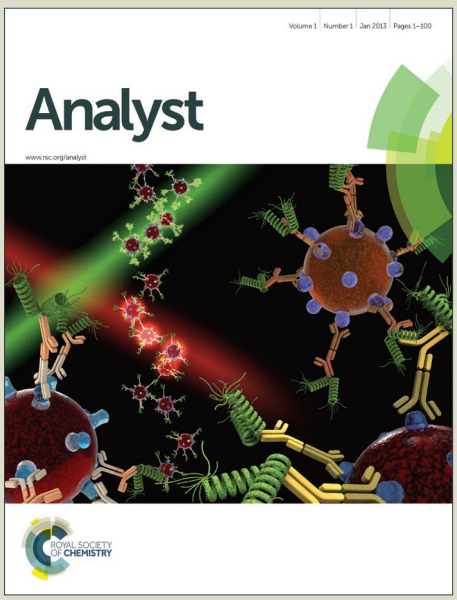
Analyst

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/analyst

Analyst

RSCPublishing

PAPER

Focus-Independent Particle Size Measurement from Streak Images: A Comparison of Multivariate Methods

Cite this: DOI: 10.1039/x0xx00000x

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/analyst

Shawna K. Tazik,[a] Megan R. Pearl,[a] Cameron M. Rekully,[a] Nicholas S. Viole,[a] Stephanie A. DeJong,[a] Timothy J. Shaw,[a] Tammi L. Richardson,[b] and Michael L. Myrick*[a]

Our laboratories have recently developed a flow-through imaging photometer to characterize and classify fluorescent particles between 3 and 47 µm in size. The wide aperture of the objective lens (0.7 NA) required for measuring spectral fluorescence of single particles restricts the depth of field, such that a large sample volume results in many particles that are out of focus. Here, we describe numerical methods for determining the size of these objects, regardless of their distance from the focal plane, using image processing and multivariate calibration. An intensity profile is extracted from the images and is used as the input for a variety of calibration methods, including partial least squares, neural networks, and support vector machines. The capabilities of these methods are examined to establish the best method for particle sizing that is independent of focus. We found that support vector machines provided the best results, with size estimation error of ± 3.1 µm.

Introduction

Determining the size of particles with accuracy and efficiency is important in a wide variety of scientific applications. In pharmaceuticals, particle size plays a large role in the effectiveness of drug delivery.[1 2 3] The bioavailability of a drug can often be increased with smaller particle sizes, particularly for low solubility drugs.[1 2] In the fuel industry, it is valuable to control the size distribution of pulverized fuels to optimize combustion.[4 5] Particle size has also been found to be a factor affecting health risks associated with the inhalation of particulate aerosols[6 7] and to influence the effectiveness of detection of trace explosives.[8] In oceanography, the size distribution of pollutant microplastics[9] and the size spectrum of sinking particles like single-celled photosynthetic algae (phytoplankton) are among topics studied. Our laboratories are interested in the latter application, where the size of the organism determines the rate at which it sinks from the surface ocean, a phenomenon related directly to the transport of carbon dioxide to the deep ocean and, thus, global climate.[10]

There are many existing methods for particle sizing. Classical light scattering, one of the more popular methods in phytoplankton size analysis, estimates cell size based on the angular dependence of light scattering.[11 12] This approach is common in flow cytometry-based instruments. Imaging provides real-space information about individual cells; the FlowCAM[13] and the Imaging FlowCytobot[14 15] are both instruments that employ imaging in a cytometry environment. Another imaging method for *in situ* plankton analysis, called the Video Fluorescence Analyzer (VFA), images fluorescent particles with a small numerical aperture to obtain a large depth of field and is suited to large, bright organisms.[16] Most rapid imaging methods used to extract particle size are arranged so that the optical depth of field is smaller than the sample volume thickness, usually by either hydrodynamic focusing or using a very small numerical aperture collection lens or both.

Thick sample volumes and high numerical aperture optics increase sample throughput and optical efficiency, but at the expense of sharp focus throughout the sample depth. Yule et al. and Kashdan et al. describe an image analysis method for determining particle sizes when images are out of focus and applied it to droplet analysis in two-phase flows.[17 18 19] In brief, their method uses adaptive two-level thresholding of an object's intensity profile to determine the degree of focus and to give an estimate of particle size. A recent multiple-threshold method described by Ju et al. was reported to increase the detectability of smaller particles.[20] In Ju et al.'s method, the apparent area of a particle is calculated at many thresholds to form a curve describing the area of the particle as a function of its intensity. This curve is then compared to a calibration database via curve matching to estimate a particle's size and distance from the focal plane. One key to the operation of this method is that the curves obtained are relatively unique, since the total absorption is a function of the particle size, and all particles studied had similar absorption characteristics.

While in most cases size alone is not sufficient to distinguish one species or class of phytoplankton from one another, it is an important variable in multivariate class discrimination[15] and in understanding the complex interactions that occur in marine ecosystems. The size of the phytoplankton cells impacts a variety of physiological and ecological processes such as metabolic rate, nutrient uptake, nutrient

diffusion, light absorption, and trophic interactions.[10] [21] The overall effect of community size structure in the marine environment is dependent on the taxonomic character of the community.[21] We recently developed a novel type of filter photometer that measures spectral fluorescence from single phytoplankton cells in a flowing stream for the purpose of taxonomic classification.[22] [23] [24] This filter photometer uses high numerical aperture optics for good fluorescence sensitivity, as well as a large sample volume and high flow rates to provide robust sampling statistics for phytoplankton in seawater. Consequently, we obtain images of phytoplankton that vary in focus quality and that form long streaks in our images. While particle shape may be difficult to obtain from these streaks, at a minimum we would like to obtain the apparent particle size from the streak profile. Since the fluorescence intensity of phytoplankton varies widely even for cells in a single monoculture[25], it was not clear whether the curve matching approach of Reference 20 would work well for estimating cell sizes when only the relative or normalized image profile of a particle (and not its absolute amplitude) contains information on its size.

In the present work we apply several forms of multivariate data analysis including curve matching to a multi-threshold fluorescent microsphere calibration set recorded with our imaging photometer to develop models for particle size. Compared to the multi-threshold work done by previous investigators, we use an objective lens with a higher numerical aperture (0.7 vs $\leq$ 0.14) and a shorter working distance (1 mm vs 150 mm) objective to capture as much fluorescence as possible. Our lens has a calculated depth of field of 1.1 µm (versus 28 µm for the lens in Reference 20), but we calibrated over a 120 µm range of focal depths. Performance of the various multivariate analysis tools is evaluated to identify the best calibration approach for particle size determination in this regime. As we report below, curve matching performs poorly on normalized fluorescent microsphere streaks, with support vector machine regression performing best. In the latter case, the performance of the calibration begins to approach the theoretical limits determined from the distribution of particle sizes for the fluorescent microspheres in our study.

Experimental

Calibration Particles

Fluorescent microspheres were purchased from Polysciences, Inc. in the following sizes: 3.14 ± 0.09 µm, 6.125 ± 0.165 µm, 10.3 ± 0.4 µm, 18.68 ± 0.73 µm, 26.93 ± 3.03 µm, 41.68 ± 3.04 µm, and 46.58 ± 3.04 µm. The average sizes and standard deviations of these particles are determined by the manufacturer by disc centrifugation (sizes 3 µm – 9 µm) or single particle optical sensing (sizes > 9 µm).

Image Collection

Fluorescence from single microspheres in a flowing stream was imaged with a custom-built imaging multivariate optical computing (IMOC) photometer.[23] In this system, a CCD array is focused at the center of a flow cell (approximately 200 µm deep) with a 60X objective lens. Radiation from a Xenon arc lamp is used to excite the calibration particles and is modulated by a chopper wheel to create streak images such as the one in Figure 1. Figure 2 presents single streaks obtained of three different particles of the same size class. It is clear that images of particles in the same size class can appear dramatically different depending on their positioning relative to the focal plane.

To apply the multi-threshold and curve matching procedures for particle sizing, a calibration database including images of calibration particles at different known depths of focus was required. This information could not be obtained with the set-up described above because it is not feasible to control the particle positions in the flow cell. Instead, the fluorescent calibration standards were adhered to a glass coverslip that had the same thickness as the face of the IMOC flow cell. A series of images was then collected for a single particle repositioned in 10 µm increments to depths of 100 µm on either side of the point of best focus. This procedure was repeated for three different particles in each size class. Streak images like the ones captured of particles in flow were simulated for each of the still particle images based on information known about the chopper wheel and pump speeds.

In most cases, the profiles of particles off the plane of best focus were found to be reasonably simple over a range of displacements, corresponding to increasing blur as displacement from the best-focus plane increased. From 100 µm inside the position of best focus to 20 µm beyond it, the profiles were rounded peaks with simple first and second derivatives. However, for displacements beyond about 20 µm further from the objective than the position of best focus, profiles for particles became consistently distorted into a peak-on-plateau shape. This type of distortion was difficult for all calibration methods to model effectively, so we confined our analysis to particles in the 120 µm range from the front of the flow cell to 20 µm beyond its center by manually discarding particle streaks displaying the unusual plateau profile (in the case of ordinary multivariate modeling) or only using still images that did not display the unusual profile to simulate streak images (for curve matching).

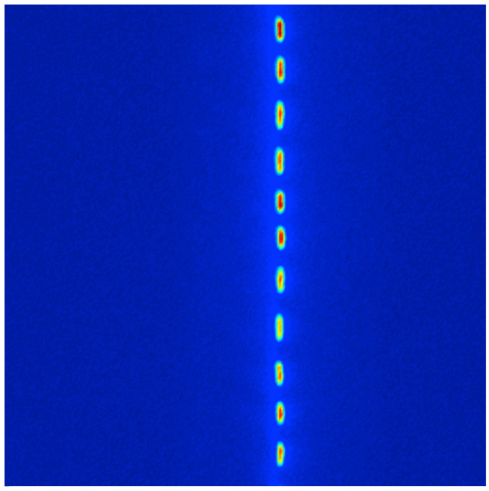


Figure 1. Streak image of a single 6.125 µm fluorescent microsphere collected with the IMOC photometer. Each fluorescent event is referred to as a streak and the collection of streaks is called a track, e.g., this image contains one track and eleven streaks.

Image Processing

The original and simulated streak images of the calibration particles were processed with an updated version of an in-house algorithm previously described by Swanstrom et al.[23] In the algorithm, images are first background-subtracted and flat-field

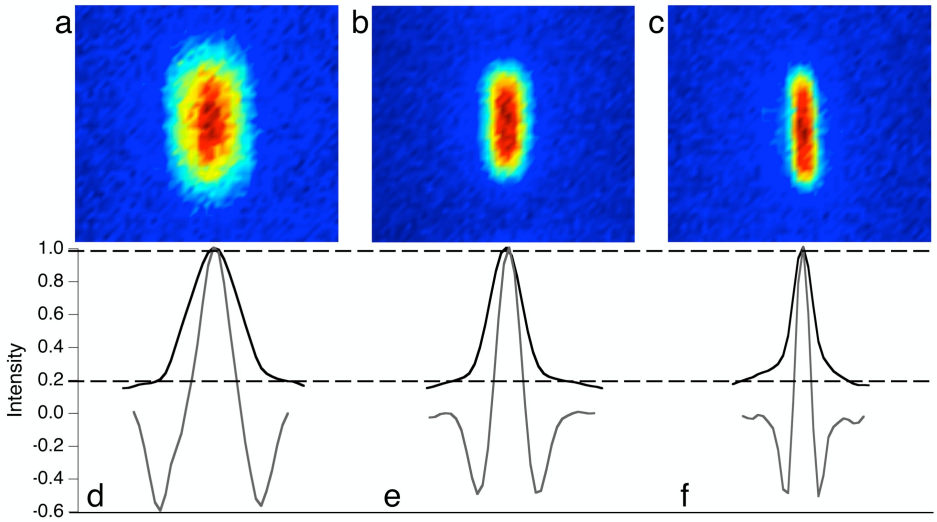


Figure 2. Single streaks from images captured in-flow of three different 10.3 µm particles at different degrees of focus: a) very out-of-focus, b) slightly out-of-focus, and c) in-focus. The bottom panel (d-f) contains the normalized intensity profiles (black trace) and profile second derivatives (gray trace) of the corresponding streaks in the top panel (a-c). The second derivatives have been inverted and scaled. The widths of the profiles are taken between 0.20 and 0.99 (dashed lines) to create a threshold curve.

corrected. An area of interest (i.e., the track of fluorescence streaks given by a single particle) is then located and the individual streak boundaries are determined. An intensity profile for each streak in a track was calculated and normalized to the maximum intensity (see Figure 2 d-f). The normalized intensity profiles of each streak in the track of a single particle were then averaged to give one representative profile for every particle passing through the field of view.

Data Preprocessing

For the application of Ju et al.'s multi-threshold and curve matching methods, thresholds were defined at every hundredth of the profile maximum from 20% to 99% of the maximum intensity and the width at each was calculated in detector units (pixels). The resulting 80 profile widths form a single reference vector of data on which modeling[26] or curve matching can be based. The profile vectors calculated from simulated streak images for a number of particles were used to build the calibration database needed for curve matching. Profile vectors calculated from particles imaged in-flow were used for multivariate calibration procedures. Figure 3 shows these curves for every size and sample used in the following calibrations.

For multivariate modeling, the calibration profiles were processed using a number of standard pretreatments, and models were created using several common methods. A second derivative pretreatment was applied using the gap derivative method[27 28 29 30] with a gap size equal to 1/4th of the full width at half maximum (FWHM) of the profile.[31] Figure 2 displays the differences between the original intensity profile and the second derivative for a fluorescent microsphere with a diameter of 10.3 µm at various positions with respect to the focal plane. Additional pretreatments included Fourier transformation of each profile and of each derivative profile.

A calibration set was compiled using images of 25 different microspheres from each of the seven size classes. Prior to compilation of the calibration set, outliers from each group were eliminated using Hotelling's T-squared test statistic with a criterion of 95% confidence. In order to make effective comparisons between the methods described below, each method was tested on the same set of sample profiles.

Methods

Figures of merit. The cross-validation (CV) error was used to compare the performances of each prediction technique. For partial least squares (PLS), support vector machines (SVM), and any models based on a single full width, the root mean square error of cross-validation (RMSECV) was calculated as

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad (1)$$

where y_i is the actual size for the i^{th} sample left-out, $\hat{y}_i$ is the predicted size with the model using the remaining samples, and N indicates the number of samples left out during CV.

In our neural network (NN) analysis, the RMSECV was calculated as

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{\sum_{j=1}^{M}(y_i - \hat{y}_{ij})^2}{M}\right)}{N}} \quad (2)$$

where y_i is as described above, $\hat{y}_{ij}$ is the predicted size of the N left-out samples with the network optimized with the j^{th} set of M validation samples More detail on CV for NNs is provided below.

Because the multi-threshold approach uses simulated calibration samples, CV was determined to be an invalid method for estimating performance on real samples. For this reason, the errors obtained when testing the calibration database on the same set of samples used in other size estimation procedures was used in method assessment and comparison. The result is best described as an error of prediction. This root mean square error of prediction was calculated as

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad (3)$$

where y_i is the actual size for the ith sample tested, $\hat{y}_i$ is the estimated size, and N indicates the number of samples tested. Although this has the same form as RMSECV above, the definition of $\hat{y}_i$ is distinct.

Models with 0 or 1 independent variable. The profile vector for each particle allows for direct estimation of the particle size by using the system magnification to convert from the apparent size in detector units to a size in µm without requiring curve matching or numerical calibration. One approach is based on the FWHM of the profile and its equivalent size in the object plane. Estimates can also be made from the full width at any other threshold level, and all of these 80 possible estimates of particle size (from 80-point threshold values) involve no variables or modeling.

A better estimation of size can be made by fitting the actual calibrated particle diameters to the full width at any threshold level to develop a univariate model. The simplest of these univariate models can include offsets, linear or higher polynomial terms; we used models no more complicated than second order polynomials. With 80 threshold values and 3 classes of fitting functions, yielding 240 possible univariate models.

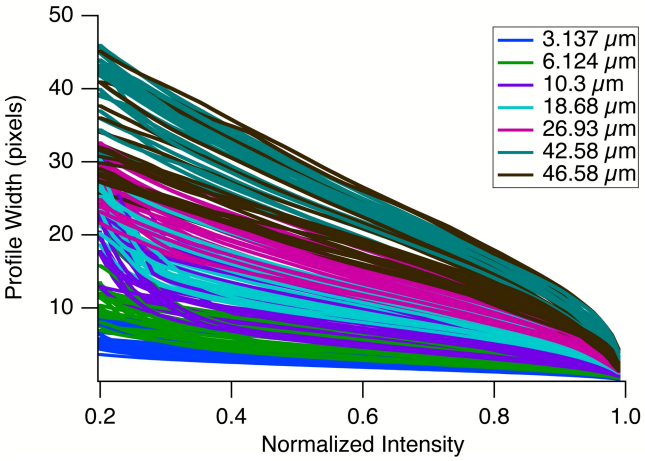


Figure 3. Threshold curves for every particle used in calibration. The profile width (pixels) is represented as function of normalized intensity measured at intervals of 0.01 between 0.20 and 0.99.

Multivariate models. Although estimates can be obtained with a univariate approach, the problem of estimating size from a blurred image is a multivariate problem because blur spreads information about the particle across multiple image channels. We show below that this problem is inherently nonlinear as well. Multivariate models can be created using some or all of the profile widths at the defined threshold levels in our data. In all models tested here, the same 80-point base data vectors for each particle were used. Two different classes of multivariate analyses were tested: conventional chemometric modeling with classic methods of pretreating the data to make predictions of particle size, and curve matching.

The curve matching method described by Ju et al. was developed as an improvement on an earlier dual threshold approach, and was applied in their report to backlit particles and droplets. As the name suggests, their multiple threshold method defines as many thresholds as possible in the particle intensity profile and determines a particle area at each threshold. The result is a curve of apparent particle area as a function of intensity. The curves of unknown particles are then compared to a calibration database via curve matching to estimate the size of the unknown particle. Of the curve matching techniques tested, Ju et al. report that continuous dynamic time warping (CDTW) produced the best results for their particle/droplet image analysis (PDIA) system. However, our implementation of curve matching with IMOC streak images differs from the application to Ju et al.'s PDIA images in several ways by necessity.

The first difference between our implementation and that in Reference 20 is that they measure the apparent area of particles as a function of threshold values, while we only measure streak width. Measuring particle area in the IMOC streak images is not practical due to motion blur in the flow direction. For this reason, only the apparent streak width as a function of intensity is readily measured and used to form a vector for curve analysis.

A second difference lies in normalization. The PDIA images reported in Reference 20 are made in transmitted light for particles with similar absorption characteristics, so the absolute attenuation of light by a particle is relevant in determining its size. Our IMOC photometer measures fluorescence intensity and was designed for application to phytoplankton, where particle size is only one of several factors that influence fluorescence efficiency. Different species of phytoplankton with different pigmentation will generally have different responses for different excitation bands. Even in a healthy monoculture, the fluorescence intensity for individual plankton can vary widely.[25] To remove any bias that absolute fluorescence intensity might introduce, the IMOC streak intensity profiles were normalized and their thresholds were defined at every percent of the peak maximum from 20% to 99% giving a vector of 80 points representing the full width in detector units (pixels) at each threshold level.

A third difference is found in the number of threshold levels being used. Due to normalization, every particle in the analysis reported here is based on a vector using the same number of threshold values. For particles imaged by the PDIA system of Reference 20, the number of thresholds calculated varies as a function of the particle size and its defocus distance.

A final difference between the two approaches is that our samples exhibit substantial motion blur as a result of a long (1 s) CCD exposure time and an average particle velocity of 3.17 mm/s through liquid medium.[23] In the PDIA work, while the particles may be moving at higher velocities of 100 mm/s [20] up to 5000 mm/s,[32] they do not exhibit significant motion blur due to a much shorter exposure time of 100 ns.[18 19 32] In the PDIA system, calibration samples could be created with fixed particles at known distances in air because the resulting calibration measurements should not significantly deviate from those of the intended application to spray particles, which are measured in air and exhibit little to no motion blur. We measured the same type of fixed calibration samples at known distances in air and then used an in-house computer program to simulate streak images from them to form our calibration data for curve matching, which obviously differ in both the surrounding medium and the noise characteristics from real streaks.

Our calibration database for curve matching consisted of all curves that could be calculated for each calibration bead size and defocus depth, ultimately totaling 714 curves that could be matched to any experimental profile. We did not attempt to interpolate to a size between calibration curves as was done in Reference 20 because the particle sizes used to build our calibration database exactly match those of the particles tested. Since we knew *a priori* that any particle tested must belong to one of the available calibration bead size classes, the error statistic for this approach is lower than if we attempted to interpolate sizes between available bead types. Hypothetically,

this should have advantaged the curve matching methods over alternatives.

Three different methods (also tested by Ju et al.) of calculating the distance between curves or the curve similarity were compared here: the Fréchet distance method, the standard deviation method (SDM), and the CDTW method. The Fréchet distance method calculates the maximum "leash length" required to connect a point on one curve to the other.[33] The SDM simply computes distance as the standard deviation of the differences between the curves. Dynamic time warping stretches or compresses the abscissa of a waveform to best match a reference waveform; the best match achievable between a measured threshold curve and a calibration threshold curve is used to determine the size class of the particle in this study.[33] The distance measure used to find the closest calibration curve was computed here as the sum of the squared differences between the stretched/compressed curve and the calibration curve. CDTW was carried out in this study with algorithms obtained from MATLAB Central File Exchange.[34]

Aside from curve matching, three classes of chemometric model types were used to estimate size from measured profiles: PLS, NN, and SVM regression.

PLS is one of the most widely used multivariate calibration techniques today. PLS works by determining the patterns in a vector of calibration variables (e.g., our 80-point intensity profiles) that best correlate with a dependent variable of the calibration set (e.g., size). PLS was carried out here with the SIMPLS algorithm[35] using the PLS toolbox 6.7.1 (Eigenvector Research Incorporated, Wenatchee, WA). Leave-one-out-cross-validation (LOO-CV) was chosen to evaluate models because each measured profile was independent. Different models were generated using the raw intensity profiles and processed profiles as described above as input variables. All variables were mean-centered before PLS calibration. The selection of latent variables for models was based on the first minimum in the plot of the RMSECV versus the number of latent variables used in each model. If the model RMSECV for this first minimum was not significantly better than one with fewer latent variables using the method of Fearn[36], the similar model with the fewest latent variables was chosen.

NNs are a pattern recognition technique often applied to classification and regression problems.[36 38 39] A NN consists of a predefined architecture connecting inputs to outputs through one or more layers of intervening nodes. The nodes apply weighting factors and transfer functions to inputs and pass the result to outputs. Typically, in NN analysis, a set of training data is used to iteratively modify the weights at each node from a randomly assigned starting point until the network outputs best match the reference outputs. Each new training run of a NN can result in different sets of weights and therefore different performances in the optimized network.

Network fitting was implemented using the Neural Network toolbox 7.0.2 (The Mathworks, Inc., Natick, MA). All NNs used in this study were fully connected feed-forward networks trained using back-propagation of errors.[37 38 40] Each NN contained three layers: the input layer, one hidden layer, and an output layer. Sigmoid and linear transfer functions were used for transport between layers. For each set of input data, a network was trained 100 times each for network architectures with 1 to 20 hidden layer nodes. To start, the samples used for NN training were randomly split into training and test sets. Within the training set, a small subset of samples, called the validation set, were used as an internal test set to decide when network training should stop. Training ended when the errors within this validation set converged to a minimum. The final network is then applied to the test set to give an error of prediction.

Implementing CV in NNs is less straightforward than in the other multivariate methods discussed here because each new training run can result in a very different network because the starting point is randomized each time. To avoid this characteristic influencing NN CV results, LOO-CV was implemented in NN analysis with a method adapted from the one described by Burden, et al.[41] In this procedure, the best network was chosen from 2000 trained to produce the lowest prediction error for the test (i.e. external validation) set. The initial weights and biases of this network were then used to seed a LOO-CV routine. For every sample left out, a network was trained multiple times with this same set of starting weights and biases; each time a new set of validation samples was chosen for use in network monitoring. Equation 2 was used to calculate an RMSECV for this process.

SVMs are learning machines that were first described by Vapnik.[42] When applied to the regression problem, they work by nonlinearly mapping input data into high-dimensional space via a kernel function. The goal is then to find a function that minimizes both prediction errors and the magnitude of regression coefficients (i.e. the cost function, C). The cost function penalizes objects (e.g., normalized intensity profiles) with errors larger than ε (a value defined by the user). These objects become support vectors that later determine the predictions for new input data. For a more thorough description of SVMs see References 42 - 45. Support vector regression was carried out in this study with the PLS toolbox 6.7.1 using a radial basis kernel function. Values of C and ε are optimized within the toolbox during CV along with the value of g, a parameter of the kernel function. All sets of input variables and references were mean-centered before SVM training and the LOO method was used for CV.

Results and discussion

Uncertainty in reference values

Overfitting became a concern during these studies. In multivariate calibration, overfitting refers to the modeling of noise or other random attributes specific to the training data.[46 47 48] In general, overfitting occurs when a model becomes too complex, i.e. including too many latent variables in PLS, too many layers or nodes in NNs, or too many support vectors in SVM. This often results in apparently good calibrations that perform poorly during prediction of future samples.[48] A prediction set or an internal test set (cross-validation) is typically used to prevent the choice of a model that is overfitting.[49]

Overfitting is not really a concern with the simple full width models (and, of course, with the "no-model" methods) because there are few or no variables. With multivariate methods like PLS, NN and SVM, steps were taken to prevent selection of overly complex models that might cause overfitting.

Despite taking these steps to prevent overfitting, the authors realized early in this study that some methods remained capable of overfitting the calibration data, resulting in anomalously low apparent errors from the mean for one or more particle sizes. This generally happened when there was an "orphan" size – a single particle size in a wide particle size range. As an example, when the largest particle size studied was nominally 47 μm diameter and the next largest was 20 μm diameter, there was little overlap in the profiles of the largest particle size with

others. Methods like NN or SVM often "recognized" these profiles as belonging to the "47 μm class of particles" instead of making a real estimation of the particle size. This problem was removed by introducing more particle sizes in the calibration. In addition, we introduced a test to ensure overfitting was not occurring. The test works by rejecting any models that cause one or more particle sizes to show an apparent RMSEC that is lower than sampling statistics should give with 95% confidence. This also tells us when a method is approaching the limits of what is possible.

The apparent RMSECV given by our calibration methods is a combination of modeling and other errors. The two most important other errors are those caused by digitization of the image into pixels and the random variability of the calibrated particle sizes. Of these, the ± 0.5 μm uncertainty imposed by the digital images is small compared to the ± 2 μm uncertainty due to the variance of the particle sizes themselves. For this reason, the theoretical lower limit to the precision of our measurements was taken as that determined by the particle sizes alone.

The average particle size of each standard reported by the manufacturer was used as the reference value in developing each calibration model and, more importantly, in assessing their performances. But these standards are not monodisperse; they contain a distribution of particle sizes also characterized and reported by the manufacturer with a standard distribution. We use the mean size of the distribution as the calibrated size value for each particle since we do not know the individual particle sizes; therefore even a hypothetical "perfect" method for determining size should have an apparent error equal to the root mean standard error of the diameter of particles actually sampled from the standard.

In order to assess the merits of a method as well as to detect possible overfitting, a lower error threshold was estimated for the overall method performance and performance within each individual size class. An individual size class threshold was calculated as the lower 95% confidence interval of the sample standard deviation assuming a normally distributed population. Kendall and Stuart provide an equation for the variance (V_s) expected for the sample standard deviation (s) of N members in terms of the population standard deviation of a parent normal distribution[50]:

$$V_s = \frac{\sigma^2}{2N} \quad (4)$$

If we assume that the mean and standard deviations of the calibration standards reported by the manufacturer are adequately representative of the particle population and that the distribution within the standard approaches normality we can estimate an uncertainty in the sample standard deviation of those particles imaged and used in calibration. Since for large numbers of samples the distribution for the variance of the second moment should also approach a normal distribution, the 95% confidence interval for the sample standard deviation was used to establish a lower threshold for the errors expected within a standard size (Table 1). These thresholds are meant to serve only as a rough guide to help determine when a method is approaching the theoretical limits of accuracy and also to detect potential overfitting. The latter is of particular importance in this study because we were unable to assess future performance with a fully external validation set due to a limited number of particle sizes and batches in the study. If a method performs noticeably better for any particle size than the lower threshold values estimated in Table 1, we assess that it is likely overfitting the input data and is not likely to make accurate predictions for future samples.

Method Performance

Figure 4 depicts the performance of each method described above (RMSEC for full width predictions, RMSEP for curve matching results, and RMSECV for all others). The solid line in Figure 4 represents the standard deviation in the reference values of the samples and is an upper bound expected for any of our calibration methods. It is the value that would be expected if our calibration returned the average particle size in the study for all measurements. This upper bound (solid line at 16.0 μm) lies at least 5 μm higher than any of the errors calculated for the size prediction methods indicating that every method tested has at least some predictive ability.

The dotted line in Figure 4 represents the lower error bound estimate from statistical considerations described above for CV of all particles in the study. None of the CV errors crossed the lower error threshold indicating, at least initially, that none of the models are likely overfitting the input data. There were three cases in which the methods produced lower CV errors than the calculated individual microsphere RMSE thresholds. Each of these incidents occurred with the 42 μm size class (with a mean diameter of 41.68 μm) which has a reported standard deviation of 3.04 μm and for which we calculated a lower threshold of 2.18 μm for the root mean error in size determination. Since the thresholds are based on normally distributed particle sizes and normally distributed sample standard deviations, and since the threshold violation was relatively minor and occurred for the same particle size but for different calibration methods, it is likely that these apparent violations result not from overfitting but from something peculiar to the 42 μm particle class in our study that violates one of our assumptions about the distributions in a minor way.

Often when comparing results for various multivariate data modeling approaches the analyst faces the task of deciding whether the differences in performance between methods are significant. Pitman and Fearn provide an approach for rendering this judgment when the modeling methods are in the same class (e.g., between different PLS models, or different NN models, but not between PLS and NN models).[36 51 52] We used Fearn's approach to determine when unbiased models using the same calibration method were not significantly different.

Table 1. Statistics for particles sampled. The mean particle sizes and uncertainties are those reported by the manufacturer. The sample standard deviation and 95% CI are calculated as described above for the particles sampled in this study. The totals describe the combined statistics.

	Mean Particle Size (μm)	Reported Uncertainty (μm)	Sample Standard Deviation (μm)	Lower 95% Confidence Interval (μm)
	3.14	±0.09	0.01	0.06
	6.125	±0.165	0.023	0.118
	10.3	±0.4	0.06	0.3
	18.68	±0.73	0.10	0.52
	26.93	±3.03	0.429	2.17
	41.68	±3.04	0.430	2.18
	46.58	±3.04	0.430	2.18
Total	21.92	2.01	0.108	1.80

Full widths. The simplest size estimation methods discussed here were those based on a single full width in the intensity profile of a particle. The first four bars in Figure 4 represent results of full width modeling with bar 1 representing errors based on using the simple FWHM value, scaled for the

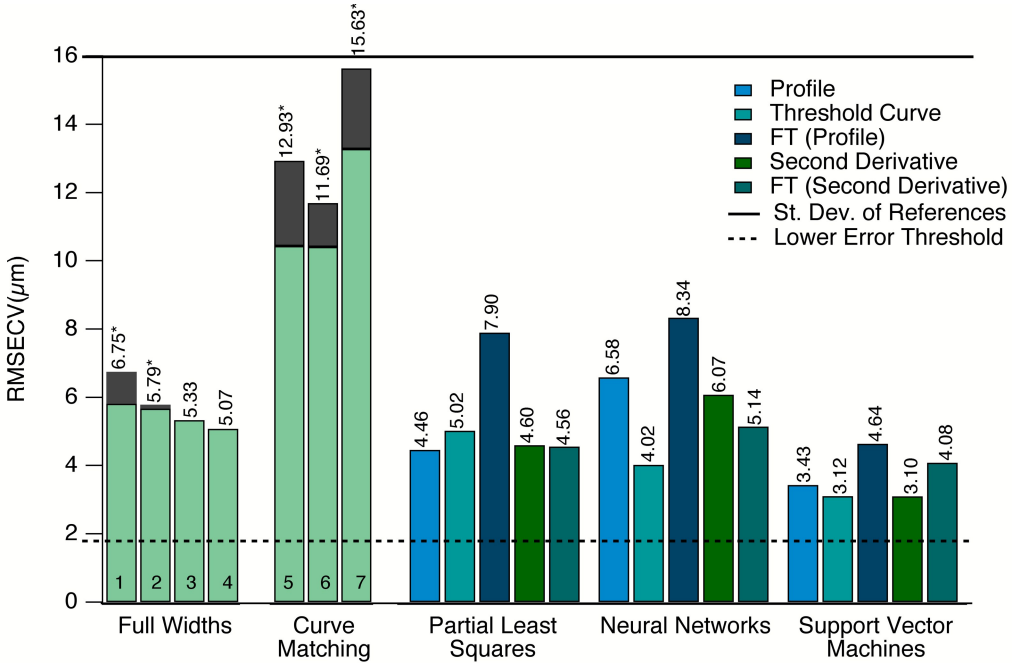


Figure 4. Error results for all size estimation methods. All reported values represent RMSECV unless denoted. RMSEP(*) is shown for methods in which cross-validation was not performed. The black lines represent the standard deviation in the reference values (solid) and the calculated lower error threshold (dashed). For the full width and curve matching methods each bar represents the error results for the following: (1) Predicting size directly from the FWHM of the intensity profile, (2) Fitting a quadratic to the FWHM of the intensity profile (3) Fitting a line to the optimum full width, (4) Fitting a quadratic to the optimum full width, (5) Using the Fréchet distance for curve matching, (6) using the standard deviation method (SDM) for curve-matching, and (7) using the continuous dynamic time warping (CDTW) method for curve matching. For the multivariate calibration methods, each bar represents the error associated with the five different inputs (coloring indicated in figure legend): The intensity profile, the threshold curve, the Fourier transform of the profile, the profile second derivative, and the Fourier transform of the profile second derivative. In some cases, an offset was applied to a size estimation method if it was found to have a significant bias. For these models, the errors without bias correction are reported and denoted by a second dark gray bar.

optical magnification of the system, to directly estimate particle size with no adjustable parameters. Using FWHM provides a biased estimate of particle size with a positive average error, and a standard error of just under 7 μm.

The FWHM for this calculation was a single value taken from the threshold curve of the particle streaks from our instrument. Other single values from points above the half-max are able to provide more accurate estimates due to lower average bias. The best single direct estimator was the full width at 57% max with an error of 5.789 μm (bar 2).

These first two estimators of particle size are not based on mathematical calibration models. The simplest calibration models based on full width use an offset to remove offset bias. These very simple univariate models gave standard errors of 5.807 μm and 5.665 μm when based on FWHM and full width at 57% max, respectively. These values are represented as the 1st and 2nd green bars in the full width section of Figure 4.

Linear (bar 3, based on full width at 80% max with two adjustable parameters) and quadratic (bar 4, based on full width at 83% max with three adjustable parameters) univariate full width models differ mostly in their prediction of small versus large particles. The linear model predicts the small particles better while the quadratic model predicts larger particles better and is better able to correct for some apparent curvature of the full width as a function of particle size.

In conclusion of this section, all univariate models with minor offset biases, or with biases corrected at any level of calibration, gave standard errors less than 1/3rd of the standard deviation in the calibrated particle standard sizes used in this study. As we show below, not all the multivariate methods perform as well, though some perform significantly better.

Curve matching. The curve matching methods described by Ju et al for sizing defocused particles were the least effective size-determination methods for the data presented in this study. The SDM curve matching method produced the lowest error among the alternative curve matching approaches (bar5; 12.93 μm). While the CDTW curve matching method produced the best results for the PDIA curves in Reference 20, it produced an error (bar 7; 15.63 μm) that approaches our upper error threshold. None of the curve matching approaches we tried were competitive with the univariate models above. All models showed some bias, but were still poor when bias was corrected. As described above, the several differences in the way the curve matching method needed modification to work with these data likely explain why it fails so spectacularly in this case.

Multivariate Calibration. The best estimates of particle size in this study came from multivariate calibration models as expected. The apparent improvements in RMSECV are 12%, 21%, and 39% for PLS, NNs, and SVMs, respectively, over the best full width method. Since the variance of the calibrated particle sizes are a source of uncertainty in the calibration that is uncorrelated with the model error, we can estimate the level of model error from the apparent error and the known variance in particle size. We estimate the model errors for multivariate calibration to be improved over univariate calibration by 15%, 25%, and 49% for PLS, NNs, and SVMs, respectively.

The best PLS model used three latent variables and was based on the particle intensity profile (left-most bar in the PLS group); none of the preprocessing approaches we tried

improved on it, and some had significant negative effects. The model based on the Fourier transform of the profile (third bar in the group) was the poorest of the PLS models, and was significantly different from the other PLS models shown in Figure 4. All other PLS models were not significantly different from the model based on intensity profiles, although their RMSECVs were slightly higher.

All the methods used here had difficulty modeling the smallest calibration particles (3.137 μm) compared to other particle sizes. Since the pixel pitch in our camera in the image plane is conjugate to a 1.86 μm spacing between object points, it is likely that the spatial resolution of our measurements is affecting calibration at least on the small end of our scale.

Of the multivariate calibration methods, NNs produced the most models with poor performance. For instance, four of the five NN models represented in Figure 4 were less predictive than the best full width predictor (quadratic model of optimal full width). However, the NN model with the best RMSECV was an improvement on the best PLS models. This model was based on the width of the profile as a function of its normalized intensity – the inverse of the profile – and gave an RMSECV of only 4.018 μm using 3 hidden layer nodes. One peculiarity of the NN models is that they tend to have a large difference between the RMSEC and RMSECV as shown in Figure 6 and this model was no exception, with RMSEC of 2.617 μm. Nevertheless, the difference was lower for this model than for the others shown. For this reason as well as its better RMSECV and the relatively low number of hidden layer nodes, the threshold curve NN model is the best among this group.

PLS is more consistent from calibration to CV predictions than NNs (see Figure 6). The squared correlation coefficient between calibration and CV errors for the same particle is greater than 0.99 for all PLS models. For NNs, this figure varies from 0.18 to 0.74. This indicates that single samples have more effect on the final NN than on the final PLS regression vector. There is a possibility that this behavior is a side effect of the LOO-CV methodology used for NN analysis in this study. Some researchers warn against using LOO-CV with NNs arguing that they will not provide accurate validation results because the models created during CV may deviate substantially from the global model.[40] However, because the starting architecture and weights for LOO-CV are not random but are chosen via tuning on a small prediction set, there is a small likelihood that the network from one left out sample to the next is highly variable or that each network created during LOO-CV is totally different from the global model. A check on calibration versus CV values revealed that one sample was unusually variable from calibration to cross validation, but removing it did not appreciably affect the RMSECV because of the large number of samples being tested.

SVM produced, on average, the lowest RMSECVs of all the methods presented here. Of the 22 methods shown in Figure 4, the 3 best models were all SVM models. This seems to indicate that SVMs are good at extracting the desired information (i.e. size) from a variety of different input forms. The most predictive of the SVMs as judged by RMSECV, as well as the best of all the prediction methods tested was the one built based on the profile second derivatives (RMSECV = 3.098 μm); models based on profiles and threshold curves were not significantly different, however. Of all the SVMs, the second derivative SVM exhibits the largest percent increase in error and the smallest correlation between calibration and CV prediction errors (r^2 = 0.497). The threshold curve model is not significantly different according to Fearn's test,[36] but had RMSECV closer to its RMSEC and an r^2 describing the correlation between calibration and CV errors of 0.874, both desirable statistics that tells us that the individual particles are not strongly affecting the modeling.

Linear methods assume the data can be decomposed into linearly additive components, and that the magnitude of a component varies linearly with the magnitude of a factor. In the problem at hand, there is a defocus blur convolution that interferes with the size factor. Further, since the defocus blur interference does not scale linearly with defocus distance, we anticipated that the calibration problem might be better solved with methods such as NN and SVM. Figure 5 illustrates that PLS is failing to completely compensate for this expected curvature. However, the inherently nonlinear SVM and NN methods we tested were not uniformly better than PLS. A major reason NNs failed to improve on PLS in these studies is likely the relative simplicity of the NN architecture that we utilized in models. SVM methods, on the other hand, were not

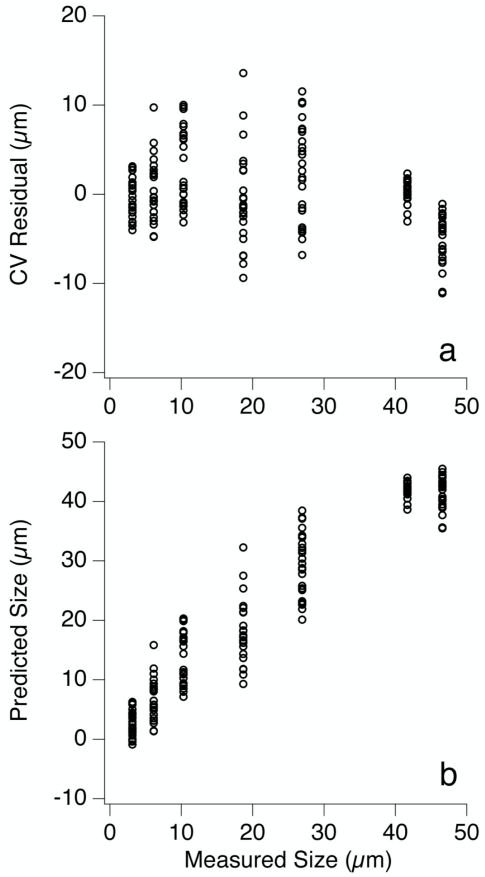


Figure 5. Residual (a) and regression (b) plots for the best PLS model based on the intensity profiles. The curvature to each suggests that there is some nonlinearity that is not accounted for by the PLS model.

constrained in their complexity. Indeed, the performance of our SVM models becomes suspect because a large portion of our samples (>75%) become support vectors in all SVM models, which is a possible indication of overfitting. Also, while SVM is more consistent from calibration to CV than NNs, it is less consistent than PLS, with squared correlation coefficients (between calibration and CV errors) ranging between 0.50 to 0.97. Still, the ability to predict samples not included in the

models via cross validation suggests that despite these apparent defects, the SVM method is not necessarily overfitting our data.

The RMSECV is an acceptable measure of how well a model will predict future samples[53], but may not fully explain how well the models can interpolate because the samples left out are all the same reference size as 24 of the samples used in the model. If the predictions on the left out samples are very poor, one could safely conclude that the model is also likely to be a poor predictor. The converse is not necessarily true: good CV predictions do not necessarily indicate good interpolation ability. We examined the ability of the SVMs to interpolate by leaving the entire 6 µm size class out of the calibration and using it as an external prediction set. This particular size class was chosen because, when left out, it does not leave a large gap in between the two size classes on either side of it. Leaving a large gap in reference values is likely to increase the leverage of the isolated reference sizes and result in a model with little resemblance to the original. As previously indicated, in early stages of this study we found that "orphan" sizes often produced suspiciously low errors. The 6 µm group poses the smallest threat to the overall model integrity, while still allowing us to investigate SVM interpolation ability. It is to be expected that a calibration model would become slightly less predictive with the absence of an entire size class. However, the prediction error of the left out size class should not be much different than the CV error of this size class in the original model. If the two figures of merit differ greatly, it would indicate that the original model is not as predictive as the RMSEC and RMSECV of the original model advertise, and is likely overfitting the data.

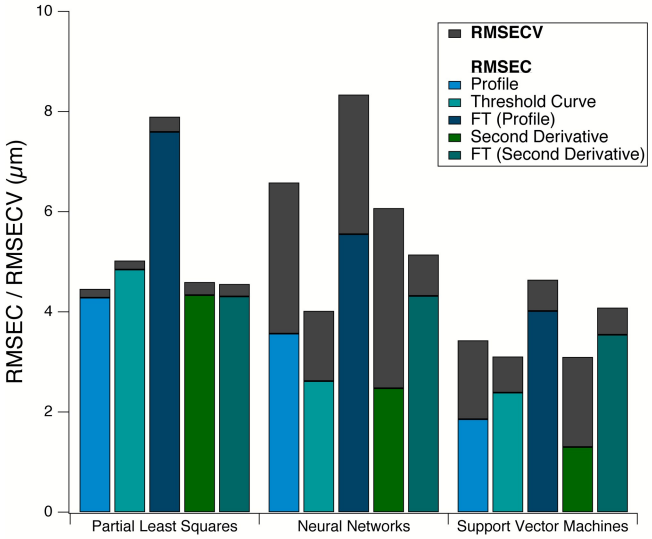


Figure 6. RMSECV (gray bars) and RMSEC (colored bars) for all multivariate calibration methods and all five inputs.

We see a 3 to 5 µm increase in the RMSECV/RMSEP of the 6 µm particles with the Fourier transform models, effectively doubling the error level. This behavior suggests that neither will interpolate well in future cases. The intensity profile and threshold curve SVM methods perform better, with errors actually decreasing somewhat for all other particle sizes and with an RMSEP for the 6 µm particles that was less than 20% larger than the RMSECV when 6 µm particles were included in the models. Second derivative models were not quite as good, with the RMSEP for 6 µm particles being 21% higher than the RMSECV. This increase was also accompanied by an 18% increase in the 3 µm particles. Because the relationships between profiles and diameters do not appear purely linear, some increase in error is expected when a gap is left between particle sizes. However, these increases fall within about 2 standard deviations of the RMSEP based on the use of Equation 4 for RMSEP as given by References 46 and 53 leading to the conclusion that the interpolation ability of the intensity profile, threshold curve, and profile second derivative SVMs are acceptable and that these models are not likely overfitting the data.

Conclusions

There are two competing factors in selecting a method for estimating particle sizes. The greatest emphasis is of course on accuracy; complexity is a competing factor. Generally we find that curve matching on these data gives poor results, and is the most complex of the methods tested in terms of both computation and experimental requirements. The need to record a range of calibration particles at a range of focal positions is undesirable and not feasible without changing the sample holder and optical system. In-house programming of curve matching is also not trivial. On the basis of complexity and poor performance, curve matching is not competitive with the other methods tested here.

Full width methods were less accurate than some other methods but were extremely simple to get working. PLS was somewhat (about 10%) more accurate than the best full width methods. Among calibration methods it had the best internal statistics (comparing calibration and cross validation prediction errors for particles) and worked well when applied directly to particle profiles, but it requires some skill in chemometric analysis to implement with completely in-house software. Overall, the benefits of PLS over the full width methods are debatable.

NN methods are more complex than PLS, more difficult to program internally, and performed relatively poorly when comparing RMSEC to RMSECV for standard particles. A method was found that improved on PLS by an additional ~10%, however.

SVM methods provided the absolute best RMSECV models (3.1 µm); they were significant improvements over NN methods in every way except perhaps being even more complex to incorporate in an in-house program. Their improvement over the best full width methods is about 39%, or even better (49%) when one considered the theoretical floor to the quality of a calibration of around 2 µm.

The SVM model based on threshold curves shows the fewest signs of overfitting among all NN and SVM approaches. It exhibits only a 30% increase between calibration and CV error, which is below the average for all 15 multivariate models. It also shows above average correlation between calibration and CV prediction errors (r^2=0.87). In the test for interpolation ability, it exhibited only a 14% increase in error for the 6 µm references. The computation of the thresholds and the implementation in calibration is also less complex than the profile second derivatives. For these reasons, the SVM model based on the threshold curves was chosen as the most appropriate model to implement in size prediction of future samples.

All models reported here were based on studies of calibrated spheres. Calibrated spheres are the simplest size calibration standards to find for many purposes, but many samples to which models might be applied are not necessarily

spherical. We have not addressed the performance of any of these models on non-spherical particles, in part because good standards are hard to come by. Nevertheless, application to non-standard particles of different shapes seem to suggest that the modeling methods are not all equal in their ability to give reasonable results for oddly shaped particles. Generally, one would expect the simplest methods to be the most bulletproof, while the nonlinear methods might give spurious results.

Acknowledgements

Funding for this study was provided by the National Science Foundation (grant number OCE0958831 to TLR, MLM, and TJS). The authors would also like to acknowledge Professor Stephen L. Morgan for helpful conversations.

Notes and references

[a] Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina 29208, USA.

[b] Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA.

* Author to whom correspondence should be addressed. Email: myrick@chem.sc.edu.

1. B. Y., Shekunov, P. Chattopadhyay, H. H. Y. Tong, and A. H. L. Chow, *Pharm. Res.*, 2007**, 24**, 203-227.
2. J. Sun, F. Wang, Y. Sui, Z. N. She, W. J. Zhai, C. L. Wang, and Y. H. Deng, *Int. J. Nanomed.,* 2012, **7**, 5733-5744.
3. Y. S. Cheng, T. D. Holmes, J. Gao, R. A. Guilmette, S. Li, Y. Surakitbanharn, and C. Rowlings, *J. Aerosol Med.,* 2001, **14**, 267.
4. L. J. Gao, Y. Yan, G. Lu, and R. M. Carter, *Flow Meas. Instrum.* 2012, **27**, 20-28.
5. H. Lu, E. Ip, J. Scott, P. Foster, M. Vickers, and L. L. Baxter, *Fuel* 2010, **89**, 1156-1168.
6. J. S. Lighty, J. M. Veranth, and A. F. Sarofim, *J. Air Waste Manage. Assoc.,* 2000, 50, 1565-1618.
7. A. A. Andersen, *J Bacteriol.,* 1958, **76**, 471-484.
8. J. R. Verkouteren, *J. Forensic Sci.,* 2007, **52**, 335-340.
9. S. L. Wright, R. C. Thompson, and T. S. Galloway, *Environ. Pollut.,* 2013, **178**, 483-492.
10. S. J. Bury, P.W. Boyd, T. Preston, G. Savidge, and N. J. P. Owens, *Deep Sea Res. Part I,* 2001, **48**, 689-720.
11. R. E. Green, H. M. Sosik, R. J. Olson, and M. D. DuRand, *Appl. Opt.* 2003, **42**, 526-541.
12. R. J. Olson, E. R. Zettler, and M.D. DuRand, in *Aquatic Microbial Ecology*, ed. P. F. Kemp, B.F. Shen and J. J. Cole, Lewis Publishers, Boca Raton, FL, 1993, pp. 175-186.
13. C. K. Sieracki, M. E. Sieracki, and C. S. Yentsch, *Mar. Ecol. Prog. Ser.,* 1998, **168**, 285-296.
14. R. J. Olson, and H. M. Sosik, *Limnol. Oceanogr. Methods,* 2007, **5**, 195-203.
15. H. M. Sosik, and R. J. Olson, *Limnol. Oceanogr. Methods,* 2007, **5**, 204-216.
16. M. Lunven, J. M. Landeira, M. Lehaitre, R. Siano, C. Podeur, M. M. Danielou, E. Le Gall, P. Gentien, and M. Sourisseau, *Limnol. Oceanogr. Methods,* 2012, **10**, 807-823.
17. A. J. Yule, N. A. Chigier, and N. W. Cox, in *Particle Size Analysis* Heyden Press, 1978, pp. 61-73.
18. J. T. Kashdan, J. S. Shrimpton, and A. Whybrew, *Part. Part. Syst. Char.,* 2004, **20**, 387-397.
19. J. T. Kashdan, J. S. Shrimpton, and A. Whybrew, *Part. Part. Syst. Char.,* 2004, **21**, 15-23.
20. D. H. Ju, J. S.; Shrimpton, and A. Hearn, *Part. Part. Syst. Char.,* 2012, **29**, 78-92.
21. A. J. Irwin, Z. V. Finkel, O. M. E. Schofield, and P. G. Falkowski, *J. Plankton Res.,* 2006, **28**, 459-471.
22. J. A. Swanstrom, L. S. Bruckman, M. R. Pearl, M. N. Simcock, K. A. Donaldson, T. L. Richardson, T. J. Shaw, and M. L. Myrick, *Appl. Spectrosc.,* 2013, **67**, 620-629.
23. J. A. Swanstrom, L. S. Bruckman, M. R. Pearl, E. Abernathy, T. L. Richardson, T. J. Shaw, and M. L. Myrick, *Appl. Spectrosc.,* 2013, **67**, 630-639.
24. M. R. Pearl, J. A. Swanstrom, L. S. Bruckman, T. L. Richardson, T. J. Shaw, and M. L. Myrick, *Appl. Spectrosc.,* 2013**, 67**, 640-647.
25. L. S. Hill, T. L. Richardson, L. T. M. Profeta, T. J. Shaw, C. J. Hintz, B. S. Twining, E. Lawrenz, and M. L. Myrick, M. L., *Rev. Sci. Instrum.,* 2010, **81**, 013103.
26. M.R. Pearl, Ph.D. Thesis, University of South Carolina, 2007.
27. W. L. Butler, and D. W. Hopkins, *Photochemistry and Photobiology* **1970,** *12*, 439-450.
28. Butler, W. L.; Hopkins, D. W., An Analysis of Fourth Derivative Spectra. *Photochem. Photobiol.,* 1970, **12**, 451-456.
29. W. L. Butler, *Methods Enzymol.,* 1979, **56**, 501-515;
30. J. E. Cahill, *Am.Lab.,* 1979, **11,** 79-85.
31. W. F. Maddams, and W. L. Mead, *Spectrochim. Acta, Part A,* 1982, **38**, 437-444.
32. J. T. Kashdan, J. S. Shrimpton, and A. Whybrew, *Opt Lasers Eng.,* 2007, **45**, 106-115.
33. A. Efrat, Q. F. Fan, and S. Venkatasubramanian, *J. Math. Imaging Vision,* 2007, **27**, 203-216.
34. P. Micó, Continuous Dynamic Time Warping, http://www.mathworks.com/matlabcentral/fileexchange/16350-continuous-dynamic-time-warping, (accessed March 28).
35. S. Dejong, *Chemom. Intell. Lab. Syst.,* 1993, **18**, 251-263.
36. T. Fearn, *NIR news,* 1996, **7**, 5-6.
37. R. O. Duda, P. E. Hart, and D. G. Stork, in *Pattern Classification*, John Wiley & Sons, Inc., New York, 2nd edn., 2001, pp. 282-349.
38. D. Svozil, V. Kvasnicka, and J. Pospichal, *Chemom. Intell. Lab. Syst.,* 1997, **39**, 43-62.
39. J. R. Long, V. G. Gregoriou, and P. J. Gemperline, *Anal. Chem.,* 1990, **62**, 1791-1797.
40. F. Despagne, and D. L. Massart, *Analyst,* 1998, **123**, 157R-178R.
41. F. R. Burden, R. G. Brereton, and P. T. Walsh, *Analyst,* 1997, **122**, 1015-1022.
42. V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
43. A. J. Smola, and B. A. Scholkopf, *Statistics and Computing,* 2004, **14**, 199-222.
44. C. J. C. Burges, *Data Mining and Knowledge Discovery,* 1998, **2**, 121-167.
45. R. G. Brereton, and G. R. Lloyd, *Analyst,* 2010, **135**, 230-267.
46. N. M. Faber, and R. Rajko, *Anal. Chim. Acta,* 2007, **595**, 98-106.
47. C. Schittenkopf, G. Deco, and W. Brauer, *Neural Networks,* 1997, **10**, 505-516.

48. T. Fearn, *NIR news,* 1997, **8**, 7-8.
49. I. V. L. Tetko, D.J. Livingstone, and A. I. Luik, *J. Chem. Inf. Comput. Sci.,* 1995, **35**, 826-33.
50. M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Macmillan Publishing Co., Inc, New York, 4th edn., 1977, vol. 1.
51. E. J. G. Pitman, *Biometrika,* 1939, **31**, 9-12.
52. Cochran, G. W. S. a. W. G., *Statistical Methods*. 8th ed.; Iowa State Univeristy Press: Ames, Iowa, 1989.
53. H. A. Martens, and P. Dardenne, *Chemom. Intell. Lab. Syst.,* 1998, **44**, 99-121.