Supplementary Information (SI) for Analytical Methods.
This journal is © The Royal Society of Chemistry 2024

Supplemental Information for Measurement Bias in spICP-TOFMS: Insights from Monte Carlo Simulations

Raven L. Buckman, Alexander Gundlach-Graham
Department of Chemistry, Iowa State University, Ames, Iowa
Address correspondence: alexgg@iastate.edu

Table of Contents

Simulation Conditions

Table S1: Summary of the parameters used in the comparison of simulated and experimental results for CeO_2, ferrocerium mischmetal, and bastnaesite minerals. The number of simulated particles, user-defined median particle diameters, log-normal shape parameters, mass fraction relative standard deviations (w_{RSD}), PSD median estimators, and multiplicative standard deviations are listed here for each particle type simulated.

Simulation Parameters	CeO_2	Ferrocerium	Bastnaesite
Number of Particles Simulated	10000	10000	10000
Median Particle Diameter (μ, nm)	42	30	35
Shape Parameter (σ)	0.33	0.27	0.80
w_{RSD}	--	0.50	0.15
Median Estimator (μ)	42.2	31.8	35.5
Multiplicative Standard Deviation (s)	1.4	1.4	1.7

*Median estimator and multiplicative standard deviations were calculated based on arithmetic mean and standard deviations as defined elsewhere.[1]

Table S2: For ferrocerium and bastnaesite simulated particles, a log-normal distribution of mass fraction values was applied with the following medians, means, and standard deviations (SD).

Particle Type	Element	Median	*Mean	*SD
Ferrocerium	Ce	0.677	0.894	0.384
	La	0.323	0.620	0.180
Bastnaesite	Ce	0.390	0.665	0.062
	La	0.196	0.493	0.033
	Nd	0.170	0.463	0.029
	Pr	0.058	0.290	0.010
	Th	0.004	0.091	0.001

*Mean and SD calculated based on the given RSD and *w* value for a log-normal distribution as defined elsewhere.[2]

Statistical Evaluation of Relevance

To evaluate the statistical similarity of the simulation data with respect to the experimental data, a Mood's median test (AKA test for equality of location) was performed.[3-5] The median was selected as the most representative measure of central tendency to avoid bias from outlier values and local modes.[4] Additionally, the Mood's median test is a nonparametric test, which allows for the direct comparison of the models without the assumption of normality.[3]

For this statistical evaluation, several Monte Carlo simulations of spICP-TOFMS data from CeO_2 NPs were performed until the center of the simulated distribution appeared, visually, to be within a reasonable range of the center of the experimental distribution. A Mood's median test was performed on subsequent simulated distributions until the median of the simulation was no longer significantly different from the experimental. This median particle diameter was determined to be 42 nm. Two simulations were performed, one with a log-normal and another with a normal PSD; descriptive statistics are summarized for both simulations and the experimental results in Table S3.

Table S3: Descriptive statistics of the Ce mass distribution data collected with Monte Carlo simulation and spICP-TOFMS measurements of CeO_2. Additionally, test statistics from the Mood's median test are provided here (Chi-Square and p-value).

	Simulation (Log-Normal)	**Simulation (Normal)**	**Experimental**
N	9651	8211	1074
Mean (fg)	0.755	0.409	0.412
Std. Dev. (fg)	2.69	0.546	0.690
Med. (fg)	0.248	0.216	0.228
Chi-Square	2.44	0.999	N/A
p > Chi-Square	0.118	0.317	N/A

A comparison of the means of the log-normal and normal distributions to the experimental results indicates that the normal distribution demonstrates a more accurate simulation. Despite this, we chose to use a log-normal distribution for a consistent comparison to the other particle types described in the main body. The Mood's median test showed a larger chi-squared value for the log-normal distribution than the normal distribution. The Ce mass distributions as well as box-and-whisker plots of the normal simulation compared to the experimental are shown in Figure S1.

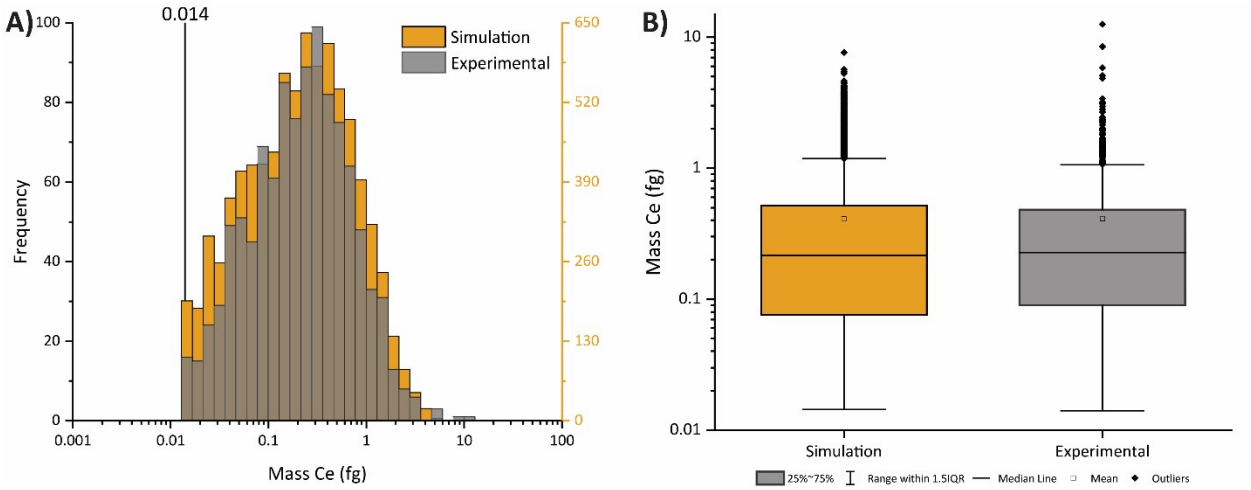

Figure S1: A) Comparison of CeO_2 Monte Carlo simulated particles (orange) to experimentally obtained measurements (grey). Here, the particles are simulated according to a normal distribution with a mean of 37 nm and a standard deviation of 25 nm. B) Box and whisker plot comparing the normal distribution of the simulated (left) and experimental (right) data for CeO_2 particles. The inner quartile range (IQR), medians, and means are within good agreement.

Mass Fraction Variance

For our Monte Carlo simulated particles with mass fractions set to a fixed value, the mass ratios closely follow the Poisson confidence interval (Figure S2, blue circles). To model the distribution of mass ratios within the ferrocerium and bastnaesite mischmetal particle types more accurately, a distribution of mass fractions was introduced to the Monte Carlo simulations (Table S2). For the purposes of this study, the mass fractions were generated according to a log-normal distribution with medians equal to those specified in Tables 1 and S2. PSD shape parameters were set equal to the product of the elemental mass fraction and the relative standard deviation. When compared to the fixed value mass fraction, the distribution clearly shows a broader range of mass ratios (Figure S2, orange circles) and a lack of clearly conserved mass ratio at higher mass values. This characteristic shape is a combination of the critical value thresholding of Ce (orange line) and La (blue line), as well as the predicted Poisson confidence interval.

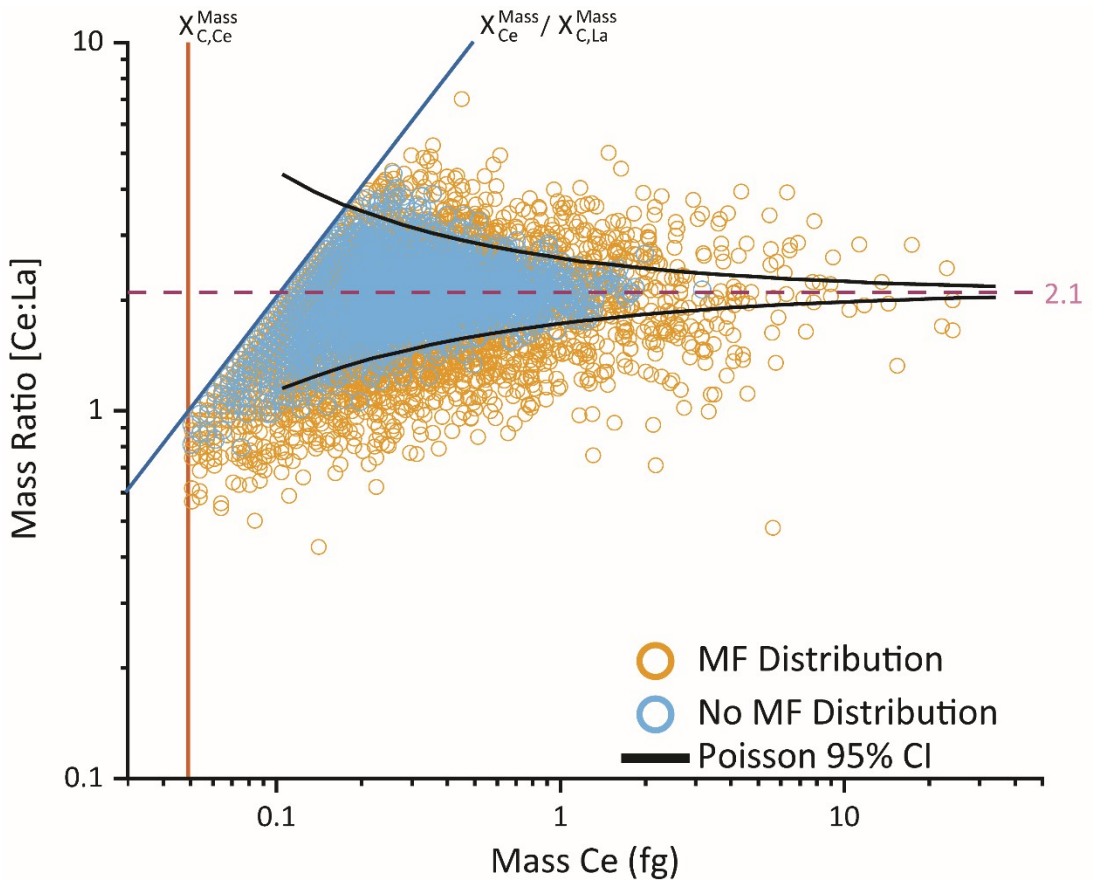

Figure S2: Simulated ferrocerium particles from a log-normal distribution with median diameter of 30 */ 1.4 nm with (orange) and without (blue) an applied distribution to the mass fractions of Ce and La. Lower bounds are dictated by the critical values of Ce (orange line) and La (blue line). Monte Carlo simulated Poisson confidence bands are shown in black.

Bastnaesite Mischmetal Simulations

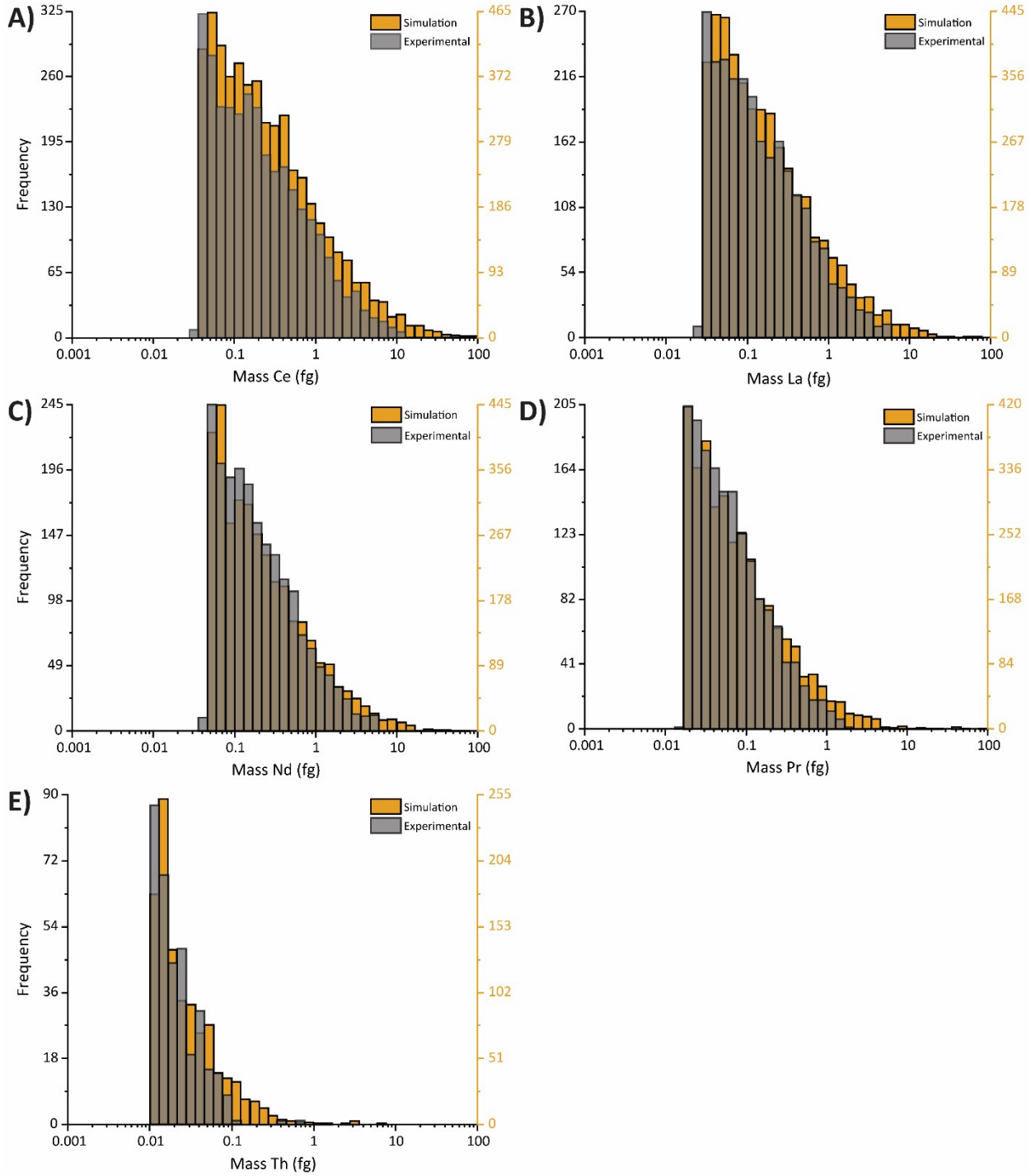

Figure S3: A comparison of the mass distributions of elements detected in Monte Carlo simulated (orange) bastnaesite particles and those found in particles measured, experimentally, by spICP-TOFMS (grey). Elements of interest in bastnaesite particles are Ce, La, Nd, Pr, and Th; these elements are shown in A, B, C, D, and E, respectively.

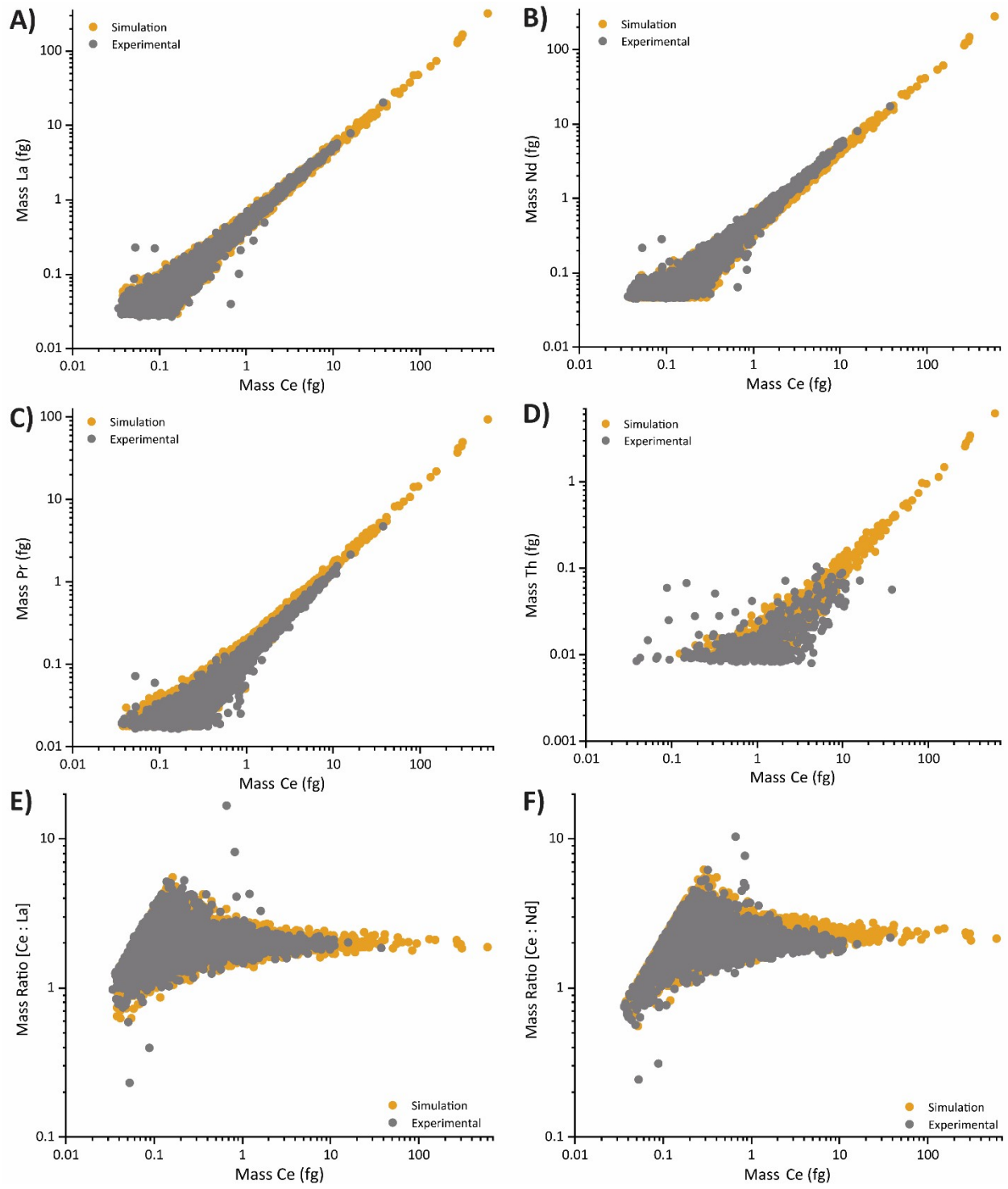

Figure S4: A comparison of the correlations between Ce and (A) La, (B) Nd, (C) Pr, and (D) Th in Monte Carlo simulated (orange) bastnaesite particles and those experimentally obtained with spICP-TOFMS (grey). Also featured here are the comparisons of the mass ratios of (E) Ce to La and (F) Ce to Nd.

Bias in spICP-TOFMS Measurements

Table S4: Summary of the simulation parameters used to investigate bias in spICP-TOFMS measurements as well as the effects of Poisson noise and critical value thresholding. The number of simulated particles, user-defined median particle diameters, log-normal shape parameters, particle size distribution median estimators, as well as multiplicative and additive standard deviations are listed here for the relevant figures.

Additional Simulations	**Figure 5A**	**Figure 5B**	**Figure 5C**	**Figure 5D**	**Figure S6**	**Figure S7**
Number of Particles Simulated	10000	10000	10000	10000	10000	10000
Median (μ)	20	35	35	60	1-200	35
Sigma (σ)	0.40	0.40	0.80	0.40	0.60	0.40
Median Estimator (μ*)	19.9	35.0	35.0	60.3	1 - 202	34.9
Multiplicative SD (s*)	1.3	1.3	1.7	1.3	1.3	1.0 - 1.9
Additive SD (s)	8.9	15.9	45.6	27.6	0.451 - 91.1	0.350 - 75.1

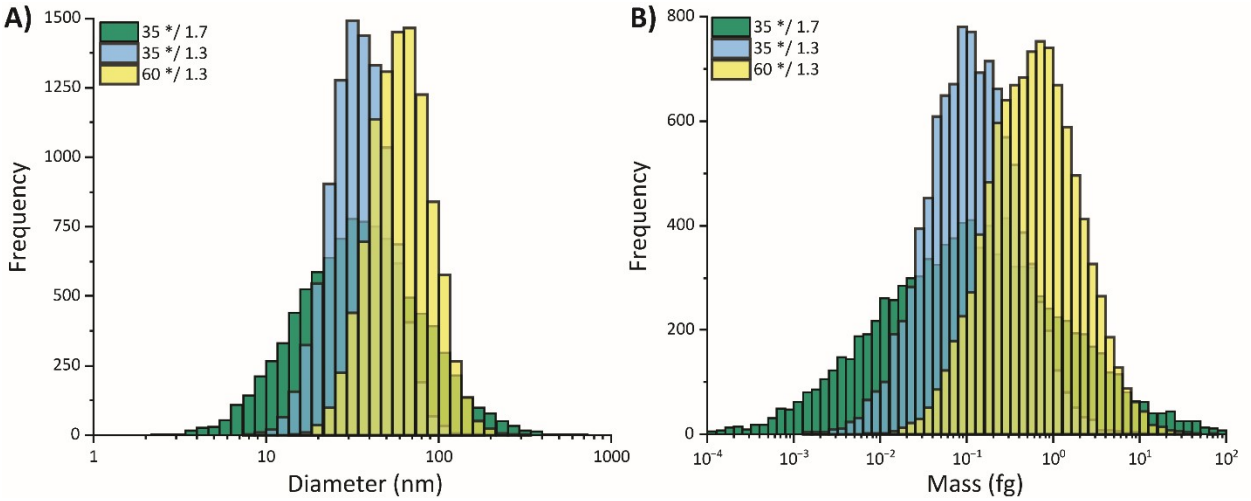

Figure S5: A visualization of the influence of simulation parameters on the particle size distribution as a function of diameter (A) and whole particle mass (B). Particles masses are calculated based on the density of bastnaesite (5.12 g cm^{-3}). Particle size distributions with a median diameter of 35 (green and blue) are compared to a distribution with a median diameter of 60 (yellow). Additionally, simulated distributions with a multiplicative standard deviation of 1.3 nm (blue and yellow) are compared to a simulation with a standard deviation of 1.7 nm.

Table S5: Log-normal cumulative distribution function[2] fit parameters for the mass recovery and number of detectable particles (No. Detected) from Figure 6A.

Model	Mass Recovery	No. Detected
Equation	$y = y_0 + A\int_0^x \frac{1}{\sqrt{2\pi}wt} e^{\frac{\ln(t-x_c)^2}{2w^2}} dt$	$y = y_0 + A\int_0^x \frac{1}{\sqrt{2\pi}wt} e^{\frac{\ln(t-x_c)^2}{2w^2}} dt$
y_0	-0.3 ± 0.4	0.1 ± 0.2
A	100.3 ± 0.5	99.8 ± 0.4
x_c	3.087 ± 0.005	3.467 ± 0.003
w	0.452 ± 0.007	0.406 ± 0.004
Reduced Chi-Square	0.363	0.171
R-Squared (COD)	0.9998	0.9999
Adj. R-Square	0.9998	0.9999

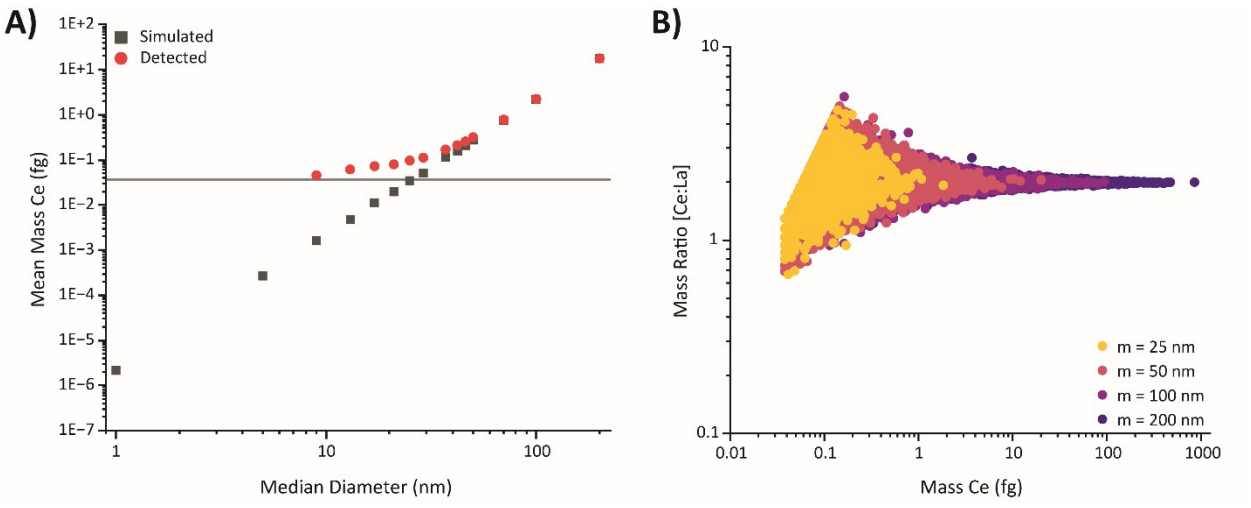

Figure S6: A) For Monte Carlo simulated spICP-TOFMS data from bastnaesite particles with variable median diameters and a multiplicative standard deviation of approximately 1.3, the detected mean mass of Ce is compared to the true mean mass of Ce. The grey horizontal line is indicative of the critical mass of Ce ($^{X}_{C,sp,Ce}^{Mass} \approx 0.036$ fg). B) The detectable mass ratios of Ce to La are plotted as a function of Ce mass (in fg) for four median diameters

In Figure S7A, as the shape parameter (sigma) increases, the PSDs broaden, and the global maxima frequency decreases. For a broad distribution of particle sizes, more extreme values are generated, and the total number of detectable particles will decrease while the mass recovery will increase (see Figure S7B). Due to the log-normal distribution, the measures of central tendency will be greatly overestimated as the low-mass particles will not be above $L_{C,sp}$ and consequently, cannot be considered in the calculation of measurands. Unlike the constant true mean observed in Figure S6, for varying sigma values the true mean values change as a function of the sigma value. This can be attributed to the log-normal distribution and the dependence of the mean on the sigma value.[2] For low dispersion PSDs, the true and detected masses of Ce are overestimated by approximately 10%. From Figure S7E, it would appear that the median is a marginally better metric for narrow distributions, however, as the PSDs become broader, the mean and medians deviate from the true values equally. In the case of extremely broad PSDs, the mean is more accurate than the median due to the dependent relationship of the mean and sigma values of a log-normal distribution.

The spread of the PSDs also has a dramatic effect on the detected mass fractions of elements within spICP-TOFMS measurements (see Figure S7F); scatter points are representative of calculated averages and error bars represent appropriate standard deviations. When sigma is between to 0.01 and 0.1 (equivalent to additive standard deviations of 0.350 and 3.50 nm, respectively) the mass fractions of all elements are grossly overestimated; these sigma values yield the greatest percentage of detectable particles with the lowest mass recoveries. These extreme deviations in mass fraction values will ultimately results in the incorrect reporting of detected particle stoichiometries and elemental correlations. While the mass recovery increases and the number of particles decrease of PSDs with sigma values between 0.1 and 1 (additive standard deviations of 3.50 and 75.1 nm), the mass continue to be overestimated. Because the low-mass particles are not detectable, the calculation of mass fractions are ultimately biased due to the skew imposed by ion-counting statistics. The data shown in Figure S7 was simulated with a median particle diameter of 35 nm, or equal to the critical diameter based on Ce. These results indicate that while the PSD changes as a function of the shape parameter, the overestimation of mass fraction values are also, in part, influenced by Poisson noise.

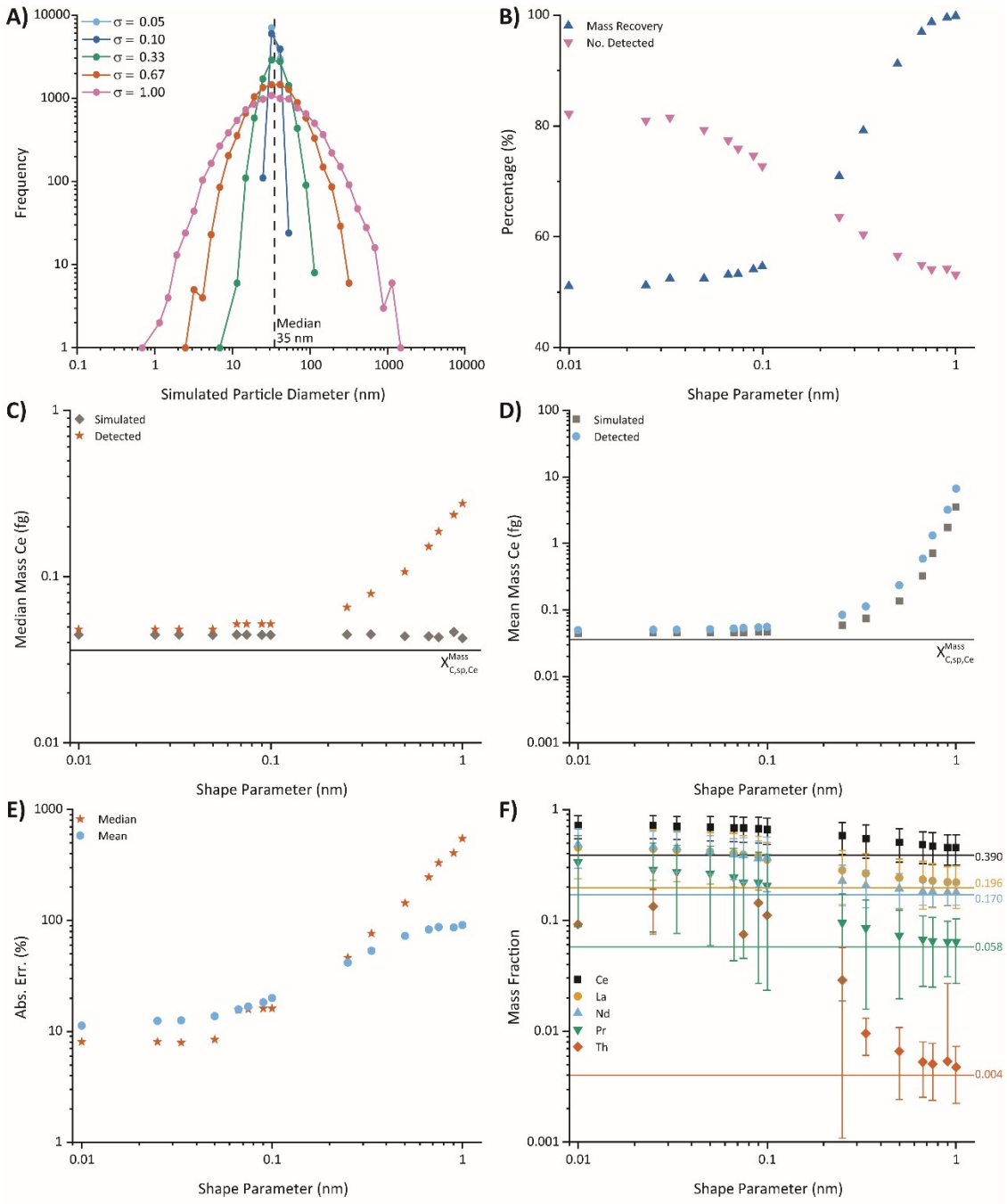

Figure S7: Summary results of Monte Carlo simulated bastnaesite particles with variable shape parameters (sigma) and median diameters of approximately 35 nm. A) PSD shapes for five sigma values. B) The number of particles and mass recovery of detectable particles. C-D) The detected median and mean masses of Ce compared to the true simulated values. The grey horizontal lines are indicative of the critical mass of Ce ($X_{C,sp,Ce}^{Mass} \approx 0.036$ fg). E) The absolute error (%) of the detected mean and medians. F) The calculated average mass fractions for each simulated element as a function of the sigma values; standard deviations from the simulation of 10,000 particle events are represented as error bars.

References

1. E. Limpert, W. A. Stahel and M. Abbt, *BioScience*, 2001, **51**, 341-352.
2. N. Heckert, J. Filliben, C. Croarkin, C. Hembree, W. Guthrie, P. Tobias and J. Prinz, in *NIST/SEMATECH e-Handbook of Statistical Methods*, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, 2002, DOI: 10.18434/M32189, ch. 1.3.6.6.9 Lognormal Distribution.
3. I. R. Savage, *Journal of the American Statistical Association*, 1957, **52**, 331-344.
4. R. W. Cooksey, in *Illustrating Statistical Procedures: Finding Meaning in Quantitative Data*, ed. R. W. Cooksey, Springer Singapore, Singapore, 2020, DOI: 10.1007/978-981-15-2537-7_5, pp. 61-139.
5. G. J. G. Upton and I. Cook, *A dictionary of statistics*, Oxford University Press, Oxford, Third edition. edn., 2014.