MESA Metals: Examining Variability within Calibration Curves

Contents

Section 1: Setup & Summary

Note that most of the code is hidden in the PDF, and we primarily look at visualizations here.

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

We were primarily interested in replicating the regression, and then examining 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.

Questions

What regression is implemented in the Excel sheet specifically, and also, what regression should we implement in our analyses? The spectroscopy software computes some transformations and additional changes which do not appear in the Excel. The regressions used here are similar to the Excel sheet (though need adjustment, see Section 2).

Section 2: Arsenic Case Study, Regression Results Replication Attempt

This section reads in just one date's one calibration iteration number, for arsenic only. Then, we can fit a regression to this (the blank corrected concentrations regressed on the standard concentrations).

Printing this regression output,

```
summary(glm_fit)
```

```
##
## Call:
## glm(formula = as_o_blk_corrected ~ std_conc_ppb, family = gaussian(link = "identity"),
##     data = arsenic_As_test)
##
## Deviance Residuals:
##       1        2        3        4        5        6
##  124.99    40.84    32.65  -139.28  -177.43   118.23
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -124.99      79.70  -1.568    0.192
## std_conc_ppb  6871.96      33.92 202.606 3.56e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 20803.51)
##
##     Null deviance: 854048044  on 5  degrees of freedom
## Residual deviance:     83214  on 4  degrees of freedom
## AIC: 80.252
##
## Number of Fisher Scoring iterations: 2
```

We can see this is not a precise replication of the first arsenic regression. This is close to the Excel sheet results but not sufficient; arsenic's 1/8/20 Calibration 1 has intercept and beta estimates of about -179.92 and 6887.72 but these show have about -124.99 and 6871.96.

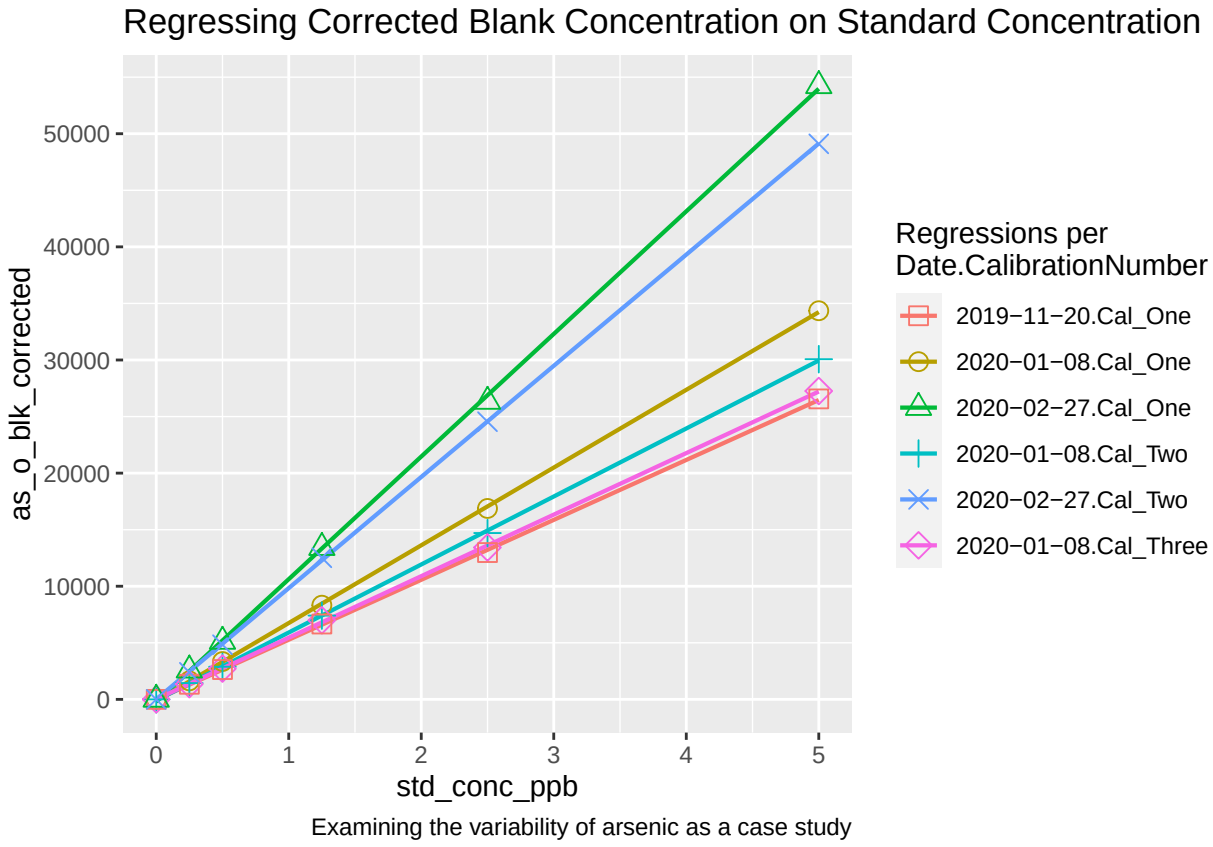
I am not able to replicate the Excel result by trying different tuning approaches (is there a weight being implemented, maybe from the last sheet?). Additionally, this is not the same regression as the spectroscopy program uses at all, as per Rony's 4/1/21 email. This must be revisited and scaled across the document, but works to start looking at the patterns.

Section 3: Iterate Regression for All Arsenic

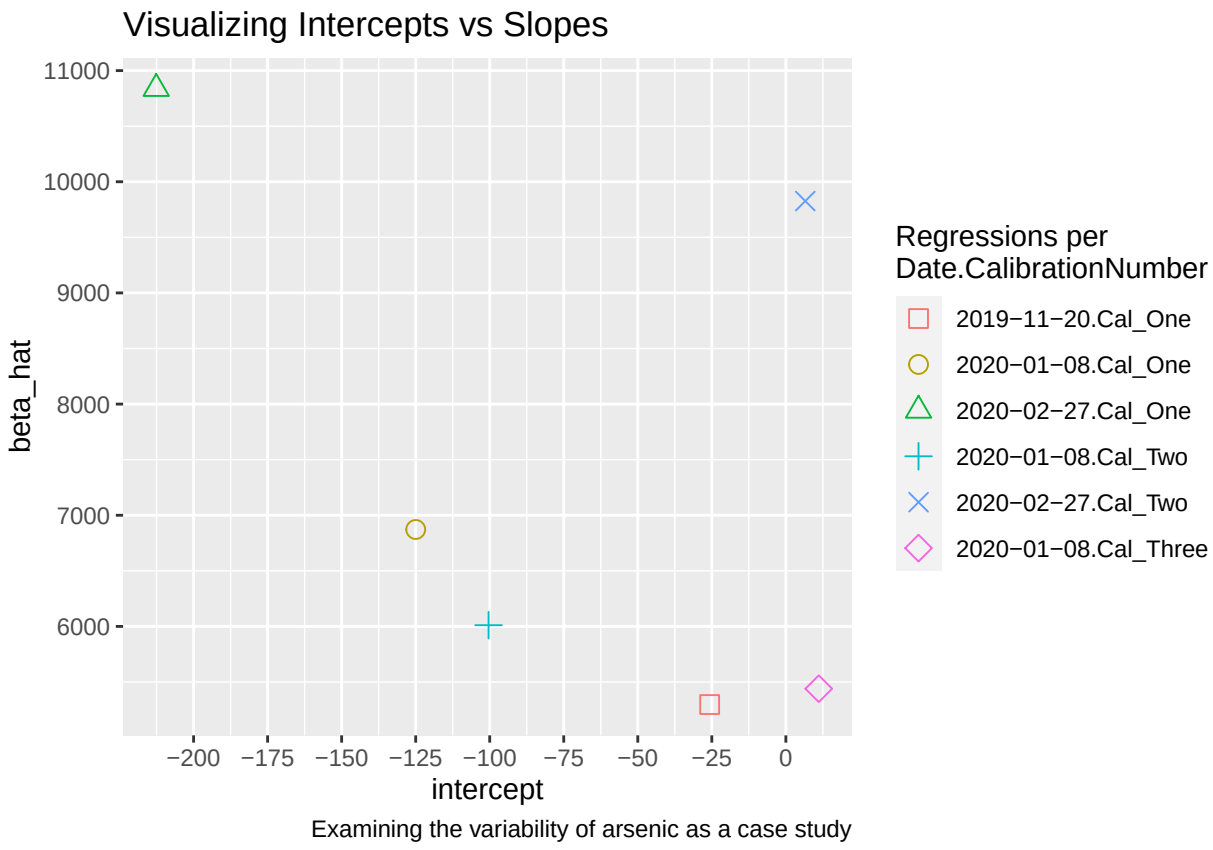
Let's iterate for each date and calibration number, for one element.

Section 4: Plot Arsenic Regression Lines and Intercept vs Slope

Then we look at these regression lines. Each unique combination of date and calibration number are their own regression line, as in the Excel.



Focusing on these regression lines, let's take a look at the intercepts against the slopes.



Section 5: All Elements Data Manipulation

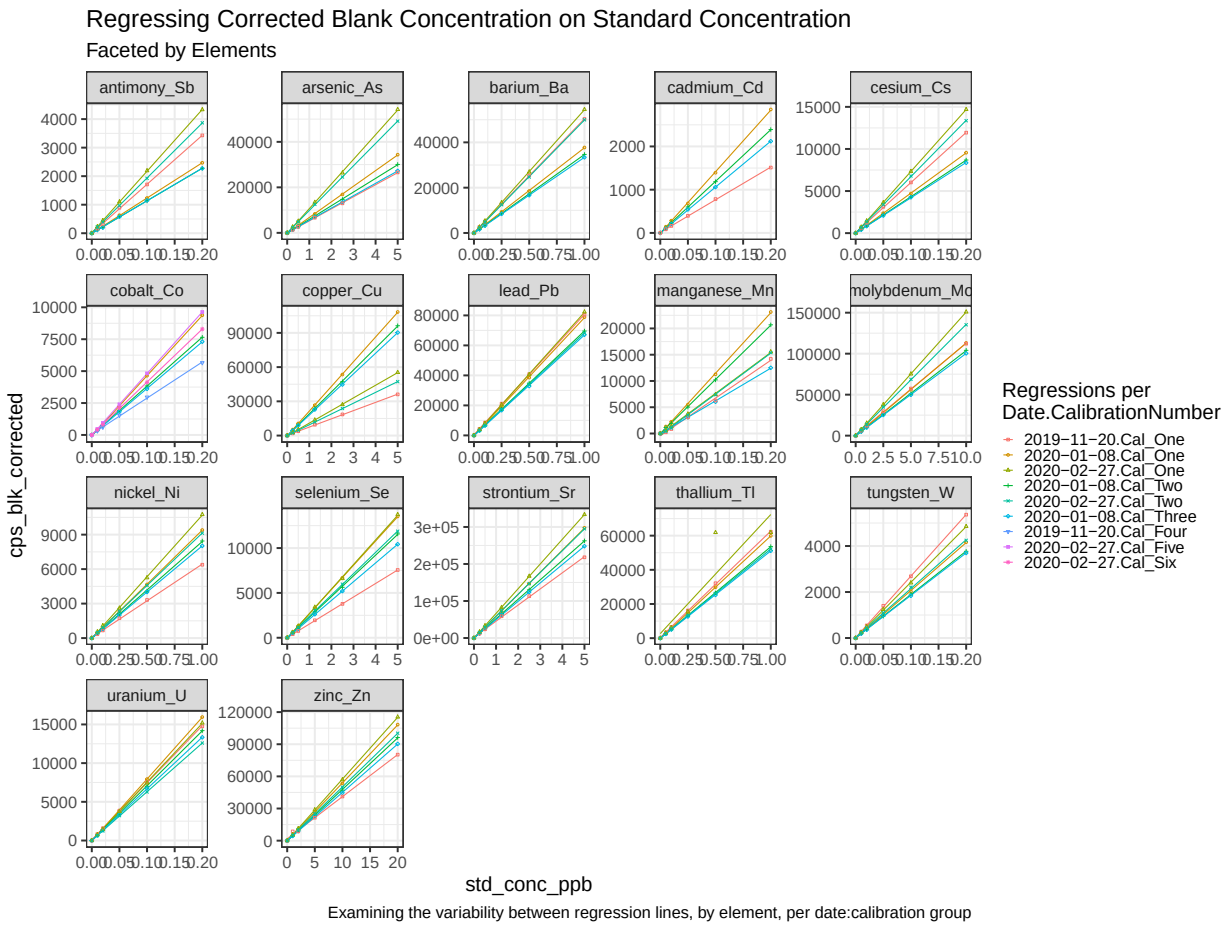
Great. Now let's work on iterating for all elements (organized per Excel sheets within the workbook).

I read in the data.

Then, iterate the regressions for each of the (date,no_cal) groups within each element.

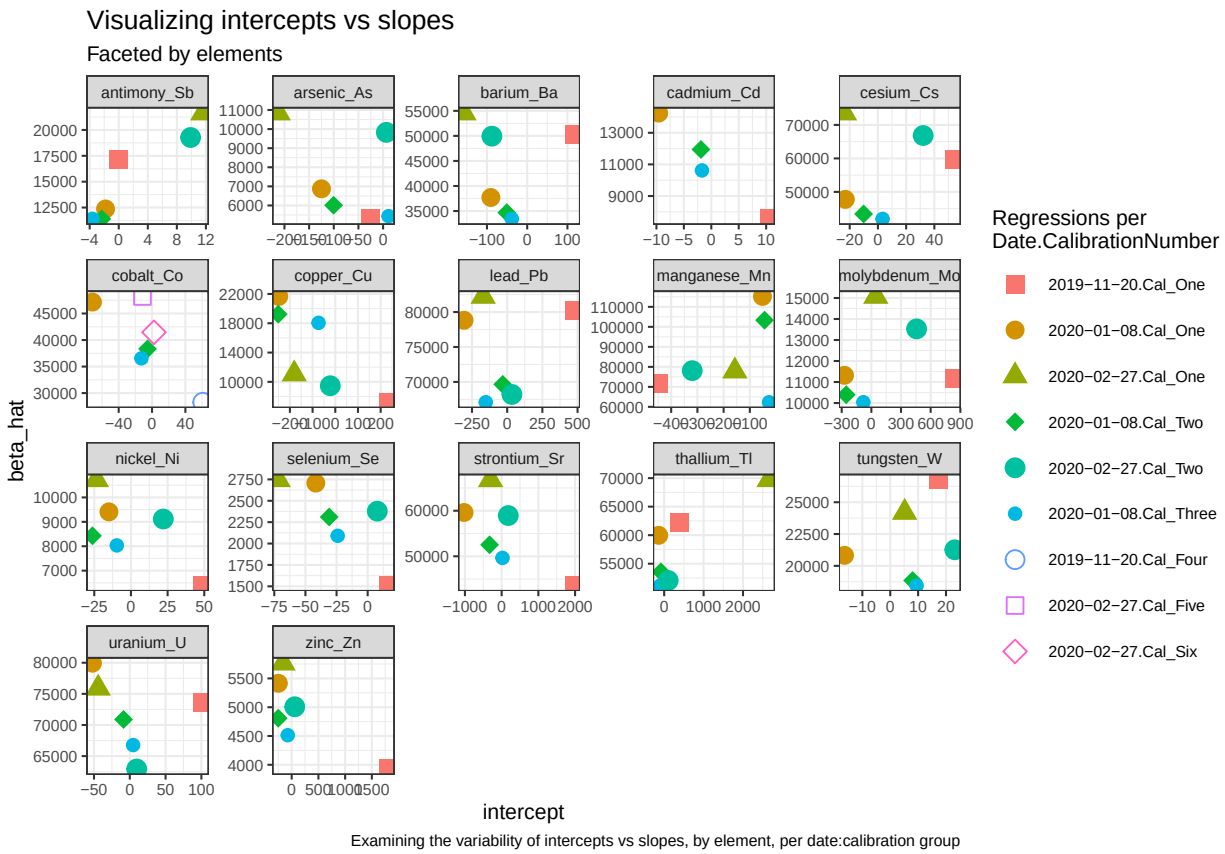
Section 6: Plot All Elements Regression Lines and Intercept vs Slope

Then, we're in a position to visualize the regression lines, and add the raw data to the plot just to see:



Interesting. There is some substantial variability between estimates within each element. Will think more on this. Note that thallium_Tl has a high value that is visible in the raw data (Excel too).

Like we did for arsenic, let's look at the intercepts vs slopes for these regression lines, per element also. Note cobalt's additional calibration numbers.



Interesting also. We're primarily interested in seeing these intercepts and slopes to see if one is correlated with the other, since this could be useful when we start extrapolating.