BLENDING OF LOG NORMAL PARTICLE SIZE DISTRIBUTION DATA FROM MULTIPLE IMAGE ANALYSES INTO A SINGLE CONTINUOUS DATA SET

As published in JVT, Winter 2011 (Vol17/No1)

Eric Olson

ABSTRACT

Image analysis is one method by which particle size may be measured. In many cases, it is considered the "gold standard" of particle size measurement and is often used to verify the results of other techniques. Sample particle size distributions are often broad, sometimes covering two to three decades of particle diameter. It is also common to analyze samples that are multi-modal (e.g., a single sample that may contain primary particles and hard agglomerates with overlapping particle size distributions). Typical results when analyzing these samples are frequently inaccurate at the particle size extremes, depending on the instrument operating parameters used. This paper describes a methodology by which data from multiple image analyses that use different operational parameters are exported from the instrument into a suitable spreadsheet. The data are then blended into a useful continuous particle size distribution with optimal accuracy.

INTRODUCTION

There are many methods of measuring particle size, one of which is by image analysis. The process of particle size measurement by image analysis is very well established and accepted, as proven by the number of International Organization for Standardization (ISO), ASTM, and other standards on this methodology. It is often the measurement to which other particle size methodologies are compared and even calibrated. Image analysis is particularly useful for non-spherical particles such as fibers and low-aspect ratio particles such as needles. This is because two of the main underlying theories in laser diffraction, Fraunhofer and Mie, are only applicable to solid spheres. Thus, the further the sample particle deviates from a solid sphere, the greater the expected error in measurement by laser diffraction. Image analysis is not governed by either Fraunhofer or Mie theory, thus, it is not subject to the same particle shape constraints.

A sample particle size distribution should ideally be monomodal and narrow. In reality, this is rarely the case. Sample particle size distributions are often broad, sometimes covering two to three decades of particle diameter. It is also common to analyze samples that are multi-modal (e.g., a single sample that may contain primary particles and hard agglomerates with overlapping particle size distributions).

Each objective of the image analysis system has an optimal range based on its resolution. The resolution is determined by the wavelength of the incident light and the numerical aperture of the lens. This means that at times, if a sample particle size distribution is very broad or multi-modal across a broad range, it may be necessary to collect data using more than one magnification level. This leads to the existence of two separate data sets, often of varying population sizes, which characterize two or more sections of the true sample particle size distribution. In order to make the two sections of data useful and continuous, a process in which the sections may be blended or combined is needed.

This discussion describes a process by which data from multiple image analyses are exported from the analytical instrument into a suitable spreadsheet such as Microsoft Excel (1). These data are then be blended into a continuous particle size distribution that provides maximum accuracy.

Image Analysis Basics

The components of a static image analysis system are very basic in concept. A magnification system typically utilizes a compound light microscope. The light source is most often diascopic light (light from below or

transmitted light); in some systems, an episcopic light (light from above or reflected light) may also be available. The detector is often either a digital camera or a (charge coupled device) CCD. The system must also have some type of sample holder such as a microscope stage. In some automated systems, there is also a dispersing system for dry powders and/or some type of wet cell capable of measuring suspended powders in a liquid dispersant.

The process begins with the sample preparation. Before automated systems, techniques were developed to disperse dry powders on a microscope slide or similar sample holder. Modern automated instruments utilize a burst of compressed air to disperse fine particles onto a sample holder. Ideally the dispersion separates the loose agglomerated material without breaking any particulate. It is also desirable for the individual particles to be cleanly separated and not touching or overlapping. Similarly for a dry sample dispersed in a liquid medium, loose agglomerates should be separated, but neither fractured nor dissolved. Ideally the sample should be contained on a sample holder with minimum depth of field and minimum particle overlap.

Once the samples are properly dispersed, the next step is to determine the proper magnification level. Again, there are trade-offs associated with this step. If too low a magnification level is chosen, small particles may go unobserved and the shape of the large particles may be skewed as a result of the detector resolution. If too high a magnification level is chosen, too few particles may be analyzed per ISO 13322 (2), leading to a biased result. Also, unless some type of automated process such as z-stacking is available, a large particle under high magnification will most often be out of focus, also leading to biased results.

If the stage is considered to reside in the x-y plane, then the z-axis would be perpendicular to the stage. Z-stacking is a process by which the stage is raised and/or lowered a finite number of steps over preset distances. For instance, if the middle of a large particle is considered 0 on the z-axis, a z-stack of -1/+1 may be employed. Thus, for every field of view, the stage would move to the middle of the particle (z = 0) as well as some predefined distance below the middle (-1) and the same predefined distance above the middle (+1). This predefined distance is generally quite small and on the order of a few microns, but is ultimately limited by the accuracy and reproducibility of the stage motor. This allows the instrument to capture in-focus images of small particles that may reside closer to the mounting surface as well as large particles whose surface may be a few microns above that of the small particles.

At some point, the microscope objective and detector system must be calibrated. This is most often accomplished by obtaining an image of a traceable stage micrometer. Once the image is obtained, the user or software is able to calculate the calibration ratio of pixels to length. It is important this be done with a traceable stage micrometer to aid in the ease of validation. It is also imperative that calibration be performed at the start of each analysis.

The next step in the process is to adjust the light intensity and particle threshold. On modern automated systems, the light intensity may be measured and adjusted automatically. Images of the dispersed fine particles are captured and digitized into pixels. The threshold, if properly set, determines the pixel intensity at which the particle boundary or edge is defined. A field of particles is then observed and the threshold value set such that fine particles of interest are not excluded, and particle images are neither dilated nor eroded.

Once the digitized particle images are captured, many software packages automatically use the calibrated pixel maps to determine a wide array of particle measurements (e.g., length, width, area, circular equivalent diameter, aspect ratio, circularity, elongation, convexity, perimeter, solidity, etc.).

EXPERIMENTAL

Sample Preparation

The sample used in this example was passed through a 250 µm sieve prior to analysis to remove extraneous pieces of matter unrelated to the particles of interest. A small portion of the example particulate sample that passed through the sieve was manually sprinkled on a microscope slide and observed at various magnifications using a Malvern

Morphologi G3S automated image analysis system. The sample at first glance appeared to have a very broad particle size distribution that was possibly multi-modal. The largest particles observed appeared to have a diameter of about 200 µm, while the smallest particles observed appeared to have a diameter near 2 µm.

Because of the two-decade difference in observed mean particle diameters, two objective magnification levels were chosen for the analysis, 2.5X and 10X. The 2.5 objective has a working range from about 4 µm to 1000 µm and the 10X objective has a working range from about 1 µm to 13 µm. This combination of objectives has an overlap from about 4 µm to 13 µm, a range of particle diameters in which they could be observed and blended.

Data Capture, Export, and Import

Particle data were collected using a suitable standard operating procedure for the instrument. The procedure utilized the sample dispersion unit (SDU) of the Malvern Morphologi G3S to ensure the particles were adequately dispersed with minimal overlap. The number of particles analyzed with the 2.5X objective was about 17,000, and the number of particles analyzed with the 10X objective was about 41,000. The number of particles to analyze is a function of the statistical confidence level and the width of the distribution. It was imperative that a sufficient number of particles be analyzed. This ensured the distributions were well developed and representative of the actual population.

Once the particle diameter data were captured, they were exported to a text document and subsequently into Microsoft Excel. One of the limitations of Microsoft Excel is that a spreadsheet may contain only 65,536 rows. This limits the maximum number of particles that can be imported in the spreadsheet directly from the text document. If a larger number of particles had been analyzed and the data subsequently exported to a text document, they would first be opened in a spreadsheet that has a higher limit of rows than Microsoft Excel such as JMP 8 (3), and the data copied and pasted into multiple columns within a Microsoft Excel spreadsheet.

Building the Histogram

Once the data are imported into Microsoft Excel, a frequency histogram can be easily generated using the Data Analysis - Histogram function which is part of the program functionality. However, before a histogram can be generated, a set of bin sizes (i.e., minimum particle diameters) must be established. In general, almost all particle size distributions are represented on a logarithmic scale. In order to generate a Gaussian particle size distribution on a logarithmic scale, the x-axis must be a geometric progression. If the geometric progression were set with a large multiplier, the bin sizes would be too broad and the system would lose resolution. However, if the multiplier were set too small, the bin sizes would be very narrow, limiting the number of particles in each bin and effectively reducing the signal to noise ratio.

In this example, the geometric multiplier was set to the value that most accurately represents the native reported bin sizes of the instrument. The default on the Malvern Morphologi G3S was 1000 bins, reported across the range 0.1 µm to 2000 µm. The multiplication factor of the geometric progression was 1.009963.

Number and Volume Weighted Statistics

In the field of particle size analysis, there are many ways to express a mean result. Two of the most common are the number mean and volume moment mean. Mathematically, these are calculated from the histogram data as follows:

$$D[1,0] = \frac{\sum n_i d_i}{\sum n_i}$$

$$D[4,3] = \frac{\sum n_i d_i^4}{\sum n_i d_i^3}$$

D[1,0] = number weighted mean
D[4,3] = volume moment mean
n = frequency of particles in each bin
d = particle diameter or bin size.

The next step in the blending process was to calculate all the values of $n_i d_i$, $n_i d_i^3$, and $n_i d_i^4$. In this example case, the volume-weighted data were more applicable to the end-user than the number weighted data. However, the overall process has been utilized with other samples on a number-weighted basis as well.

RESULTS AND DISCUSSION

Data Smoothing and Overlay Plot

A plot of the volume-weighted differential raw histogram data over the range of the blending region is illustrated in Figures 1 and 2, which was very jagged and somewhat difficult to interpret. Figure 1 shows the raw volume-weighted differential plot over the full range of the analysis. Inset in Figure 1 is box indicating the approximate region over which the blending process occurs. Figure 2 shows the raw volume-weighted differential plot, zoomed in to the region over which the blending process occurs.

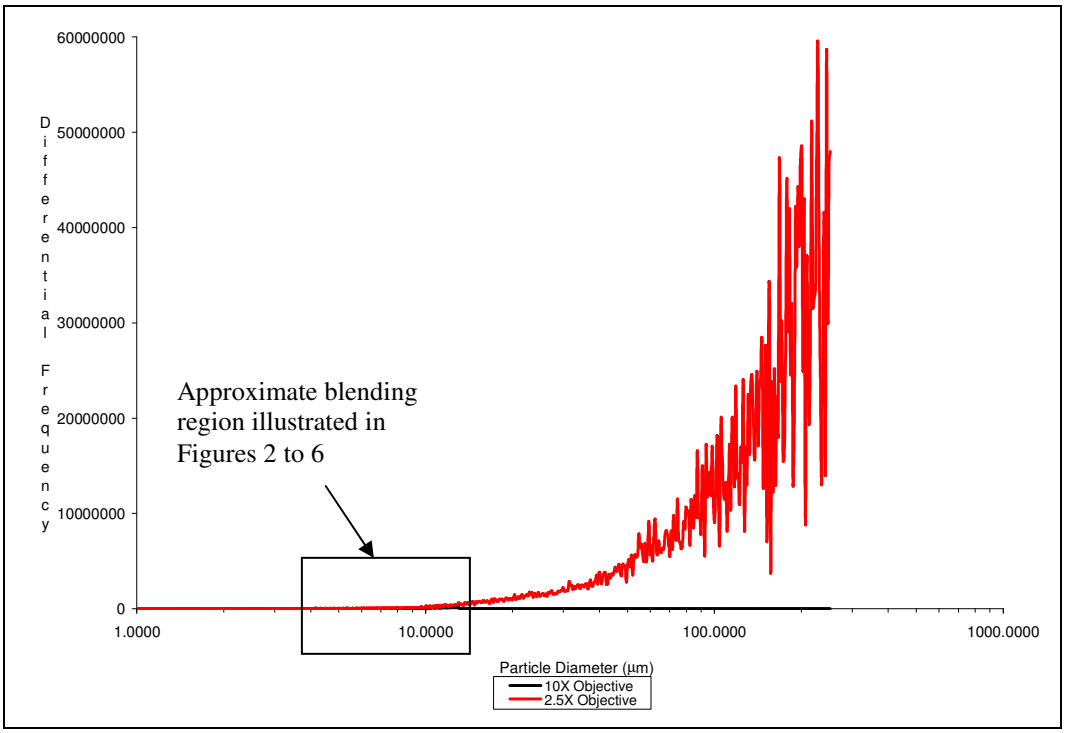

Figure 1. Volume-weighted differential particle diameter - raw eata - full range

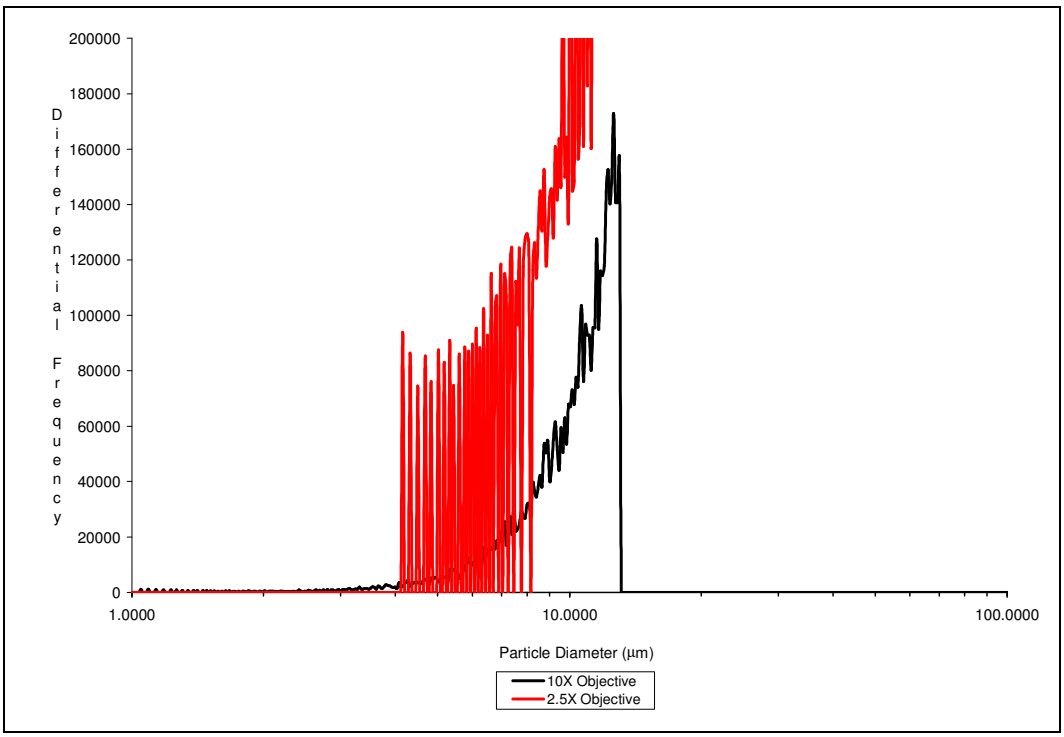

Figure 2. Volume-weighted differential particle diameter - raw data – blending region

One possible solution that has been implemented was to use smoothing or apodization filters. Of the numerous filters available, the two that have been proven the most useful are the moving average and Savitzky-Golay (4).

Plots of the volume-weighted differential histogram data after a 5-point and 11-point moving average over the same range are shown in Figures 3 and 4, respectively. Comparison of Figures 2, 3, and 4 indicate that as the number of points in the moving average increases, the plot of the volume-weighted differential particle diameter becomes smoother.

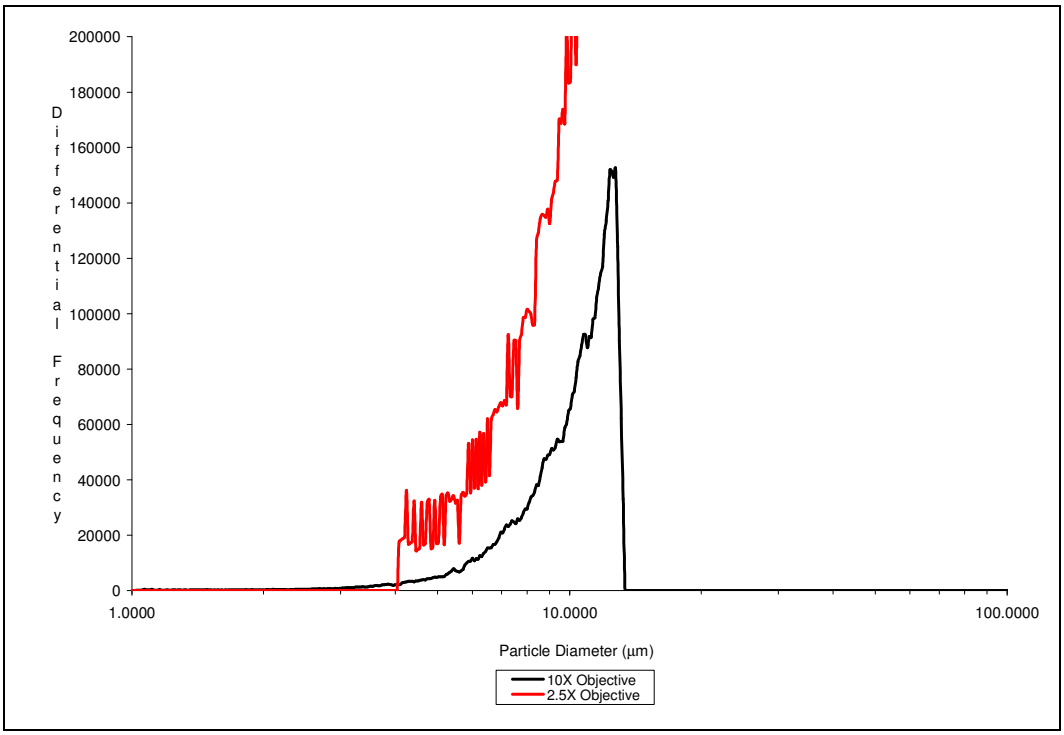

Figure 3. Volume-weighted differential particle diameter - 5-point moving average – blending region

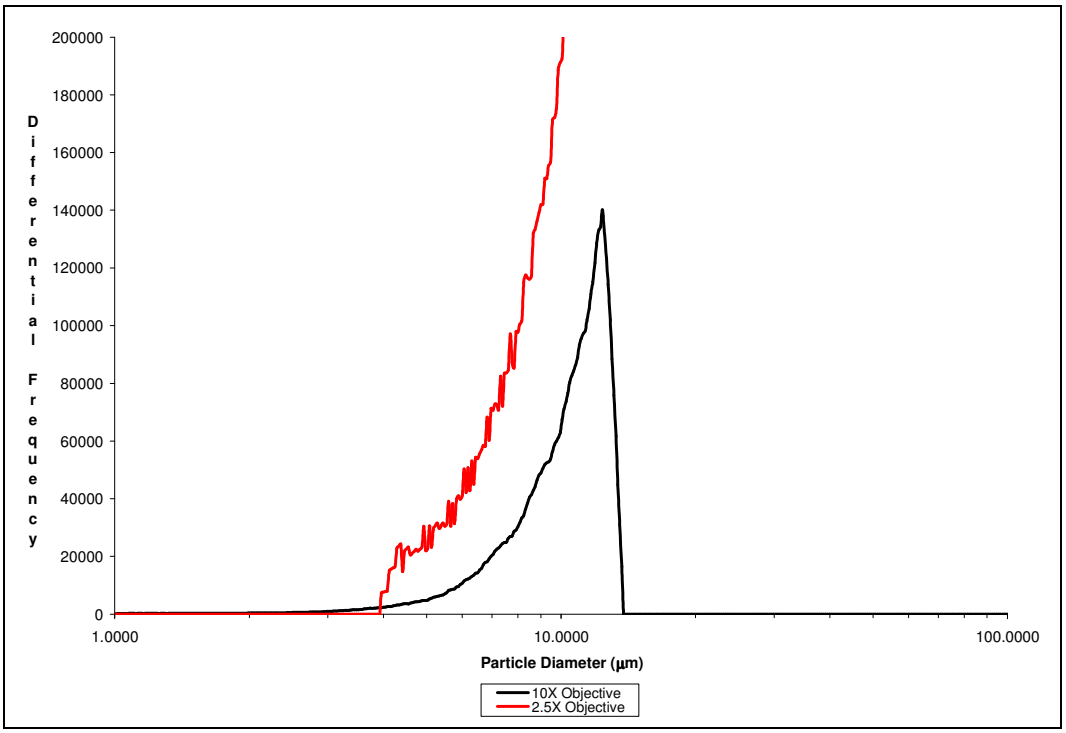

Figure 4. Volume-weighted differential particle diameter - 11-point moving average – blending region

Shown here are only the plots of the raw, 5-point, and 11-point moving average smoothed data for illustrative purposes. However, calculations were also completed for 7-, 9-, 13-, and 17-point filters and the corresponding Savitzky-Golay filtered data as well.

Blending

In this example case, the high magnification (small particle) data effectively supplement the much more broad and encompassing low magnification (large particle) data. This was accomplished by blending on the upward slope of the small particle tail of the low magnification (large particle) data. However, this was not the ideal situation. In general, it is best to blend data across a range with minimal change such as when the first derivative approaches zero (i.e., a peak in the distribution).

The ordinate data of the low magnification (large particle) data set was multiplied by a normalization constant. The normalization constant was determined by minimizing the sum of the squared differences between the two data sets over the designated blending range. Concurrently, the blending range was optimized by minimizing the sum of the squared differences between the maximum high magnification (small particle) point and the minimum of the low magnification (large particle) point. The maximum and minimum points were found by taking the first derivative of the respective data sets. Both sums of squared differences were determined using the Solver Tool in Microsoft Excel.

In this example case, the calculated blending range was 4.4 to 12.0 µm, which utilized 101 bins. That was nearly the entire range of overlap between the 2.5X and 10X objectives. The normalization constant was determined to be 0.33258. The plot of the normalized data sets is shown in Figure 5.

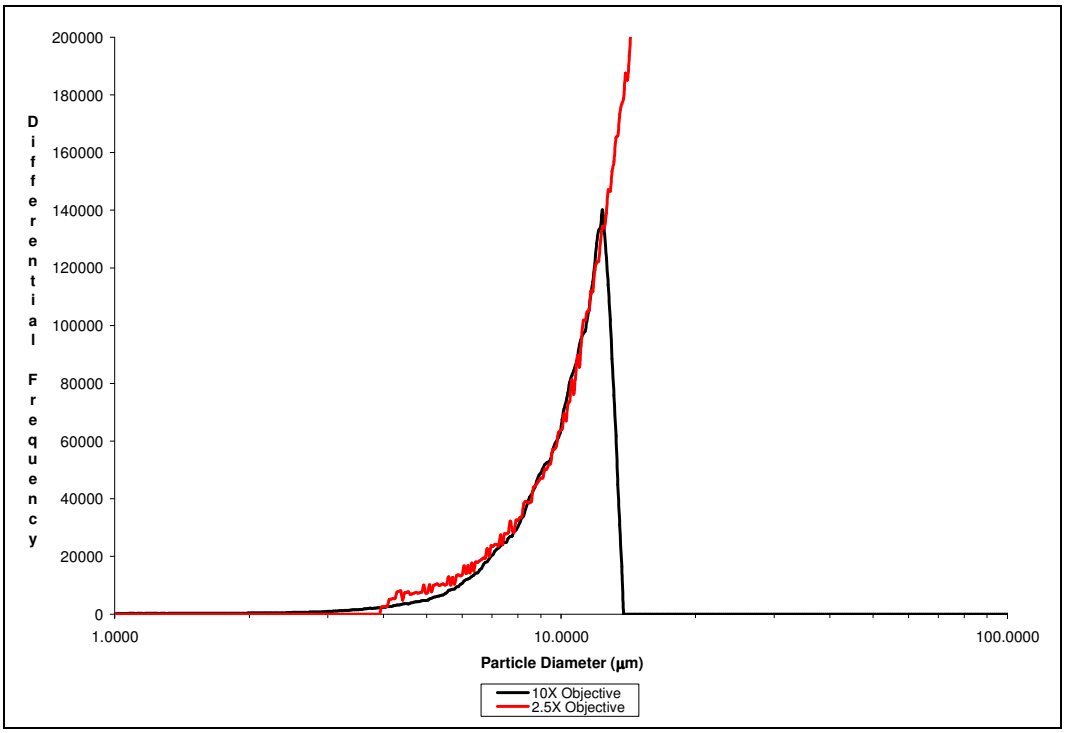

Figure 5. Volume-weighted differential particle diameter - 11-point moving average after normalization – blending region

Plot and Calculate Final Results

Once the blending region has been established and the normalization constant calculated, the final differential and cumulative plots may be created and all the summary statistics calculated. This was begun by blending the two normalized data sets. This was accomplished using the following logic:

If	Then
$x < x_{\text{blending region}}$	$y = y_{\text{high magnification}}$
$x = x_{\text{blending region}}$	$y = (y_{\text{high magnification}} + y_{\text{low magnification}}) / 2$
$x > x_{\text{blending region}}$	$y = y_{\text{low magnification}}$

A plot of the blended data for this example case is demonstrated in Figure 6. A plot of the same data, now showing the entire range is shown in Figure 7. A cumulative % curve was also generated and is shown in Figure 8. Various statistics as given in Table I may be calculated once the data are in a convenient spreadsheet form.

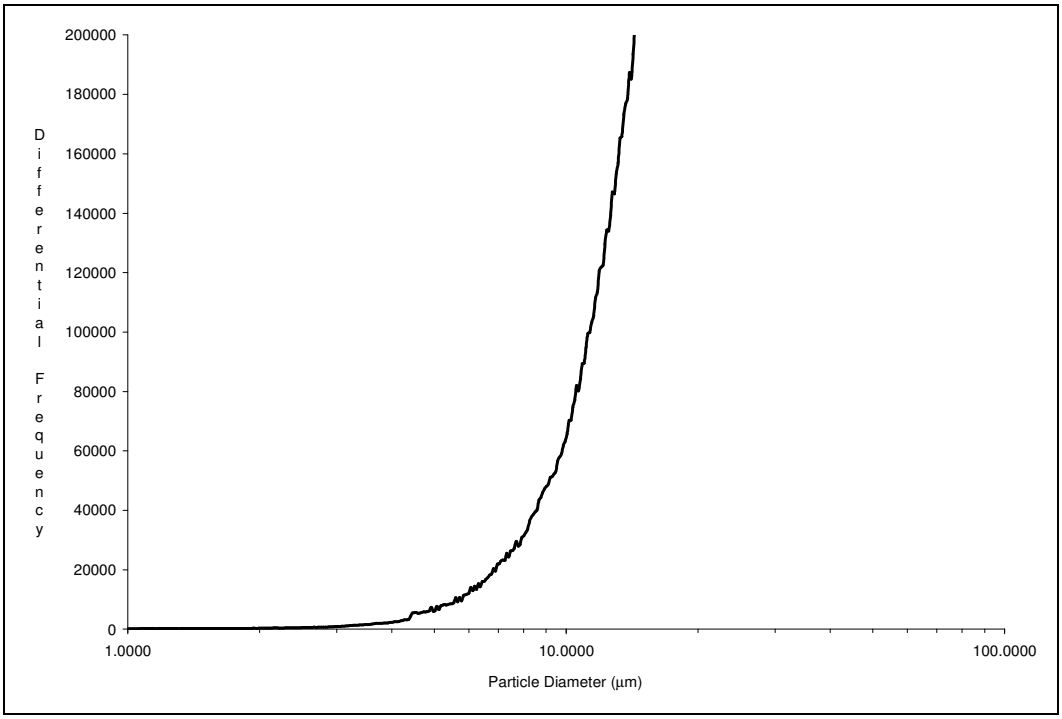

Figure 6. Volume-weighted differential particle diameter after blending – blending region

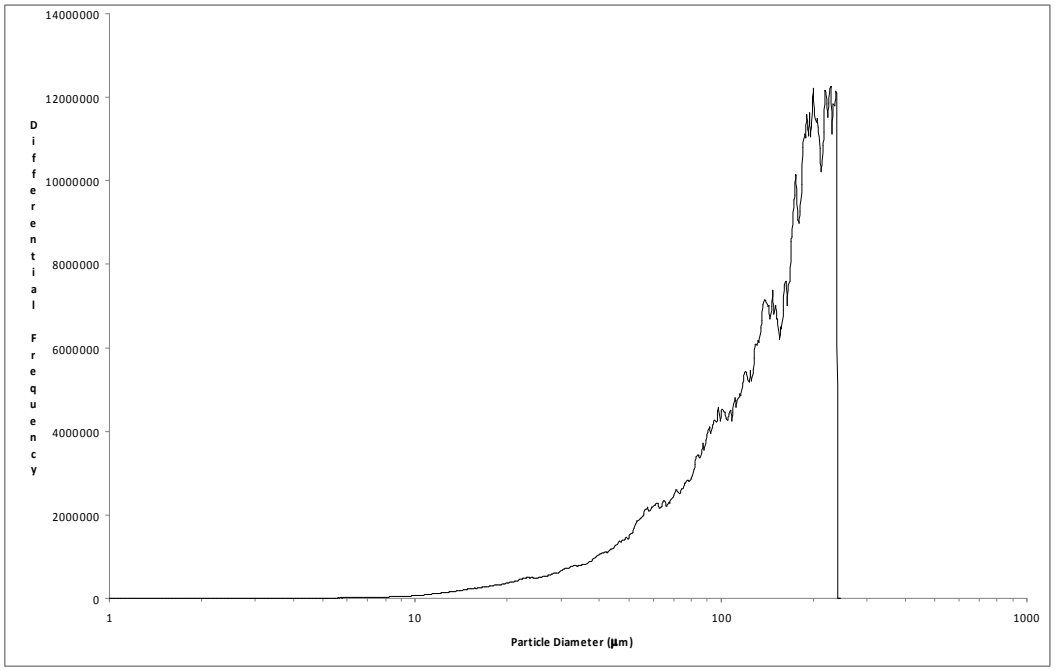

Figure 7. Volume-weighted differential particle diameter after blending – full range

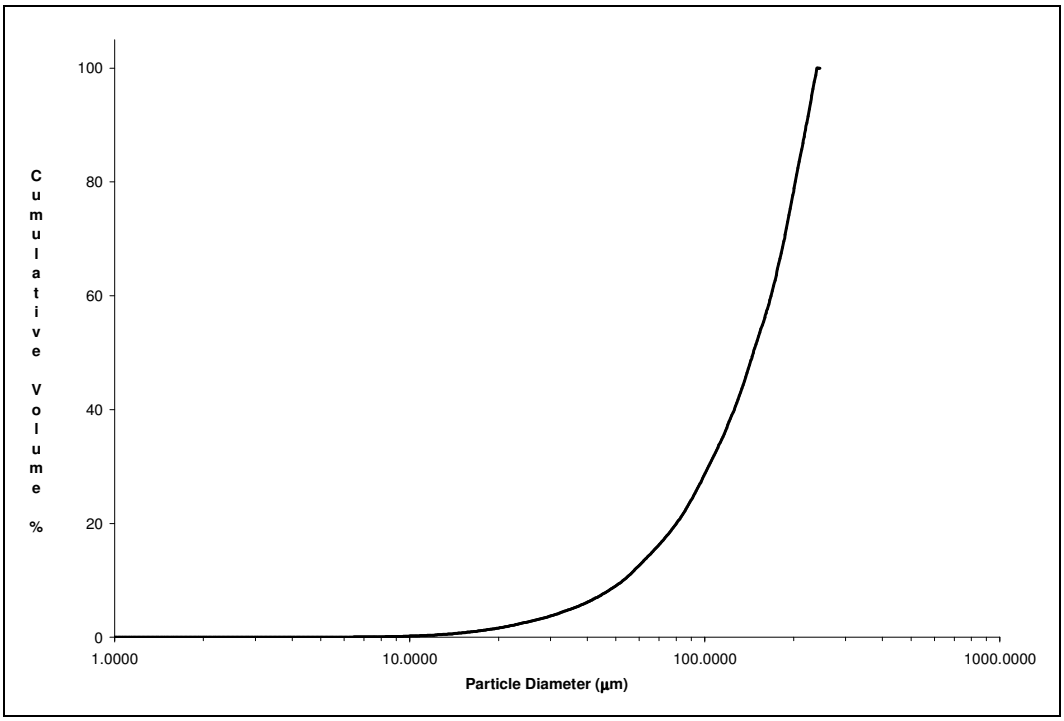

Figure 8. Cumulative volume % of the blended sample population

Table I. Volume and number weighted statistics

Parameter	µm
D[4,3]	142.1
D[v,0.1]	52.9
D[v,0.5]	145.4
D[v,0.9]	220.5
D[1,0]	11.6
D[n,0.1]	1.1
D[n,0.5]	5.5
D[n,0.9]	24.7

Discussion

As demonstrated in Figures 5 to 8, it was evident how much of the entire sample population was characterized by the low magnification (large particle) data, and how very little of the volume was characterized by the high magnification (small particle) data. In this example, less than 1% of the total sample volume was characterized by the high magnification (small particle) data. This small volume may seem insignificant, but the significance is dependent on how the data relate to the question being asked. For instance, in the case of pulmonary drug delivery, there exists a maximum in the deposition fraction for particles between 2 and 5 µm (5). In this example case, the start of the blending region was 4.4 µm, which is very near the lower limit of detection for the 2.5X objective. If the question being asked of this analysis were to pertain to bioavailability and dosage of a pulmonary drug, the accuracy and precision of the data below 5 µm would be critical, hence, necessitating the use of the 10X objective and a robust blending procedure.

The original observed particle size range was from about 2 µm to 200 µm, and the actual measured distribution had a range from 1 µm to 251 µm, which was in good agreement. Although this work focused on calculating the volume-weighted statistics and data, the number-weighted statistics and data are also shown in Figure 9 and Table I. It was of interest to note 17% of the particles by number were <4 µm, while only 0.1% of the particles by number were >200 µm. As one would expect, the number-weighted differential distribution plot was skewed very heavily towards the small particles as opposed to the volume-weighted differential distribution plot, which was skewed very heavily towards the large particles.

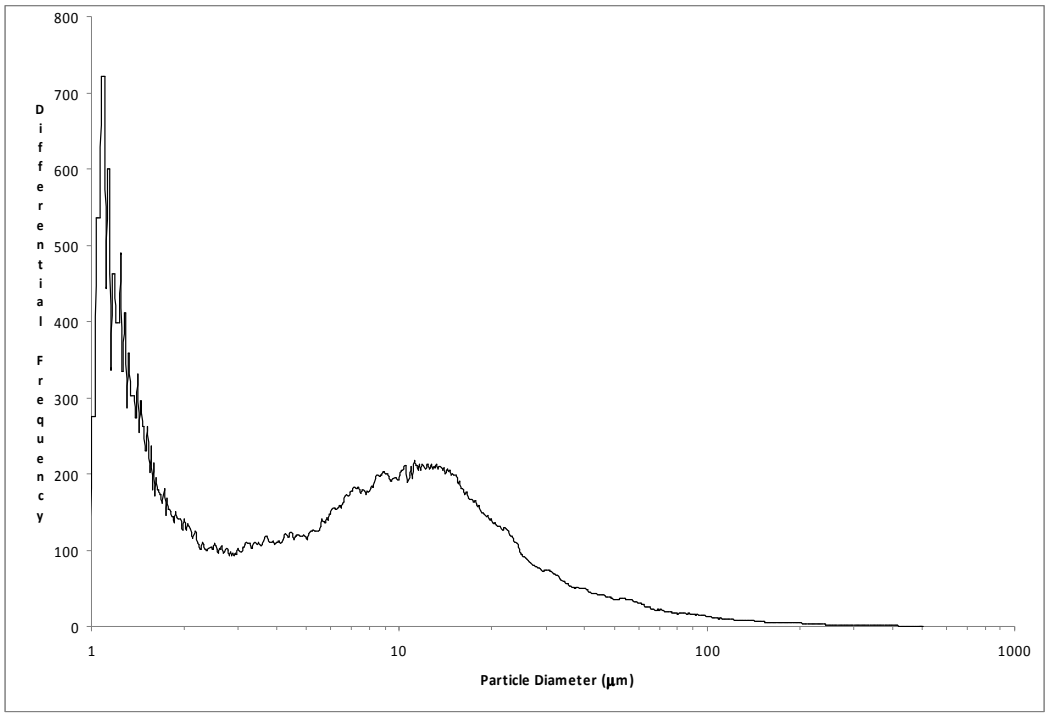

Figure 9. Number-weighted differential particle diameter after blending – full range

A total of about 57,000 particles were analyzed. A lognormal distribution function was fit to the volume-weighted data and the geometric standard deviation was approximated to be 1.6. At the 95% confidence level, this suggests about 33,000 particles were needed for the analysis. Thus, a sufficient number of particles were analyzed. As previously mentioned, the number of channels over which the blending occurred in this example case was 101. However, it is important to note that these 101 channels accounted for 20,327 particles of the total 41,000 (nearly 50%) that were analyzed.

As previously mentioned, both the moving average and Savitsky-Golay filters were investigated for smoothing the raw data. Ultimately, the decision was made to use the moving average. This decision was made based on two reasons: the moving average was easier to use, thus prone to less human error, and that visually, the moving average appeared to be more efficient at smoothing the data given the same number of smoothing points.

Likewise, a range of smoothing points from 5 to 17 was investigated. As shown in Figure 3, the 5-point smoothing filter made an improvement to the plot as opposed to the raw data (Figures 1 and 2), but it was not enough to clearly determine the optimum blending range. Application of the 17-point smoothing filter resulted in a plot that was exceptionally smooth. However, as the number of points in the smoothing filter increased, the D[4,3] also increased (see Table II and Figure 10) until the number of smoothing points reached 11, at which point the D[4,3] remained virtually constant.

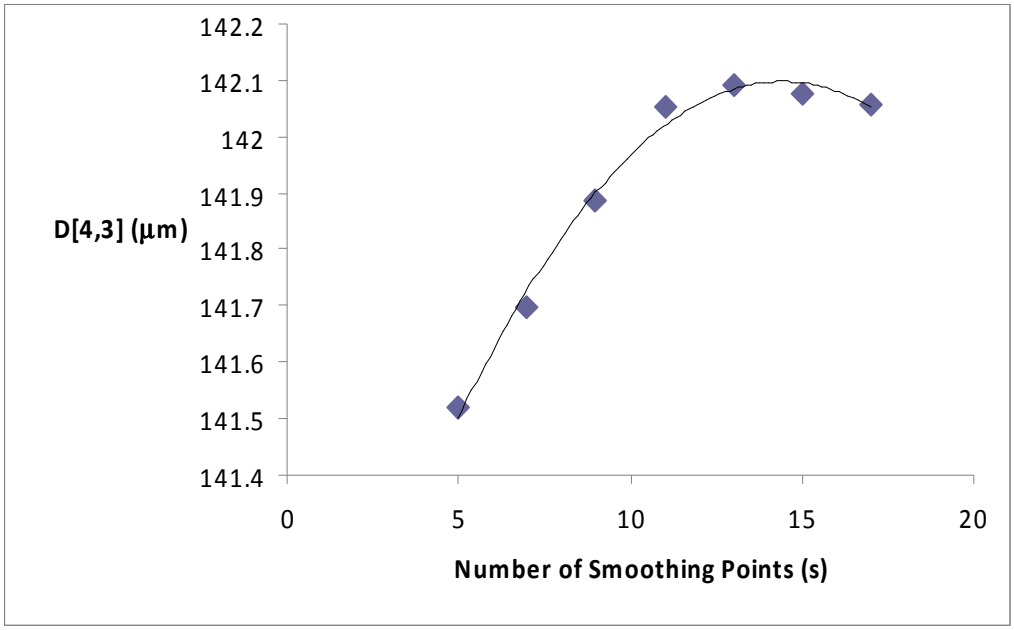

Figure 10. Effect of the number of smoothing points (s) on the D[4,3]

Table II. Effects of the application of a smoothing filter

Number of Points (s)	Normalization Constant	D[4,3](µm)
5	0.33220	141.521
7	0.33226	141.696
9	0.33237	141.888
11	0.33258	142.053
13	0.33294	142.091
15	0.33325	142.078
17	0.33345	142.056

A plot of the normalization constant vs. the number of smoothing points is given in Figure 11. When the data were fit with the cubic polynomial function as shown, the correlation coefficient (R^2) was 0.9963, indicating a good fit. The first derivative of the data in Figure 11 can be roughly fit ($R^2 = 0.84$) by a parabola with an inflection point near s = 11, where s is the number of smoothing points as shown in Figure 12. Since this was an inflection point, it can be thought of as a point with the least amount of change or that, which was most stable. Hence, the number of smoothing points in the filter was chosen to be 11.

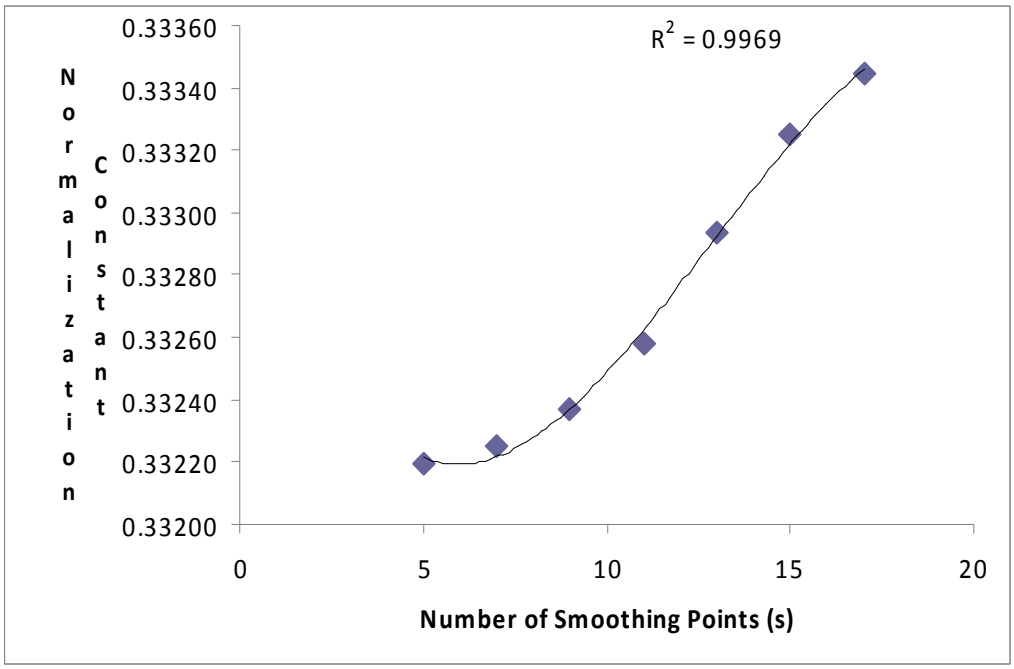

Figure 11. Normalization constant as a function of the number of smoothing points

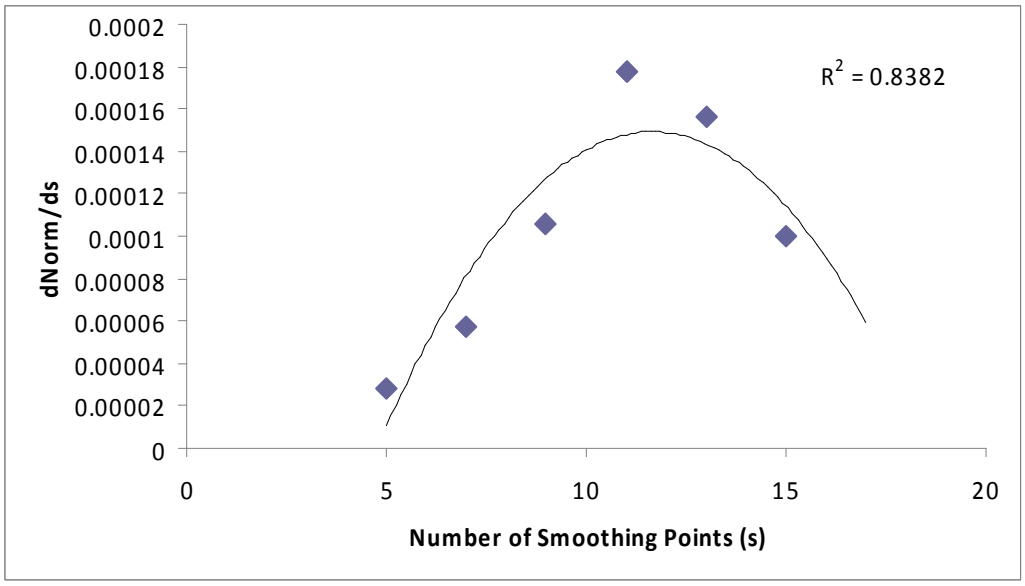

Figure 12. First derivative of the normalization constant as a function of the number of smoothing points

The logic function used to construct the final blended data could potentially lead to errors, the greatest of which, would be expected at the points of discontinuity. However, this can generally be minimized by ensuring the proper blending region is chosen.

CONCLUSIONS

This paper has described methodology by which particle size data with different magnification levels are blended into a single particle size distribution. This methodology provides optimal accuracy in determining the particle size distribution. Typical particle size analytical procedures (i.e., non-blended procedures) are biased to either small or large particle size regions of the distribution. This bias could lead to significant errors in judgment depending on the objective of the analysis.

REFERENCES

1. Microsoft Office Excel 2003 SP3, Part of Microsoft Office Professional Edition 2003, Copyright 1985-2003 Microsoft Corporation.
2. ISO 13322-1:2005, Particle Size Analysis. Image Analysis Methods. Static Image Analysis Methods.
3. JMP 8.0.1, Copyright 2009 SAS Institute, Inc.
4. Savitzky, A., Golay, M., "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," Analytical Chemistry, 36, 1964, 1627–1639.
5. Dolovich MB, Newhouse MT. Aerosols. Generation, methods of administration, and therapeutic applications in asthma. In: Middleton E Jr, Reed CE, Ellis EF, Adkinson NF Jr, Yunginger JW, Busse WW, editors, *Allergy Principles and practice.* 4. St Louis: Mosby Year Book, Inc.; 1993. pp. 712–739.

ARTICLE ACRONYM LISTING

CCD Charge Couples Device
ISO International Organization for Standardization
SDU Sample Dispersion Unit