MESA Metals 2: Examining Variability within Calibration Curves

Contents

Section 1: Setup & Summary

Let's revisit the regression analysis and compare old and new calibrations, using the new (5/10/21) dataset. Mostly, this is to make sure there aren't systematic differences (between old and new calibrations), and then we can move into the lower range to determine values that were below the LOD.

Summary

Overall, the procedure is as follows: 1. Run a regression for a single calibration and single data of a single chemical (arsenic chosen) 2. Run a regression for all calibration runs of a single chemical (also arsenic) 3. Run regressions of all runs on all chemicals. This step is likely of the most interest. 1. Print plots of these regressions, and also print plots of the intercepts versus their slopes.

By plotting, we can examinine the variability between regression lines per each calibration run, as well as examining the slopes versus the intercepts (since this may be helpful later). See Section 6.

Section 2: Arsenic Case Study, Regression Results Replication Attempt

This section reads in just one elements calibration data (arsenic only). Then, we fit a regression to this (the raw counts divided by internal standard, minus the calibration blank).

Printing the regression output,

```
summary(glm_fit)
```

```
##
## Call:
## glm(formula = net_signal_old ~ concentration_ug_old, family = gaussian(link = "identity"),
##     data = As_df_test)
##
## Deviance Residuals:
##          1           2           3           4           5           6
## -0.0002612  -0.0004349   0.0007673   0.0003127  -0.0001543  -0.0002296
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.0002763  0.0002791    0.99    0.378
## concentration_ug_old 0.0137662  0.0001188  115.91 3.32e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 2.550935e-07)
##
##     Null deviance: 3.4280e-03  on 5  degrees of freedom
## Residual deviance: 1.0204e-06  on 4  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: -70.495
##
## Number of Fisher Scoring iterations: 2
```

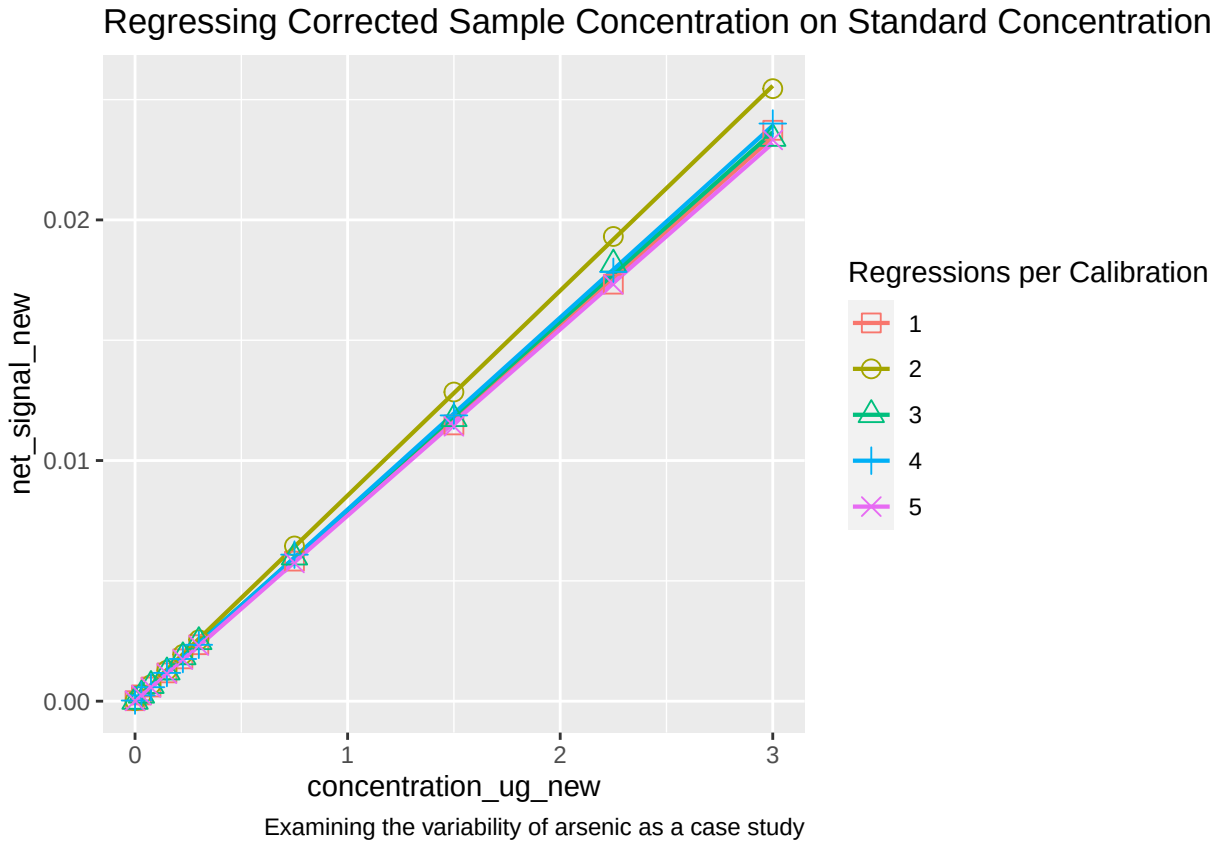
It would be helpful to compare this result to what Kathrin and Rony produce using the spectroscope software, since this is not in the Excel sheet. It should match exactly.

Section 3: Iterate Regression for All Arsenic

Let's iterate the regression for each calibration run, for one element (As).

Section 4: Plot Arsenic Regression Lines and Intercept vs Slope

Then we look at these regression lines for arsenic. Each unique calibration run is its own regression line, as in the Excel structure. For brevity, we just do the newer arsenic data (and then we'll look at all old and new data for all elements later).



Focusing on these regression lines based on the newer arsenic data, let's take a look at the intercepts against the slopes. Note that the vertical line at 0 is for reference.

Section 5: All Elements Data Manipulation

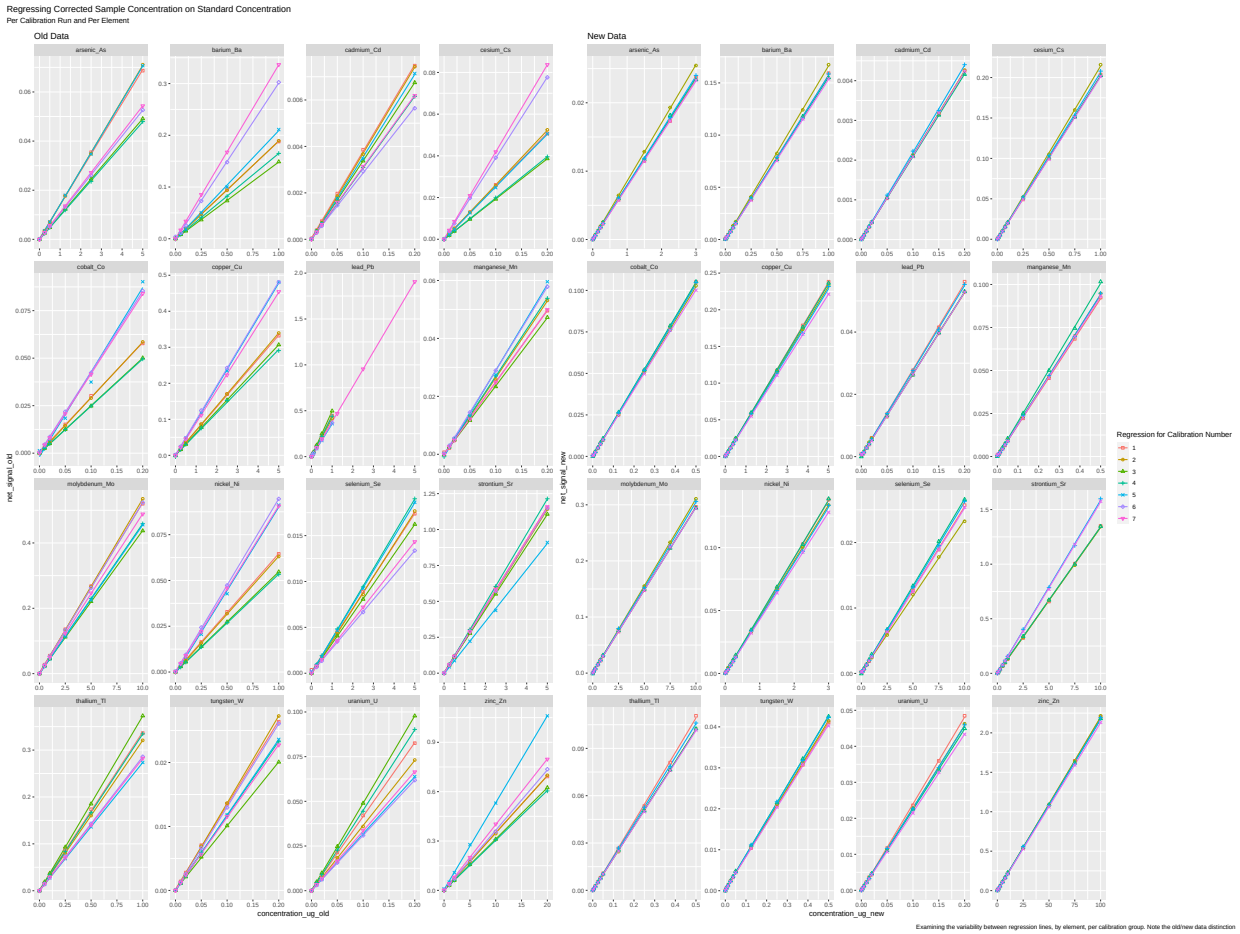
Great. Now we's work on iterating for all elements (organized per Excel sheets within the workbook). I read in the data, then, iterate the regressions for each of the (calibration) groups within each element.

Section 6: Plot All Elements

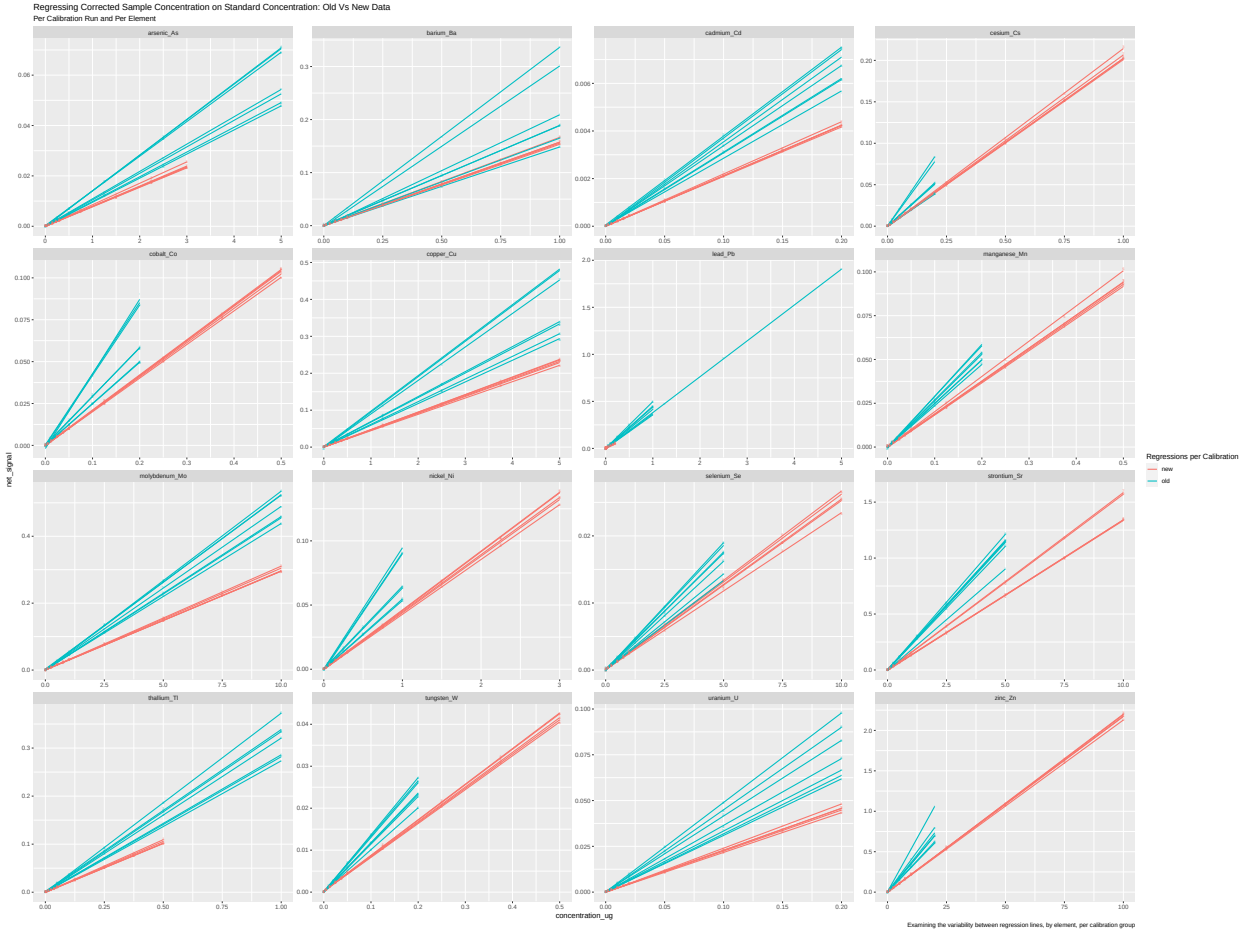
Then, we're in a position to visualize the regression lines, and we also add the raw data to the plot. Note that the calibration ranges have changed for some elements, and this impacts the way the results look. However, even with that note, there does appear to be more consistency in the new calibration method.

Section 6a: Regression Lines

First, let's look at each calibration run separately. The old data have 6 data points per calibration run, and 7 calibration runs per element. The new data have 10 daata points per calibration run, and 5 calibration runs per element. There are 16 elements. Resolution is set very high to fit it on one page, please zoom in.

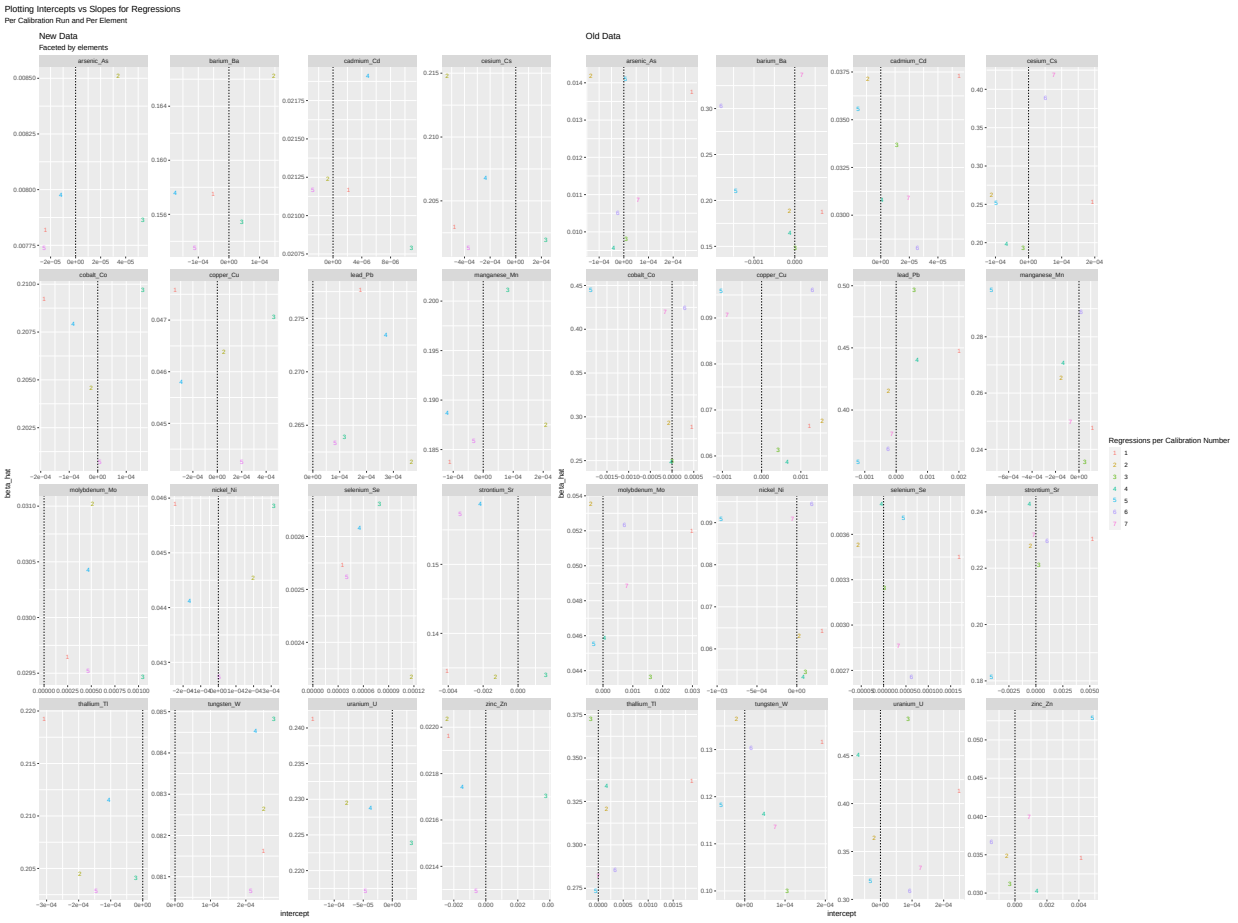


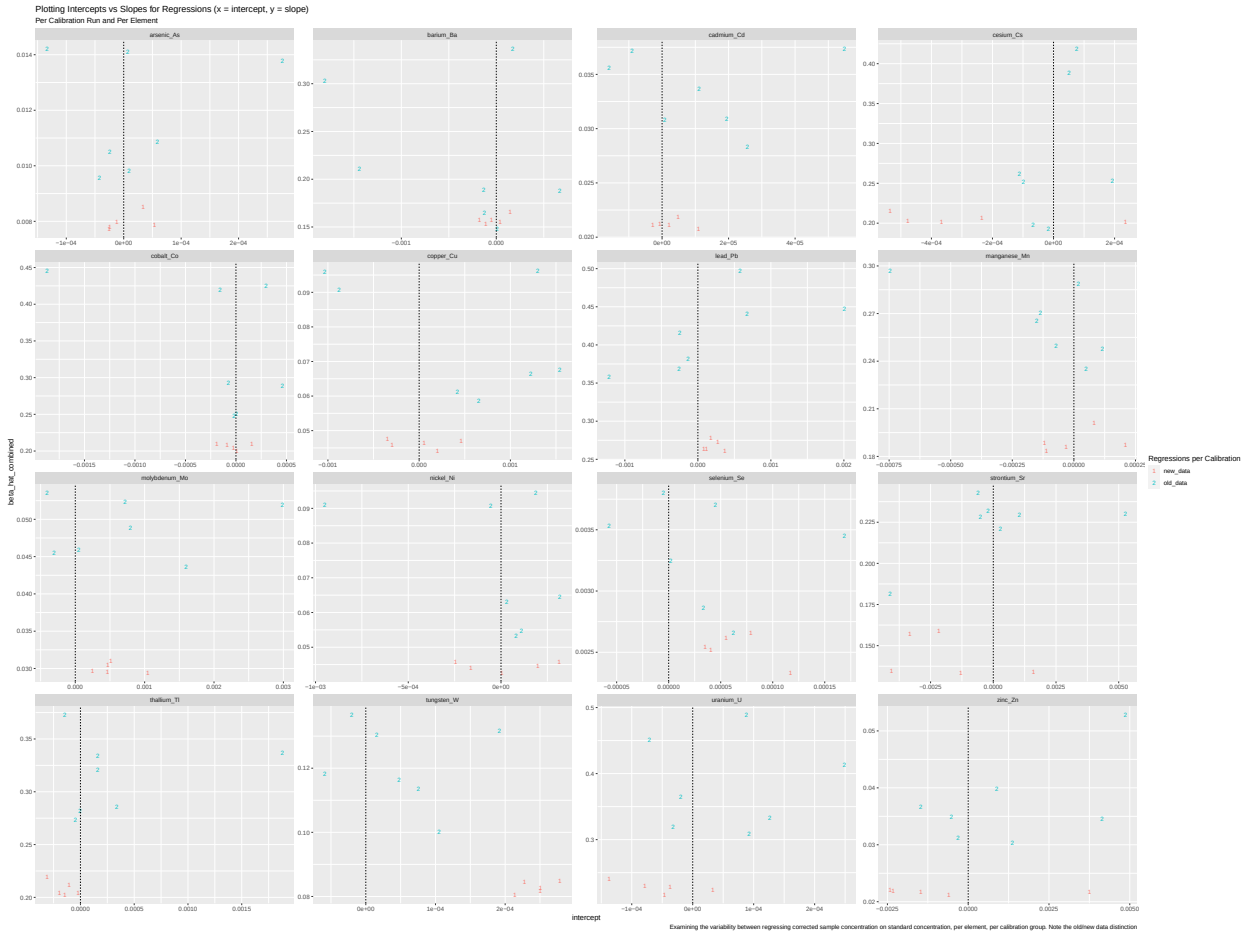
Then, let's look at the new versus the old data, maintaining each regression line, but color-coding simply so can compare easily. While the standardized concentration ranges (x-axis) vary from the old to the new data, we overlay the graphs anyways, to get a sense of what consistency is there. Note the special case of lead (Pb).



Section 6b: Intercept and Slope Coefficients

Now, similar to the first regression lines plot, we look at the intercepts versus slopes (x-axis and y-axis respectively) for each of the regressions. This first separates old and new data, per calibration run and per element, and then overlays old and new data (with a simple color-code).


Plotting Intercepts vs Slopes for Regressions
Per Calibration Run and Per Element
New Data
Faceted by elements
Old Data
arsenic_As
barium_Ba
cadmium_Cd
cesium_Cs
cobalt_Co
copper_Cu
lead_Pb
manganese_Mn
molybdenum_Mo
nickel_Ni
selenium_Se
strontium_Sr
thallium_Tl
tungsten_W
uranium_U
zinc_Zn
beta_hat
intercept
Regressions per Calibration Number
1 1
2 2
3 3
4 4
5 5
6 6
7 7
Examining the variability between regressing corrected sample concentration on standard concentration, per element, per calibration group. Note the old/new data distinction.


Plotting Intercepts vs Slopes for Regressions (x = intercept, y = slope)
Per Calibration Run and Per Element
arsenic_As
barium_Ba
cadmium_Cd
cesium_Cs
cobalt_Co
copper_Cu
lead_Pb
manganese_Mn
molybdenum_Mo
nickel_Ni
selenium_Se
strontium_Sr
thallium_Tl
tungsten_W
uranium_U
zinc_Zn
beta_hat_combined
intercept
Regressions per Calibration
1 new_data
2 old_data
Examining the variability between regressing corrected sample concentration on standard concentration, per element, per calibration group. Note the old/new data distinction