

MESA Metals 2: Examining Variability within Calibration Curves

Contents

Section 1: Setup & Summary	2
Summary	2
Section 2: Arsenic Case Study, Regression Results Replication Attempt	3
Section 3: Iterate Regression for All Arsenic	3
Section 4: Plot Arsenic Regression Lines and Intercept vs Slope	3
Section 5: All Elements Data Manipulation	5
Section 6: Plot All Elements Regression Lines and Intercept vs Slope	5

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 examine 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
```

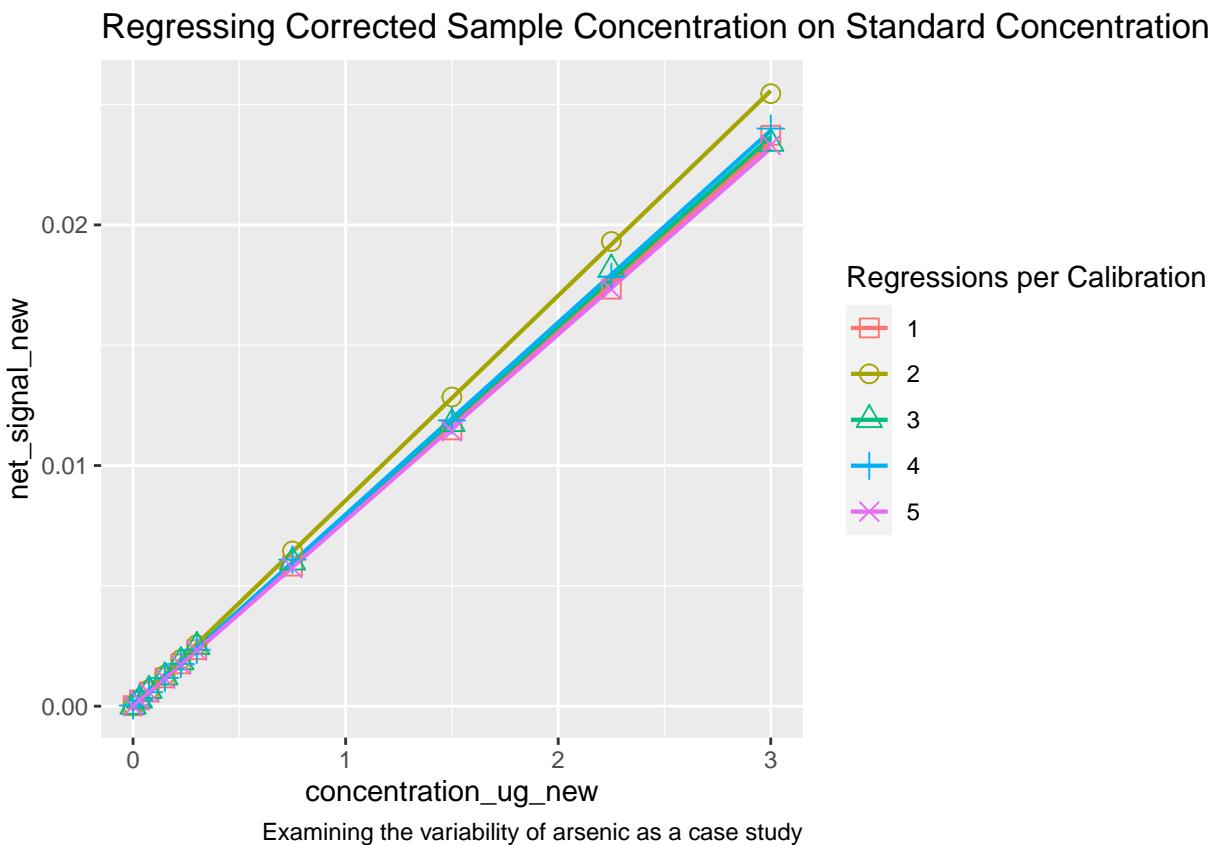
It would be helpful to compare this result to what Kathrin and Rony produce using the spectroscope software, since this is not in the Excel sheet. It should match exactly.

Section 3: Iterate Regression for All Arsenic

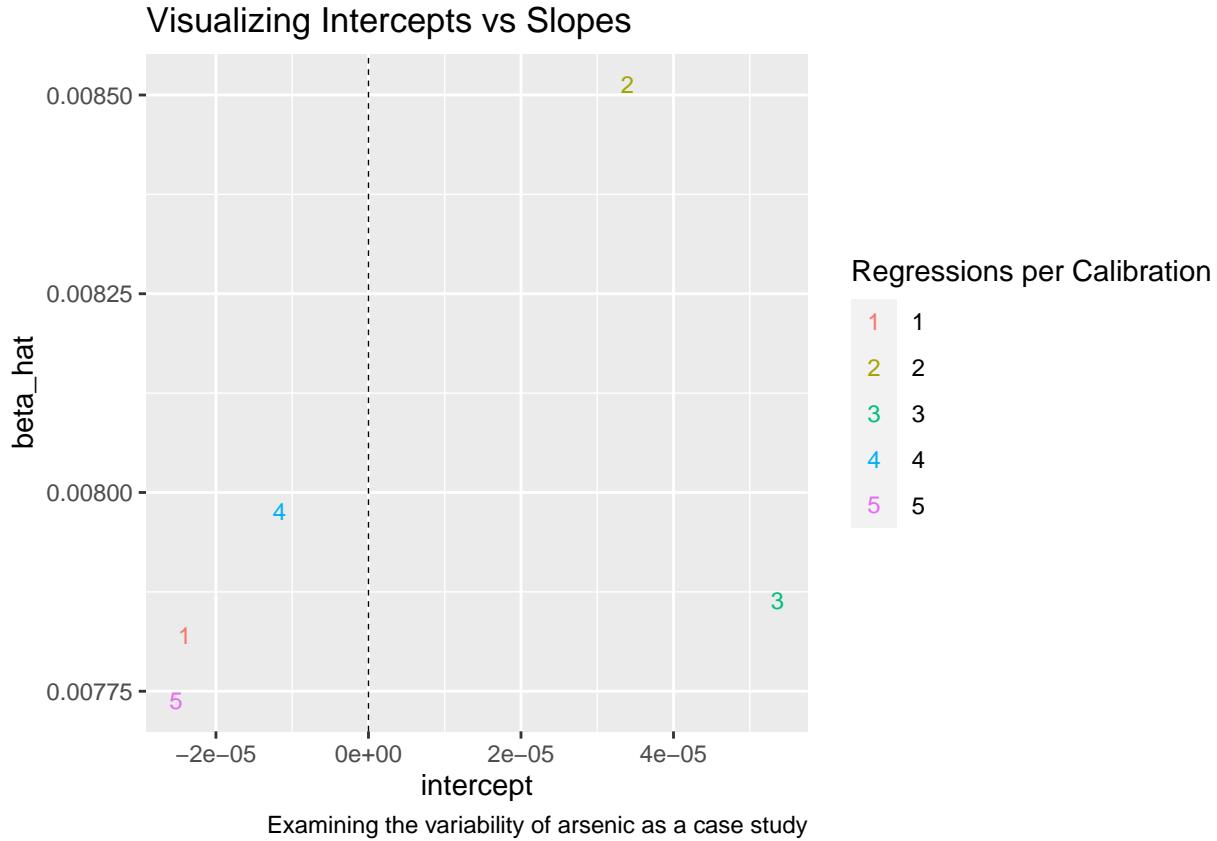
Let's iterate the regression for each calibration run, for one element (As).

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're 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.

```
mesa_metals_df = mesa_metals_df %>% rename(no_cal_old = x3) %>
  mutate(no_cal_new = NA) %>% group_by(elements) %>% mutate(no_cal_old = rep(1:7,
  each = 7, length.out = length(elements)), no_cal_new = rep(1:5,
  each = 11, length.out = length(elements))) %>% ungroup(elements) %>%
  select(elements, no_cal_old, concentration_ug_old, net_signal_old,
  no_cal_new, concentration_ug_new, net_signal_new, everything())
```

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

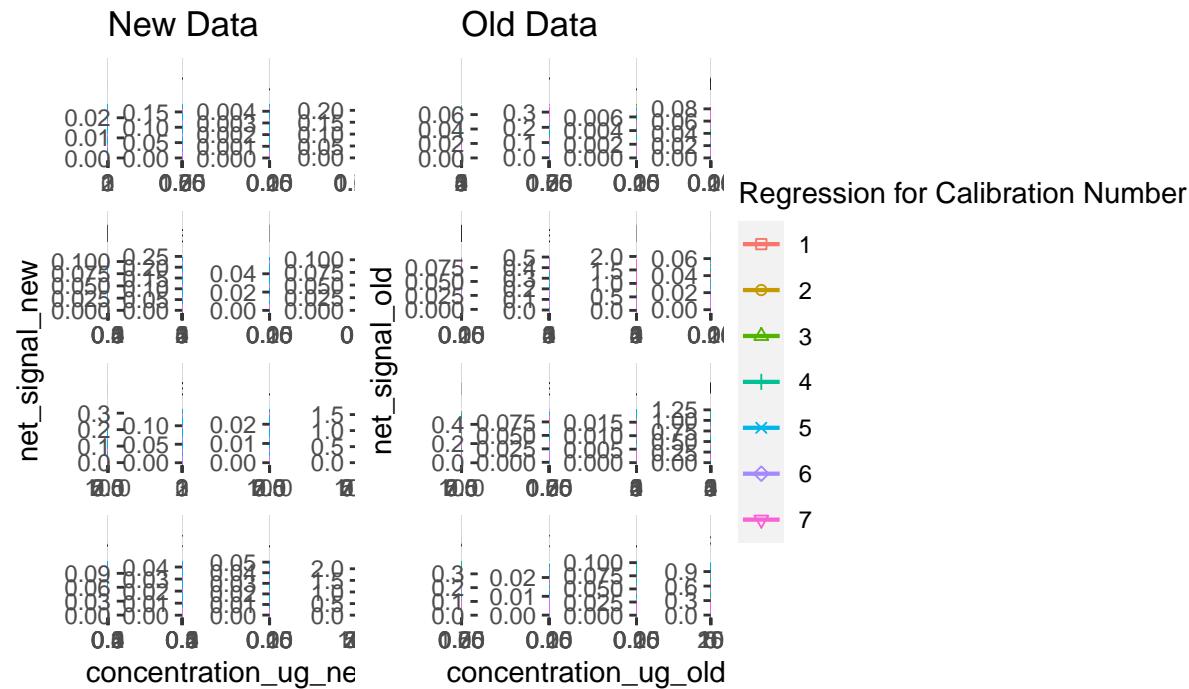
Then, we're in a position to visualize the regression lines, and we also add the raw data to the plot. Note that the calibration ranges have changed for some elements, and this impacts the way the results look. However, even with that note, there does appear to be more consistency in the new calibration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

```
patchwork_old_calibrations = mesa_metals_df %>% mutate(no_cal_old = as.factor(no_cal_old)) %>%
  ggplot(aes(x = concentration_ug_old, y = net_signal_old,
             group = no_cal_old, color = no_cal_old, shape = no_cal_old,
             )) + stat_smooth(method = "lm", formula = y ~ x, se = F,
             size = 0.75) + geom_point() + labs(color = "Regression for Calibration Number",
             shape = "Regression for Calibration Number", title = "Old Data") +
  facet_wrap(~elements, scales = "free") + scale_shape_manual(values = seq(0,
  9))

patchwork_new_calibrations = mesa_metals_df %>% mutate(no_cal_new = as.factor(no_cal_new)) %>%
  ggplot(aes(x = concentration_ug_new, y = net_signal_new,
             group = no_cal_new, color = no_cal_new, shape = no_cal_new,
             )) + stat_smooth(method = "lm", formula = y ~ x, se = F,
             size = 0.75) + geom_point() + labs(color = "Regression for Calibration Number",
             shape = "Regression for Calibration Number", title = "New Data") +
  facet_wrap(~elements, scales = "free") + scale_shape_manual(values = seq(0,
  9)) + theme(legend.position = "none")

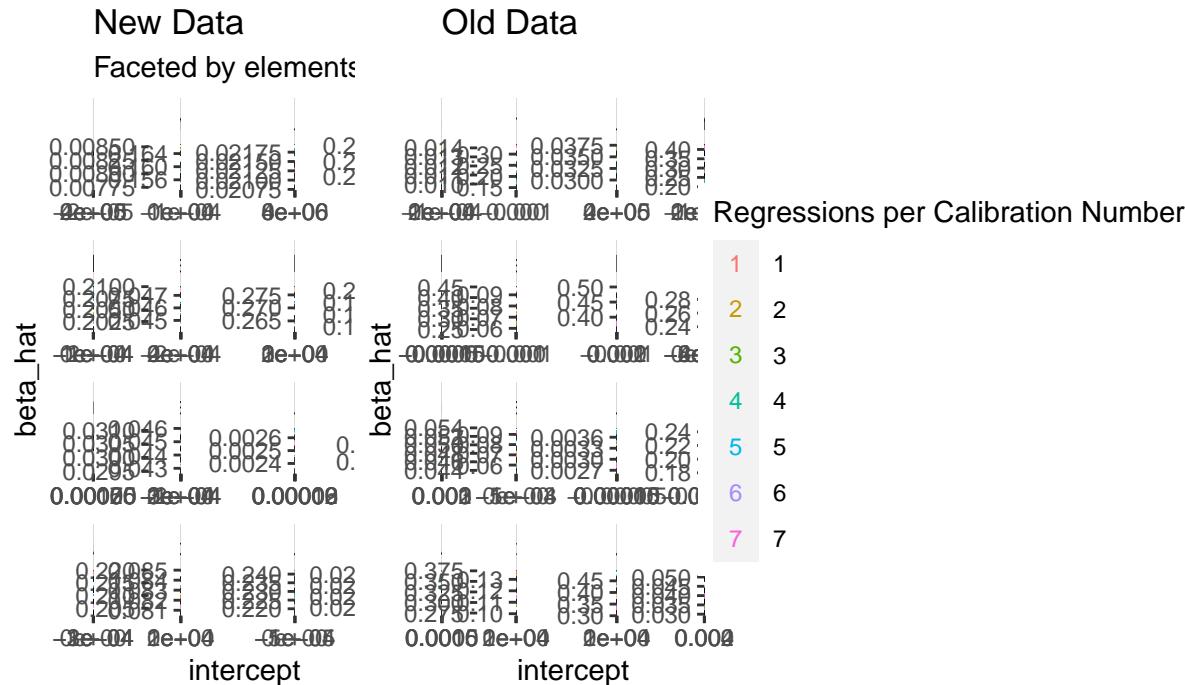
(patchwork_new_calibrations + patchwork_old_calibrations) + plot_layout(guides = "collect") +
  plot_annotation(title = "Regressing Corrected Sample Concentration on Standard Concentration",
                 subtitle = "Per Calibration Run and Per Element", caption = "Examining the variability between regression lines, by element, per calibration run")
```

Regressing Corrected Sample Concentration on Standard Concentration
Per Calibration Run and Per Element



Examining the variability between regression lines, by element, per calibration group. Note the old/new data distinction

Plotting Intercepts vs Slopes for Regressions
Per Calibration Run and Per Element



multiple concentration on standard concentration, per element, per calibration group. Note the old/new data distinction.

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.