Supplementary Information

"Improved Uranium Particle Analysis by SIMS using O3- Primary Ions"

Evan E. Groopman*[1], Todd L. Williamson[1], and David S. Simons[2]

[1]Materials Measurement Science Division, National Institute of Standards and Technology, Gaithersburg, MD, 20899, USA
[2]MELE Associates, Inc., Rockville, MD, 20850, USA
*evan.groopman@nist.gov

A. Appendix: Biased and Unbiased Weighted Mean Uncertainties

1. Biased and Unbiased Standard Errors of the Weighted Mean

Typically, the weighted mean (WM) and its variance are given by:

$$WM = \bar{x} = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\sum_{i=1}^{n} w_i} \#(1)$$

$$Var(x) = \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} w_i} \times \frac{n}{n-1} \#(2)$$

where x_i and w_i are the individual measurement and its weight, and n is the number of measurements. For isotope ratios, the weights are typically the inverse square of the measurement uncertainties. The standard error of the weighted mean (SE_{WM}) follows as $\sqrt{Var}/\sqrt{n}$. However, this may be biased and underestimated when the uncertainties on the data points are not equal. Imagine, as an extreme example, a dataset (e.g., isotope ratios) consisting of a very precisely known point and several very imprecise points. Naturally, the weighted mean will be heavily influenced by the precise datum and will be negligibly influenced by the others. In this case, n does not reflect the true degrees of freedom of the weighted mean, which is better approximated by some *effective n* much closer to 1 [1, 2]. Therefore, calculating the SE_{WM} using n would underestimate the true uncertainty. The *effective n* is calculated:

$$n_{eff} = \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} \left(w_i^2\right)} \#(3)$$

n_{eff} tends towards the number of dominant points with equal weights; in our hypothetical, this is close to 1, but may be some number up to n, depending on the data. n_{eff} is then substituted for n in the equations for the variance and SE_{WM} to calculate the unbiased values. In keeping with our hypothetical scenario, one

could conceivably produce highly uncertain isotope ratios by reducing the measurement time to some infinitesimally small value. These are real data, not fabrications, and they do not affect the weighted mean when combined with the higher precision point(s). However, it would clearly be improper to suppose that the uncertainty on the weighted mean would be better constrained by adding more data with arbitrarily large uncertainties. Since $n_{eff} \leq n$, the unbiased SE_{WM} will always be larger than the biased version with unequal data weights. The data in this paper are not so highly biased, but the uncertainties on the WMs are expanded relative to their biased estimates due to the presence of unequal weights on isotope ratios (e.g., from analyses of smaller particles, which have fewer atoms).

The next two sections demonstrate the difference in the unbiased and biased SE_{WM} on simulated data from the extreme case discussed above, and from real data shown in Figure 2 from the main paper.

2. Simulated Extreme Data

Table A-1 shows simulated data that illustrate an extreme example of divergence between the biased and unbiased SE_{WM}. The simulated data consist of two precise isotope ratios and 25 imprecise ratios. The precise data points are the ratio of two isotopes, X and Y, with average intensities of 10^4 and 10^6 counts·s^{-1}, respectively, each measured for 1 s (highlighted green). The given counts were randomly drawn from Poisson distributions matching these criteria. These produce ratios of approximately X/Y = 0.010 ± 1×10^{-4} each. The imprecise points represent 25 measurements of the same signal intensities for durations of 10 µs each. Twenty-five random samples (counts) were drawn from similar Poisson distributions. For the X

Table A-1: Simulated data illustrating the difference between the *biased* and *unbiased* SE_{WM}. The simulated data consist of: 1) two 1 s measurement of two isotopes, X and Y, that have mean count rates of 10^4 and 10^6 cps, respectively, and 2) 25 measurements of the same isotopes with identical signal intensities lasting 10 µs each. The 10 µs measurements are highly uncertain and add a total of 5 counts to the X isotope, but reduce the *biased* SE_{WM} by a factor of ~2.4×. The *unbiased* SE_{WM} is expanded to accommodate the uncertainty of averaging many imprecise data. "--" indicates no data.

N	X counts	Y counts	FC mid X	FC er X	FC mid Y	FC er Y	X (×10^4 cps)	Y (×10^5 cps)	X/Y	er	|W| (×10^{-8})
1	10055	1001700	--	--	--	--	1.01	10.02	0.010	1×10^{-4}	0.5×10^8
2	9925	999100	--	--	--	--	0.99	9.99	0.010	1×10^{-4}	0.5×10^8
3	1	8	1.19	1.56	3.01	8.31	11.92	3.01	0.396	1.210	0.3
4	1	12	1.19	1.56	3.73	12.56	11.92	3.73	0.319	1.153	0.4
5	1	10	1.19	1.56	3.51	10.29	11.92	3.51	0.339	1.087	0.4
6	0	13	0.64	0.64	4.01	13.29	6.38	4.01	0.159	0.550	1.7
7	0	13	0.64	0.64	4.01	13.29	6.38	4.01	0.159	0.550	1.7
8	1	10	1.19	1.56	3.51	10.29	11.92	3.51	0.339	1.087	0.4
9	0	16	0.64	0.64	4.23	16.57	6.38	4.23	0.151	0.609	1.4
10	0	4	0.64	0.64	2.22	4.56	6.38	2.22	0.287	0.656	1.2
11	0	14	0.64	0.64	4.01	14.31	6.38	4.01	0.159	0.590	1.5
12	0	13	0.64	0.64	4.01	13.29	6.38	4.01	0.159	0.550	1.7
13	0	10	0.64	0.64	3.51	10.29	6.38	3.51	0.181	0.561	1.6
14	0	5	0.64	0.64	2.53	5.28	6.38	2.53	0.252	0.584	1.5
15	0	9	0.64	0.64	3.23	9.56	6.38	3.23	0.198	0.618	1.3
16	0	13	0.64	0.64	4.01	13.29	6.38	4.01	0.159	0.550	1.7
17	0	8	0.64	0.64	3.01	8.31	6.38	3.01	0.212	0.623	1.3
18	0	6	0.64	0.64	2.73	6.55	6.38	2.73	0.234	0.608	1.4
19	0	10	0.64	0.64	3.51	10.29	6.38	3.51	0.181	0.561	1.6
20	0	10	0.64	0.64	3.51	10.29	6.38	3.51	0.181	0.561	1.6
21	0	6	0.64	0.64	2.73	6.55	6.38	2.73	0.234	0.608	1.4
22	1	12	1.19	1.56	3.73	12.56	11.92	3.73	0.319	1.153	0.4
23	0	12	0.64	0.64	3.73	12.56	6.38	3.73	0.171	0.600	1.4
24	0	5	0.64	0.64	2.53	5.28	6.38	2.53	0.252	0.584	1.5
25	0	8	0.64	0.64	3.01	8.31	6.38	3.01	0.212	0.623	1.3
26	0	15	0.64	0.64	4.00	15.32	6.38	4.00	0.159	0.630	1.3
27	0	12	0.64	0.64	3.73	12.56	6.38	3.73	0.171	0.600	1.4
								n (n_{eff})	**WM**	**SE_{WM}**	
							Biased	27	0.010	2.2×10^{-5}	
							Unbiased	2.00	0.010	1.1×10^{-4}	
							% diff.		0%	500%	

isotope, this resulted in 5 measurements of a single count and 20 measurements with no counts. Feldman-Cousins (FC) confidence intervals at the 68% level were used to estimate the uncertainties for each isotope [3, 4]. The midpoint of each FC confidence interval was set as the number of counts and half the interval width was assigned to be the uncertainty. This is not an accurate interpretation of the confidence intervals, but suffices for this demonstration to produce uncertainties that can be incorporated into the WM. In all instances, the uncertainties on the X/Y ratios are larger than the ratio value. The last column shows the normalized weight, |W|, for each measurement. The first measurements are essentially 0.5, which the remainder are on the order of 10^{-8}.

The resulting biased SE_{WM} was 2.2×10^{-5}, remarkably ~2.4× less than the SE_{WM} of only the first two points (5.2×10^{-5}). In contrast, the average number of X and Y counts, respectively, over the 25 short measurements were 0.2 ± 0.4 (1sd) and 10.2 ± 3.3 (1sd). This corresponded to an X/Y ratio of 0.02 ± 0.04. Using the FC estimates, the unweighted average and standard error of the X/Y ratios was 0.21 ± 0.02 (sd = 0.9). It is well known that biases can be introduced in isotope ratios when either averaging ratios or summing all counts and then averaging, but this is beside the point here [5]. There is no legitimate reason why these imprecise data should *reduce* the uncertainty on the WM when their unweighted average and variance should be perturbing the distribution's center away from the WM. By including the imprecise data in our biased SE_{WM}, we implicitly assert that the data have more degrees of freedom than they truly do. An inspection of the infinitesimally small sample weights and the resulting n_{eff} shows that this assumption is misplaced, so these data should not count equally towards reducing the uncertainty on the WM.

The *unbiased* SE_{WM} of the data was 1.1×10^{-4}. This was ~2.1× larger than the SE_{WM} of only the two precise points, and 5× larger than the biased SE_{WM}. Despite their small weights, the imprecise points did contribute to the overall variance. The variance of the two precise points only was 5.41×10^{-9}, whereas it was 1.54×10^{-8} for all of the points. With $n_{eff} = 2.00$, both variances were reduced by the same factor, yielding a larger SE_{WM} for all the data together. Qualitatively this makes sense, since adding uncertain data to the distribution should only decrease confidence in its mean, even if only slightly.

Note, for this example, two precise measurements were chosen instead of one. If only one were chosen, then the *n/n-1* term in the weighted variance would become extremely large resulting in an unbiased SE_{WM} of 0.14 when taken with the rest of the uncertain data. This would be correct (i.e., taking the weighted mean of effectively one point should not be well constrained) but it was simply less illustrative than the two-point example where the unbiased SE_{WM} was closer in scale to the uncertainties of the two dominant points.

While this scenario would be implausible, it was realistic in the sense that these data could be collected easily on a mass spectrometer or similar instrument that integrates counts over a discrete time. Calculating the traditional, biased WM and SE_{WM} of such data may seem absurd to an incredulous scientist, however, this supports the argument that a more robust, unbiased estimator should be used universally instead of at the scientist's discretion based upon the data.

3. Real Th/U RSF Data

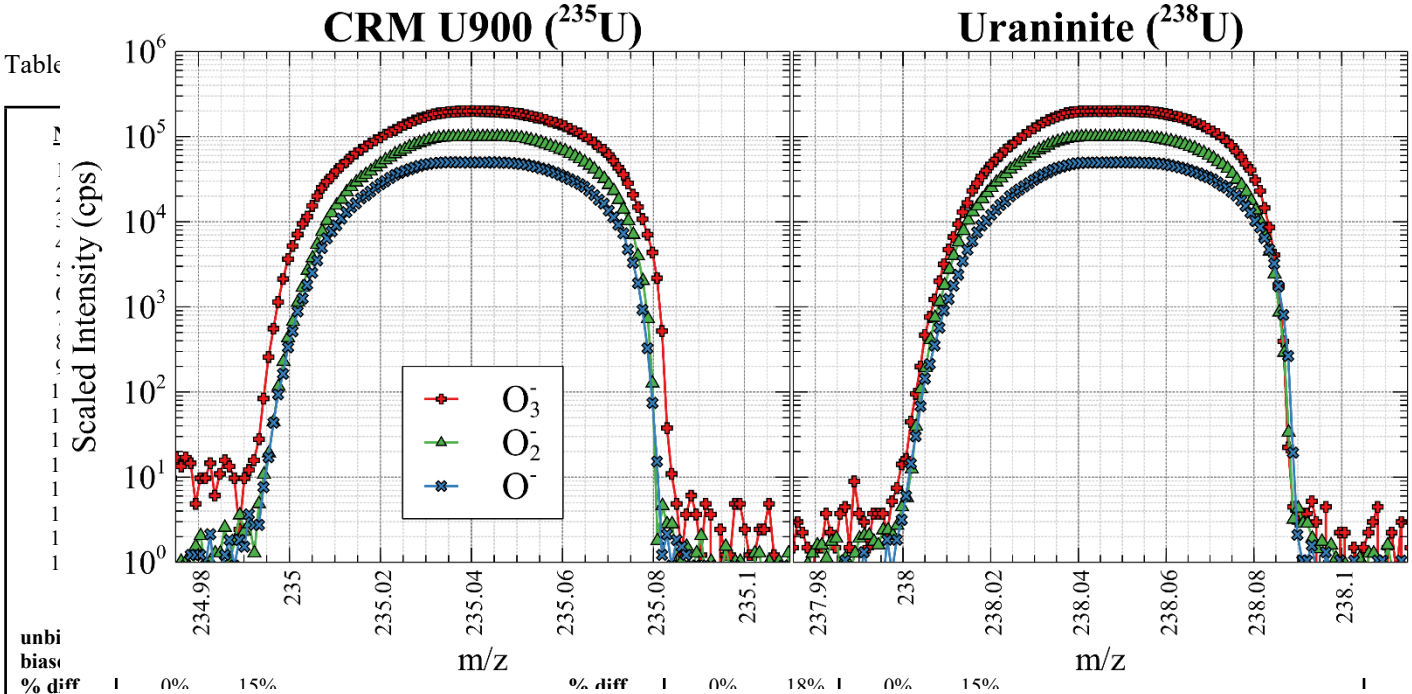
Here we compared the magnitudes of the biased an unbiased SE_{WM} using real data reported in this paper. Figure 2 of the parent paper shows the Th/U RSF values for repeated measurements on CRM U900 particles with the three different primary beam species. Data for the particles on graphite were reproduced in Table A-2. The unbiased SE_{WM} were larger than the biased ones by approximately 15-18%. In addition, when calculating the 95% CI, equal to SE_{WM} × *t*-value × Sqrt(MSWD), the *t*-value would be larger when using n_{eff} instead of *n*. For example, Table A-2 shows n_{eff} = 11.1 and *n* = 15 for O_2^-. The two *t*-values for α = 0.05 would be 2.225 and 2.145, respectively, an additional expansion of the 95% CI by 4%.

B. Appendix: Additional Figures

Mass scans of $^{235}U^+$ and $^{238}U^+$ on CRM U900 and uraninite, respectively.

1. J. Kirchner, *Journal*, 2006.

Table



unbi
biase
% diff 0% 15% % diff 0% 18% 0% 15%
n
n_{eff}

Figure B- 1: High MRP (M/ΔM = 3500, 50 eV energy bandpass) scans of major U isotopes using each primary beam. There is not difference in MRP or peak shape based upon primary beam species. Peak profiles are offset for visual clarity.

2. P. Meier, *Biometrics*, 1953, **9**, 59-73.
3. G. J. Feldman and R. D. Cousins, *Physical Review D*, 1998, **57**, 3873-3889.
4. C. Szakal, D. S. Simons, J. D. Fassett and A. J. Fahey, *Analyst*, 2019, **144**, 4219-4232.
5. R. C. Ogliore, G. R. Huss and K. Nagashima, *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 2011, **269**, 1910-1918.