Alternative approach for defining the particle population requirements for static image analysis based particle characterization methods

Clarke, James; Gamble, John F.; Jones, John W.; Tobyn, Mike; Greenwood, Richard; Ingram, Andy

DOI:
10.1016/j.apt.2019.02.006

Document Version
Peer reviewed version

Citation for published version (Harvard):
Clarke, J, Gamble, JF, Jones, JW, Tobyn, M, Greenwood, R & Ingram, A 2019, 'Alternative approach for defining the particle population requirements for static image analysis based particle characterization methods', *Advanced Powder Technology*, vol. 30, no. 5, pp. 920-929. https://doi.org/10.1016/j.apt.2019.02.006

+Title page

Alternative approach for defining the particle population requirements for static image analysis based particle characterization methods

James Clarke[1], John F. Gamble[2*], John W. Jones[2], Mike Tobyn[2], Richard Greenwood[1], Andy Ingram[1]

[1] School of Chemical Engineering, University of Birmingham, Birmingham, B15 2TT, UK

[2] Bristol-Myers Squibb, Reeds Lane, Moreton, Wirral, CH46 1QW, UK

[*] Corresponding Author:

John F. Gamble

Bristol-Myers Squibb

Reeds Lane, Moreton, Wirral, UK.

Phone: +44 (0)151 552 1646

E-mail: john.gamble@bms.com

Declarations of interest: none

ABSTRACT

For image based particle characterisation approaches one of the most common discussion points is determining the number of particles required to have statistical confidence that the measurement is able to adequately describe the distribution of the sample. This topic becomes significantly more challenging when applied to the extraction of single component size distributions from multi-component samples.

The aim of this work was to propose a means to accurately assess the particle number requirements using a method specific approach. The method applies a sub-sampling method to the original imaged dataset in order to provide an understanding of the impact of sub-sampling on the ability of accurately reproduce the original distribution.

The method was applied to understand the particle number requirements for two batches of theophylline with varied particle size distributions, using the input size distribution to guide the requirements for the subsequent multi-component samples of both materials.

The results demonstrate the utility of the method to determine the appropriate number particles required to recreate the size distributions. Whilst the minimum number of particles required to be sampled can be calculated, how those particles are sampled can also affect the validity of the measurement and must be taken into consideration.

KEYWORDS

Particle size, particle shape, image analysis, sampling, spectroscopy

1. INTRODUCTION

For image based particle characterisation approaches, one of the most common discussion points is determining the number of particles required to have statistical confidence that the measurement is able to adequately describe the distribution of the sample [1-4]. There have been a number of reported approaches [1, 2, 5] which have culminated into an ISO guidance for static image analysis [6] which is based upon utilising the distribution width in order to provide an estimation of the required number of particles.

Whilst such approaches are useful in providing guidance to the number of particles required for statistical confidence, without some prior thought about the true nature of the sample, the analysis method utilised and the underlying assumptions made in the associated calculations, the results can be misleading [7].

Firstly, the calculations assume a perfect log-normal distribution [5, 6]. This may be true, in particular for highly milled materials, however, many materials do not demonstrate idealised log-normal distributions [8]. Such distributions can range from skewed (positive or negative) to multi-modal distributions, and consequently the predicted number of particles can be prone to varying degrees of error.

Additionally, the absolute number of particles does not necessarily correlate with confidence of accuracy and/or precision [7]. Many dynamic imaging systems introduce the sample by means of a cycling loop; for such systems the number of particles measured may not reflect the number of

independent particles passing the optics but the number of cycles the particles make around the loop.

Static image analysis systems often utilise a multi-lens configuration in order to enable selection of a particular size range in line with the particles being characterised. However, such approaches can lead to confusion as to the relevance of the particle number requirements [7, 9-11].

Consider the analysis of a sample with a wide particle size distribution; if we measure the sample using a high resolution lens we can expect to measure a high number of particles as the lens will be sensitive to the fine particles, of which there are generally high numbers. However, care needs to be taken to ensure that the field of view is large enough to capture the coarse particles and the analysis area large enough to capture sufficient number of such particles. If the distribution is wide, this approach can lead to a method which has low precision due to a high sensitivity to the sampling of a small population of high volume particles even with high particle counts. A means to solve this precision issue is to utilise a lower resolution lens which would enable a larger area of the sample to be analysed therefore the sensitivity to the high volume particles reduces. This approach, however, would also lower the sensitivity of the method to the fine particles and so we now have a situation where the number of particles counted could be lower, but the precision and accuracy greater. Clearly, particle number alone can be misleading and so some thought needs to be applied to what is being measured, for what purpose and by which means [7].

This all becomes notably more complex when dealing with the characterisation of individual components within multi-component systems. In

recent work [8, 12, 13] the particle size and shape distributions of individual components within multi-component pharmaceutical systems has been extracted in order to monitor the change in the physical nature of formulated drug particles during manufacture. Such approaches utilise spectroscopic tools to chemically classify the particles thus enabling extraction of the characteristics of the individual components. Similar work is being reported in a wide range of fields including characterisation of forensic samples, including counterfeit drugs, gunshot residues and soil samples [14, 15], nasal sprays [16], environmental microplastics [17] and therapeutic antibody products [18] to name but a few.

One challenge with the approach is that the analysis time for such methods can be high (in the order of days) and the resulting particle counts notably lower than for the comparable single component analysis. Therefore, assuming appropriate sampling from the bulk powder, the question of how many particles required to provide statistical confidence that a particle size distribution can be accurately described becomes significantly more important. This concern was highlighted by Kippax et al. [16] noting that “the results tend to be relatively subjective with low statistical significance, pragmatism limiting the number of particles that can be measured”.

The aim of this work is to propose a means to accurately assess the sensitivity of a measurement to particle number for individual samples using a method specific approach. The approach is explained and applied to batches of a test material, theophylline anhydrous, with varied particle size distributions both before and after incorporation into a formulated matrix. The impact of replicate analysis on the measurement outcome is also evaluated.

2. MATERIALS AND METHODS

2.1 MATERIALS

Two batches of theophylline anhydrous (Sigma Aldrich, Gillingham, UK), batch A (MKBV6764V) and batch B (MKCC0719), were used in this study. The materials were also incorporated into a formulation containing microcrystalline cellulose (Avicel PH102) (FMC Corp., Philadelphia, PA) and magnesium stearate (Mallinckrodt Inc., Philipsberg, NJ).

2.2 Blend Preparation

Theophylline was incorporated into a formulation as a model active pharmaceutical ingredient (API) containing microcrystalline cellulose and magnesium stearate; the weight percentages of the three components were 50 %w/w, 49.5 %w/w and 0.5 %w/w, respectively. Theophylline and microcrystalline cellulose were first blended for 150 rotations (10 minutes at 15 rpm) in a 20 L intermediate bulk container using an MB100 tumble blender (Pharmatech, Coleshill, UK). The magnesium stearate was passed through a 1 mm aperture screen prior to addition to the blend and the material was then blended for a further 75 rotations (5 minutes at 15 rpm).

2.3 Image based particle characterisation (Single component)

Particle size analysis for each batch was determined using a Malvern Morphologi G3-ID particle characterisation system (Malvern Instruments Limited, Malvern, UK). Samples were dry dispersed (air pressure of 2 bar, injection time of 20 ms and a settling time of 180 s) using the systems automated sample dispersion unit onto a glass plate. Verification that the air

dispersion did not change the size and/or shape of the particles was conducted by dispersing the particles in octane and the suspension pipetted onto a microscope slide and left to air dry. Both methods provided equivalent data for both size and shape and as such the more rapid and simpler dry dispersion approach was utilised thereafter.

Particle Imaging was conducted using an x5 magnification lens (6.5 – 420 µm) with z-stacking enabled to take 3 planes above the initial point of focus (equivalent to 147 µm) to account for 3-dimensionality within the sample. Morphological filtering was applied to the raw image data in order to remove partially imaged/overlapping particles using a combination of convexity and solidity filters[19].

For the initial API samples the largest possible scan area was selected in order to maximise the number of particles imaged and ensure the particle count was more than sufficient to adequately describe the sample distributions.

2.4 Image based particle characterisation (Multi-component samples)

Characterisation of the theophylline size distribution after incorporation into the formulation was conducted using a Morphologi G3-ID particle characterisation system (Malvern Instruments Limited, Malvern, UK), an integrated static image analyser with Raman spectroscopic capabilities.

The formulated samples were dry dispersed (air pressure of 2 bar, injection time of 20 ms and a settling time of 180 s) using the systems automated sample dispersion unit onto a glass plate and image analyse

conducted using the same method conditions as the initial theophylline samples, albeit, a smaller scan area and a single additional plane during z-stacking (equivalent to 48.9 µm) was used to limit spectral variation due to focal depth. Morphological filtering was applied to the raw image data in order to remove partially imaged/overlapping particles using a combination of convexity and solidity filters.

Raman analysis of the particles was conducted using the integrated Kaiser Raman system (Kaiser Optical Systems Inc, Ann Arbor, MI) with a 785nm laser with a 3 µm spot size.

For the analysis of single components within multi-component samples, consideration of the aims of the analysis must be considered prior to analysis to ensure that the approach is suitable for purpose. As previously highlighted, the analysis time of these experiments are high and the particle sample size lower than typically utilised for image based characterisation methods. The samples described in this work were to be utilised to investigate their propensity to undergo process induced attrition during manufacture in subsequent work.

As particle size is generally reported in terms of the geometric (volume weighted) distribution, the results are more sensitive to large volume particles. These large particles are also most prone to attrition[20]. As the number of particles within samples typically reduces as the particle size increases, smaller sampling populations can lead to non-representative sampling of the high volume particles.

Replicate analysis can be utilised to increase the particle count, however, the requirement for replicate analyses will greatly impact the time taken to analyse a series of samples. A second option is to remove particles below a particular size threshold; this approach removes high numbers of low volume particles thus reducing the risk of under-sampling high volume particles whilst having minimal effect on the overall geometric distribution.

Combinations of the two strategies have been utilised in previous work in this area, but the key question is how many particles are required to accurately describe the distribution and how does the removal of fine particles affect that number. This was an additional focus of the reported approach, as was the impact of replicate sampling strategies.

For each analysis run, 10,000 particles were randomly selected for Raman analysis by the system. The selection of particles was conducted by equally spaced sampling of particles throughout the dataset; the particles in the dataset were sequenced in the order they were imaged from a randomly dispersed sample plate thus providing a randomised sampling protocol.

For the Raman analysis, particles were analysed for 10 s (two sequential 5s scans) over a range of 150 – 1850 cm^{-1}. The measured spectra for each particle was compared to a reference library spectra and the correlation coefficient used to classify the individual components into chemical classes thus enabling particle size distributions of individual component within the formulated sample to be extracted.

Replicate data sets were generated and combined using a MATLAB script and evaluated for their geometric D[v,0.5] values. To aid comparison, the values were normalised with respect to the initial theophylline size.

2.5 Scanning Electron Microscopy

Samples were sputter coated using a JFC-1300 auto fine coater (Jeol Inc, MA, USA) and then imaged using a Neoscope JCM-500 (Jeol Inc, MA, USA) using an accelerating current of 10-15 kV.

2.6 Estimation of Sample Size in Accordance with ISO Guidelines

The ISO guidelines [6, 21] are based on the approach reported by Masuda and Iinoya [22] in order to determine the required sample size to represent a distribution. This method assumes a log-normal distribution and is based upon the number weighted distribution of the population.

The logarithmic mean diameter for the number distribution of the population and the population standard deviation are calculated from equations 1 and 2 respectively.

$$\mu^0 = \int_{-\infty}^{\infty} \ln D_p f(\ln D_p, \mu^0, \sigma^2) d \ln D_p = \ln D_{[n,0.5]} \quad (1)$$

$$\sigma = \ln \sigma_g = \ln D_{[n,0.84]} - \ln D_{[n,0.5]} \quad (2)$$

where μ^0 is the logarithmic mean diameter for the arithmetic (number weighted) distribution of the population, D_p is the particle diameter, σ is the standard deviation of the population, σ_g is the geometric standard deviation and D[n,0.5] and D[n,0.84] are the diameters below which 50% and 84% of the number of particles reside, respectively.

In this method the number of particles required to describe a distribution can be calculated using the following equation:

$$\log n^* = -2\log\delta + \log\omega \quad (3)$$

where ω is a parameter described by:

$$\omega \equiv u^2\alpha^2\sigma^2(2c^2\sigma^2 + 1) \quad (4)$$

where n* is the number of particles required in a measurement, δ is the relative error, u is a parameter dependent upon P, the probability that the experimental data may be in the range of relative error, –δ to +δ and c is described by:

$$c = \left(\frac{\alpha}{2}\right) + \beta \quad (5)$$

where α is a constant and β is the basis number; this is equal to 0 for a count basis and 3 for a volume basis.

When determining the Mass Median Diameter (MMD), the parameter ω is calculated using equation 6:

$$\omega = 36u^2\sigma^2(18\sigma^2 + 1) \quad (6)$$

In this work a relative error value of 0.05 and P value of 0.95 were used, with the corresponding value of u being 1.96. The number of particles required was calculated for each method of calculating the parameter ω for both batches of theophylline.

2.7 Proposed method background

A method was developed to randomly sample particles from a previously analysed size distribution dataset in order to determine the number of particles required to describe the original distribution. The approach was designed remove the requirement to assume a log-normal distribution.

In essence, the approach is a modified bootstrap method [23, 24]; such methods work on the basis that without any other knowledge about a population, the distribution values found in a random sample of size *n* from the population is the best guide to the distribution of the population distribution of values in a random sample [25]. Therefore, to assess the impact of re-sampling of the population, it is sensible to re-sample the sample [26]. The proposed approach utilises an adapted, non-replacement bootstrap methodology.

The Morphologi measures multiple size and shape characteristics for each particle within a sample, the details of which can be extracted in a readable data format for external data analysis. A MATLAB (Mathworks) script was developed to take this data, remove particles below a selected circle equivalent (CE) diameter and then, using an in-built function to randomly sample a given number of particles from this filtered sub-set, create a new sub-set of data. The randomly sampled data sub-set was then used to determine statistical parameters (i.e. D[v,0.5], D[v,0.9] etc.); normalization of the results to that of the initial dataset were typically utilized to provide a scale independent measure of accuracy / precision. The random sampling, modelling and comparison steps were repeated for a number of iterations to assess the expected precision.

In this report several particle population sizes were tested. For assessments of the impact of CE diameter thresholding the original data set for each batch was subjected to the following CE diameter filters: 3, 5, 10 and 25 µm prior to sub-sampling. The sub-set results were then compared to that

of the equivalently constrained initial dataset. The sampling step was repeated six times for each condition tested.

3. Results

3.1 Characterisation of theophylline batches

The geometric particle size distributions (PSD) for both batches of theophylline (Figure 1, upper plot) demonstrate that the distribution for batch A (#MKBV6764V) is narrower, containing a higher relative volume percentage of coarse particles, than that of batch B (#MKCC0719V).

While batch A's geometric PSD maybe more coarse and uniform than that of batch B the underlying arithmetic distribution is very different (Figure 1, lower plot). Batch A was observed to have a skewed/bimodal distribution containing a large number of fines in addition to a smaller population of coarse particles, whereas the corresponding arithmetic distribution for batch B was observed to be more uniform in shape than batch A, containing a larger population of particles in the 5-40 µm size range. The observations from the image analysis were corroborated by the corresponding SEM images (Figure 2).

Based on this data, one could suggest that whilst batch A has a narrower geometric distribution, the high degree of skew observed in the underlying arithmetic distribution would suggest that the number of particles that would need to be taken in order to adequately sample the coarse population would be higher than one may expect based purely on the geometric data. In other words, as the bulk of the volume of the sample is contained in a small population of high volume particles but the vast majority

of particles in that sample are notably finer, i.e. contain little volume and as such will be less impactful with respect to ‘accurately’ describing the volume weighted size distribution, the statistical chance of sampling the high volume particles is generally low.

In contrast, the wide geometric distribution for batch B would initially suggest a larger sampling requirement than batch A, but the underlying arithmetic data suggests that the particles are log-normally distributed which could reduce the sampling requirements.

To investigate this further, the particle dataset was sub-classified into 3 sub-classes: ‘Fine’ (< 30 µm), ‘Median’ (≥30<µm>60) and ‘Coarse’ (≥ 60 µm). This would provide an initial means to better compare the arithmetic and geometric nature of the two input API lots.

Comparison of the arithmetic and geometric percentages within each of the pre-defined size classes for the two theophylline batches (Figure 3) demonstrates that the fine particles (< 30µm) account for the majority of the number of particles measured for both batches (88.4% and 88.8% respectively), however, for batch A this arithmetic majority accounts for just 0.8% of the total sample volume whereas for batch B the volume percentage is an order of magnitude higher (8%).

In the coarse size class (≥ 60 µm), it is observed that 97% of the volume is contained in just 6% of the particles arithmetically for batch A whereas the same size class constitutes just 76% of the volume for batch B , however, this is contained in just 2% of the particles arithmetically.

The results suggest that for batch A the coarse particle class, containing over 95% of the total sample volume, almost singularly describe

the geometric particle size distribution whereas for batch B, approximately 24% of the volume is distributed in the median and fine populations meaning sampling of these fractions is more 'consequential'. Whilst less volume is present in the coarse size class for batch B, the arithmetic sampling frequency of this class is much lower than batch A adding further sampling challenges for this sample.

The results clearly demonstrate that an understanding of the inter-relationship between the arithmetic and geometric distributions is required to fully elucidate the sampling requirements.

3.2 Application of ISO Guidance approach

The number of particles, n*, required to estimate the Mass Median Diameter (MMD), within a given degree of relative error and confidence limit (10% and 95% respectively), were calculated for both theophylline batches using the approach outlined in the ISO guidance (Table 1).

Table 1: Statistical parameters and estimated number of particles required to estimate (with a 10% relative error at a 95% confidence limit) the MMD for each batch

Material	Logarithmic Mean, $\mu^{(0)}$	Standard Deviation, σ ($\sigma=\ln \sigma_g$)	Geometric Standard Deviation, σ_g	n* (MMD)
Batch A	1.88	1.22	3.40	578,000
Batch B	2.45	0.79	2.20	105,000

Figure 4 compares the theoretical log-normal distribution as described by the logarithmic mean and standard deviation (Table 1) with the measured

distribution for each of the theophylline batches. The arithmetic PSD of batch B is described better by the log-normal distribution than that of batch A where almost half of the theoretical distribution resides below the measurement capabilities of the utilized system.

As a consequence, the calculated number of particles required to estimate the MMD (equivalent to D[v,0.5]) is significantly greater for batch A than batch B.

Clearly, whilst the approach does provide a useful guide to population requirements, the underlying assumption of a log-normal distribution can introduce inaccuracies in determining a representative measure of the number of particles required for a particular measurement approach.

3.3 Application of proposed method

The proposed method removes the assumption of normality requirement by repeatedly sub-sampling the original dataset in order to provide a sample/method specific handle on the number of particles required to accurately describe the original distribution with adequate precision. Figure 5 shows the results generated by the method for batches A and B; a normalization of the D[v,0.5] against the original dataset was utilized in order to enable comparison of materials of varying particle size.

The results show that as the sample size increases the normalised D[v,0.5] value approaches unity, indicating that the sampled sub-dataset closely matches that of the original, whilst the variation between the iterations decreases demonstrating increasing precision as a consequence of decreasing sensitivity of the measured result to sample size.

Comparing the results for the two batches, the data for batch A shows convergence to unity at a far lower sample size than batch B; the data suggests that for batch A an ‘accurate’ result could be achieved with as few as 1000 particles, however, to ensure greater precision a sample population of closer to 10,000 is required.

For batch B the story is a very different one; the low sample population of the coarse particles and the higher volume ratio in the finer classes means that the method predicts that much higher particle populations are required to achieve an ‘accurate’ result. The method suggests that a sample population of approximately 10,000 is required, whereas the result is prone to more variability below particle populations of approximately 50,000.

The results are closely aligned with the expectations based upon the initial characterization; the variation between iterations is much lower for batch A as it has a higher arithmetic percentage of coarse particles and lower arithmetic percentage of fine and mid-range particles than batch B. There is therefore a higher probability of the method sampling these coarse particles, which contain the majority of the volume. Correspondingly, for batch B the variability is higher due to the larger number of fine and mid-range particles, which also contain a higher proportion of the overall sample volume.

3.4 Application of the method to multi-component analysis approaches

One use of the method has been to support the characterization of single components within multi-component samples. In the approach utilized to date, a sub-set of randomly selected particles from the full imaged dataset

and chemically classified using Raman spectroscopy. The analysis takes up a lot of instrument time per sample and the particles of interest are ‘diluted’ in the formulated matrix; as a consequence, the API particle population sampled is generally much smaller than typically utilized for image characterization of single components. The importance of understanding the number of particles required to have confidence that the distribution is accurately determined, or that any changes are real and not due to sampling, is therefore critical.

Achieving the required number of particles may require numerous replicate runs and thereby significant instrument time. An alternative means of shifting the sampling statistics is to subtract the fine particle classes from the dataset; such size classes often constituting very high fraction of the particle count but very little in terms of the sample volume. By applying such filters the number of particles within the population required to be sampled to describe the size distribution can be significantly reduced, with minimal effect on the volume percentage of the mid-range and coarse particle size classes.

Such an approach does assume that the geometric distribution is of interest and that any changes of interest will be occurring to the higher volume particles, i.e. in the case of particle attrition, the approach could be utilized to track the change in size of the coarse/median particles (which are more prone to attrition) rather than the increase in fines as a consequence of the attrition.

Selection of a suitable threshold for the fines cut-off, and thereby ensuring that the selection statistics are shifted such that reduced number of particles can be utilized to accurately describe the size distribution whilst minimizing the percentage of the overall sample volume removed, is therefore critical to understand.

To assess the impact of such an approach the method was adapted to apply a range of size class thresholds to sample datasets prior to the random sampling stage. For this work, size classes below 3, 5, 10 and 20 µm were assessed and compared to the unrestrained (full) dataset. In all cases, the threshold applied to the subsets was also applied to the reference dataset. The effect of each size threshold applied on the arithmetic and geometric percentages for both batches is detailed in Table 2.

Table 2 Effect of CE particle size threshold implementation on particle number and volume

CE Diameter Filter Applied [µm]	Batch A			Batch B		
	Number of Particles	Numerical Reduction [%]	Volumetric Reduction [%]	Number of Particles	Numerical Reduction [%]	Volumetric Reduction [%]
0 (unconstrained)	135,000	-	-	93,000	-	-
3	110,000	18.8	0.002	89,000	5.0	0.002
5	82,000	39.3	0.008	79,000	15.4	0.02
10	48,000	64.4	0.06	53,000	43.1	0.3
20	24,000	82.0	0.3	23,000	74.9	2.9

The results for the arithmetic and geometric size classifications for both API batches (Figure 6) show how the filters affect the sampling statistics. As previously addressed, it would have been predicted that Batch A would be less sensitive to the size filtering as 97% of the volume is contained in the coarse particles and thus removal of fines particles would not greatly affect the volume percentages in the geometric distribution whilst the impact on

sampling would improve the expected precision at lower sampling populations.

The results clearly corroborate this demonstrating that the volume distribution of the sample is generally unaffected by any of the size class filters, however, the shift in the arithmetic frequency for the coarse particle class is observed to shift significantly from 6% up to approximately 35% for the 20 µm filter. This would suggest that for this batch, the 20 µm filter could be utilized without affecting the ability to describe the geometric size distribution. The particle number data from the method for this batch (Figure 7) clearly shows that the use of the filters greatly enhances the precision of the smaller population sizes suggesting a lower particle population could be applied.

In a similar manner, it could be predicted that batch B would be more sensitive to the size filtering as approximately 25% of its volume is contained in both the median and fine particle fractions. As with batch A, application of a size threshold was observed to have little impact on the volume percentage of each of the size classes, however, there was an approximate fourfold increase in the arithmetic frequency of the median and coarse particles when a 20 µm size threshold is applied. Accordingly, it can be seen that the whilst the precision does increase, due to the smaller population size, the relative precision at any given sample size was less than that which would be expected for batch A; even with a 20 µm size threshold the fine particle class still accounts for 55% of the number of particles in batch B and the coarse just 8%; the equivalent arithmetic frequencies for batch A were approximately 35% in both classes.

The results demonstrate that, for batch B, only the 20 µm size threshold has any significant effect on the arithmetic sampling frequency and consequently the precision of at any given sample size appears to remain unchanged from that of the unconstrained sample. For this sample it could be suggested that a higher threshold value could have been applied.

The approach enables a means to provide a sample/method appropriate estimation of the number of particles required to accurately represent the geometric size distribution whilst the variance between the iterations gives an indication of the expected precision due to the selected sample size.

The results demonstrate that by removing fine particles from the dataset the sample size required to accurately represent the distribution is significantly reduced whilst causing little change on the geometric distribution. Using this method it can be seen that when a 20 µm is applied the number of particles to represent the volume distribution with a degree of accuracy is in the range of 5,000 particles for batch A and 10,000 for batch B.

These values are order of magnitude lower than those calculated using the ISO guidance method; with respect to batch A, even without removal of fines, this approach would provide justification that 10,000 particles would be sufficient to provide statistical confidence of appropriateness and therefore having the consequence of vastly reducing the required analysis time when compared to the target population of 500,000 that the ISO guidance would suggest for the same sample. In addition it can clearly be seen that less particles are required accurately represent batch A than batch B, contrary to the values calculated using the ISO guidance method.

3.5 Verification of the method for multi-component analysis

To experimentally verify the prediction output from the method batch B was incorporated into a formulated blend. The multi-component sample was then analysed using the Morphologi G3-ID, an integrated imaged based characterization system with Raman spectroscopic capability, to size and classify the particles in terms of their chemistry. Based on the earlier method data a 20 µm size threshold was utilized for the analysis. Eight replicate samples were analysed in order to generate a large enough sample population; each individual replicate measured approximately 2,000 API particles.

In order to assess the inherent variability due to sampling approach, permutated combinations of the replicates were generated using the below equation:

$$No.of\ samples = \sum_{i=1}^{n-1} \frac{i!}{(n-i)!\,i!} + 1$$

where n is the number of measurements taken and i is the number of measurements, thus creating 225 different data sets of varying sample population.

These were normalised with respect of the initial data set (with a 20µm filter applied). Figure 8 shows the results; the dotted lines show the error limits due to sample size for a log-normal distribution which were calculated using the method described by Yoshida et al.[5] using the logarithmic mean and standard deviation of batch B.

It can clearly be seen that as the replicates are combined to create greater particle populations the variation between similar sized samples is reduced, with the variability within the expected limits. The results show a convergence of the normalised D[v,0.5] values at a particle count of approximately 13,000, however, the convergence occurs at a normalised value offset from unity.

Although the measured size is observed to converge at an value offset from unity, it can be seen that the number of particles at which the precision reaches an acceptable level is in good agreement with the earlier prediction of the method; the data converges at a sample size of 13,000 which is slightly higher than the 10,000 predicted by the method method, but still far lower than the 105,000 sample size calculated using the ISO guidance method (albeit for a unconstrained dataset).

Whilst the data validates the reduced particle count requirements, the presence of an offset for batch B does raise a question about the ‘accuracy’ of combining replicate samples below the particle requirement threshold. One possible reason for the observed offset could be the sampling approach utilised to generate the final datasets. The method suggests that randomly selecting approximately 10,000 particles is required to provide statistical confidence that the distribution can be recreated, however, how those particles are sampled may be just as important as the number itself.

In the current approach, the final particle population is a merged dataset collated from multiple (n was generally between 4 and 6) measurements of replicate dispersions, with approximately 2000-3000 API particles sampled from each dispersion all of which is then combined. For

batch A, the population size of each sample replicate is greater than the requirement for adequate sampling obtained from the method, consequently, good reproducibility is observed. For batch B, however, the approach assumes that whilst the API particle count for each individual replicate is below the required particle population sampling threshold suggested by the method, the summation the aliquots will provide an adequate sample population indicative of the bulk, but is this a correct assumption?

To test this, six randomly sampled particle populations, each consisting of exactly 2000 particles selected from the initial sample datasets for each of the batches were extracted; the 20 micron thresholded datasets were utilised for the purpose of this test in order to match the conditions for the original observations. In order to obtain an approximation of the expected variability for the multiple dispersion approach, these sub-sets were then combined into all possible non-repeating permutations, thereby creating five separate populations of 10,000 particles. Additionally, a single randomly selected population of 10,000 particles was sampled from the same original datasets, representing sampling from a single dispersion; this sampling approach was also replicated five times to represent the expected variability.

The results (Figure 9) provide verification that for batch B, where the sampling from each dispersion is below the required particle sampling size suggested by the earlier analysis, the sampling approach does have a notable impact on the final result. The data demonstrated that the variability of each dispersion is high and combination of the individual sub-samples did result in an over-estimation of the measured size, simulating the observations in the

real data. In comparison, utilising a single large sample was observed to correct the previously observed offset.

This anomaly was not observed for batch A as the particle population requirement was below the actual sample size for each individual dispersion sample thus the results demonstrate minimal variation. This data further strengthens the validity of the method output.

As a final verification, the above sampling method was applied to the blended sample of batch B in order to verify that the overall approach could extract a meaningful (equivalent) API particle size distribution from the multi-component sample. The results (Figure 10) demonstrate that the obtained volume weighted cumulative particle size distribution (~10,000 particles) was observed to match closely with that of the input API.

This work demonstrates that whilst a minimum number of particles required to be sampled can be calculated, how those particles are sampled can also affect the validity of the measurement and must be taken into consideration prior to analysis.

4. Conclusions

A method has been developed to estimate the number of particles required to describe geometric particle size distributions using static image based approaches. The approach removes the need for an assumption of log-normality and instead applies a sub-sampling method to the original imaged dataset in order to provide an understanding of the impact of sub-sampling on the ability of accurately reproduce the original distribution.

The application was also utilized to assess the impact of size thresholding samples on the arithmetic and geometric size distributions as a means to reduce particle sampling requirements for the utility in multi-component particle characterization approaches. The method was validated experimentally and it has been shown that the approximate number particles required to recreate the geometric distribution can be predicted.

The work also demonstrates that whilst a minimum number of particles required to be sampled can be calculated, how those particles are sampled can also affect the validity of the measurement and must be taken into consideration prior to analysis.

Acknowledgements

The authors would like to acknowledge the EPSRC for funding. The authors would also like to thank Ana Ferreira and Srini Sridharan (Bristol-Myers Squibb) for their support during this study.

References

[1] H. Masuda, K. Gotoh, Study on the sample size required for the estimation of mean particle diameter, Advanced Powder Technology, 10 (1999) 159-173.
[2] A.J. Paine, Error Estimates in the Sampling From Particle Size Distributions, Particle & Particle Systems Characterization, 10 (1993) 26-32.
[3] E. Vigneau, C. Loisel, M.F. Devaux, P. Cantoni, Number of particles for the determination of size distribution from microscopic images, Powder Technology, 107 (2000) 243-250.
[4] G. Heinicke, J.B. Schwartz, Particle size distributions of inert spheres and pelletized pharmaceutical products by image analysis, Pharmaceutical Development and Technology, 9 (2004) 359-367.
[5] H. Yoshida, Y. Mori, H. Masuda, T. Yamamoto, Particle size measurement of standard reference particle candidates and theoretical estimation of uncertainty region, Advanced Powder Technology, 20 (2009) 145-149.
[6] ISO 13322-1:2014 Image analysis methods Part 1: Static image analysis methods, in.
[7] J.F. Gamble, M. Tobyn, R. Hamey, Application of Image-Based Particle Size and Shape Characterization Systems in the Development of Small Molecule Pharmaceuticals, Journal of Pharmaceutical Sciences, (2015).
[8] J.F. Gamble, M. Hoffmann, H. Hughes, P. Hutchins, M. Tobyn, Monitoring process induced attrition of drug substance particles within formulated blends, International Journal of Pharmaceutics, 470 (2014) 77-87.
[9] S. Almeida-Prieto, J. Blanco-Méndez, F.J. Otero-Espinar, Microscopic image analysis techniques for the morphological characterization of pharmaceutical particles: Influence of process variables, Journal of Pharmaceutical Sciences, 95 (2006) 348-357.
[10] F. Podczeck, S.R. Rahman, J.M. Newton, Evaluation of a standardised procedure to assess the shape of pellets using image analysis, International Journal of Pharmaceutics, 192 (1999) 123-138.
[11] J.P. Zingerman, S.C. Mehta, J.M. Salter, G.W. Radebaugh, Validation of a computerized image analysis system for particle size determination. Pharmaceutical applications, International Journal of Pharmaceutics, 88 (1992) 303-312.
[12] M. Hoffmann, P.S. Wray, J.F. Gamble, M. Tobyn, Investigation into process-induced de-aggregation of cohesive micronised API particles, International Journal of Pharmaceutics, 493 (2015) 341-346.
[13] J.F. Gamble, A.B. Dennis, P. Hutchins, J.W. Jones, P. Musembi, M. Tobyn, Determination of process variables affecting drug particle attrition withinmulti-component blends during powder feed transmission, Pharmaceutical Development and Technology, 22 (2017) 904-909.
[14] B.W. Kammrath, A. Koutrakos, J. Castillo, C. Langley, D. Huck-Jones, Morphologically-directed Raman spectroscopy for forensic soil analysis, Forensic Science International, 285 (2018) e25-e33.
[15] B.W. Kammrath, A. Koutrakos, P.E. Leary, J. Castillo, J. Wolfgang, D. Huck-Jones, Morphologically directed Raman spectroscopic analysis of forensic samples, Spectroscopy (Santa Monica), 33 (2018) 46-53.
[16] P. Kippax, D. Huck, C. Levoguer, A. Virden, J. Suman, Characterising a nasal spray formulation from droplet to API particle size, Pharmaceutical Technology Europe, 23 (2011) 34-39.
[17] W.J. Shim, S.H. Hong, S.E. Eo, Identification methods in microplastic analysis: A review, Analytical Methods, 9 (2017) 1384-1391.
[18] N.N. Pardeshi, W. Qi, K. Dahl, L. Caplan, J.F. Carpenter, Microparticles and Nanoparticles Delivered in Intravenous Saline and in an Intravenous Solution of a Therapeutic Antibody Product, Journal of Pharmaceutical Sciences, 106 (2017) 511-520.
[19] J.F. Gamble, W.S. Chiu, M. Tobyn, Investigation into the impact of sub-populations of agglomerates on the particle size distribution and flow properties of conventional microcrystalline cellulose grades, Pharmaceutical Development and Technology, 16 (2011) 542-548.
[20] K. Kendall, The impossibility of comminuting small particles by compression, Nature, 272 (1978) 710-711.
[21] ISO 9276-6:2008 Representation of results of particle size analysis Part 6: Descriptive and quantitative representation of particle shape and morphology.
[22] H. Masuda, K. Iinoya, Theoretical study of the scatter of experimental data due to particle-size-distribution, Journal of Chemical Engineering of Japan, 4 (1971) 60-66.
[23] T. Mattfeldt, A brief introduction to computer-intensive methods, with a view towards applications in spatial statistics and stereology, Journal of Microscopy, 242 (2011) 1-9.
[24] T. Matsuyama, An application of bootstrap method for analysis of particle size distribution, Advanced Powder Technology, 29 (2018) 1404-1408.

[25] A.P. Ferreira, T.P. Alves, J.C. Menezes, Monitoring complex media fermentations with near-infrared spectroscopy: Comparison of different variable selection methods, Biotechnology and Bioengineering, 91 (2005) 474-481.
[26] B.F.J. Manly, Randomization, bootstrap and Monte Carlo methods in biology, in, Chapman & Hall, London, 1997, pp. 399.

FIGURES

Figure 1 Volume (top) and number (bottom) weighted particle size distributions of two batches of theophylline. Batch A (Dashed line) and Batch B (Solid line)

Figure 2 SEM images taken of batches A (left) and B (right)

Figure 3 Comparison of the number (top) and volume (bottom) weighted frequencies in manually defined size classes

Figure 4 Arithmetic and fitted log-normal distributions for batch A (left) and batch B (right)

Figure 5 Normalised D[v,0.5] versus particle count for batches A (left) and B (right)

Figure 6 Effects of filtering on the number and volume percentage of different size classes for each batch

Figure 7 Normalised D[v,0.5] values versus particle count for batches A (top) and B (bottom) with different applied size constraints

Figure 8 Normalised D[v,0.5] values for combinations of replicates of varying particle count

Figure 9 Comparison of measured particle size using differing sampling populations and sampling regimes

Figure 10 Comparison of volume weighted cumulative particle size distribution plots for batch B (Solid line = Input API size (unconstrained); Dot/Dash line = Input API size (CE>20 µm constraint); Dashed line = API extracted from blended samples)