of the models developed for all configuration (cases described in Table 6.3). In general, poor model performance is observed, with errors of calibration above 1 wt.%. However, comparing this analysis with the analysis performed using the vis-NIR data (Table 6.8), the absorbance NIR spectra results in lower RMSECV values. The configurations which display the lowest calibration errors when using the normal incident light are from collecting fibres at 0.6 and 0.9 mm from the light source. For both the angular sources analysed, the detectors at the farthest distances exhibited the best performance, leading to a lower RMSECV value and to higher correlation coefficient. The detector selection to estimate particle concentration differs from the findings for particle estimation using NIR data, and may be due higher sensitiveness of the farthest fibres to distinguish solid concentration. Even though the quality of the signal may be poorer due to the weaker signal, the photons collected travel farther, increasing the chance of interaction with more particles and consequently, increasing the amount of light being absorbed, leading to better sensitivities and better models.

As explained earlier, considering that the system consists of two independent variables, models using less than 2 LVs are discarded from analysis.

The best model performance for the single detectors was obtained when using the 30° angular source and capturing the spectra by the farthest distance, 1.2 mm, which needed 4 LVs to describe the model with an error of 1.2 wt.% and a coefficient of correlation of 0.868, indicating a reasonable fit.

Table 6.16: Summary of the results obtained from PLSR models when using the absorbance NIR spectra to estimate particle concentration.

Case No.	No of LVs	RMESCV (wt.%)	R^2
1	1	2.90	0.237
2	4	1.60	0.769
3	3	2.02	0.629
4	4	2.35	0.510
5	4	1.88	0.696
6	3	1.46	0.805
7	5	1.30	0.848
8	4	1.20	0.868
9	5	1.96	0.681
10	2	2.43	0.474
11	6	1.65	0.771
12	3	1.56	0.778

Figure 6.22 shows the RMSECV versus LVs profile from the cases 1-12 when using the original NIR spectra to estimate particle concentration. Overall, it is exhibited an almost flat profile, indicating the need of few LVs to explain the data variance in the models.

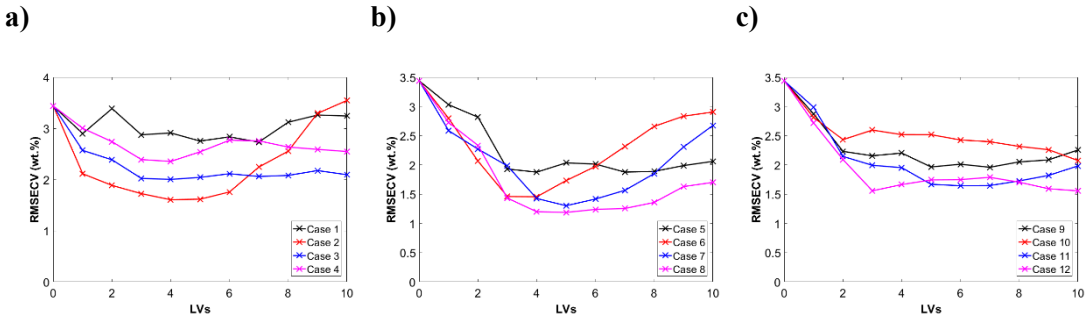


Figure 6.22: RMSECV versus LV obtained from the PLSR analysis using the absorbance NIR spectra from each one of the detectors at different distances from the light source at **a)** 0°, **b)** 30° and **c)** 45°, for estimating particle concentration.

Table 6.17 shows the measurement combinations selected to investigate whether the calibration models to estimate particle concentration can be improved. Table 6.18 summarises the PLSR calibration model performance built using the absorbance NIR spectral data for estimating particle concentration and compares the performance of two data combining approaches, data augmentation and data co-adding. The cases 13 to 19 are the same combinations studied in the vis-NIR and NIR region, using all the detectors from each light source used and then making different combinations among the different light sources. It is observed that the combination of all detecting fibres at different distances from the normal incident light, Case 13, leads to an improvement of the model performance for the augmented and co-added data in comparison to the single configuration. To simplify the discussion, the cases from 20 to 26 are based on the best model performances using different combinations of the 6 detectors selected previously, 2 per each light source used, based on the results summarised in Table 6.16. As the combination D1:D4 – L1 led to a better performance model, combinations with selected detecting fibres was also investigated.

Table 6.17: Measurement configurations cases used to build PLSR models and compare the data co-adding and augmentation approaches for particle size estimation using the absorbance NIR spectra. These were constructed using spectra from different detector distances (D1 – 0.3 mm, D2 – 0.6 mm, D3 – 0.9 mm, D4 – 1.2 mm) from the light source 0°, 30° and 45°.

	Angular source		
Case nº	**0°**	**30°**	**45°**
13	D1:D4		
14		D1:D4	
15			D1:D4
16	D1:D4	D1:D4	
17		D1:D4	D1:D4
18	D1:D4		D1:D4
19	D1:D4	D1:D4	D1:D4
20	D2:D3	D4	D3:D4
21	D2:D3	D4	D4
22	D2:D3	D3:D4	D3:D4
23	D1:D4		D4
24	D1:D4	D4	
25	D1:D4		D3
26	D2		D3:D4

Table 6.18 summarises the PLSR calibration model performance built using the absorbance NIR spectral data for estimating particle concentration, and compares the performance of two data combining approaches, data augmentation and data co-adding. In general, the data augmentation approach leads to fewer LVs in comparison to the co-adding approach. These results differ from the previous results using vis-NIR data and NIR data to estimate particle size, where the co-adding approach leads to less complicated models and higher errors of calibration. As most of the cases studied use the detecting fibres at the longest distance from the light source, the signal is weaker and less resolved. The co-adding approach averages all the spectra to improve the signal-to-noise ratio, but in the cases of higher number of LVs, it is modelling the spectral noise. In the data augmentation approach, there was an improvement of the model performance in comparison to the single detectors, with an exception for the cases 15, 17, 18 and 19, although it was not significant. The best model performance in data augmentation was found from Case 20 showing an RMSECV of 0.99 wt.% with an R^2 of 0.915.

Table 6.18: Summary of the PLSR calibration model performance for estimating particle concentration using data augmentation and co-adding approaches on absorbance NIR data.

	Augmented			Co-added		
Case No.	**No of LVs**	**RMESCV (wt.%)**	**R^2**	**No of LVs**	**RMESCV (wt.%)**	**R^2**
13	5	0.99	0.910	4	1.10	0.889
14	4	1.13	0.88	5	1.23	0.863
15	7	1.81	0.722	5	2.33	0.562
16	6	1.10	0.892	6	1.56	0.882
17	4	1.61	0.770	8	1.63	0.773
18	5	1.31	0.843	7	1.43	0.829
19	4	1.30	0.844	6	1.39	0.831
20	5	0.96	0.915	9	0.81	0.939
21	5	0.96	0.914	7	0.83	0.936
22	5	0.99	0.911	8	0.88	0.931
23	4	0.99	0.910	3	1.12	0.884
24	4	0.99	0.911	3	1.17	0.875
25	5	1.00	0.909	4	1.13	0.881
26	6	1.01	0907	5	0.98	0.914

The lowest errors of calibration were obtained using the co-adding approach where case 21 displayed the best relationship between the number of LVs and RMSECV. Although case 20 shows an error of 0.81 wt.%, the error decrease from 0.84 wt.% (case 21) to 0.81 wt.% (case 20) is not significant to justify the need of 2 extra LVs.

Figure 6.23 shows the RMSECV curves for Cases 13-26. The combination of different source-detectors results in an improved and smoother RMSECV profile in comparison to those from single configurations shown in Figure 6.22.

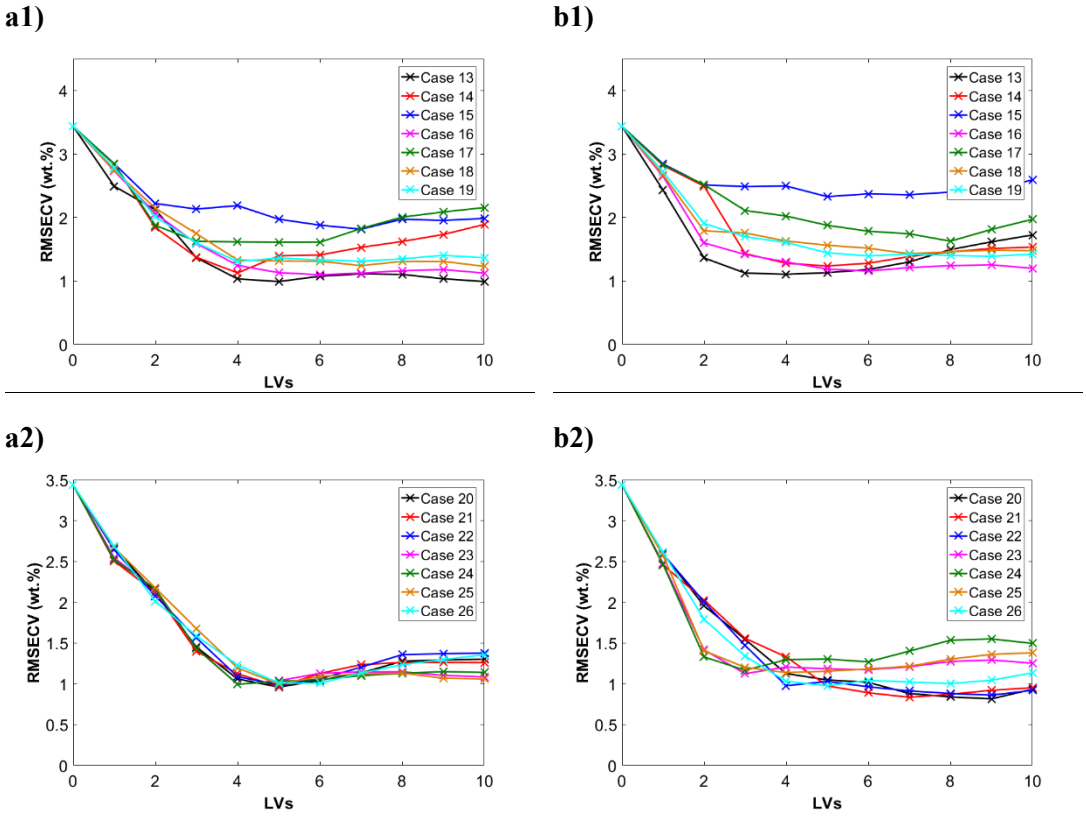


Figure 6.23: Cross-validation curves of PLSR models built for estimation of particle concentration using **a)** data augmentation and **b)** co-adding approach for **(1)** cases 13-19 and **(2)** cases 20-26. The models were built using the absorbance NIR spectral data.

Similar to the previous study, the impact of the pre-processing methods, namely, SNV, MSC, EMSC and ISC, on the calibration models for particle concentration, was evaluated on the augmented and co-added data. The best results for each pre-processing technique are summarised in Table 6.19. For both data combining approaches, the models developed using pre-processed data showed similar performance. However, in the co-adding approach, these models resulted in a reduction of the number of LVs used to explain the model, reducing its complexity.

Overall, there was not a significant improvement in the model performance in comparison to the unprocessed data. The results indicate the methods applied to correct the scattering and baseline effects in the absorbance NIR dataset may not be robust enough to recognise or correct for the undesired effects in the spectra.

Table 6.19: Impact of pre-processing methods (SNV, MSC and ISC) on calibration models built to estimate particle concentration from the absorbance NIR spectra. The table presents the cases (measurement configurations) in which the pre-processing method resulted in the lowest RMSECV using data augmentation and data co-adding approaches.

	Augmented				Co-added			
Pre-processing method	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**
None	**20**	5	0.96	0.915	**21**	7	0.83	0.936
SNV	**14**	4	1.13	0.887	**22**	5	1.03	0.903
MSC	**24**	6	1.18	0.879	**21**	5	1.05	0.903
EMSCL	**24**	6	1.20	0.875	**20**	7	0.94	0.920
ISC	**24**	4	1.13	0.883	**21**	5	1.08	0.894

6.2. Part 2 – Multi-sensor calibration

Particle size

Section 6.1 presents a logical analysis of the PLSR model performance for finding the best model performance for each technique, FBRM and SAR-DRM. In the case of SAR-DRM, two wavelength regions were investigated, vis-NIR (500-900 nm) and NIR (1000-1850 nm). For each region, it was investigated the impact of the combination of configurations on the calibration models by two data approaches, data augmentation and co-adding. Table 6.20 summarises the best PLSR calibration models performance for estimating particle size and concentration for FBRM and SAR-DRM found in Section 6.1.

Table 6.20: Summary of the best PLSR calibration models for the estimation of particle size and concentration for each technique and wavelength region investigated.

Particle size estimation						
Instrument	**Configuration**	**Data combination approach**	**Pre-processing**	**No of LVs**	**RMECV (µm)**	**R^2**
FBRM	Square weighted CLD	-----	SNV	8	18.3	0.993
SAR-DRM (vis-NIR)	D2 – L1 D3 – L2	Augmentation	None	5	30.2	0.977
SAR-DRM (NIR)	D1 & D2 – L1 D1 & D2 – L2 D4 – L3	Augmentation	None	9	16.1	0.994
Particle concentration estimation						
Instrument	**Configuration**	**Data combination approach**	**Pre-processing**	**No of LVs**	**RMECV (wt.%)**	**R2**
FBRM	Square weighted CLD	-----	None	9	0.45	0.982
SAR-DRM (vis-NIR)	D2 – L1 D2 – L2 D3 – L3	Augmentation	None	8	0.73	0.952
SAR-DRM (NIR)	D2 & D3 – L1 D4 – L2 D3 & D4 – L3	Augmentation	None	5	0.96	0.915
SAR-DRM (NIR)	D2 & D3 – L1 D4 – L2 D4 – L3	Co-added	None	7	0.83	0.936

This section 6.2 investigates further whether the calibration models can be improved by combining the data generated by the different equipment and spectrometers. The dataset described in Table 6.20 was combined to form the following cases of study: FBRM and vis-NIR; FBRM and NIR; FBRM, vis-NIR and NIR; vis-NIR and NIR. For data augmentation and co-adding approaches the blocks were mean centred. Similarly to the analysis in section 6.1, leave-one-out cross-validation was used.

Table 6.21 summarises the PLSR calibration model performances for estimating particle size using the augmented and co-added approaches. In all cases, the data augmentation approach resulted in a lower error of cross-calibration and higher

correlation coefficient. However, when combining the FBRM data with NIR data, the co-adding approach results in a simpler model, using fewer LVs to explain the variance in the model with an RMSECV error relatively similar to the corresponding augmented dataset. Figure 6.24 displays the cross-validation curves profiles for every combination and approach used. Overall, a smooth profile was obtained.

Table 6.21: PLSR calibration model performance for estimating particle size using data augmentation and co-adding approaches on data fusion from different techniques.

	Strategy	No of LVs	RMSECV (µm)	R^2
FBRM + vis-NIR	**Augmented**	9	10.3	0.997
	Co-added	8	16.5	0.993
FBRM + NIR	**Augmented**	8	12.1	0.997
	Co-added	6	13.8	0.995
FBRM + vis-NIR + NIR	**Augmented**	8	11.7	0.997
	Co-added	6	29.0	0.979
vis-NIR + NIR	**Augmented**	10	21.3	0.989
	Co-added	10	27.6	0.983

The best model performance was obtained when combining the square weighted CLDs with the source-detectors from the NIR described in Table 6.20, using a co-adding approach. This model only needs 6 LVs to describe the model with an error of 13.8 µm and coefficient of correlation of 0.995, indicating a good fit.

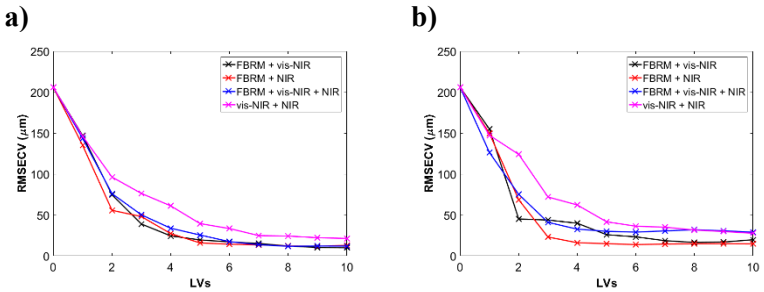


Figure 6.24: Cross-validation curves of PLSR models built for estimation of particle size using **a)** data augmentation and **b)** co-adding approach for data fusion.

The same strategy was used to build models to estimate particle concentration. Table 6.22 summarises the calibration model performance built to estimate particle concentration. Overall, models built on augmented data present lower RMSECV values with a relatively similar number of LVs to the co-added approaches. Figure 6.25 shows the cross-validation curves of PLSR models built for particle concentration estimation do not exhibit a very smoother although showing a constant decrease in error, suggesting the high number of LVs to be used to explain the remaining variance of the model. However, for the conditions evaluated, more than 10 LVs are not justified. The best model performance was obtained when combining the information from FBRM and from the source-detectors NIR (combinations described in Table 6.20), using either an augmentation or co-added approach. This model needs 9 LVs to describe the model with an error of 0.58 wt.% and the coefficient of correlation of 0.970, indicating a good fit. The number of LVs is justified by the high number of source-detectors used to build the model, with each one of them bringing extra information to the model. The models built on the co-added approach were expected to need less LVs as the information from multiple detectors is averaged to form just one spectrum.

Table 6.22: PLSR calibration model performance for estimating particle concentration using data augmentation and co-adding approaches on data fusion from different techniques.

	Strategy	No of LVs	RMSECV (wt.%)	R^2
FBRM + vis-NIR	**Augmented**	10	0.91	0.924
	Co-added	7	1.18	0.874
FBRM + NIR	**Augmented**	9	0.58	0.969
	Co-added	9	0.58	0.970
FBRM + vis-NIR + NIR	**Augmented**	10	0.80	0.941
	Co-added	8	1.06	0.899
vis-NIR + NIR	**Augmented**	9	0.67	0.959
	Co-added	8	1.03	0.903
FBRM + NIR-Data co-added	**Augmented**	9	0.67	0.958
	Co-added	9	0.56	0.972
FBRM + vis-NIR + NIR-Data co-added	**Augmented**	9	0.82	0.938
	Co-added	8	1.05	0.901
vis-NIR + NIR-Data co-added	**Augmented**	9	0.61	0.966
	Co-added	8	0.93	0.921

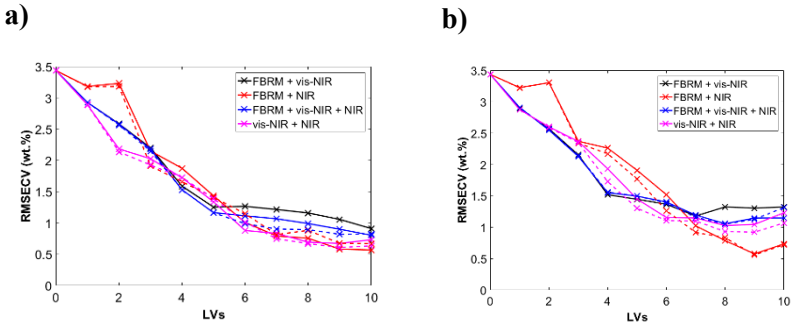


Figure 6.25: Cross-validation curves of PLSR models built for estimation of particle concentration using **a)** data augmentation and **b)** co-adding. The models using the NIR-data co-added approach are represented by the dashed line.

In general, a combination of information from multiple sensors has shown to improve the calibration models for estimating either particle size or particle concentration.

6.3. Conclusions

In this chapter, it was investigated the impact of fusing information from different measurement platforms on the calibration models performance for estimating particle size and concentration of polystyrene samples.

Initially, calibration regression models were built for the in-line tools FBRM and SAR-DRM, individually. The FBRM data was divided into two datasets, one containing the unweighted CLD and the second the square weighted CLD information. The best model performance, to estimate both particle size and concentration, was achieved using the square weighted CLD. As the square weighted CLD takes into to account larger particles weight more and occupy a larger volume, deemphasising the fine particles and simplifying the CLD to unimodal distributions in the majority of the cases, makes it easier for the model to detect a shift in the peaks with the particle size. The use of SNV in this data, removes the interference of the counts, improving the linearity of the model. This led to a decrease of the calibration error from 207 μm to 18.3 μm and improvement to an R^2 of 0.993. The square weighted CLD was also essential to build PLSR models to estimate particle concentration. In this particular case, the best model was achieved without using any pre-processing method, which demonstrates that besides the peak shifting and peaks shape (as the system is dependent on both particle size and concentration), extra variance was being captured from the number of counts in each condition tested. The best model performance resulted in an RMSECV of 0.45 wt.% and an R^2 of 0.982, using 9 LVs.

Calibration models were also developed for the vis-NIR and NIR spectra acquired from SAR-DRM. The investigation indicates an improvement in PLS regression model performance when using the extra information captured the different source-detector distances by SAR-DRM. Overall, the data augmentation approach led to the improvement of the model performance. The extra information is better utilised when the dataset is built by data augmentation rather than co-adding the spectra. The sample features seem to be better preserved by using extra variables to build the models. The increase in signal-to-noise by co-adding additional signal from different combinations

of detectors and angular sources seems to have a negligible effect on the model performance.

In general, the models developed for the vis-NIR and NIR region performed better using the raw data or log-transformed data. The application of pre-processing methods to correct the vis-NIR and NIR seemed not to be robust enough to discern or correct undesired effects in the spectra and consequently, improve the model.

This study also shows that particle size and concentration can be best estimated by fusing the information from different techniques and spectrometers. The combination strategy is, however, dependent on the system and conditions in study.

Chapter 7 - Data fusion methodologies for prediction of particle size and concentration of pharmaceutical suspensions

In the previous chapter, calibration models were built to estimate particle size and concentration of polystyrene samples. A data fusion approach to combine in-line sensors was proposed to improve the models. Nevertheless, due to the small dataset used (calibration set of 25 samples), the model performance is only indicative.

In this chapter, multivariate regression analysis is employed to build predictive models to estimate particle size and particle concentration of α-lactose monohydrate suspensions in acetone using a calibration set of 80 samples and a test set of 40 samples. The approach proposed in Chapter 6 is applied. As the dataset is bigger, the models developed can be considered robust and with more accurate median values as a predictive response. This model system also allows us to assess SAR-DRM response to crystalline systems and its potential applicability in the pharmaceutical industry for the in-line monitoring of processes such as crystallisation or milling, for example, along with the data fusion strategy.

7.1. Introduction

In order to study the capability of SAR-DRM for in-line monitoring of processes such as crystallisation or milling, α-lactose monohydrate suspended in acetone was selected to form a model system. Acetone with a purity of 99.5% was chosen as the non-solvent medium for this study since the solubility of lactose is very close to zero [147].

The α-lactose monohydrate crystals often present a unique tomahawk crystal habit but they can also show alternative morphologies such as diamond-shaped plates and pyramidal shape, depending on the conditions of the crystallisation (e.g., supersaturation, growth rate, temperature) and drying method, for example [148-150].

In anti-solvent crystallisation of α-lactose monohydrate, water is used as the solvent. Several studies have been reporting the effects of using not only different anti-solvents but also the proportion of anti-solvent/solvent in the system. Brito and Giulietti, for example, have reported that different water-acetone compositions lead to crystals with different average diameters, shape factors, and recovered mass [147]. However, the solid forms of lactose studied were not specified. The α-lactose monohydrate samples used in this work were obtained through DFE pharma and we are not aware of the crystallisation conditions used. Although the tomahawk crystal shape is expected to be dominating, different particle shapes may be observed during the experimental measurements. Moreover, the use of water in our system to preserve the monohydrate form would likely induce some changes in the particle size distribution as small fragment would likely to be dissolved as well as part of the crystals surface, and could lead to events such as crystal growth. Hence, using lactose suspensions in acetone is expected to minimize these risks and preserve the particle size distribution and shape of the crystals during the in-line measurements.

However, acetone is known to induce dehydration and the formation of the anhydrous form. Chen *et al*., studied the transformation of α-lactose monohydrate to α-lactose anhydrous using thermal analysis [151]. In this study, the samples prepared from dewatered acetone revealed that α-lactose monohydrate should only gradually lose its water of crystallisation at temperatures above 100ºC and that the water loss is expected to be completed at 145ºC which is accompanied by a change of its crystalline structure.

A change in our samples polymorph could lead to different refractive indexes, shape and size. However, in our study, α-lactose monohydrate suspensions were kept at 20 ºC, which is not high enough to drive off the water from the crystals.

Garnier *et al*., suspended α-lactose monohydrate large single crystals in acetone at room temperature for two weeks without stirring [152]. The authors reported a slow formation of 'whisker-like' crystals, which were limited to the surface of the initial particle. The dendritic growth mechanism was due to the miscibility of water from the hydrated phase with the organic solvent. The authors also reported that the smooth conditions used (no stirring) led to the formation of anhydrous β-lactose and attributed it to the departure of water molecules accompanied by partial dissolution and solvation of α-lactose molecules.

The study of polymorph was not the scope of this thesis, however, if the multi-inline sensor approach is applied to monitor crystallisation processes, more studies should be performed. The application of an in-line Raman would help to track any polymorph change. Due to the experimental conditions used in this work, stirring speed, temperature and time of analysis, it is assumed there was no significant change in the crystalline structure of the commercial grades obtained. A change in size and shape would be likely to be observed in FBRM over time. A change in the sample refractive index would be likely to affect the diffused reflected signal collected from SAR-DRM equally for each one of the experiments.

This chapter is divided into three sections. In section 7.2, the PSD of the lactose sieved samples is evaluated and it is shown the typical data collected from FBRM and SAR-DRM. In section 7.3., multivariate regression analysis is employed to each in-line technique to build predictive models to estimate particle size and particle concentration from α-lactose monohydrate suspensions in acetone. Differently from the polystyrene study, a larger range of particle sizes (<38 to 400 μm) and solid loading (0.5 to 25 wt.%) were evaluated to produce a sufficient number of training and test samples to evaluate the proposed analysis approach. A test set for external validation was used to improve model performance and stability. The experimental setup is described in Chapter 4 Section 4.3.2. In case of SAR-DRM, the impact of different combinations of source-detectors on the model performance were assessed. The CLD and spectral

data were subjected to different pre-treatments, log-transformation, SNV, MSC and ISC according to the type of data analysed in each subsection and they are specified in the respective section. Section 7.4 explores the impact of fusing the FBRM and SAR-DRM data on the predictive capability of the models.

7.2. Off-line characterisation and in-line analysis on α-lactose monohydrate suspensions

The PSDs of the sieved lactose crystals were characterised using a laser diffraction device. The mean particle size was extracted and used as reference values to build multivariate regression models to predict particle size.

In this section, it is also shown the typical data collected from the lactose suspensions using the in-line tools, FBRM, PVM and SAR-DRM. Both particle size and concentration effects were analysed qualitatively.

7.2.1. Off-line analysis

Laser diffraction

The α-lactose monohydrate crystals size distribution was measured using a laser diffraction device. As it was observed during the polystyrene study, in Chapter 6, the resemblance between sieving and laser diffraction measurements can be poor due to the mechanics involving sieving. Long and thin particles will pass through a mesh and assumed to be the size of the mesh, whereas when long and thin particles are in a suspension, they are rotating, the full projection of the particle will result in much larger particle size. Moreover, during sieving, electrostatic forces can lead to particles agglomeration. The suspension of these particles in a dispersion medium will lead to its separation, as a result they will be measured as smaller particles. Although Morphologi G3 is advantageous to provide quantitative information from each individual particle, such as size and shape, the size measurement can differ from laser diffraction, as it only takes into account the 2D nature of the particles. Even though

Morphologi G3 makes use of a dispersion unit to help to separate the particles, the electrostatic forces can still lead to particles agglomeration issues. In addition, this off-line analysis is more time consuming than laser diffraction analysis. For these reasons, and due to the fact that laser diffraction is the most commonly used technique in industry for determination of particle size, the laser diffraction measurements were used as reference values to build particle size prediction models.

Figure 7.1 shows the volume-weighted distribution of the Equivalent Sphere Diameter (ESD) obtained from the LD measurement for the sieved fractions of lactose. It is observed a distinctive PSD curve for each size fraction. An increase in the crystal size range led to a shift of the PSD towards larger sizes, with each distribution curve displaying a good agreement with the sieved size range.

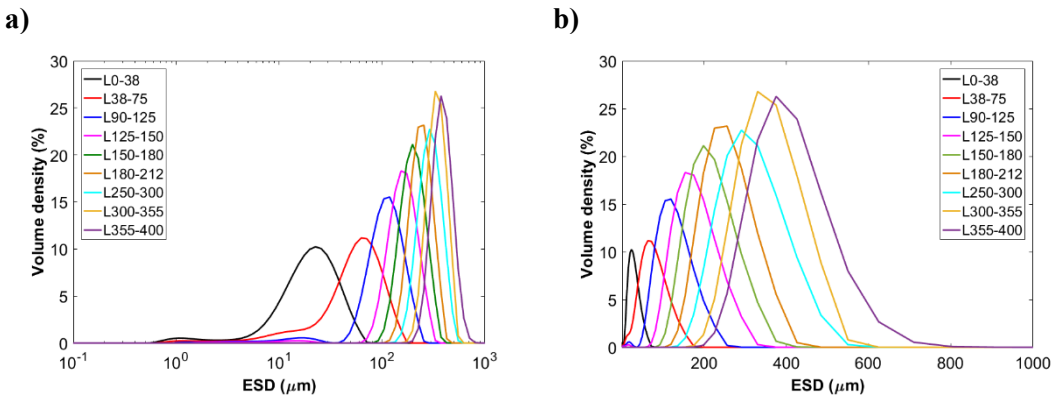


Figure 7.1: Equivalent Sphere Diameter (ESD) obtained from LD analysis in **a)** log and **b)** linear scale.

Figure 7.1 also shows that the presence of fine particles seems to have an impact on the particle size distribution curves, with the samples L0-38, L38-75, L90-124 and L125-150 exhibiting PSD curves with relatively small shoulders towards smaller sizes. Figure 7.2 exhibits an example of the images acquired during the LD analysis. In general, a significant amount of fines is detected in each of the particle size ranges evaluated. These fine particles may be related to particle attrition, which induces physical damage and generation of fragment daughter particles, resulting in particle size reduction. Figure 7.3 exhibits three possible mechanisms of attrition: shattering,

fracture of the particles resultant from high collision energy; abrasion, the collision energy is only capable to remove small amounts of material from the particle surface; and chipping, partial fracture of the particles, resulting in fewer but larger particles than abrasion [153]. Although a combination of the three mechanisms usually occurs, one of the mechanisms may dominate the process. The attrition mechanism taking place can be indicated by the nature of the change of the particle size distribution. Shattering usually leads to a large shift of the PSD main peak towards smaller sizes, changing the D50 significantly and abrasion results in a bimodal distribution from the generation of the fines with little impact on the D50 [153]. Laser diffraction is a quick particle size measurement device. During the few minutes taken to analyse the samples particle size, the nature of the particle distributions remained the identical, and consequently, little or no change was observed in the D10, D50 and D90 values obtained. This indicates the absence of attrition effects or partial solubilisation of the sample.

The particle size analysis of the commercial grades of lactose ordered from the manufacturer using an offline imaging microscope, Morphologi G3, before the samples were sieved, revealed the presence of a high number of fine particles (<10 μm), as shown in Figure C.1(a), which exhibits the PSD as a number weighted ECD distribution. The different lactose grades acquired from the manufacturer are likely to be obtained through a milling or sieving process. Commonly, finer PSD are produced by milling whereas coarse grades are separated by sieving [154]. The particles fragments observed in the images can, therefore, be a product of attrition mechanisms during the crystallisation/milling processes which were not removed during the sieving process or a result from particles attrition during sieving. In the case of the coarser particles, the latter one could have happened twice (first by the manufacture and second by further sieving in our lab). Moreover, powders of smaller particle sizes are more challenging to obtain narrower PSD distributions through sieving as the particles tend to adhere to each other surface. In that case, wet sieving may be preferable. However, in our study, in order to subject all the samples to the same conditions, only dry sieving was performed. The resuspension of particles in acetone could lead to the separation of the crystal fragments adhered to larger crystals surface.

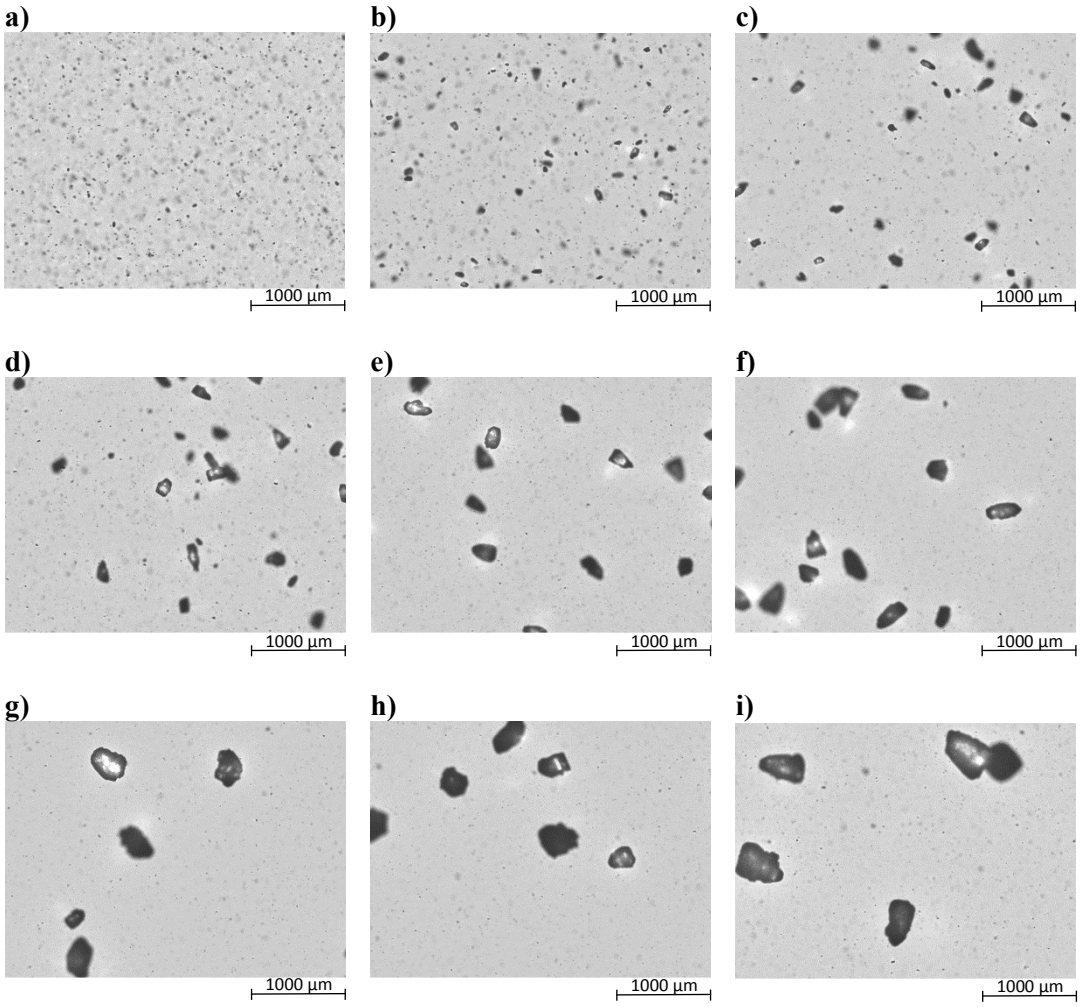


Figure 7.2: Hydro Sight images of **a)** L0-38, **b)** L38-75, **c)** L90-125**, d)** L125-150, **e)** L150-180**, f)** L180-212, **g)** L250-300, **h)** L300-355 and **i)** L355-400.

A study conducted by Agrawal *et al.* on lactose monohydrate crystals suspensions using a laser diffraction device, Mastersizer, revealed that particle size and impeller speed significantly affect attrition [153]. Two stirring speeds were used, 2000 and 3000 rpm. Their results have shown coarse particles are more likely to undergo shattering, chipping and abrasion whereas small particles only underwent abrasion, and that higher impeller speed also increases the production of new fragments. The LD measurements presented in this current study were acquired using a stirring speed of 2000 rpm. The short time of analysis and low stirring speed may be the reason why no attrition mechanisms were observed. However, considering the in-line measurements

are collected during larger periods of time, although with a lower stirring speed, at 700 rpm, the effects of attrition were more closely analysed in the next section.

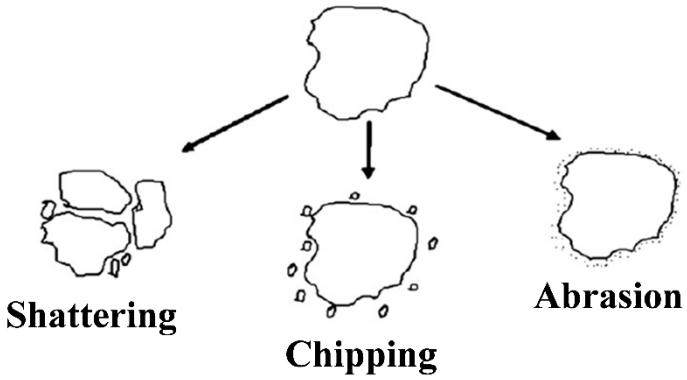


Figure 7.3: Attrition mechanisms *[153]*.

7.2.2. In-line analysis

The effect of crystal size and concentration was evaluated using FBRM, PVM and SAR-DRM. The attrition mechanism taking place was mainly evaluated using FBRM, by assessing its impact on the CLD measurements over time. Although the effects of attrition and particle size have been studied for all solid loadings evaluated, the results shown below are for 10 wt.% solid loading, as an example of the information that can be collected from all techniques.

FBRM & PVM

In addition to providing the unweighted and square weighted CLD distributions of the particulate suspension, FBRM software displays the number of counts detected for individual size classes over time. This feature makes FBRM an attractive tool to monitor mechanisms such as growth, agglomeration and attrition. To assess the particle attrition mechanism, samples L0-38, L90-125, L250-300 and L355-400 are used as an example. Figure 7.4 (a) shows the trend of the counts with chord lengths <10 μm, 10-100 μm, 100-1000 μm, and of the corresponding sieved size range, (1) <40 μm, (2) 90-125 μm, (3) 250-300 μm and (4) 355-400 μm, over time. A decrease

in the target size range or an increase of the fines would be expected in case of attrition. Figure 7.4 (b) shows the corresponding unweighted and square weighted CLD distributions at specific times selected, typically, 5, 10, 20 and ~25min after the measurement has started, to simplify the observation. In general, smaller particle sizes exhibited a constant number of counts over time for each individual size classes evaluated, as shown in Figure 7.4 (a1)-(a2). The corresponding CLD curves remain identical during the analysis, Figure 7.4 (b1)-(b2). No evidence of particles shattering, chipping or abrasion is perceived during the 25 min of sample measurement. However, the increase in particle size seems to lead to particle abrasion. In Figure 7.4 (a3) is shown a small increase in the number of particles counts with chord lengths below 10 μm over time, for the sample L250-300. This suggests some abrasion effect in the sample. Increasing the particle size range of analysis, L355-400, this event becomes more prominent. Figure 7.4 (a4) exhibits not only an increase in the counts with a chord length below 10 μm but also an increase in the number of chords between 10-100 μm, which suggests particles abrasion and particles chipping. Nevertheless, these effects were not sufficient to shift the mean chord length towards smaller sizes or to change the CLD curve shape during the 40 min of sample agitation, as observed in Figure 7.4(b4), meaning the effect of abrasion is negligible for the measurements in this study. The same conclusions were drawn from the remaining samples and concentrations analysed. It was concluded the measurements and setup were reliable and suitable to be used to build predictive models. Particle sizes above the range 355-400 μm were not included in this study as shattering and chipping mechanism started dominating the measurements under the chosen experimental conditions.

An example of the effect of particle size in both unweighted and square weighted CLDs is presented in Figure 7.5. In general, it exhibits a shift of the CLD curve towards larger size when increasing the particle size range. The difference between the CLDs is more apparent in the lactose samples than in those from the polystyrene beads. The large number of particle fragments below 10 μm leads to broader CLDs in the unweighted mode, as particles are assumed to have equal weight. These small fragments are overlooked in all the square weighted CLDs, which emphasises the chord length range of interest by approximating the CLD to the sieved particle size

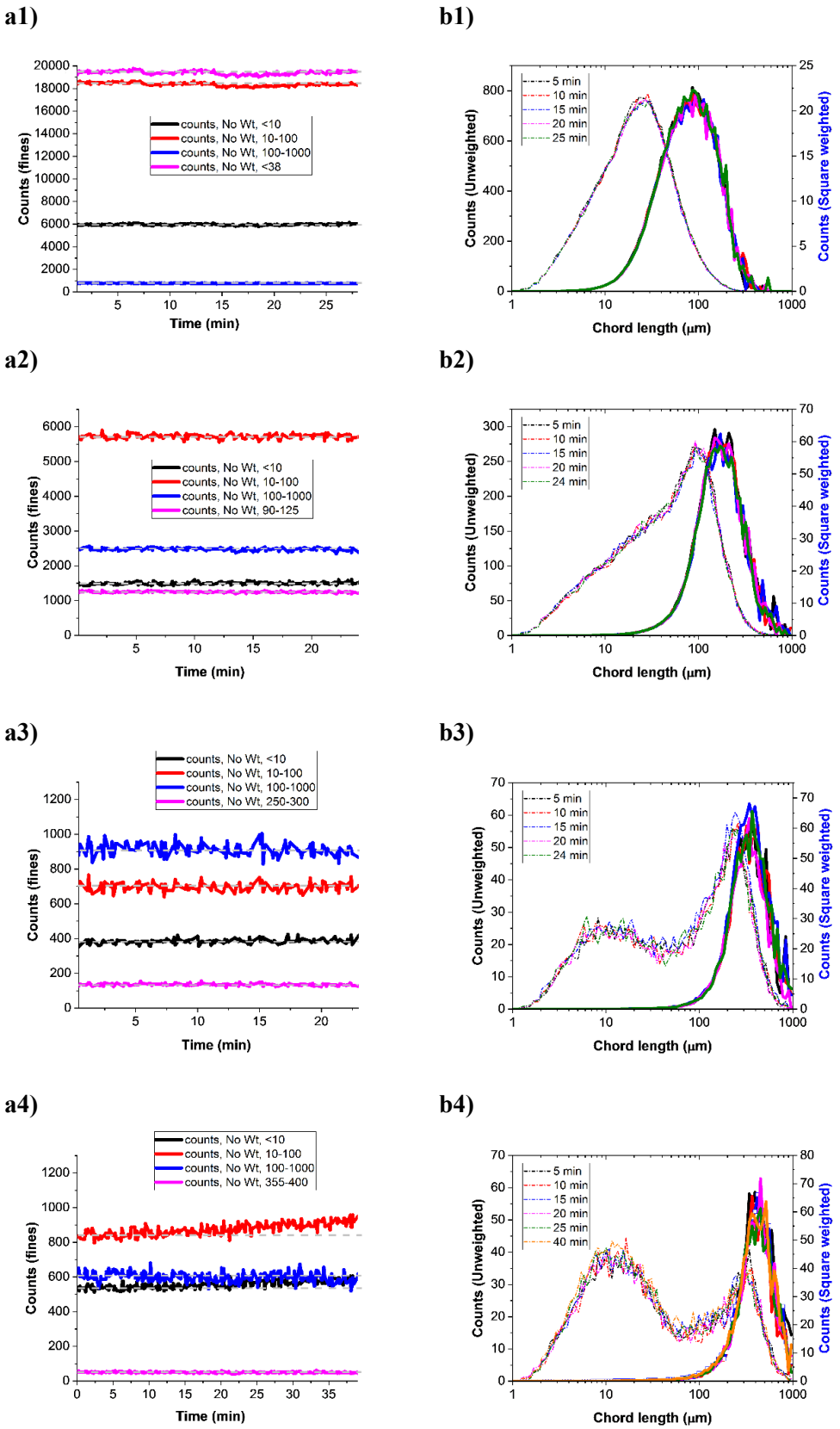


Figure 7.4: (a) Number of counts detected per each individual size classes over time and respective **(b)** unweighted (dashed-dotted line) and square weighted CLD (solid line) for 10 wt.% solid loading suspension of **(1)** L0-38, **(2)** L90-125, **(3)** L250-300 and **(4)** L355-400.

ranges. The particle size effect for the remaining concentrations can be observed in Figure C.2 and Figure C.3 in Appendix B.

The CLD inversion algorithm has also been applied to lactose data and it seems to successfully track the change in particle size during the in-line measurements. Figure C.4 shows the effect of particle size on the volume-weighted ECD distribution, inverted from the unweighted CLD, for the different solid loading tested. It is observed a shift towards larger sizes with the increase of crystal sizes. The inversion results in a bimodal distribution for particle sizes above 400 µm, in which the CLD suggests the crystal shattering. However, in general, the D50 obtained from the inversion algorithm is higher than the D50 from laser diffraction measurements, as observed in Figure C.5. For laser diffraction measurements, it is required a relatively low turbid suspension, in case of FBRM, the high solid content could lead to particles close to each other being detected as a single particle. This effect on the CLD would be carried on in the inversion to PSD, leading consequently to higher D50 values.

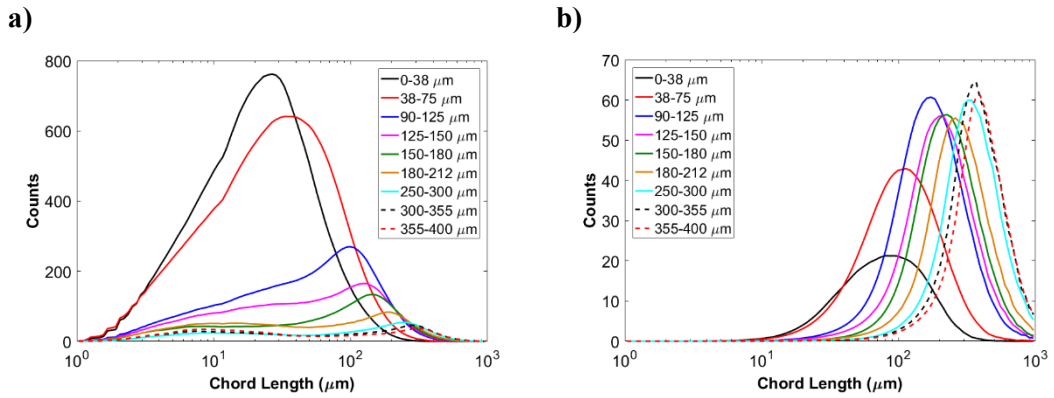


Figure 7.5: a) Unweighted and **b)** Square weighted CLD of 10wt.% of lactose crystals of various particle size.

Figure 7.6 shows examples of PVM images from samples of 10 wt.% solid loading for the different particle size ranges sieved. Similar to the analysis of polystyrene bead suspensions, it was used a reflector Teflon cap on PVM to obtain a transmission image. The smaller the particles, the higher the number of particles needed to achieve a solid loading of 10 wt.%, and therefore, samples L0-38 and L38-75 resulted in an overcrowded system for such condition. In Figure C.6, it is shown the PVM images

from a low turbid medium, containing 1wt.% of lactose particles, for visual comparison. From the pictures collected from larger particle size groups at 10 wt.% solid loading, it is possible to observe some crystals with the expected tomahawk shape. The other shapes observed could be related to the projection of the tomahawk shaped crystals, as images are acquired while particles are rotating. Due to the constant movement in space of particles, the 3D nature and shape of particles, it is difficult to obtain a good resolution 2D picture of the crystal sharp edges. The images also reveal the presence of the small crystal fragments detected by FBRM.

The analysis of large particles regardless the solid loading used would need thousands of pictures to be collected in order to find a statistically valid number of particles in focus and not touching the frame or overlapping, in order to extract PSD accurately. Hence, the PVM imaging algorithm was not applied in this study since it would not be able to captured particles in the majority of the conditions tested.

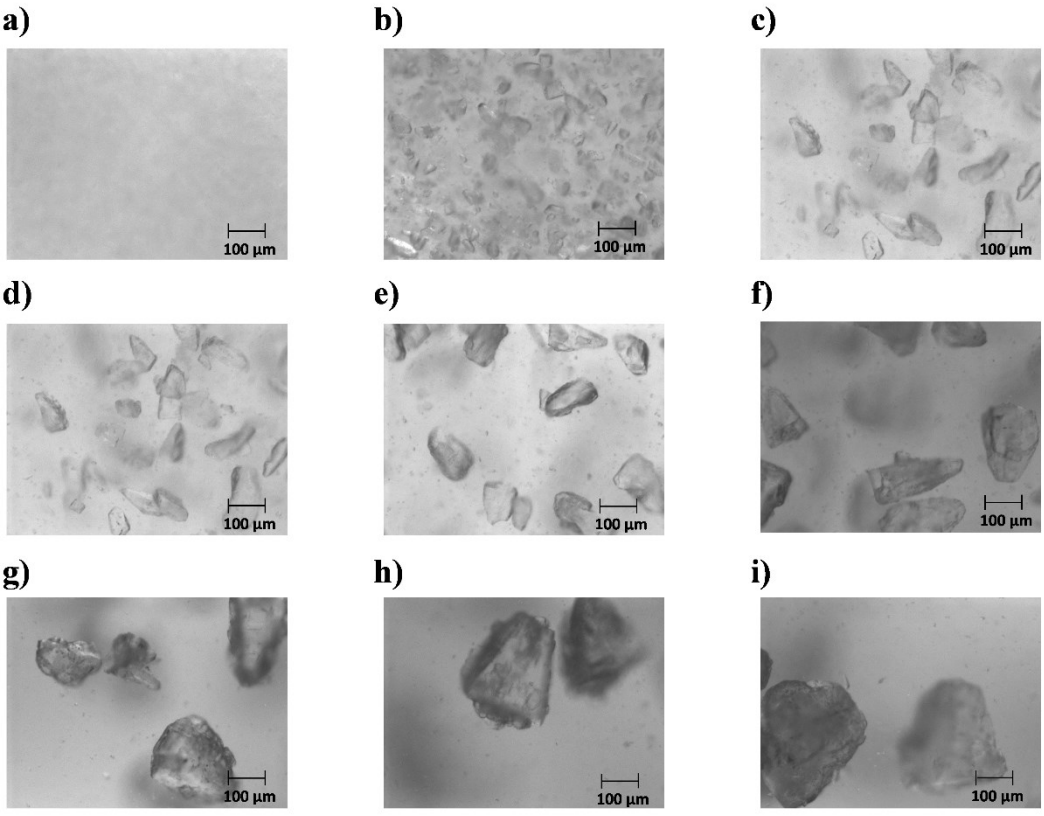


Figure 7.6: PVM images of lactose crystal using a suspension of 10 wt.% a) L0-38, **b)** L38-75, **c)** L90-125, **d)** L125-150, **e)** L150-180, **f)** L180-212, **g)** L250-300, **h)** L300-355 and **i)** L355-400.

Figure 7.7 displays the typical CLDs obtained when using different solid loadings of lactose crystals with a particle size range of 125-150 µm. An increase of the number of particles in suspension leads to an increase of the particle counts by FBRM, which is accompanied by a shift of the unweighted CLD towards smaller sizes. The same effects were previously observed in Chapter 6 when analysing polystyrene bead suspensions. A large number of particles present in the medium may interfere the flow of the suspension, leading to small particles to crowd FBRM window and to be measured more often, resulting in a shift of the CLD towards smaller sizes.

The crystals of non-spherical shape, its sharp edges, and the presence of fine particles in the suspension lead to the appearance of a shoulder in the main peak at ~130 µm, for the unweighted CLDs for L125-150. Although an increase in the solid loading results in the broadening of the square weighted CLD, the majority of the conditions evaluated for the same particle size range resulted in CL peak at similar chords/position. This shows that the increase of small crystals fragments does not affect the square weighted CLD. Due to the size of these fragments, their interference is this type of distribution is minimal.

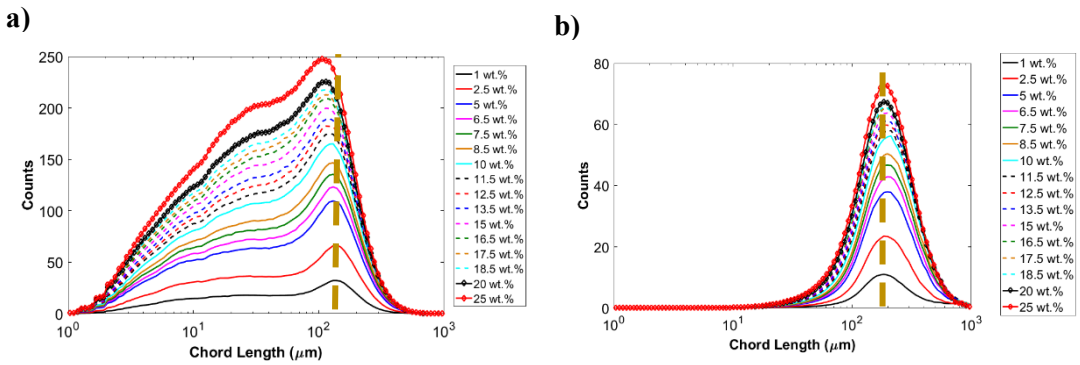


Figure 7.7: Effect of solid loading expressed as **a)** unweighted and **b)** square weighted CLD of L125-150.

Figure 7.8 shows the non-linear relationship between the total number of counts, obtained from the unweighted CLD distribution, and the concentration of particles studied. The counts of L38 and L38-75 exponentially increase with the concentration

and it seems to be reaching a plateau stage. The same behaviour is observed for the remaining particle sizes, although the process seems to be slower due to their larger size, and consequently, fewer particles in the medium for the same concentration of L38 and L38-75. The number of total counts decreases with the increase of particle size, regardless the concentration. As large particles weight more, fewer particles will be added to the suspension. An exception to this trend is observed for the samples L300-355 and L355-400 in which abrasion and chipping of particles start taking place. The increase in the number of fines in these conditions leads to slightly higher number of total counts. For the sample L38-75, when using a particle concentration above 15 wt.%, an unexpected behaviour was observed. A concentration of 16.5 wt.% led to lower counts than 15 wt.%. As the suspension was left overnight in the vessel and experiments of 16.5 wt.% or higher concentrations were run in the next day, the lower number of counts could be due to some deposit of particles at the bottom of the vessel. Although the sample was left stirring for about 15 min before starting the experiments, the number of counts did not increase during the 30 min of measurements taken. As it is difficult to assess whether this effect is real, if the longer time of the particles in the suspension induced some particle changes, if it is due to particles deposit or changes in the suspension flow due to the high number of particles, the measurements taken for 16.5-20 wt.% were not included in the dataset used to build multivariate regression models.

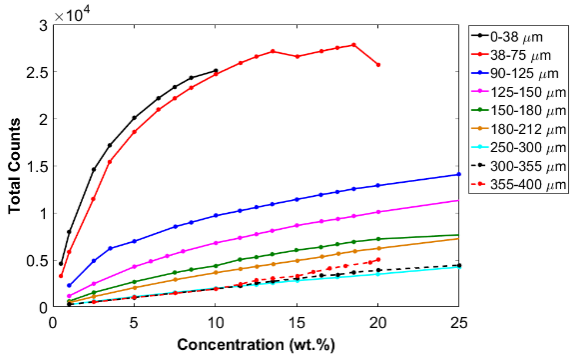


Figure 7.8: Total number of CL counts per each solid loading of the different lactose crystal size ranges.

SAR-DRM

SAR-DRM measurements were acquired simultaneously with FBRM and PVM. Two spectral regions were evaluated, vis-NIR and NIR, for each one of the conditions tested. However, the NIR signal acquired was saturated in a significant number of conditions tested and therefore, it was excluded from the multivariate analysis.

The aim of this section is to show an example of the typical vis-NIR spectra acquired from lactose suspensions which is later used to build predictive models. In order to simplify the discussion, the effect of particle size and concentration are only shown in the spectra collected from the closest fibre detector, 0.3 mm, when the light is emitted from the different angular sources, 0, 30 and 45°, as shown in Figure 7.9 and Figure 7.10 respectively. The step observed at 760 nm is due to the removal of the wavelength range 760 to 800 nm from the analysis in order to eliminate FBRM interference from the spectra (at ~785 nm), as mentioned in Chapter 4 Section 4.4 and Section 4.7.1.

Similarly to what was observed in SAR-DRM spectra from polystyrene suspensions, the change in the vis-NIR spectra with particle size is not straightforward. In Figure 7.9 is observed that the only samples with a clear signal separation are related to the smaller particle size ranges, L0-38 and L38-75. In a 10 wt.% suspension, the particle density number of these samples is significantly higher compared to the samples of larger sizes. This results in an increased number of scattering events which are translated into higher diffuse reflected intensity. In both normal incident and angular 30° source, the intensity seems to decrease monotonically from <38 to 125-150 μm. When using the 45° angular source, the intensity seems to decrease monotonically from <38 to 300-355 μm. While in polystyrene beads suspension, the largest particle size range evaluated showed the lowest spectral intensity, here is not the case. Although the second-largest particle size range (L300-355) has the lowest intensity when the light is emitted from the angular sources, 30 and 45, the intensity of L355-400 is higher. This could be due to the small crystal fragments observed in FBRM and PVM.

The effect of particle size, shape, and perhaps the increased number of crystal fragments with the increased size range seems to have an impact on the magnitude of

the signal and complicating the visual analysis. Multivariate regression analysis will be applied to assess whether a linear relationship can be captured among all the samples.

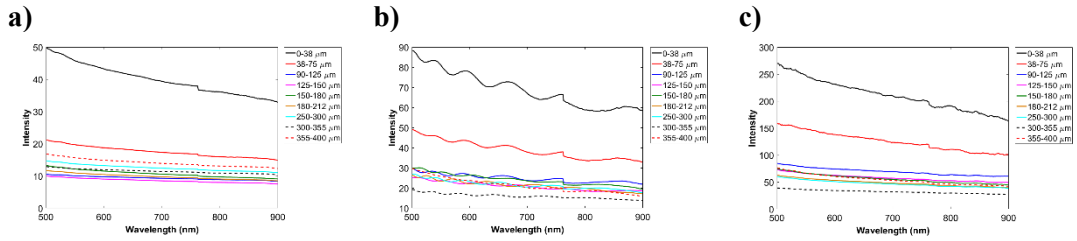


Figure 7.9: Effect of particle size in the vis-NIR region, of 10 wt.% suspensions of α-lactose monohydrate, when the light is emitted from the different angular sources, **a)** 0, **b)** 30 and **c)** 45°, and collected by the closest detector to the light source, at 0.3 mm.

The increase of solid loading resulted in a monotonical increase of the diffusely reflected light intensity, as shown in Figure 7.10 for all incident light sources. In general, incident angle results in a wider spectral separation between the different solid loading.

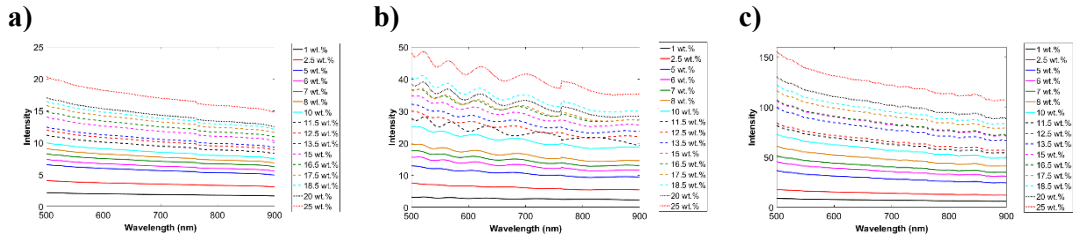


Figure 7.10: Effect of solid concentration in the vis-NIR region, of an α-lactose monohydrate sample with a particle size range of 125-150 μm, when the light is emitted from the different angular sources, **a)** 0, **b)** 30 and **c)** 45°, and collected by the closest detector to the light source, at 0.3 mm.

In both Figure 7.9 and Figure 7.10, it is observed an increase in the incident angle in leads to an increase of the reflected intensity, similar to the observations in polystyrene beads. The shorter light path length is a result of the shallower penetration depth that the photons travel in the particulate systems when using angular incident light, leading to a stronger reflectance signal.

Contrary to previous observations in the polystyrene system, the spectra collected when using the 30° incident angle revealed a wavy shape. This was observed in all the conditions tested. The fact that all detecting fibres at different distances to the source fibre showed the same pattern suggests that this is unlikely to be caused by the detecting fibres. Moreover, if the pattern was due to the quality of the detecting fibres, it would also be observed when using other incident light sources. Another possibility could be due to damage in the 30° incident source. At the beginning of the experiment, an integrating sphere is used to normalise the signal collected from the samples. The integrating sphere uniformly scatters the incident light. If the wavy spectral shape was related with a damage of the 30° incident source, the wavy shape would be seen in the spectra collected from the integrating sphere measurement. As the unprocessed sample signal is divided by the integrating sphere signal, the shape of the signal would be carried away and likely to be removed. Therefore, the impact of the damaged light source should be removed through this normalisation and the angular source would have equal performance as those using other incident angles. In Figure 7.11 is shown the spectra from the integrating sphere and the unprocessed signal from the sample L125-150 at 25 wt.% solid loading collected from the configuration of D1–L2. It is observed that the prominent wavy shape is only presented in the spectra from the sample. The green dash light highlights a prominent shoulder in the unprocessed sample signal which is not observed in the integrating sphere. A subtle shoulder in the integrating sphere seems to appear at higher wavelengths. The peaks highlighted by the orange and black line although they seem to appear in the same region, they show higher intensity in the unprocessed sample. However, in Figure 7.10 is observed a shift of the waves towards smaller wavelengths from most of the concentration evaluated. An exception is found in the pattern from a suspension of 11.5 wt.% of lactose. This seems to indicate the pattern is not related with a damage on the 30° incident light.

Another possible cause for the wavy shape could be due to the probe position in the reactor, with the 30° incident light illuminating the impeller, the reactor walls or the other in-line probes, FBRM and PVM, and cause reflection. However, such spectral pattern was not observed during the polystyrene suspensions measurements, in which a smaller 250ml reactor was used. The lactose experiments were run in a 1L reactor, providing more space to set the probes in the reactor. Nevertheless, one of the main

differences from the polystyrene measurements is the material of the impeller used. In the 250ml reactor, the pitch-blade impeller was PTFE coated while in the 1L vessel the impeller was made of glass. However, if the interference would arise from the impeller, reactor wall or other probes, it would be expected this interference to be minimised in high turbid mediums, as the light emitted would not be in direct contact with any type of interference, only the particulate suspension. Yet, the wavy pattern is observed in high turbid mediums.

Although SAR-DRM probe was always set in the same place in the reactor, on the day of the experimental measurements, the probe would have to be removed to perform the calibration measurements for all the angular sources. Every time SAR-DRM probe was replaced in the reactor, its position would be assessed to minimise such interferences. It is unlikely that the probe had exactly the same position for all the 120 measurements performed which all revealed the same wavy pattern.

Multivariate regression analysis will be applied to assess whether there is a linear relationship from the sample measurements.

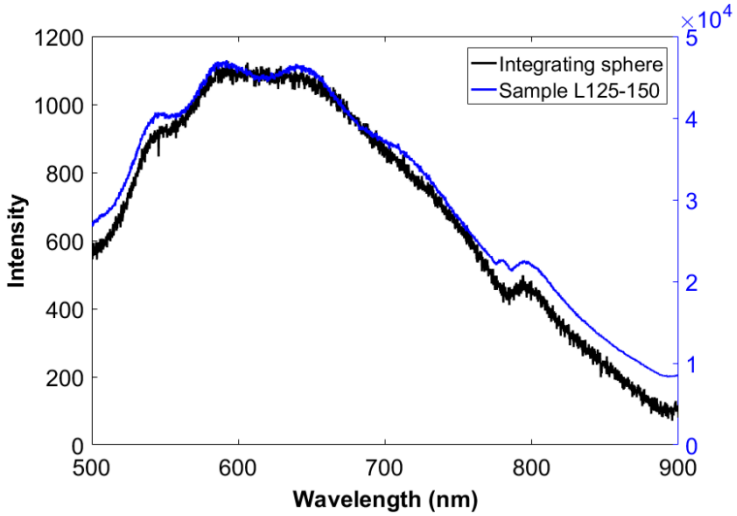


Figure 7.11: Spectra from collected from integrating sphere and sample L125-150 at 25 wt.%, from the 30° incident source. The detecting fibre is positioned at 0.3 mm from the light source.

7.3. Part 1 – Individual techniques calibration

In this section, PLSR analysis is performed on the unweighted and square weighted CLD, obtained from FBRM, and on the vis-NIR spectra, obtained from SAR-DRM, to predict the crystal size and solid loading. Individual models are initially developed for single source-detectors. A combination of different source-detector separations and angular sources is investigated to assess whether predictive improvements can be obtained, whether complementary information can help the models to capture distinct information about the composition, shape and solid loading of the lactose suspensions. Data augmentation and co-added strategies are explored. Different pre-processing methods are applied to investigate if further removal of baseline shifts and multiplicative effects are translated in better model performance. Finally, a data fusion approach is proposed. The data treatment/combination that leads to the best model performance for FBRM and SAR-DRM are selected to fuse by a data augmentation and co-adding approaches.

The dataset consists of nine distinct mean particle sizes, from <38 to 400 μm, suspended at different solid loadings in the suspension, from 0.5 to 25 wt.%. In total, 120 samples were evaluated. The dataset was split in 80 samples for the calibration set and 40 samples for the test set. The 40 samples in the test set were randomly selected. A detailed description of the dataset is shown in described in Chapter 4, Section 4.4. The split of the datasets was performed prior to any data pre-processing steps. Hence, the pre-processing parameters determined by the calibration dataset were applied to the test set. The model of validation employed is the leave-one-out cross-validation. The criteria followed in the establishment of best models is described in Chapter 4, Section 4.5.

7.3.1. FBRM

From Figure C.2 and Figure C.3 in Appendix B, it is observed that all the samples displayed a relatively smooth CLD profile. Hence no smoothing was applied to the dataset. This profile is also seen in Figure 7.12, which shows the unweighted and

square weighted CLDs collected from the 120 samples of lactose crystals. Scaling would amplify the noise and compress the main region of interest. In FBRM data only two types of pre-processing were applied, log transformation and SNV. From Figure 7.12, it is observed skewed distributions mainly towards larger chord lengths. This is related to the presence of crystals of different sizes and shapes in the suspensions. Although the main peak is at larger crystal sizes, the large number of fines will have an impact on the distribution, especially in the unweighted distribution where fine and coarse crystals have the same weight and same contribution to the distribution. The square weighted minimizes the impact of the fine crystals present in the suspension, however, they still contribute for the distribution. Besides that, these distributions, especially the unweighted, results in a high variation of the total counts, depending on the particle size and concentration used, as previously observed in Figure 7.8. The log transformation of the y-axis can be used to convert the variables into a more comparable scale, aiming to improve their linear relationship. The SNV is another method also used to reduce the count variation of the CLD measurements and increase the interpretability of the patterns such as peak shift and shape.

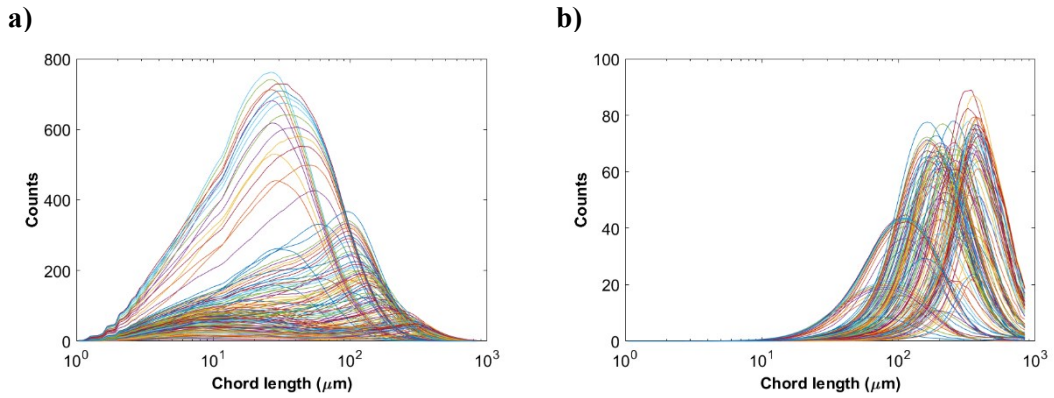


Figure 7.12: a) Unweighted and **b)** square weighted CLD of the 120 samples of lactose analysed.

Particle Size

Table 7.1 presents a summary of the best PLSR models using the unweighted and square weighted CLD with and without data pre-processing. Overall, the application of pre-processing methods led to models with better performance. For the unweighted

CLD, it is observed that the log and SNV transformation of the data, as well as the combination of both transformations, leads to lower errors of calibration, lower predictive errors and higher correlation coefficients, using a fewer number of LV to explain the data variance. These models exhibited similar performance and indicate that the data transformations helped the model to find a more linear relationship between the variables for the reasons described above, simplifying the model. The loadings from the combination of log and SNV transformations appeared to be capturing noise above 3 LVs.

Table 7.1: Summary of the results obtained from PLSR models in which the unweighted and square weighted CLD were used to determine lactose mean particle size.

Estimation of particle size				
Unweighted CLD				
Pre-processing	**No of latent variables**	**RMSECV (µm)**	**R^2**	**RMSEP (µm)**
None	5	70.2	0.707	84.4
Log	4	22.6	0.959	23.2
SNV	4	31.8	0.918	33.9
Log & SNV	3	23.5	0.955	24.0
Square weighted CLD				
None	2	67.6	0.73	82.2
Log	5	21.9	0.961	20.4
SNV	3	19.4	0.976	21.7
Log & SNV	4	24.9	0.950	18.2

The models built on the transformed square weighted CLD also led to lower errors of calibration and predictive errors. The removal of the effect of counts on the CLD by SNV led to a significant decrease in the RMSECV from 67.6 to 19.4 µm using only one extra variable. The linear relationship between the measured and estimated crystal size improved significantly and is shown by the increase of the R^2 from 0.730 to 0.976. SNV seems to significantly reduce variance ‘‘within sample’’ in the CLD leading to a separation of samples with different sizes by analysing the samples in a similar/comparable scale. The models built using log-transformed data needed higher number of LVs to explain the data variance than when using the SNV method. Figure 7.13 shows the SNV transformation of the square weighted CLD from the calibration

set, results in a clearer separation of the CLD peaks in comparison to the unprocessed data in Figure 7.12(b). Figure 7.14(a) shows the RMSECV curves of the PLSR models summarised in Table 7.1. In general, the transformed variables led to a smoother LVs profile. Table 7.1 also summarises the RMSEP of each model, which shows a good consistency to the RMSECV of the model. It is worthy to emphasise that the selection of the best model for each pre-processing does not include the consistency of RMSEP to RMSECV as a criterion, the results observed from Table 7.1 indicates the methodology setup for the calibration and test dataset is fair and robust for the purpose of this study.

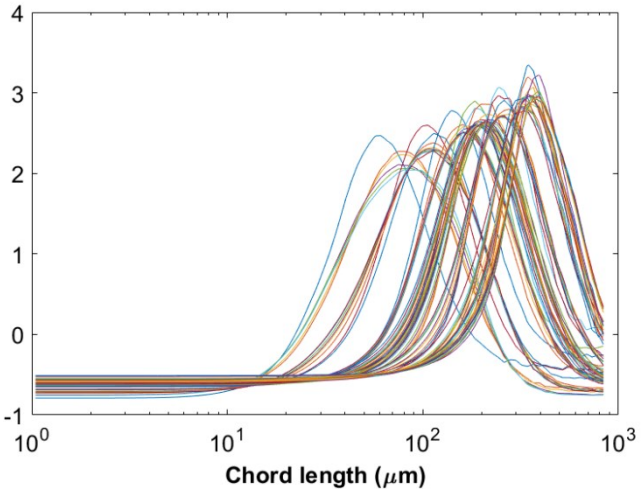


Figure 7.13: SNV transformation of the square weighted CLD of the calibration set.

Overall, the best model performance was achieved when using the SNV transformation of the square weighted CLD, which estimated the crystal size with an error of 19.4 µm and an R^2 of 0.976 using only 3LVs. Examining the loadings of the first three LVs in Figure 7.15(a1)-a(3), they seem be capturing the CL peak shifts with the particle. Figure 7.15 b(1)-b(3) displays the scores of the three LVs. The first LV seems to be capturing most of the information from particle sizes above 100 µm, while the second LV seems to be capturing extra information from the largest particle size groups (above 200 µm). The third LV seems to be capturing the remaining information from the crystals with a size smaller than 100 µm).

Figure 7.16(a1) shows the estimated particle size versus actual particle size using 3 LVs, in which an R^2 of 0.976 was obtained. The residuals plot Figure 7.16(b1) shows that samples L250-300 and L355-400 resulted in larger errors. Figure 7.16(a2) and b2)

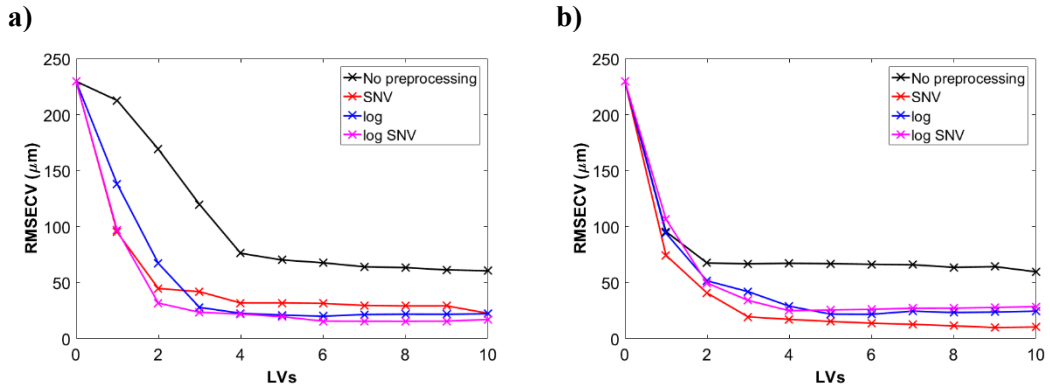


Figure 7.14: Cross-validation curves of PLSR models built on the **a)** unweighted and **b)** square weighted CLD for estimation of crystal size.

show the estimated versus actual particle size and the residual plot when using 4LVs to build the same model. The results are similar to the model using 3 LVs. Furthermore, a systematic error showing a U-shape in the residuals for particles above 200 µm clearly seen in Figure 7.16(b1), and also in (b2). This suggests that the CLD features correspond to larger particles are not well described in the model, and cannot be easily described by adding an extra LV to the model. On the contrary, the additional LV seems to cause the smaller particles (< 200 µm) to form a consistent, U-shaped pattern in Figure 7.16 (b2). The difficulty in determining the number of LVs from the RMSECV curve, coupled with the systematic error observed in Figure 7.16, indicates the non-linearity CLDs which presents a challenge in building a linear regression model on the data alone.

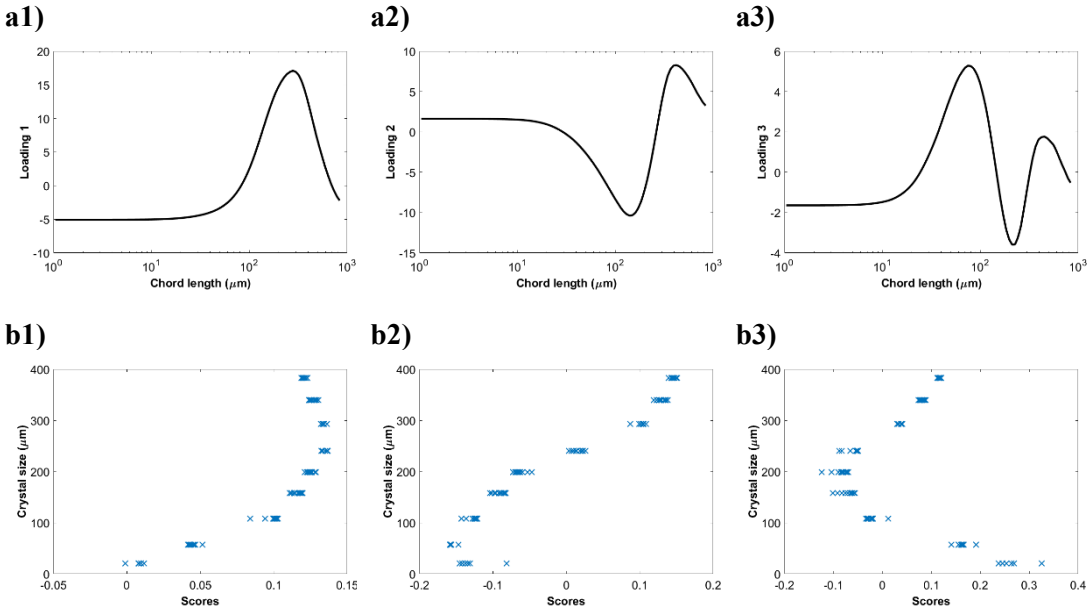


Figure 7.15: (a1)-(a3) loading curves and **(b1)-b(3)** scores for the first 3 LVs of the model built on the SNV transformation of the square weighted CLD for estimation of lactose crystal size.

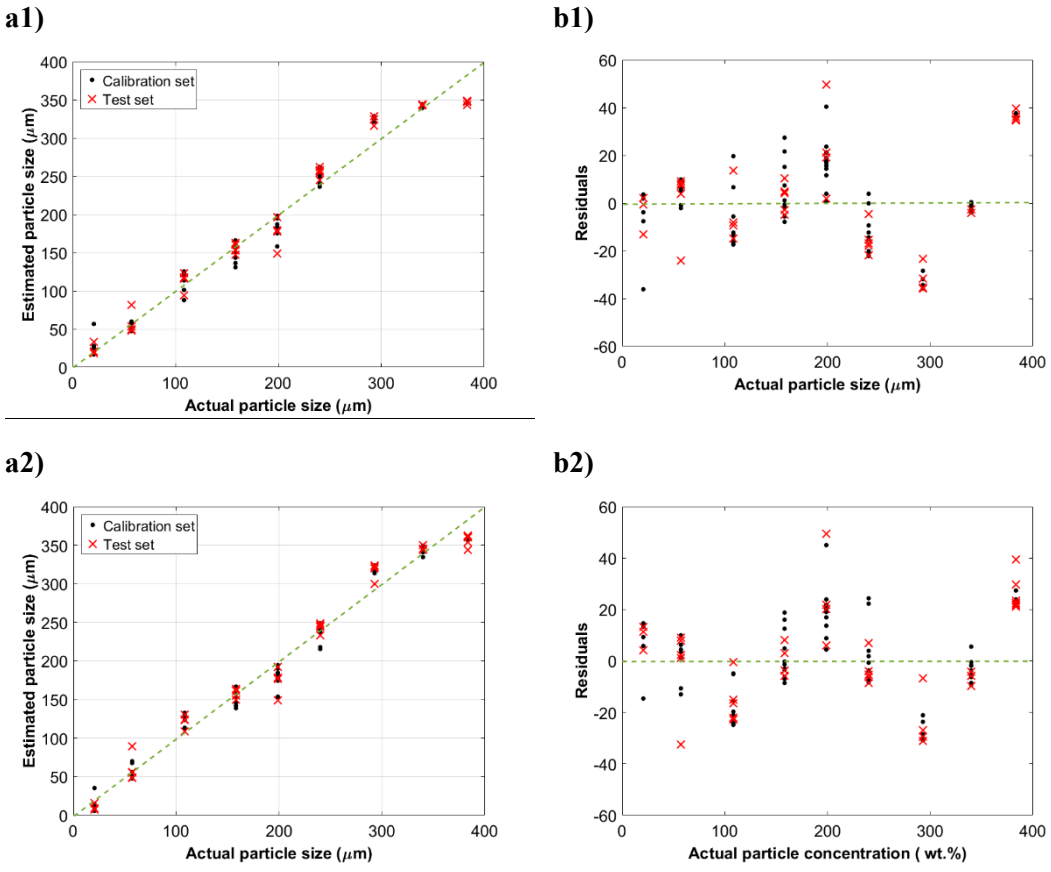


Figure 7.16: Prediction models build on SNV corrected square-weighted CLD with different number of LVs. **a)** Estimated versus actual particle size from lactose suspensions and **b)** residuals in the calibration dataset using **(1)** 3 LVs and **(2)** 4 LVs to explain the data variance.

Particle Concentration

The same methodology was followed to estimate the crystal concentration of the lactose suspensions. Table 7.2 summarises the results from the PLSR regression to estimate particle concentration when using the log and SNV transformations of the data. Although the SNV removes the effect of counts which are related to the solid concentration, it can be used to track the shifts in the CLD related to particle concentration of samples with the same size range, as shown in Figure 7.7, and assess whether there is a linear relationship. Overall, the model performance was poor. In comparison to the unprocessed unweighted and square weighted CLD, the use of the log transformation led to higher errors of calibration and lower coefficient of correlation. The rescaling of the y-axis though this method seems to lead to the loss of distinct features of the CLD.

The removal of the count interference from the model, using SNV or the combination of log and SNV on the unweighted CLD, led to a more complex model, with 9 and 10 LVs respectively. Figure 7.17 displays the RMSECV curves of the PLSR models built on the unweighted and square weighted CLD using the different pre-processing methods. In this figure it is clearly observed the additional number of LVs required to explain the models using the SNV and both log and SNV transformation from the remaining models.

In Figure 7.7(a) and Figure 7.8 it was observed an increase in the number of the counts with the solid loading. The elimination of this effect forces the model to look for subtle changes in the peaks shifts. In Figure 7.7(a) is observed a small but clear shift towards smaller CLs with the increase of the concentration in the unweighted distribution. Hence, a higher number of LVs is needed to capture the peaks shift variance for all the concentrations from the different particle sizes used. The use of SNV and the combination of log and SNV transformation on the unweighted and square-weighted CLDs resulted in higher RMSECV in comparison to the respective raw CLDs. These transformations and pre-processing methods result in even higher RMSECV for the square-weighted CLD as the shift of the CLs bands with the solid loading is not as straightforward as in the unweighted CLD, as observed in Figure 7.7(b). The majority of the peaks seem to be overlapping. This increases the model complexity and the

struggle to find a linear relationship between the concentration and the peaks shift. Even though the best PLSR model built on the SNV transformation of the square weighted only needed 3 LVs to explain the model, the RMSECV is twice of that for the models using the raw data.

Table 7.2: Summary of the results obtained from PLSR models in which the unweighted and square weighted CLD were used to determine lactose crystal concentration in acetone.

Estimation of particle concentration				
Unweighted CLD				
Pre-processing	**No of latent variables**	**RMECV (wt.%)**	**R^2**	**RMSEP (wt.%)**
None	5	2.86	0.828	2.85
log	4	3.58	0.685	3.67
SNV	9	3.28	0.736	3.42
Log & SNV	10	3.13	0.769	3.06
Square weighted CLD				
None	3	3.00	0.808	3.01
log	4	3.59	0.685	3.61
SNV	3	5.9	0.135	6.24
Log & SNV	10	3.37	0.728	2.83

The two models reporting the lowest errors of calibration were the models built using the raw unweighted and raw square weighted CLD, with errors of 2.86 and 3.00 wt.% respectively. However, the decrease of 0.14 wt.% does not justify the use of the two extra LVs obtained when using the unprocessed unweighted CLD. Hence, the best model to estimate particle concentration using the FBRM data is achieved when using the unprocessed square weighted distribution, which showed an RMSECV of 3 wt.% and an R^2 of 0.808, using 3 LVs which leads to an RMSEP of 3.01 wt.%. Examining the loadings and scores of these three latent variables in Figure 7.18, they appear to explain variations that affect the CLD peaks shift and their variation in counts with the concentration. It was found that the scores of the first latent variable are capture mainly the variance in solid concentration from samples with particle sizes above 90 µm, and the scores of the second latent variable capture the remaining information from the largest particle sizes, from 250-400 µm, which correspond to the peak position evidenced in the loading of the respective LV while compensating the shift in the other

direction for smaller size groups by the negative score. The third LV captures information from the samples with the smallest particle size group.

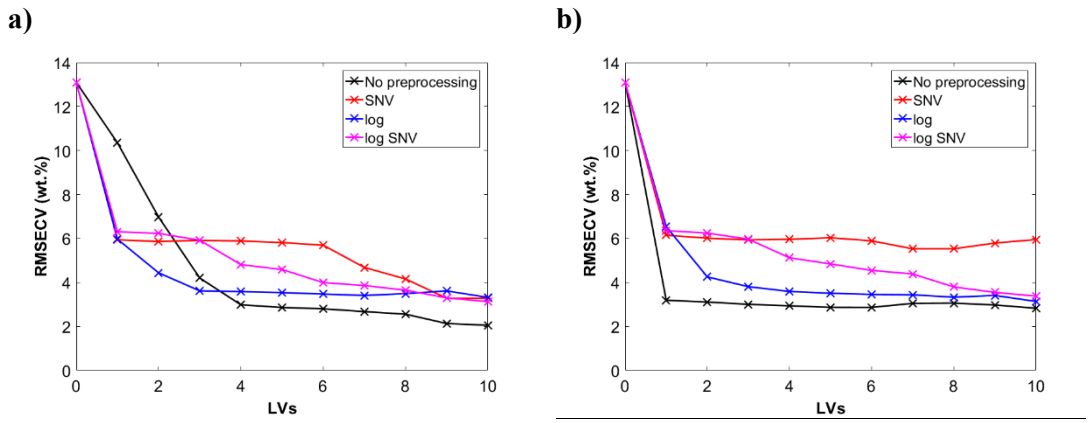


Figure 7.17: Cross-validation curves of PLSR models built on the **a)** unweighted and **b)** square weighted CLD for estimation of crystal concentration.

Figure 7.19 shows the estimated versus actual particle concentration, in which an R^2 of 0.828 was obtained, and the residuals plot of the model built using the raw square weighted CLD. Clearly, the residuals are concentration-dependent. The curvature in both plots strongly suggests the need for a non-linear model to explain better the data variance. Additional LVs only result in a decrease of the RMSECV of ~0.1 wt.% per LV extra, meaning that the model be improved by linear regression analysis. This is due to the non-linear nature of the counts, as shown in Figure 7.8, and peak shift with the increase of solid loading. In Figure 7.7(b) for example, it is observed a small shift to higher CLs with the increase of concentration, which seems to be shifting to the left after it reaches an optimum concentration.

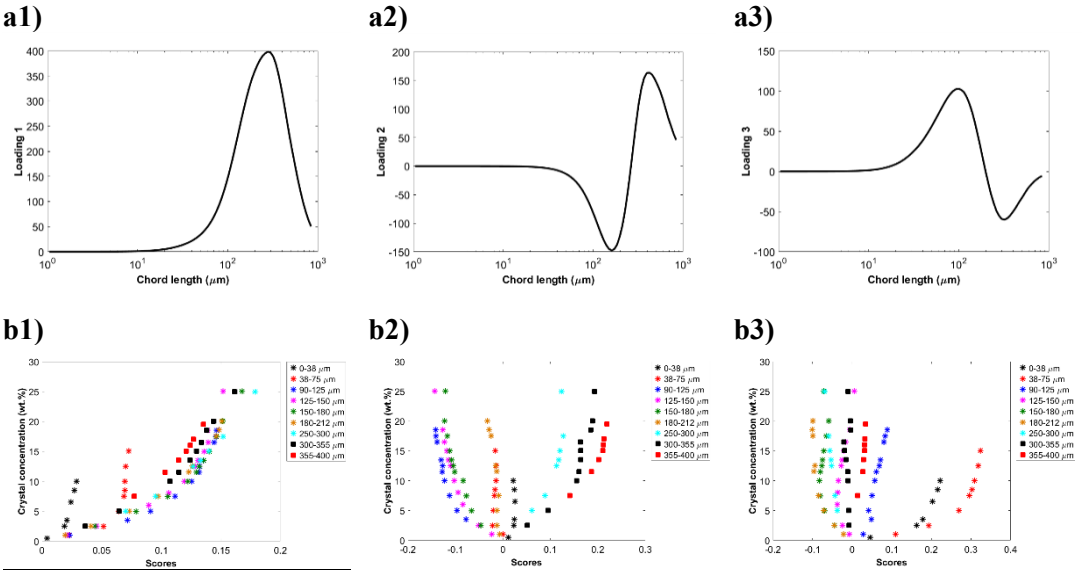


Figure 7.18: (a1)-(a3) loading curves and **(b1)-b(3)** scores for the first 3 LVs of the model built on the SNV transformation of the square weighted CLD for estimation of lactose solid concentration.

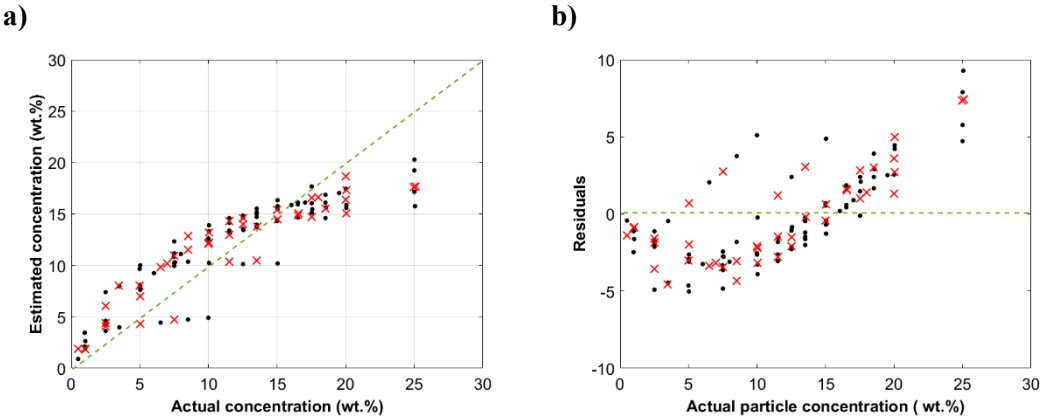


Figure 7.19: a) Estimated versus actual particle concentration from lactose suspensions using the unprocessed square weighted CLD. **b)** Residuals in the calibration dataset.

These results are in agreement with the results obtained from polystyrene beads, in which the SNV transformation of the square weighted CLD led to a better performance model to predict particle size and the raw data was better to predict particle concentration.

7.3.2. SAR-DRM

PLSR models to estimate particle size and concentration were built using the vis-NIR data collected from SAR-DRM. The wavelength range selected for analysis was from 500 nm to 900 nm due to the quality of the signal acquired in this region. The performance of each configuration used is assessed. Measurement combinations from different source-detectors are investigated, through a data augmentation and co-adding approaches, to improve the model performance.

In all datasets, a smoothing window of 15 points using a least-squares second-order polynomial fit was applied. All the datasets were subjected to mean centring.

7.3.2.1. Vis-NIR data fusion – mean particle size and concentration

Particle size

Similarly to the polystyrene study in Chapter 6, the first steps of the analysis are to select the best performing configurations for each incident angle used, which will subsequently be used to study different combination of configurations on the model performance. Table 7.3 summarises the detectors and angles evaluated, where D represents the detector distance at (1) 0.3, (2) 0.6, (3) 0.9 and (4) 1.2 mm from the light source while attributing a case number to each one of them.

Table 7.4 presents the PLSR analysis models results for the estimation of lactose mean particle size using the raw and log-transformed data. Figure 7.20 shows the number of LVs needed to explain each of the models built using raw data (solid line) and log-transformed data (dashed line). Overall, all models result in large RMSECV errors.

In Table 7.4, it is observed that although the log transformation of the spectra led to, in general, lower RMSECV, the models developed using the raw vis-NIR spectra is less complicated as it employs a fewer or equal number of LVs to describe the data variance in the model. Cases 1, 4, 5, 7 and 10 needed higher number of LVs to explain the variance in the log-transformed data. The predictive performance of the models developed using the spectra collected from the 30° incident light was relatively poor,

resulting in the highest errors of RMSEP for both raw and log-transformed spectra. An exception is observed in Case 7. Further investigation on the loading curves suggests that the models are not able to capture and distinguish the spectral features related to the wavy pattern observed in Figure 7.9(b).

Table 7.3: Initial measurement configurations cases used to build PLSR models. These were constructed using spectra from different detector distances (D1 – 0.3 mm, D2 – 0.6 mm, D3 – 0.9 mm, D4 – 1.2 mm) from the light source 0°, 30° and 45°.

	Angular source		
Case nº	**L1**	**L2**	**L3**
1	D1		
2	D2		
3	D3		
4	D4		
5		D1	
6		D2	
7		D3	
8		D4	
9			D1
10			D2
11			D3
12			D4

While the models developed using the raw spectra had relatively similar performance to those using the log transformation, there is a configuration for each angular source which exhibits an outstanding performance (i.e., Cases 1, 7 and 10) with a 50% decrease in the RMSECV. These cases, however, needed 3 or 4 extra LVs to explain the data variance in the models in comparison to the models built using the respective raw spectra. Despite their better model performance than the other cases, the RMSECV is still noticeably higher than those obtained from models built using CLDs. Therefore, further analysis is focussed on the possibility to improve the model performance using multiple configurations.

Table 7.4: Summary of the results obtained from PLSR models when using the vis-NIR spectra and its log transformation, to determine the lactose mean particle size.

	No transformation				Log transformation			
Case No.	**No of LVs**	**RMESCV (µm)**	**R^2**	**RMSEP (µm)**	**No of LVs**	**RMESCV (µm)**	**R^2**	**RMSEP (µm)**
1	5	58.4	0.727	63.4	8	36.2	0.896	35.5
2	7	56.0	0.749	54.9	7	68.2	0.655	53.7
3	6	56.8	0.739	61.6	6	64.5	0.674	55.3
4	7	77.2	0.55	67.8	8	77.0	0.563	67.9
5	5	76.6	0.532	86.5	6	61.9	0.694	78.65
6	6	66.2	0.651	93.5	6	67.6	0.638	80.8
7	6	51.5	0.786	76.6	10	43.2	0.851	49.4
8	6	56.6	0.742	76.0	6	68.4	0.627	71.84
9	10	58.9	0.726	63.3	10	48.2	0.821	37.3
10	6	56.6	0.742	57.7	10	23.0	0.957	22.2
11	8	41.8	0.860	49.2	7	43.5	0.846	44.6
12	7	42.7	0.853	51.4	7	40.1	0.871	44.0

The fact the raw spectra led to models of similar performance, with the fewer or same number of LVs than used for the log-transformed data, suggests these models are more robust. Hence, the selection of the two best detectors of each incident light was based on the models built using the raw vis-NIR spectra.

For the normal incident light, D1 and D2 were selected. D2-L1 (Case 2) was selected because it resulted in the lowest calibration error and lowest predictive error from the group. Log-transformed D1-L1 and D3-L1 were selected for a similar reason. Although D3-L1 exhibit a lower RMSECV and RMSEP than D1-L1, the decrease of RMSECV from 58.4 to 56.4 µm did not justify the use of an extra LV to explain the model. Moreover, the models developed for the polystyrene system in Section 6.1.2 and the study conducted by Chen *et al*. [9], have shown that the detectors at the shortest distances from the normal incident light source led to better models performance. In these cases, the diffused reflected light is a result of the shorter light path length taken,

resulting in spectra with high intensity values. The light beam is attenuated when passing through particulate systems, due to the increase of scattering events and absorption effects. Moreover, large particles, in the micrometre range, exhibit a preferred forward direction, the strong forward scatter events will be. The farther the detector from the incident light, the longer is the light path length taken. As a result of long path, the light penetrates deeper in the sample and may contain chemical and physical information which is not available at the samples superficie. Hence, if the signal collected from a distance fibre have a good quality, signal-to-noise ratio, it can unveil important information of the system.

When angular incident light was used, the configurations exhibiting the best model performance were those at the largest distances from the light source, D3 and D4. In both scenarios, the models resulted in the lowest errors of calibration and a higher correlation coefficient from the group. These could be due to the fact the light traveling shallower in the suspensions than when light is emitted by the normal incident light. Even though the light has gone through an increased number of scattering events and therefore, travelled longer path lengths, the results suggest it contains meaningful information of the crystal size in the system. The shortest distance detectors, contain only information from short light path lengths and it might be contaminated by single scattering events.

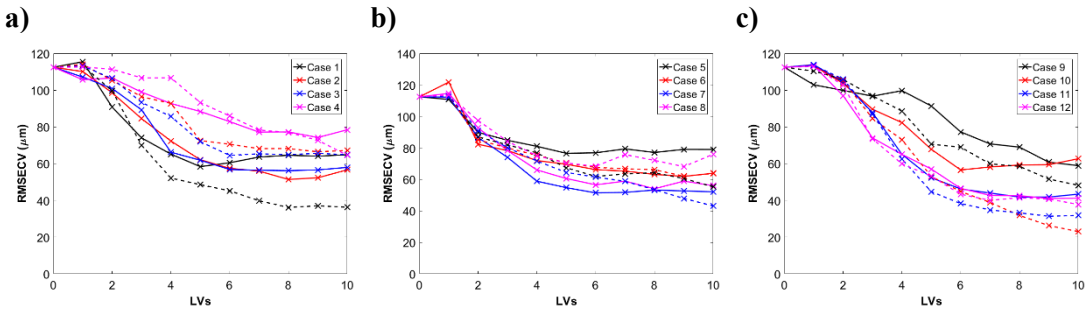


Figure 7.20: RMSECV versus LV obtained from the PLSR analysis using the spectra from each one of the detectors at different distances from the light source at **a)** 0°, **b)** 30° and **c)** 45°. The dashed line represents the RMSECV obtained when transforming the respective spectra (cases) in a logarithm scale.

The selected configurations were used to form combinations of multiple configuration in order to assess whether an improvement in the PLSR model performance could be achieved. Similar to the analysis in Chapter 6, due to a large number of possible combinations, only the configuration exhibiting the best model performance are chosen to form the combinations to simplify the discussion. Table 7.5 summarises the combinations studied where data were combined using either augmentation and co-adding approaches. Cases 13-15 examine the impact of combining all configurations from each incident light angle. Cases 16-19 examine different combinations of all spatial configuration from different incident light angles. Cases 20-26 are the combinations formed using the configurations selected from Table 7.4.

Table 7.5: Measurement configurations cases used to build PLSR models and compare the data co-adding and augmentation approaches. These were constructed using spectra from different detector distances (D1 – 0.3 mm, D2 – 0.6 mm, D3 – 0.9 mm, D4 – 1.2 mm) from the light source 0°, 30° and 45°.

	Angular source		
Case nº	**0°**	**30°**	**45°**
13	D1:D4		
14		D1:D4	
15			D1:D4
16	D1:D4	D1:D4	
17		D1:D4	D1:D4
18	D1:D4		D1:D4
19	D1:D4	D1:D4	D1:D4
20		D4	D3
21	D1		D3
22	D1		D3:D4
23	D1	D4	D3
24	D2	D4	D3
25	D1:D2	D4	D4
26	D2	D4	D3:D4

The best PLSR models from these combinations are summarised in Table 7.6. When combining data from all spatial configurations with different incident light angles, i.e., Cases 13 to 19, it was observed that the co-adding approach leads to more complex

models, using more LVs to explain the data variance, and to higher RMSECV and RMSEP. Cases 20-26 had a relatively similar performance in both data approaches. However, the co-adding approach led to, in general, higher RMSECV. This suggests the data augmentation approach is capturing patterns from the system that are essential to distinguish samples of different solid loadings.

Table 7.6: Summary of the PLSR calibration model performance for estimating lactose particle size using data augmentation and co-adding approaches in the log-transformed vis-NIR spectra.

	Augmented				Co-added			
Case No.	**No of LVs**	**RMESCV (µm)**	**R^2**	**RMSEP (µm)**	**No of LVs**	**RMESCV (µm)**	**R^2**	**RMSEP (µm)**
13	2	47.8	0.815	62.8	6	55.0	0.759	61.5
14	3	52.6	0.776	60.4	5	69.5	0.613	87.1
15	6	57.2	0.738	63.0	6	75.7	0.557	71.9
16	5	42.9	0.851	53.0	5	65.9	0.651	82.8
17	6	40.8	0.865	49.7	6	66.4	0.652	82.7
18	4	46.1	0.828	54.3	6	72.0	0.595	70.8
19	7	39.1	0.876	47.8	6	64.7	0.669	81.0
20	8	38.0	0.883	51.1	8	39.9	0.871	51.4
21	7	39.9	0.871	47.2	8	42.4	0.856	50.1
22	8	41.0	0.864	48.6	8	48.7	0.815	53.8
23	8	39.2	0.876	47.4	8	40.4	0.869	52.5
24	8	40.3	0.868	46.7	8	40.7	0.866	50.2
25	7	40.9	0.865	52.8	8	40.0	0.871	47.9
26	8	38.9	0.878	50.9	5	52.4	0.779	61.9

Overall, the combination of information from additional configurations did not lead to a significant improvement on the PLSR models performance, as the spectra collected by the individual detectors at 0.9 and 1.2mm from the 45° incident light resulted in models with similar performance, as shown in Cases 11 and 12 in Table 7.4.

Nevertheless, the combination of all detectors and sources utilised in the experiment lead to slightly lower errors of calibration and prediction using fewer LVs.

Figure 7.21 displays the RMSECV curves of the PLSR models built for Cases 13-26. In the data augmentation approach is observed that the use of all detector-sources distances (Cases 13-19) quickly decreases the RMSECV and reaches a stable level of error, regardless of which incident angle was used. In the data augmentation approach, the combinations of the selected configuration (Cases 20-26) lead to a smoother profile but more complexity in the curves. This seems to be related to the model involving data from L2, suggesting the challenges remains for describing the wavy pattern in the spectra. On the other hand, the similarity in RMSECV curve in Figure 7.22 (d) suggested that the co-added spectra could be dominated by D3-L3 and D4-L3. This can be seen as a typical characteristic, and a drawback, of conventional co-adding approach where the information in the combined spectra can be dictated by a signal response which is not necessarily the most sensitive to the changes in the sample condition.

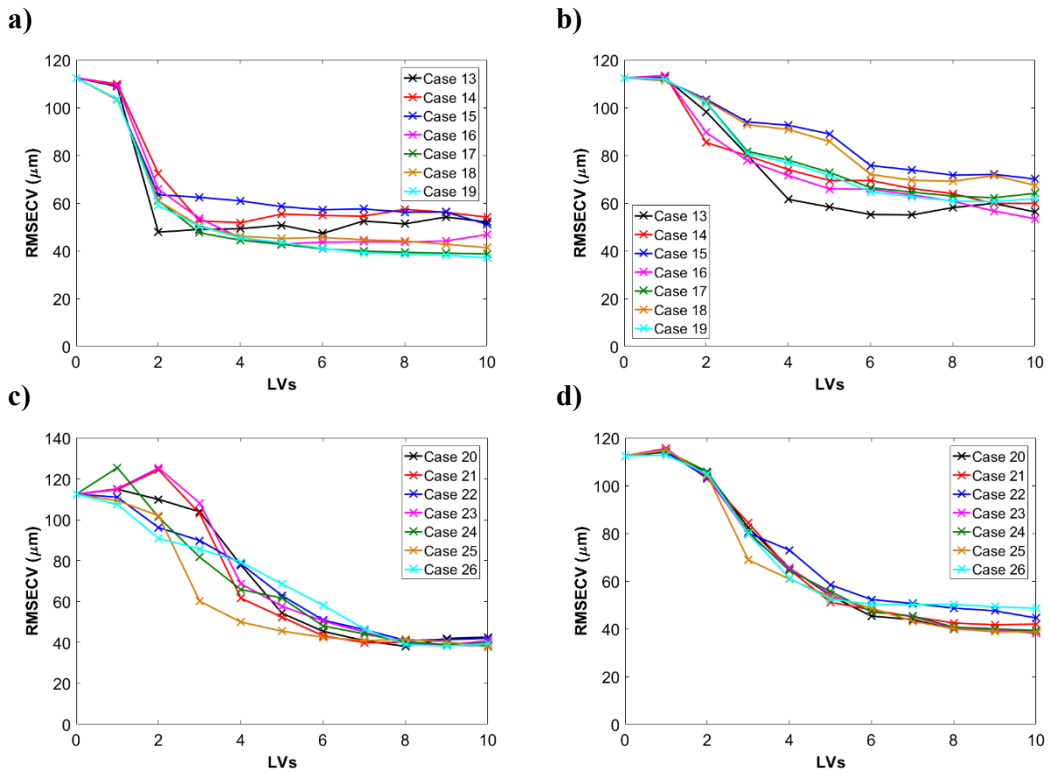


Figure 7.21: Cross-validation curves of PLSR models built for estimation of lactose particle size using **a)** data augmentation and **b)** co-adding approach for **(1)** Cases 13-19 and **(2)** Cases 20-26.

In order to assess whether empirical pre-processing methods can improve further the model performance, the following methods were used: log transformation, SNV, MSC, EMSCW, EMSCL and ISC in all the cases presented before. Differently from the polystyrene study, two variations of the MSC have been studied. As the results in Chapter 6 suggests that the use of MSC-based method could have an impact on improving model performance, it was decided to study the effect of using Extended MSC. More details of the methods can be found in Chapter 3. Both data augmentation and co-adding approaches were studied. Although the log transformation was discretised to select the single detectors with the best performance, it suggested that lower RMSECV and RMSEP could be obtained. Hence, it is here applied to compare model performance. To simplified the discussion, only the best model performances are shown.

Table 7.7 shows the impact of the different pre-processing methods on calibration models built to estimate lactose particle size. In both approaches, the application of correction methodologies led to lower errors RMSECV in comparison to the untreated data. The largest improvement by applying pre-processing methods was observed when applied to the co-added approach, however, it is also observed that the co-added data needs more LVs in the model. Even though the majority of the variance in the spectra, is captured by the first three LVs in the majority of the cases, it needs to use more LVs to capture information from the Y variable. The co-added approach of the different pre-processing methods also resulted in lower RMSECV and RMSEP than the data augmentation approach. Comparing the co-added results in Table 7.7 with the results in Table 7.6, for Case 22, the pre-processing methods led to not only lower RMSECV but also RMSEP.

For both approaches, the models built using different pre-processing methods show similar performance. The use of SNV, MSC and ISC lead to a better consistency among different models than the models developed to estimate mean particle size from polystyrene bead suspensions in water, in Chapter 6 Section 6.1.2. Differently from the polystyrene study, the spectra analysed are more consistent, showing a similar slope in all the conditions tested and a smoother signal, in the different pre-processing methods, as shown in Figure 7.22. This could be due to the range of particle

concentration used (from 2.5 wt% instead of 1wt% as the case for polystyrene system) and the similar shape of crystals analysed in each condition, as the shape could also influence the scattering effect, hence vis-NIR spectra. In the polystyrene study, the 1 wt.% of PS630-800 seem to lead to the highest noise in the spectra due to the low turbidity in the system. Increasing the number of particles in the suspension seems to reduce the spectral noise. These seem to lead to more comparable spectra for the lactose suspension and eliminate the issue of model deterioration by the noise.

Table 7.7: Impact of pre-processing methods (SNV, MSC and ISC) on calibration models built to estimate lactose particle size. The table presents the cases (measurement configurations) in which the pre-processing method resulted in the lowest RMSECV using data augmentation and data co-adding approaches.

	Augmented					Co-added				
Pre-processing method	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**	**RMSEP (µm)**	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**	**RMSEP (µm)**
None	**19**	7	39.1	0.876	47.8	**25**	8	40.0	0.871	47.9
log	**21**	8	23.0	0.957	21.6	**22**	7	37.5	0.897	44.8
SNV	**22**	4	35.5	0.898	53.0	**22**	7	25.2	0.948	42.8
MSC	**22**	4	34.8	0.902	54.2	**22**	7	25.1	0.949	43.6
EMSCW	**18**	5	28.9	0.933	39.2	**22**	6	25.4	0.948	28.2
EMSCL	**15**	4	35.8	0.989	53.3	**22**	7	23.7	0.955	40.1
ISC	**22**	4	34.8	0.902	53.8	**22**	7	25.6	0.947	35.0

The best model performance was achieved by using the EMSCW and a data augmentation approach, obtaining an RMSECV of 28.9 µm with an R^2 of 0.933 and an RMSEP of 39.9 µm, using 5 LVs to explain the model. Although the log transformation led to the lowest error of calibration, the decrease in the RMSECV from 28.9 to 23.0 µm does not justify the use of 3 extra LVs.

Although the co-added approach leads to lower RMSECV values, the data augmentation leads needs more LVs to explain the data variance. This suggests the

signal-to-noise ratio could be improved to enhance the model performance, by using co-added approach.

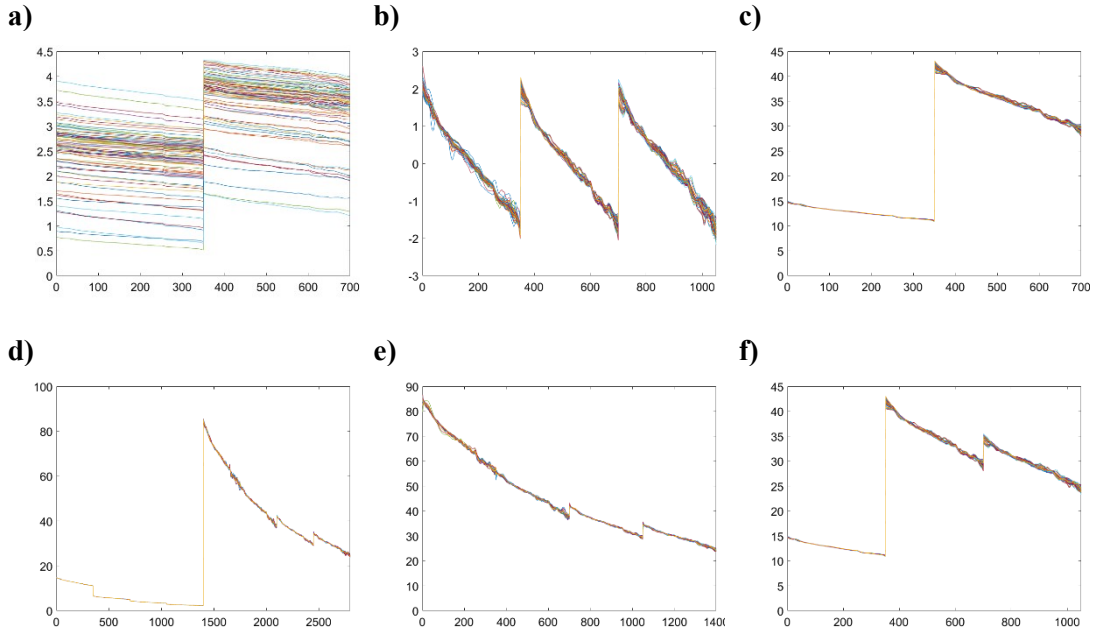


Figure 7.22: Impact of the pre-processing methods on the vis-NIR spectra. **a)** Log transformation, **b)** SNV, **c)** MSC, **d)** EMSCW, **e)** EMSCL and **f)** ISC.

All the pre-processing methods revealed a similar performance as they are here mainly used to remove the baseline shift of the spectra by only using different parameters linearly fit the data. This is due to the fact of the spectra being modelled being featureless, and with a consistent slope, which means the different parameters are only improving this linear relationship.

The use of pre-processing tools seems to be robust enough to distinguish or correct for the undesired effects in the vis-NIR spectra.

Particle Concentration

Similar to estimating the particle size, the PLSR analysis on particle concentration estimation were performed in the raw vis-NIR spectra and log-transformed data. In

general, using the log-transformed spectra led to higher RMSECV, hence it is not shown here.

Table 7.8 presents the PLSR results built for each source-detector distance described in Table 7.3. Overall, the errors of calibration and prediction were lower than the ones obtained using the CLD data reported in Table 7.2. This suggests SAR-DRM can track and discern better the patterns related to solid loading than FBRM. The two configurations for each incident light angle which exhibited the best models showed relatively similar performance. However, the models built using data from L2, i.e., Cases 5-8, resulted in the largest RMSEP and are significantly higher than the RMSECV of the models. This is in agreement with the results obtained for particle size estimation using SAR-DRM, as shown in Table 7.4. Considering the spectra from L2 exhibit a wavy pattern, the discrepancy between the RMSECV and RMSEP observed in Case 5-8 could suggest that the models do not capture the full variation in the samples, hence leading to the poor outcomes from the independent test set.

Table 7.8: Summary of the results obtained from PLSR models when using the original vis-NIR spectra to determine lactose crystals concentration.

Case No.	No of LVs	RMESCV (wt.%)	R^2	RMSEP (wt.%)
1	6	2.08	0.894	1.48
2	6	1.12	0.969	1.21
3	7	1.67	0.931	1.45
4	8	2.31	0.873	2.04
5	6	1.46	0.940	2.26
6	4	1.89	0.912	2.51
7	4	2.70	0.821	3.15
8	7	2.28	0.873	2.45
9	8	2.05	0.898	1.68
10	6	1.32	0.957	1.23
11	6	1.56	0.940	1.31
12	6	1.74	0.926	1.47

Figure 7.23 displays the RMSECV curves for each case in Table 7.8. In Figure 7.23(b) it is observed a clear performance separation between the detectors at the shortest and longest distances from the light source. In Figure 7.23(c) suggests the need of 10 LVs to explain the data variance of the detector at the shortest distance, 0.3 mm, however, the analysis on the loading curves (not shown here) suggests that only noise was capture for LVs higher than 8. The RMSECV of the detectors at 0.6 and 0.9 mm from the light source exhibit a smoother profile.

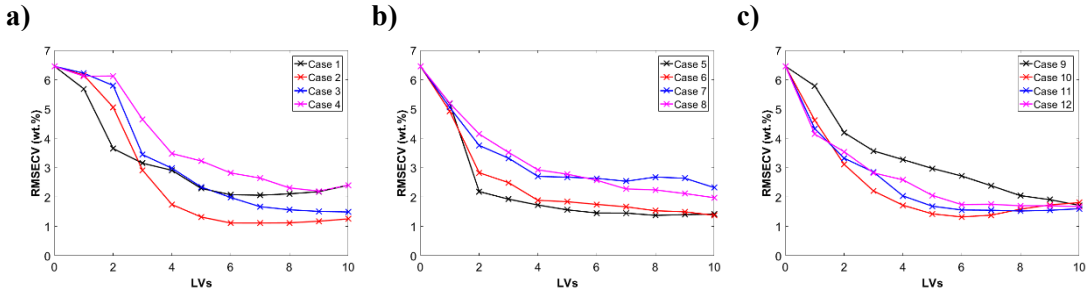


Figure 7.23: RMSECV versus LV obtained from the PLSR analysis to estimate particle concentration, using the spectra from each one of the detectors at different distances from the light source at **a)** 0°, **b)** 30° and **c)** 45°.

As conducted in the previous sections, the two best performing configurations using either normal and 45º incident light were selected. For the configurations using 30º incident light as the source, only D1 is selected to avoid further complication in the combined data due to the inclusion of multiple wavy patterns. The selected configurations are highlighted in colour in Table 7.8, and the combinations studied are summarised in Table 7.9.

Table 7.9: Measurement configurations cases used to build PLSR models and compare the data co-adding and augmentation approaches for lactose particle concentration estimation. These were constructed using spectra from different detector distances (D1 – 0.3 mm, D2 – 0.6 mm, D3 – 0.9 mm, D4 – 1.2 mm) from the light source 0°, 30° and 45°.

	Angular source		
Case nº	**0°**	**30°**	**45°**
13	D1:D4		
14		D1:D4	
15			D1:D4
16	D1:D4	D1:D4	
17		D1:D4	D1:D4
18	D1:D4		D1:D4
19	D1:D4	D1:D4	D1:D4
20		D1	D2
21		D1	D2:D3
22	D2	D1	D2
23	D2:D3	D1	D2
24	D2	D1	D2:D3
25	D3	D1	D2
26	D3	D1	D2:D3

Table 7.10 summarises the PLSR calibration and prediction model performance for estimating particle concentration using data augmentation and co-adding approaches. Similar to the estimation of lactose particle size using the vis-NIR, the combination of multiple configurations did not change significantly the model performance. This could be due to the small difference between the combinations analysed, leading to high consistency in the model performance. This is further evidenced by the RMSECV curved shown in Figure 7.24(a2) and (b2). The small variation in the configurations used in different cases suggests a similar co-added spectrum could be obtained for each case, leading to the similarity observed in the RMSECV curve in (b2).

Table 7.10: Summary of the PLSR calibration model performance for estimating particle concentration using data augmentation and co-adding approaches.

	Augmentation				Co-added			
Case No.	**No of LVs**	**RMESCV (wt.%)**	**R^2**	**RMSEP (wt.%)**	**No of LVs**	**RMESCV (wt.%)**	**R^2**	**RMSEP (wt.%)**
13	8	1.45	0.948	1.30	8	1.59	0.938	1.35
14	6	1.32	0.957	1.89	4	1.57	0.939	2.11
15	7	1.00	0.975	0.87	7	1.19	0.965	0.91
16	7	1.31	0.958	1.98	4	1.71	0.928	2.35
17	9	0.88	0.981	1.25	6	1.10	0.970	1.24
18	9	0.95	0.978	0.95	7	1.14	0.968	0.88
19	9	0.95	0.978	1.24	6	1.15	0.967	1.29
20	5	1.08	0.971	1.55	4	1.28	0.960	1.56
21	6	1.10	0.970	1.48	5	1.32	0.957	1.52
22	7	1.01	0.975	1.51	4	1.31	0.958	1.58
23	6	1.07	0.972	1.57	4	1.33	0.956	1.57
24	8	1.00	0.975	1.42	5	1.34	0.955	1.53
25	7	1.01	0.975	1.51	4	1.30	0.959	1.54
26	5	1.08	0.971	1.55	5	1.32	0.957	1.52

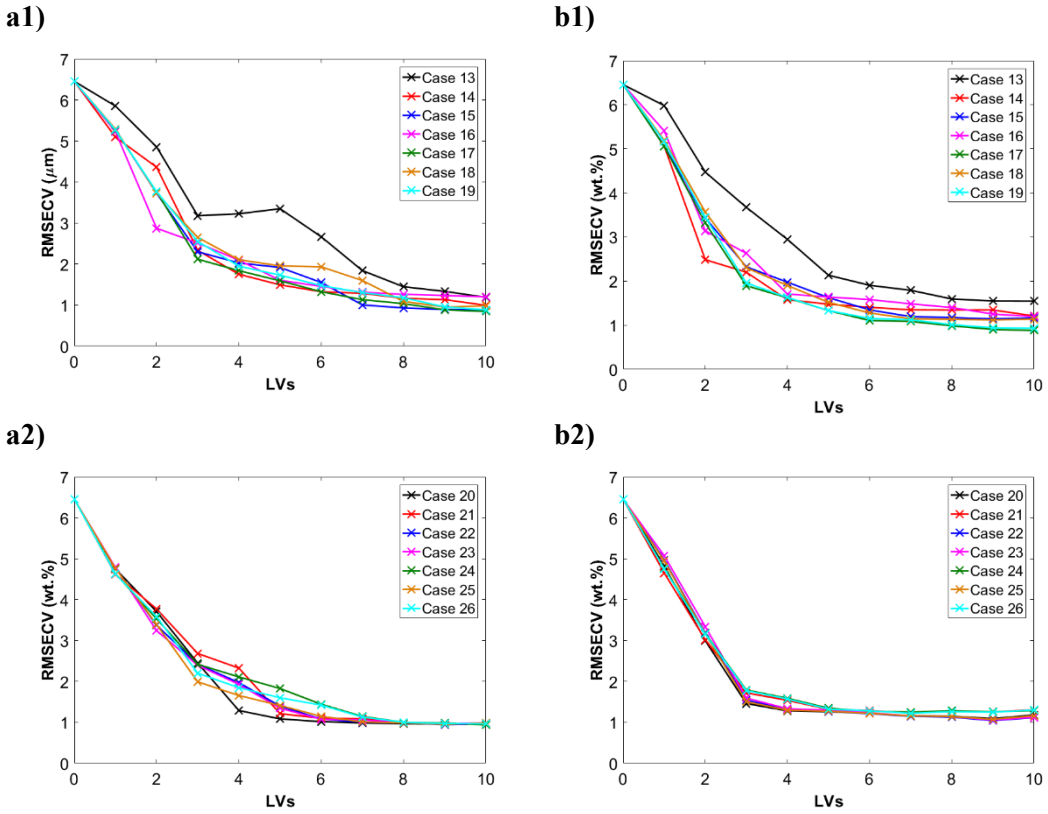


Figure 7.24: Cross-validation curves of PLSR models built for estimation of lactose particle concentration using **a)** data augmentation and **b)** co-adding approaches for **(1)** Cases 13-19 and **(2)** Cases 20-26.

The impact of pre-processing on the calibration models described in Table 7.11 was evaluated by using the SNV, MSC and ISC techniques, by the data augmentation and co-adding approaches. To simplify the discussion, only the best results for each pre-processing technique are shown and summarised in Table 7.11.

Overall, the use of pre-processing methods led to poor calibration models. As expected the data augmentation led to a higher number of LVs, extracting extra information provided by the multiple detectors. The co-adding approach, as it averages the multiple spectra in a single one, needing fewer LVs. However, the errors of calibration obtained through the co-adding approach are even higher.

In each approach, the SNV, MSC and ISC had a similar performance, presenting similar and high RMSECV values among them. These results indicate the pre-processing methods are not adequate applied are not robust enough to distinguish or

correct for the undesired effects in the spectra. This indicates a simple model can be built using only the raw spectral data.

Table 7.11: Impact of pre-processing methods (SNV, MSC and ISC) on calibration models built to estimate lactose particle concentration. The table presents the cases (measurement configurations) in which the pre-processing method resulted in the lowest RMSECV using data augmentation and data co-adding approaches.

	Augmentation					Co-added				
Pre-processing method	**Case No.**	**No of LVs**	**RMESCV (wt.%)**	**R^2**	**RMSEP (wt.%)**	**Case No.**	**No of LVs**	**RMESCV (µm)**	**R^2**	**RMSEP (wt.%)**
None	**15**	7	1.00	0.975	0.87	**18**	7	1.14	0.968	0.88
SNV	**18**	8	2.28	0.88	2.24	**13**	4	4.07	0.605	3.41
MSC	**18**	7	2.81	0.822	2.4	**18**	6	3.97	0.643	2.95
ISC	**18**	7	2.70	0.834	2.00	**18**	7	3.88	0.678	2.88

Although both data augmentation and co-adding approaches had relatively similar performances, using the same number of LVs to explain the data variance, the lowest error of calibration and prediction as achieved by using the data augmentation approach. Hence, the best model performance was obtained using the measurement combination described by Case 15 in Table 7.10 (using all the detectors distances from the 45° incident light), with a calibration error of 1 wt.%, with a correlation coefficient of 0.975 and a prediction error of 0.87 wt.%. Figure 7.25 exhibits the prediction and the residuals plots of the best performing models. There was a significant improvement from the models built using the CLD data (Figure 7.19). Using the vis-NIR data, the models were able to produce a linear relationship between the spectra collected and the particle solid loading.

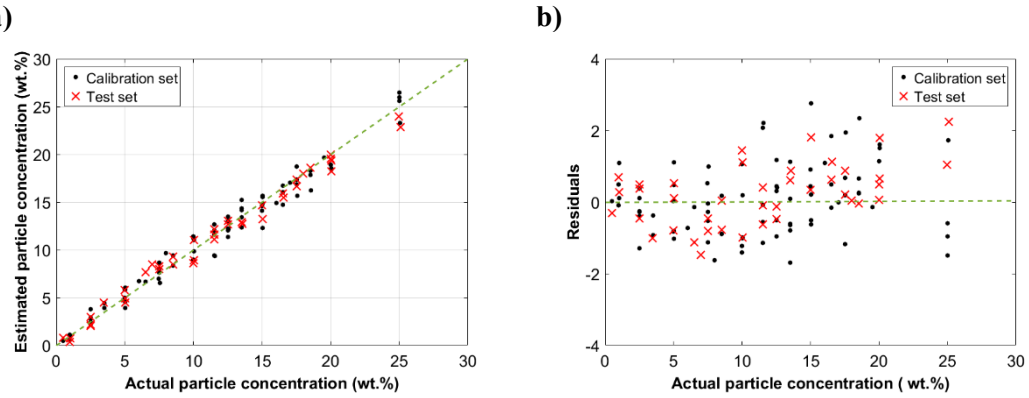


Figure 7.25: a) Estimated versus actual particle concentration from lactose suspensions using the unprocessed vis-NIR spectra. **b)** Residuals in the calibration dataset.

Figure 7.26 shows the vis-NIR spectra measurement combination used in Case 15, the loading curves and the scores for the first 3 LVs. The loadings appear to be explaining variation that affects the vis-NIR pattern, as it seems to capture information from spectra with different slopes, while the scores suggest the complicated response in the loadings to describe the intensity variation with the concentration for LV2 and 3. The scores of the first LV captured mainly the mean concentration from most size group with stronger correlation with L0-90. More complicated correlation between the scores and sample conditions were observed in, for example, Figure 7.27 (c2) and (c3), with the LV shows stronger correlation with certain particle size group. This reflects the non-linearity and complexity in the data and in each configuration used to collect the multiple scattered light.

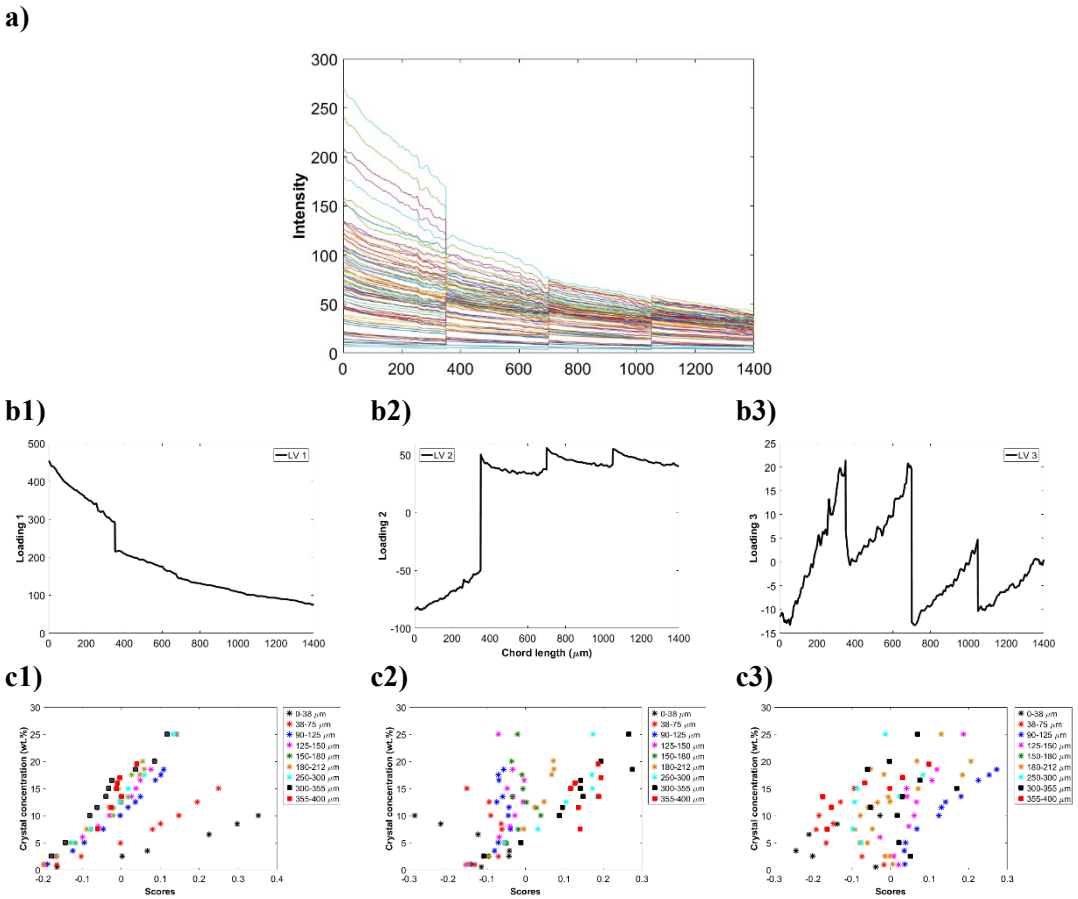


Figure 7.26: (a) The augmented vis-NIR spectra used for Case 15. **(b1)-(b3)** loading curves and **(c1)-c(3)** scores for the first 3 LVs of the model built a PLSR model for estimation of lactose solid concentration.

7.4. Part 2 – Multi-sensor calibration

Similarly to the polystyrene study, the first part of this chapter consisted of finding the best model performance for lactose mean size and solid loading estimation for each technique, FBRM and SAR-DRM. For FBRM, it was assessed the impact of using the raw unweighted and square weighted CLDs and the impact of pre-processing methods on the model performance. For the vis-NIR spectra acquired from SAR-DRM, it was investigated the impact of source-detector combinations on the calibration models by two data approaches, data augmentation and co-adding, as well as the impact of different pre-processing techniques. Table 7.12 summarises the best PLSR calibration models performance for estimating particle size and concentration for FBRM and SAR-DRM found in Part 1 of chapter 7.

Table 7.12: Summary of the best PLSR calibration models for estimation of α-lactose monohydrate suspensions particle size and concentration for each technique and wavelength region investigated.

Particle size estimation							
Instrument	**Configuration**	**Data combination approach**	**Pre-processing**	**No of LVs**	**RMECV (µm)**	**R^2**	**RMSEP (µm)**
FBRM	Square weighted CLD	-----	SNV	3	19.4	0.976	21.7
SAR-DRM (vis-NIR)	D1 – L1 D1:D4 – L3	Co-added	EMSW	6	25.4	0.948	28.2
Particle concentration estimation							
Instrument	**Configuration**	**Data combination approach**	**Pre-processing**	**No of LVs**	**RMECV (wt.%)**	**R2**	**RMSEP (wt.%)**
FBRM	Square weighted CLD	-----	None	3	3.00	0.808	3.01
SAR-DRM (vis-NIR)	D1:D4 – L3	Augmentation	None	7	1.00	0.975	0.87

In this section, the FBRM and SAR-DRM data which produced the best model performances are combined using co-adding and augmentation approaches. Table 7.13 summarises the PLSR model performances for estimating particle size and concentration. A significant improvement of the model performance for estimating particle size is observed when combining the square-weighted CLDs with the best performing configuration from vis-NIR spectra, reducing the RMSECV from 25 to ~9 µm. Both models built on augmented and co-added data exhibit a very similar performance, as shown in the smooth RMSECV curve displayed in Figure 7.27(a). This is likely due to the smaller size of co-added vis-NIR data block, hence reduce the differences between the fused data. On the contrary, the difference between the co-added and augmented fusion strategy is more prominent from models estimating particle concentration. In this case, the SAR-DRM data is an augmented data block, and when combined with FBRM, one could anticipate that the SAR-DRM data block would have larger influence than the FBRM data to the final model in the case of data fused by the augmentation approach. The best model to predict particle size is built on

the augmented dataset. Although the co-added approach led to a lower RMSECV than those built on the augmented data using the same number of LVs, the models built on the augmented data resulted in a higher R^2 and lower RMSEP. Figure 7.28 shows the prediction and residuals plots from the model built on the augmented data, which evidences the good fitting of the data, the accuracy and robustness of the model.

Table 7.13: PLSR calibration model performance for estimating particle size of α-lactose monohydrate suspensions using data augmentation and co-adding approaches on data fusion from different techniques.

Particle size estimation					
	Strategy	**No of LVs**	**RMSECV (µm)**	**R^2**	**RMSEP (µm)**
FBRM + vis-NIR	**Augmented**	5	9.49	0.995	8.07
	Co-added	5	9.10	0.993	9.07
Particle concentration estimation					
	Strategy	**No of LVs**	**RMSECV (µm)**	**R^2**	**RMSEP (wt.%)**
FBRM + vis-NIR	**Augmented**	10	0.93	0.979	0.80
	Co-added	8	1.24	0.962	1.28

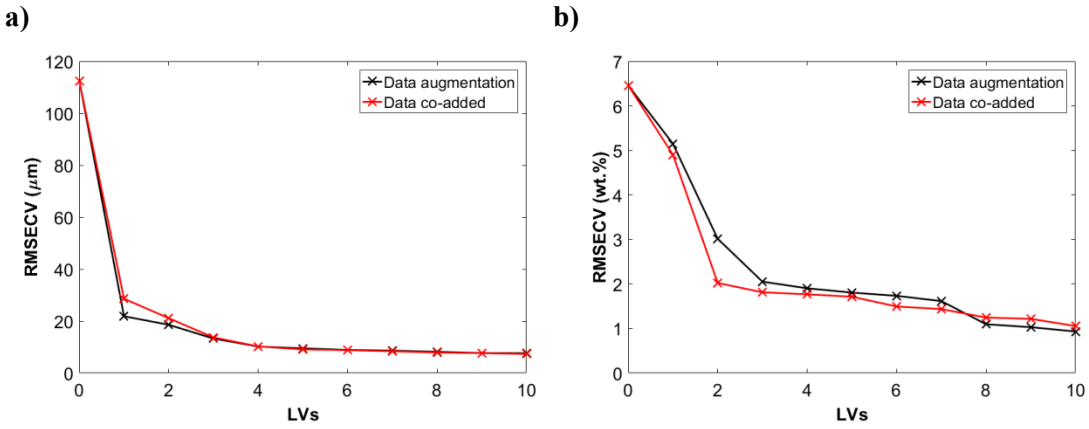


Figure 7.27: Cross-validation curves of PLSR models built for estimation of **a)** particle size and **b)** particle concentration using data augmentation and co-adding approaches for data fusion.

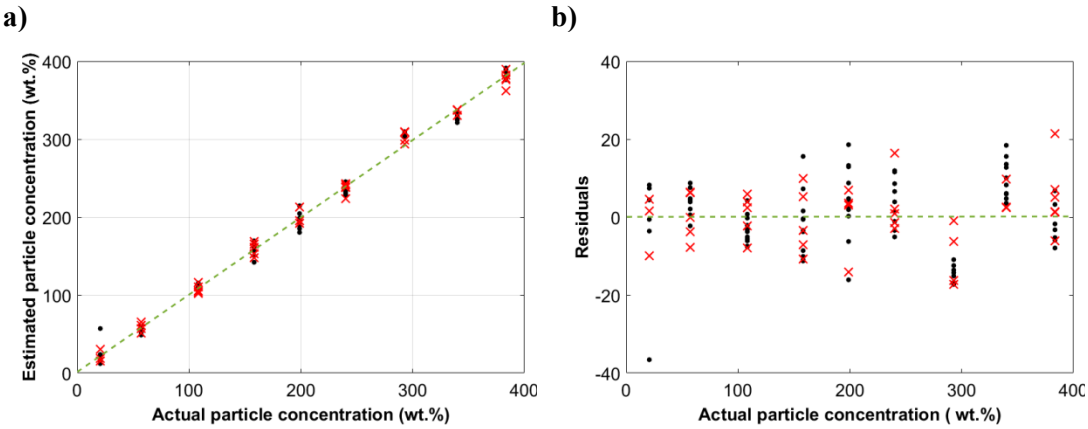


Figure 7.28: a) Estimated versus actual particle size and residuals from the model built on the augmented data.

The combination of the square-weighted distribution obtained from FBRM with the vis-NIR data did not have a significant improvement on the model performance for estimating particle concentration, in comparison to the best model built using only SAR-DRM data given in Table 7.12. As it was observed in Section 7.3.1., the FBRM data led to poor prediction performance, indicating the non-linear relationship between the CLD measured and the concentration. The addition of FBRM data to the best performing SAR-DRM data increases the number of LVs required for the model. This could be due to the model attempting to explain the variance in FBRM data block which does not vary linearly with the concentration. Figure 7.29 displays the prediction and residual plots for the model built on the augmented FBRM and SAR-DRM data. Although it is perceived a significant improvement from the model only using the FBRM data (Figure 7.19), the residuals still show the impact of FBRM data on the model, slightly resembling a curvature. Figure 7.29(c) shows the loadings from the respective model and confirms the increased model complexity as even the loading of the 10th LV can still be related to the spectral/CL feature in the data. These results highlight the limitation of FBRM to deal with high turbid samples and show that to assess highly concentrated SAR-DRM outperforms FBRM.

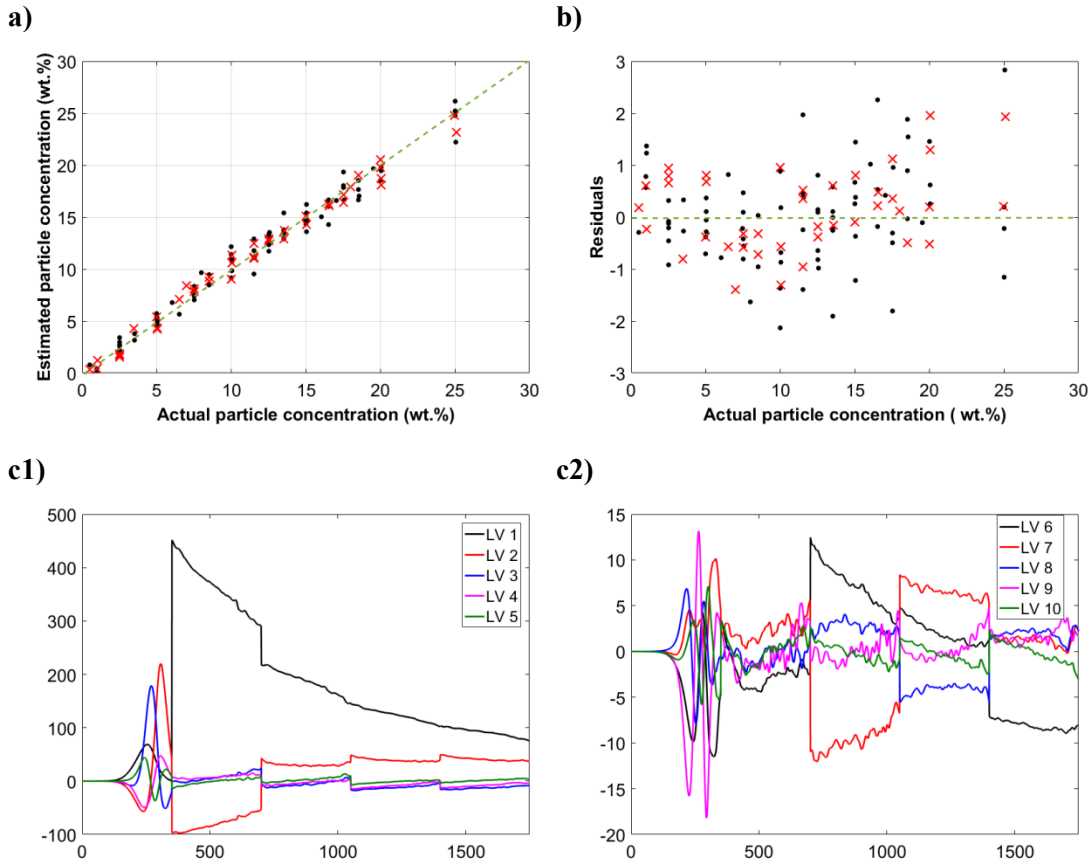


Figure 7.29: a) Estimated versus actual particle concentration and residuals from the model built on augmented data. Loadings from **c1)** LV1-5 and **c2)** LV 6-10.

7.5. Conclusions

The aim of this chapter is to investigate SAR-DRM response to crystals, using α-lactose monohydrate suspensions in acetone as a model system, and to assess whether accurate and robust models can be built to predict particle size and concentration using the spectra collected from the vis-NIR region. Its performance was compared and combined with FBRM.

Firstly, the FBRM data were analysed in order to investigate particle attrition mechanisms taking place while acquiring the measurement. It was observed that particles above 300 µm start showing abrasion. However, it was shown that up to 400 µm this effect seems negligible for this model system and the particle size distributions are not affected.

PLSR models were built for FBRM and SAR-DRM individually. In the case of FBRM, the removal of count interference by applying SNV to the CLDs led to an improvement of the model performance for estimating particle size. For estimation of particle concentration, poor performance models were obtained regardless the data treatment used. The results suggested a non-linear relationship among both counts and peak shift with the particle concentration which is difficult for the linear regression analysis such as PLSR to describe.

Overall, the models built using SAR-DRM vis-NIR spectral data showed that there is no correlation between the source-detecting fibre configuration and model performance. However, for both particle size and concentration estimation, the improvement of the predictive models is dependent on combining the best detecting fibres found.

The results show that better particle size predictive models can be obtained when fusing the FBRM and SAR-DRM data. For particle concentration estimation, it was observed SAR-DRM outperforms FBRM, providing more accurate and robust results when the models are built only using key combinations of configurations.

Chapter 8

8. Conclusions and Future Work

8.1. Conclusions

Real-time monitoring of particulates still presents a great challenge for pharmaceutical industries, for example, where particle size is either measured through off-line techniques, or represented by FBRM, which presents a CLD rather than a PSD. No information on the solid loading is typically obtained from any commercial technique applied. To date, only a few techniques are capable of in situ/in-line analysis for real-time process control and optimisation, and little has been reported on the comparison of the performance and reliability of different particle analysis methods. Hence, multiple off-line and in-line particle size analysis technologies were here compared. This comparison also allowed us to evaluate the potential of SAR-DRM to measure particle size and solid concentration. In this thesis, SAR-DRM was for the first time applied to monitor micrometre particles from <90 to 800 μm (previously, only turbid polystyrene particle suspensions between 100 to 500 nm were tested). As scattering and absorption coefficients are wavelength dependent, SAR-DRM signal can be related to chemical composition, concentration and particle size. Partial least square regression (PLSR) analysis was applied to provide a robust means of quantifying particle size and concentration and different data approaches and data fusion were evaluated to improve the models performance.

The investigation was carried out on two model systems, polystyrene beads suspensions in water and alpha lactose monohydrate suspensions in acetone, for size and concentration ranges relevant to many pharmaceutical processes. Both suspensions were monitored in situ using FBRM, PVM and SAR-DRM. The results form polystyrene samples demonstrate the challenging in characterising non-spherical particles, in comparing and obtaining consistent results from different particle size analysers. In Chapter 5, the off-line measurements (Mastersizer and Morphologi G3) and PVM demonstrated a better performance at differentiating the samples, even when

particles of similar size were present. However, the off-line methods require sampling, sample transportation which can in some cases change the sample properties and do not allow the instant feedback control. For the in-line tool, FBRM and PVM, it was observed that transparent and large particles, mainly above 250/300 μm, could cause chord length splitting in FBRM and exceed the field of view for PVM. High turbid medium can lead FBRM to measure particle aggregates, instead of individual particle, and result in loss of imaging resolution in PVM for further analysis. Chapter 5 also indicates the SAR-DRM good performance in such conditions, suggesting it to be used in linearly relating the SAR-DRM signal to the solid loadings. In this Chapter 5 it is also shown that SAR-DRM seems to be capable of differentiating suspensions of different particle sizes.

In order to extract quantitative information of particle size and concentration using the in-line technologies, robust calibration models were established for the individual technologies and for the combined information, for both polystyrene and lactose systems.

In the case of FBRM, the square-weighted CLD typically result in the best performing calibration models. For particle size estimation, the data pre-processing with SNV to remove the CL count interference had enhanced the interpretability of the data and led to improved calibration and prediction models for particle size estimation in both model systems investigated. For particle concentration, the best performing models were built on the raw CLDs. In the lactose system, as a larger variation in the particle concentration was included in the model, it is observed that the there was no linear relationship between the concentration and any feature of the CLD, indicating that a non-linear model would be more suitable.

In the case of SAR-DRM, the source-detecting fibres distances influence the signal quality and amount of information used. An increase in each spatial distance increases the probability of absorption and scattering events due to the longer light path length travelled, resulting in a weaker signal. Yet, as the light penetrates deeper in the sample, the photons may contain important chemical or physical information not available at the sample surface. The detectors at the shortest distance collect the diffused reflected light that travelled a shorter path which results in higher intensity signals. In the

calibration models built, no pattern was found between the source-detecting fibre configuration and the model performance. However, it was observed that by selecting and combining the best performing configurations, a significant improvement of the model performance for both particle size and concentration estimation is generally achieved. The combination of the configurations was performed using data augmentation and co-adding approaches. Overall, the data augmentation enhances model performance. This is because it allows extra variables that contain important information and features to be preserved and utilised by the PLSR model.

In general, the pre-processing of the spectra seems not effective to discern or correct undesired effects and consequently, improve the calibration models. The raw spectra and the log-transformed are usually the datasets leading to good model performance.

The results also suggest that the accuracy of multivariate calibration models can be improved by combining key SAR-DRM configurations with FBRM data. The fusion of information from these two techniques led to an improvement of the calibration models for particle size estimation. The results also show that SAR-DRM can outperform FBRM for particle concentration estimation. Such capability is related to the spectra captured by SAR-DRM containing more complete information about the light scattering properties of samples, which can be related to the particle size and concentration.

As SAR-DRM prefers high turbidity samples, i.e., samples with higher solid content, the study shows that SAR-DRM technology can be a potential complementary to other PAT tools such as FBRM and PVM, which tend to perform better for low turbidity samples, to achieve robust process monitoring and enable improved control and optimisation of manufacturing processes. This thesis shows SAR-DRM is capable of handling high solid loading over a large range of particle size and that is a promising technique to complement other in-line analysis methods for process control and optimization, especially in high turbid mediums.

8.2. Recommendations for Future Work

Following the conclusions of this thesis, some recommendations for future work are given in this section.

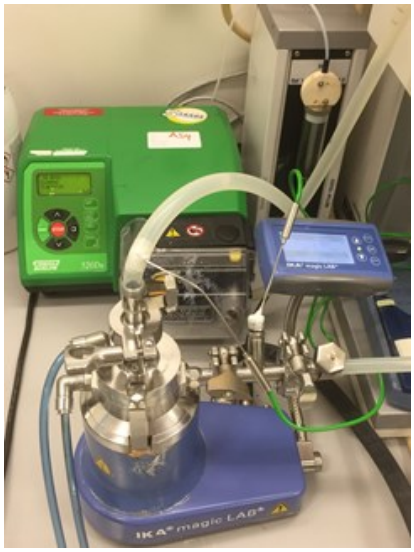
8.2.1. Optimisation of the wet milling study setup

Wet milling is commonly used in pharmaceutical industries to improve the solubility and bioavailability of poorly water-soluble drugs by causing the breakdown of coarse particles into finer ones through mechanical forces. Enhancing the particle attributes monitoring and control throughout the process can help to deliver consistent particulate products. Preliminary studies were performed to study both SAR-DRM response and the multi-sensor modelling strategy to monitor particle size and concentration in a wet milling process.

The model system used was α-lactose monohydrate suspension in acetone, with a particle size range of 212-250 μm, with a solid loading of 5 and 10 wt.%. A Mettler Toledo OptiMax™ workstation consisting of a 1 L stirred tank crystalliser (STC) equipped with an inline Hastelloy® Pt100 temperature sensor was used for all experiments with the system controlled by the *iControl* v5.4 software. FBRM, PVM and SAR-DRM were integrated within the reactor. An *IKA MagicLab (Module UTL)* wet mill unit was used for the inline milling process. Mill outlet temperature was monitored on the IKA operating unit display. The rotational speed used as 6000 rpm, as the lactose crystals can be easily fragmented. Figure 8.1 shows the full setup. However, this setup used needs to be optimised.

a)

b)



c)

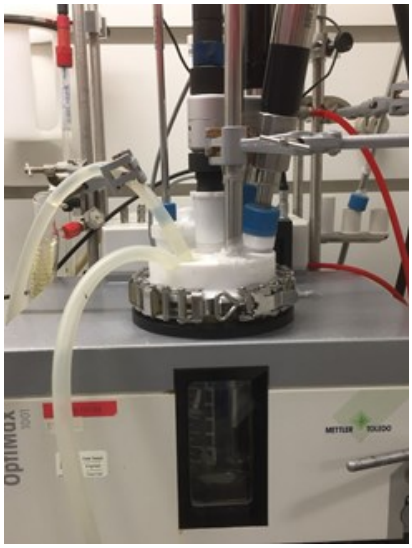


Figure 8.1: a) Full experimental set-up, **b)** wet milling machine and **c)** probes and wet milling tubes set-up.

In the set of experiments ran, it was noticed the constant linking of acetone from the wet mill unit, which would change the solid loading of the starting material. The studies performed in this thesis showed a dependent relationship between particle size and solid loading measured and the signal acquired, hence, a change in the solid loading could lead to high errors of predictions in the model for estimation of size and concentration. The material sealing the milling unit needs to be replaced. Although

blockage of the mill unit was only critical on the milling process run by steps, it was also observed high amount of the crystals deposited on the milling unit at the end of the continuous milling experiment. Even though the material is continuously pumped back to the reactor, the high density of the material compared to the density of the acetone and the fact that the suspension needs to go up against gravity, lead us to realise the flowrate used is not be powerful enough to move the material and avoid material accumulation. An alternative to this, would be get the pump and the milling unit at the same level of the reactor or a slightly higher. This way, gravity would help with the material flow.

During the preliminary studies it was also observed a curvy vis-NIR spectra rather than the typical straight spectra. This could be due to the presence of the rubber tubes immersed in the vessel, which are needed to feed the mill rotator with sample and return the sample to the vessel. Due to the packed reactor, these tubes are very close to SAR-DRM as observed in Figure 8.1(c) and are likely to be interfering with the scattering properties of the medium and contaminating the spectra. The angular sources are more likely to emit light that hits directly the tubes, hence, the detectors at different distances from these light sources are more sensitive in collecting diffuse reflectance light not only from particles but also from the tube.

Due to the speed of the milling process, in case of SAR-DRM, it is not feasible to use all 16 detectors as it would take about 5 to 10 minutes from the first detector to the last one, i.e., depending on the integrating times used for each region. The spectra collection in the vis-NIR region is fast and therefore, the detectors used to build the best calibration and prediction models, described in Chapter 8, could be used to monitor the process. As NIR measurements need longer integration times, it should be discarded from analysis. This would allow the in-line monitoring of the milling process, especially during the first minutes of milling in which the number of particle fragments seem to increase quickly and drastically, avoiding the milling process by stages. In a continuous milling process, another improvement would be the sampling of material for off-line measurements to match the in-line measurements, as all the reactor ports are occupied with probes and tubes.

8.2.2. Extraction particle size distribution and application of non-linear multivariate models

PLSR models could be built to extract the particle size distribution of polystyrene and lactose suspensions and its performance could be compared with off-line techniques, Mastersizer and Morphologi G3, and with the CLD inversion and imaging algorithm outputs, from FBRM and PVM respectively. During this work, it was observed a non-linear relationship between the CLD and counts with the particle concentration in study. Non-linear calibration models such as PLS-2, locally weighted regression (LWR) or artificial neural networks could be addressed to improve the models performance. A data fusion approach using Multiblock PLS could be explored and compared to data augmentation and co-added approaches, in order to assess which approach leads to the best models performance.

8.2.3. Extraction of absorption and scattering coefficients

The pre-processing methods applied in this thesis to remove the scatter effects from the spectra lead to models of poor performance when estimating particle concentration. Although models of more consistent performance were obtained for particle size estimation, the models could perhaps be improved by incorporating terms that describe the samples light propagation, such as their bulk optical properties μ_a, μ_s and anisotropic g factor (amount of forward scattering). It has been shown that SAR-DRM spectra can be inverted to obtain bulk optical properties [10] and subsequently to estimate PSD [11] of systems containing sub-micron particles. A similar approach should be investigated for particles in the micrometre range, which could make the SAR-DRM measurements free from calibration models.

8.2.4. Optimisation of the NIR integration time selection

Before starting the experimental work, the integration time for both vis-NIR and NIR wavelength regions was studied. Different integration times were studied for the same fibres to alter the noise in the signal. The defined conditions were applied to the remaining experiments. However, the spectral intensities vary throughout the experiments depending on the particle size and concentration of particles in the system.

The selected integrating time can be either too low, leading to low signal-to-noise ratio, or too high and lead to signal saturation. An automatic procedure to alter the integrating times should be incorporated as well as a software warning to inform the signal saturation.

8.2.5. Automation of the illuminating source fibres

At the moment, each angular light source needs to be manually change while taking measurements. Incorporating a controlled stepper motor and a rotary block to switch the light source according to a pre-defined sequence and time interval set by the user, would speed up the acquisition of measurements and make the equipment more independent from the user.

References

[1] A.N. Saleemi, Strategic feedback control of pharmaceutical crystallization systems, (2011) 1-194.

[2] Part One Process Control and Quality Assurance, Wiley-VCH Verlag GmbH & Co. KGaA, (2012).

[3] Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, http://www.fda.gov/downloads/Drugs/Guidances/ucm070305.pdf.

[4] R. Guenard, G. Thurau, Implementation of Process Analytical Technologies, Process Analytical Technology, John Wiley & Sons, Ltd2010, pp. 17-36.

[5] L.X. Liu, I. Marziano, A.C. Bentham, J.D. Litster, E.T.White, T. Howes, Effect of particle properties on the flowability of ibuprofen powders, International Journal of Pharmaceutics, 362 (2008) 109-117.

[6] B. Ahmed, C.J. Brown, T. McGlone, D.L. Bowering, J. Sefcik, A.J. Florence, Engineering of acetaminophen particle attributes using a wet milling crystallisation platform, International Journal of Pharmaceutics, 554 (2019) 201-211.

[7] H.G. Morrison, W. Tao, W. Trieu, S.D. Walker, S. Cui, S. Huggins, K. Nagapudi, Correlation of Drug Substance Particle Size Distribution with Other Bulk Properties to Predict Critical Quality Attributes, Organic Process Research & Development, 19 (2015) 1076-1081.

[8] P. Narayan, B.C. Hancock, The relationship between the particle properties, mechanical behavior, and surface roughness of some pharmaceutical excipient compacts, Materials Science and Engineering: A, 355 (2003) 24-36.

[9] Y.-C. Chen, D. Foo, N. Dehanov, S.N. Thennadil, Spatially and angularly resolved spectroscopy for in-situ estimation of concentration and particle size in colloidal suspensions, Analytical and Bioanalytical Chemistry, 409 (2017) 6975-6988.

[10] K. Thomson, Stoliarskaia, D., Tiernan-Vandermotten, S., Lue, L., Chen, Y-C., Simulation of diffuse reflectance for characterisation of particle suspensions, in: G.L.

Coté (Ed.) SPIE Photonics West 2017, Society of Photo-optical Instrumentation Engineers (SPIE), Bellingham, Washington, 2017.
[11] A. Postelmans, B. Aernouts, W. Saeys, Estimation of particle size distributions from bulk scattering spectra: sensitivity to distribution type and spectral noise, Opt. Express, 26 (2018) 15015-15038.
[12] J. Cardona, C. Ferreira, J. McGinty, A. Hamilton, O.S. Agimelen, A. Cleary, R. Atkinson, C. Michie, S. Marshall, Y.-C. Chen, J. Sefcik, I. Andonovic, C. Tachtatzis, Image analysis framework with focus evaluation for in situ characterisation of particle size and shape attributes, Chemical Engineering Science, 191 (2018) 208-231.
[13] B.Y. Shekunov, P. Chattopadhyay, H.H.Y. Tong, A.H.L. Chow, Particle Size Analysis in Pharmaceutics: Principles, Methods and Applications, Pharmaceutical Research, 24 (2007) 203-227.
[14] D. Elder, G. Pacowski, P. Nelson, S. Tindal, The importance of particle size analysis, European Pharmaceutical Review, (2017).
[15] A. Rawle, Basic Principles of Particle Size Analysis, Malvern Instruments Ltd.,Technical Paper, Worcestershire, UK, 1993.
[16] A.F.T. Silva, A. Burggraeve, Q. Denon, P. Van der Meeren, N. Sandler, T. Van Den Kerkhof, M. Hellings, C. Vervaet, J.P. Remon, J.A. Lopes, T. De Beer, Particle sizing measurements in pharmaceutical applications: Comparison of in-process methods versus off-line methods, European Journal of Pharmaceutics and Biopharmaceutics, 85 (2013) 1006-1018.
[17] M. Panalytical, Morphologi G3, 2015.
[18] Particle size and shape measurement using image analysis, Malvern Instruments Limited, Worcestershire, UK, 2016, pp. 1-4.
[19] R. Zhang, C.Y. Ma, J.J. Liu, X.Z. Wang, On-line measurement of the real size and shape of crystals in stirred tank crystalliser using non-invasive stereo vision imaging, Chemical Engineering Science, 137 (2015) 9-21.
[20] H.G. Merkus, Particle Size Measurements: Fundamentals, Practice, Quality, 1 ed., Springer Netherlands, Springer, 2009.
[21] L. Reimer, Scanning Electron Microscopy, Springer1998.

[22] S. Borrett, L. Hughes, Reporting methods for processing and analysis of data from serial block face scanning electron microscopy, Journal of Microscopy, 263 (2016) 3-9.
[23] C.J. Peddie, L.M. Collinson, Exploring the third dimension: Volume electron microscopy comes of age, Micron, 61 (2014) 9-19.
[24] P. Barrett, B. Glennon, Characterizing the Metastable Zone Width and Solubility Curve Using Lasentec FBRM and PVM, Chemical Engineering Research and Design, 80 (2002) 799-805.
[25] X. Liu, D. Sun, F. Wang, Y. Wu, Y. Chen, L. Wang, Monitoring of antisolvent crystallization of sodium scutellarein by combined FBRM–PVM–NIR, Journal of Pharmaceutical Sciences, 100 (2011) 2452-2459.
[26] Y.-H. Luo, Y.-R. Tu, J.-L. Ge, B.-W. Sun, Monitoring the Crystallization Process of Methylprednisolone Hemisuccinate (MPHS) from Ethanol Solution by Combined ATR-FTIR- FBRM- PVM, Separation Science and Technology, 48 (2013) 1881-1890.
[27] D. Treffer, P.R. Wahl, T.R. Hörmann, D. Markl, S. Schrank, I. Jones, P. Cruise, R.-K. Mürb, G. Koscher, E. Roblegg, J.G. Khinast, In-line implementation of an image-based particle size measurement tool to monitor hot-melt extruded pellets, International Journal of Pharmaceutics, 466 (2014) 181-189.
[28] P. Cruise, C. Cortazzo, J. Fröhlich, L. Sprich, R. Pila, I. Jones, Monitoring Fluid Bed Granulation Processes In-Line with Real Time Imaging, http://www.glatt.com/fileadmin/user_upload/content/pdf_downloads/News/Glatt_IPL_Fluid_Bed_Granulation.pdf.
[29] A.S. El Hagrasy, P. Cruise, I. Jones, J.D. Litster, In-line Size Monitoring of a Twin Screw Granulation Process Using High-Speed Imaging, Journal of Pharmaceutical Innovation, 8 (2013) 90-98.
[30] Z. Ma, H.G. Merkus, J.G.A.E. de Smet, C. Heffels, B. Scarlett, New developments in particle characterization by laser diffraction: size and shape, Powder Technology, 111 (2000) 66-78.
[31] ISO 13320:2009 Particle Size Analysis—Laser Diffraction Methods. Part 1: General Principles, 2009.

[32] D. Wiegel, G. Eckardt, F. Priese, B. Wolf, In-line particle size measurement and agglomeration detection of pellet fluidized bed coating by Spatial Filter Velocimetry, Powder Technology, 301 (2016) 261-267.
[33] A. Burggraeve, T. Van Den Kerkhof, M. Hellings, J.P. Remon, C. Vervaet, T. De Beer, Evaluation of in-line spatial filter velocimetry as PAT monitoring tool for particle growth during fluid bed granulation, European Journal of Pharmaceutics and Biopharmaceutics, 76 (2010) 138-146.
[34] A.N. Saleemi, G. Steele, N.I. Pedge, A. Freeman, Z.K. Nagy, Enhancing crystalline properties of a cardiovascular active pharmaceutical ingredient using a process analytical technology based crystallization feedback control strategy, International Journal of Pharmaceutics, 430 (2012) 56-64.
[35] S.C. Barthe, M.A. Grover, R.W. Rousseau, Observation of Polymorphic Change through Analysis of FBRM Data: Transformation of Paracetamol from Form II to Form I, Crystal Growth & Design, 8 (2008) 3316-3322.
[36] H. Li, M.A. Grover, Y. Kawajiri, R.W. Rousseau, Development of an empirical method relating crystal size distributions and FBRM measurements, Chemical Engineering Science, 89 (2013) 142-151.
[37] E. Kougoulos, A.G. Jones, K.H. Jennings, M.W. Wood-Kaczmar, Use of focused beam reflectance measurement (FBRM) and process video imaging (PVI) in a modified mixed suspension mixed product removal (MSMPR) cooling crystallizer, Journal of Crystal Growth, 273 (2005) 529-534.
[38] J.A. Boxall, C.A. Koh, E.D. Sloan, A.K. Sum, D.T. Wu, Measurement and Calibration of Droplet Size Distributions in Water-in-Oil Emulsions by Particle Video Microscope and a Focused Beam Reflectance Method, Industrial & Engineering Chemistry Research, 49 (2010) 1412-1418.
[39] M. Li, D. Wilkinson, K. Patchigolla, Comparison of Particle Size Distributions Measured Using Different Techniques, Particulate Science and Technology, 23 (2005) 265-284.
[40] J. Heinrich, J. Ulrich, Application of Laser-Backscattering Instruments for In Situ Monitoring of Crystallization Processes - A Review, Chemical Engineering & Technology, 35 (2012) 967-979.

[41] M. Li, D. Wilkinson, Determination of non-spherical particle size distribution from chord length measurements. Part 1: Theoretical analysis, Chemical Engineering Science, 60 (2005) 3251-3265.
[42] O.S. Agimelen, P. Hamilton, I. Haley, A. Nordon, M. Vasile, J. Sefcik, A.J. Mulholland, Estimation of particle size distribution and aspect ratio of non-spherical particles from chord length distribution, Chemical Engineering Science, 123 (2015) 629-640.
[43] J. Worlitschek, T. Hocker, M. Mazzotti, Restoration of PSD from Chord Length Distribution Data using the Method of Projections onto Convex Sets, Particle & Particle Systems Characterization, 22 (2005) 81-98.
[44] W. Yu, K. Erickson, Chord length characterization using focused beam reflectance measurement probe - methodologies and pitfalls, Powder Technology, 185 (2008) 24-30.
[45] A. Ruf, J. Worlitschek, M. Mazzotti, Modeling and Experimental Analysis of PSD Measurements through FBRM, Particle & Particle Systems Characterization, 17 (2000) 167-179.
[46] O.S. Agimelen, A. Jawor-Baczynska, J. McGinty, J. Dziewierz, C. Tachtatzis, A. Cleary, I. Haley, C. Michie, I. Andonovic, J. Sefcik, A.J. Mulholland, Integration of in situ imaging and chord length distribution measurements for estimation of particle size and shape, Chemical Engineering Science, 144 (2016) 87-100.
[47] D.J. McClements, Ultrasonic characterisation of emulsions and suspensions, Advances in Colloid and Interface Science, 37 (1991) 33-72.
[48] P.S. Epstein, R.R. Carhart, The Absorption of Sound in Suspensions and Emulsions. I. Water Fog in Air, The Journal of the Acoustical Society of America, 25 (1953) 553-565.
[49] C.M. Atkinson, H.K. Kytömaa, Acoustic wave speed and attenuation in suspensions, International Journal of Multiphase Flow, 18 (1992) 577-592.
[50] J.R. Allegra, S.A. Hawley, Attenuation of Sound in Suspensions and Emulsions: Theory and Experiments, The Journal of the Acoustical Society of America, 51 (1972) 1545-1564.

[51] D.M. Forrester, J. Huang, V.J. Pinfield, Characterisation of colloidal dispersions using ultrasound spectroscopy and multiple-scattering theory inclusive of shear-wave effects, Chemical Engineering Research and Design, 114 (2016) 69-78.
[52] I. 20998-1:2006, Measurement and characterization of particles by acoustic methods -- Part 1: Concepts and procedures in ultrasonic attenuation spectroscopy, 2006.
[53] A. Shukla, A. Prakash, S. Rohani, Online measurement of particle size distribution during crystallization using ultrasonic spectroscopy, Chemical Engineering Science, 65 (2010) 3072-3079.
[54] A. Pankewitz, H. Geers, In-line Crystal Size Distribution analysis in industrial crystallisation by Ultrasonic Extinction, Sympatec GmbH, System-Partikel-Technik, Clausthal-Zellerfeld, Germany.
[55] G. Bonacucina, D.R. Perinelli, M. Cespi, L. Casettari, R. Cossi, P. Blasi, G.F. Palmieri, Acoustic spectroscopy: A powerful analytical method for the pharmaceutical field?, International Journal of Pharmaceutics, 503 (2016) 174-195.
[56] R. Steponavičius, S.N. Thennadil, Extraction of Chemical Information of Suspensions Using Radiative Transfer Theory To Remove Multiple Scattering Effects: Application to a Model Multicomponent System, Analytical Chemistry, 83 (2011) 1931-1937.
[57] X.Y. Woo, Z.K. Nagy, R.B.H. Tan, R.D. Braatz, Adaptive Concentration Control of Cooling and Antisolvent Crystallization with Laser Backscattering Measurement, Crystal Growth & Design, 9 (2009) 182-191.
[58] M. Barrett, M. McNamara, H. Hao, P. Barrett, B. Glennon, Supersaturation tracking for the development, optimization and control of crystallization processes, Chemical Engineering Research and Design, 88 (2010) 1108-1119.
[59] Y. Hu, J.K. Liang, A.S. Myerson, L.S. Taylor, Crystallization Monitoring by Raman Spectroscopy: Simultaneous Measurement of Desupersaturation Profile and Polymorphic Form in Flufenamic Acid Systems, Industrial & Engineering Chemistry Research, 44 (2005) 1233-1240.
[60] A.N. Saleemi, G. Steele, N.I. Pedge, A. Freeman, Z.K. Nagy, Enhancing crystalline properties of a cardiovascular active pharmaceutical ingredient using a

process analytical technology based crystallization feedback control strategy, International Journal of Pharmaceutics, 430 (2012) 56-64.
[61] S. Baset, H. Akbari, H. Zeynali, M. Shafie, Size measurement of metal and semiconductor nanoparticles via uv-vis absorption spectra, Digest Journal of Nanomaterials and Biostructures 6(2011) 709-716.
[62] M.-T. Celis, A. Forgiarini, M.-I. Briceño, L.H. García-Rubio, Spectroscopy measurements for determination of polymer particle size distribution, Colloids and Surfaces A: Physicochemical and Engineering Aspects, 331 (2008) 91-96.
[63] M.C. Pasikatan, J.L. Steele, C.K. Spillman, E. Haque, Near Infrared Reflectance Spectroscopy for Online Particle Size Analysis of Powders and Ground Materials, Journal of Near Infrared Spectroscopy, 9 (2001) 153-164.
[64] N. Ghosh, S.K. Mohanty, S.K. Majumder, P.K. Gupta, Measurement of optical transport properties of normal and malignant human breast tissue, Applied Optics, 40 (2001) 176-184.
[65] D. Arifler, C. MacAulay, M. Follen, R. Richards-Kortum, Spatially resolved reflectance spectroscopy for diagnosis of cervical precancer: Monte Carlo modeling and comparison to clinical measurements, Biomedo, 11 (2006) 064027.
[66] P. Hjalmarsson, S.N. Thennadil, Determination of glucose concentration in tissue-like material using spatially resolved steady-state diffuse reflectance spectroscopy, 2008, pp. 685508-685508-685509.
[67] R.M. Doornbos, R. Lang, M.C. Aalders, F.W. Cross, H.J. Sterenborg, The determination of in vivo human tissue optical properties and absolute chromophore concentrations using spatially resolved steady-state diffuse reflectance spectroscopy, Physics in Medicine and Biology, 44 (1999) 967.
[68] T.J. Farrel, M.S. Patterson, A diffusion theory model of spatially resolved, steady-state diffuse reflectance for noninvasive determination of tissue optical properties in vivo., Medical Physics, 1992, pp. 879.
[69] M. Born, E. Wolf, A.B. Bhatia, D. Gabor, A.R. Stokes, A.M. Taylor, P.A. Wayman, W.L. Wilcock, Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light, Cambridge University Press2000.

[70] D.J. Burgess, E. Duffy, F. Etzler, A.J. Hickey, Particle size analysis: AAPS workshop report, cosponsored by the Food and Drug Administration and the United States Pharmacopeia, The AAPS Journal, 6 (2004) 23-34.
[71] K. Sommer, 40 Years of Presentation Particle Size Distributions–Yet Still Incorrect?, Part. Part. Syst. Charact., 18 (2001) 22–25.
[72] A. Abbas, D. Nobbs, J.A. Romagnoli, Investigation of on-line optical particle characterization in reaction and cooling crystallization systems. Current state of the art, Measurement Science and Technology, 13 (2002) 349-356.
[73] O.S. Agimelen, V. Svoboda, B. Ahmed, J. Cardona, J. Dziewierz, C.J. Brown, T. McGlone, A. Cleary, C. Tachtatzis, C. Michie, A.J. Florence, I. Andonovic, A.J. Mulholland, J. Sefcik, Multi-sensor inline measurements of crystal size and shape distributions during high shear wet milling of crystal slurries, Advanced Powder Technology, 29 (2018) 2987-2995.
[74] E. Skibsted, S.B. Engelsen, Spectroscopy for Process Analytical Technology (PAT), in: J.C. Lindon (Ed.) Encyclopedia of Spectroscopy and Spectrometry (Second Edition), Academic Press, Oxford, 2010, pp. 2651-2661.
[75] F. van den Berg, C.B. Lyndgaard, K.M. Sørensen, S.B. Engelsen, Process Analytical Technology in the food industry, Trends in Food Science & Technology, 31 (2013) 27-35.
[76] M. Alcalà, M. Blanco, M. Bautista, J.M. González, On-line monitoring of a granulation process by NIR spectroscopy, Journal of Pharmaceutical Sciences, 99 (2010) 336-345.
[77] J. Rantanen, H. Wikström, R. Turner, L.S. Taylor, Use of In-Line Near-Infrared Spectroscopy in Combination with Chemometrics for Improved Understanding of Pharmaceutical Processes, Analytical Chemistry, 77 (2005) 556-563.
[78] J.P. Nielsen, D. Bertrand, E. Micklander, P. Courcoux, L. Munck, Study of NIR Spectra, Particle Size Distributions and Chemical Parameters of Wheat Flours: A Multi-Way Approach, Journal of Near Infrared Spectroscopy, 9 (2001) 275-285.
[79] W.B. Lee, E. Widjaja, P.W.S. Heng, L.W. Chan, Near infrared spectroscopy for rapid and in-line detection of particle size distribution variability in lactose during mixing, International Journal of Pharmaceutics, 566 (2019) 454-462.

[80] M.C. Sarraguça, A.V. Cruz, H.R. Amaral, P.C. Costa, J.A. Lopes, Comparison of different chemometric and analytical methods for the prediction of particle size distribution in pharmaceutical powders, Analytical and Bioanalytical Chemistry, 399 (2011) 2137-2147.
[81] V. Pauli, Y. Roggo, P. Kleinebudde, M. Krumme, Real-time monitoring of particle size distribution in a continuous granulation and drying process by near infrared spectroscopy, European Journal of Pharmaceutics and Biopharmaceutics, 141 (2019) 90-99.
[82] A.F. Santos, E.L. Lima, J.C. Pinto, In-line evaluation of average particle size in styrene suspension polymerizations using near-infrared spectroscopy, Journal of Applied Polymer Science, 70 (1998) 1737-1745.
[83] D.A. Whitaker, F. Buchanan, M. Billham, M. McAfee, A UV-Vis spectroscopic method for monitoring of additive particle properties during polymer compounding, Polymer Testing, 67 (2018) 392-398.
[84] G. Moreno-Martin, M.E. León-González, Y. Madrid, Simultaneous determination of the size and concentration of AgNPs in water samples by UV–vis spectrophotometry and chemometrics tools, Talanta, 188 (2018) 393-403.
[85] J.G. Emidio, Spatially and angularly resolved diffuse reflectance spectroscopy for in-situ monitoring of suspension polymerisation reactions, Department of Chemical and Process Engineering, University of Strathclyde, 2017.
[86] D.A. Gómez, J. Coello, S. Maspoch, The influence of particle size on the intensity and reproducibility of Raman spectra of compacted samples, Vibrational Spectroscopy, 100 (2019) 48-56.
[87] Y. Hu, H. Wikström, S.R. Byrn, L.S. Taylor, Analysis of the Effect of Particle Size on Polymorphic Quantitation by Raman Spectroscopy, Applied Spectroscopy, 60 (2006) 977-984.
[88] H. Wang, C.K. Mann, T.J. Vickers, Effect of Powder Properties on the Intensity of Raman Scattering by Crystalline Solids, Applied Spectroscopy, 56 (2002) 1538-1544.
[89] M.V. Pellow-Jarman, P.J. Hendra, R.J. Lehnert, The dependence of Raman signal intensity on particle size for crystal powders, Vibrational Spectroscopy, 12 (1996) 257-261.

[90] W.H. Doub, W.P. Adams, J.A. Spencer, L.F. Buhse, M.P. Nelson, P.J. Treado, Raman Chemical Imaging for Ingredient-specific Particle Size Characterization of Aqueous Suspension Nasal Spray Formulations: A Progress Report, Pharmaceutical Research, 24 (2007) 934-945.
[91] D. Iacobucci, M.J. Schneider, D.L. Popovich, G.A. Bakamitsos, Mean centering helps alleviate “micro” but not “macro” multicollinearity, Behavior Research Methods, 48 (2016) 1308-1317.
[92] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Analytical Chemistry, 36 (1964) 1627-1639.
[93] P.A. Gorry, General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method, Analytical Chemistry, 62 (1990) 570-573.
[94] S.R. Delwiche, J.B. Reeves, A Graphical Method to Evaluate Spectral Preprocessing in Multivariate Regression Calibrations: Example with Savitzky—Golay Filters and Partial Least Squares Regression, Applied Spectroscopy, 64 (2010) 73-82.
[95] T. Naes, T. Isaksson, T. Fearn, T. Davies, A User-friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester UK, 2002.
[96] C. Pizarro, I. Esteban-Díez, A.-J. Nistal, J.-M.a. González-Sáiz, Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy, Analytica Chimica Acta, 509 (2004) 217-227.
[97] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Applied Spectroscopy, 43 (1989) 772-777.
[98] T. Isaksson, T. Naes, Effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy, Applied Spectroscopy, 42 (1988) 1273-1284.
[99] P. Geladi, D. MacDougall, H. Martens, Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat, Applied Spectroscopy, 39 (1985) 491-500.
[100] H. Martens, E. Stark, Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy, Journal of Pharmaceutical and Biomedical Analysis, 9 (1991) 625-635.

[101] Y.-C. Chen, S.N. Thennadil, Insights into information contained in multiplicative scatter correction parameters and the potential for estimating particle size from these parameters, Analytica Chimica Acta, 746 (2012) 37-46.
[102] S.N. Thennadil, H. Martens, A. Kohler, Physics-Based Multiplicative Scatter Correction Approaches for Improving the Performance of Calibration Models, Applied Spectroscopy, 60 (2006) 315-321.
[103] P. Bassan, A. Kohler, H. Martens, J. Lee, H.J. Byrne, P. Dumas, E. Gazi, M. Brown, N. Clarke, P. Gardner, Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples, Analyst, 135 (2010) 268-277.
[104] H. Martens, J.P. Nielsen, S.B. Engelsen, Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures, Analytical Chemistry, 75 (2003) 394-404.
[105] S.N. Thennadil, E.B. Martin, Empirical preprocessing methods and their impact on NIR calibrations: a simulation study, Journal of Chemometrics, 19 (2005) 77--89.
[106] Å. Rinnan, F.v.d. Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC Trends in Analytical Chemistry, 28 (2009) 1201-1222.
[107] I.S. Helland, T. Næs, T. Isaksson, Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data, Chemometrics and Intelligent Laboratory Systems, 29 (1995) 233-241.
[108] J. Camacho, J. Picó, A. Ferrer, Data understanding with PCA: Structural and Variance Information plots, Chemometrics and Intelligent Laboratory Systems, 100 (2010) 48-56.
[109] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics and Intelligent Laboratory Systems, 2 (1987) 37-52.
[110] B.M. Wise, N.B. Gallagher, The process chemometrics approach to process monitoring and fault detection, Journal of Process Control, 6 (1996) 329-348.
[111] N. Kumar, A. Bansal, G.S. Sarma, R.K. Rawal, Chemometrics tools used in analytical chemistry: An overview, Talanta, 123 (2014) 186-199.
[112] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems, 58 (2001) 109-130.

[113] M. Blanco, A. Peguero, An expeditious method for determining particle size distribution by near infrared spectroscopy: Comparison of PLS2 and ANN models, Talanta, 77 (2008) 647-651.
[114] A.J. O'Neil, R.D. Jee, A.C. Moffat, Measurement of the percentage volume particle size distribution of powdered microcrystalline cellulose using reflectance near-infrared spectroscopy, Analyst, 128 (2003) 1326-1330.
[115] J.A. Westerhuis, P.M.J. Coenegracht, Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares, Journal of Chemometrics, 11 (1997) 379-392.
[116] J.A. Lopes, J.C. Menezes, J.A. Westerhuis, A.K. Smilde, Multiblock PLS analysis of an industrial pharmaceutical process, Biotechnology and Bioengineering, 80 (2002) 419--427.
[117] M. Jing, W. Cai, X. Shao, Quantitative Determination of the Components in Corn and Tobacco Samples by Using Near-Infrared Spectroscopy and Multiblock Partial Least Squares, Analytical Letters, 43 (2010) 1910-1921.
[118] V. Gaydou, J. Kister, N. Dupuy, Evaluation of multiblock NIR/MIR PLS predictive models to detect adulteration of diesel/biodiesel blends by vegetal oil, Chemometrics and Intelligent Laboratory Systems, 106 (2011) 190-197.
[119] L.P. Brás, S.A. Bernardino, J.A. Lopes, J.C. Menezes, Multiblock PLS as an approach to compare and combine NIR and MIR spectra in calibrations of soybean flour, Chemometrics and Intelligent Laboratory Systems, 75 (2005) 91-99.
[120] G. Kos, H. Lohninger, R. Krska, Validation of chemometric models for the determination of deoxynivalenol on maize by mid-infrared spectroscopy, Mycotoxin Research, 19 (2003) 149-153.
[121] F. Westad, F. Marini, Validation of chemometric models – A tutorial, Analytica Chimica Acta, 893 (2015) 14-24.
[122] J.E. Gledson, Spatially and angularly resolved diffuse reflectance spectroscopy for in-situ monitoring of suspension polymerisation reactions, Chemical and Process Engineering Department, University of Strathclyde, Glasgow, UK, 2017.
[123] N. Sultanova, S. Kasarova, I. Nikolov, Dispersion properties of optical polymers, *Acta Physica Polonica A*2009.
[124] Lactose - Some basic properties and characteristics, DFE pharma

[125] CAS No: 5989-81-1 Alpha-D-Lactose monohydrate.
[126] D.R. Lide, CRC Handbook of Chemistry and Physics, 84th Edition, Taylor & Francis2003.
[127] J. Rheims, J. Köser, T. Wriedt, Refractive-index measurements in the near-IR using an Abbe refractometer, Measurement Science and Technology, 8 (1997) 601-605.
[128] J. Cardona, C. Tachtatzis, ImagingApp: image analysis framework for particle size and shape characterisation, 2018.
[129] O.S. Agimelen, A. Jawor-Baczynska, J. McGinty, J. Dziewierz, C. Tachtatzis, A. Cleary, I. Haley, C. Michie, I. Andonovic, J. Sefcik, A.J. Mulholland, Integration of in situ imaging and chord length distribution measurements for estimation of particle size and shape, Chemical Engineering Science, 144 (2016) 87-100.
[130] O.S. Agimelen, A.J. Mulholland, J. Sefcik, Software for transforming measured chord length distribution data to particle size distribution, 2018.
[131] R. Hennessy, W. Goth, M. Sharma, M.K. Markey, J.W. Tunnell, Effect of probe geometry and optical properties on the sampling depth for diffuse reflectance spectroscopy, BIOMEDO, 19 (2014) 107002-107002.
[132] H. Arimoto, M. Egawa, Y. Yamada, Depth profile of diffuse reflectance near-infrared spectroscopy for measurement of water content in skin, Skin Research and Technology, 11 (2005) 27-35.
[133] R. Watté, B. Aernouts, R. Van Beers, A. Postelmans, W. Saeys, Computational optimization of the configuration of a spatially resolved spectroscopy sensor for milk analysis, Analytica Chimica Acta, 917 (2016) 53-63.
[134] Y.-C. Chen, D. Foo, N. Dehanov, S.N. Thennadil, Spatially and angularly resolved spectroscopy for in-situ estimation of concentration and particle size in colloidal suspensions, Analytical and Bioanalytical Chemistry, 409 (2017) 6975-6988.
[135] R. Steponavičius, S.N. Tennadil, Full Correction of Scattering Effects by Using the Radiative Transfer Theory for Improved Quantitative Analysis of Absorbing Species in Suspensions, Applied Spectroscopy, 67 (2013) 526-535.
[136] R. Steponavičius, S.N. Tennadil, Extraction of Chemical Information of Suspensions Using Radiative Transfer Theory to Remove Multiple Scattering Effects:

Application to a Model Two-Component System, Analytical Chemistry, 81 (2009) 7713-7723.
[137] K. Vay, W. Frieß, S. Scheler, Understanding reflection behavior as a key for interpreting complex signals in FBRM monitoring of microparticle preparation processes, International Journal of Pharmaceutics, 437 (2012) 1-10.
[138] R.G. Sparks, C.L. Dobbs, The Use of Laser Backscatter Instrumentation for the on-line measurement of the particle size distribution of emulsions, Particle & Particle Systems Characterization, 10 (1993) 279-289.
[139] N. Kail, H. Briesen, W. Marquardt, Advanced Geometrical Modeling of Focused Beam Reflectance Measurements (FBRM), Particle & Particle Systems Characterization, 24 (2007) 184-192.
[140] A. Senaputra, F. Jones, P.D. Fawell, P.G. Smith, Focused beam reflectance measurement for monitoring the extent and efficiency of flocculation in mineral systems, AIChE Journal, 60 (2014) 251-265.
[141] W. Haiss, N.T.K. Thanh, J. Aveyard, D.G. Fernig, Determination of Size and Concentration of Gold Nanoparticles from UV−Vis Spectra, Analytical Chemistry, 79 (2007) 4215-4221.
[142] B. Aernouts, R. Watté, R. Van Beers, F. Delport, M. Merchiers, J. De Block, J. Lammertyn, W. Saeys, Flexible tool for simulating the bulk optical properties of polydisperse spherical particles in an absorbing host: experimental validation, Opt. Express, 22 (2014) 20223-20238.
[143] L.K.H. Bittner, N. Heigl, C.H. Petter, M.F. Noisternig, U.J. Griesser, G.K. Bonn, C.W. Huck, Near-infrared reflection spectroscopy (NIRS) as a successful tool for simultaneous identification and particle size determination of amoxicillin trihydrate, Journal of Pharmaceutical and Biomedical Analysis, 54 (2011) 1059-1064.
[144] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Analytica Chimica Acta, 185 (1986) 1-17.
[145] M.B. Seasholtz, B.R. Kowalski, The effect of mean centering on prediction in multivariate calibration, Journal of Chemometrics, 6 (1992) 103-111.
[146] J.W. Einax, Paul Gemperline (Ed.): Practical guide to chemometrics, 2nd Ed, Analytical and Bioanalytical Chemistry, 388 (2007) 511-512.

[147] A.B.N. Brito, M. Giulietti, Study of lactose crystallization in water-acetone solutions, Crystal Research and Technology, 42 (2007) 583-588.
[148] X.M. Zeng, G.P. Martin, C. Marriott, J. Pritchard, The Influence of Crystallization Conditions on the Morphology of Lactose Intended for Use as a Carrier for Dry Powder Aerosols, Journal of Pharmacy and Pharmacology, 52 (2000) 633-643.
[149] C.S. Omar, R.M. Dhenge, J.D. Osborne, T.O. Althaus, S. Palzer, M.J. Hounslow, A.D. Salman, Roller compaction: Effect of morphology and amorphous content of lactose powder on product quality, International Journal of Pharmaceutics, 496 (2015) 63-74.
[150] H. Larhrib, G.P. Martin, D. Prime, C. Marriott, Characterisation and deposition studies of engineered lactose crystals with potential for use as a carrier for aerosolised salbutamol sulfate from dry powder inhalers, European Journal of Pharmaceutical Sciences, 19 (2003) 211-221.
[151] J. Chen, J. Wang, R. Li, A. Lu, Y. Li, Thermal and X-ray Diffraction Analysis of Lactose Polymorph, Procedia Engineering, 102 (2015) 372-378.
[152] S. Garnier, S. Petit, G. Coquerel, Dehydration Mechanism and Crystallisation Behaviour of Lactose, Journal of Thermal Analysis and Calorimetry, 68 (2002) 489-502.
[153] S.G. Agrawal, A. Balandier, A.H.J. Paterson, J.R. Jones, Study on lactose attrition inside the mixing cell of a laser diffraction particle sizer using a novel attrition index, Powder Technology, 208 (2011) 669-675.
[154] G.A. Hebbink, B.H.J. Dickhoff, Chapter 5 - Application of lactose in the pharmaceutical industry, in: M. Paques, C. Lindner (Eds.) Lactose, Academic Press2019, pp. 175-229.

Appendix A

A1. Sample conditions used in the lactose calibration and test datasets

Table A.1: D50 and concentration of the samples used in the lactose calibration and test datasets.

Calibration dataset		Test dataset	
D50 (µm)	**Concentration (wt.%)**	**D50 (µm)**	**Concentration (wt.%)**
20.35	0.50	20.35	5.01
20.35	2.51	108.00	8.52
20.35	3.50	292.67	11.52
20.35	6.50	108.00	15.01
20.35	8.50	240.00	25.08
20.35	10.01	20.35	7.50
57.32	1.00	292.67	2.52
57.32	2.49	198.67	8.50
57.32	5.00	108.00	20.01
57.32	7.50	158.00	5.00
57.32	8.51	340.00	12.53
57.32	10.02	158.00	20.03
57.32	12.53	383.33	20.03
57.32	15.03	158.00	11.50
108.00	1.01	240.00	15.06
108.00	3.51	198.67	16.51
108.00	5.01	57.32	11.53
108.00	7.52	57.32	6.50
108.00	10.02	57.32	13.53
108.00	11.51	383.33	2.51
108.00	12.51	240.00	10.04
108.00	13.51	57.32	3.48
108.00	16.51	57.32	0.50
108.00	17.51	340.00	7.52
108.00	18.52	240.00	13.57
158.00	1.00	20.35	1.01
158.00	2.50	198.67	18.52
158.00	6.01	292.67	10.02
158.00	8.01	198.67	1.00
158.00	10.01	240.00	16.57

158.00	12.51	158.00	17.53
158.00	13.51	383.33	18.02
158.00	15.02	108.00	24.99
158.00	16.53	383.33	10.01
158.00	18.53	383.33	5.01
158.00	25.04	383.33	12.53
198.67	2.50	158.00	7.01
198.67	4.99	340.00	17.54
198.67	7.48	292.67	20.00
198.67	10.01	108.00	2.51
198.67	11.50		
198.67	12.50		
198.67	13.50		
198.67	14.99		
198.67	17.51		
198.67	20.03		
198.67	25.01		
240.00	1.02		
240.00	2.52		
240.00	5.03		
240.00	7.54		
240.00	11.56		
240.00	12.57		
240.00	17.58		
240.00	18.57		
240.00	20.07		
292.67	5.02		
292.67	7.51		
292.67	12.52		
292.67	13.52		
292.67	15.03		
292.67	17.52		
292.67	25.00		
340.00	2.51		
340.00	5.02		
340.00	10.02		
340.00	11.53		
340.00	13.53		
340.00	15.03		
340.00	16.53		
340.00	18.52		
340.00	20.02		
340.00	25.00		
383.33	7.52		

383.33	11.53	
383.33	13.53	
383.33	15.03	
383.33	16.03	
383.33	17.03	
383.33	19.53	

Appendix B

B1. Effect of particle size in laser diffraction

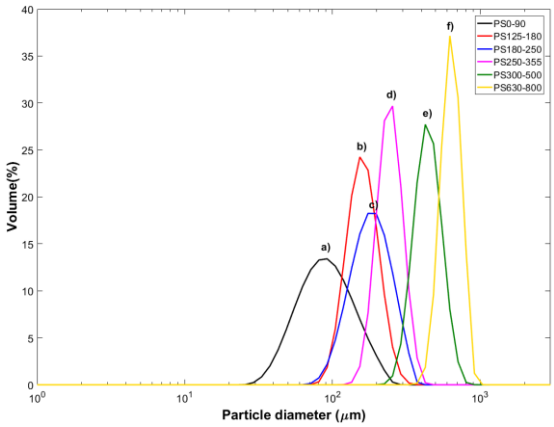


Figure B.1: Volume distribution plots of the PS beads of various size ranges, obtained from LD analysis.

B2. Effect of particle size and solid loading in FBRM

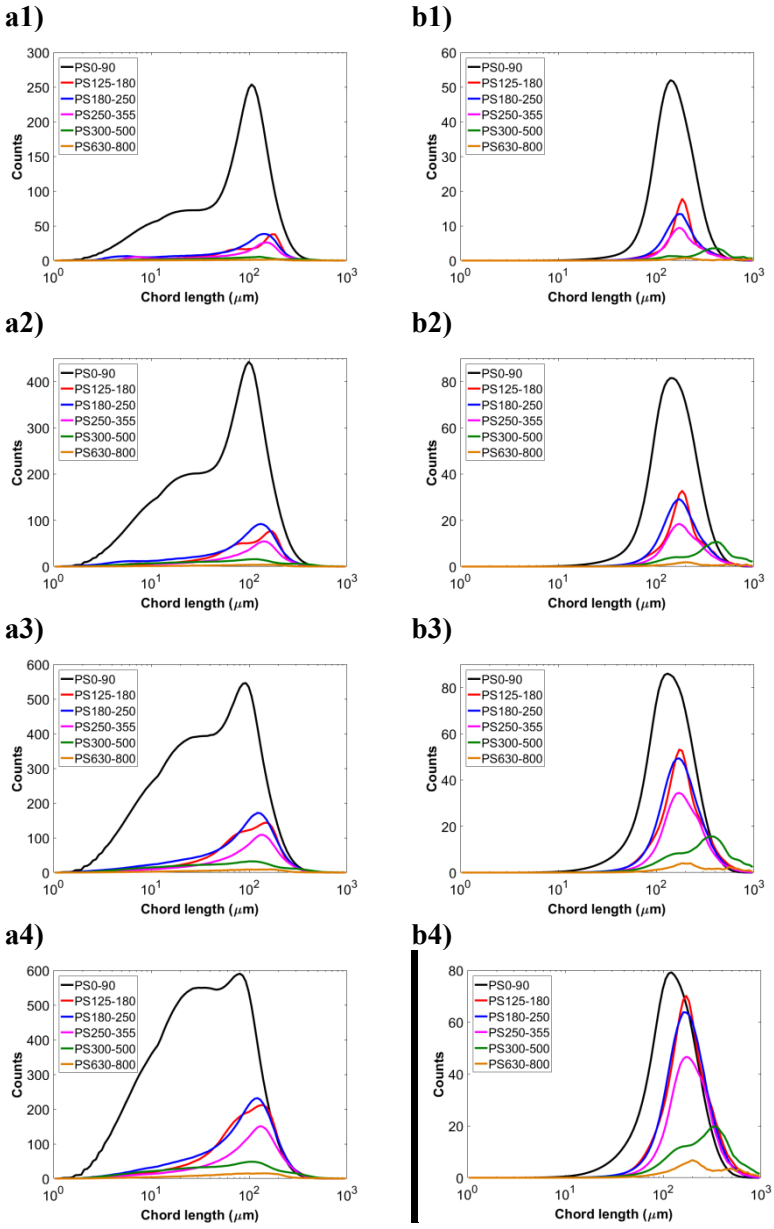


Figure B.2: Effect of particle size on the **a)** unweighted and **b)** square-weighted CLD in the solid loading of **(1)** 1wt.%, **(2)** 2.5 wt.%, **(3)** 5wt.% and **(4)** 7.5 wt.%.

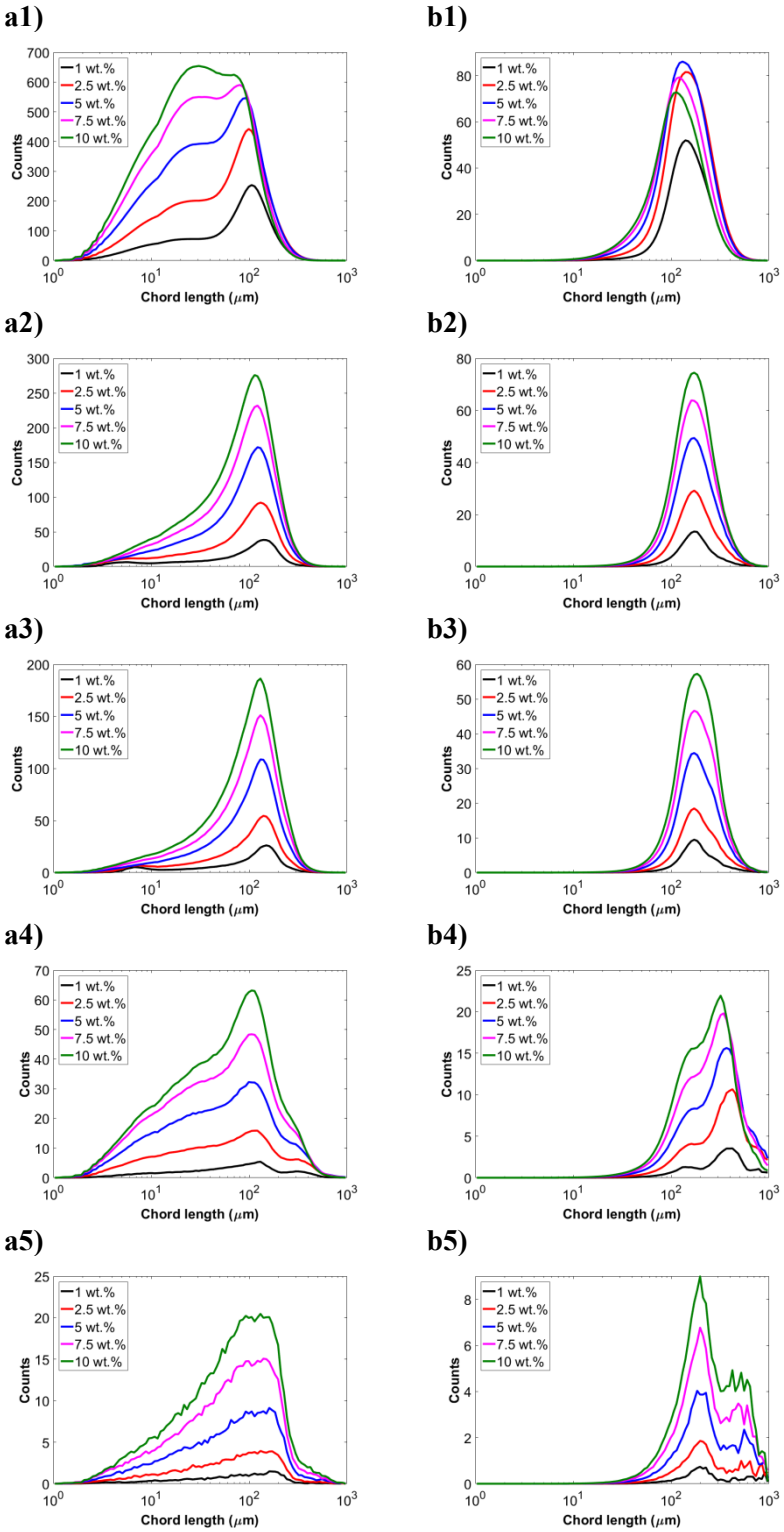


Figure B.3: Effect of solid loading on the **a)** unweighted and **b)** Square-weighted CLD using **(1)** PS0-90, **(2)** PS180-250, **(3)** PS250-355, **(4)** PS300-500 and **(5)** PS630-800. The dashed line corresponds to the most frequent CL in the lowest solid loading.

B3. Effect of solid loading in PVM

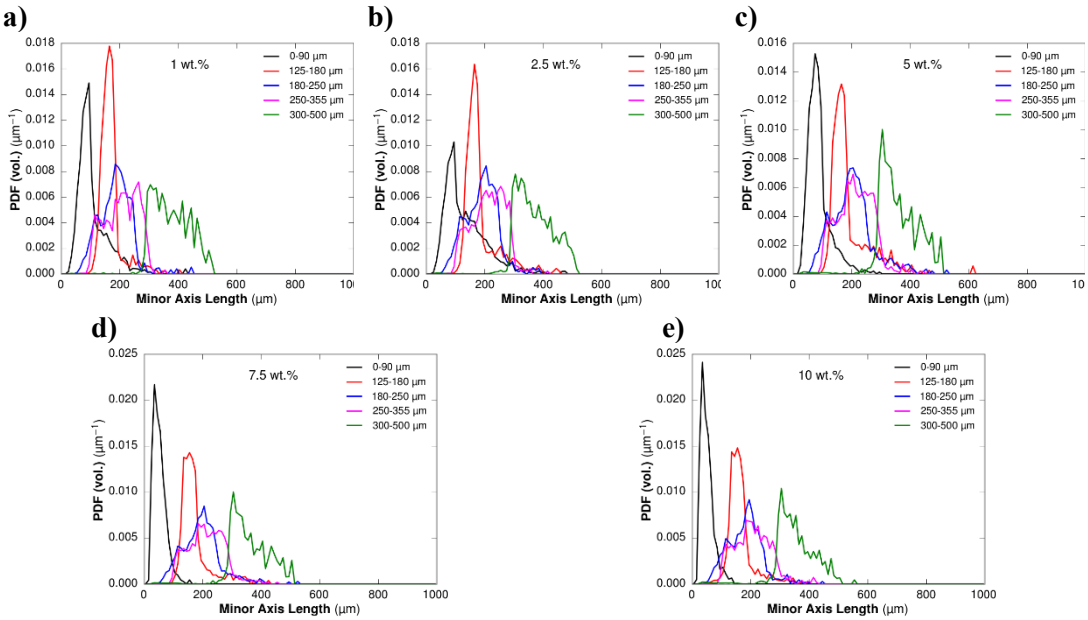


Figure B.4: Effect of particle size on the volume-weighted minor axis length obtained from the PVM image processing algorithm [35] when using **(a)** 1wt.%, **(b)** 2.5 wt.%, **(c)** 5wt.%, **(d)** 7.5 wt.% and **(e)** 10 wt.% of PS samples. The sample PS630-800 is not shown here as it was not possible to capture a representative number of particles for a valid analysis.

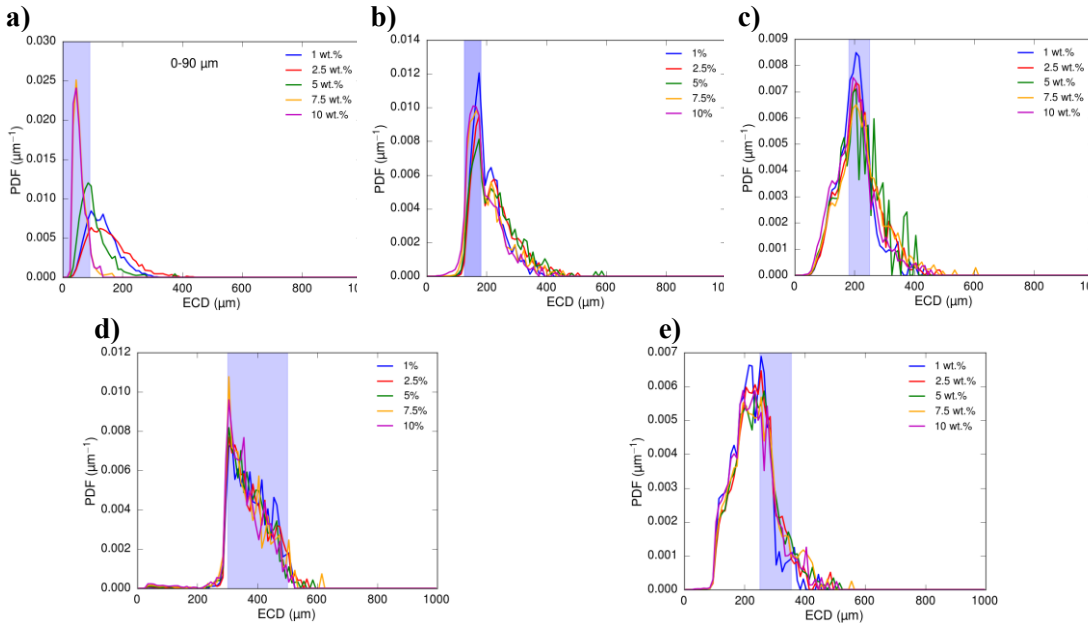


Figure B.5: Effect of solid loading on the volume-weighted ECD, obtained from the PVM image processing algorithm, for the samples **(a)** PS0-90, **(b)** PS125-180, **(c)** PS180-250, **(d)** PS250-355 and **(e)** PS300-500.

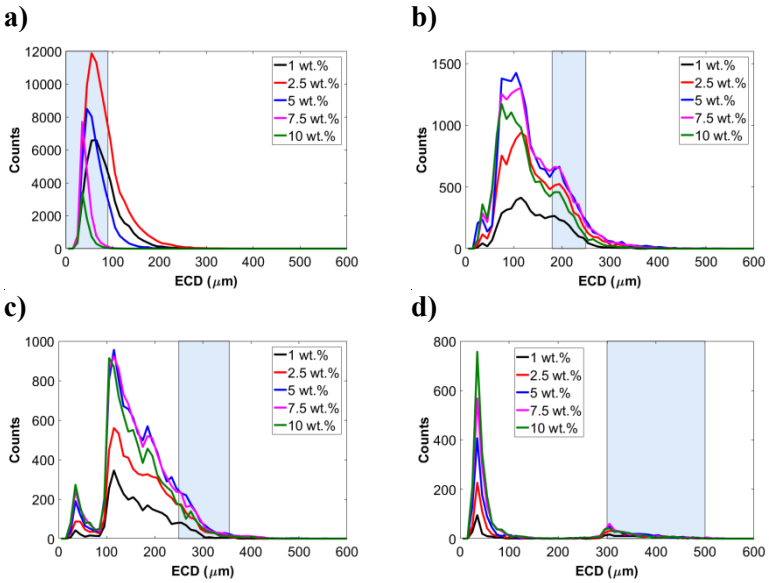


Figure B.6: Effect of solid loading on the number-weighted ECD obtained from the PVM image analysis. **(a)-(e)** corresponds to PS0-90, PS180-250, PS250-355 and PS300-500, respectively. The shaded size range corresponds to those defined by sieving.

B4. Effect of particle size and solid loading on SAR-DRM spectra

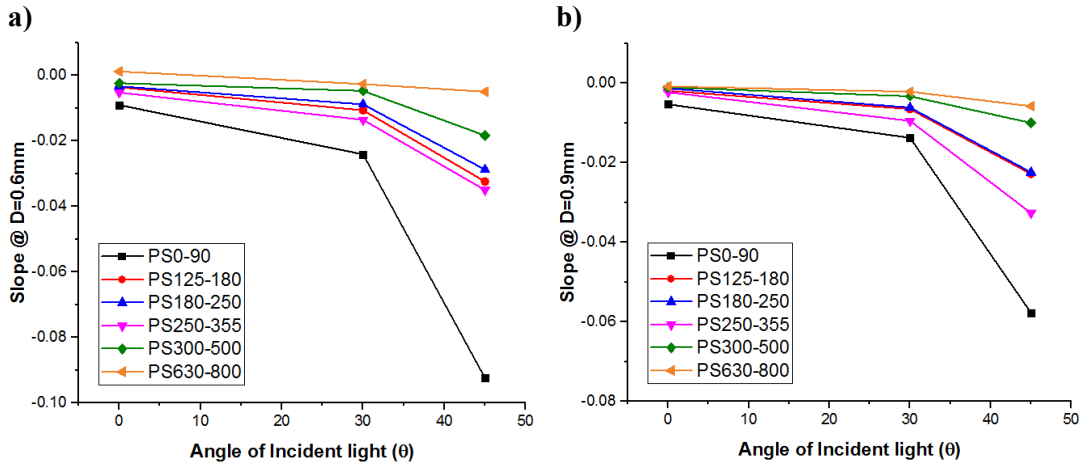


Figure B.7: Changes in the slope of SAR-DRM spectra at 10 wt.% solid loading with angular incident light at 0, 30 and 45°. The spatial arrangement is **a)** 0.6 mm and **b)** 0.9 mm.

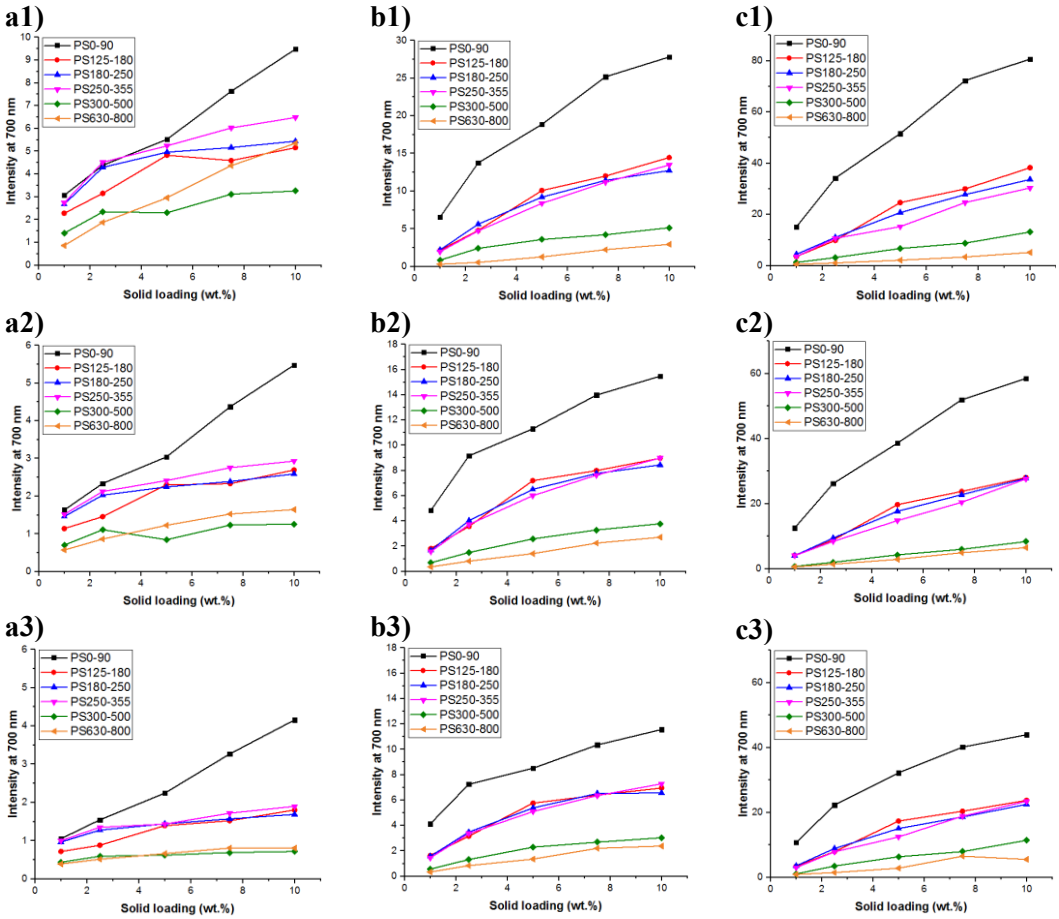


Figure B.8: Changes in the SAR-DRM intensity at the wavelength of 700 nm under **(a)** normal and angular incident light at **(b)** 30 and **(c)** 45°. **(1)-(3)** corresponds to the changes with spatial distances of 0.6, 0.9 and 1.2 mm, respectively.

Appendix C

C1. Effect of particle size in laser diffraction

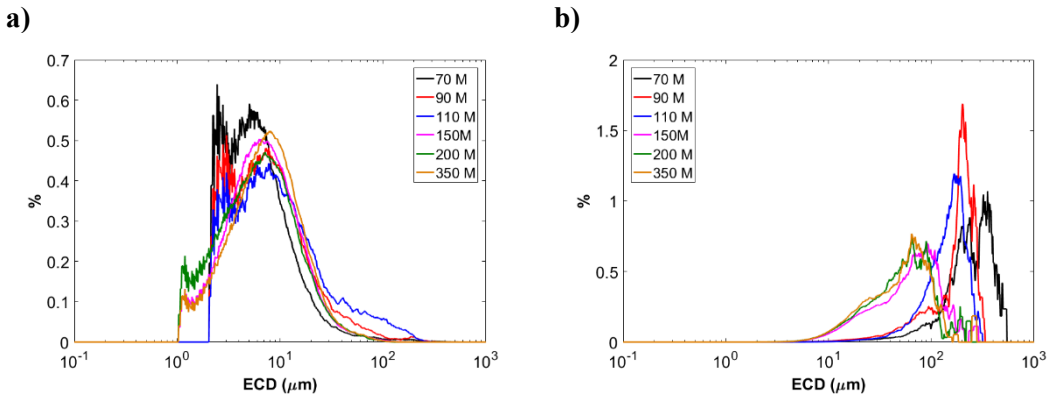
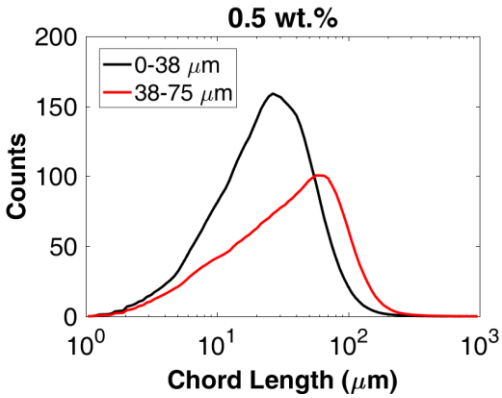
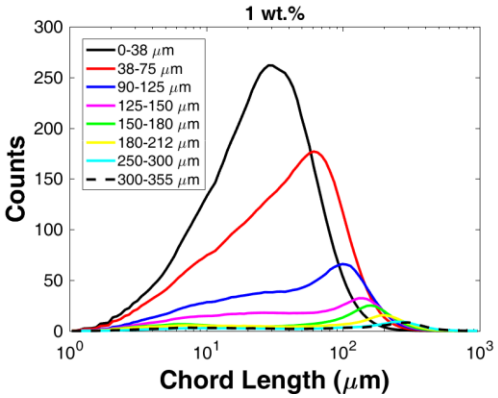
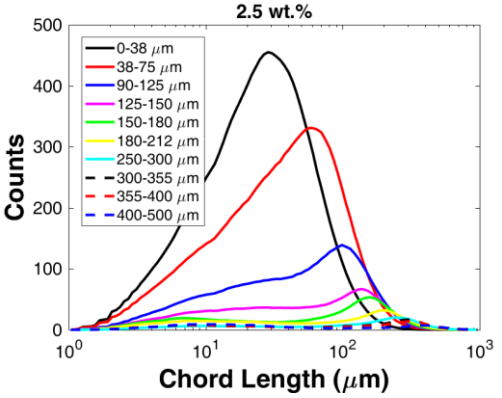


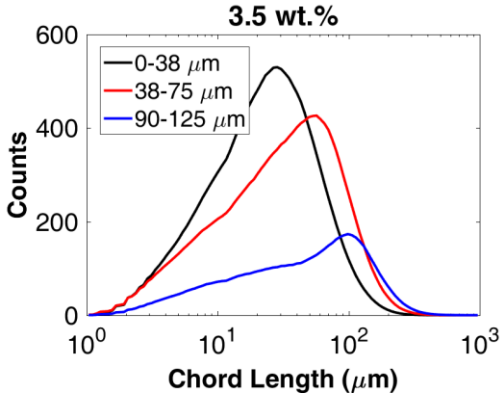
Figure C.1: a) number and **b)** volume based ECD distributions of different lactose grades obtained from Morphologi G3.

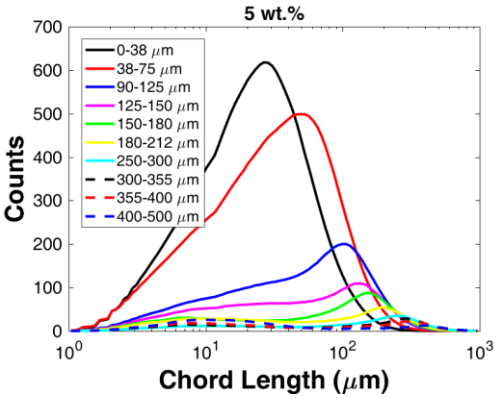
Figure C.1 shows the number and volume based ECD distributions from Morphologi G3, of the different lactose grades ordered from DFE pharma. According to the manufacturer, each grade corresponds to samples with a different D50. These samples were sieved into the ranges described in Chapter 4 Section 4.1. The number based ECD, Figure C.1 (a), shows an overlapping of the ECD distributions and the large number of fines in each one of the grades prior sieving is emphasised. The cut at smaller sizes is related with the microscope limits of optical resolution when using different magnifications. The volume based ECD, Figure C.1(b), emphasises the differences in the samples PSD, with exception for the samples 150 M, 200 M and 350 M.

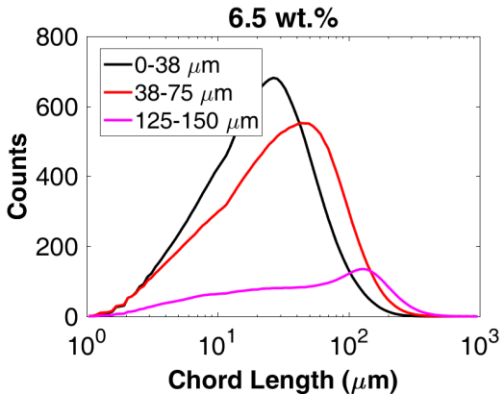

0.5 wt.%
0-38 μm
38-75 μm
Counts
Chord Length (μm)

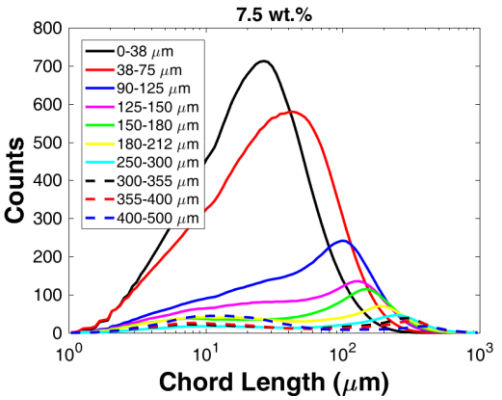

1 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
Counts
Chord Length (μm)

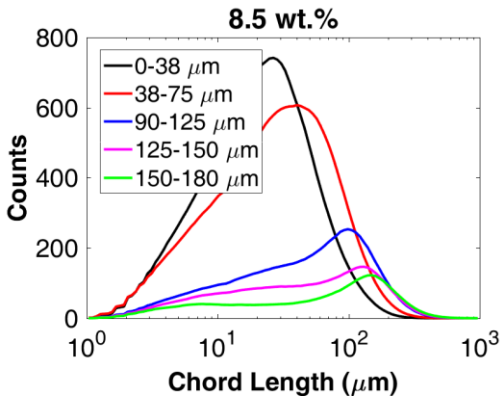

2.5 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

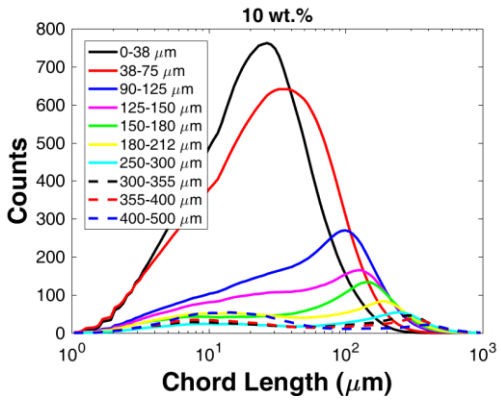

3.5 wt.%
0-38 μm
38-75 μm
90-125 μm
Counts
Chord Length (μm)

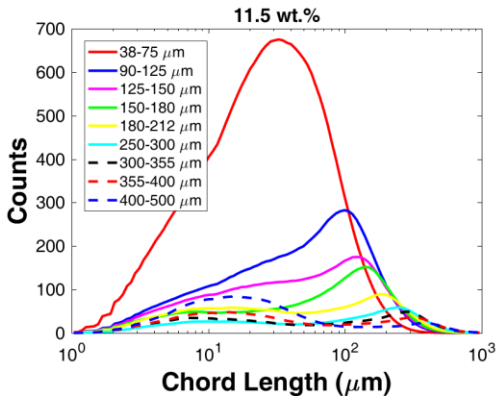

5 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

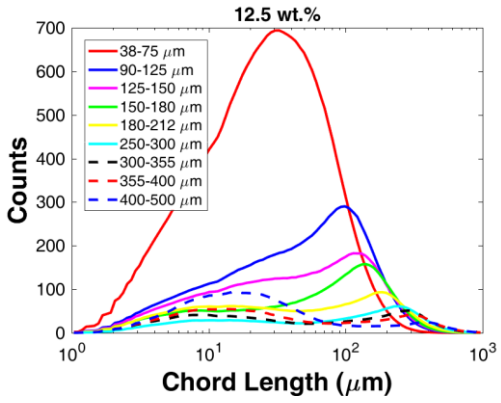

6.5 wt.%
0-38 μm
38-75 μm
125-150 μm
Counts
Chord Length (μm)

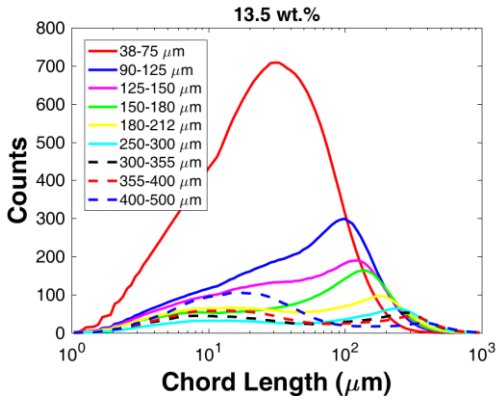

7.5 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

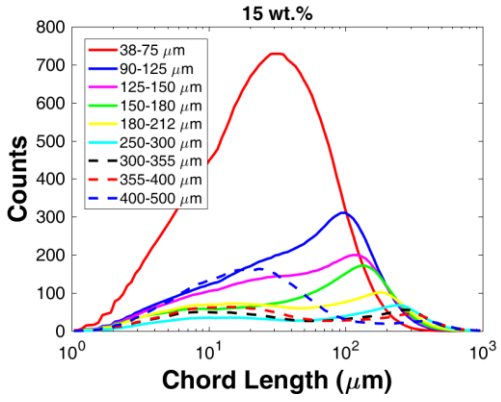

8.5 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
Counts
Chord Length (μm)

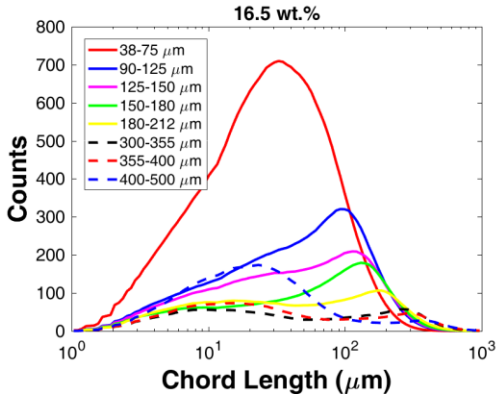

10 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

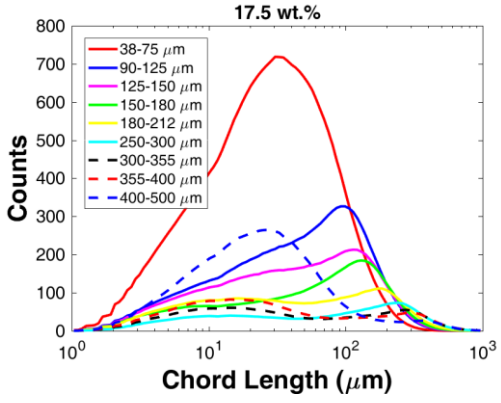

11.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

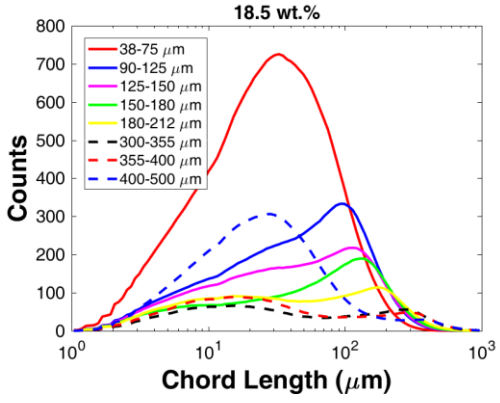

12.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


13.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


15 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


16.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


17.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


18.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

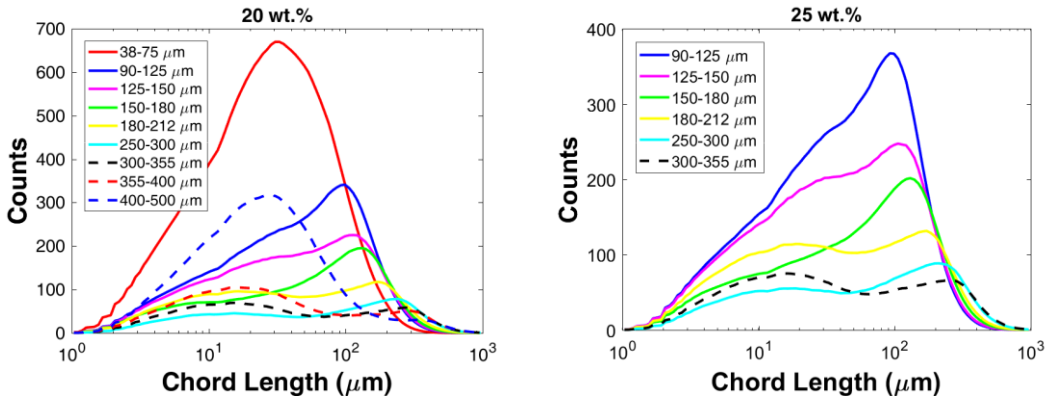
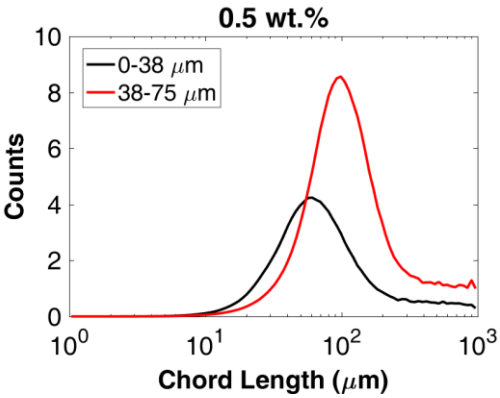
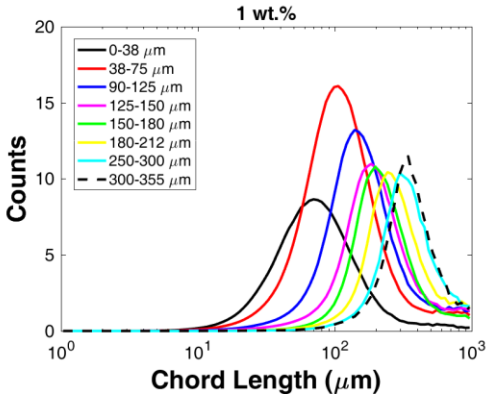
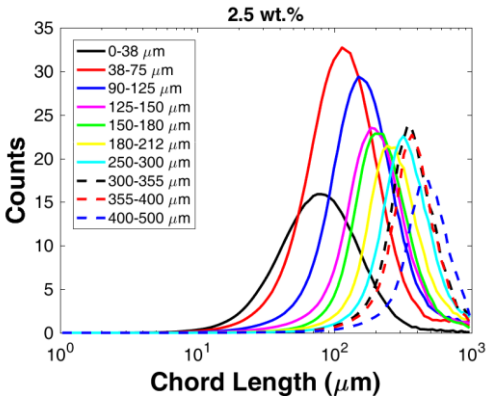
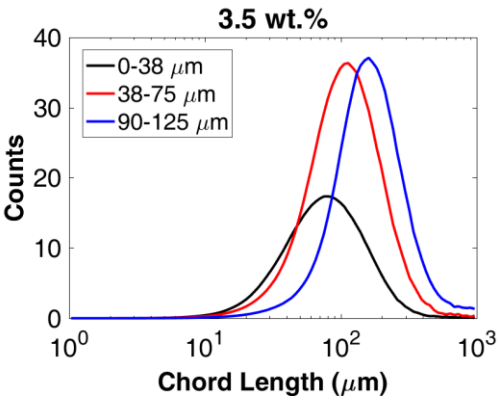


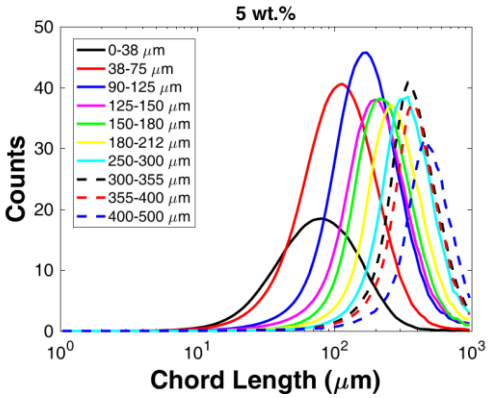
Figure C.2: Effect of particle size on the **u**nweighted CLD for the different solid loadings tested.

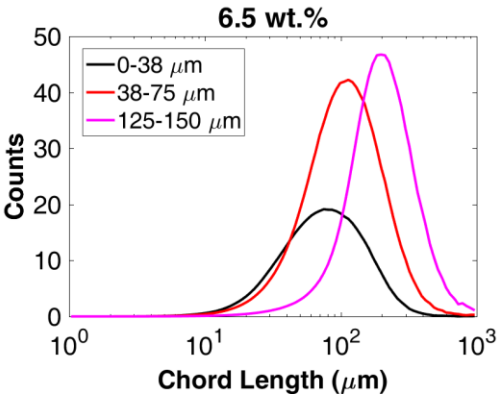

0.5 wt.%
0-38 μm
38-75 μm
Counts
Chord Length (μm)

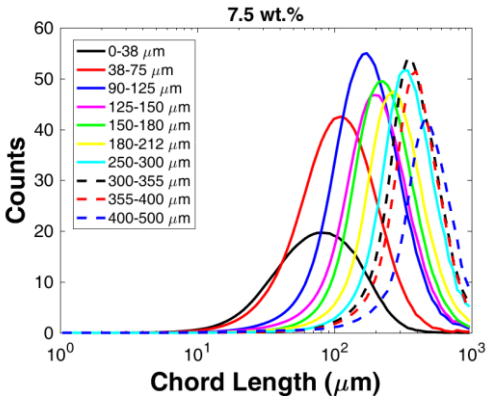

1 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
Counts
Chord Length (μm)

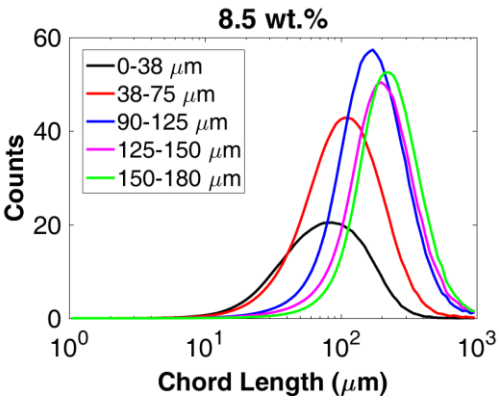

2.5 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

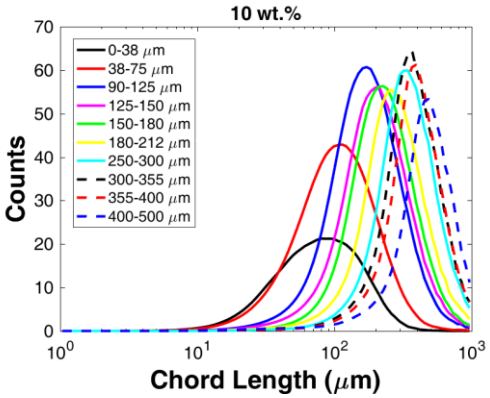

3.5 wt.%
0-38 μm
38-75 μm
90-125 μm
Counts
Chord Length (μm)

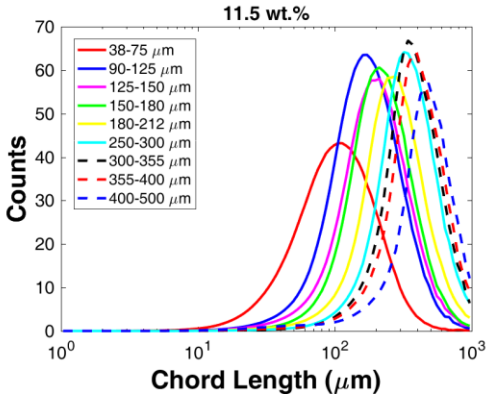

5 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

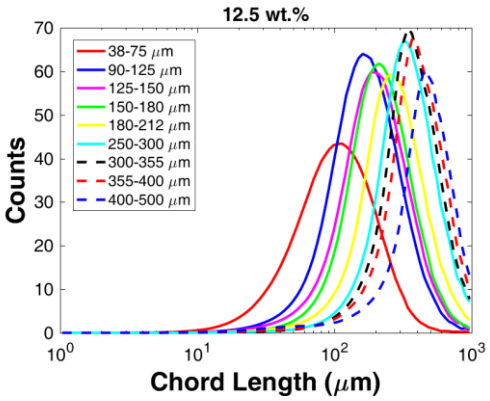

6.5 wt.%
0-38 μm
38-75 μm
125-150 μm
Counts
Chord Length (μm)

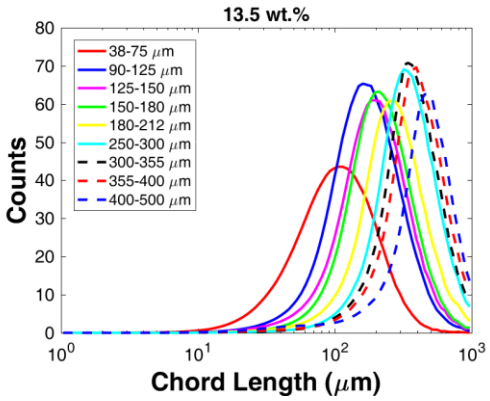

7.5 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

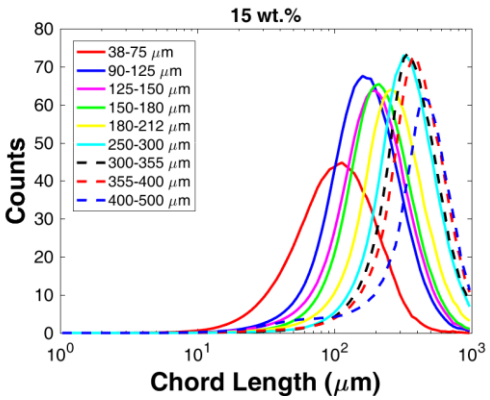

8.5 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
Counts
Chord Length (μm)

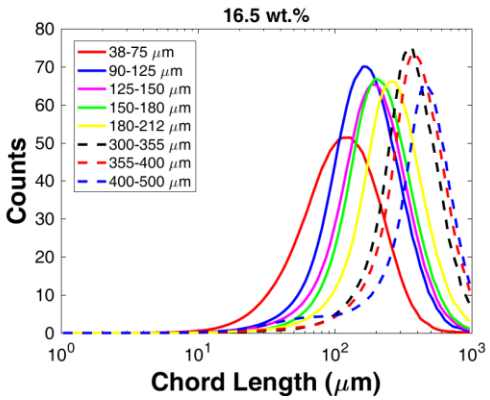

10 wt.%
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

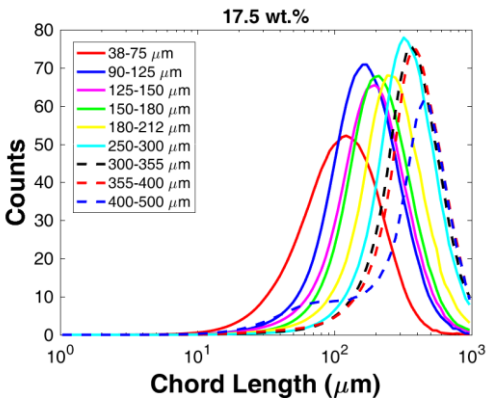

11.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

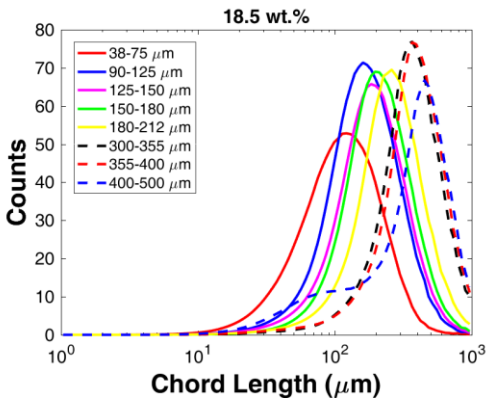

12.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


13.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


15 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


16.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


17.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)


18.5 wt.%
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
300-355 μm
355-400 μm
400-500 μm
Counts
Chord Length (μm)

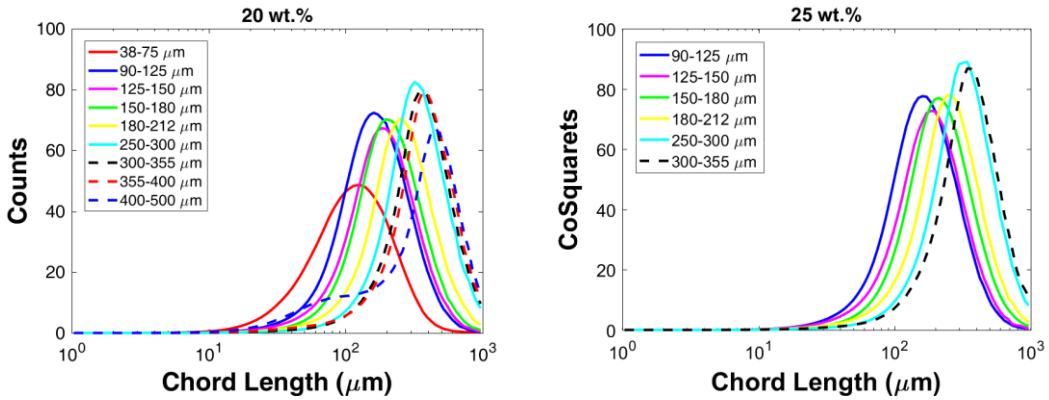
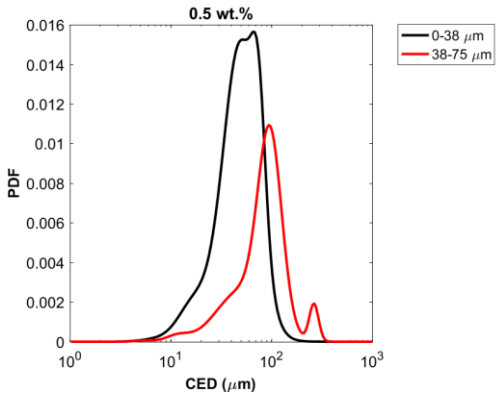
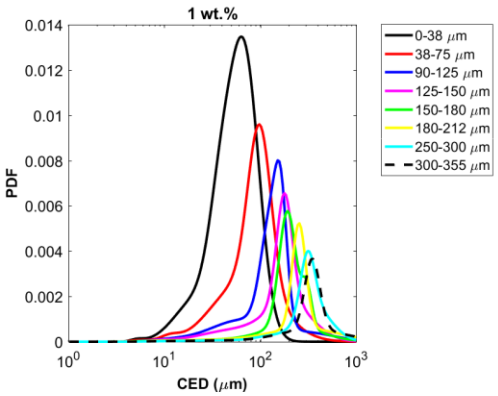
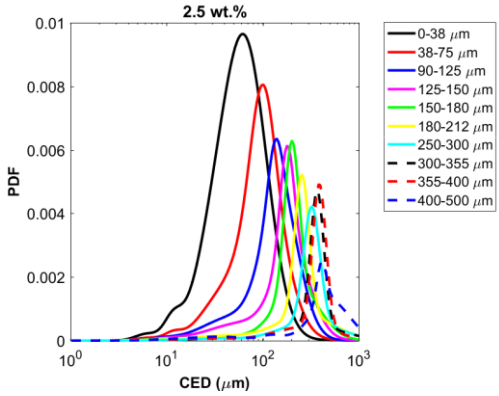
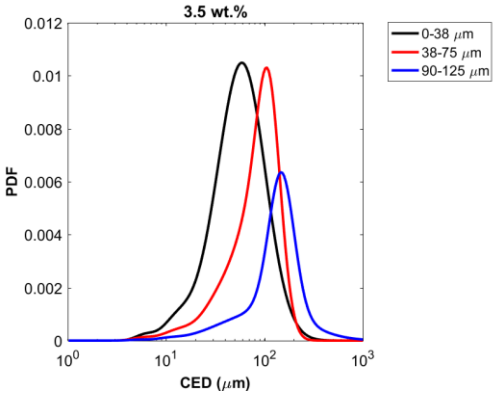


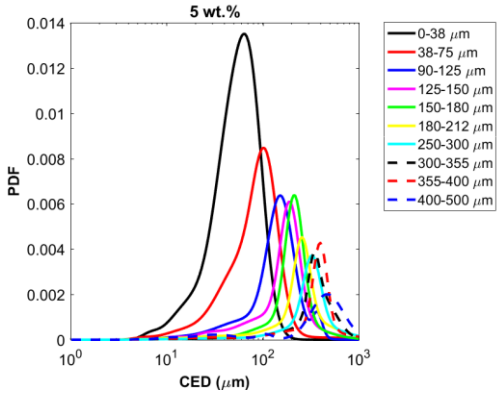
Figure C.3: Effect of particle size on the square weighted CLD for the different solid loadings tested.

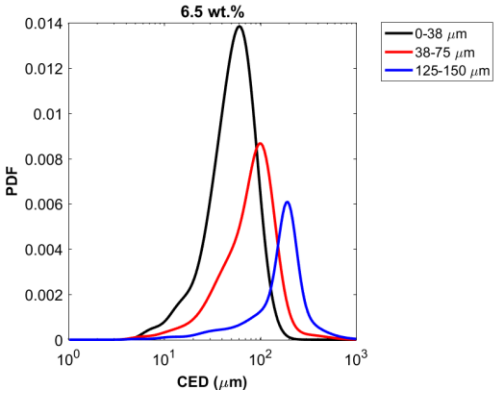

0.5 wt.%
PDF
CED (μm)
0-38 μm
38-75 μm

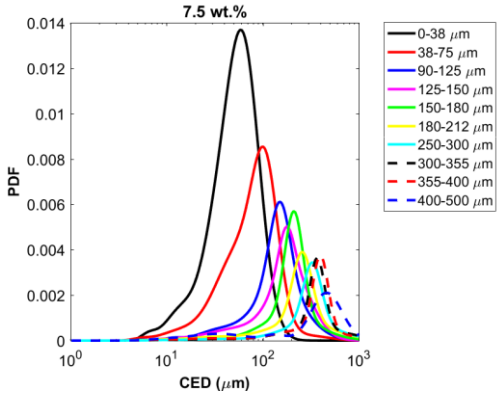

1 wt.%
PDF
CED (μm)
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm

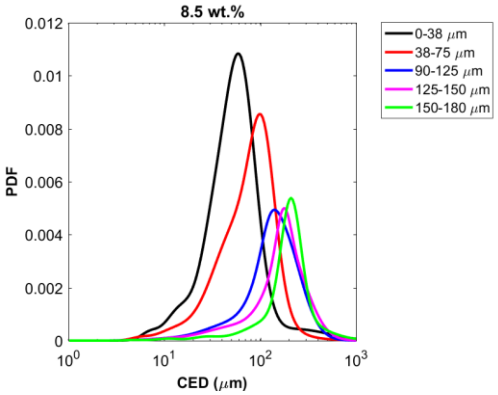

2.5 wt.%
PDF
CED (μm)
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm

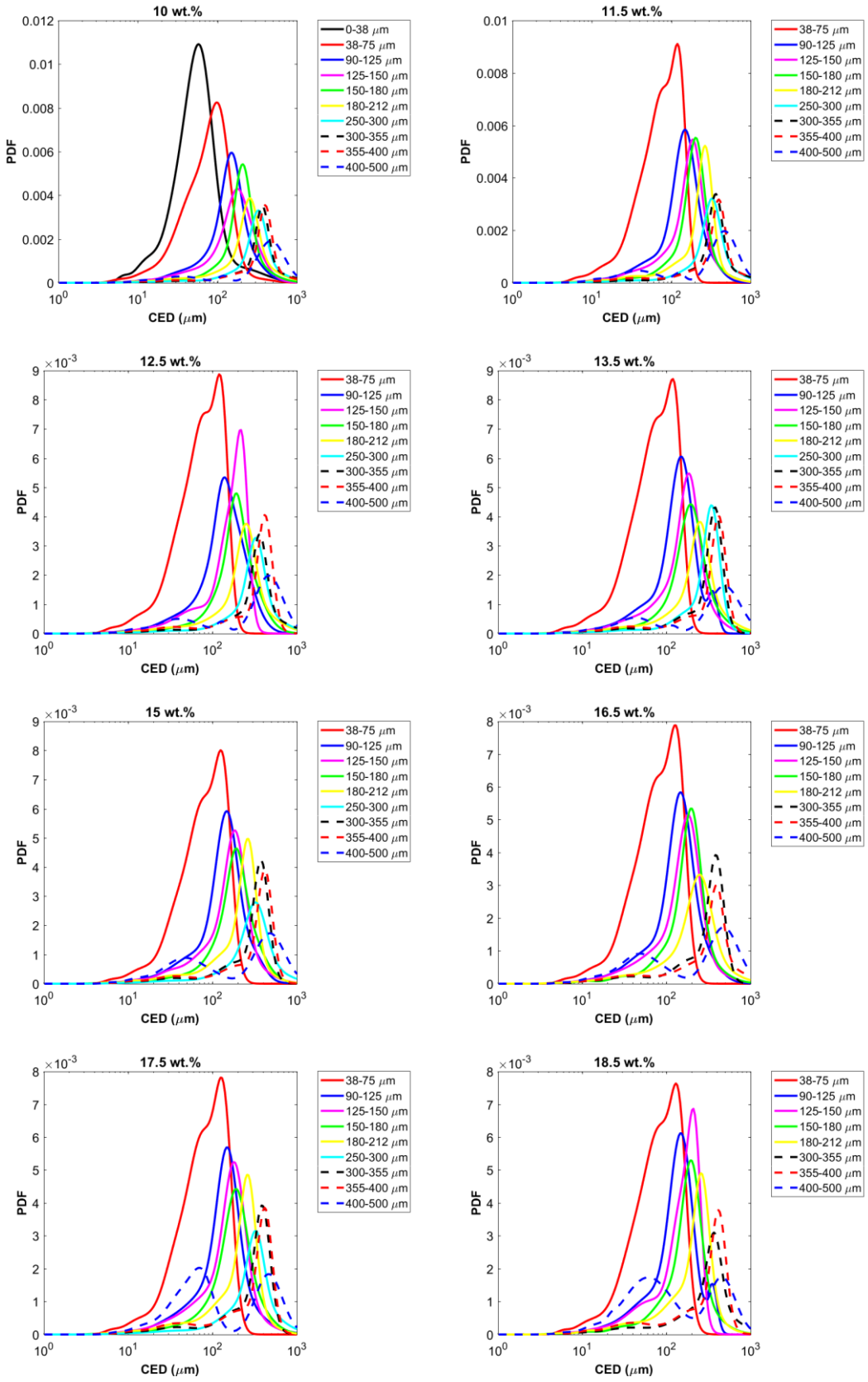

3.5 wt.%
PDF
CED (μm)
0-38 μm
38-75 μm
90-125 μm


5 wt.%
PDF
CED (μm)
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm


6.5 wt.%
PDF
CED (μm)
0-38 μm
38-75 μm
125-150 μm


7.5 wt.%
PDF
CED (μm)
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm


8.5 wt.%
PDF
CED (μm)
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm


10 wt.%
11.5 wt.%
12.5 wt.%
13.5 wt.%
15 wt.%
16.5 wt.%
17.5 wt.%
18.5 wt.%
PDF
CED (μm)
0-38 μm
38-75 μm
90-125 μm
125-150 μm
150-180 μm
180-212 μm
250-300 μm
300-355 μm
355-400 μm
400-500 μm

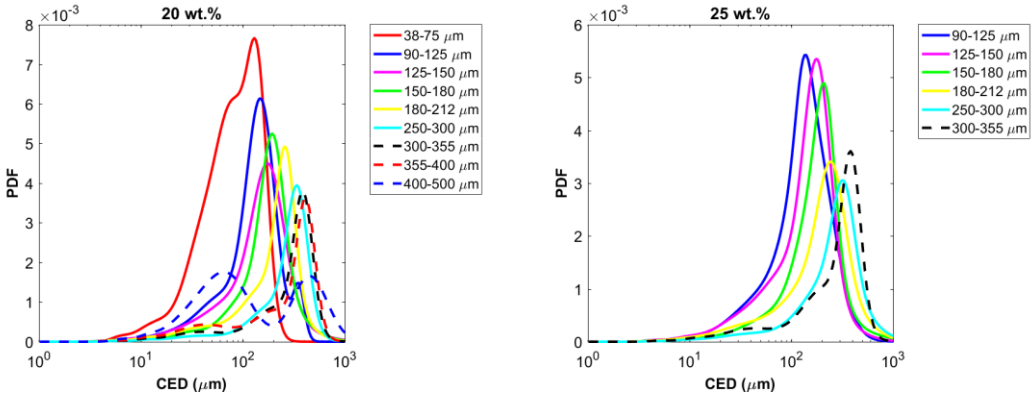


Figure C.4: Effect of particle size on the volume weighted ECD distribution, obtained from the inversion algorithm app, for the different solid loadings tested.

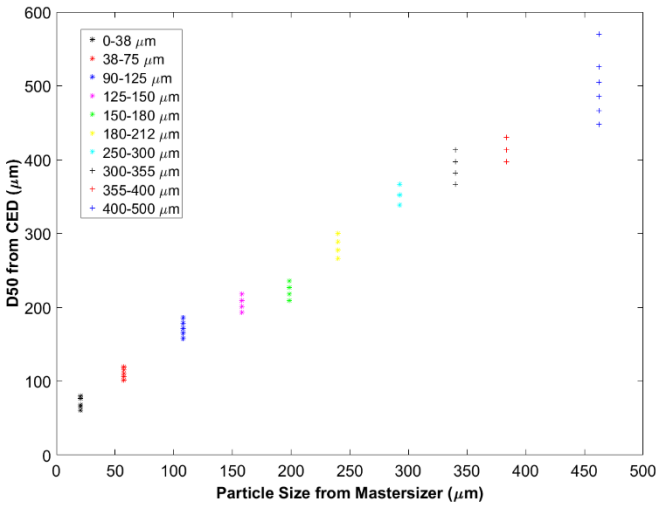


Figure C.5: D50 from the inverted CLD versus D50 obtained for Mastersizer for all particle concentration evaluated.

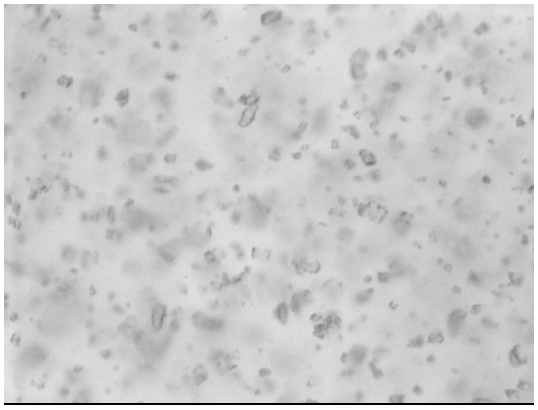


Figure C.6: PVM images of lactose crystal using a suspension of 1 wt.% of L0-38.