

Alternative Statistical Methods for Spectral Data Processing: Applications to Laser-Induced Breakdown Spectroscopy of Gaseous and Aerosol Systems

L. A. ÁLVAREZ-TRUJILLO, A. FERRERO, J. J. LASERNA, and D. W. HAHN*

Department of Analytical Chemistry, Faculty of Sciences, University of Málaga, 29071 Málaga, Spain (L.A.Á.-T., A.F., J.J.L.); and Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6300 (D.W.H.)

A new spectral data processing scheme based on the standard deviation of collected spectra is compared with the traditional ensemble-averaging of laser-induced breakdown spectroscopy (LIBS)-based spectral data for homogenous (i.e., pure gas phase) systems and with a LIBS-based traditional conditional spectral analysis scheme for non-homogenous (e.g., aerosol system) analyte systems under discrete particle loadings. The range of conditions enables quantitative assessment of the analytical approaches under carefully controlled experimental conditions. In the homogeneous system with gaseous carbon dioxide producing the carbon atomic emission signal, the standard deviation method provided a suitable metric that is directly proportional to the analyte signal and compares favorably with a traditional ensemble averaging scheme. In contrast, the applicability of the standard deviation method for analysis of non-homogenous analyte systems (e.g., aerosol systems) must be carefully considered. It was shown both experimentally and via Monte Carlo simulations that the standard deviation approach can produce an analyte response that is monotonic with analyte concentration up to a point at which the analyte signal starts to transition from a non-homogeneous system to a homogeneous systems (i.e., around a 50% sampling point for aerosol particles). In addition, the standard deviation spectrum is capable of revealing spectral locations of non-homogeneously dispersed analyte species without *a priori* knowledge.

Index Headings: Laser-induced breakdown spectroscopy; LIBS; Aerosol; Statistical methods.

INTRODUCTION

Laser-induced breakdown spectroscopy (LIBS) is a spectroscopic technique applicable to the elemental analysis of solids, liquids, and gaseous samples, including aerosol-laden systems. Historically, the more common method of data analysis used in LIBS has been the ensemble-average of spectra, resulting from many hundreds to thousands of laser pulses, depending on the requirements of a given experiment, as has been reported in numerous research and review papers.^{1–5} Such an approach greatly improves the signal-to-noise ratio for relatively homogeneous systems due to the pulsed nature of LIBS, which yields spectral fluctuations associated with the variability of the plasma energy on a shot-to-shot basis combined with the very nonlinear laser–material coupling and noise associated with the detection (notably with intensified charge-coupled device (CCD) systems). These noise sources are significantly diminished by spectral averaging (ideally as the square root of the number of averaged spectra), thereby improving the sensitivity of the LIBS technique. This approach, however, reaches a natural limit when there are low analyte concentrations and a non-homogeneous analyte distribution, as is often

the case with aerosol systems due to the discrete nature of individual aerosol particles suspended in a gas matrix.⁶ For example, the particle sample rates (defined as the percentage of pulses for which a particle is present in the plasma volume) for aerosol loading conditions may be on the order of 0.1%, as was verified experimentally under various conditions, including the analysis of ambient air particles.^{6,7} Such a condition will result in a large component of collected spectra with no spectral information regarding the targeted species; hence, the traditional ensemble-averaging of LIBS spectra will produce poor signal-to-noise ratios and consequently poor species detection.

To date many efforts have been made to enhance LIBS sensitivity and precision based on traditional ensemble-averaging of spectra as a means to overcome the extensive spectral fluctuations observed on a laser shot-to-shot basis and subsequent single-shot analysis.^{1,8–10} A comprehensive approach that significantly increases the signal-to-noise ratio, and therefore the analyte detection limits, based on the rejection of all null spectral data using a suitable conditional data analysis has been reported by Hahn et al.^{7,11} The further evolution of laser-induced breakdown spectroscopy as an effective technique for the analysis of aerosol systems, including discrete particle analysis, is dependent on the ability to accurately quantify the presence and concentration of analyte species within individual and small groups of particles. The primary factors challenging LIBS-based aerosol analysis include the fact that small aerosol particles (i.e., in micrometer to nanometer size range) may contain only a few femtograms of analyte, while laser-induced plasmas are known to experience large shot-to-shot fluctuations in combination with significant spectral noise, thereby reducing the associated method detection limit. While the use of conditional analysis schemes is helpful, the specific details of thresholding remain difficult and the subject of research.¹² A new method of spectral data analysis was recently reported by Laserna et al. that takes advantage of the spectral fluctuations associated with LIBS spectroscopic detection in combination with the high variability of analyte signals for non-homogeneous systems, as is the case for aerosol systems.¹³ This method is based on the standard deviation of each pixel (or column of pixels if binning is used) of the CCD detector array associated with a set of collected spectral data. The scheme was demonstrated to be a robust method of real-time data analysis for qualitative information in stand-off LIBS-based aerosol particle detection.¹³ Specifically, the authors were able to produce an analyte signal from remote, individual plasma breakdown events of sodium-rich aerosol particles. The study was primarily designed as a proof of principle, with the limitations toward quantitative analysis, as implemented, noted in the study.

In the current paper, the standard deviation approach is

Received 12 February 2008; accepted 31 July 2008.

* Author to whom correspondence should be sent. E-mail: dwhahn@ufl.edu.

compared with the traditional ensemble-averaging of spectral data for a homogenous (i.e., pure gas phase) analyte system, and with traditional conditional spectral analysis for an aerosol system under discrete particle loadings. The range of conditions enables quantitative assessment of the two approaches under carefully controlled experimental conditions.

EXPERIMENTAL METHODS

Laser-Induced Breakdown Spectroscopy Configuration.

The LIBS system has been described in an earlier publication.^{11,14} The laser-induced plasma was generated using a Q-switched Nd:YAG laser (Big Sky CFR 400) at its fundamental wavelength of 1064 nm, with a 10 ns pulse width, 210 mJ pulse energy, and at a 2 Hz pulse repetition rate. The expanded laser beam (12 mm diameter) was focused at the center of the sample chamber using the 75 mm lens mounted directly on the sample cross. The plasma emission was collected via backscatter using a pierced mirror and focused onto a fiber-optic bundle. The light was then dispersed by a 0.275 m Czerny–Turner spectrometer (Acton Spectra Pro-275) and finally recorded by a 256×1024 element intensified CCD array. The spectrometer used a 2400 lines/mm grating, which provided an approximately 30 nm wavelength range and 0.12 nm spectral resolution.

Analyte Generation. All analyte samples flowed through the standard six-way cross at atmospheric pressure, which functioned as the LIBS sample chamber as described above. For the gas-phase measurements, various flow rates of pure carbon dioxide gas were mixed with a constant co-flow of 42 lpm of purified air, after which the mixture was introduced to the LIBS sample chamber. The purified air co-flow was passed through an activated alumina dryer, which removed nearly all humidity and most of the CO₂, leaving a residual of approximately 50 ppm of carbon dioxide. In addition, all gases were passed through HEPA filter cartridges to remove any particulate matter. Both the co-flow air and the added carbon dioxide flow rates were controlled with mass flow controllers, appropriately sized to the actual mass flow rates. Carbon dioxide flow rates were varied such that the resulting added carbon dioxide concentration varied from 225 to 900 ppm (by volume) in the sample chamber. The carbon dioxide and air experiments will be referred to as the *gas-phase experiments*.

The second set of measurements was made for a micrometer-sized aerosol flow. The aerosol generation system was described in an earlier publication.^{14,15} Briefly, the aerosol particles were introduced by nebulizing a suspension of approximately 2 μm sized (± 0.5 μm) borosilicate glass microspheres (Duke Scientific, #9002) in ultra-purified water. The suspension was nebulized and subsequently mixed with a 42 lpm co-flow of purified air, with the resulting aerosol passed directly to the LIBS sample chamber. Based on previous analysis, the calcium concentration within the glass microspheres was determined to be about 2% (by mass), or an average of about 400 fg of calcium per particle, which yields a strong calcium atomic emission signal.¹⁵ The particle concentration in suspension was adjusted so that the resulting borosilicate particle number density produced in the LIBS sample chamber resulted in a particle sampling rate ranging from about 10% to 50%. The experiments with the glass microspheres will be referred to as the *aerosol particle experiments*.

Spectral Measurements. For all experiments, the external

Q-switch sync from the laser was used to trigger the intensified CCD (iCCD) controller, with the exact detector delay and gate width optimized for each analyte (i.e., carbon or calcium). For gas-phase experiments, carbon dioxide was used to generate atomic carbon, and the neutral carbon (C I) emission line at 247.8 nm was used for spectral analysis. A temporal delay and gate width of 6 μs were used for carbon analysis, as based on previous optimization studies.¹⁶ For each given carbon dioxide concentration, spectral data were acquired in sets of 1000 spectra corresponding to 1000 sequential laser shots. For each concentration, the measurements were repeated a total of four times; hence, final spectral data consisted of four separate 1000-spectra sets. For the gas-phase measurements, the spectral data were processed using two different methods: namely, (1) using the traditional ensemble-average of each set of 1000 spectra acquired for a given concentration of carbon dioxide, and (2) by calculating the standard deviation of each column of pixels (i.e., binned pixels) of the CCD array, for each set of 1000 spectra, as described previously.¹³ Specifically, for each column pixel, the standard deviation (SDV) of the 1000 spectra ($N = 1000$) is calculated as

$$SDV = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (p_i - \bar{p})^2} \quad (1)$$

where p_i is the intensity value of the given column pixel of the i th spectrum, and \bar{p} is the average intensity value of the given column pixel of the 1000 spectra. It is noted that vertical binning was used for CCD readout, thereby producing a single linear spectral array for a given laser shot. The term column pixel will be used to refer to the resulting value of the vertically binned pixels. This method yields a new array in which each pixel value is the standard deviation of the corresponding column pixel values for all N shots. The resulting standard deviation array may then be plotted in the corresponding wavelength space, thereby generating a “spectrum” such that each pixel corresponds to the standard deviation of the actual set of spectra.

For the aerosol phase experiments (i.e., glass microspheres), the analyte peak of interest was the first ionized calcium (Ca II) atomic emission line at 393.4 nm. For calcium analysis, both the detector delay and the gate width were set to 25 μs. For these experiments, spectral data were collected in sets of 3000 spectra corresponding to consecutive laser shots, for a given particle concentration. A total of six different particle concentrations were investigated; hence, the final data corresponds to six sets of 3000 spectra. The spectral data were then processed using two different methods: namely, (1) using the single-shot conditional data analysis as reported by Hahn et al.^{7,11,12} to identify each calcium-based particle hit, and (2) using the standard deviation analysis in which the standard deviation was calculated for each set of 3000 spectra in the same way as described for the gas-phase experiments. Therefore, for each particle concentration, a standard deviation spectrum was produced corresponding to a set of 3000 spectra.

RESULTS AND DISCUSSION

In the present paper, the standard deviation method of spectral analysis is directly compared to two established methodologies, namely, the ensemble-averaging for the homogeneously dispersed analyte associated with the gas-

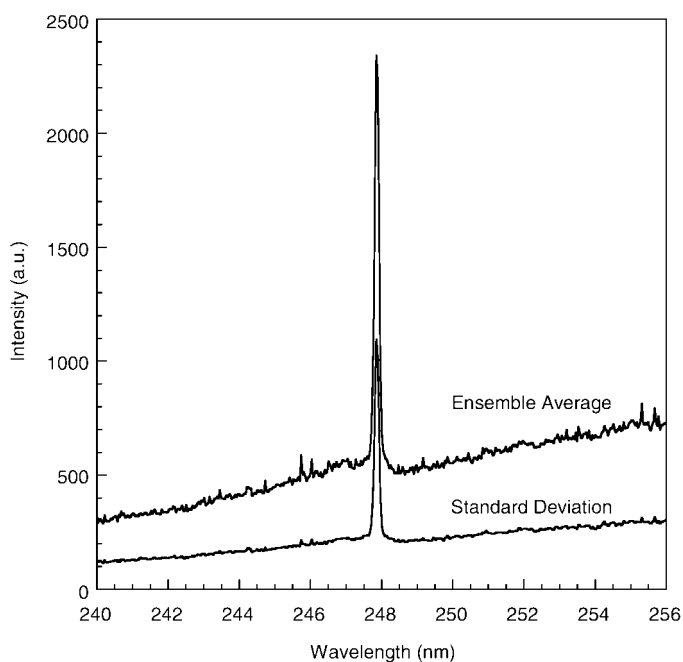


FIG. 1. Spectra showing the carbon atomic emission line at 247.8 nm. The data correspond to ensemble-average (upper spectrum) and standard-deviation (lower spectrum) of 1000 spectra recorded in 450 ppm (volume) of CO₂. Both spectra have the same scale.

phase experiments, and conditional analysis for the non-homogeneously dispersed analyte associated with the aerosol experiments.

Spectral Analysis of the Gas-Phase System. For the gas-phase experiments, the standard deviation method of data analysis was compared with the traditional ensemble-average commonly used in homogeneous or well-mixed analyte systems. As an example, carbon emission spectra for both the ensemble-average and the standard deviation methods are shown in Fig. 1 for the addition of 400 ppm of carbon dioxide, where it is observed that the signal-to-noise ratio (SNR) of the carbon peak looks comparable in both spectra. The SNR will be quantified below, but first additional thoughts are offered regarding the relative nature of spectral noise. As observed in the figure, the standard deviation curve appears to scale with the ensemble-averaged spectrum across the entire spectral window, which suggests a similar correlation in the degree of shot-to-shot variability of both the carbon line intensity (i.e., atomic emission) and the broadband continuum emission intensity (i.e., recombination and Bremsstrahlung emission). To further explore this phenomenon, the relative standard deviation (RSD) of the data was calculated by dividing the standard deviation spectrum by the ensemble-averaged spectrum on a column pixel by column pixel basis across the CCD array. Figure 2 shows the resulting RSD of the Fig. 1 data, in addition to the resulting RSD of the 225 ppm of carbon dioxide measurements. Both RSD curves are observed to have a nearly constant value across the entire spectral window. In fact, the RSD of the carbon emission line deviates by only about 10% from the RSD of the continuum emission. Such behavior implies that the standard deviation follows the actual analyte signal value¹³ and, furthermore, demonstrates that the plasma does not reveal any significant difference between the plasma continuum emission and atomic emission processes for a

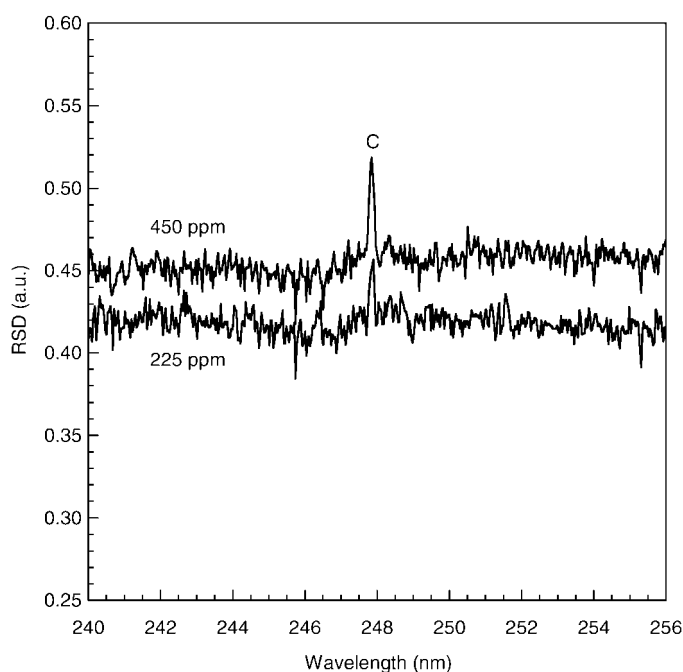


FIG. 2. Relative standard deviation (RSD) as calculated from the quotient of the standard-deviation and the ensemble-average for 1000 spectra recorded in 225 and 450 ppm (volume) of CO₂. The 450 ppm case corresponds to the data in Fig. 1. Both spectra have the same scale but have been offset vertically, and the location of the 247.8 nm C line is labeled.

perfectly homogeneous analyte sample, as is the case in the gas-phase system.

To quantify the analyte emission signal, both the peak-to-base ratio (P/B) and the SNR were calculated using the 247.8 nm C I spectral line for both the ensemble-average and the standard deviation methods, for each added carbon dioxide concentration. The P/B ratio is a metric that provides a precise measure of analyte atomic emission intensity by normalizing to the absolute plasma continuum emission signal as noted previously.^{1,16} For the present analysis the P/B is calculated as the ratio of the integrated atomic emission line intensity (full-width peak area) to the average intensity of the adjacent featureless continuum emission (i.e., baseline). As an analytical figure of merit, the SNR is perhaps the more relevant metric. For this study, the SNR was calculated as the ratio of the integrated atomic emission line intensity (full-width peak area) to the average root mean square (rms) noise of the adjacent featureless continuum intensity (i.e., rms noise) multiplied by the number of pixels summed for the peak area. The rms noise was calculated from the deviation of a least squares fit over two spectral continuum regions corresponding to 50 pixels on either side of the carbon atomic emission line.

The P/B ratio and the SNR were calculated as a function of added carbon dioxide concentration, with the results shown in Figs. 3 and 4, respectively. It is observed in Fig. 3 that both the ensemble-average and the standard-deviation methods provide equivalent values of P/B for each concentration of carbon dioxide, and each of the resulting calibration curves results in correlation coefficients (*R*) near unity, namely, greater than 0.999.

Error bars represent the experimental variability of the P/B over the four 1000-shot sequences at each concentration. The P/B of the standard deviation signal scales with the P/B of the

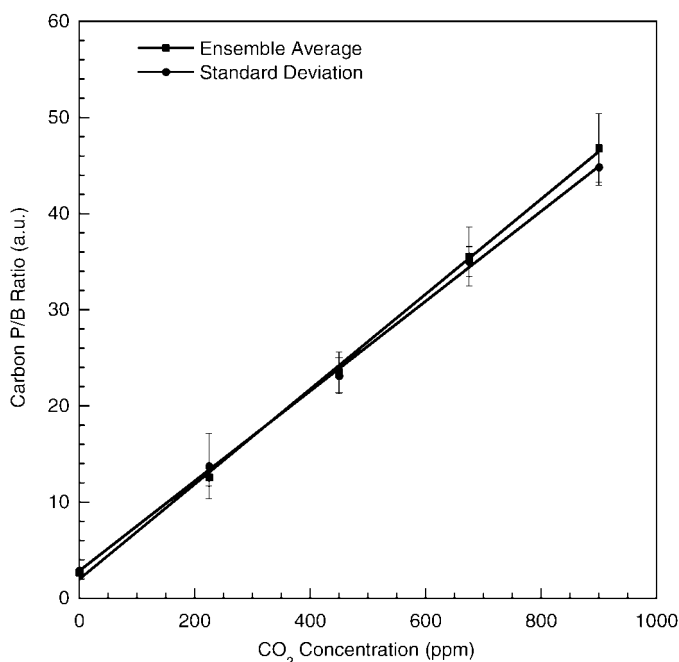


FIG. 3. Calibration curves based on the peak-to-base (P/B) ratio of the 247.8 nm C I atomic emission line intensity as a function of added carbon dioxide concentration for both the ensemble-averaged data and the standard-deviation data. Linear least squares fits are included for both data sets, with regression coefficients >0.999 for both curves. Errors bars represent one standard deviation of the P/B value based on four independent measurements of 1000 spectra each.

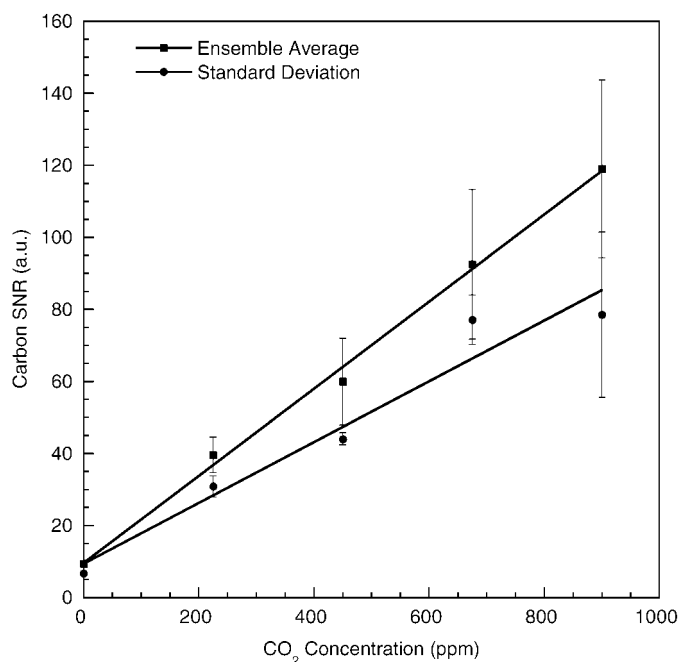


FIG. 4. The signal-to-noise ratio (SNR) of the 247.8 nm C I atomic emission line intensity as a function of added carbon dioxide concentration for both the ensemble-averaged data and the standard-deviation data. Linear least squares fits are included for both data sets, with regression coefficients equal to 0.998 and 0.975 for the ensemble-averaged and standard-deviation data sets, respectively. Errors bars represent one standard deviation of the SNR value based on four independent measurements of 1000 spectra each.

ensemble-average spectrum because the standard deviation is directly proportional to the magnitude of the respective column pixel. As is apparent in Eq. 1, a linear increase in column pixel value directly scales to a linear increase in the corresponding standard deviation so long as the overall distribution of the analyte population remains constant. This is certainly the case for a gas-phase analyte species. Clearly, quantification of the standard-deviation based spectrum is reflective of the analyte signal, as based on the actual ensemble-averaged spectrum, thereby establishing equivalency between the two approaches for a perfectly dispersed analyte. One additional comment regarding Fig. 3 is noted, namely, the non-zero y -intercept value. The independent variable represents the concentration of *added* carbon dioxide but does not reflect the ~ 50 ppm of residual carbon dioxide in the purified air co-flow stream. This residual carbon results in the offset of the absolute carbon signal; hence, the y -intercept is consistent with ~ 50 ppm carbon dioxide at zero added carbon dioxide.

Consideration of the SNR data in Fig. 4 reveals a slight difference in slope and precision between the SNR values calculated for the two methodologies. It is clear in Fig. 4 that both the ensemble-average and the standard-deviation based SNR values follow a linear trend with the analyte concentration; however, unlike with the P/B ratio, the slopes are slightly different, as are the correlation coefficients (0.998 versus 0.975 for the ensemble-averaged and standard-deviation data, respectively).

The above results (Figs. 3 and 4) demonstrate that the standard deviation of a set of spectra is essentially equivalent to the ensemble-average of the same spectral data set (as quantified by the SNR and P/B ratios) for such a homogeneous analyte sample as realized with the carbon dioxide and air

mixtures. It is noted that the standard deviation approach captures the fluctuations in the detection of a spectroscopic signal in a set of spectra recorded, although these fluctuations do not reflect only the variability of the analyte signal. The calculation takes into account the fluctuation in the intensity level of each spectrum recorded, which includes fluctuations due to the photon shot noise inherent in any spectroscopic measurement due to the quantum nature of light. The shot noise depends of the number of photons that arrive at the CCD array, such that the related root mean square variation is directly proportional to the root square of the number of photons incident on the detector. Due to additional shot-noise related fluctuations, each single spectrum recorded has a slightly different intensity level over the spectral window detection array; hence, the peak area of the analyte is not constant. Such fluctuations in the spectroscopic signal associated with the analyte of interest appear in the spectrum resulting from both the standard deviation method as well as ensemble-averaging. To provide some additional insight into the nature of spectral noise for the two techniques, Fig. 5 presents a plot of the log of the SNR as a function of the log of the absolute analyte signal for the Fig. 3 and Fig. 4 carbon dioxide data, as detailed by Ingle and Crouch.¹⁷ Both curves show a linear response with a slope near unity (slope = 0.96 and 1.09 for the ensemble-average and standard-deviation data, respectively). A slope of unity corresponds to blank-noise dominant spectral data, which holds for both the ensemble-averaged and the standard-deviation data. This is consistent with the Fig. 2 calibration curves and related discussion, namely, the equivalence of analyte response under these conditions.

Spectral Analysis of Aerosol Particle Experiments. For the aerosol particle experiments, the standard deviation method

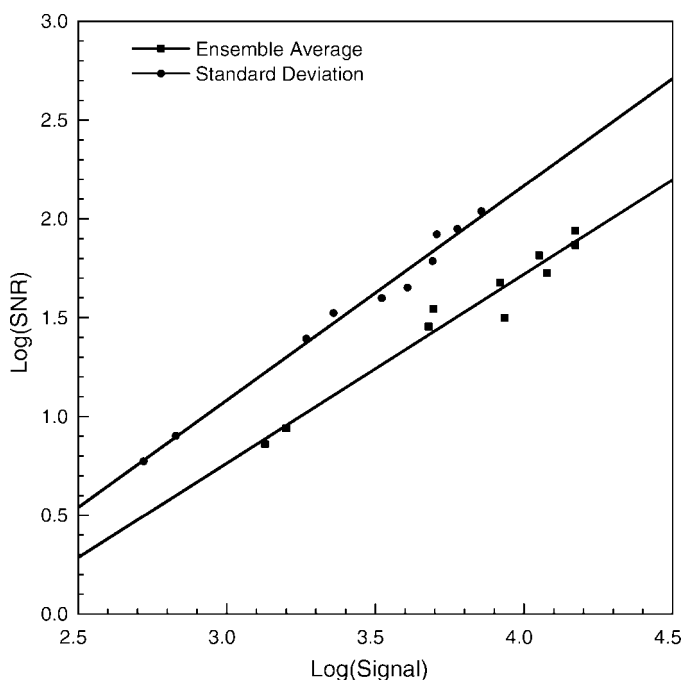


FIG. 5. The log of the signal-to-noise ratio (SNR) as a function of the log of the signal for the 247.8 nm C I atomic emission line intensity data of Figs. 3 and 4. Linear least squares fits are included for both data sets, with slopes equal to 0.96 and 1.09 for the ensemble-average and standard-deviation data sets, respectively.

was compared to the conditional data analysis scheme as established by Hahn et al.^{6,12,18} As discussed above, a typical ensemble-average based method has the potential to lose significant analyte information for aerosol systems when the analyte signal is derived from aerosol particles under discrete (i.e., low particle concentration) conditions. For the current experiments, the goal was to directly compare the resulting particle-derived analyte signal for these two different methodologies.

To perform the traditional conditional data analysis of the aerosol system, it is first necessary to select an appropriate threshold value for the analyte P/B ratio to quantify a given spectrum as corresponding to an actual particle hit. For all aerosol experiments, the analyte signal was the 393.4 nm Ca II line derived from the borosilicate glass microspheres. In the absence of any microspheres (i.e., purified air only), a threshold P/B value based on the spectral location of the 393.4 nm Ca II line was set to produce a false-hit rate of 0.4%, as discussed previously.^{6,12,18} This was done using a set of 4000 spectra recorded in purified air only, which yielded exactly 16 false particle hits for the selected P/B threshold value.

Using this conditional analysis threshold, experiments were conducted by collecting a set of 3000 spectra for a given particle concentration. For these experiments, the particle loading was expressed in terms of the average calcium concentration (ppb by mass) based on the average of 400 fg calcium per particle. The actual calcium concentration reported for each measurement is based on the average of two calculations. The calcium concentration was first calculated using the measured particle hit rate and the previously calculated calcium mass per particle and plasma sampling volume. Alternatively, the calcium concentration was calculated based directly on the resulting P/B value of the 393.4 nm

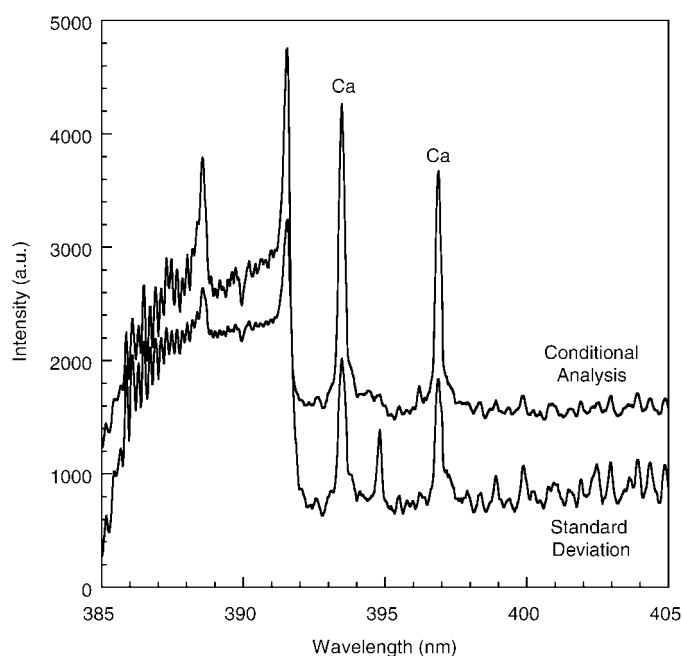


FIG. 6. Spectra showing the Ca II atomic emission lines at 393.4 and 396.9 nm (as labeled), one calculated from the standard-deviation analysis of 3000 sequential spectra (lower spectrum), another based on the ensemble-average of a subset of 314 particle hits obtained by applying the conditional analysis routine triggered on the 393.4 nm calcium line (upper spectrum). Both spectra have the same scale but have been displaced vertically.

calcium emission line for the ensemble-averaged spectra and a corresponding calibration. Given the emission intensity of the calcium line in combination with the moderate sample rates, the emission line was distinct in the ensemble-averaged spectrum, noting that this latter approach eliminates all issues associated with thresholding and particle hit detection.

These two approaches agreed within an average of 7% over all reported concentrations, although there was a trend of convergence with higher concentration. This is not unexpected, as ensemble-averaging and conditional-processing methodologies converge exactly at 100% sampling rate. The conditional analysis threshold was used to identify spectra corresponding to actual particle hits, which were subsequently separated from the set of 3000 total spectra. The spectra corresponding to particle hits were then averaged together, and the actual number of hits and corresponding frequency of hits (total hits divided by total shots) were then stored. In addition, the standard deviation of each set of 3000 spectra was calculated for each calcium concentration in an identical manner as done for the carbon dioxide experiments.

Figure 6 shows representative spectra based on the conditional analysis method and on the standard deviation method for a single calcium concentration. Specifically, Fig. 6 contains the spectrum corresponding to the ensemble-average of 314 particle hits identified from one 3000-shot sequence and the standard deviation of the same set of all 3000 spectra. In addition to the two Ca II atomic lines, both spectra reveal emission features at 391.4 nm and 388.4 nm, which are attributed to the N_2^+ first negative system, and the CN violet system, primarily at 388.3 and 387.1 nm.¹⁹ The continuum and molecular features are generally rather similar, which is consistent with the results of the gas-phase carbon dioxide experiments. In contrast, the aerosol particle-derived calcium

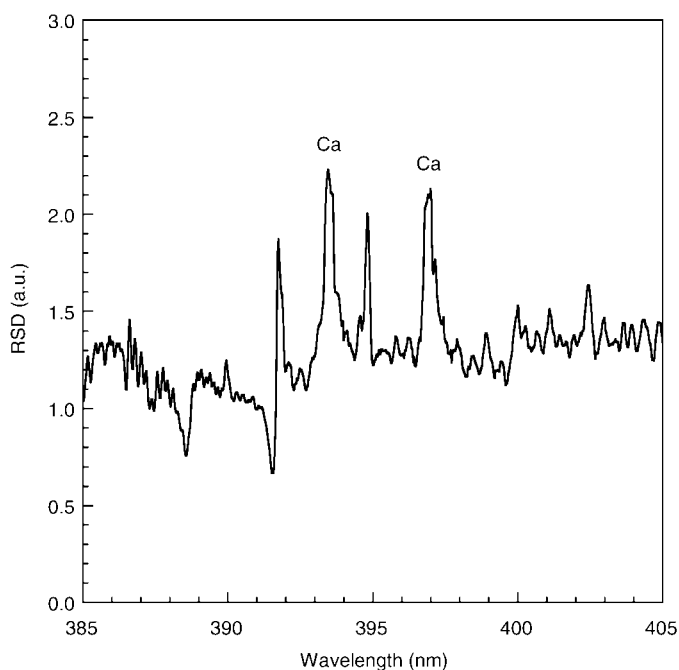


FIG. 7. Relative standard deviation (RSD) as calculated from the quotient of the standard-deviation and the ensemble-average for 3000 sequential spectra recorded in the presence of calcium-containing particles. The data correspond to the spectral data presented in Fig. 6. The location of the Ca II atomic emission lines at 393.4 and 396.9 nm are labeled.

emission intensity realized with the standard deviation method is observed to be less intense than in the spectrum obtained with conditional data analysis. Due to the fact that the standard deviation method responds to any fluctuations associated with each individual shot, somewhat higher continuum fluctuations appear in the standard deviation spectrum due to the greater variability of the plasma in the presence of aerosol particles. In contrast, the conditional data analysis generated a sufficient number of real hits to diminish (i.e., average out) the spectral noise.

Figure 7 shows the RSD for the standard deviation spectrum of Fig. 6. Unlike with the gas-phase experiments, the particle-phase analyte lines at 393.4 and 396.9 nm do show an increased RSD with respect to the continuum emission. This is certainly to be expected, as the variability of particle hits, and hence of the calcium emission signal, is expected given the particle hit rate of about 10% (314/3000). Furthermore, it is noted that the actual analyte response of an aerosol particle is a localized phenomenon within the plasma; hence, spatial variation is to be expected beyond what is realized for a gas-phase (i.e., well-dispersed) analyte.^{19–21} Such spatial variations may therefore contribute to the RSD in addition to the contribution of simple particle hits and misses. An additional peak is noted between the two calcium emission lines that is not attributed to any specific element but was consistently observed in the standard deviation spectra.

For quantitative analysis of the aerosol data, the 393.4 nm Ca II emission line was used to calculate the P/B and the SNR for both the conditional data analysis spectra (i.e., average of the particle hits) and the standard-deviation based spectra as a function of the calcium concentration in the aerosol sample stream. As noted in earlier publications,^{11,18} the actual calcium concentration is proportional to the product of the P/B of the

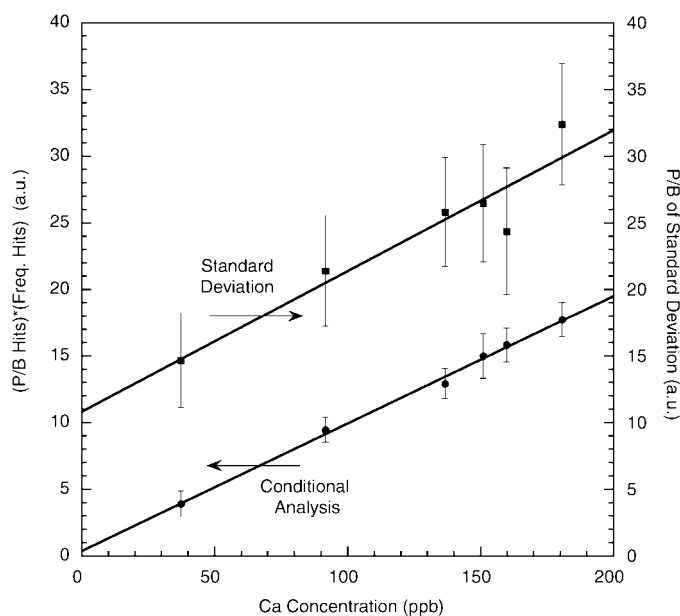


FIG. 8. Calibration curves based on the peak-to-base ratio of the 393.4 nm Ca II atomic emission line as a function of calcium mass concentration in the aerosol stream. Curves are presented based (1) on conditional analysis of the calcium-based hits, in which the P/B of the hits are scaled by the frequency of particle hits, and (2) on the standard deviation of the sequentially recorded set of 3000 spectra based on 12 250-shot sequences. The corresponding least squares fits are shown, which yield regression coefficients of 0.998 and 0.946 for the conditional analysis and standard deviation analysis, respectively. Error bars represent the standard deviation of the 12 250-shot sequences.

calcium-based hits spectrum and the particle sampling frequency (i.e., number of hits/number of shots). As the particle sampling frequency approaches 100%, this method converges to the traditional ensemble-averaging approach. Figure 8 presents the calcium analyte signal based on the standard deviation method and the conditional data analysis method as a function of calcium concentration, along with the linear least squares fits. The resulting correlation coefficients (R) of the fits are 0.998 and 0.946 for the conditional data analysis and standard deviation methods, respectively. The error bars in the standard deviation curve and conditional data analysis correspond to the standard deviation of the 12 250-shot spectral sets among the overall set of 3000 spectra for a given concentration. It is clear that the variability associated with the standard deviation method is larger than in the case of conditional data analysis due to the fluctuations associated with the spectral measurement, as noted above. Similar linear trends were also observed for the SNR calculations for both methods of analysis.

Perhaps the most striking feature of the Fig. 8 data concerns the apparent offset of the standard-deviation based calibration curve with respect to the conditional-analysis based curve. Specifically, the linear fit to the conditional analysis data results in a nearly ideal zero intercept, even though the lowest concentration measured was about 40 ppb. It is noted that the curve fit was not forced through zero. However, the linear fit of the standard deviation data exhibits a considerable y-intercept value. Clearly the analyte signal of the standard-deviation based approach converges toward the continuum intensity as the analyte concentration is reduced to zero, as demonstrated with the carbon experiments (see Fig. 3). If such a result is to occur with the aerosol data, a nonlinear analyte response must

be necessary to achieve such a zero intercept. Because it is difficult to control the statistics of particle sampling and loadings at near-zero particle loading, the convergence of the standard deviation and conditional analysis routines were explored numerically using a Monte Carlo based approach to simulate discrete particle sampling.

Monte Carlo Simulation of Aerosol Sampling. A Monte Carlo scheme was used to simulate spectral data consistent with LIBS-based discrete particle analysis. Specifically, a random number generator was used to assign each spectrum as either a particle hit or miss based on a predefined average particle hit rate. For each designated hit, the analyte peak was assigned additional random noise based on an average rms value, while the continuum emission value was also assigned additional random noise. For each designated miss, both the analyte peak and the continuum emission value were assigned the same magnitude along with additional random noise. The process was repeated for a total of 1000 simulated spectra. The simulated spectra were then processed in a manner similar to the actual aerosol data discussed above. Namely, the standard deviation of the 1000 spectra was calculated, and the resulting P/B of the analyte line was directly evaluated. For the conditional analysis, the particle hits were separated from the spectra using a threshold, and the P/B of the average spectrum of resulting particle hits was calculated and then multiplied by the particle hit frequency. For each defined particle sampling rate, this process was repeated a total of eight times; hence, final results correspond to 8000 simulations for each prescribed sampling rate.

The results of the numerical simulation are presented in Fig. 9, where P/B ratios for the two methods are presented as a function of defined particle sampling rate. For this data set, the average intensity of an analyte hit was equal to 35 times the value of the continuum rms noise, with the actual intensity of a given hit assigned a random noise variation within a range of $\pm 10\%$ of the average. The resulting thresholding allowed for perfect accuracy in distinguishing between particle hits and non-hits (i.e., miss). In the figure, the error bars represent the variability based on the eight 1000-shot simulations for each particle sampling rate. Several features are noted from the numerical results. First, as the particle sampling rates approach zero (0.1% lowest prescribed value), both the standard deviation method and conditional analysis method yield an analyte signal that correctly converges to zero. Secondly, the conditional analysis method produces a linear trend of analyte signal as a function of particle sampling rate, which is in excellent agreement with the experimental data reported in Fig. 8. Such linear behavior is expected if the analyte signal per particle remains constant with increasing single-particle hit rate, as is the case with both the simulations and experiments.

In contrast, however, the standard-deviation based data analysis method is seen to produce a trend that is nearly parabolic, with an apparent extrema occurring at a sampling rate of 50%. Additional calculations (not shown) ranging to a near 100% sampling rate reveal that the 50% value is in fact the maximum. While such a trend may appear unusual for an analyte response, the nature of the standard deviation method readily accounts for the curve shape. Specifically, when zero or near-zero particle hits are encountered, the standard deviation of the analyte peak is essentially reduced to the standard deviation of the off-peak continuum region (i.e., equivalent peak and base signals), and hence the P/B ratio is properly zero

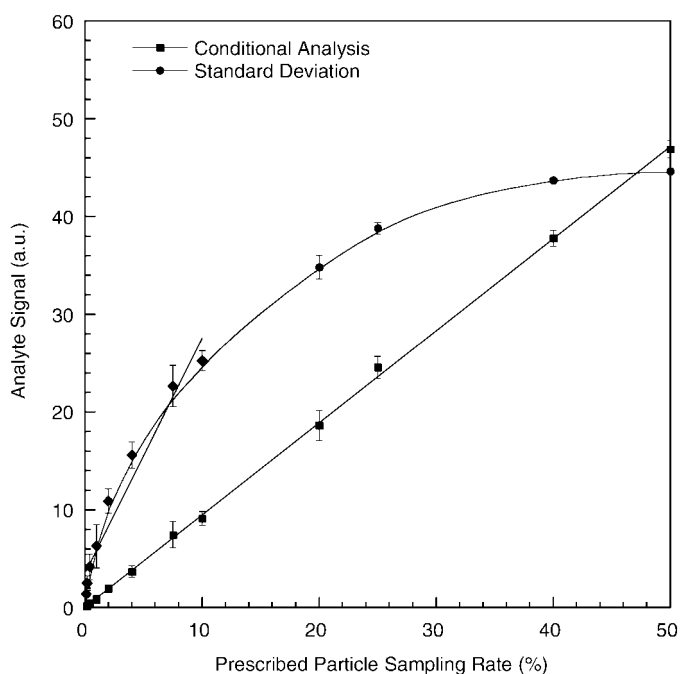


FIG. 9. Monte Carlo simulation showing the simulated analyte response (P/B ratio) as a function of particle sampling rate for a conditional analysis data processing scheme and for a standard deviation data processing scheme. The average analyte intensity of the particle hits is equal to 35 times the rms of the continuum intensity, with a random noise fluctuation of $\pm 10\%$. Each sampling rate corresponds to a 1000-shot simulation repeated eight times, with corresponding error bars based on the standard deviation of the 8 1000-shot simulations. The solid line is a linear fit of the standard-deviation data for the sampling rate data of 10% or less.

or near-zero. As particle hits are added as an increasing percentage of total shots, the standard deviation must increase due to the variation in analyte peak signals between the hits and non-hits. Therefore, the P/B of the standard deviation increases significantly with particle sampling rate. However, the maximum standard deviation must occur at a particle sample rate of 50%, which corresponds to half of the spectra belonging to particle hits and half of the spectra being non-hits. Beyond a 50% hit rate, the *variability* of the analyte peak is reduced as an increasing number of spectra are characterized by analyte peak signals consistent with particle hits. For 100% hits, the P/B is reduced to a value consistent with a homogeneous sample of particles. Overall, the simulations reveal nonlinear analyte response for discrete particle sampling rate with the standard deviation method, although the resulting function is monotonic up to particle sampling rates of 50%.

The above simulations were performed for what might be considered the case of significant or high analyte signals for a given hit, in that the simulation was perfect in identifying particle hits. However, it is also useful to explore the behavior of the two approaches for conditions representative of particle-derived analyte signals much closer to the continuum noise level. Therefore, the Monte Carlo simulations were repeated for the average intensity of an analyte hit equal to 10 times the value of the continuum rms noise, with the actual intensity of a given hit assigned a random noise variation within a range of $\pm 70\%$ of the average. The resulting thresholding allowed for 92.7% accuracy in distinguishing between particle hits and non-hits (i.e., miss). In other words, 7.3% of assigned particle hits were not detected by the conditional analysis threshold and

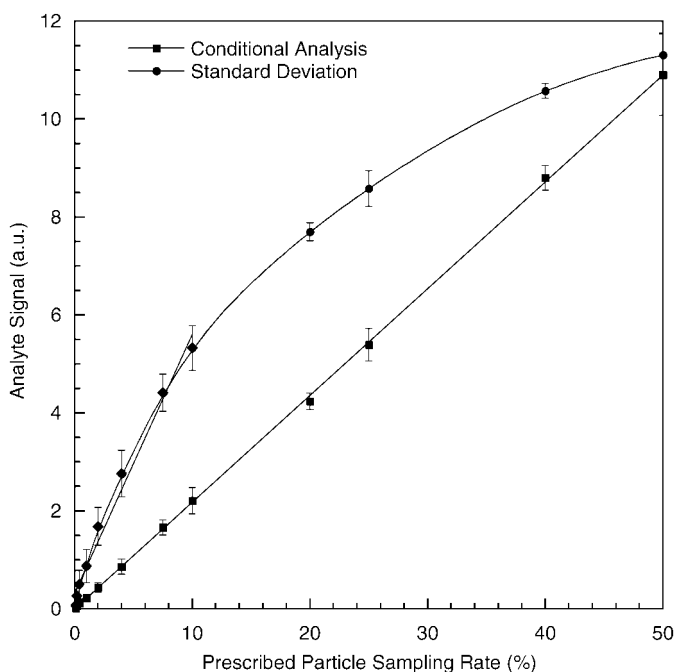


FIG. 10. Monte Carlo simulation showing the simulated analyte response (P/B ratio) as a function of particle sampling rate for a conditional analysis data processing scheme and for a standard deviation data processing scheme. The average analyte intensity of the particle hits is equal to 10 times the rms of the continuum intensity, with a random noise fluctuation of $\pm 71\%$. Each sampling rate corresponds to a 1000-shot simulation repeated eight times, with corresponding error bars based on the standard deviation of the eight 1000-shot simulations. The solid line is a linear fit of the standard-deviation data for the sampling rate data of 10% or less.

were subsequently left out of the resulting analyte signal. In contrast, such events can still affect the standard-deviation based analyte signal, as they are statistically greater than the average continuum signal.

The results of the second Monte Carlo simulation are presented in Fig. 10 for a set of eight 1000-shot simulations. While the overall results are similar to the Fig. 9 data (i.e., linear response for the conditional analysis data and a parabolic response for the standard deviation data), a trend is observed concerning the region of data corresponding to low sampling rates. To illustrate this effect, a linear fit is made to the particle sampling rates data less than or equal to 10%. For the Fig. 9 data, the correlation coefficient (R) is equal to 0.977 and the offset (y -intercept) is equal to 38% of the conditional-analysis signal at 10% sample rate. In contrast, the Fig. 10 data reveal a correlation coefficient equal to 0.993 and an offset (y -intercept) value equal to 14% of the conditional-analysis signal at 10% sample rate. This suggests that as the analyte signal becomes weaker and/or noisier (i.e., a signal regime where conditional analysis thresholding is expected to produce more false hits and/or to miss more real hits), the standard deviation procedure may be useful, noting the near-linear response and greater analyte-response slope under such conditions.

In view of the data shown in Figs. 9 and 10, one may reconsider the results of Fig. 8 in further consideration of the actual particle sampling rates. The lowest calcium concentration of Fig. 8 corresponds to an average particle sampling rate of 10%, while the greatest concentration corresponds to 44%. Interestingly, if one considers the Fig. 9 and 10 data only between sampling rates of 10 and 44%, one might easily draw

a linear curve that would roughly parallel the conditional analysis curve, but with a significant y -intercept value. Essentially, this is in excellent qualitative agreement with the experimental sampling data of Fig. 8 when comparing the two methodologies. In particular, the Monte Carlo simulations did not allow for multiple-particle sampling (i.e., two or more particles per shot), although clearly the sampling of multiple particles is to be expected as the overall sampling rate approaches 40–50%. Multiple particle hits would tend to mitigate the leveling off in the standard deviation curve near the 50% sample rate, as multiple particles in the plasma volume would further increase the analyte emission, and therefore the shot-to-shot variability and the corresponding standard deviation. This is consistent with the greater variability and nonlinear nature of the standard deviation data for the higher calcium concentrations.

Additional effects may contribute to the differences observed in the Fig. 8 aerosol data. Recall that the photon shot noise rms value is directly proportional to the square root of the number of photons incident on the detector. However, with the standard deviation method, no averaging is performed that diminishes the influence of photon shot noise with respect to the analyte signal, as happens with conditional data analysis, since the number of real hits are ensemble-averaged to diminish this effect.

CONCLUSION

Several important findings are revealed from the present experiments and simulations regarding the use of the standard deviation as an analytical metric for LIBS or other spectroscopic techniques. First, in a homogeneous system, the standard deviation method provides a suitable metric that is directly proportional to the analyte signal, as is the case with the gaseous carbon dioxide experiments. Mathematically, the scaling of the standard deviation of the analyte signal with increasing analyte concentration is in agreement with the actual definition of the standard deviation. Accordingly, the equivalence of the standard deviation and traditional ensemble-averaging is consistent with the problem formulation and is validated by experimental data. In contrast, the applicability of the standard deviation method for analysis of non-homogeneous analyte systems (e.g., aerosol systems), where the analyte signal is highly dispersed with regard to the analytical sample volume (i.e., the laser-induced plasma volume), must be carefully considered. It was shown both experimentally and via Monte Carlo simulations that the standard-deviation approach can produce an analyte response that is monotonic with analyte concentration up to a point at which the analyte signal begins to transition from a non-homogeneous system to a homogeneous system (i.e., $\sim 50\%$ particle sampling rate for aerosol particles). Clearly, the standard deviation scheme is well suited for semi-quantitative analysis, although an absolute calibration will be problematic without additional knowledge regarding the overall analyte sample rate.

While the above comments show the equivalency of the standard deviation approach and the traditional ensemble-average under high analyte loading, and some limitations in comparison to conditional analysis for aerosol systems, an additional key feature of the method is noted. Namely, the use of conditional processing for LIBS-based analysis of non-homogeneous systems necessitates the pre-selection of target analyte emission lines such that appropriate thresholds can be

established, and all spectra are then processed by comparing each expected analyte peak to the threshold. This requires *a priori* knowledge about the spectral location of targeted analyte lines, as well as a general expectation of what analyte species might be present. This might be considered a limitation of conditional processing, in that locating unexpected or unknown analyte lines can be problematic. In contrast, the standard deviation method requires no prior knowledge of the spectral location of analyte lines or what exact species are expected. Rather, the technique produces a standard-deviation spectrum that reveals all spectral locations of non-homogeneously dispersed analyte species. Clearly such broad-based spectral screening with no prior knowledge of analyte species is beneficial, noting that the approach is applicable for aerosol systems as well as for liquids (e.g., suspensions) and non-homogeneous solids. For example, one can imagine using the standard deviation method to perform qualitative screening to efficiently identify analyte species present, which could then be followed by a guided conditional analysis routine. Clearly, the use of LIBS for analysis of highly variable analyte systems, such as aerosols or occluded solids, presents interesting challenges to traditional approaches; hence, there remains motivation to develop and optimize alternative spectral processing and analyte routines. The standard deviation approach can very well fill an important role in these efforts.

ACKNOWLEDGMENTS

This work was supported in part by the Dirección General de Investigación, Ministerio de Educación y Ciencia (Spain), through grant CTQ2007-60348, which provided a travel grant for Luis Álvarez. The authors would like to thank Soupy Dalyander (University of Florida) for her generous assistance with the

data analysis software, and Prof. Nicolo Omenetto (University of Florida) for useful discussions regarding the role of spectral noise.

1. L. J. Radziemski, T. R. Loree, D. A. Cremers, and N. M. Hoffman, *Anal. Chem.* **55**, 1246 (1983).
2. A. Molina, P. M. Walsh, C. R. Shaddix, S. M. Sickafoose, and L. G. Blevins, *Appl. Opt.* **45**, 4411 (2006).
3. J. Sneddon and Y. Lee, *Anal. Lett.* **32**, 2143 (1999).
4. D. A. Rusak, B. C. Castle, B. W. Smith, and J. D. Winefordner, *Crit. Rev. Anal. Chem.* **27**, 257 (1997).
5. W. B. Lee, J. Y. Wu, and J. Sneddon, *J. Appl. Spectrosc. Rev.* **39**, 27 (2004).
6. D. W. Hahn, W. L. Flower, and K. R. Hencken, *Appl. Spectrosc.* **51**, 1836 (1997).
7. D. W. Hahn, *Appl. Phys. Lett.* **72**, 2960 (1998).
8. I. Schechter, *Anal. Sci. Technol.* **8**, 779 (1995).
9. I. B. Gornushkin, B. W. Smith, G. E. Potts, N. Omenetto, and J. D. Winefordner, *Anal. Chem.* **71**, 5447 (1999).
10. M. D. Cheng, *Fuel Process. Technol.* **65-66**, 219 (2000).
11. D. W. Hahn and M. M. Lunden, *Aerosol Sci. Technol.* **33**, 30 (2000).
12. J. E. Carranza, K. Iida, and D. W. Hahn, *Appl. Opt.* **42**, 6022 (2003).
13. L. A. Álvarez-Trujillo, A. Ferrero, and J. J. Laserna, *J. Anal. At. Spectrom.* **23**, 885 (2008).
14. J. E. Carranza, G. R. Arsenault, H. A. Johnsen, K. R. Hencken, and D. W. Hahn, *Rev. Sci. Instrum.* **72**, 3706 (2001).
15. V. Hohreiter and D. W. Hahn, *Anal. Chem.* **78**, 1509 (2006).
16. J. E. Carranza and D. W. Hahn, *Spectrochim. Acta, Part B* **57**, 779 (2002).
17. J. D. Ingle and S. R. Crouch, "Signal-to-Noise Ratio Considerations", in *Spectrochemical Analysis* (Prentice Hall, New Jersey, 1988), Chap. 5.
18. J. E. Carranza, B. T. Fisher, G. D. Yoder, and D. W. Hahn, *Spectrochim. Acta, Part B* **56**, 851 (2001).
19. V. Hohreiter and D. W. Hahn, *Anal. Chem.* **77**, 1118 (2005).
20. G. A. Lithgow and S. G. Buckley, *Spectrochim. Acta, Part B* **60**, 1060 (2004).
21. P. K. Diwakar, P. B. Jackson, and D. W. Hahn, *Spectrochim. Acta, Part B* **62**, 1466 (2007).