Post hoc Interlaboratory Comparison of Single Particle ICP-MS Size Measurements of NIST Gold Nanoparticle Reference Materials

Journal:	*Analytical Chemistry*
Manuscript ID:	ac-2015-01741n.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Montoro Bustos, Antonio; National Institute of Standards and Technology, Petersen, Elijah; National Institute of Standards and Technology, Possolo, Antonio; National Institute of Standards and Technology, Winchester, Michael R.; National Institute of Standards and Technology,

Post hoc Interlaboratory Comparison of Single Particle ICP-MS Size Measurements of NIST Gold Nanoparticle Reference Materials

Antonio R. Montoro Bustos,[1]* Elijah J. Petersen, [1] Antonio Possolo[2], Michael R. Winchester[1]
[1]Material Measurement Laboratory, [2]Information Technology Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-1070.

ABSTRACT

Single particle inductively coupled plasma-mass spectrometry (spICP-MS) is an emerging technique that enables simultaneous measurement of nanoparticle size and number quantification of metal-containing nanoparticles at realistic environmental exposure concentrations. Such measurements are needed to understand the potential environmental and human health risks of nanoparticles. Before spICP-MS can be considered a mature methodology, additional work is needed to standardize this technique including an assessment of the reliability and variability of size distribution measurements and the transferability of the technique among laboratories. This paper presents the first *post hoc* interlaboratory comparison study of the spICP-MS technique. Measurement results provided by six expert laboratories for two National Institute of Standards and Technology (NIST) gold nanoparticle reference materials (RM 8012 and RM 8013) were employed. The general agreement in particle size between spICP-MS measurements and measurements by six reference techniques demonstrates the reliability of spICP-MS and

validates its sizing capability. However, the precision of the spICP-MS measurement was better for the larger 60 nm gold nanoparticles and evaluation of spICP-MS precision indicates substantial variability among laboratories, with lower variability between operators within laboratories. Global particle number concentration and Au mass concentration recovery were quantitative for RM 8013 but significantly lower and with a greater variability for RM 8012. Statistical analysis did not suggest an optimal dwell time, because this parameter did not significantly affect either the measured mean particle size or the ability to count nanoparticles. Finally, the spICP-MS data were often best fit with several single non-Gaussian distributions or mixtures of Gaussian distributions, rather than the more frequently used normal or lognormal distributions.

INTRODUCTION

Engineered nanomaterials (ENMs), objects with at least one external dimension in the range from approximately 1 nm to 100 nm,[1] often possess exceptional properties and are expected to be increasingly used in commercial applications in the future. However, their release during manufacturing, product use and disposal[2] has aroused global concern regarding their potential adverse impacts to the environment[3] and human health.[4]

To assess these risks accurately, innovative and reliable analytical methods are needed for ENM characterization and quantification at extremely low environmental concentrations (on the order of ng L^{-1})[5] and for the implementation of nanomaterial safety regulations.[6,7] In particular, the European Commission has recommended a definition of nanomaterial as "natural, incidental or manufactured material containing particles, in an unbound state or as an aggregate or as an agglomerate and where, for 50 % or more of the particles in the number size distribution, one or

more external dimensions is in the size range 1 nm to 100 nm".[8] Application of this definition requires the availability of reliable, accurate, and efficient measurement methods of ENM size distribution and number concentrations. However, none of the currently available analytical techniques can satisfactorily fulfill this regulatory definition for all types of ENMs.[9,10] Thus, new or enhanced analytical approaches for sizing and counting individual nanoparticles (NPs, defined as nanomaterials having all three dimensions in the nanoscale)[1] in aqueous solutions are required.

Several well-established analytical techniques such as transmission electron microscopy (TEM) and dynamic light scattering (DLS) can be used to characterize the size distribution of NPs. However, these techniques are challenged by polydispersed samples, drying artifacts, and the analysis of a sufficiently high number of NPs for representative microscopic measurements.[10,11] Newer hyphenated techniques with mass spectrometry detection, which utilize a size separation method such as field flow fractionation or size exclusion chromatography prior to elemental analysis, hold significant promise for NP size distribution measurements, but typically involve long analysis times, interactions with a solid phase, do not always yield quantitative recoveries, and require additional instrumentation, all of which hinder widespread use and standardization.[12,13]

Single particle inductively coupled plasma mass spectrometry (spICP-MS) is one technique that offers significant promise in overcoming these limitations because of its capability to quantify NP number and size distribution simultaneously in aqueous samples. First developed by Delguedre et al.,[14] spICP-MS performs "particle by particle" measurements by operating in time-resolved analysis (TRA) with short dwell times (integration time of one reading by the detector) and provides valuable information about the chemical composition, number concentration, size

and size distribution of metal containing NPs at mass concentration levels down to ng L^{-1}.[15,16] In contrast to hyphenated techniques, spICP-MS offers greater promise owing to its accessibility for any ICP-MS owner. To analyze samples using spICP-MS, very dilute suspensions of NPs are nebulized so that each recorded pulse represents a single NP. The frequency of the pulses is directly related to the NP number concentration and the signal intensity is proportional to the mass of the particle, which can be translated into size assuming a particular ENM geometry (i.e. spherical). Thus, spICP-MS exhibits the unique ability to differentiate between both dissolved and particulate species simultaneously without additional sample preparation, as well as to characterize polydispersed samples.[17,18] For metal NPs, spICP-MS reported size and mass detection limits are about 20 nm and a few hundred attograms per NP, and in terms of NP number concentration, detection limits are in the range of 10^6 L^{-1} using current ICP-MS instruments.[16] Recently, spICP-MS has been used to characterize various commercialized ENMs in matrices ranging from algal growth media, wastewater, macroinvertebrates and animal tissues.[18-22]

Despite numerous advances with this technique in recent years,[19,23-29] some important limitations still remain and spICP-MS cannot yet be considered a mature methodology.[13,16,23,29-31] Additional work is needed to standardize this technique such as an assessment of the robustness of various parameters and the transferability of the technique among laboratories. In a recent round robin study using silver nanoparticles (AgNPs) in water and ethanol, there was dramatic variability among the laboratories with some participants unable to detect AgNPs.[32] It was unclear to what extent this variability was a result of heterogeneity of the initial AgNP suspensions or changes to the AgNPs during sample processing such as dissolution, or variability from the calibration approaches utilized. However, there was general agreement for the particle

sizes if outliers were excluded. Thus, additional work is needed to improve the understanding of the various sources of uncertainty. One approach often used to achieve this goal involves the use of reference materials (RMs) with carefully evaluated uncertainties for key parameters which are critical for enabling metrological traceability. In addition, evaluating the reliability of spICP-MS with a more stable nanomaterial with regards to dissolution would remove this particular source of variability. In this context, the recommended use of RMs has been widely recognized because the well-defined mean sizes and particle number concentrations enable a correct calibration and determination of nebulization efficiency which support a reliable and accurate spICP-MS analysis. In addition, determining the best statistical and qualitative methods to compare among spICP-MS size distributions and what statistical models best fit spICP-MS data were also unclear from the current literature.

In this study we discuss the first *post hoc* interlaboratory comparison study to evaluate the reliability and variability of the sizing and counting capabilities of the spICP-MS technique using National Institute of Standards and Technology (NIST) gold nanoparticle RMs (8012 and 8013); the term *post hoc* indicates that data already published were predominately used for the interlaboratory comparison to identify patterns that were not evident *a priori*. The interlaboratory comparison that we have undertaken is similar to the meta-analyses that are frequently conducted in the medical field to compare treatments or procedures based on published results from multiple studies.[33] This approach is most relevant in the absence of a standard procedure. The main goal of this study was to evaluate the interlaboratory reproducibility, precision, trueness, and variability of the size determination using spICP-MS. The influence of dwell time on spICP-MS size measurements, a key parameter in recent spICP-MS studies,[24,28,29,34] was also assessed. While individual studies have assessed the influence of different parameters (e.g., dwell time) or

sources of uncertainty on spICP-MS measurements, interlaboratory comparisons can still yield important insights because the impact of different factors may differ among instruments, operators, and laboratories which often are not predictable *a priori*. In addition to qualitative and statistical comparisons of the means of the distributions and the sources of variance, uncommon statistical approaches were applied to compare if the entire spICP-MS distributions were statistically similar. Lastly, single probability distribution models, and mixtures of normal distributions were fitted to the size distributions to determine the optimal model. While normal and lognormal distributions are often used to fit nanoparticle size distribution data, these distributions may not yield the best fit.

EXPERIMENTAL

Participants.

Single particle ICP-MS measurements of NIST AuNP reference materials from recent publications were provided by three academic, one governmental, and two commercial laboratories. All these laboratories were recruited based on their spICP-MS expertise in NP size characterization. Different users and different experimental conditions within a given laboratory were also included to assess the within laboratory repeatability. Three of the laboratories were European, and three laboratories were from the USA. To preserve confidentiality, the labs are referred to as Labs 1 to 6. Each laboratory and user set their own criteria for the background subtraction prior to size analysis. This choice for background subtraction may have impacted the detection limit, a parameter that was not explicitly explored in this study.

Sample selection.

Gold nanoparticle (AuNP) reference materials RM 8012 and RM 8013 with nominal diameters of 30 nm and 60 nm, respectively, sold by NIST (Gaithersburg, USA) starting in 2007, were selected for this study. Both reference materials consist of aqueous suspensions of negatively charged, citrate-stabilized, monodispersed AuNPs with approximately spherical shape, although a small percentage of larger particles and distinctly non-spherical shapes (i.e., faceted) may also be present. Advantages of using these RMs include their chemical and colloidal stability and homogeneity and the fact that the mean particle sizes and particle size distributions were characterized by six different sizing techniques, although no certified values were assigned owing to the metrological challenges for value assignment in the nanoscale.[35,36] Using these homogenous and stable materials effectively reduces the variability in test results, making it easier to interpret measurement results and characterize the sources of the observed variability**.** Recently, an interlaboratory study using NIST RM 8012 for particle size distribution evaluation by transmission electron microscopy (TEM) was performed.[37] In the present study we mainly used TEM and scanning electron microscopy (SEM) results provided in the NIST Report of Investigation, because scientists conducting spICP-MS analysis most frequently compare their results to the reported SEM and TEM sizes and often use TEM sizes for spICP-MS calibration.

Instrument parameters.

In this study, multiple ICP-MS instruments, operators, sample preparation procedures, calibration approaches, and different experimental conditions (dilution levels, sample flow rate, acquisition time and transport efficiency) were employed to analyze the NIST AuNPs. The diversity in these parameters (see Table 1) is reflective of the expected variability among different expert laboratories. This enabled an evaluation of the relative importance of different

experimental parameters, including dwell time, a subject that is presently receiving considerable attention.[29,34]

Data Evaluation.

Statistical analysis of the particle size data was used for three purposes: evaluation of different sources of observed variability for particle size measurements, quantitative comparison of different size distributions, and fitting reference models to the size distribution data.

Analysis of variance (ANOVA). A mixed effects model[38] was used to evaluate the between-laboratory, within-laboratory, between-operator (within a laboratory) and total variability of the measured particle diameter distributions, amounting to a top-down evaluation of measurement uncertainty as is often done in chemistry.[39] Additional details for how this analysis was conducted are available in the Supporting Information.

Comparing distributions. Cramer-Von Mises (CVM)[40] and Kolmogorov-Smirnov (KS)[41] statistical tests of the equality of continuous distributions were used to make pairwise comparisons between sample distributions. These tests differ from more common statistical approaches such as t-tests and ANOVA in that the equality of the entire distribution was compared. Additional information about the statistical comparison is available in the Supporting Information.

Fitting reference models to data. Normal and lognormal distributions are commonly used models for nanoparticle size distributions. Although ideal NP size distributions are considered lognormal,[42] all the experimental results included in this study showed NP size histograms with a large central distribution and with smaller secondary distributions in the tails. In addition to the presence of a significant percentage of larger particles suggested in the Reports of Investigation

for RMs 8012 and 8013,[35,36] spICP-MS measured particle sizes greater than the primary distribution can be attributed to particle coincidence effects and/or to agglomeration.[34]

In this study, several single distribution models and mixtures of Gaussian models were fitted to the spICP-MS data. The Bayesian Information Criterion (BIC)[43] value was used to select the best model; additional details of this analysis are provided in the Supporting Information.

RESULTS AND DISCUSSION

Results from a total of fifteen independent users/dwell time combinations from the six participant laboratories were submitted. A set of 68 and 77 independent spICP-MS runs for RM 8012 and RM 8013, respectively, were processed in this study.

Results for mean particle diameter, spread level and boxplots of each single spICP-MS run for the RM 8013 60 nm AuNPs and for RM 8012 30 nm AuNPs are shown in Fig 1 and Fig S-1, respectively. Tables S-1 and S-2 summarize main results for each laboratory, user and dwell time for RM 8013 and RM 8012 spICP-MS analysis, respectively, as well as number of replicates, number of events and interquartile ranges.

Precision. Precision refers to the mutual agreement between measurement results obtained under conditions of repeatability.

Mean particle diameter. All-laboratory mean (average of all data obtained) and the results of the ANOVA for the different identified sources of variability are listed in Table S-3 for both RMs. The data from Lab4 at the second dwell time (td2) value were excluded from the precision evaluation of RM 8012 owing to their anomalously high dispersion. After elimination of outliers, global (all laboratories) results for RM 8013 followed a relatively narrow unimodal distribution; results ranged from 36 nm to 69 nm with an all-laboratory mean of 55.9 nm. RM 8012 followed

a very broad unimodal distribution after removal of outliers, with global results ranging from 10 nm to 47.5 nm with an all-laboratory mean of 28.2 nm. The variability of spICP-MS results for both RMs, expressed as 95% coverage intervals, was generally much lower than that reported for the reference size techniques (see Fig 1A and S-1A) as a result of the large number of particles measured. Importantly, the global spICP-MS variability between labs for RM 8013 was moderate, but it was significantly higher for RM 8012 (Table S-3). The difference in precision among particle diameters for the two RMs was also evident by the median absolute deviation plot (Fig 1B, MAD). Despite the larger expanded uncertainties for SEM and TEM, MAD values of SEM and TEM were smaller than for the participant spICP-MS results, indicating that the SEM and TEM results were, on average, closer to the median value. The higher dispersion of particle diameters for the smaller AuNPs was also evident from the plots of each individual spICP-MS run measurement (Fig S-1B) which showed a larger range for the boxplot whiskers.

The results of the ANOVA revealed that, while there was substantial variability among labs, the within-lab standard deviation was significantly lower for both sizes of AuNPs, indicating that each lab provided internally consistent results (Table S-3). The ANOVA revealed that the main factor affecting the precision for both RMs was the residual contribution: the residual contribution, between laboratory variability, and within laboratory variability were 6 nm, 5 nm and 1 nm for RM 8012 and 7 nm, 2 nm, and 0.8 nm for RM 8013, respectively (see Table S-3). The residual term includes all sources of variability other than the between and within-lab variabilities such as the dispersion of values obtained by the users, which may be attributed to multiple uncontrolled factors, including the heterogeneity of the samples.

Particle size distribution. Smooth histograms (kernel density estimates)[44] (see Fig S-2, S-3, S-4, and S-5), were used to depict the variability of the particle diameter; histograms are included

in Fig S-6 and S-7 for comparison. A qualitative comparison of the probability density functions reveals substantial variability in the shape, height and maximum peak position among the different laboratories involved. These between-laboratory differences were even more pronounced in the case of RM 8012, revealing that it is more difficult to characterize the size distribution of smaller AuNPs.

To compare particle size distributions statistically, CVM and KS statistical tests were utilized. Analysis of the spICP-MS data from all participants and TEM and SEM distributions showed that these distributions were significantly different in all cases for RM 8012 data and in 95 % of the comparisons for RM 8013. Thus, almost all of the size distributions between laboratories differ from one another, which corroborated the qualitative analysis.

However, the distributions were more similar within each laboratory (see Table S-4 and S-5 and Figures S-2 and S-3). For RM 8013, the null hypothesis was accepted in 84% of the samples analyzed by each user, and in 48% for RM 8012, showing again more variability in the size characterization of the smaller-sized AuNP RM. The substantially higher similarity in distributions within laboratories than between laboratories suggested that the diversity in ICP-MS instruments, experimental conditions and data processing approaches used in each laboratory contributed more to the variability than heterogeneity of the AuNPs.

Accuracy or trueness. Accuracy or trueness (that is, unbiasedness) refers to the closeness of agreement between the arithmetic mean of a large number of test results and the true or accepted reference value.

Particle size diameter. Evaluation of trueness of the particle diameter size is difficult because it is a method-specific measurand; various sizing techniques rely on different fundamental principles.[12,32] Thus, the discrepancy among particle size measurement techniques hampers the

determination of a single true particle diameter, and this confounds classical accuracy assessment.[20] According to NIST, the dimensional reference values provided in the Reports of Investigation for the AuNPs (expressed as a mean particle diameter in solution, as an aerosol, or after deposition on a substrate, depending on the technique) are at best only an estimate of the true value, because all known or suspected sources of bias have not been fully investigated. Nevertheless, comparing among various techniques for NP sizing is still recognized as a valuable and valid approach for evaluating trueness and for method validation[45] even though there is not a single recognized true value. Trueness of the spICP-MS AuNP data obtained in this study was evaluated using size results from the six different sizing techniques reported in the Reports of Investigation (Table S-6, Fig 1A and S-1A). The particle size given on the Reports of Investigation for DLS at 90° scattering angle was not included in this study, because of its high expanded uncertainty.

While general agreement of the average particle size was observed across all techniques for both RMs, the means of the global spICP-MS size measurements were within 4% and 1% for RM 8012 and 8013, respectively, of the average value of the six reference mean values. This key finding validates the capability of spICP-MS for mean particle size determination. Moreover, spICP-MS results were in closest agreement with TEM and DMA results for both RMs, and for RM 8013, also with the SEM results. When comparing among the techniques, the mean size range was 3.3 nm and 3.4 nm for RM 8012 and 8013, respectively, indicating again a much smaller relative variability for the RM 8013 given the substantially larger size. The fact that spICP-MS offered the lowest relative expanded uncertainty values (expanded uncertainty divided by mean) for both RMs was not surprising because the number of particles globally sized in this

study was more than two orders of magnitude larger than for the other sizing techniques; an advantage of the spICP-MS is its capability to rapidly acquire large numbers of particle events.

Particle size distribution. In addition to the mean particle size, the reliability of the size distribution is another key parameter measurable using spICP-MS, but it has not yet been studied at a between laboratory level. Global probability density estimates (PDEs) of spICP-MS particle size diameter including raw data from all participants for RM 8012 and 8013 were compared in Fig 2 to the corresponding TEM and SEM distributions obtained from the size distributions provided in the Reports of Investigation. The PDEs are kernel density estimates and do not rely on specific parametric models for the data. Despite incorporating data from multiple laboratories, the spICP-MS PDE results for 8012 were only 121% and 133% broader than the PDEs for the TEM and SEM distributions, respectively. Widths of the size distributions for each laboratory can be directly compared to the widths for the SEM and TEM results in tables S-1 and S-2. For RM 8013 (60 nm AuNPs), the width of the global distribution obtained by spICP-MS was 51% and 66% broader than the original TEM and SEM distributions, respectively, a result in close agreement with the 53 % value previously reported by a single laboratory analysis compared with TEM.[24] However, an alternative qualitative comparison carried out in another laboratory did not reflect such a broadening effect.[28]

This broadening of the global distribution obtained from spICP-MS in comparison to those from TEM and SEM is not surprising because of the variability in the experimental conditions of the participating spICP-MS laboratories in contrast to the SEM and TEM results which were obtained by a single operator using the same experimental conditions. However, the broader spICP-MS size distribution may also be partly attributable to the following sources of uncertainty related to instrument performance: ion counting, introduction and ionization of the

nanoparticles in the plasma, partial detection of the pulses, and coincident particle events.[24,26] Small differences in particle size diameter and particle size distribution may be partly attributable to the small percentage of non-spherical shapes contained in the RM AuNPs; all particles including the non-spherical shapes are included in the spICP-MS results while these may be avoided in microscopy results depending upon the circularity value used in the image analysis software.

Particle number concentration and Au mass concentration recovery. The spICP-MS capability to assess the trueness of particle number concentration and Au mass concentration (expressed as recovery (ratio between determined and initial values)) important requirements for the recent EU nanomaterial definition,[8] was also studied (Tables S-7 and S-8 for RM 8013 and 8012, respectively). While definite values for the particle number concentration in the original RM suspension were not available and each participant chose their own sample dilution level, the AuNP number concentration recovery was still considered a reliable indicator because initial values of this parameter could be easily obtained from two independently assigned values in the Reports of Investigation: Au mass concentration information and reference particle mean size (TEM mean diameter was selected).[35,36] Therefore, the calculated recovery for the particle number and Au mass concentrations are the same. Global recovery for RM 8013 averaged (86 ± 12) % (uncertainty indicates one standard deviation), with 7 of the 14 recoveries in the last column of Table S-7 being indistinguishable from quantitative recovery at the 95 % confidence level. The recoveries for RM 8012 were significantly lower, with a global average of (60 ± 29) % (uncertainty indicates one standard deviation), and with a greater dispersion of values (Table S-8). Moreover, only 3 of the 15 values were in statistical agreement with quantitative recovery at the 95 % confidence level. This difference revealed that it is more difficult to count smaller

NPs using spICP-MS, an effect already described in the literature.[24,32] The capability to count small particles could be lower for two reasons: 1) an existing fraction of particles were closer the size limit of detection, and 2) the transport efficiency correction, typically computed using larger NPs, might be biased against smaller particles and could thus overlook a significant fraction of smaller particles.

Losses in the AuNP concentration from adsorption onto the container sidewalls could theoretically occur in the ampoule or in a separate container after a dilution is made from the stock suspension. The NIST RMs were originally certified until October 2015, and all of the spICP-MS used in our study were acquired well before this initial expiration date. Subsequently, the expiration dates for both RMs have been extended to October 2018. Based on the expiration dates and the fact that long and deep experience with the RMs within our laboratory and many others have not indicated any problems, we can reasonably rule out the possibility of significant losses in the ampoules. Losses to the side walls of secondary containers, which are certainly possible if the stock solution is transferred or diluted, have been observed for silver NPs,[46] and are an important topic for future studies. However, the quantitative recovery for the 60 nm AuNPs suggests that losses to the sidewalls were low for these NPs.

There was no significant difference in recovery as a function of measured particle numbers or Au mass concentration values for both RMs. This result indicates that spICP-MS counting ability is independent of particle number and Au mass concentration levels within the range used in this study. This finding reflects positively on the robustness of the method.

Influence of dwell time on particle diameter, particle size distribution and particle number and Au mass concentration recovery. The wide range of dwell time values used by the participants in the study (see Table 1) enabled the comparison between millisecond and

microsecond spICP-MS. The advantages and limitations of these dwell times are discussed in the Supporting Information and are summarized in Table S-9. Overall, the choice of dwell time will depend on the flux of particles entering the plasma, and the intensity of the dissolved analyte, which also depend on the nebulization system and the sample uptake rate used.

It was not possible to study the effect of dwell time rigorously, because, for the most part, different laboratories used different dwell times, thus confounding the effect of dwell time with between-laboratory variability. To investigate general trends among the different dwell times, the data were grouped using 4 and 5 different dwell time value categories for RM 8013 and RM 8012, respectively.

The mode, median and mean particle diameters and recoveries determined for each dwell time category for RM 8013 and RM 8012 are compared in Table 2 and S-10, respectively. The measured particle size distributions were similar across all the dwell times and did not show a systematic bias in particle size determination with dwell time. This finding indicates the robustness of the method. Interestingly, the relative range of values was again larger (15 %) for RM 8012 than for RM 8013 (4 %), revealing a higher variability for sizing smaller NPs at different dwell times. Nanoparticle number and Au mass concentration recoveries (sixth column in Table 2 and S-10) for the different dwell time values ranged from 79.3 % to 93.9 % for RM 8013, indicating an ability to count NPs independent of the dwell time. In the case of RM 8012, the recoveries were significantly lower with a larger range (38.5 % to 102.2 %), which once again revealed a higher difficulty to empirically count smaller NPs.

A qualitative comparison of the probability density functions obtained for different dwell time values, depicted in Fig 3 and S-6 for RM 8013 and RM 8012 respectively, revealed a more similar distribution (shape, height and maximum peak position) for the different dwell times with

RM 8013 than for RM 8012. However, statistical analysis using the CVM and KS tests showed that the distributions of particle size for each dwell time value were all significantly different from one another for both RMs. The width of the distributions (fifth column in Table 2 and S-10) were narrowest for 10 ms dwell time, being 35 % and 55 % lower than the other dwell times for RM 8012 and RM 8013, respectively. This finding may be the result of a higher probability of false positives for microsecond dwell times and lower probability of split particle events at higher dwell times.

Overall, an optimal dwell time value could not be determined. While the data for the 10 ms dwell time showed narrower distributions, these data were from a single laboratory and more data are needed. The average size and recovery were similar for all dwell times. This indicates that all dwell times can provide equally reliable sizing and counting results for RM 8013 (and size measurements for RM 8012) if dwell time specific corrections (split events, false positive events, etc.) and sample preparation steps are taken. Nevertheless, there is still a clear need for the development of fit-for-purpose standard spICP-MS protocols as a result of the many considerations necessary for accurate measurements.

Fitting reference models to data.

Univariate fitting models. The best candidate single probability distribution models for RM 8013 and RM 8012 are listed in Tables S-11 and S-12, respectively. Results demonstrated that normal and lognormal models were rarely optimal. Moreover, spICP-MS, SEM and TEM particle size distributions for both RMs were not fitted best by a single distribution from those that we considered. For RM 8013, only three models provided the best fits: the Hyperbolic, Generalized Tukey Lambda, and Generalized Hyperbolic for 65 %, 30 %, and 5 % of the

samples, respectively. For RM 8012, the best two models were the Hyperbolic (47 % of the samples) and the 3-parameter Weibull (18 % of the samples) distributions.

Figures 4 and S-9 show a comparison of the probability density of a sub-sample of 2500 NPs from analyses conducted by Lab 2 with the best fitting univariate model for RM 8013 and RM 8012, respectively. The hyperbolic distribution was the best fitting model for RM 8013 and provided a reasonably accurate representation of the measured size distribution, with the same actual and fitted mean (56.5 nm) and very close fitted standard deviation (5.2 nm vs. 5.3 nm). A similar high quality fit was observed for RM 8012 with the normal distribution. QQ-plots in both figures confirm the high quality fit.

Mixture of Gaussian distribution models. The particle size distributions were also modeled with mixtures of Gaussian distributions that could track their complex shapes such as the bumpy tail reflecting the presence of larger particles or coincident effects. The components of the optimal mixture of Gaussian distribution models for RM 8013 and RM 8012 were listed in Table S-13 and S-14, respectively. Overall, the mixture models produced the best fits for TEM, SEM and spICP-MS size distributions for the 70 % and 47 % of the samples for RM 8013 and RM 8012, respectively even though the BIC value calculation penalized more complex models.

The higher quality fit of the mixture Gaussian models compared to single distribution models is illustrated for Lab 2 spICP-MS results in Fig S-10 and S-11 for RM 8013 and RM 8012, respectively. The parameters of the mixture model for Lab 2 were included in Table S-13 and S-14. The visual comparison for RM 8013 of the optimal 3-component Gaussian mixture (Fig S-10) shows that the first component almost perfectly matches the main size distribution of the sample, whereas the other two components modeled the larger NPs in the tail. The more

comprehensive fitting strategy utilized in this study could improve the accuracy of modeling the size distributions of commercial NPs.

CONCLUSIONS

This paper summarizes the assessment of reliability and variability of the spICP-MS technique in the first *post hoc* interlaboratory study using NIST AuNP RMs (8012 and 8013). At first glance, the fact that the different users of the six participating laboratories did not follow a definite protocol may seem to be a shortcoming of the present study. However, it may also be viewed as a strength, in that the current state-of-the-art in spICP-MS analysis among knowledgeable and experienced laboratories can be evaluated and the relative importance of the various experimental parameters in the reliability and variability spICP-MS results can be investigated.

Several general important findings were made. The evaluation of the precision shows a generally lower variability of spICP-MS mean particle diameter than that reported for the reference size techniques for both NPs. While there is substantial variability among laboratories, the significantly lower within-laboratory standard deviation indicates consistent results. ANOVA reveals that the main factor affecting the global precision of spICP-MS is the residual contribution attributable to multiple uncontrolled factors. Nevertheless, the global spICP-MS results had similar mean particle diameter sizes and similar but broader size distributions compared to the reference techniques (measured in a single laboratory). The statistical approaches used to compare the nanoparticle size distributions in this study may also be useful for scientists needing to make size distributions comparisons in other studies.

Precision and trueness results revealed one of the existing challenges in spICP-MS, the greater difficulty in the size characterization and number quantification for smaller NPs. Robust results were obtained among different dwell times with regard to sizing and counting results but each dwell time has unique considerations that need to be taken into account for accurate measurements. BIC and QQ-plot criteria results demonstrated that normal and lognormal models were rarely the optimal model and that mixture models produce the best quality fits for the spICP-MS lab and user participants, albeit with small differences in BIC compared to the single distribution models. Thus, scientists should not assume that normal and lognormal distributions will best fit their data especially given that there was a small subpopulation of larger size particles.

Overall, the following further improvements of spICP-MS methodology are highly desirable: reducing the NP size detection limit, implementation of new software tools for more automated data processing, more robust determination of nebulization efficiency, and further development of standard protocols. The results from this study are intended to serve as a next step toward standardization of spICP-MS and to support ongoing spICP-MS standardization activities by the ISO Technical Committee 229 Nanotechnologies that will enable the transferability of the technique among laboratories.

TABLES

Table 1. Instrument factors and experimental conditions.

ICP-MS instrument	Agilent: 7500ce, 7700x Perkin Elmer: ELAN DRC-e, NexION 300Q, NexION 350D Thermo Scientific: X series 7, X series II, ELEMENT 2
Dwell time, td (ms)	0.1, 1, 3, 5, 10
Acquisition time (s)	10 to 360
Sample flow rate (mL min^{-1})	0.2 to 1.4
Transport efficiency, η (%)	1.4 to 17
Particle number concentration (Np L^{-1})	1.5×10^7 to 4.6×10^9
Au mass concentration (ng L^{-1})	3 to 950
Sonication time (min)	0 to 15

Table 2. Influence of dwell time on size spICP-MS results: mode, median and mean particle diameter for RM 8013.

Dwell time (ms)	Mode (nm)	Median (nm)	Mean (nm)[a]	Width (nm)[b]	NP Number & Au Mass Conc. Recovery (%)[c]
0.1	55.0	56.7	56.5 ± 0.1	12.3	86.6 ± 7.5
3	54.7	55.0	55.8 ± 0.2	12.0	93.9 ± 11.5
5	54.5	54.2	54.2 ± 0.1	11.9	NA
10	53.0	56.0	56.1 ± 0.1	7.8	79.3 ± 16.0

[a] Values indicate the mean and uncertainties are the expanded uncertainty of the mean on a 95% confidence level.

[b] The width of the size distribution is defined as the distance between the 84th and 16th percentiles of the particle diameters since it can be considered a robust analog of the length of the interval going from - 1 σ to + 1 σ.

[c] Uncertainties correspond to single standard deviations.

FIGURES

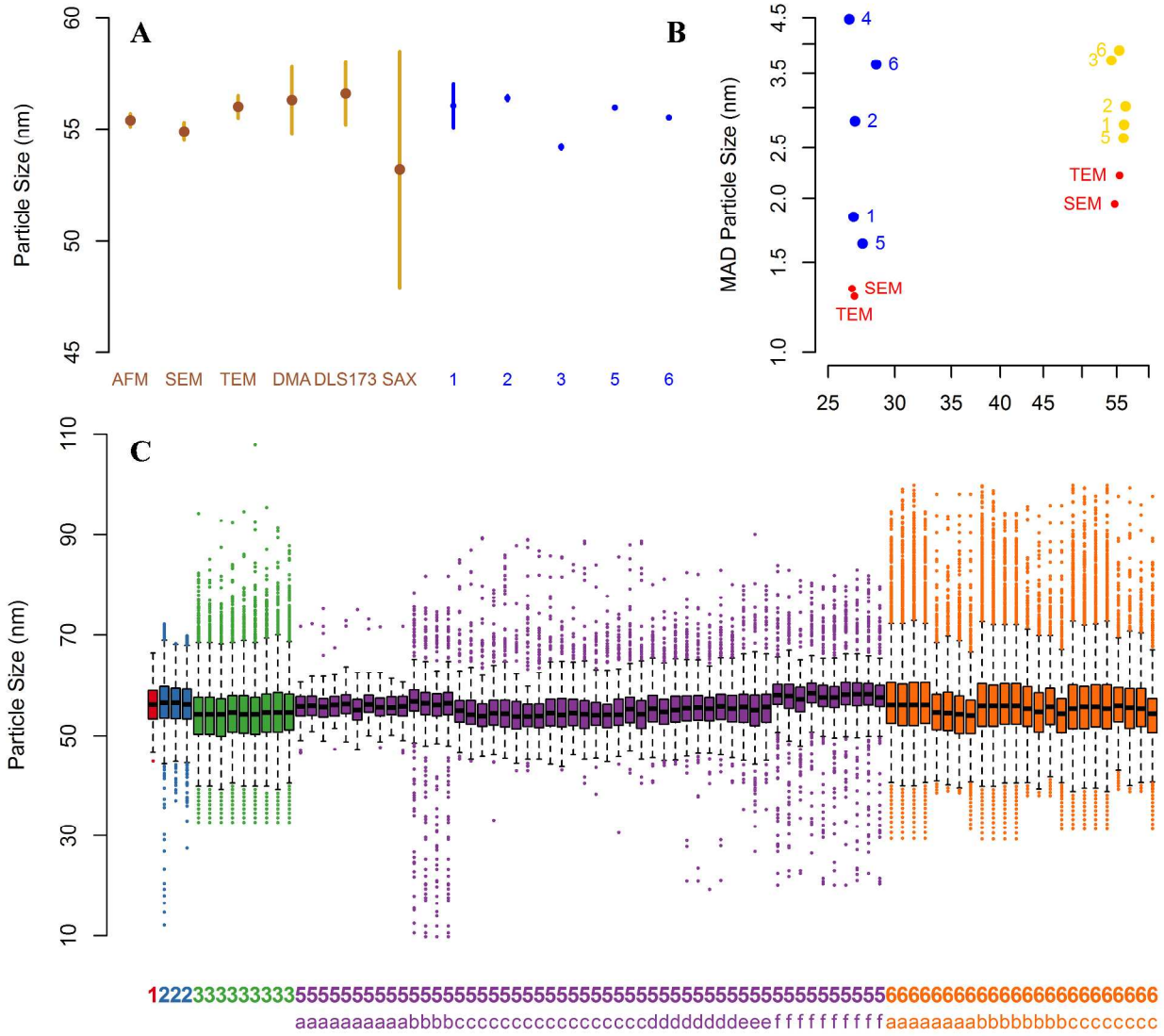


Figure 1. (A) Mean particle diameter for RM 8013 obtained by spICP-MS for all the laboratories (small blue dots), and as listed on the Report of Investigation for Atomic Force Microscopy (AFM), SEM, TEM, Differential Mobility Analysis (DMA), Dynamic Light Scattering at 173 degrees (DLS173) and Small Angle X-ray Scattering (SAX) (large brown dots). The vertical bars represent approximate 95 % coverage intervals for the true value of the diameter. For the spICP-MS results for Lab 5 and Lab 6 the vertical bars are smaller than the dots. **(B)** Spread versus

level plot of the median absolute deviation (MAD) vs level (indicated by the median of the measured diameters) for both RMs. Note that both axes have logarithmic scales. **(C)** Boxplots of each single spICP-MS run measurement of particle diameter included in this study for RM 8013. The thick horizontal line across each box marks the median of the corresponding particle size distribution; and the bottom and top of the box indicate the 25th and 75th percentiles. The whiskers of each box are 1.5 times the interquartile range and small dots correspond to data farther than the end of the whiskers. The x-axis indicates the laboratories (from 1 to 6) and the users within each laboratory (from a to f). Multiple boxplots for a laboratory/user indicate multiple replicates. Two outliers larger than 110 nm are not represented.

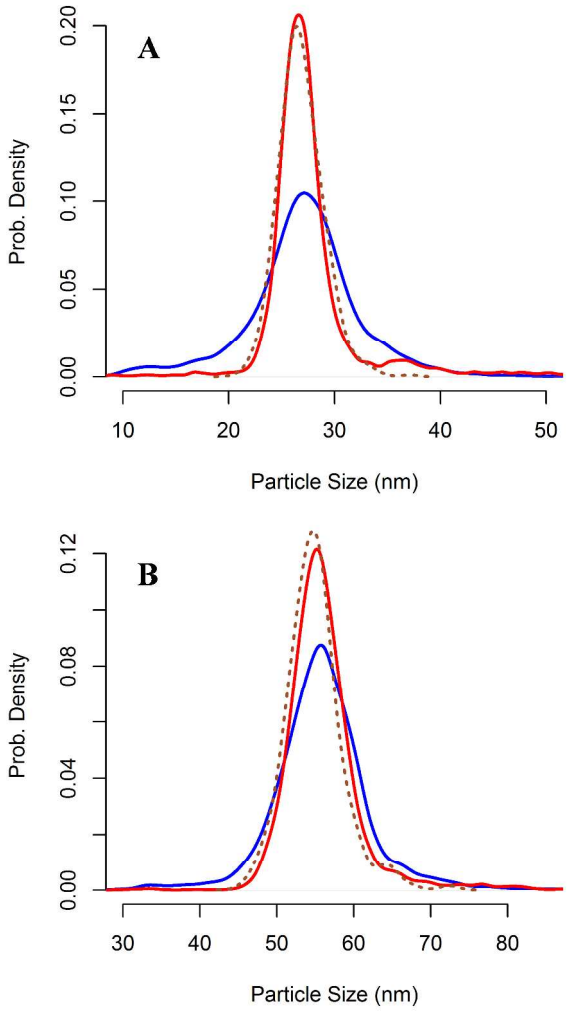


Figure 2. Probability density estimates for all the sets of data of spICP-MS (blue solid line) and assigned references TEM (red solid line) and SEM (red dotted line) particle size distributions for RM 8012 **(A)** and RM 8013**(B)**. The global probability density estimates were based on a pooled sample built from subsamples drawn uniformly at random, with replacement, from each participant's results.

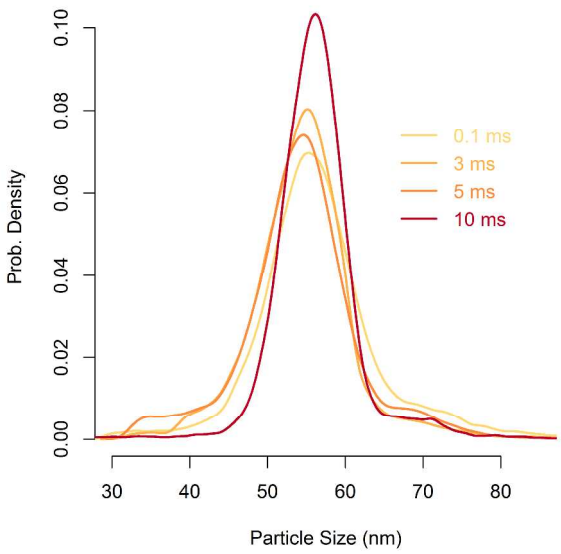


Figure 3. Probability density estimates for all the sets of data of spICP-MS measurements at different dwell times: 10, 5, 3, and 0.1 ms for RM 8013.

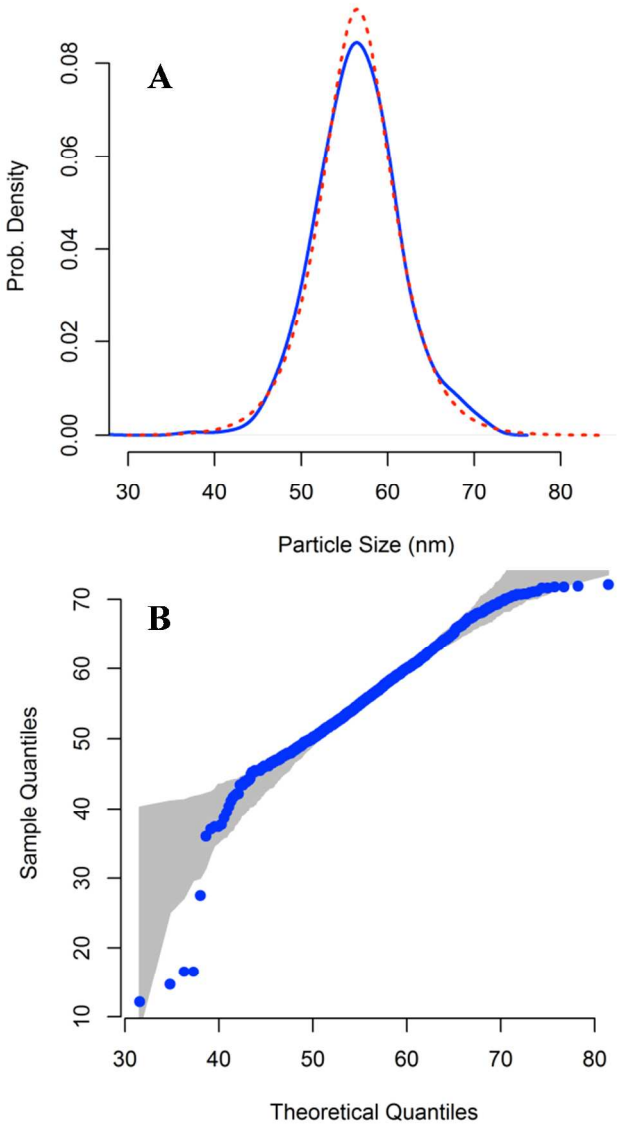


Figure 4. (A) Probability density estimates for particle size distribution results for RM 8013 from Lab 2 (blue line) and the probability density of the best fitting Hyperbolic model (dotted red line). **(B)** QQ-plot decorated with an approximate 95 % confidence band suggesting a general adequacy of the model to the data. A sub-sample of 2500 particles was used for the analysis.

ASSOCIATED CONTENT

Supporting Information Available

Additional information on data evaluation, influence of dwell time on spICP-MS results, tables and figures are provided. This material is available free of charge via the Internet at:

AUTHOR INFORMATION

Corresponding Author

*A.R. Montoro Bustos, phone: +1-3019754139 Fax: +1-3018690413 E-mail: antonio.montorobustos@nist.gov

ACKNOWLEDGMENTS

The authors wish to acknowledge the study participants, who so willingly shared their spICP-MS research results and information; we deeply appreciate their good cooperation and openness: Sebastian Sannac (Agilent, Les Ulis, FR), James Ranville (Colorado School of Mines, Department of Chemistry/Geochemistry, Golden, CO, USA), Chady Stephan (PerkinElmer Canada, Woodbridge, ON, CA), Geert Cornelis (University of Gothenburg, Department of Chemistry and Molecular Biology, Göteborg, SE), Francisco Laborda (University of Zaragoza, Institute of Environmental Sciences, Zaragoza, ES) and Jingyu Liu, Karen E. Murphy, Monique E. Johnson and Hind El Hadri (National Institute of Standards and Technology, Gaithersburg, MD, USA). Certain commercial products or equipment are described in this paper in order to specify adequately the experimental procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that it is necessarily the best available for the purpose.

REFERENCES

(1) International Organization for Standardization (ISO), ISO TS 27687, *Geneva, Switzerland,* **2008**.

(2) Nowack, B., Ranville, J.F., Diamond, S., Gallego-Urrea, J.A., Metcalfe, C., Rose, J., Horne, N., Koelmans, A.A., Klaine, S. J. *Environ. Toxicol. Chem.,* **2012,** 31, 50-59.

(3) Petersen, E., Henry, T.B., Zhao, J., MacCuspie, R.I., Kirschling, T.L., Dobrovolskaia, M.A., Hackley, V., Xing, B., White, J.C. *Environ. Sci. Technol.,* **2014,** 48, 4226-4246.

(4) Nel, A., Xia, T., Mädler, L., Li, N. *Science,* **2006,** 311, 622-627.

(5) Gottschalk, F.; Sonderer, T.; Scholz, R.W.; Nowack, B. *Environ. Sci. Technol.,* **2009,** 43, 9216-9222.

(6) Winchester, M.R., Sturgeon, R.E., Costa-Fernandez, J.M. *Anal. Bioanal. Chem.*, **2010,** 396, 951-952.

(7) Howard, A.G. *J. Environ. Monit.,* **2010,** 12, 135-142.

(8) European Commission, 2011/696/EU: Commission Recommendation of 18 October 2011 on the definition of nanomaterial, *Off. J. Eur. Communities: Legis.,* **2011,** 275, 38-40.

(9) Linsinger, T., Roebben, G., Gilliland D., Calzolai L., Rossi F., Gibson N., Klein C. Requirements on measurements for the implementation of the European Commission definition of the term 'nanomaterial'. *Joint Research Centre, Institute for Reference Materials and Measurements,* **2012**.

(10) Calzolai, L., Gilliland, D., Rossi, F. *Food Addit. Contam., Part A,* **2012**, 29, 37-48.

(11) Hassellöv, M., Readman, J.M., Ranville, J.F., Tiede, K. *Ecotoxicology*, **2008,** 17, 344-361.

(12) Tiede, K., Boxall, A.B.A.; Tear, S.P., Lewis, J., David, H., Hassellöv, M. *Food. Addit. Contam*., **2008,** 25, 795-821.

(13) Montoro Bustos, A.R., Ruiz Encinar, J., Sanz-Medel, A. *Anal. Bioanal. Chem.,* **2013**, 405, 5637-5643.

(14) Degueldre, C.; Favarger, P.Y.; Wold, S. *Anal. Chim. Acta,* **2006,** 555, 263-268.

(15) Sadik, O.A., Du, N., Kariuki, V., Okello, W., Bushlyar, V. *ACS Sustainable Chem. Eng.,* **2014,** 2, 1707-1716.

(16) Laborda, F., Bolea, E., Jiménez-Lamana, J. *Anal. Chem.,* **2014,** 86, 2270-2278.

(17) Laborda, F., Jiménez-Lamana, J., Bolea, E., Castillo, J.R. *J. Anal. At. Spectrom.,* **2011,** 26, 1362-1371.

(18) Reed, R.B.; Higgins, C.P.; Westerhoff, P.; Tadjiki, S.; Ranville, J.F. *J. Anal. At. Spectrom.,* **2012,** 27, 1093-1100.

(19) Tuoriniemi, J.; Cornelis, G.; Hassellöv, M. *Anal. Chem.,* **2012,** 84, 3965-3972.

(20) Pace, H.E.; Rogers, N.J.; Jarolimek, C.; Coleman, V.A.; Gray, E.P.; Higgins, C.P.; Ranville, J.F. *Environ. Sci. Technol.,* **2012**, 46,12272-12280.

(21) Gray, E.P.; Coleman, J.G.; Bednar, A.J.; Kennedy, A.J.; Ranville, J.F.; Higgins, C.P. *Environ. Sci. Technol.,* **2013,** 47, 14315-1423.

(22) Peters, R.B., Rivera, Z., Bemmel, G., Marvin, H.P., Weigel, S., Bouwmeester, H. *Anal. Bioanal. Chem.,* **2014,** 406, 3875-3885.

(23) Pace, H.E.; Rogers, N.J.; Jarolimek, C.; Coleman, V.A; Higgins, C.P.; Ranville, J. F. *Anal. Chem.,* **2011,** 83, 9361-9369.

(24) Laborda, F., Jiménez-Lamana, J., Bolea, E., Castillo, J.R. *J. Anal. At. Spectrom.,* **2013,** 28, 1220-1232.

(25) Tuoriniemi, J., Cornelis, G., Hassellöv, M. *J. Anal. At. Spectrom.,* **2014,** 29, 743-752.

(26) Olesik, J. W., Gray, P.J. *J. Anal. At. Spectrom.,* **2012,** 27, 1143-1155.

(27) Cornelis, G., Hassellöv, M. *J. Anal. At. Spectrom.,* **2014,** 29, 134-144.

(28) Liu, J., Murphy, K.E., MacCuspie, R.I., Winchester, M. R. *Anal. Chem.,* **2014,** 86, 3405-3414.

(29) Montaño, M.D., Badiel, H. R., Barzagan, S., Ranville, J.F. *Environ. Sci. Nano,* **2014,** 1, 338-346.

(30) Mitrano, D. M., Barber, A., Bednar, A., Westerhoff, P., Higgins, C. P., Ranville, J. F. *J. Anal. At. Spectrom.,* **2012,** 27, 1131-1142.

(31) Krystek, P., Ulrich, A., Garcia, C.C., Manohar, S., Ritsema, R. *J. Anal. At. Spectrom.,* **2011,** 26, 1701-1721.

(32) Linsinger, T.P.J., Peters, R.B., Weigel, S. *Anal. Bioanal. Chem.,* **2014,** 406, 3835.

(33) Cooper, H., Hedgesm L.V., Valentine, J.C. *The Handbook of Research Synthesis and Meta-Analysis, Ed. Russell Sage Foundation Publications,* **2009**.

(34) Hineman, A., Stephan, C. *J. Anal. At. Spectrom.,* **2014,** 29, 1252-1257.

(35) NIST, Reference Material® 8012 Gold Nanoparticles, Nominal 30 nm Diameter; *National Institute of Standards and Technology,* **2012**.

(36) NIST, Reference Material® 8013 Gold Nanoparticles, Nominal 60 nm Diameter; *National Institute of Standards and Technology,* **2012**.

(37) Rice, S.B., Chan, C., Brown, S.C., Eschbach, P., Han, L., Ensor, D.S., Stefaniak, A.B., Bonevich, J., Vladar, A.E., Walker, A.R.H., Zheng, J.W., Starnes, C., Stromberg, A., Ye, J., Grulke, E.A. *Metrologia,* **2013,** 50, 663-678.

(38) Pinheiro, J.C., Bates, D.M., *Mixed-Effects Models in S and S-Plus, Ed. Springer-Verlag,* **2014**.

(39) ISO, ISO 5725, *Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 1: General Principles and Definitions*, *Geneva, Switzerland,* **1994**.

(40) Anderson, T. W. *Ann. of Math. Stat.,* **1962,** *33,* 1148-1159.

(41) Chakravarti, I. M., Laha, R.G., Roy, J. *Handbook of methods of applied statistics, Ed. John Wiley and Sons,* **1974,** 1, 392.

(42) Kiss, L. B., Söderlund, J., Niklasson G.A., Granqvist, C.G. *Nanotechnology,* **1999,** 10, 25-28.

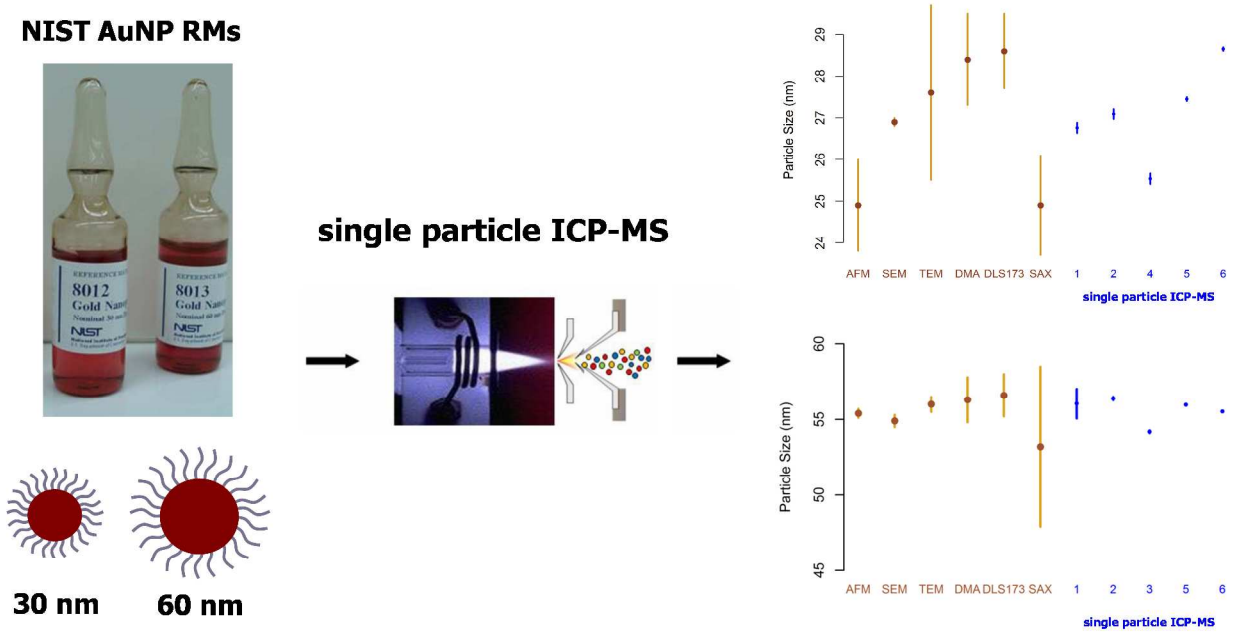
(43) Burnham, K.P., Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Ed. Springer-Verlag,* **2002**.

(44) Silverman, B. M. *Density Estimation for Statistics and Data Analysis, Ed. Chapman and Hall,* **1986**.

(45) Wood, R. *Trends in Anal. Chem.,* **1999,** 18, 624-632.

(46) Sekine, R., Khurana, K., Vasilev, K., Lombi, E., Donner, E. *Nanotoxicology,* **2015,** *in press* doi:10.3109/17435390.2014.994570.

For TOC only


NIST AuNP RMs
8012
8013
single particle ICP-MS
30 nm
60 nm
Particle Size (nm)
AFM SEM TEM DMA DLS173 SAX 1 2 4 5 6
single particle ICP-MS
Particle Size (nm)
AFM SEM TEM DMA DLS173 SAX 1 2 3 5 6
single particle ICP-MS