



ELSEVIER

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Integration of in situ imaging and chord length distribution measurements for estimation of particle size and shape

Okpeafoh S. Agimelen^{a,*}, Anna Jawor-Baczynska^a, John McGinty^a, Jerzy Dziewierz^b, Christos Tachtatzis^b, Alison Cleary^c, Ian Haley^d, Craig Michie^c, Ivan Andonovic^c, Jan Sefcik^e, Anthony J. Mulholland^{f,*}^a EPSRC Centre for Innovative Manufacturing in Continuous Manufacturing and Crystallisation, Department of Chemical and Process Engineering, University of Strathclyde, James Weir Building, 75 Montrose Street, Glasgow G1 1XJ, United Kingdom^b The Centre for Ultrasonic Engineering, Department of Electronic and Electrical Engineering, University Of Strathclyde, Royal College Building, 204 George Street, Glasgow G1 1XW, United Kingdom^c Centre for Intelligent Dynamic Communications, Department of Electronic and Electrical Engineering, University Of Strathclyde, Royal College Building, 204 George Street, Glasgow G1 1XW, United Kingdom^d Mettler-Toledo Ltd., 64 Boston Road, Beaumont Leys Leicester LE4 1AW, United Kingdom^e EPSRC Centre for Innovative Manufacturing in Continuous Manufacturing and Crystallisation, Department of Chemical and Process Engineering, University of Strathclyde, James Weir Building, 75 Montrose Street, Glasgow G1 1XJ, United Kingdom^f Department of Mathematics and Statistics, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH, United Kingdom

HIGHLIGHTS

- Methods for estimating PSD from CLD and images obtained in situ are developed.
- The particle shape (aspect ratio) is estimated from images obtained in-line.
- Method 1 reduces uncertainty in the aspect ratio at which PSD is estimated from CLD.
- Method 2 decouples aspect ratio from the PSD estimation from CLD.

ARTICLE INFO

Article history:

Received 12 May 2015

Received in revised form

1 October 2015

Accepted 7 January 2016

Available online 14 January 2016

Keywords:

Chord length distribution

Particle size distribution

Particle shape

Focused beam reflectance measurement

Imaging

ABSTRACT

Efficient processing of particulate products across various manufacturing steps requires that particles possess desired attributes such as size and shape. Controlling the particle production process to obtain required attributes will be greatly facilitated using robust algorithms providing the size and shape information of the particles from in situ measurements. However, obtaining particle size and shape information in situ during manufacturing has been a big challenge. This is because the problem of estimating particle size and shape (aspect ratio) from signals provided by in-line measuring tools is often ill posed, and therefore it calls for appropriate constraints to be imposed on the problem. One way to constrain uncertainty in estimation of particle size and shape from in-line measurements is to combine data from different measurements such as chord length distribution (CLD) and imaging. This paper presents two different methods for combining imaging and CLD data obtained with in-line tools in order to get reliable estimates of particle size distribution and aspect ratio, where the imaging data is used to constrain the search space for an aspect ratio from the CLD data.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

One of key steps in the manufacture of particulate products in the pharmaceuticals and fine chemicals industry is crystallisation,

which is widely used for separation and purification of intermediates, fine chemicals and active pharmaceutical ingredients. The crystals come in different sizes and shapes. Subsequent steps in the manufacturing process, such as filtration, drying, blending and formulation of final products, require that the particle sizes and shapes lie within some desirable range. In order to provide monitoring and control of crystallisation processes it is necessary to develop techniques for estimating the shape and size

* Corresponding authors.

E-mail addresses: okpeafoh.agimelen@strath.ac.uk (O.S. Agimelen), anthony.mulholland@strath.ac.uk (A.J. Mulholland).

distribution of particles in situ. There are a number of offline tools (Washington, 1992) that can be used to estimate the particle size distribution¹ (PSD) of crystals produced in a crystallisation process. However, of particular importance to the control of a crystallisation process are tools that can be used in situ. These tools should be suitable for estimating size and shape information of particles dispersed in a slurry without the need for sampling and/or dilution. Examples of such instruments are the focused beam reflectance measurement (FBRM), the three dimensional optical reflectance (3D-ORM) (Heinrich and Ulrich, 2012) and the particle vision and measurement (PVM) (Barrett and Glennon, 2002) sensors.

In-line sensors such as FBRM and 3D-ORM measure a chord length distribution (CLD)² which is related to the size and shape of the particles in a slurry. It has been a long standing challenge to be able to deduce the actual PSD and particle shape from experimental CLD data. In order to do this, an inverse problem needs to be solved. This is usually achieved by suitably discretising both CLD (which is already measured as a discrete distribution) and PSD and then constructing an appropriate transformation matrix relating these two distributions (Wynn, 2003; Li and Wilkinson, 2005; Agimelen et al., 2015). The transformation matrix depends on the choice of size bins used to discretise the two distributions and the corresponding size ranges as well as the shape of particles. The transformation matrix is usually not known in advance and needs to be estimated along with the corresponding PSD (discretised) so that the convolution of the transformation matrix with the PSD yields a CLD which agrees with the experimentally measured CLD. However, this problem is ill posed. There are a number of different transformation matrices and PSDs whose convolutions give rise to the same CLD. Hence the challenge is how to estimate a combination of transformation matrix and PSD whose convolution will agree with an experimentally measured CLD for a given slurry as well as the PSD estimated being physically reasonable and representative of the particles in the slurry.

The approach which was used in previous works (Ruf et al., 2000; Worlitschek et al., 2005; Li et al., 2014; Kail et al., 2009) when estimating the transformation matrix was to assume the same shape (quantified by a metric referred to as aspect ratio) for all the particles in the slurry, and then use a previously estimated³ range of particle sizes in the slurry to construct the transformation matrix. This approach is not suitable for monitoring a crystallisation process where nucleation and/or growth of particles is present as neither the range of particle sizes nor their aspect ratio would be known in advance. A technique which was suitable for estimating the range of particle sizes in a slurry in situ was presented in our previous work (Agimelen et al., 2015). However, like in other previous works, the transformation matrix was constructed with a single aspect ratio for all particles. This leaves open a possibility that the transformation matrix is constructed with inappropriate aspect ratio or that there is a wider range of aspect ratios present for particles of same or different sizes. It was demonstrated in our previous work (Agimelen et al., 2015) that it was still possible to calculate different CLDs that all had a very good agreement with an experimentally measured CLD even though some of the transformation matrices were constructed at aspect ratios that were far from the shape of the particles

described. However, it was also shown that as the aspect ratio deviated further from the true shape of the particles, then the corresponding PSD became increasingly noisy. This situation led to the introduction of a penalty function in order to eliminate unrealistic aspect ratios. However, when there is a wide variation of aspect ratios of the particles in the slurry, there is a need to introduce further constraints on the aspect ratio to reduce the search space and regularise the inverse problem. One way to do this is to get estimates of aspect ratio (within some reasonable bounds) using imaging, and then use this information to constrain the search for a representative aspect ratio. However, the imaging needs to be done in situ in order to develop techniques for estimation of PSD and particle shape which could be used for real time monitoring and control of particle production processes.

While it would be desirable to get good estimates of both PSD and particle aspect ratio using in situ imaging alone, this is currently not the case. The currently available in-line imaging tools (for example, the PVM used in this work) produce 2D projection images. Furthermore, the objects in the images may be partially or completely out of focus, parts of imaged object may cross the image frame or objects may overlap each other.⁴ Although advanced measurement equipment have been developed which can be used to capture 3D images of particles in a slurry and make good estimates of PSD and shape of particles, it requires sampling and dilution flow loops⁵ to allow capturing 3D images of individual particles in a flow-through cell (Eggers et al., 2008; Kempkes et al., 2010; Schorsch et al., 2012, 2014). Therefore this approach may not be generally applicable for in-line monitoring of particle manufacturing processes. Hence the current situation is that PSD cannot be estimated to a good degree of accuracy using routinely available in-line imaging tools. To overcome this challenge, we propose to combine in-line CLD measurements with imaging data to provide more reliable estimates of quantitative particle attributes.

In this paper, we present two different methods for combining imaging data with CLD data for particle size and shape estimation. The first method presented here calculates an estimate of the mean aspect ratio of all the particles in the slurry and then uses this information to constrain the search space for size and shape estimation from the CLD data. In the second method, a distribution of aspect ratios for each particle size is used for the PSD estimation. The distribution of aspect ratios is based on the data from the captured images.

2. Experiment

To demonstrate the technique for estimating particle size and shape information using a combination of the CLD and imaging data, experiments were performed in slurries containing particles of different shapes. The materials and procedure are described below.

2.1. Materials

Three different samples were used for the measurements. Sample 1 consisted of polystyrene (PS) microspheres purchased from EPRUI nanoparticles and Microspheres Co. Ltd. with batch number 2012-5-7, and 0.2 g of the PS microspheres were dispersed in 100 g of isopropanol (IPA) purchased from VWR (20842.323) giving a concentration of 0.2% by weight. Sample 2 consisted of cellobiose octaacetate (COA) particles obtained from GSK. The particles were dispersed in

¹ The term particle size distribution is broadly used here to refer both to continuous analytical probability density functions for particle sizes and discretised probability histograms of the particle sizes.

² Similar to the case of PSD, the term chord length distribution is used to cover both continuous analytical functions and discretised probability histograms.

³ The approach was to estimate the range of particle sizes in a sample by techniques such as sieving, laser diffraction, microscopy or use information supplied by the manufacturer before suspending the particles.

⁴ The issue of objects overlapping each other would not be a problem if an appropriate image processing algorithm which can resolve the objects is used.

⁵ The dilution is necessary to avoid instances of overlapping particles in images.

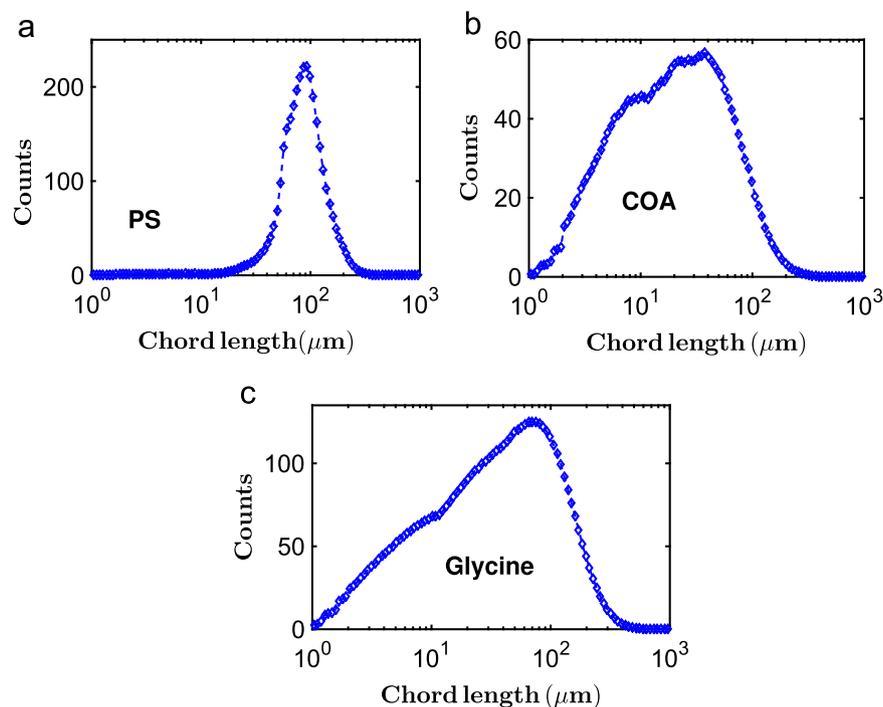


Fig. 1. The CLD (measured with the FBRM G400 sensor) for the (a) PS, (b) COA and (c) glycine samples.

methanol (purchased from VWR (20847.307)) with the same concentration as in sample 1. Sample 3 consisted of glycine (Glycine) crystals obtained by cooling crystallisation from an aqueous solution. The solution with glycine concentration of 340 mg/ml was prepared using glycine (purchased from Sigma-Aldrich (G8898, $\geq 99\%$ TLC)) and deionised water (from an in-house Millipore Water System (18 M Ω /cm)). The solution was cooled from a temperature of 90 °C to a temperature of 43 °C at a rate of 3 °C/min. During this process the glycine crystallised out of solution until an equilibrium particle size distribution was reached. The crystallisation of glycine was monitored with the FBRM probe which showed an initial increase (in time) of chord lengths before eventually reaching a steady state.

2.2. Experimental setup

The suspension of particles for all samples was made in the Mettler Toledo EasyMax 102 system. The EasyMax system consists of a cylindrical jacketed vessel of volume 100 ml with different stirrer and blade options. An anchored overhead stirrer with pitched (45° pitch angle) blades was used in all the experiments in this work. The stirrer shaft and probes were inserted into the slurries through ports located at the top of the setup. The stirring speed was set at 400 rpm in all experiments.

The CLD measurements were made with a Mettler Toledo FBRM G400 probe. The images of the particles were captured with a Mettler Toledo PVM V819 probe during the period of CLD measurement. The FBRM sensor consists of a system of lenses which focus a laser beam onto a spot near the probe window in the slurry. The laser spot moves in a circular trajectory and the back scattered light is detected. The chord length is then calculated as the speed of the laser spot multiplied by the duration of the back scattered light as the laser traverses a particle. The FBRM sensor records the lengths of the chords for a pre-set duration after which the CLD is reported (Heinrich and Ulrich, 2012; Kail et al., 2007, 2008; Worlitschek and Mazzotti, 2003; Agimelen et al., 2015).

The PVM is an in situ microscope which consists of eight laser sources enclosed in a cylindrical tube. The six forward lasers and two back lasers (achieved by reflecting two lasers off a Teflon cap

at the probe window) illuminate the particles in the slurry. The back scattered light is detected by a CCD element from which greyscale images are constructed. The image frame of the CCD array consists of 1360 \times 1024 pixels with a pixel size of 0.8 μ m. The PVM V819 sensor has a maximum acquisition rate of 5 images per second, although lower rates of acquisition could be set depending on requirements. The depth of the focal zone is restricted to about 50 μ m so that all objects that are in focus result in images that have identical magnification levels. Each of the lasers can be switched on or off so that different degrees of illumination can be achieved.

3. Experimental data

The CLD data measured with the FBRM sensor for PS is shown in Fig. 1(a). Images were captured with the PVM sensor during the CLD measurement. A representative image from the PS sample is shown in Fig. 2(a).

Similarly, the CLD data recorded for COA is shown in Fig. 1(b) while that of Glycine is shown in Fig. 1(c). Representative images for COA and Glycine are shown in Fig. 2(b) and (c) respectively. A total of 600 images were acquired each for PS and COA, and 121 images for glycine.

4. Image analysis

As mentioned in the introductory section, an estimate of the PSD and particle shape can be made from images alone without the need to include CLD data. However, due to the reasons discussed in the introductory section, this approach is not always convenient. The techniques presented here utilise images obtained with an in-line measuring tool. However, due to the limitations in the images (as discussed in the introductory section), it is necessary to combine the imaging data with CLD to obtain reasonable aspect ratio and/or size estimates.

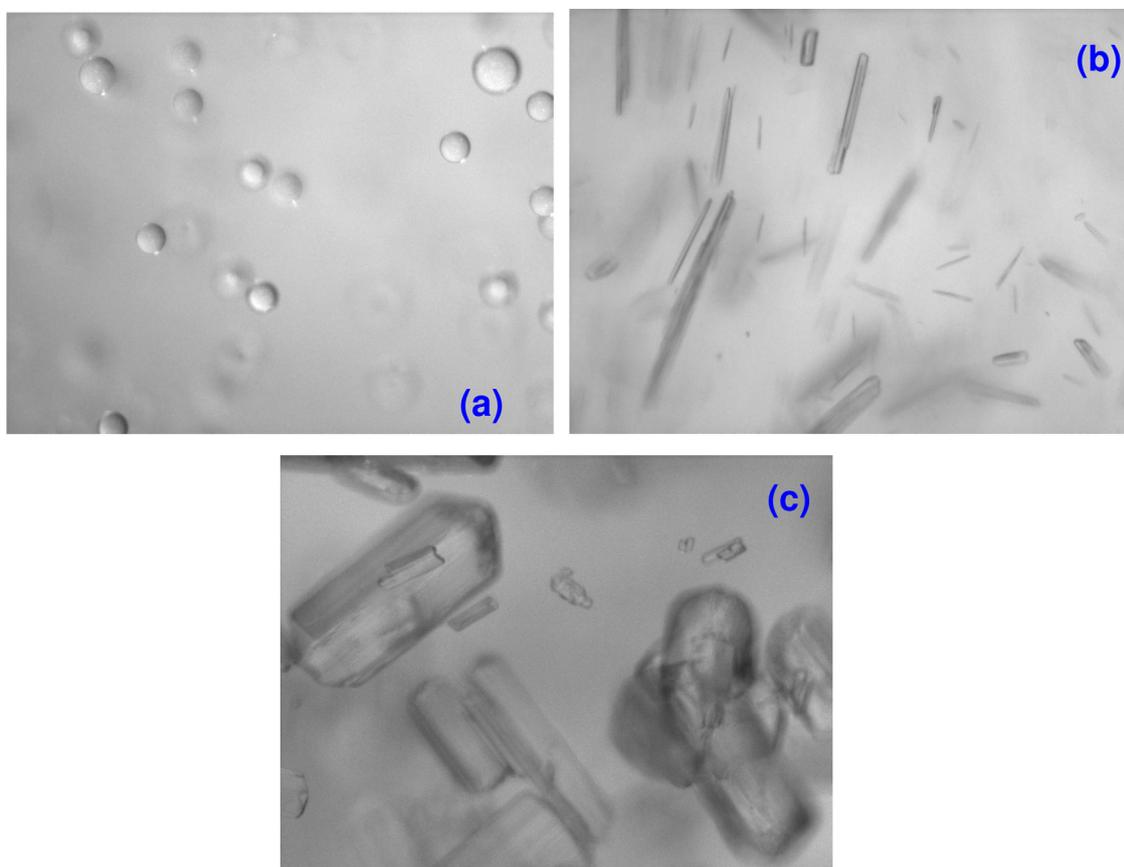


Fig. 2. Representative images (obtained with the PVM V819 sensor) for the (a) PS, (b) COA and (c) glycine samples. The images have the same width of 1088 μm and height of 819 μm .

The images captured with the PVM are processed in order to detect the objects contained in them and hence obtain information about the shape of the particles in the slurry. However, the image processing algorithm used in this work does not have features to resolve an object completely when it does not lie entirely in focal plane of the PVM sensor. Also, it does not have functionalities to resolve overlapping objects in images. For this reason the samples used in this work were deliberately prepared dilute⁶ to reduce instances of overlapping objects. The parameters⁷ of the image processing algorithm were tuned to reject most of the particles that were not entirely in the focal plane of the PVM sensor. However, the imaging data still has some degree of inaccuracy as seen in the error bars of the data (see Section 4.6 of the Supplementary Information). This limitation notwithstanding, the data from the images was sufficiently accurate to demonstrate the techniques developed in this work. Furthermore, images which do not contain objects that are contained completely within the image frame were also discarded. This situation of having to discard some images reduces the number of data sets that can be gathered from the images. However, it can be shown (see Section 2 of the Supplementary Information) that with a sample size (number of objects) of about 500 the error incurred in estimating the aspect ratio is reduced to a reasonable extent. However, for a

more robust estimate of the PSD using imaging data, a larger number of objects will be required. The number⁸ of detected objects used in this work is just sufficient to demonstrate the methods developed here. The issue of objects not completely in focus can be dealt with if additional functionalities are added to the image processing algorithm, but this is beyond the scope of this work. The key steps for detecting objects in the images captured by the PVM sensor are summarised in Section 4.1.

4.1. Object detection

The raw greyscale images from the PVM sensor are passed through a median filter to remove speck noise from the image background which is homogeneous. At this stage objects on the boundary of the image frame are removed. Any object with surface area below 900 pixels⁹ is considered noise and excluded from processing. Finally, a closing operation with a disk structural element is used to join broken edges. The resulting blob properties such as area, centroid, eccentricity, convex area, and major and minor axes can be obtained. These steps are summarised in Fig. 3. The tunable parameters of the image processing algorithm are summarised in Section 4.5 of the Supplementary Information.

⁶ Low slurry densities have been used here for the purpose of methods development. Future work will involve the investigation of the applicability of the methods developed here at higher slurry densities (with a more advanced image processing algorithm) using suspensions of particles of known PSD, then the degree of deviation of the results from the known PSD can be quantified at different slurry densities.

⁷ The parameters of the algorithm need to be tuned for different samples due to variation of contrast.

⁸ A total of 1393 objects were detected for PS, 1810 for COA and 526 for glycine.

⁹ The surface area of 900 pixels represents length dimensions of approximately 30×30 pixels (assuming a square geometry). This implies that objects that are smaller than approximately $24 \mu\text{m}$ are rejected by the image processing algorithm. The consequence is that there is no estimate of aspect ratio for these small objects. However, the particles used in this work have sizes mostly in the range of $100 \mu\text{m}$ so that the effects of this are minimal.

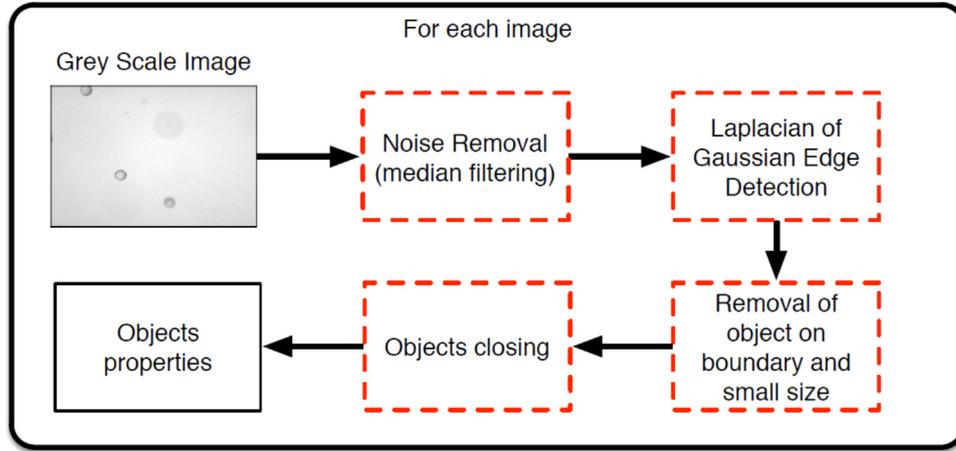


Fig. 3. A schematic for the key steps of the image processing algorithm used in this work.

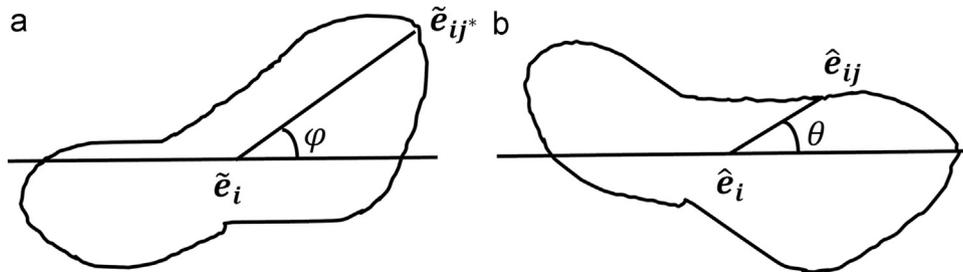


Fig. 4. A sketch of the particle characterisation procedure. The boundary pixels before rotation are shown in (a). The line joining the farthest boundary pixel \tilde{e}_{ij}^* to the centroid \tilde{e}_i makes an angle φ with the horizontal axis. (b) A rotation is performed so that the farthest pixel from the centroid lies on the horizontal axis. The indices of the remaining pixels are assigned relative to this pixel.

4.2. Characterising particle shape

The following procedure is used to obtain a shape descriptor, which is then used to characterise the shape of each particle. Each boundary pixel j of object i has coordinates $\tilde{e}_{ij} = [\tilde{e}_{ij,x}, \tilde{e}_{ij,y}]$, where the Cartesian coordinates system has been used. The centroid of object i has coordinates \tilde{e}_i . The distance \tilde{d}_{ij} of the centroid of object i to the pixel j is given as

$$\tilde{d}_{ij} = \|\tilde{e}_i - \tilde{e}_{ij}\|. \quad (1)$$

Then for each object i , locate the pixel (with index j^*) whose distance from the centroid \tilde{d}_{ij^*} is largest as

$$\tilde{d}_{ij^*} = \max_j \{\tilde{d}_{ij}\}. \quad (2)$$

The angle between the line joining this farthest pixel to the centroid and the horizontal axis is φ as shown in Fig. 4(a). Then, for each object i , transform the coordinates of each pixel by performing a rotation through the angle φ so that the line joining the centroid to the farthest pixel becomes parallel to the horizontal axis. This operation transforms the coordinates of each pixel j to \hat{e}_{ij} and the centroid to \hat{e}_i . The line joining the pixel j to the centroid now makes an angle θ with the horizontal axis as shown in Fig. 4 (b). Due to the rotation, the farthest pixel from the centroid corresponds to $\theta = 0$. The distance of each pixel j from the centroid is the same of course.

Since the sample rate with respect to θ is not uniform, the pixel distances were resampled with N_p uniformly spaced θ values constructed as $\theta_p = p\Delta\theta$, $\Delta\theta = 2\pi/N_p$, $p = 1, 2, \dots, N_p$. This allows the vector of all pixels distances (from the centroid) for object i to be written as

$$\mathbf{d}_i = [\tilde{d}_{i1}, \tilde{d}_{i2}, \dots, \tilde{d}_{iN_p}]. \quad (3)$$

Finally, the average vector $\bar{\mathbf{d}}$ of all pixel distances for all objects detected by the image processing algorithm can be calculated as

$$\bar{\mathbf{d}} = \frac{1}{N_{obj}} \sum_{i=1}^{N_{obj}} \mathbf{d}_i, \quad (4)$$

where N_{obj} is the number of objects detected from all the images analysed.

A plot of $\bar{\mathbf{d}}$ versus angle (θ) can be made as shown in Fig. 5(a)–(c). For near spherical particles, the shape descriptor $\bar{\mathbf{d}}$ is nearly constant as in the case of PS in Fig. 5(a). However, for elongated particles, the shape descriptor $\bar{\mathbf{d}}$ has two minima and maxima as in the cases of CoA and glycine in Fig. 5(b) and (c), respectively. The shape descriptor shown in Fig. 5(a)–(c) is similar to the type described in Schorsch et al. (2012).

In an ideal situation, the shape descriptor for spherical particles will be constant at a value representing the radius of the spherical particles. However, since the PS particles are not perfectly spherical there is slight variation in the dimensions so that an average aspect ratio r (the ratio of the minor to the major dimension) can be estimated. Similarly, the maxima in the shape descriptors for the CoA and glycine particles (in Fig. 5(b) and (c)) represent the major dimension of the particles while the minima represent the minor dimension of the particles.

For the case of the PS particles the mean aspect ratio r is estimated from the minimum dimension \mathcal{L}_{min} and maximum dimension \mathcal{L}_{max} (see Fig. 5(a)) as

$$r = \frac{\mathcal{L}_{min}}{\mathcal{L}_{max}}. \quad (5)$$

In the case of elongated particles, the maximum dimension (average length of particles) is given as $\mathcal{L}_{max} = \mathcal{L}_{1max} + \mathcal{L}_{2max}$ (shown in Fig. 5(b) and (c)) and the minimum dimension (average

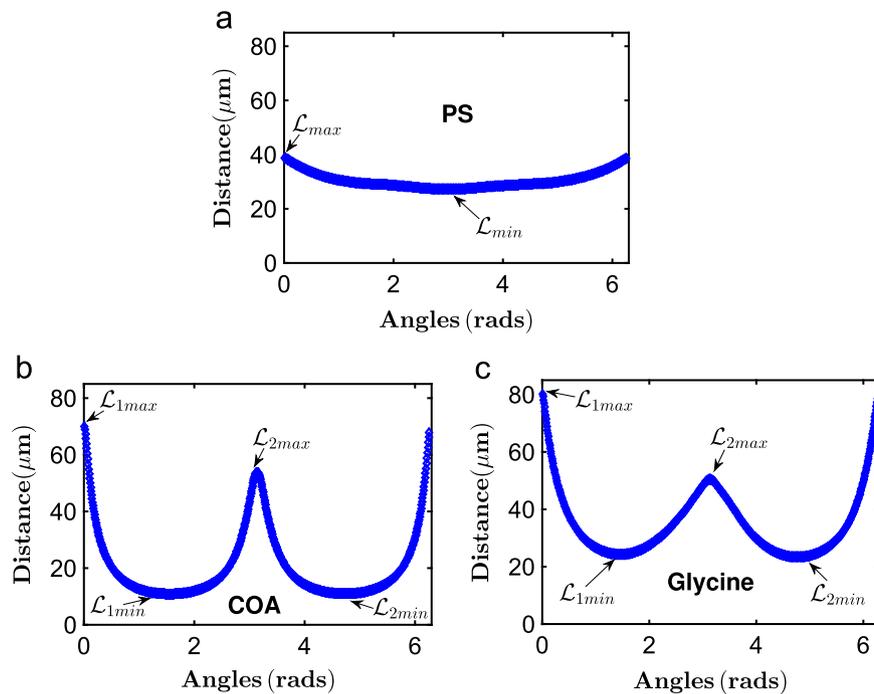


Fig. 5. The shape descriptor (\bar{d} in Eq. 4) as a function of the angle (θ) for the (a) PS, (b) COA and (c) glycine samples.

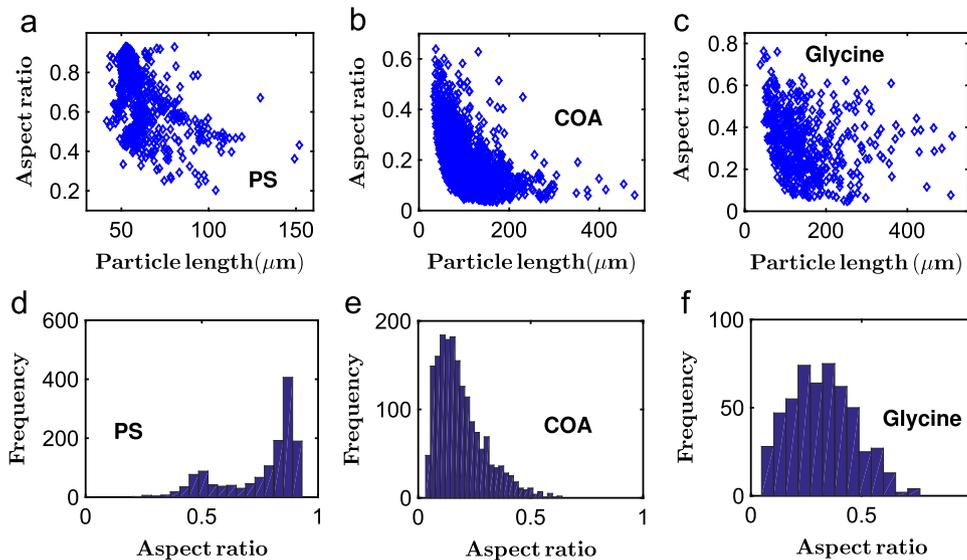


Fig. 6. Scatter plots of the aspect ratio versus particle length for individual objects for (a) PS, (b) COA and (c) glycine. The histograms of the aspect ratios are shown in (d) PS, (e) COA and (f) glycine.

width of particles) is given as $\mathcal{L}_{min} = \mathcal{L}_{1min} + \mathcal{L}_{2min}$. So that the average aspect ratio can then be estimated using Eq. (5).

However, the aspect ratios for individual particles will be different from r . The aspect ratio for each particle is estimated from the shape descriptor corresponding to that particle. Once the aspect ratios of individual particles are estimated, then a scatter plot of aspect ratio versus particle length can be made as shown in Fig. 6(a)–(c). The shape descriptors for individual particles are not always smooth as in the cases shown in Fig. 5(a)–(c). They contain different degrees of variation due to imperfections in the particles and images (see Section 4.6 of the Supplementary Information for details).

Furthermore, the aspect ratios estimated from the shape descriptor for individual particles contain some artefacts (as can be seen in Fig. 6) due to a number of factors. These factors include

deformations in the particles (that is, particles whose shapes deviate from the majority of particles in the slurry), impurity objects in the monitored slurry and particles not completely in focus.¹⁰ For example, aspect ratios as low as about 0.2 in the case of the spherical particles in Fig. 6(a) or the minor peak at aspect ratio $r \approx 0.5$ in Fig. 6(d) are artefacts.

Fig. 6 (b) and (c) suggests the presence of particles of sizes up to about 500 μm in the COA and glycine samples respectively. However, the number of data points (in Fig. 6(b) and (c)) corresponding to these large particles may not be representative of the

¹⁰ The image processing algorithm parameters are tuned to remove particles that are out of focus. However, when the particle is only partially in focus, the image processing algorithm detects only part of the particle and this leads to an error in the estimated aspect ratio for that particle.

actual number of these large particles in the slurry. This is because the image processing algorithm has been designed to remove objects making contact with the image frame, and larger particles have a higher probability of making contact with the image frame. In the situation where the PSD is to be estimated from image data alone, this probability will need to be taken into account (BSI, 2014). However, since the objective here is just to estimate the aspect ratio of particles of different sizes, this is not a crucial issue.

The histogram (in Fig. 6(d))¹¹ for the spherical particles has a dominant mode close to the aspect ratio of 0.85 which is close to the average aspect ratio $r \approx 0.8$ estimated from the average shape descriptor (in Fig. 5(a)) for this slurry. Similarly the modes of the histograms (in Fig. 6(e) and (f)) for COA and glycine occur close to aspect ratios of 0.2 and 0.4 respectively. These values are close to the average aspect ratios $r \approx 0.2$ (for COA) and $r \approx 0.4$ (for glycine) obtained from their respective shape descriptors in Fig. 5 (b) and (c).

5. Modelling chord length distribution

The sizes of particles in a slurry can be represented by the equivalent spherical diameter as was done in Agimelen et al. (2015). However, a characteristic length L could also be used, which can be chosen as the distance between the two extreme points in the particles' geometry. Since the estimated sizes from the images is L , this metric is used here for consistency with the image data. Once the metric for particle sizes has been chosen, then the PSD can then be expressed in terms of the chosen particle size metric. The PSD \mathbf{X} is related to the CLD \mathbf{C} by means of a convolution function (Hobbel et al., 1991; Heinrich and Ulrich, 2012; Agimelen et al., 2015) and the relationship can be written in matrix form as Agimelen et al. (2015)¹²

$$\mathbf{C}(s) = \mathbf{A}(s, L)\tilde{\mathbf{X}}(L), \quad (6)$$

where s is the chord length, and $\tilde{\mathbf{X}}$ is the length weighted PSD (Agimelen et al., 2015) given as

$$\tilde{X}_i = \bar{L}_i X_i, \quad i = 1, 2, 3, \dots, N. \quad (7)$$

The PSD X_i (which is actually a histogram) consists of N bins. The characteristic size \bar{L}_i of particle size bin i is the geometric mean of sizes L_i and L_{i+1} as $\bar{L}_i = \sqrt{L_i L_{i+1}}$. The bin boundaries are calculated as

$$L_i = L_{\min} \omega^{i-1}, \quad i = 1, 2, \dots, N+1 \quad (8)$$

where

$$\omega = \left(\frac{L_{\max}}{L_{\min}} \right)^{1/N}, \quad (9)$$

where L_{\min} is the left boundary of the first particle size bin and L_{\max} is the right boundary of the last particle size bin.

In previous works (Ruf et al., 2000; Worlitschek et al., 2005; Li et al., 2005, 2013, 2014; Yu et al., 2008) the values of L_{\min} and L_{\max} were estimated from suitable measurements. However, the technique of estimating L_{\min} and L_{\max} directly from the bin boundaries of the CLD histogram using a moving window technique has been demonstrated to yield more accurate results (Agimelen et al., 2015). This window technique is more suitable for estimating the

sizes of particles in-line in a process where particle size information is obtained from the CLD (Agimelen et al., 2015).

The length weighting applied to the PSD \mathbf{X} in Eq. (7) is necessary because the CLD for a population of particles is biased towards particles of larger sizes (Agimelen et al., 2015; Hobbel et al., 1991; Simmons et al., 1999; Vaccaro et al., 2007). The forward problem in Eq. (6) is implemented by considering a chord length histogram C_j of M bins where the characteristic chord length \bar{s}_j of bin j is the geometric mean of the chord lengths of s_j and s_{j+1} as outlined in Li and Wilkinson (2005) and Agimelen et al. (2015).

If the PSD for a population of particles is known, then the CLD can be calculated using Eq. (6). However, in practical situations, the particle size histogram X_i is not known in advance resulting in the inverse problem of calculating an unknown PSD X_i from a known CLD C_j . For this reason, the forward problem in Eq. (6) is reformulated as

$$\mathbf{C}(s) = \tilde{\mathbf{A}}(s, L)\mathbf{X}(L), \quad (10)$$

where the matrix $\tilde{\mathbf{A}}$ is obtained from matrix \mathbf{A} by multiplying each column of \mathbf{A} by the corresponding particle length as described in Agimelen et al. (2015).

Each column i of matrix A_{ji} is calculated from the CLD of a single particle (single particle CLD) of length L_i and given aspect ratio r_i . In the current work, the single particle CLD used in constructing the columns of matrix \mathbf{A} are obtained from the analytical Li and Wilkinson (LW) model (Li and Wilkinson, 2005).¹³

The process of calculating the single particle CLD involves computing the relative likelihood of obtaining a chord of length s from a particle of a given length and aspect ratio (Li and Wilkinson, 2005). The LW model gives a probability density function (PDF) which can be used in making this calculation for ellipsoidal shaped particles.¹⁴ The PDF is derived from an ellipsoid of semi major axis length a , semi-minor axis length b and aspect ratio $r=b/a$ (Li and Wilkinson, 2005). The LW model gives the probability $p_{L_i}(s_j, s_{j+1})$ of obtaining a chord whose length lies between s_j and s_{j+1} from an ellipsoid of characteristic length $\bar{L}_i = 2a_i$ (see Section 1 of the Supplementary Information and Agimelen et al., 2015; Li and Wilkinson, 2005 for the mathematical expression for $p_{L_i}(s_j, s_{j+1})$). Once the probabilities are calculated, then for each row j of matrix \mathbf{A} the columns are constructed as

$$A_j = \left[p_{L_1}(s_j, s_{j+1}), p_{L_2}(s_j, s_{j+1}), \dots, p_{L_i}(s_j, s_{j+1}), \dots, p_{L_N}(s_j, s_{j+1}) \right]. \quad (11)$$

6. Incorporating aspect ratio from images

As stated in Section 5 the calculation of a column of matrix \mathbf{A} requires the characteristic size of the particle size bin corresponding to that column, as well as the aspect ratio r of the particle of that characteristic size. In the previous work (Agimelen et al., 2015), all particles were assumed to have the same aspect ratio, and its value was estimated using an algorithm based solely on CLD data. The single aspect ratio approach will be used in Section

¹¹ The uniform bin widths of the histograms in Fig. 6(d)–(f) were estimated using the Freedman–Diaconis rule (Izenman, 1991). The number of bins was then estimated as 18 bins for PS, 28 bins for COA and 13 bins for glycine.

¹² Note that the symbol L was used to represent the length of a chord in Agimelen et al. (2015). However, the symbol L is used to represent the characteristic length and s the length of a chord in this work. The CLD \mathbf{C} and PSD $\tilde{\mathbf{X}}$ in Eq. (2) have been discretised. As such they are not continuous probability density functions and the term distribution is used in this work for simplicity.

¹³ Even though the model by Vaccaro et al. (2007) gave estimates of particle aspect ratios that were closer to the estimates from images in Agimelen et al. (2015), the LW model is used in this work for the CLD calculation. The reason is that the Vaccaro model is restricted to small values of aspect ratios $r \leq 0.4$, whereas the image data in Fig. 3 cover aspect ratios of $r \approx 0.1$ to $r \approx 0.9$. The LW model covers the entire range from $r=0$ to $r=1$.

¹⁴ The shapes of the COA and glycine particles in this work have been approximated as ellipsoids. However, this is only an approximation as Fig. 2 (c) clearly shows that the glycine particles are faceted. The use of ellipsoids to represent faceted objects introduces some discrepancy between the single particle CLDs of both objects (see Section 8 of the Supplementary Information for details). However, the ellipsoid approximation used here is sufficient to illustrate the methods presented here.

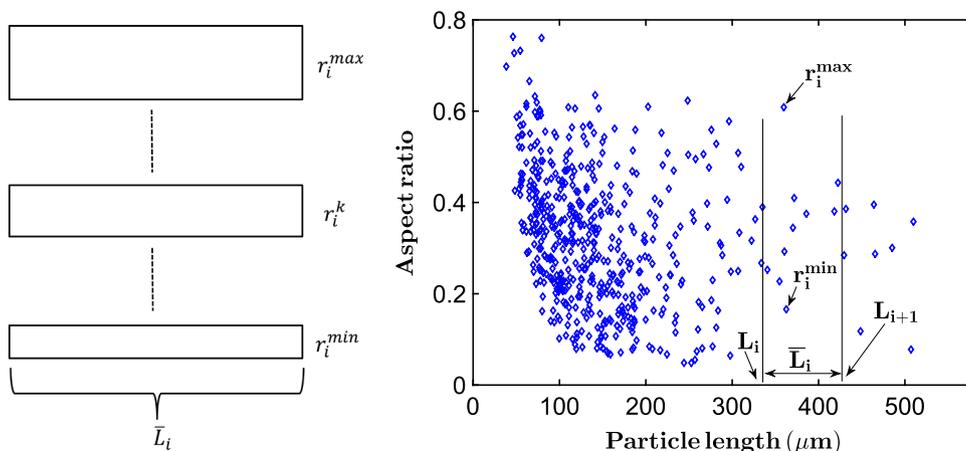


Fig. 7. Left: a schematic to illustrate the assignment of different aspect ratios to particles of the same length. Right: a scatter plot to illustrate the maximum aspect ratio r_i^{\max} and minimum aspect ratio r_i^{\min} for a particular bin i for calculating the columns (of slice i) of the 3 D matrix in Eq. (15).

6.1, where the single aspect ratio value is estimated from imaging data. This approach is most suitable for the case of spherical particles where the aspect ratios of the individual particles are tightly packed around some mean value. However, for the case of particles where there is a wider spread of aspect ratios, a variation of aspect ratios for different particle sizes can also be used. The corresponding technique is outlined in Section 6.2.

6.1. Method 1: population CLD with a single aspect ratio

When the aspect ratios are tightly packed around some mean value as in the case of spherical (or near spherical) particles in Fig. 6(a) and (d), it may be desirable to use a mean value for all particles. Although, the mean aspect ratio estimated from the images provides the best estimate based on available data, there is an uncertainty due to sampling limitations and various artefacts discussed above. This necessitates a search (within a suitable confidence interval) around the mean aspect ratio estimated from the images for an aspect ratio which best matches the experimentally measured CLD. This results in the search space being narrowed down leading to significantly lower computation time and less uncertainty in subsequent calculations.

In this section, the transformation matrix $\tilde{\mathbf{A}}$ in Eq. (10) is constructed with a single aspect ratio for all its columns. The aspect ratio r is chosen from a range given as

$$r \in [\bar{r} - \mathcal{N}\sigma_r, \bar{r} + \mathcal{N}\sigma_r], \quad (12)$$

where σ_r is the standard deviation of all aspect ratios estimated from images and \mathcal{N} is the number of standard deviations chosen. The purpose of Eq. (12) is to constrain the search space for the inverse problem. The details of the inverse problem will be given in Section 7. Hence, Method 1 corresponds to the case described in Agimelen et al. (2015) where each column of matrix $\tilde{\mathbf{A}}$ consists of the single particle CLD of a particle of length \bar{L} and aspect ratio r , where the search space for r is reduced by means of Eq. (12).

6.2. Method 2: population CLD with multiple aspect ratios

The Method 1 presented in the previous subsection assigns the same aspect ratio to all the particles in the slurry. This method is capable of getting reasonable estimates of the PSD. However, to take the variation of particle shape into account, a second method is presented here in which multiple aspect ratios are assigned to particles of the same size. This method is particularly relevant for particles whose shape is needle-like or near rectangular as illustrated on the left of Fig. 7. The second method is outlined below.

In Method 2, the particles in bin i are assigned the same characteristic size \bar{L}_i but different aspect ratios. The aspect ratios assigned to the particles run from r_i^{\min} to r_i^{\max} as illustrated in Fig. 7. The procedure is to divide the particles in bin i into N_r subgroups ($N_r=50$ in this case, see Section 4.1 of the Supplementary Information for details), and the particles in subgroup k are assigned the same aspect ratio r_i^k . The aspect ratios of the subgroups are uniformly spaced so that the aspect ratios are given as

$$r \in \{r_i^1 = r_i^{\min}, r_i^1 + \Delta r, \dots, r_i^k, \dots, r_i^{N_r} = r_i^{\max}\}, \quad (13)$$

where

$$\Delta r = \frac{r_i^{\max} - r_i^{\min}}{N_r - 1}. \quad (14)$$

However, the aspect ratios r_i^k of the different subgroups need to be weighted by different amounts as the numbers of particles with given aspect ratios in the images may not necessarily be uniform from r_i^{\min} to r_i^{\max} . The data on the right of Fig. 7 suggests an approximately uniform distribution of the aspect ratios from r_i^{\min} to r_i^{\max} , hence the aspect ratios were assigned to each of the subgroup k with equal weight in this case.

Once aspect ratios have been assigned to different particle size bins,¹⁵ then a 3D probability array \mathcal{A}_{jki} is constructed. Each slice i of the array corresponds to a particle of size \bar{L}_i , and each column k of slice i contains the probabilities $p_{\bar{L}_i^k}(s_j, s_{j+1})$ of obtaining chords whose lengths lie between s_j and s_{j+1} from a particle of size \bar{L}_i and aspect ratio r_i^k . Hence the 3D array consists of M rows, N_r columns and N slices. The transformation matrix A_{ji} in Eq. (10) is then obtained from the 3D array by averaging over the slices as

$$A_{ji} = \frac{1}{N_r} \sum_{k=1}^{N_r} \mathcal{A}_{jki}. \quad (15)$$

This simple averaging is carried out since the aspect ratios r_i^k are assigned to the subgroups k with equal weights. This is the simplest way to construct the matrix A_{ji} from the 3D array \mathcal{A}_{jki} , and this approach is supported by the data from the images as shown on the right of Fig. 7. It is possible to introduce a probability distribution for the aspect ratios assigned to the subgroups, but this simple approach has been used here for the purpose of illustrating

¹⁵ Details of the technique for combining the aspect ratios with the windowing technique developed in Agimelen et al. (2015) can be found in Section 5 of the Supplementary Information.

the technique. Once the matrix A_{ji} has been constructed, then the forward problem in Eq. (10) can be solved for a given PSD X_i .

7. PSD estimation

As mentioned in Section 5, the problem encountered in practical situations is the estimation of the PSD X_i corresponding to an experimentally measured CLD C_j^* . This is the inverse problem to the forward problem given in Eq. (10). One of the key steps in the process of the PSD estimation is to determine the transformation matrix \tilde{A} in Eq. (10) as accurately as possible. The level of accuracy of the matrix \tilde{A} depends on the values of L_{min} and L_{max} as well as the aspect ratio(s) used in calculating its columns.

To determine the best possible values of L_{min} and L_{max} , the forward problem in Eq. (10) is rewritten as

$$\mathbf{C} = \tilde{\mathbf{A}}\mathbf{X} + \boldsymbol{\epsilon}, \quad (16)$$

where $\boldsymbol{\epsilon}$ is an additive error between the calculated and experimentally measured CLD. Then for given values of L_{min} , L_{max} , values of the fitting parameter γ are found¹⁶ which minimises the objective function f_1 given as

$$f_1 = \sum_{j=1}^M \left[C_j^* - \sum_{i=1}^N \tilde{A}_{ji} X_i \right]^2, \quad (17)$$

where

$$X_i = e^{\gamma_i}, \quad i = 1, 2, 3, \dots, N. \quad (18)$$

A trial solution of $\gamma_i = 0$ was used in the calculation of the vector X_i from Eq. (17). Once the solution vector X_i is obtained, then it is used to solve the forward problem in Eq. (10) to obtain a calculated CLD C_j , $j = 1, 2, \dots, M$, where M is the number of bins in the CLD histogram.¹⁷ The procedure is repeated until the optimum values of L_{min} and L_{max} are found for which there is the best match between the calculated and experimentally measured CLD.

The objective function f_1 given in Eq. (17) is suitable for estimating the optimum values of L_{min} and L_{max} whether the same aspect ratio is assigned to all particles or a distribution of aspect ratios is assigned to particles of the same size. In the case where a single aspect ratio is assigned to all particles, the objective function f_1 is not suitable for picking out the best aspect ratio within the confidence interval of aspect ratios. This task is accomplished with another objective function f_2 (to be introduced in Section 7.1). The calculation for estimating the best aspect ratio (within the confidence interval of aspect ratios) using the objective function f_2 is carried out with the values of L_{min} and L_{max} estimated with the objective function f_1 . However, in the other case where a distribution of aspect ratios is assigned to particles of the same size, the objective function f_2 is not used as the problem of determining the best aspect ratio has been removed.

Once the optimum values of L_{min} and L_{max} and (in the case of Method 1) the best aspect ratio have been estimated, then the PSDs (both number and volume based) can be calculated.

¹⁶ The Levenberg–Marquardt (LM) algorithm as implemented in Matlab was used in this work to solve the optimisation problem here. The PSD X_i is estimated by means of the parameter γ_i . An initial value of γ_i is passed on to the LM algorithm which then searches for the optimum value of γ_i to fit the given CLD. Since the PSD X_i is defined as an exponential function in Eq. (18), the parameter γ_i can take values in the interval $(-\infty, +\infty)$ and still give $X_i \geq 0$. This implies that the non-negativity requirement on the PSD is maintained by the formulation of X_i in Eq. (18). Therefore the LM algorithm was run without the use of lower or upper bounds as the parameter γ_i is defined in $(-\infty, +\infty)$.

¹⁷ The value of $M=100$ was used in all the calculations here to mimic the number of bins set in the FBRM G400 sensor. The values of $N=70$ and $N=50$ were used in Methods 1 and 2 respectively. See Section 4.2 of the Supplementary Information for more details on the choice of the values of N for the two methods.

However, these PSDs may not be reasonably smooth, showing non-physical oscillations as is often the case when solving ill-posed problems. In such cases, a third objective function f_3 (to be introduced in Section 7.3) is used to calculate smooth PSDs. The calculation is carried out using the optimum values of L_{min} and L_{max} obtained with the objective function f_1 and in the case of Method 1, the best aspect ratio obtained with the objective function f_2 . The calculation of the smooth PSDs with the objective function f_3 is done using suitable criteria described in Section 4.4 of the Supplementary Information.

As stated above, the objective function f_1 is used to obtain the optimum values of L_{min} and L_{max} for both Methods 1 and 2. A given pair of L_{min} and L_{max} is said to be optimum when the corresponding calculated CLD \mathbf{C} has the best match with the experimentally measured CLD \mathbf{C}^* . The level of agreement between the calculated and experimentally measured CLD is assessed by computing the L_2 norm

$$\|\mathbf{C}^* - \mathbf{C}\| = \sqrt{f_1}. \quad (19)$$

The values of L_{min} and L_{max} for which the L_2 norm in Eq. (19) reaches a minimum are chosen as the optimum values.

7.1. PSD estimation for Method 1

In Method 1 the search for the optimum values of L_{min} and L_{max} using the objective function f_1 is done at each aspect ratio within the confidence interval in Eq. (12). However, for particles of a given shape, the L_2 norm in Eq. (19) initially decreases with increasing aspect ratio and then becomes level (see Section 4.3 of the Supplementary Information for details). This leads to non-uniqueness in determining the optimum value of r (Agimelen et al., 2015). This problem of non-uniqueness is removed by using a modified objective function f_2 which contains a penalty term to control the size of the calculated PSD vector as

$$f_2 = \sum_{j=1}^M \left[C_j^* - \sum_{i=1}^N \tilde{A}_{ji} X_i \right]^2 + \lambda_1 \sum_{i=1}^N X_i^2, \quad (20)$$

where the parameter λ_1 sets the level of imposed penalty. The value of λ_1 is chosen by comparing the magnitudes of the terms in Eq. (20) (see Section 4.3 of the Supplementary Information). The aspect ratio at which the objective function f_2 reaches a minimum is then chosen as the optimum.

The solution vector (which is a number based PSD) obtained from Eq. (20) is not necessarily smooth as the penalty imposed on the solution vector only restricts its magnitude. With this penalty function, the LM algorithm could settle on a solution vector that contains some local fluctuations but whose value of f_2 is slightly less than a nearby solution that is smooth. For this reason, a new objective function f_3 (see Section 7.3 for details) which contains a penalty term to control the second derivative (to improve the smoothness of the solution vector) of the solution vector is used to estimate a number based PSD whose corresponding CLD is compared with the experimental data.

7.2. PSD estimation for Method 2

In Method 2 (described in Section 6.2) particles of different characteristic sizes \bar{L}_i are assigned a range of aspect ratios as outlined in Section 6.2. This eliminates the need to search for the best global aspect ratio which is the situation in Method 1. Hence in Method 2, it is only necessary to search for the best values of L_{min} and L_{max} using the objective function f_1 . Once the optimum transformation matrix is obtained using the objective function f_1 , then the corresponding smoothed solution is obtained with the objective function f_3 (given in Eq. (21)).

7.3. Volume based PSD

It is often necessary to recast the PSD X_i (which is number based) in Eq. (17) as a volume based PSD since most instruments for measuring PSD give the data in terms of a volume based PSD. A new technique which allows suitable penalties to be imposed on the calculated volume based PSD \mathbf{X}^v was introduced in Agimelen et al. (2015). In the current work, a smoothing penalty (referred to in Section 7.1, see also Roths et al., 2001) is imposed. This is because the estimated volume based PSD \mathbf{X}^v may contain significant non-physical oscillations even though the corresponding number based PSD contains none or minor fluctuations. The objective function f_3 (see Eq. (21)) used to impose smoothness on the volume based PSD can also be used to obtain a smooth number based PSD. Hence the function f_3 is given in Eq. (21) in terms of generic quantities depending on whether the number based or volume based PSD is being computed.

The function f_3 is given as

$$f_3 = \sum_{j=1}^M \left[C_j^+ - \sum_{i=1}^N A_{ji}^+ X_i^+ \right]^2 + \lambda_2 \sum_{i=1}^N [\nabla_h^2 X_i^+]^2. \quad (21)$$

In the case of a number based PSD, the CLD $C_j^+ = C_j^*$ (the experimentally measured CLD), the matrix $A_{ji}^+ = \tilde{A}_{ji}$,¹⁸ and the solution vector $X_i^+ = X_i$. However, in the case of the volume based PSD, the vector $X_i^+ = X_i^v$ (in the case where smoothing is not required,¹⁹ then the volume based PSD X_i^v is obtained from an objective function similar to f_1 in Eq. (17) as described in Section 3 of the Supplementary Information, otherwise Eq. (21) is used), the matrix A_{ji}^+ will be scaled accordingly (see Section 3 of the Supplementary Information for details), and the vector $C_j^+ = \hat{C}^*$ where the transformed CLD \hat{C}^* is calculated as

$$\hat{C}_j^* = \sum_{i=1}^N \tilde{A}_{ji} \hat{X}_i, \quad (22)$$

where

$$\hat{X}_i = \frac{X_i}{\sum_{i=1}^N X_i}. \quad (23)$$

The operator ∇_h^2 is a finite difference approximation to the second derivative of the vector X_i^+ given as (Veldman and Rinzema, 1992)²⁰

$$\nabla_h^2 = \frac{h_- X_{i+1}^+ - (h_+ + h_-) X_i^+ + h^+ X_{i-1}^+}{\frac{1}{2} h^+ h_- (h_+ + h_-)}, \quad (24)$$

where

$$h_- = \bar{L}_i - \bar{L}_{i-1}, h_+ = \bar{L}_{i+1} - \bar{L}_i \quad (25)$$

and X_i^+ has been treated as a function of the characteristic particle size \bar{L} . The parameter λ_2 sets the level of penalty imposed on the second derivative of X_i^+ . If the value of λ_2 is sufficiently large, then

¹⁸ The matrix \tilde{A}_{ij} is the transformation matrix in the forward problem in Eq. (10) initially estimated with the objective function f_1 in Eq. (17). The smoothed solution is then calculated using the function f_3 in Eq. (21) at a unique aspect ratio determined using the function f_2 in Eq. (20). The solution vector X_i from Eq. (17) is used to construct a trial solution as $\gamma_i = \ln(X_i)$ in the case of the number based PSD. For the volume based PSD, the corresponding number based PSD from Eq. (21) is used to construct a trial solution as described in Section 3 of the supplementary information.

¹⁹ The vector X_i^+ is defined as an exponential function of a fitting parameter similar to Eq. (18) for X_i . The optimisation is then performed to obtain the optimum value of the fitting parameter using the LM algorithm similar to the case of X_i in Eq. (18).

²⁰ The form of the central difference approximation to the second derivative of the vector X_i^+ given in Eq. (18) is necessary since the grid for \bar{L} is non-uniform as seen in Eq. (8).

the penalty on the second derivative causes the LM algorithm to search for a solution vector which is smooth thereby avoiding solutions with localised oscillations (see Section 4.4 of the Supplementary Information).

The volume based PSD obtained from Eq. (21) is normalised and converted to a probability density distribution as

$$\bar{X}_i^v = \frac{\bar{X}_i^v}{(L_{i+1} - L_i) \sum_{i=1}^N \bar{X}_i^v}. \quad (26)$$

8. Results and discussion

The results obtained with the two methods outlined in Sections 6 and 7 are presented in this Section. More details of the choice of parameter values are presented in the supplementary information.

8.1. Results from Method 1

Fig. 8(b) shows the objective function f_2 as a function of the aspect ratio r for PS. The function reaches a minimum at $r=1$ suggesting spherical particles. This is consistent with the shape of the particles in Fig. 2(a) and the mean aspect ratio of $\bar{r} \approx 0.8$ obtained from the shape descriptor in Fig. 5(d) for this sample. This is also in agreement with the histogram in Fig. 6(d) which suggests that the majority of the particles in the sample are near spherical. Hence the aspect ratio predicted with Method 1 gives a reasonable description of the shape of the particles in the population as previously established (Agimelen et al., 2015).

The calculation in Fig. 8(b) was done with $\lambda_1 \approx 0.95$ in Eq. (20) (see Sections 4.3 and 4.4 of the Supplementary Information for details on how the values of λ_1 and λ_2 are chosen in Eqs. (20) and (21)) using the optimum transformation matrix from Eq. (17). Using this optimum transformation matrix and $r=1$, a number based PSD is calculated from Eq. (21) with the smoothness penalty set by $\lambda_2 = 10^{-2}$. The CLD corresponding to this number based PSD is shown by the solid line in Fig. 8(a). Furthermore, the volume based PSD (calculated at $r=1$) is shown in Fig. 8(c). In this case, the volume based PSD was calculated from the objective function f_2 (with the CLD C_j^* replaced with transformed CLD \hat{C}_j^* and the matrix \tilde{A}_{ji} rescaled as described in Section 3 of the Supplementary Information) at $\lambda_1 \approx 0.95$. The objective function f_3 was not used in calculating the volume based PSD in this case as smoothing was not required.

The calculated CLD in Fig. 8(a) has a near perfect match with the measured CLD for PS which is shown by the symbols in Fig. 8(a). The calculations in Fig. 8 were done at a value of $\mathcal{N} = 2$ (where \mathcal{N} is defined in Eq. (12)). This value was sufficient to give a wide enough range of aspect ratios to find a good match to the experimentally measured CLD. If the value of \mathcal{N} is not large enough, then the calculated CLD may not match the experimentally measured CLD as the particles do not have exactly the same shape and Method 1 only uses a single aspect ratio to describe the shape of all the particles in the population. The single aspect ratio chosen will then not be representative of all the particles in the population. However, the imaging data narrows down the search space for a representative aspect ratio, and hence reduces the risk of predicting an unreliable aspect ratio.

Figs. 9 and 10 are similar to Fig. 8 but for COA and glycine respectively. Fig. 9(b) shows the objective function f_2 (in Eq. (20)) with aspect ratio for COA. The function f_2 in Fig. 9(b) predicts an aspect ratio $r=0.3$ for COA. This is reasonable when compared with crystals in Fig. 2(b) and the shape descriptor in Fig. 5(b). Also the mode of the histogram in Fig. 6(e) is close to the aspect ratio $r=0.3$. The predicted aspect ratio of $r=0.3$ in Fig. 9(b) is also close

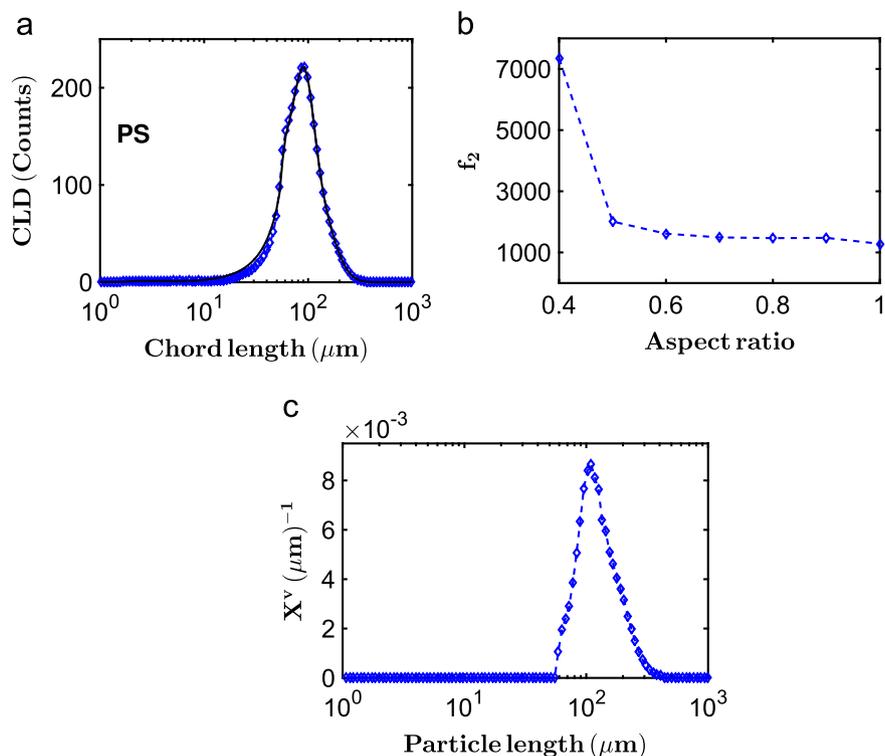


Fig. 8. (a) Experimentally measured (symbols) and calculated (solid line) CLD for PS. The calculated CLD was obtained by solving the forward problem in Eq. (10) using the number based PSD calculated with the objective function f_3 (using $\lambda_2 = 10^{-2}$) in Eq. (21). The calculation was done at the aspect ratio $r=1$ where the objective function f_2 (in Eq. (20) using $\lambda_1 \approx 0.95$) reaches a minimum. The matrix \mathbf{A} in Eq. (10) was calculated by Method 1 (described Section 6.1). (b) The objective function f_2 (using $\lambda_1 \approx 0.95$) in Eq. (20) for different aspect ratios for PS. (c) Calculated (by Method 1) volume based PSD for the PS sample. The volume based PSD was calculated using the objective function f_2 (at $\lambda_1 \approx 0.95$). The objective function f_3 was not used in the calculation of the volume based PSD in this case as smoothing was not required.

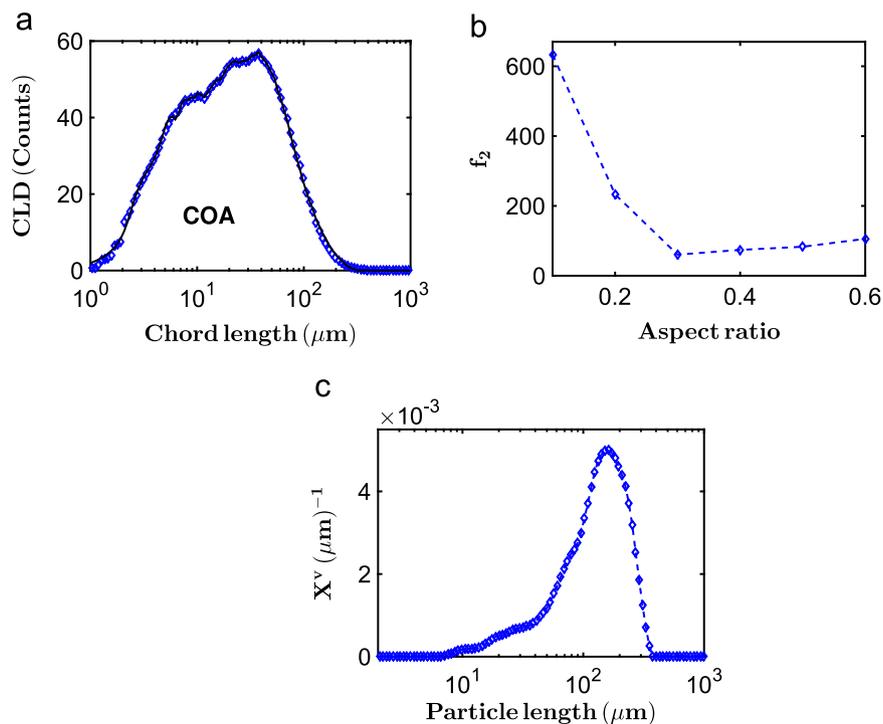


Fig. 9. Similar to Fig. 8 but for COA. The calculation was done at the aspect ratio $r=0.3$ where the objective function f_2 (using $\lambda_1 = 0.54$) reaches a minimum in (b). The number based PSD (used for calculating the CLD) was calculated using $\lambda_2 = 0.05$ in Eq. (21) while the volume based PSD (shown in (c)) was calculated using $\lambda_2 = 10^{-7}$ in Eq. (21).

to the estimated aspect ratio of $r \approx 0.2$ from the shape descriptor in Fig. 5(b). Furthermore, the calculated (in a manner similar to the case of PS in Fig. 8(a)) CLD for COA shown by the solid line in Fig. 9 (a) has a near perfect match with the measured CLD for the sample

shown by the symbols in Fig. 9(a). The calculations were done with $\mathcal{N} = 4$ in Eq. (12).

The volume based PSD for COA (calculated in a manner similar to the case of PS in Fig. 8(c)) is shown in Fig. 9(c). The calculated

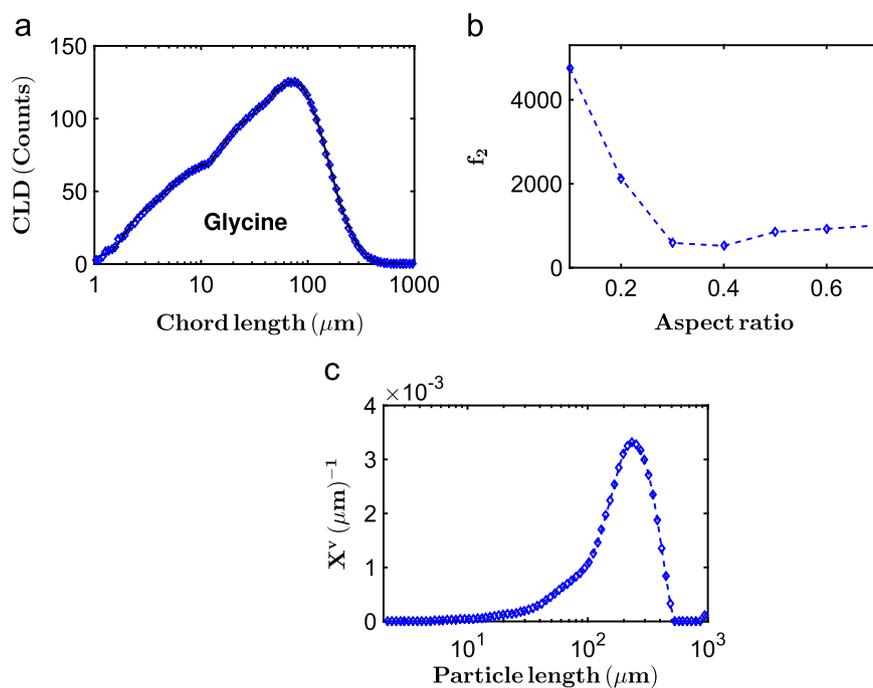


Fig. 10. Similar to Fig. 9 but for glycine with calculations done at $r=0.4$. The values of $\lambda_1 = 0.41$ and $\lambda_2 = 0.01$ were used in Eqs. (20) and (21) respectively for the number based PSD. The value of $\lambda_2 = 10^{-6}$ was used in Eq. (21) for the volume based PSD.

PSD in Fig. 9(c) has a left shoulder extending to needle lengths of about $10\ \mu\text{m}$. This gives a hint of the presence of a significant number of short needles in the COA sample. Some of these short needles can be seen in the image of Fig. 2(b).

The calculations for glycine in Fig. 10 are similar to the cases of PS and COA in Figs. 8 and 9 respectively. The predicted particle shape represented by $r=0.4$ in Fig. 10(b) is consistent with the particles in Fig. 2(c) and shape descriptor (which yields $r \approx 0.4$) in Fig. 5(c) as well as the histogram in Fig. 6(f). The calculated CLD (solid line Fig. 10(a)) also matches the experimentally measured CLD for the glycine sample (symbols in Fig. 10(a)). The calculated volume based PSD for this sample at $r=0.4$ is shown in Fig. 10(c). The volume based PSD in Fig. 10(c) also has a left shoulder extending to about $10\ \mu\text{m}$ similar to the case of COA in Fig. 9(c). The calculations in Fig. 10 for glycine were done with $\mathcal{N} = 2$ which was sufficient to get a good match for the measured CLD.

8.2. Results from Method 2

The aspect ratios in Fig. 6(b) and (c) show a spread over a significant range of particle sizes. Hence the technique referred to as Method 2 in Sections 6.2 and 7.2 was also applied in the analysis of the data from COA and glycine.

The solid line in Fig. 11(a) shows the calculated CLD for COA using Method 2. The calculation was done by searching for the optimum values of L_{min} and L_{max} while constructing different transformation matrices as outlined in Section 6.2. The search for the optimum values of L_{min} and L_{max} (and hence the optimum transformation matrix) is done by minimising the objective function f_1 in Eq. (17). Once the optimum transformation matrix is found, then a number based PSD is calculated by minimising the objective function f_3 (with $\lambda_2 = 0.1$) in Eq. (21). The CLD corresponding to this number based PSD is shown by the solid line in Fig. 11(a). The transformed CLD \hat{C}_j^* in Eq. (22) is calculated using the optimum transformation matrix and the number based PSD obtained with Eq. (17).

The calculated CLD in Fig. 11(a) (solid line) has a near perfect match with the experimentally measured CLD (shown by the

symbols in Fig. 11(a)) for COA. This is similar to the situation in Fig. 9(a) where the calculation was done with Method 1. The degree of agreement of the calculated CLD in Fig. 11(a) with the experimentally measured CLD demonstrates the level of accuracy that can be achieved with Method 2. Note that the aspect ratios of each of the subgroups of each bin (in Fig. 7) were assigned equal weights; a simple approach that is sufficient for reasonable results in this case. The volume based PSD for COA (obtained by Method 2) suggests particle sizes from about $3\ \mu\text{m}$ to about $400\ \mu\text{m}$ (Fig. 11(b)). This is close to the prediction of particle sizes from about $7\ \mu\text{m}$ to about $400\ \mu\text{m}$ by Method 1. Even though the ranges of particle sizes predicted by both methods are close, Method 2 has the advantage that aspect ratio is not used as a fitting parameter which removes the issue of estimating the optimum aspect ratio from the problem. The aspect ratio is assumed to vary according to imaging data available.

Although the particle sizes estimated from these 2D images are not very accurate because of the focusing problem highlighted earlier, a comparison of the estimated PSD from the images with the volume based PSD obtained by both methods can still be made. This comparison shows good agreement of the estimated volume based PSD from the images with the volume based PSD estimated by Methods 1 and 2. Details are given in Section 6 of the Supplementary Information. The peak of the volume based PSD obtained by Method 2 is higher than that obtained by Method 1 in this case. This is because the volume based PSD by Method 2 is slightly narrower within the size range of about $50\ \mu\text{m}$ to about $200\ \mu\text{m}$ (Fig. 11(b)) so that the main peak gets higher to satisfy the normalisation constraint in Eq. (26). The main peak is accompanied by a smaller peak at a particle length close to $30\ \mu\text{m}$ (Fig. 11(b)) suggesting a bimodal distribution for the COA particles. However, this feature of a bimodal distribution is not picked up by Method 1 (Fig. 11(b)). This could be because Method 2 is more efficient in picking out bimodal distributions in a population of particles where there is a variation of aspect ratio for particles of different sizes (see Section 7 of the Supplementary Information for details) than Method 1.

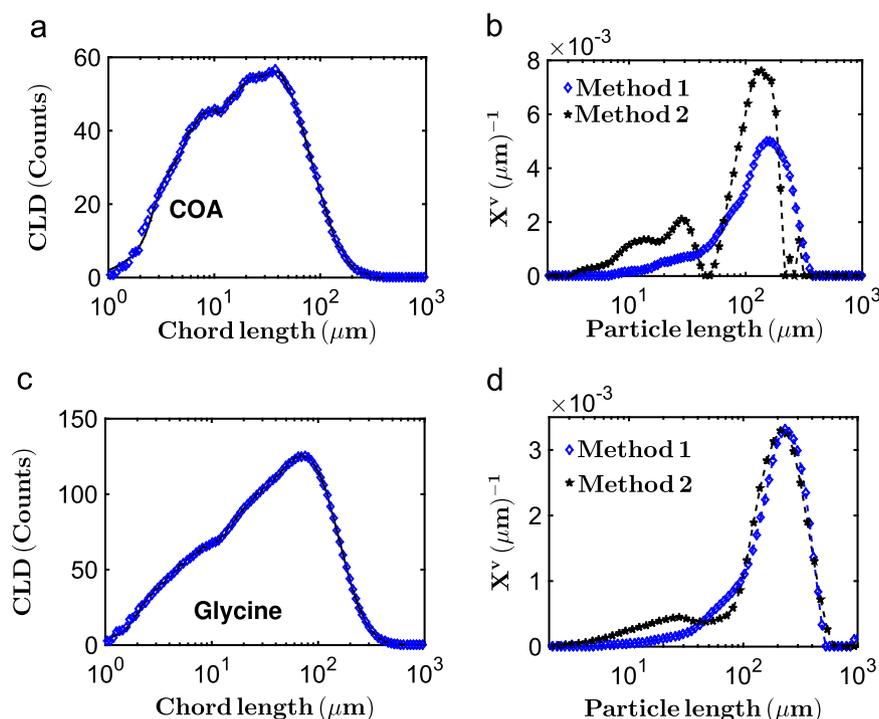


Fig. 11. (a) Experimentally measured (symbols) and calculated (solid line) for COA. The calculated CLD was obtained by solving the forward problem in Eq. (10), where the matrix $\hat{\mathbf{A}}$ was calculated by Method 2 as outlined in Section 6.2. The number based PSD used in solving the forward problem was obtained from the objective function f_3 in Eq. (21) for $\lambda_2 = 0.1$. (b) The blue diamonds are the calculated (by Method 1) volume based PSD for COA shown in Fig. 9(c). The black asterisks are the volume based PSD calculated by Method 2 for COA. The volume based PSD by Method 2 was calculated from Eq. (21) at $\lambda_2 = 3 \times 10^{-6}$. (c) Similar to (a) but for glycine. The number based PSD was obtained from Eq. (21) at $\lambda_2 = 0.01$. (d) Similar to (b) but for glycine, where the value of $\lambda_2 = 10^{-5}$ has been used for the volume based PSD calculated by Method 2. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

The situation for glycine is similar to that of COA. The solid line in Fig. 11(c) shows the calculated (calculated in a manner similar to the case of COA in Fig. 11(a)) CLD for glycine. The calculated CLD in Fig. 11(c) also has a near perfect match with the experimentally measured (symbols) CLD in Fig. 11(c). This is similar to the case of COA in Fig. 11(a). The calculated volume based PSD (by Method 2) in Fig. 11(d) for glycine also covers about the same range of particle sizes as in the case of Method 1 (Fig. 11(d)) with the PSD very similar from both methods.

The volume based PSD for COA obtained by the two Methods (Fig. 11(b)) cover a size range of $\leq 10 \mu\text{m}$ to about $400 \mu\text{m}$, while the CLD data for COA (in Fig. 11(a)) shows a maximum chord length of about $300 \mu\text{m}$.²¹ The PSD in Fig. 11(b) agrees with the image data in Fig. 6(b) for COA where the scatter plot is dense in the region between about $30 \mu\text{m}$ and about $300 \mu\text{m}$ with a small number of particles of sizes $\geq 300 \mu\text{m}$. Particles of small sizes below about $30 \mu\text{m}$ are not picked up by the image processing algorithm because objects smaller than that are rejected by the algorithm to reduce the risk of processing background noise as real objects. This is the reason why particles of sizes $\leq 30 \mu\text{m}$ do not contribute to the scatter plot of Fig. 6(b) even though the volume based PSD for COA in Fig. 11(b) suggests the presence of these particles.

The situation with glycine is similar to that of COA as seen in Fig. 11(d). The volume based PSD obtained by both methods cover a size range from about $10 \mu\text{m}$ to about $500 \mu\text{m}$ in agreement with the CLD for glycine (in Fig. 11(c)) which shows the longest chord to be about $400 \mu\text{m}$. The data in Fig. 11(d) also agrees with the image data in Fig. 6(c) which shows particle sizes up to about $500 \mu\text{m}$. There may be a larger number of large particles (of sizes close to

$500 \mu\text{m}$) in the glycine slurry than in the COA slurry so that their contribution to the CLD is more significant.

9. Conclusions

Two different methods have been developed to constrain the search space of aspect ratio(s) for particle size estimation using CLD and imaging data. Both methods estimate aspect ratio from images and then use the information in the estimation of aspect ratio and/or PSD from CLD data.

In the first method, the PSD estimation from CLD data is carried out using a single representative aspect ratio for all the particles in the slurry. However, the search space for this representative aspect ratio is reduced by means of data from the images of the particles captured in-line during the process. This reduces the risk of predicting an aspect ratio which is not representative of the particles in the slurry, and hence an unreliable PSD.

In the second method, a range of aspect ratios (also estimated from the images of the particles captured in-line) is assigned to particles of different sizes. This takes aspect ratio estimation out of the problem, and hence eliminates the risk of estimating a PSD at an aspect ratio which is not representative of the particles in the slurry.

The techniques presented in this work have been developed to be applied in situ, and an in-line imaging tool has been used in this work. The currently available in-line imaging tools are not suitable (when used on their own) for obtaining accurate PSD and aspect ratio due to various issues outlined in the text. The limitations of using images from these in-line tools alone to get aspect ratio and/or PSD estimates also show up in the large error bars in Fig. 14(d)–(f) in Section 4.6 of the Supplementary Information. Hence the methods presented here combine imaging and CLD data obtained in-line to obtain more robust estimates of PSD and aspect

²¹ Note that CLD is number based and therefore much less sensitive to presence of a small number of large particles.

ratio. Note that the methods presented here can be applied to combine CLD with imaging captured with any in situ tools. The images need to be of sufficient quality so that aspect ratio information can be obtained from them using a suitable image processing algorithm.

Acknowledgements

This work was performed within the UK EPSRC funded project (EP/K014250/1) 'Intelligent Decision Support and Control Technologies for Continuous Manufacturing and Crystallisation of Pharmaceuticals and Fine Chemicals' (ICT-CMAC). The authors would like to acknowledge financial support from EPSRC, AstraZeneca and GSK. The authors are also grateful for useful discussions with industrial partners from AstraZeneca, GSK, Mettler-Toledo, Perceptive Engineering and Process Systems Enterprise. The authors would also like to acknowledge discussions/suggestions from Alison Nordon and Jaelyn Dunn.

Appendix A. Supplementary data

Supplementary data associated with this paper can be found in the online version at <http://dx.doi.org/10.1016/j.ces.2016.01.007>.

References

- Agimelen, O.S., Hamilton, P., Haley, I., Nordon, A., Vasile, M., Sefcik, J., Mulholland, A. J., 2015. Estimation of particle size distribution and aspect ratio of non-spherical particles from chord length distribution. *Chem. Eng. Sci.* 123, 629–640.
- Barrett, P., Glennon, B., 2002. Characterizing the metastable zone width and solubility curve using lasentec FBRM and PVM. *Chem. Eng. Res. Des.* 80, 799–805.
- BSI, 2014. Particle Size Analysis-Image Analysis Methods, BS ISO 13322-1:2014, BSI Standards Publication.
- Eggers, J., Kempkes, M., Mazzotti, M., 2008. Measurement of size and shape distributions of particles through image analysis. *Chem. Eng. Sci.* 63, 5513–5521.
- Heinrich, J., Ulrich, J., 2012. Application of laser-backscattering instruments for in situ monitoring of crystallization processes—a review. *Chem. Eng. Technol.* 35 (6), 967–979.
- Hobbel, E.F., Davies, R., Rennie, F.W., Allen, T., Butler, L.E., Waters, E.R., Smith, J.T., Sylvester, R.W., 1991. Modern methods of on-line size analysis for process streams. *Part. Part. Syst. Character.* 8, 29–34.
- Izenman, A.J., 1991. Recent developments in nonparametric density estimation. *J. Am. Stat. Assoc.* 86 (413), 205–224.
- Kail, N., Briesen, H., Marquardt, W., 2007. Advanced geometrical modeling of focused beam reflectance measurements (FBRM). *Part. Part. Syst. Character.* 24, 184–192.
- Kail, N., Briesen, H., Marquardt, W., 2008. Analysis of FBRM measurements by means of a 3D optical model. *Powd. Technol.* 185, 211–222.
- Kail, N., Marquardt, W., Briesen, H., 2009. Estimation of particle size distributions from focused beam reflectance measurements based on an optical model. *Chem. Eng. Sci.* 64, 984–1000.
- Kempkes, M., Vetter, T., Mazzotti, M., 2010. Measurement of 3d particle size distributions by stereoscopic imaging. *Chem. Eng. Sci.* 65, 1362–1373.
- Li, M., Wilkinson, D., 2005. Determination of non-spherical particle size distribution from chord length measurements. Part 1: theoretical analysis. *Chem. Eng. Sci.* 60, 3251–3265.
- Li, M., Wilkinson, D., Patchigolla, K., 2005. Determination of non-spherical particle size distribution from chord length measurements. Part 2: experimental validation. *Chem. Eng. Sci.* 60, 4992–5003.
- Li, H., Grover, M.A., Kawajiri, Y., Rousseau, R.W., 2013. Development of an empirical method relating crystal size distributions and FBRM measurements. *Chem. Eng. Sci.* 89, 142–151.
- Li, H., Kawajiri, Y., Grover, M.A., Rousseau, R.W., 2014. Application of an empirical FBRM model to estimate crystal size distributions in batch crystallization. *Cryst. Growth Des.* 14, 607–616.
- Roths, T., Marth, M., Weese, J., Honerkamp, J., 2001. A generalized regularization method for nonlinear ill-posed problems enhanced for nonlinear regularization terms. *Comput. Phys. Commun.* 139, 279–296.
- Ruf, A., Worlitschek, J., Mazzotti, M., 2000. Modeling and experimental analysis of PSD measurements through FBRM. *Part. Part. Syst. Character.* 17, 167–179.
- Schorsch, S., Vetter, T., Mazzotti, M., 2012. Measuring multidimensional particle size distributions during crystallization. *Chem. Eng. Sci.* 77, 130–142.
- Schorsch, S., Ochsenbein, D.R., Vetter, T., Morari, M., Mazzotti, M., 2014. High accuracy online measurements of multidimensional particle size distributions during crystallization. *Chem. Eng. Sci.* 105, 155–168.
- Simmons, M.J.H., Langston, P.A., Burbidge, A.S., 1999. Particle and droplet size analysis from chord distributions. *Powd. Technol.* 102, 75–83.
- Vaccaro, A., Sefcik, J., Morbidelli, M., 2007. Modeling focused beam reflectance measurement and its application to sizing of particles of variable shape. *Part. Part. Syst. Character.* 23, 360–373.
- Veldman, A.E.P., Rinzema, K., 1992. Playing with nonuniform grids. *J. Eng. Math.* 26, 119–130.
- Washington, C., 1992. Particle Size Analysis in Pharmaceuticals and Other Industries. Ellis Horwood Limited, Chichester, England.
- Worlitschek, J., Mazzotti, M., 2003. Choice of the focal point position using Lasentec FBRM. *Part. Part. Syst. Character.* 20, 12–17.
- Worlitschek, J., Hocker, T., Mazzotti, M., 2005. Restoration of PSD from chord length distribution data using the method of projections onto convex sets. *Part. Part. Syst. Character.* 22, 81–98.
- Wynn, E.J.W., 2003. Relationship between particle-size and chord-length distributions in focused beam reflectance measurement: stability of direct inversion and weighting. *Powd. Technol.* 133, 125–133.
- Yu, Z.Q., Chow, P.S., Tan, R.B.H., 2008. Interpretation of focused beam reflectance measurement (FBRM) data via simulated crystallization. *Org. Process Res. Dev.* 12, 646–654.