Journal of Aerosol Science 114 (2017) 317–326

Contents lists available at ScienceDirect

Journal of Aerosol Science

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

Repeatability and reproducibility of TEM soot primary particle size measurements and comparison of automated methods

Paul M. Anderson[a,*], Haiqing Guo[b], Peter B. Sunderland[c]

[a] *Department of Mechanical Engineering, 2181 Glenn L. Martin Hall, Building 088, University of Maryland, College Park, MD 20742, USA*
[b] *C-Far Services, Federal Aviation Administration Technical Center, Atlantic City, NJ 08405, USA*
[c] *Department of Fire Protection Engineering, 3106 J.M. Patterson Building, University of Maryland, College Park, MD 20742, USA*

ARTICLE INFO

Keywords:
Gage R & R
Uncertainty analysis
Transmission electron microscopy
Aerosols
Statistics

ABSTRACT

Transmission electron microscopy (TEM) is used widely to measure the size and structure of aerosols, however the repeatability and reproducibility of these measurements has never been investigated. In this study, the primary particle size distributions of three soot samples from a propylene/air co-flow diffusion flame are measured from TEM images by four independent operators. The overall uncertainty, accounting for repeatability and variation across samples and operators, is quantified through gage repeatability and reproducibility (Gage R & R) analysis. A randomized measurement procedure coupled with a semi-automated algorithm called center-selected edge scoring, is found to limit operator bias while also taking a fifth as long as manual measurement. Accounting for all sources of uncertainty, the 95% confidence intervals determined from a single measurement of primary particle mean and standard deviation are ± 14% and ± 33%, which is greater than is commonly reported. Since Gage R & R analysis requires more replicate measurements than is often practical, a simplified approach for obtaining confidence intervals is presented which can capture Gage R & R uncertainties while requiring only a few replicate measurements of a given sample. The Gage R & R results are used to evaluate three automated methods, two employing a Circular Hough Transform (CHT) algorithm and one utilizing Euclidian distance mapping (EDM). The CHT methods underestimate the mean primary particle diameter, while the EDM method overestimates it. Both the CHT method with Canny edge detection and the EDM method assess reasonably well the standard deviation of the size distribution.

1. Introduction

Many aerosols, such as combustion generated soot, exhibit a fractal-like geometry resulting from the aggregation of small, nearly spherical, primary particles. Accurate knowledge of the size distribution of these primary particles is critical to quantifying a wide range of aerosol properties. Primary particle size is needed to derive optical properties from Rayleigh–Debye–Gans fractal aggregate theory (Dobbins & Megaridis, 1991; Sorensen, 2001; Williams, Shaddix, Jensen, & Suo-Anttila, 2007), and charging efficiency in electrical mobility analysis depends on primary particle size (Bau, Zimmermann, Payet, & Witschger, 2015; Lall & Friedlander, 2006). Morphological properties, such as fractal dimension, surface area, and volume, also depend on primary particle size distribution (Bau, Witschger, Gensdarmes, Rastoix, & Thomas, 2010; Köylü, Faeth, Farias, & Carvalho, 1995; Park, Kittelson, Zachariah, & McMurry, 2004).

* Corresponding author.
E-mail addresses: pmanders@umd.edu (P.M. Anderson), haiqing.ctr.guo@faa.gov (H. Guo), pbs@umd.edu (P.B. Sunderland).

http://dx.doi.org/10.1016/j.jaerosci.2017.10.002
Received 13 June 2017; Received in revised form 24 September 2017; Accepted 3 October 2017
Available online 06 October 2017
0021-8502/

Nomenclature

a_1, a_2	limits of bias correction distribution
B	expected value of bias correction
d	scale diameter (Eq. (8))
d_p	primary particle diameter
$d_{p,min}$	specified minimum primary particle diameter
$d_{p,max}$	specified maximum primary particle diameter
$\overline{d_p}$	mean primary particle diameter
$\overline{d_{pg}}$	geometric mean primary particle diameter
E_{ijk}	measured difference from mean due to random error
k	coverage factor
n	number of measured primary particles
N	number of circumference pixels
o	number of operators
O_j	measured difference from mean due to operator effect
P_i	measured difference from mean due to sample effect
r	number of replicate measurements
s	standard deviation
s_{dp}	standard deviation of primary particle diameter
u	standard uncertainty
w	statistical weight equaling the inverse variance
$\lvert x_n \rvert$	magnitude of pixel intensity gradient
Y	the true value of a measurand
Y_{ijk}	a single measured value of a measurand
$\overline{Y_{***}}$	the mean measured value of a measurand

Greek letters

α	significance level
β, Ω	calibration constants (Eq. (8))
$\Delta\theta_n$	difference in pixel intensity gradient orientation from that of a perfect circle
μ	true mean
$\hat{\mu}$	point estimator of mean
σ^2	true variance
$\hat{\sigma}^2$	point estimator of variance
σ_{pg}	geometric standard deviation of primary particle diameter

Subscripts

A	a given sample
B	bias correction
C	combined result
dp	primary particle diameter
$\overline{\mathrm{dp}}$	mean primary particle diameter
E	random effect
i	sample index
j	operator index
k	replicate measurement index
M	overall measurement system effect
n	circumference pixel index
O	operator effect
P	sample effect
s	standard deviation
Y	a measurand

Manual analysis of transmission electron microscopy (TEM) images is the most common method for measuring primary particle size, and the results are often used to calibrate or validate other measurements. De Iuliis, Maffi, Cignoli, and Zizak (2010) employed manual TEM measurements to validate the results of three-angle scattering and extinction. Laser induced incandescence (LII) measurements of primary particle size have often relied on manual TEM measurements to calibrate (Vander Wal, Ticich, & Stephens, 1999) or validate results (Cenker et al., 2015; Kock, Tribalet, Schulz, & Roth, 2006). However, a recent comparison of TEM and LII primary particle measurements by Kholghy et al. (2017) showed that additional factors such as shielding and bridging of particles due to high aerosol concentration can complicate validation. Attempts at automating the TEM measurement process have also relied on manual measurements to calibrate (Bescond et al., 2014) or validate results (Grishin, Thomson, Migliorini, & Sloan, 2012; Kook et al., 2016).

Given the large number of measurements needed to quantify morphological properties with statistical significance, and the laborious nature of manually obtaining them, there is a strong incentive to automate the TEM measurement process. Most attempts at automating measurement of primary particles have followed one of two approaches. The first is built around the Hough Transform (Hough, 1962), a well-known algorithm that identifies pre-defined geometries, such as a line or circle, by the strength of features in a parameter space. Grishin et al. (2012) were the first to apply the Circular Hough Transform (CHT) to fractal aggregates. Because the algorithm relies on accurate edge detection, primary particle detection was restricted to the perimeter of a binarized aggregate image. Other studies (Kook et al., 2016; Wang et al., 2016), extended the technique to interrogate the interior of an aggregate using a combination of image self-subtraction and Canny edge detection (Canny, 1986). Canny edge detection is known for its ability to detect weak edges and is therefore thought to be well-suited to discerning primary particles in the interior of a soot aggregate. A drawback of the CHT approach is that it requires accurate pre-knowledge of several input parameters, including the range of primary particle diameters and the CHT sensitivity level.

The second automation method is based upon Euclidian distance mapping (EDM) of a binarized aggregate image (Bescond et al., 2014; De Temmerman, Verleysen, Lammertyn, & Mast, 2014). EDM is a technique whereby pixels within a binarized image object are assigned grayscale intensities corresponding to their distance from the object's nearest edge. The procedure outlined by Bescond et al. (2014) involves thresholding the EDM image of an aggregate at a specific pixel intensity, then inverting and repeating the EDM-threshold sequence. Repeating this process across all possible threshold levels, a sigmoidal function emerges relating the number of white pixels remaining after processing, to threshold value. Following calibration, this can be translated into a primary particle size distribution function.

The EDM approach has the advantage of not requiring an assumed geometry or size range for detection. However, its results must be calibrated against manual TEM measurements, and the shape of the size distribution must be assumed. Additionally, one must assume that the calibration holds across samples with potentially dissimilar morphologies.

Despite its importance and use as a calibration/validation source, uncertainties associated with manual TEM soot measurements have received little attention. If reported at all, the standard uncertainty of the mean primary particle diameter is typically < 3% and represents the standard deviation of the mean of a single measurement replicate by a single operator on a single sample (e.g., Köylü & Faeth, 1992; Hu, Yang, & Köylü, 2003; Bescond et al., 2014). A study by Kondo, Aizawa, Kook, and Pickett (2013) observed that different operators, especially those with less experience, were the greatest source of fluctuation among primary particle measurements. However, to date there has been no formal attempt to quantify the uncertainty of TEM soot measurements resulting from different operators, samples, and measurement replicates. In addition, opportunities for bias are introduced at nearly every stage of the measurement process. These include bias in choosing which regions of the TEM grid to investigate, which aggregates to image from those regions, which images to select for measurement, which image areas and primary particles to consider, and how to best interpret the size of a given primary particle. With a few exceptions (e.g., Wozniak, Onofri, Barbosa, Yon, & Mroczka, 2012; Lapuerta, Martos, & Expósito, 2013; Toth, Farrer, Palotas, Lighty, & Eddings, 2013), studies rarely employ safeguards against operator bias.

The objective of this investigation is to provide guidance for minimizing TEM measurement uncertainties, present a statistically robust estimate of the true uncertainty associated with primary particle TEM measurements, and then provide guidance for how others may efficiently account for these uncertainties in their own measurements. Assessment of the true TEM measurement uncertainty is accomplished by applying analysis of variance (ANOVA) gage repeatability and reproducibility (Gage R & R) analysis to the semi-automated measurement of mean primary particle diameter and standard deviation. The confidence intervals obtained from Gage R & R analysis are used to evaluate the performance of several different automated primary particle measurement methods.

2. Methods

2.1. Experimental apparatus

Soot was generated in an atmospheric pressure, laminar propylene/air diffusion flame supported by a co-flow burner (Santoro et al., 1983), consisting of concentric brass tubes with inside diameters of 14 and 101.6 mm. A flow of 2.1 g/s of propylene through the inner tube was surrounded by 1.18 g/s of co-flowing air. The propylene flame had a luminous length of about 50 mm and emitted soot in a vertical column. Additional details may be found in Guo, Anderson, and Sunderland (2015). Soot was thermophoretically sampled by rapidly inserting a probe carrying a 200 mesh copper TEM grid with a carbon support film into the flame (Dobbins & Megaridis, 1987). TEM imaging was performed on a JEOL JEM 2100 LaB_6 at a magnification of 10,000, which was high enough to capture aggregate details without excluding larger aggregates from the viewing window. The TEM length scale was calibrated against a gold diffraction grating replica (Ted Pella, part no. 607), and the standard uncertainty of the calibration is estimated to be 0.8% at this magnification.

2.2. Preliminary measurements without bias controls

To assess the effect of different operators in the absence of bias limiting procedures, three operators measured primary particle diameters on the image sets of two samples collected at a height of 13.4 cm above the co-flow fuel port, or 1.3 cm above the second

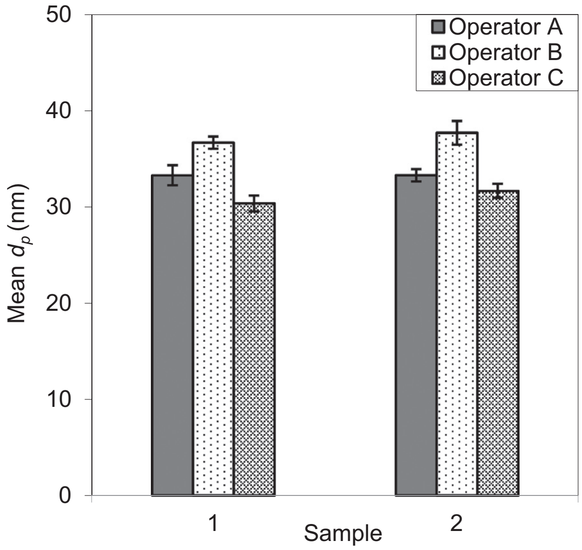

Fig. 1. Preliminary measurements of mean primary particle diameter without bias controls. Error bars reflect the precision of the mean at the 95% confidence level.

stage ring burner of the ternary flame system detailed in Guo et al. (2015). The operators were not told which images or regions therein to measure. The results (Fig. 1) demonstrate considerable variability, which cannot be attributed to statistical precision alone. Moreover, the consistent relative trend among the three operators indicates a significant between-operator effect. The overall fluctuation, calculated as the measurement range over its average, was 18.9% and 17.6% for samples 1 and 2, respectively. This is in line with the fluctuation across all operators reported by Kondo et al. (2013).

2.3. *Semi-automated measurements*

2.3.1. *Bias controls*

Three soot samples were drawn on separate occasions at a height of 10 cm above the co-flow fuel port. Coordinates for imaging on the TEM grid plane were randomly generated. If soot was present at a coordinate, an image was recorded with no attempt to center the viewing window on any particular image element. Individual aggregates were identified in the images and then randomly chosen for analysis. Four independent operators, three of whom made the preliminary measurements, performed image analysis using a custom program developed in MATLAB. The program randomly selected 100 × 100 pixel regions of interest (ROIs) on each soot aggregate image. To ensure that more weight was not given to primary particles of smaller aggregates, the number of ROIs was proportional to an automated estimate of the aggregate's projected area. Operators were directed to measure every element within these ROIs they considered to reasonably constitute a primary particle using a semi-automated algorithm called center-selected edge scoring (CSES) which is detailed in Section 2.3.2. If not satisfied with the algorithm's output, the operator could reject it and remeasure by selecting a different pixel or restricting the diameter range. The semi-automated measurement therefore reflects an operator's best judgement, but is also expected to reduce bias since it quantifies image features uniformly. Fig. 2 shows a typical soot aggregate, whose primary particles have been measured using the CSES algorithm within random ROIs. Each operator performed 2 replicate measurements on each sample's image set, yielding a total of 24 measurement sets with an average of 290 measured primary particles per set. The randomly chosen images and ROIs used for a given replicate were the same for all operators.

2.3.2. *CSES algorithm*

The existing methods of CHT and EDM rely on detectible edges in an image to infer circular features. By contrast, center-selected edge scoring (CSES) begins from an estimate of a primary particle's center point, and then, like a manual operator, considers image features over the entire circumference of a primary particle, not just those parts detectable as "edges." Fig. 3 outlines this process. The operator manually selects a point on the image that is judged to be the center of a primary particle. A square neighborhood of pixels around the selected point is calculated (Fig. 3a). Candidates for a best-fit circle are those within an operator-specified diameter range, $d_{p,\min}$ to $d_{p,\max}$, centered at any pixel within this neighborhood. A neighborhood measuring $1 + 2[\mathrm{round}(0.025 d_{p,\max} + 1)]$ on a side was found to match operators' intuitive expectations for candidate center point tolerance. Using a 3 × 3 Sobel filter, the algorithm computes a gradient map of the grayscale intensity at each image pixel (Fig. 3b). For each candidate circle, all pixels in the gradient

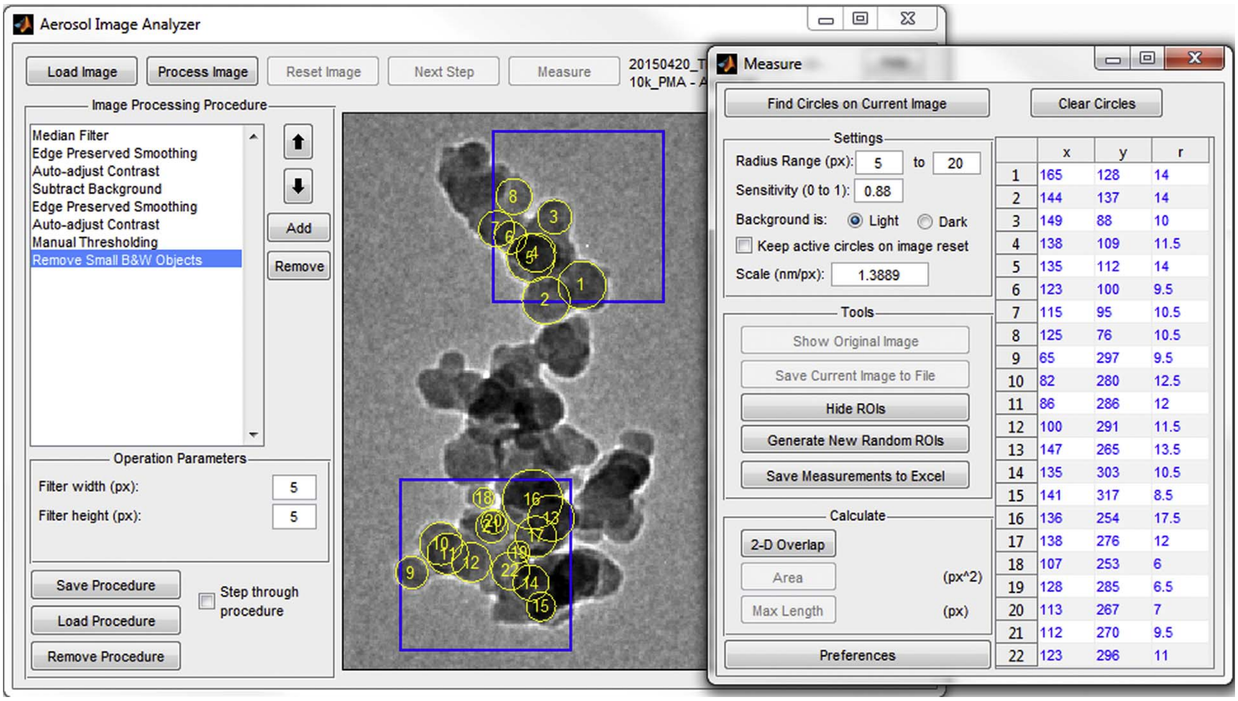

Fig. 2. Screenshot of the MATLAB-based image analysis program. Primary particles have been measured using the CSES algorithm within randomly generated ROIs.

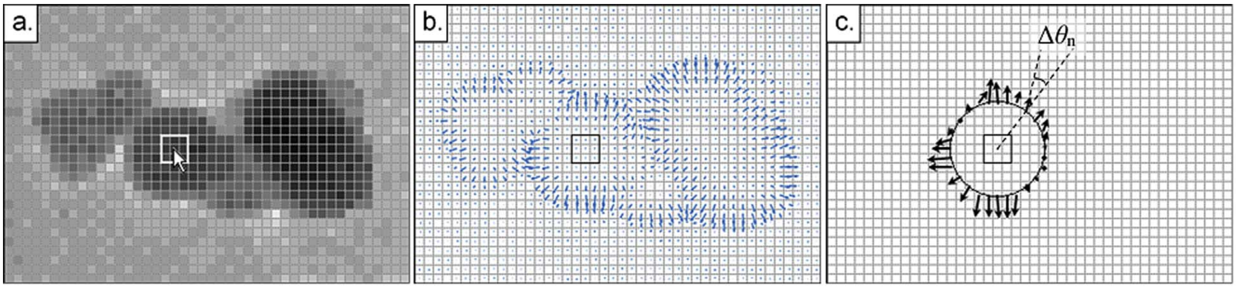

Fig. 3. Steps in the CSES algorithm. A low magnification image is shown to clearly illustrate individual pixels. (a) Pixels of candidate center points around an operator-selected pixel. (b) Grayscale intensity gradient map with vectors denoting gradients. (c) The magnitude and orientation (shown by vectors) of the gradient of a candidate circle's edge pixels.

map are eliminated except those that comprise the candidate circle's circumference (Fig. 3c). The fit of each circle is scored using each pixel's gradient magnitude $|x_n|$, and the difference in its gradient orientation from that of a perfectly circular primary particle, $\Delta\theta_n$ according to:

$$\text{Score} = \frac{1}{N}\sum_{n=1}^{N} |x_n| \cos(\Delta\theta_n), \tag{1}$$

where N is the number of circumference pixels. The best-fit circle is chosen as that with the greatest score, according to Eq. (1). The CSES code and supporting files are provided as Online Supplemental information, and the custom image analysis program is available from the authors upon request. Though developed using soot TEM images, the CSES algorithm has application to any aerosol that is spheroidal or comprised of spheroidal primary particles.

2.3.3. Gage R & R analysis

The true uncertainty of the measurement method described above and the extent to which it eliminates operator bias, was assessed through gage repeatability and reproducibility analysis. The Gage R & R model appropriate for this study is a balanced two-factor crossed random model (Burdick, Borror, & Montgomery, 2005):

$$Y_{\mathrm{ijk}} = \mu_{\mathrm{Y}} + P_{\mathrm{i}} + O_{\mathrm{j}} + E_{\mathrm{ijk}}. \tag{2}$$

Here, Y_{ijk} is the measured value of sample i, by operator j, the k^{th} time, and μ_{Y} is the true mean of the measurand Y from the population of all samples producible by the given process (i.e., the specified co-flow flame). The terms P_{i}, O_{j}, and E_{ijk}, are jointly independent normal random variables with means of zero and variances σ_{P}^2, σ_{O}^2, and σ_{E}^2, corresponding to the variation away from μ_{Y} due to effects of sample, operator, and random error, respectively. The variances σ_{O}^2 and σ_{E}^2 are the reproducibility and repeatability, and their sum is the variance of the measurement system, σ_{M}^2.

The model given in Eq. (2) assumes that operator-sample interaction is insignificant and that overall repeatability uncertainty is greater than measurement resolution. The measurands here are the mean, $\overline{d_{\mathrm{p}}}$, and standard deviation, s_{dp}, of the distribution, not individual primary particle measurements. The resolution of $\overline{d_{\mathrm{p}}}$ and s_{dp} estimated from n measured primary particle diameters are given by their respective standard deviations: $s_{\overline{\mathrm{dp}}} = s_{\mathrm{dp}}/\sqrt{n}$ and $s_{s_{\mathrm{dp}}} = s_{\mathrm{dp}}/\sqrt{2(n-1)}$. The latter expression assumes a normal distribution of primary particles, however uncertainty due to any divergence from normality is expected to be negligible compared to other sources of uncertainty.

The values of these variances and μ_{Y} found with Gage R & R analysis are point estimators (so denoted $\hat{\sigma}^2$) of the true values, and are themselves subject to uncertainty. Uncertainty distributions for these terms were simulated following the generalized inference approach of Hamada and Weerahandi (2000). Sampling from these distributions, Monte Carlo methods (Solaguren-Beascoa Fernández, Alegre Calderón, & Bravo Díez, 2009) were employed to compute confidence intervals for the means and standard deviations of the measured size distributions.

To combine operator measurements and obtain confidence intervals for each sample, a balanced one factor random ANOVA model is used:

$$Y_{\mathrm{Ajk}} = \mu_{\mathrm{Y_A}} + O_{\mathrm{Aj}} + E_{\mathrm{Ajk}}. \tag{3}$$

For a given sample A, Y_{Ajk} is the k^{th} replicate measurement by operator j, and $\mu_{\mathrm{Y_A}}$ is the true mean of the measurand Y_{A}. Since the sample is fixed, and O_{Aj} and E_{Ajk} are assumed to be normally distributed about a mean of zero, $\mu_{\mathrm{Y_A}}$ also represents the true value of Y_{A}, for which the best estimate, $\hat{\mu}_{\mathrm{Y_A}}$, is the mean of all measurements, $\overline{Y_{\mathrm{A}**}}$, for the sample. The confidence interval for Y_{A} is then found from

$$Y_{\mathrm{A}} = \hat{\mu}_{Y_A} \pm \sqrt{s_{\mathrm{O}}^2 F_{1-\alpha:1,o-1}/ro}\,, \tag{4}$$

where $s_{\mathrm{O}}^2 = (r/(o-1))\sum_{j=1}^{o}(\overline{Y_{\mathrm{Aj}*}} - \overline{Y_{\mathrm{A}**}})^2$ is the mean squares for o operators each performing r replicates, $\overline{Y_{\mathrm{Aj}*}}$ is the mean of measurements by operator j, and $F_{1-\alpha:1,o-1}$ is the F-statistic at significance level α, having 1 and $o-1$ degrees of freedom (Burdick et al., 2005).

2.3.4. A simplified approach to obtaining sample confidence intervals

As a practical matter, it would be desirable to estimate the uncertainty of a sample's primary particle distribution without the relatively large number of replicates required by ANOVA. One approach treats the combination of operator results as an inter-laboratory evaluation in which operators are equally competent and subject to independent random and systematic effects (Kacker, Datla, & Parr, 2002). The measurements of each operator j for a given sample are pooled, and the resulting distribution will have a mean and standard deviation with standard random uncertainties, equal to $s_{\overline{\mathrm{dp}},\mathrm{j}}$ and $s_{s_{\mathrm{dp}},\mathrm{j}}$, as given in Section 2.3.3. For a given sample A, the results Y_{Aj} from each operator j may then be combined using a weighted mean, where the weight is defined as $w_{\mathrm{Aj}} = 1/s^2_{Y_{\mathrm{Aj}}}$. The combined result, $Y_{\mathrm{A,C}}$, is then:

$$Y_{\mathrm{A,C}} = \sum_{j=1}^{o} w_{\mathrm{Aj}} Y_{\mathrm{Aj}} / \sum_j w_{\mathrm{Aj}}, \tag{5}$$

for which the standard uncertainty is:

$$u_{Y_{\mathrm{A,C}}} = 1/\left(\sum_j w_{\mathrm{Aj}}\right)^{1/2}. \tag{6}$$

If the interval $2u_{Y_{\mathrm{A,C}}}$ of this combined mean fails to encompass a non-negligible fraction of operator results, an expanded confidence interval that includes an additive bias correction with its own expectation value and uncertainty distribution is appropriate. Since each operator's results are believed to be equally probable, this correction is assumed to follow a rectangular distribution on the interval $(-a_1,\ a_2)$, where $-a_1 = \min_{\mathrm{j}}(Y_{\mathrm{Aj}}) - Y_{\mathrm{A,C}}$ and $a_2 = \max_{\mathrm{j}}(Y_{\mathrm{Aj}}) - Y_{\mathrm{A,C}}$. The expected value for this correction is $B = (a_2 - a_1)/2$ with a standard uncertainty $u_{\mathrm{B}} = (a_2 + a_1)/\sqrt{12}$. Finally, the confidence interval for the measurand Y_{A} is:

$$Y_{\mathrm{A}} = (Y_{\mathrm{A,C}} + B) \pm k\left(u^2_{Y_{\mathrm{A,C}}} + u^2_{\mathrm{B}}\right)^{1/2}, \tag{7}$$

where a value of 2 for the coverage factor k corresponds to a 95% confidence level.

2.4. Automated measurements

2.4.1. Circular Hough Transform

Three different automated measurement methods were applied to the TEM image sets obtained in Section 2.3.1. The first, proposed by Grishin et al. (2012), applies CHT to binarized images of soot aggregates, and will be referred to as the binary-CHT method. Techniques for automating the image binarization step have been proposed (Grishin et al., 2012; Park, Huang, Ji, & Ding, 2013), however, we found no automated procedure that reliably separated aggregate from background. We had greatest success with a general procedure that employed edge-preserved smoothing (e.g., median filtering) and contrast enhancement followed by background removal via a rolling-ball Transform and manual thresholding. Grishin et al. (2012) used a modified CHT algorithm to reduce processor load and eliminate falsely identified primary particles produced by convex fragments on the aggregate border. As these were not issues in our implementation, we used the unmodified CHT found in MATLAB's imfindcircles function.

The second method, suggested by Kook et al. (2016), also uses the CHT algorithm, however with different pre-processing. First, the image undergoes inversion and self-subtraction to enhance contrast between aggregate and background. Primary particle edges are then enhanced by negative Laplacian filtering, after which Canny edge detection is applied. We used the MATLAB script provided by Kook et al. (2016), which also utilizes the imfindcircles function. This method is referred to here as the Canny-CHT method.

Both CHT methods require the user to specify a diameter range and a sensitivity level. The sensitivity level is a value between 0 to 1, where higher sensitivities interpret more image features as circular. The Canny-CHT method also requires specified values for self-subtraction level and negative Laplacian filter shape parameter. Kook et al. (2016) observed that parameter values, especially the lower-bound diameter, and CHT sensitivity, substantially affected results. Without knowledge of the primary particle size distribution *a priori*, the choice of parameter values represents a source of uncertainty for all CHT methods. In practice, one would likely estimate the diameter range manually before performing automated analysis. Therefore, we used the diameter ranges found from the Gage R & R measurements as inputs to both CHT algorithms. For the binary-CHT method, we varied the sensitivity value between 0.8 and 0.92, based on our own experience. For the Canny-CHT method, we used parameter ranges found by Kook et al. (2016) to optimally detect soot primary particles produced by five different diesel combustion facilities. Sensitivity ranged from 0.75–0.79, self-subtraction level ranged from 0.8–1.2, and alpha was 0.1. Each of the input parameters to both CHT algorithms were then varied individually to assess sensitivity of the response variables.

2.4.2. Euclidian distance mapping

The EDM method of Bescond et al. (2014) was evaluated. This approach has the advantage of requiring no input parameters, however results must be calibrated and a distribution shape assumed. Bescond et al. (2014) assumed a log-normal distribution and computed calibration parameters that adequately fit EDM results from soot produced by both an aircraft engine and an ethylene laminar diffusion flame. Given the very different types of soot produced by these systems, the calibration is expected to be robust and was applied to the present study. Measurement was performed in the NIH software ImageJ using the plug-in provided by Bescond et al. (2014). The resulting measurements were expected to follow a sigmoidal curve:

$$\frac{S(d)}{S(0)} = \left\{1 + \exp\left[\frac{(\ln d - \ln \overline{d_{pg}})/\ln \sigma_{pg} - \beta}{\Omega}\right]\right\}^{-1} \tag{8}$$

The left-hand side of this equation is the normalized average aggregate surface area after the EDM Transformation at scale diameter, d, itself a function of the EDM threshold value. The calibration constants are $\beta = 1.90 \pm 0.03$ and $\Omega = 0.80 \pm 0.03$ (Bescond et al., 2014). The geometric mean, $\overline{d_{pg}}$, and geometric standard deviation, σ_{pg}, were determined from a fit of the sigmoidal curve to the data, and were then converted to their arithmetic values for comparison to results from other methods.

3. Results and discussion

3.1. Gage R & R results

The mean primary particle diameter, $\overline{d_p}$, and standard deviation, s_{dp}, for each measurement replicate is given in Fig. 4, with the corresponding Gage R & R results in Table 1. An F-test ($\alpha = 5\%$), confirmed the assumption that operator-sample interaction was negligible for both $\overline{d_p}$, and s_{dp}, ($p = 0.156$ and 0.102, respectively). Repeatability variances exceeded individual measurement precision, ($s^2_{\overline{dp}} \sim 0.6$ nm and $s^2_{s_{dp}} \sim 0.3$ nm), justifying the assumption that measurement resolution is subsumed within overall repeatability. It is also clear from the differences in replicate measurements in Fig. 4 that inhomogeneities in the sample account for some of the repeatability uncertainty. A follow-up test, in which 2 operators performed 5 replicate measurements on the same group of ROIs on successive days, found that inhomogeneities accounted for about two-thirds of the overall repeatability variance. Contrary to Kondo et al. (2013), there was no discernable trend in results based on the experience level of the operator.

Sampling from the variance distributions produced by the generalized inference method, Monte Carlo simulations were used to construct 95% confidence intervals for key metrics. The true mean of the primary particle diameter means producible by this process, $\mu_{\overline{dp}}$, is 32.5 ± 5.9 nm (± 18%), and the true mean of the standard deviation, $\mu_{s_{dp}}$, is 12.3 ± 4.3 nm (± 35%). The values of μ_{dp} and σ_{dp} of a sample for which a single measurement has been performed may be known within ± 14%, and ± 33%, respectively. The confidence intervals for μ_{dp} and σ_{dp} for any *unmeasured* sample produced by this process are 32.5 ± 8.9 nm (± 27%) and 12.3 ± 6.7 nm (± 54%), respectively, however a larger number of samples beyond the three used here would likely narrow these intervals.

For both $\overline{d_p}$, and s_{dp}, the total variance is almost evenly split between sample and measurement system effects. Nearly all the measurement system variation for $\overline{d_p}$ is due to repeatability, while for s_{dp}, it is mostly due to operator effects. A comparison of operator measurements of identical image regions indicates that operators had consistent preferences for the overall range of valid primary particles. Still, the muted influence of operator on $\overline{d_p}$ variance suggests that randomization and CSES semi-automation were effective in reducing operator bias.

Additionally, the time per primary particle needed to perform measurements using CSES was compared to a manual measurement on the same images in ImageJ. Twice measuring ~100 primary particles on two randomly selected image sets, CSES was found to require only about 20% as much operator time as the manual measurement. This improvement comes from the fact that a pixel can be selected at a perceived center point much more quickly than a circle can be drawn to fit a perceived circumference.

The two methods for combining operator measurements for each sample produced comparable results as shown in Table 2. The approach outlined in Eqs. (5)–(7) therefore offers a way to estimate the values and uncertainties of μ_{dp} and σ_{dp} in the absence of a large number of measurement replicates. Additionally, if measurements by multiple operators cannot be obtained, replicate

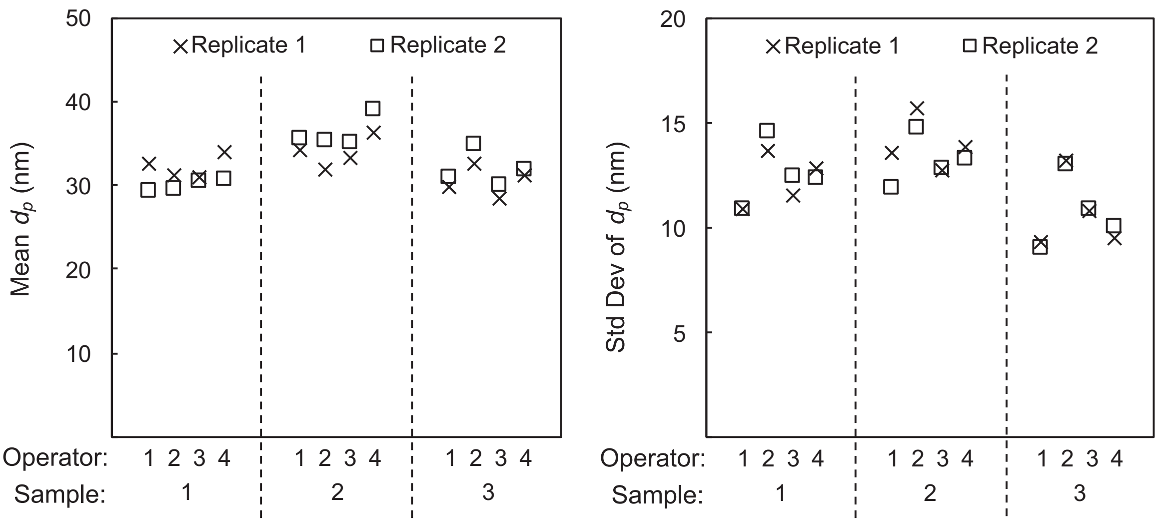

Fig. 4. Primary particle diameter mean and standard deviation for each Gage R & R replicate.

Table 1
Components of variance in the measurement of primary particle diameter mean and standard deviation as quantified by Gage R & R analysis.

Variance Component	$\overline{d_p}$ Gage R & R Results		s_{dp} Gage R & R Results	
	Variance	% Variance	Variance	% Variance
$\hat{\sigma}_P^2$	4.82	57.4%	2.03	48.1%
$\hat{\sigma}_O^2$	0.53	6.3%	1.82	43.1%
$\hat{\sigma}_E^2$	3.05	36.3%	0.37	8.8%
$\hat{\sigma}_M^2$	3.58	42.6%	2.19	51.9%
$\hat{\sigma}_{Tot}^2$	8.40	100%	4.22	100%

Table 2
Primary particle diameter mean and standard deviation of each sample using two methods for combining operator measurements. Intervals are at the 95% confidence level.

Sample	Mean d_p (nm)		Standard Deviation of d_p (nm)	
	Eq. (4)	Eqs. (5–7)	Eq. (4)	Eqs. (5–7)
1	31.1 ± 4.3%	31.3 ± 4.4%	12.5 ± 17.1%	12.6 ± 15.0%
2	35.1 ± 8.1%	35.7 ± 6.5%	13.6 ± 13.8%	13.7 ± 14.4%
3	31.3 ± 9.8%	31.5 ± 8.4%	10.8 ± 25.7%	11.2 ± 20.7%

measurements by a single operator utilizing bias limiting methods may serve as an adequate surrogate. The overall uncertainty of the combined result of r replicates on given sample A, could then be estimated from Eqs. (5)–(7) by incorporating $\hat{\sigma}_O^2$ from Table 1 into Eq. (6):

$$u_{Y_{A,C}} = \left(1/\left(\sum_{k=1}^{r} w_{Ak}\right) + \hat{\sigma}_O^2\right)^{1/2}. \tag{9}$$

3.2. Automated results

The three automated methods were applied to each of the 37, 41, and 73 aggregate images of samples 1, 2, and 3, respectively. The Gage R & R measurements found diameters ranging 6–79 nm, 8–86 nm, and 8–82 nm for these three samples. These ranges were supplied as inputs to the two CHT-based algorithms. The remaining parameters for each algorithm were varied as discussed in Section 2.4.1. The results of all three automated measurements are shown in Fig. 5 alongside the Gage R & R results combined across operators using Eq. (4). Both CHT-based algorithms drastically underpredicted $\hat{\mu}_{dp}$ for all three samples. The binary-CHT method also

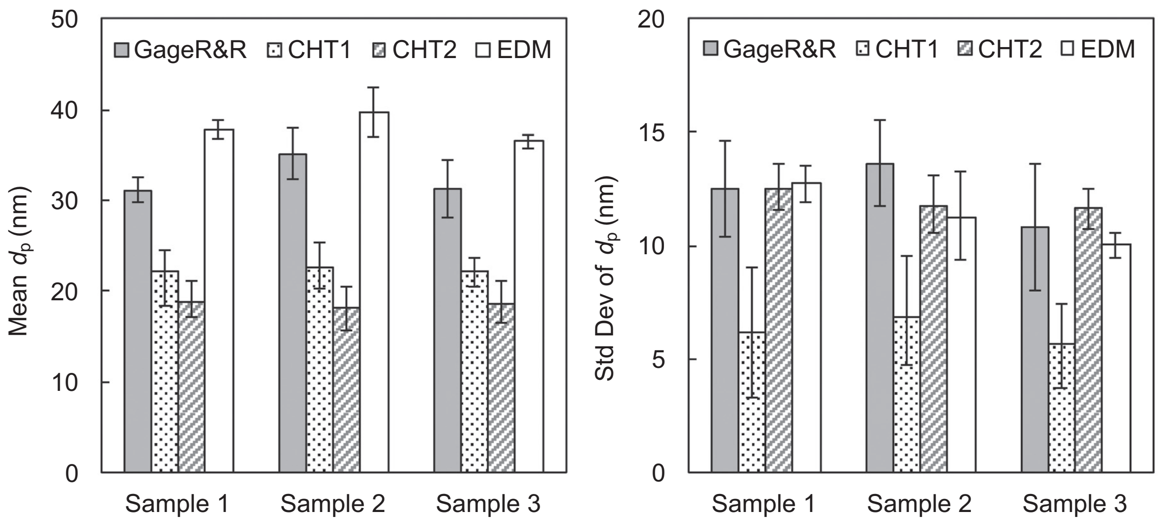

Fig. 5. Primary particle diameter mean and standard deviation for each sample as found by CSES (Gage R & R), binary-CHT (CHT1), Canny-CHT (CHT2) and Euclidian distance mapping (EDM) methods.

substantially underestimated $\hat{\sigma}_{dp}$, while the Canny-CHT method generally matched the Gage R & R $\hat{\sigma}_{dp}$ results. While performing better than CHT, the EDM method consistently overpredicted $\hat{\mu}_{dp}$. For $\hat{\sigma}_{dp}$, the results are statistically equivalent to the Canny-CHT results. Inaccuracy of the EDM results can be credited to the model's assumed calibration values and lognormal distribution. The latter is questionable given that the distributions found from the semi-automated measurements almost uniformly failed standard tests for both normality and lognormality.

The error bars in Fig. 5 are at the 95% confidence level. For the CHT results, the error bars reflect the statistical precision of the mean combined with uncertainty caused by varying input parameters other than diameter range. Error bars for the EDM method include the published uncertainty of the calibration constants (Bescond et al., 2014), and uncertainty in the sigmoidal fits, but do not consider repeatability and reproducibility uncertainty of the calibration.

The two CHT methods showed sensitivity to several of the input parameters, most prominently, the lower-bound diameter. For the binary-CHT method, the value of $\hat{\mu}_{dp}$ rose 1 nm for each 1 nm increase in $d_{p,min,}$ and for Canny-CHT, $\hat{\mu}_{dp}$ increased at twice that rate. This large dependence on $d_{p,min}$ is problematic. To force either CHT method to yield a value for $\hat{\mu}_{dp}$ that matches the Gage R & R results for all three samples, a value for $d_{p,min}$ between 17 and 19 nm must be specified. There is no practical reason to choose an input value within this range without having foreknowledge of the distribution. Choice of such a value would also truncate the bottom 16% of the distribution. One possibility for overcoming this difficulty is through a weighted scoring scheme for the Hough Transform that is a function of diameter.

4. Conclusions

A Gage R & R study of soot primary particle measurements on TEM images was performed utilizing a semi-automated method called center-selected edge scoring (CSES), which required only 20% as much operator time as manual measurement. Coupled with randomization of image measurement regions, the study showed only a weak operator influence on $\hat{\mu}_{dp}$. In contrast, different operators were the principle source of variation for $\hat{\sigma}_{dp}$. Monte Carlo sampling of Gage R & R variance distributions found 95% confidence intervals for a single estimate of μ_{dp} and σ_{dp}, were ± 14% and ± 33%, respectively. These are substantially greater than commonly reported uncertainties, which do not consider variation due to the measurement method. Uncertainties can be reduced by combining measurements across operators (Eqs. 5–7) or across replicates of a single operator (Eq. (9)). Doing so incorporates an estimate of repeatability and reproducibility uncertainties without demanding a large number of replicates.

All three automated methods considered were unable to produce results consistent with the Gage R & R measurements. The EDM method came closest, but consistently overestimated $\hat{\mu}_{dp}$. A calibration tailored to the soot produced by a specific process might improve this method's performance, although the need to assume a distribution shape will remain a source of uncertainty. On the other hand, both CHT methods greatly underestimated $\hat{\mu}_{dp}$ and the binary-CHT method also underestimated $\hat{\sigma}_{dp}$. Inaccuracies in the CHT methods are mostly attributable to a high sensitivity to the algorithms' lower-bound diameter input. Less sensitivity to this and other input parameters is needed to make this approach robust enough to use broadly.

Given the difficulties presented by existing automated techniques, manual measurement (or a semi-automated approach like CSES), with randomization to reduce bias, remains the best method for evaluating soot primary particle size distributions from TEM images.

Acknowledgements

We acknowledge the support of the Maryland NanoCenter and its AIMLab.
This work was supported by the National Science Foundation (NSF) Grant no. CBET0954441.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jaerosci.2017.10.002.

References

Bau, S., Witschger, O., Gensdarmes, F., Rastoix, O., & Thomas, D. (2010). A TEM-based method as an alternative to the BET method for measuring off-line the specific surface area of nanoaerosols. *Powder Technology, 200*(3), 190–201.

Bau, S., Zimmermann, B., Payet, R., & Witschger, O. (2015). A laboratory study of the performance of the handheld diffusion size classifier (DiSCmini) for various aerosols in the 15–400 nm range. *Environmental Science: Processes & Impacts, 17*(2), 261–269.

Bescond, A., Yon, J., Ouf, F. X., Ferry, D., Delhaye, D., Gaffié, D., & Rozé, C. (2014). Automated determination of aggregate primary particle size distribution by TEM image analysis: Application to soot. *Aerosol Science and Technology, 48*(8), 831–841.

Burdick, R. K., Borror, C. M., & Montgomery, D. C. (2005). *Design and analysis of gauge R & R studies: Making decisions with confidence intervals in random and mixed ANOVA models.* Philadelphia: Society for Industrial Applied Mathematics; Alexandria: American Statistical Association.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* (6), 679–698 (PAMI-8).

Cenker, E., Kondo, K., Bruneaux, G., Dreier, T., Aizawa, T., & Schulz, C. (2015). Assessment of soot particle-size imaging with LII at diesel engine conditions. *Applied Physics B, 119*(4), 765–776.

De Temmerman, P.-J., Verleysen, E., Lammertyn, J., & Mast, J. (2014). Semi-automatic size measurement of primary particles in aggregated nanomaterials by transmission electron microscopy. *Powder Technology, 261*, 191–200.

De Iuliis, S., Maffi, S., Cignoli, F., & Zizak, G. (2010). Three-angle scattering/extinction versus TEM measurements on soot in premixed ethylene/air flame. *Applied Physics B, 102*(4), 891–903.

Dobbins, R. A., & Megaridis, C. M. (1987). Morphology of flame-generated soot as determined by thermophoretic sampling. *Langmuir, 3*(*2*), 254–259.
Dobbins, R. A., & Megaridis, C. M. (1991). Absorption and scattering of light by polydisperse aggregates. *Applied Optics, 30*(*33*), 4747–4754.
Grishin, I., Thomson, K., Migliorini, F., & Sloan, J. J. (2012). Application of the Hough Transform for the automatic determination of soot aggregate morphology. *Applied Optics, 51*(*5*), 610–620.
Guo, H., Anderson, P. M., & Sunderland, P. B. (2015). A ternary flame system for soot oxidation studies. *Combustion Science and Technology, 187*(*11*), 1836–1840.
Hamada, M., & Weerahandi, S. (2000). Measurement system assessment via generalized inference. *Journal of Quality Technology, 32*(*3*), 241–253.
P.V.C. Hough, Method and means for recognizing complex patterns, 1962.
Hu, B., Yang, B., & Köylü, Ü.Ö. (2003). Soot measurements at the axis of an ethylene/air non-premixed turbulent jet flame. *Combustion and Flame, 134*(*1–2*), 93–106.
Kacker, R., Datla, R., & Parr, A. (2002). Combined result and associated uncertainty from interlaboratory evaluations based on the ISO guide. *Metrologia, 39*, 279–293.
Kholghy, M. R., Afarin, Y., Sediako, A. D., Barba, J., Lapuerta, M., Chu, C., & Thomson, M. J. (2017). Comparison of multiple diagnostic techniques to study soot formation and morphology in a diffusion flame. *Combustion and Flame, 176*, 567–583.
Kock, B. F., Tribalet, B., Schulz, C., & Roth, P. (2006). Two-color time-resolved LII applied to soot particle sizing in the cylinder of a diesel engine. *Combustion and Flame, 147*(*1–2*), 79–92.
Kondo, K., Aizawa, T., Kook, S., & Pickett, L. (2013). *Uncertainty in sampling and TEM analysis of soot particles in diesel spray flame. (SAE Technical Paper No. 2013-1-908)*. Warrendale, PA: SAE International.
Kook, S., Zhang, R., Chan, Q. N., Aizawa, T., Kondo, K., Pickett, L. M., & Nordin, E. Z. (2016). Automated detection of primary particles from Transmission Electron Microscope (TEM) images of soot aggregates in diesel engine environments. *SAE International Journal of Engines, 9*(*1*), 279–296.
Köylü, Ü., & Faeth, G. M. (1992). Structure of overfire soot in buoyant turbulent diffusion flames at long residence times. *Combustion and Flame, 89*(*2*), 140–156.
Köylü, Ü., Faeth, G. M., Farias, T. L., & Carvalho, M. G. (1995). Fractal and projected structure properties of soot aggregates. *Combustion and Flame, 100*(*4*), 621–633.
Lall, A. A., & Friedlander, S. K. (2006). On-line measurement of ultrafine aggregate surface area and volume distributions by electrical mobility analysis: I. Theoretical analysis. *Journal of Aerosol Science, 37*(*3*), 260–271.
Lapuerta, M., Martos, F. J., & Expósito, J. J. (2013). Morphological characterization of diesel soot agglomerates based on the Beer–Lambert law. *Measurement Science and Technology, 24*(*3*), 035405.
Park, C., Huang, J. Z., Ji, J. X., & Ding, Y. (2013). Segmentation, inference and classification of partially overlapping nanoparticles. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(*3*) (1–1).
Park, K., Kittelson, D. B., Zachariah, M. R., & McMurry, P. H. (2004). Measurement of inherent material density of nanoparticle agglomerates. *Journal of Nanoparticle Research, 6*(*2*), 267–272.
Solaguren-Beascoa Fernández, M., Alegre Calderón, J. M., & Bravo Díez, P. M. (2009). Implementation in MATLAB of the adaptive Monte Carlo method for the evaluation of measurement uncertainties. *Accreditation and Quality Assurance, 14*(*2*), 95–106.
Sorensen, C. M. (2001). Light scattering by fractal aggregates: A review. *Aerosol Science and Technology, 35*(*2*), 648–687.
Toth, P., Farrer, J. K., Palotas, A. B., Lighty, J. S., & Eddings, E. G. (2013). Automated analysis of heterogeneous carbon nanostructures by high-resolution electron microscopy and on-line image processing. *Ultramicroscopy, 129*, 53–62.
Vander Wal, R. L., Ticich, T. M., & Stephens, A. B. (1999). Can soot primary particle size be determined using laser-induced incandescence? *Combustion and Flame, 116*(*1–2*), 291–296.
Wang, C., Chan, Q. N., Zhang, R., Kook, S., Hawkes, E. R., Yeoh, G. H., & Medwell, P. R. (2016). Automated determination of size and morphology information from soot transmission electron microscope (TEM)-generated images. *Journal of Nanoparticle Research, 18*(*5*), 127.
Williams, T. C., Shaddix, C. R., Jensen, K. A., & Suo-Anttila, J. M. (2007). Measurement of the dimensionless extinction coefficient of soot within laminar diffusion flames. *International Journal of Heat and Mass Transfer, 50*(*7–8*), 1616–1630.
Wozniak, M., Onofri, F. R. A., Barbosa, S., Yon, J., & Mroczka, J. (2012). Comparison of methods to derive morphological parameters of multi-fractal samples of particle aggregates from TEM images. *Journal of Aerosol Science, 47*, 12–26.