EPA ULSD Qualification and Round Robin Test Program Results

Chris Laroo

US EPA Office of Transportation and Air Quality
Assessment and Standards Division

ASTM D02
Norfolk, VA
December 6, 2005

Overview

- Motivation for test program.
- Overview of instrument qualification process and results.
- Overview of round robin test program participation.
- Review of test program sample analysis.
- Review of test program data analysis.
- Review of test program results.
- Conclusions.

Motivation for ULSD Round Robin Program

- ULSD FRM allows for a 2 ppm downstream test tolerance on sulfur measurements.
- We heard concerns that actual reproducibility (R) may be > 2 ppm.
  - 4.4 ppm historically as published in ASTM D 5453-03a.
  - 3-4 ppm in 2004 and 2005 ASTM crosscheck program for D 5453.
- If real world reproducibility is higher, then industry feared it would force down pipeline standards and refinery production targets, impacting cost and supply.

Motivation for ULSD Round Robin Program cont.

- EPA was concerned that current data is not reflective of what is possible/likely in 2006.
  - If we set the tolerance based on historical reproducibility, and significant improvement occurred, it would have the effect of relaxing the 15 ppm standard in-use.
  - None of the labs in the ASTM ILCP were qualified for measuring sulfur in the 15 ppm range for precision and accuracy.
- Committed to conduct our own round-robin test program limited to just EPA qualified laboratories and adjust the test tolerance accordingly as necessary.
  - We developed the test program and analytical protocol with industry stakeholders and received their buy-in May 2005.
- This test program has been completed and the results will be presented here.

ULSD Round Robin Program Qualification

- All laboratories participating in this round robin test program were required to qualify their sulfur measurement methods with EPA.
- This meant that the labs must meet the precision and accuracy requirements per 40 CFR 80.580 - 80.585.

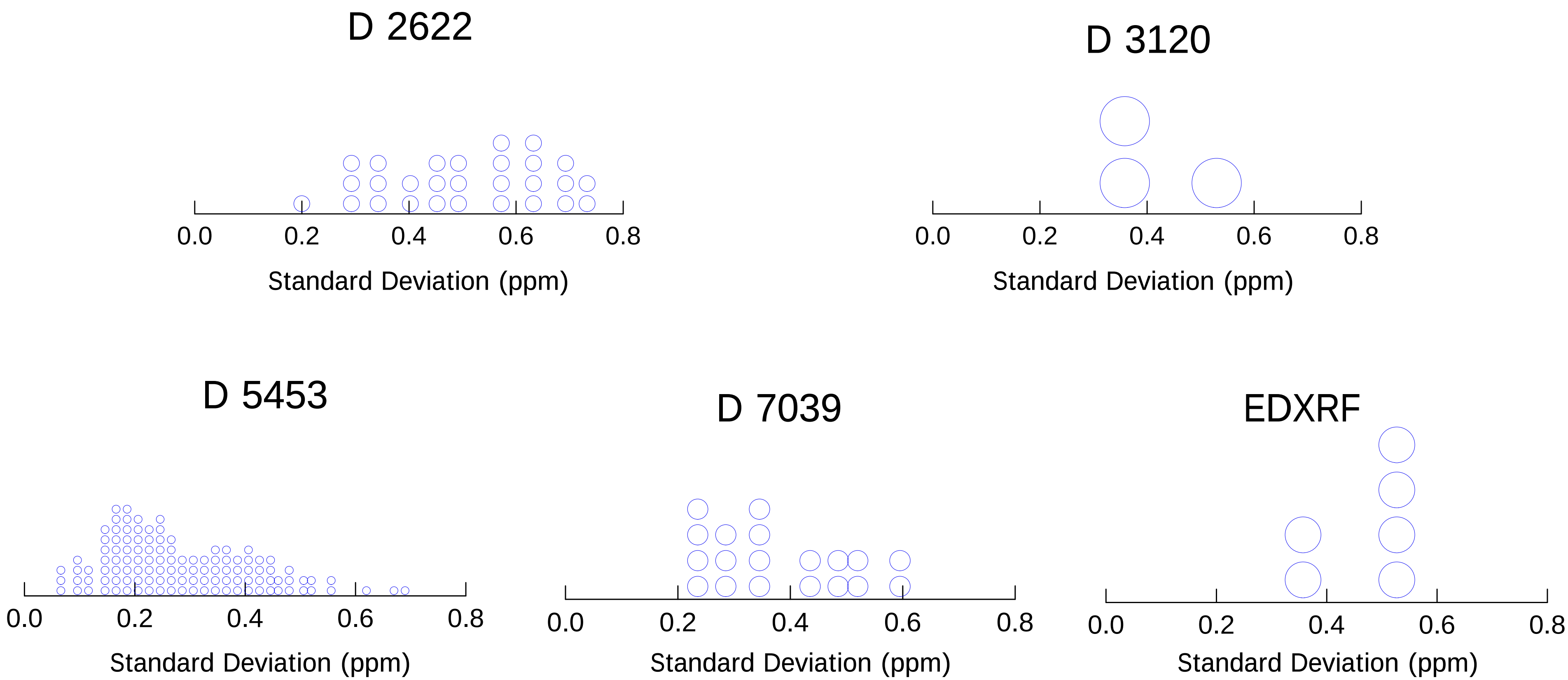
Qualification Results

- The qualification results by method are as follows (as of 10/13/05)*:

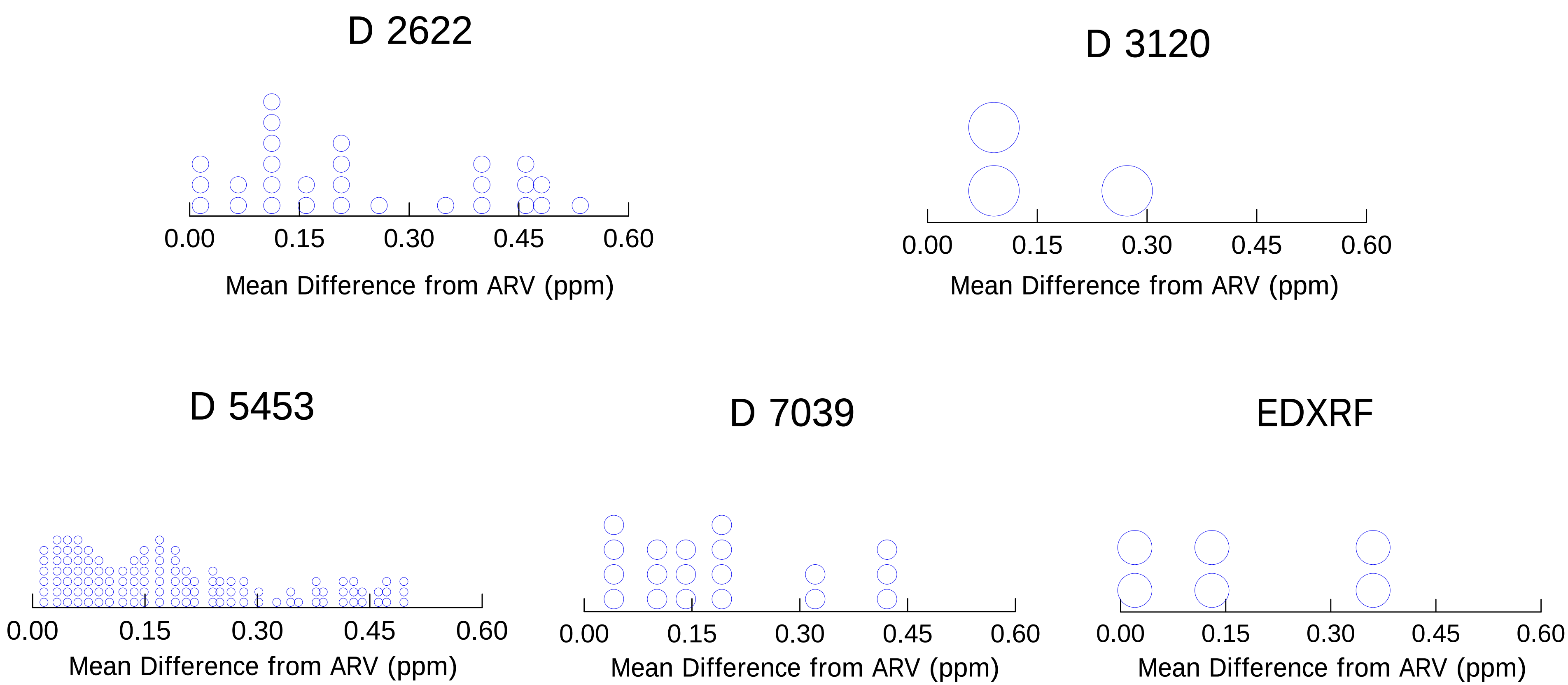
Test Method	D 5453	D 7039	D 2622	D 3120	EDXRF	Average Across Methods	**CFR Req.**
Number of Inst. (Total = 173)	116	19	28	3	6		
Average Precision	0.29	0.38	0.50	0.39	0.47	0.34	**0.72**
Average Accuracy (1 - 10 ppm)	0.20	0.18	0.24	0.14	0.16	0.20	**0.54**
Average Accuracy (10 - 20 ppm)	0.20	0.20	0.24	0.20	0.24	0.21	**0.54**

*Not all of the qualified labs participated in the RR test program.

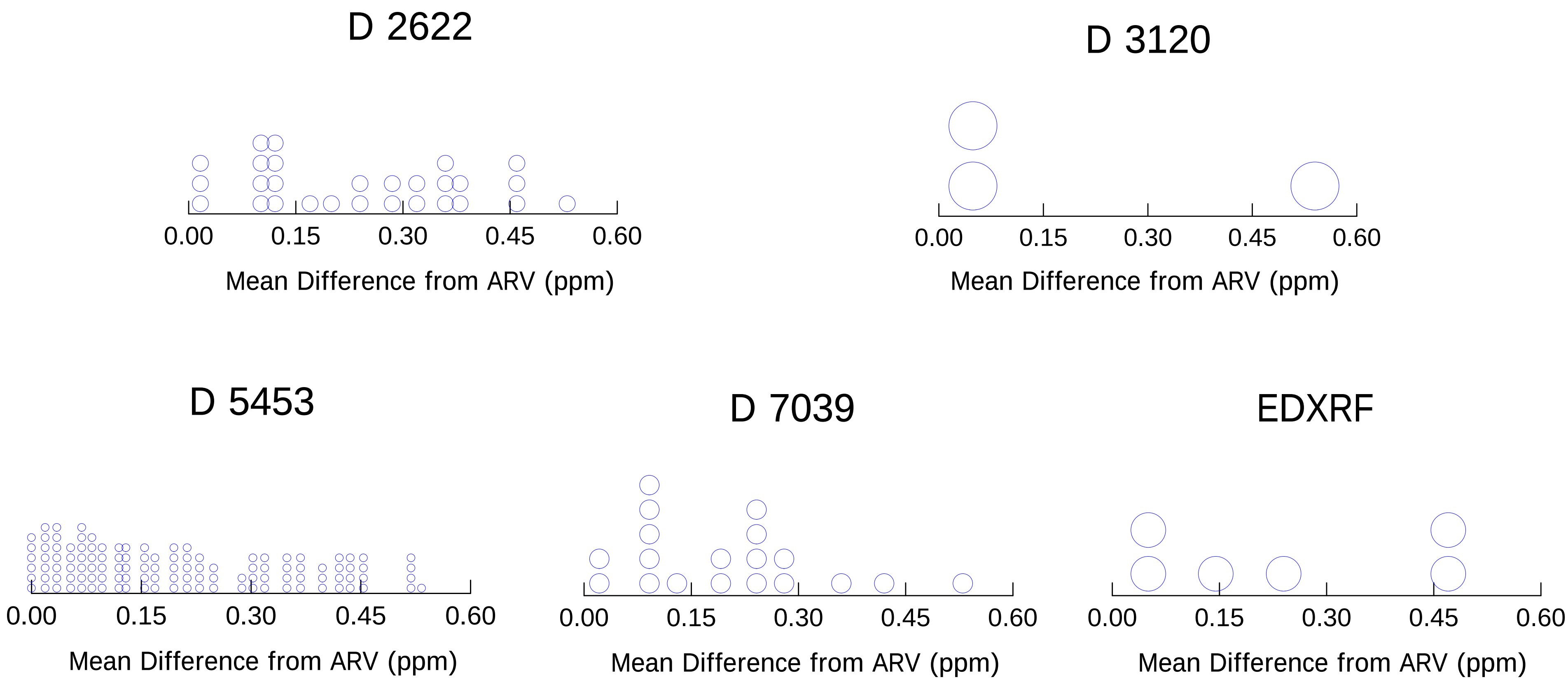
Dot Plot of Qualification Method Specific Precision Results



Dot Plot of Qualification Method Specific 1 to 10 ppm Accuracy Results

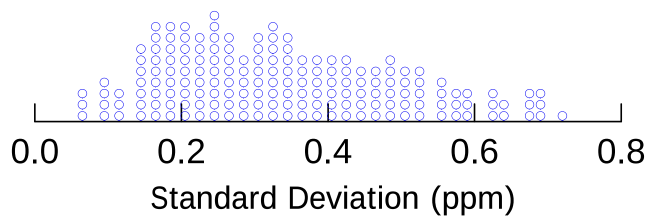


Dot Plot of Qualification Method Specific 10 to 20 ppm Accuracy Results

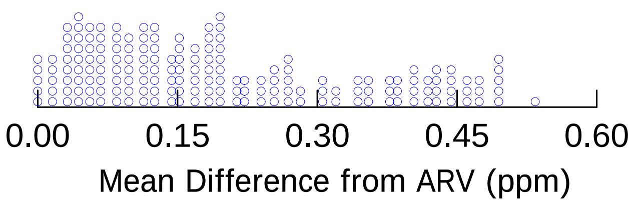


Dot Plot of Qualification Composite Precision and Accuracy Results

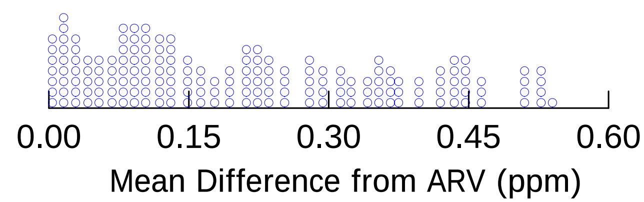
Precision



1-10 ppm Accuracy



10-20 ppm Accuracy



ULSD Round Robin Program Qualification Conclusions

- Qualification criteria easily met by the newest methods, D 5453 and D 7039.
- Precision means of 0.29 and 0.38 ppm for D 5453 and D 7039 respectively were well below the CFR limit of 0.72 ppm.
- Accuracy means for D 5453 and D 7039 respectively were well below the CFR limit of 0.54 ppm.
  - 0.20 and 0.18 (1 to 10 ppm gravimetric std.)
  - 0.20 and 0.20 (10 to 20 ppm gravimetric std)

ULSD Round Robin Test Program

ULSD Round Robin Program Participation

- Initially 161 labs utilizing 208 instruments registered to participate in the program.
- Some labs failed to qualify (2).
- Others determined during the qualification process that they would not pass and abandoned testing.
- Most of these labs started looking into procuring new instrumentation.
- Overall, 59, or 28% of the instruments that registered for the program dropped out.
  – This left 129 labs participating with 149 instruments.

ULSD Round Robin Program Participation cont.

Test Method	July 2005	August 2005	Dropped Out
D 5453	98	93	27
D 2622	25	24	23
D 7039	16	16	3
EDXRF	6	6	1
D 3120	3	3	3
D 7041	1	1	0
D 4294	0	0	2
Total Instruments	149	143	59
Total Labs	129	125	32

ULSD Round Robin Program Fuel Samples

- Five fuel samples were sent out in the months of July and August 2005.
  - Fuel sample sulfur values were unknown to the test labs.
- EPA targeted blending samples in the 7 to 15 ppm range.
- The samples were not sent out for independent analysis.
- The actual concentrations turned out to be in the 7 to 21 ppm range.
- One blend was sent out both months as sample #5.
- A blind gravimetric was sent out each month as fuel #4 - NIST SRM 1616b, 8.41 ppm sulfur in kerosene.

ULSD Round Robin Program Fuel Samples cont.

- The target fuel sample concentration and actual concentration based on composite robust mean are as follows:

	July Blend Target	July Composite Robust Mean*	August	August Composite Robust Mean*
Fuel #1	7	7.31	9	10.05
Fuel #2	11	10.71	13	14.42
Fuel #3	16	20.86	17	17.80
Fuel #4**	8.41	8.32	8.41	8.32
Fuel #5***	15	14.69	15	14.76

*This mean is the average of the two composite robust means taken from the in-house and NIST data.

** This fuel was the gravimetric both months and was actually NIST SRM 1616b.

*** This fuel blend was sent out both months as fuel #5.

ULSD Round Robin Program Sample Analysis

- NIST SRMs were sent out each month with the blind fuel samples.
- Laboratories were required to measure the blind fuel samples in triplicate using two different calibration curves.
  – Based on their own individual in-house calibration standards (presumably used for qualification).
  – Based on four EPA provided NIST SRMs.

ULSD Round Robin Program Sample Analysis cont.

- SRMs used in 4-point calibration curve generation are as follows:
  - RM 8771 0.07 ± 0.014 ppm S in diesel fuel
  - SRM 1616b 8.41 ± 0.12 ppm S in kerosene
  - SRM 2723a 11.0 ± 1.1 ppm S in diesel fuel
  - SRM 2770 41.57 ± 0.39 ppm S in diesel fuel

ULSD Round Robin Program Data Analysis

- Data analysis was performed under contract by SwRI.
- Outliers were determined two ways.
  - Based on the results of the measurement of the blind gravimetric fuel sample (SRM check standard).
    - Analogous to the use of a calibration check standard in normal day-to-day test operations.
    - Possible when known gravimetric standards exist.
  - Using the two-stage robust procedure identical to that used in the ASTM inter-laboratory crosscheck program (ILCP).
    - It does not require known fuel sulfur values for any of the sample fuels.

Gravimetric Outlier Deletion Method

- Used the 8.41 ppm SRM as the calibration check standard.
  - This SRM was one of the same SRMs used to calibrate the instrument.
  - The SRM was dyed yellow to “blend in” with other samples.
  - Sulfur contribution of the dye to the SRM was 0.000516 ppm.

- Compute the average (AVG) of the three repeat tests taken on the 8.41 ppm SRM for a given month by a given lab.
  - Fuel #4 in both July and August.

- Obtain the accepted reference value (ARV) of the standard fuel.
  - ARV=8.41 ppm in this study.

- Classify the data collected on all five sample fuels for a given month by a given lab as outliers and delete the entire set of lab data if

$$|AVG - 8.41| > 0.90.$$

Gravimetric Outlier Deletion Method cont.

- We allowed a ±0.90 ppm deviation since it was an average of three measurements.
  - Instead of 0.54 ppm qualification accuracy criteria over 10 measurements.
  - This compares to the actual means of 0.20 and 0.21 from the actual qualification results.
- The value takes into consideration the 95% two-sided confidence interval for three repeat measurements, as well as real bias and gravimetric standard uncertainty (GSU).

$$= 0.54 - 95\%\ CL_{10-1} + 95\%\ CL_{3-1} + GSU$$
$$= (0.54 - 0.298 + 0.543 + 0.12) = 0.905$$

95% CL calculations assume infinite degrees of freedom and use 0.48 as the std. dev. (0.48 is std. dev. of D 3120 @ 15 ppm). GSU = 0.12

Robust Outlier Deletion Method

- Follows the procedure used in the ASTM inter-laboratory crosscheck program.

- Compute robust mean, RM, and robust standard deviation, RSD, for each combination of fuel sample, test method and calibration curve using a procedure that limits the influence of unusually large or small values.

- Classify an individual lab repeat value, Y, as an outlier and delete the value if

$$|Y – RM| > 3*RSD.$$

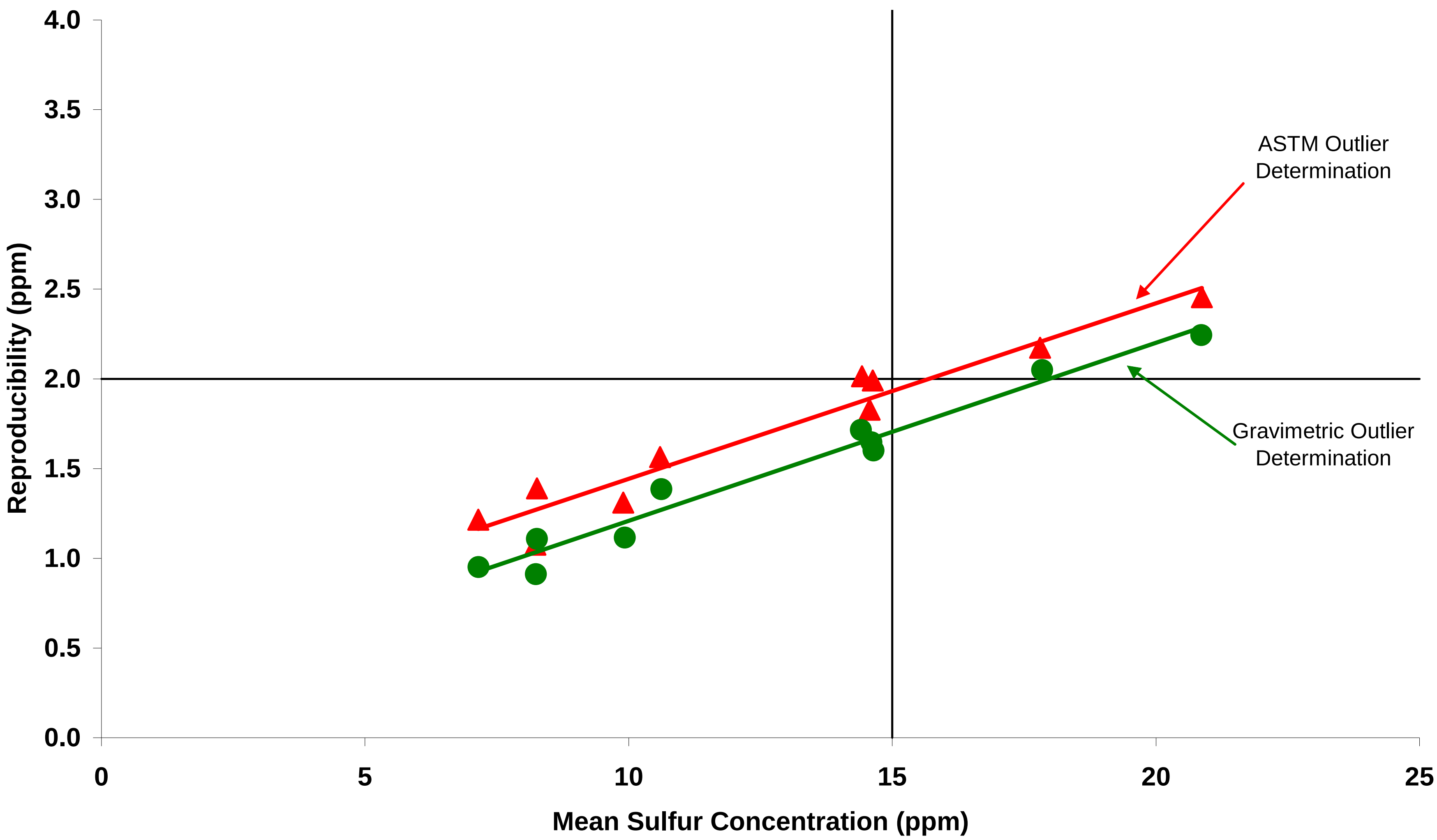
R&r Analysis Methods

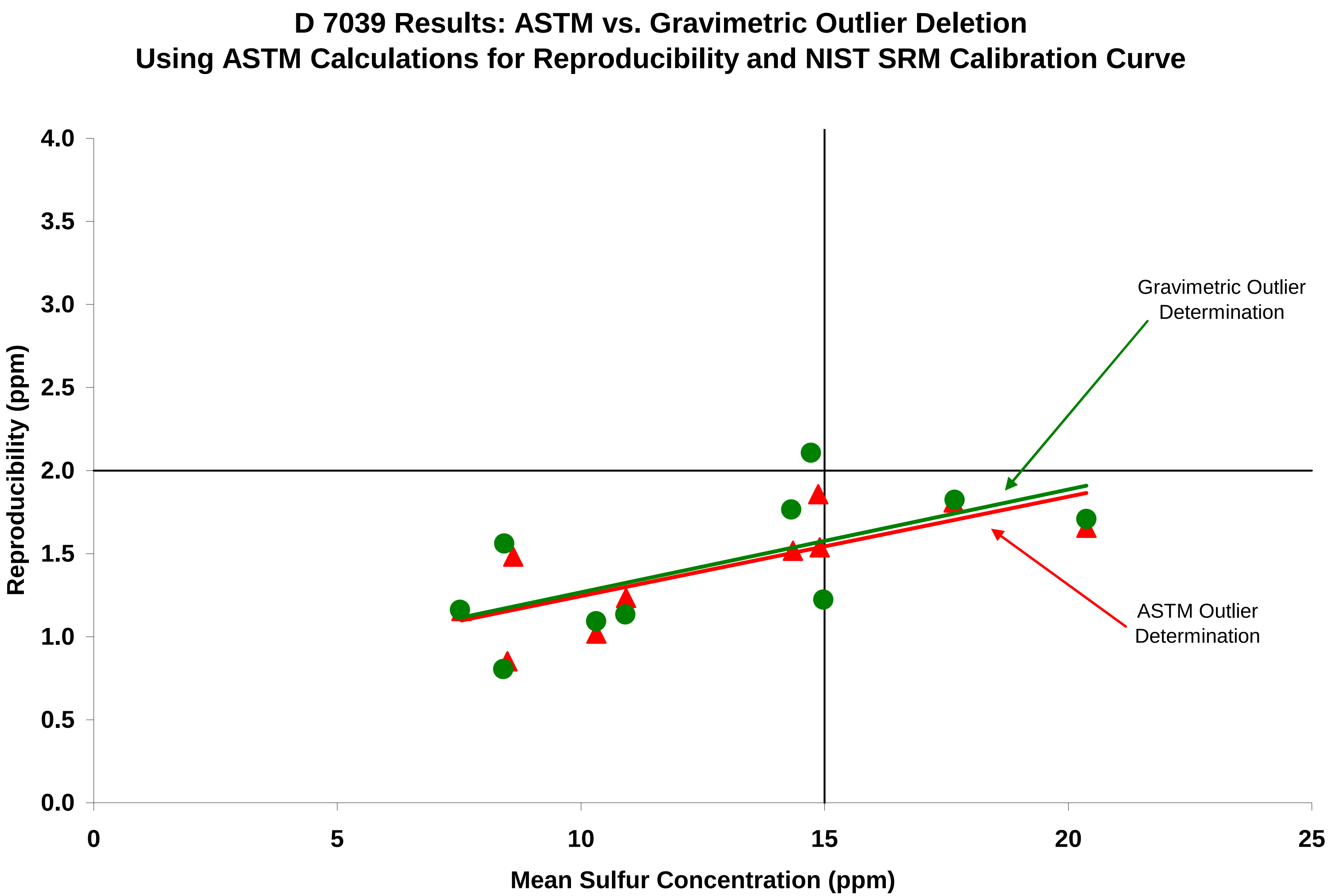
- Calculate R and r in two ways
  - Robust calculation identical to the ASTM crosscheck program.
  - Analysis of Variance (ANOVA) method.
- ANOVA results were different, but no clear advantage/disadvantage was evident.
- Therefore, only results using the robust ASTM calculation will be presented here.

Results

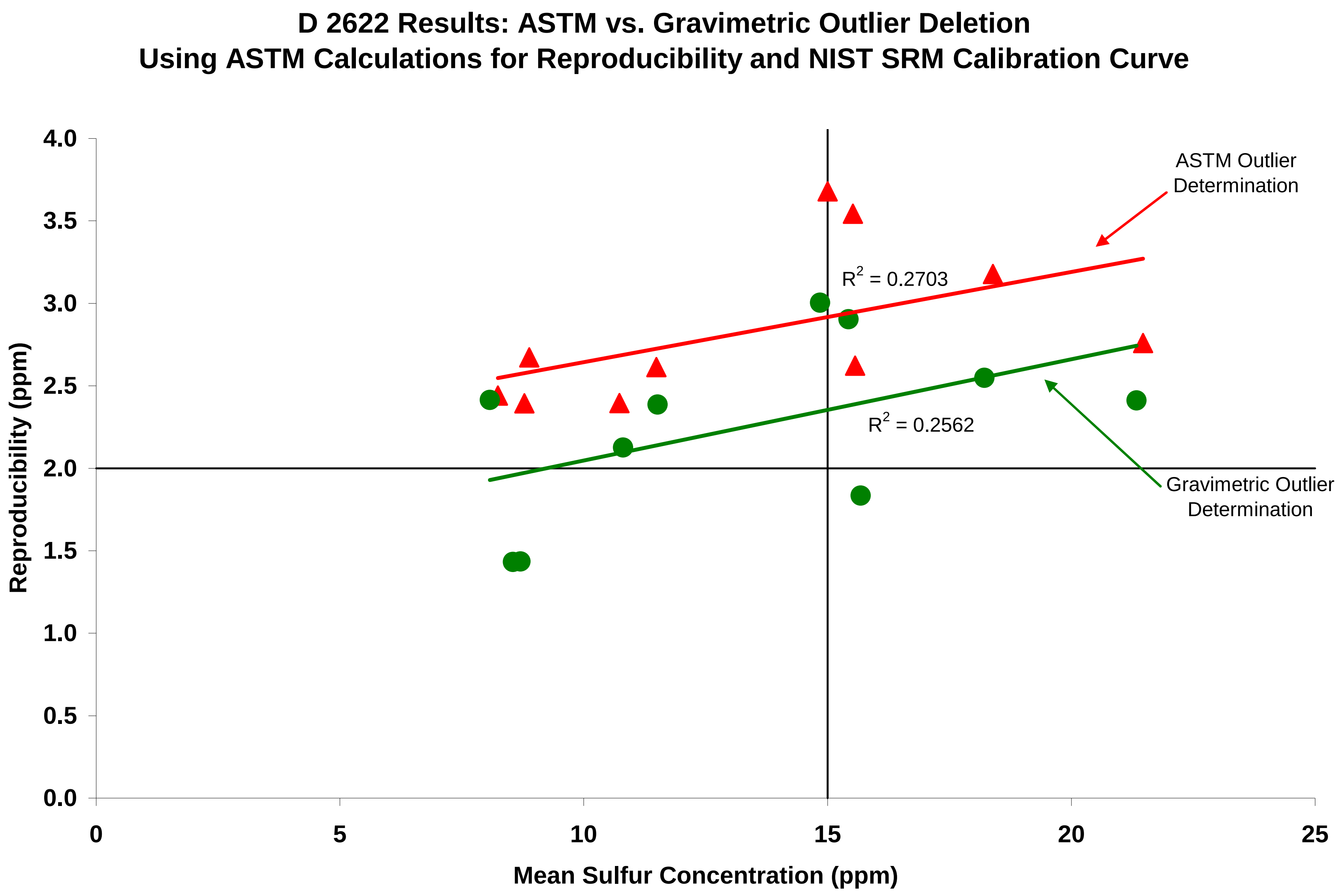
ASTM Robust Outlier Determination vs. Gravimetric Outlier Determination - Using ASTM Reproducibility Calculation

D 5453 Results: ASTM vs. Gravimetric Outlier Deletion
Using ASTM Calculations for Reproducibility and NIST SRM Calibration Curve





D 7039 Results: ASTM vs. Gravimetric Outlier Deletion
Using ASTM Calculations for Reproducibility and NIST SRM Calibration Curve
Reproducibility (ppm)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
Gravimetric Outlier Determination
ASTM Outlier Determination



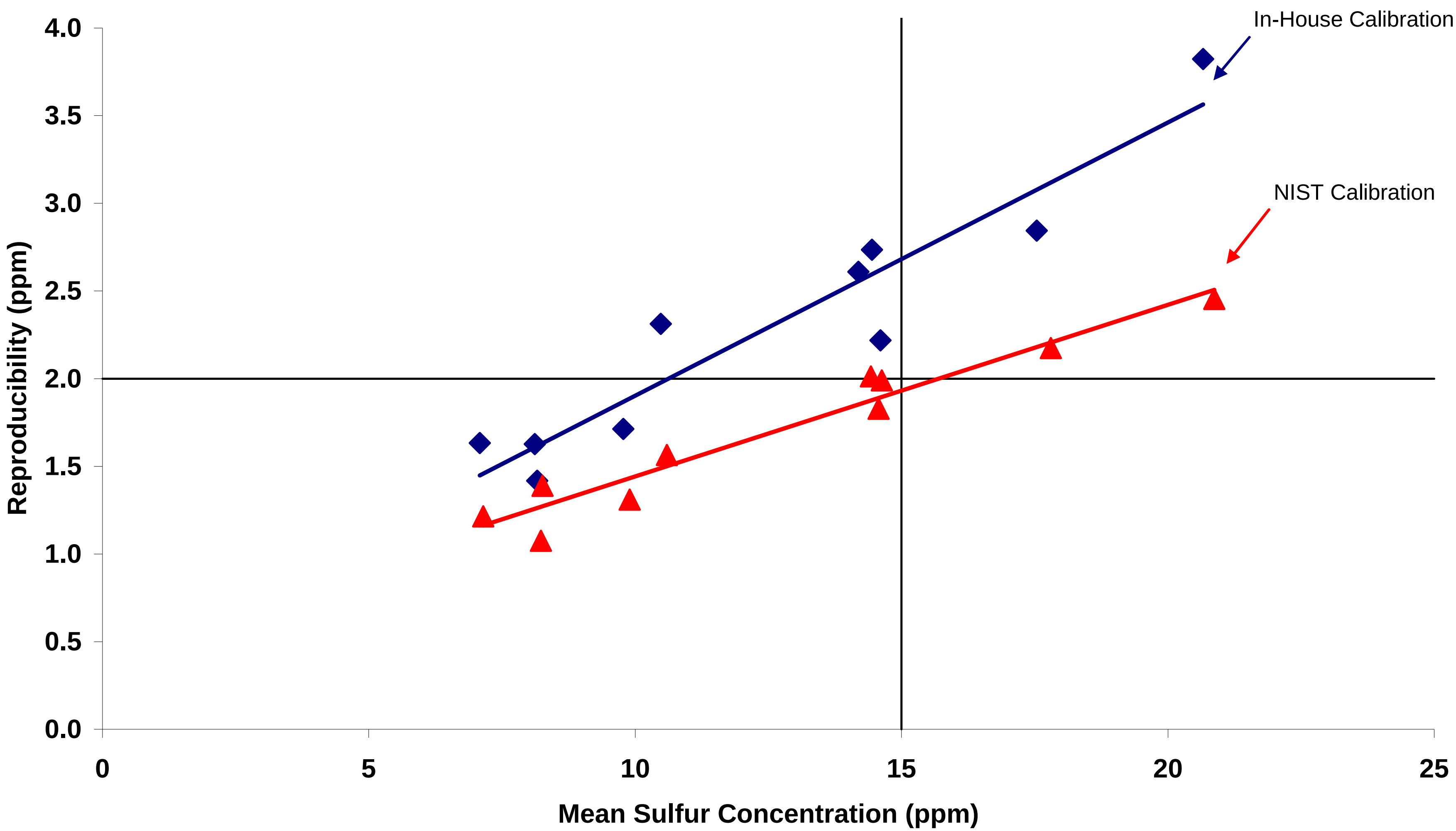
D 2622 Results: ASTM vs. Gravimetric Outlier Deletion
Using ASTM Calculations for Reproducibility and NIST SRM Calibration Curve
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Reproducibility (ppm)
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
ASTM Outlier Determination
R^2 = 0.2703
R^2 = 0.2562
Gravimetric Outlier Determination

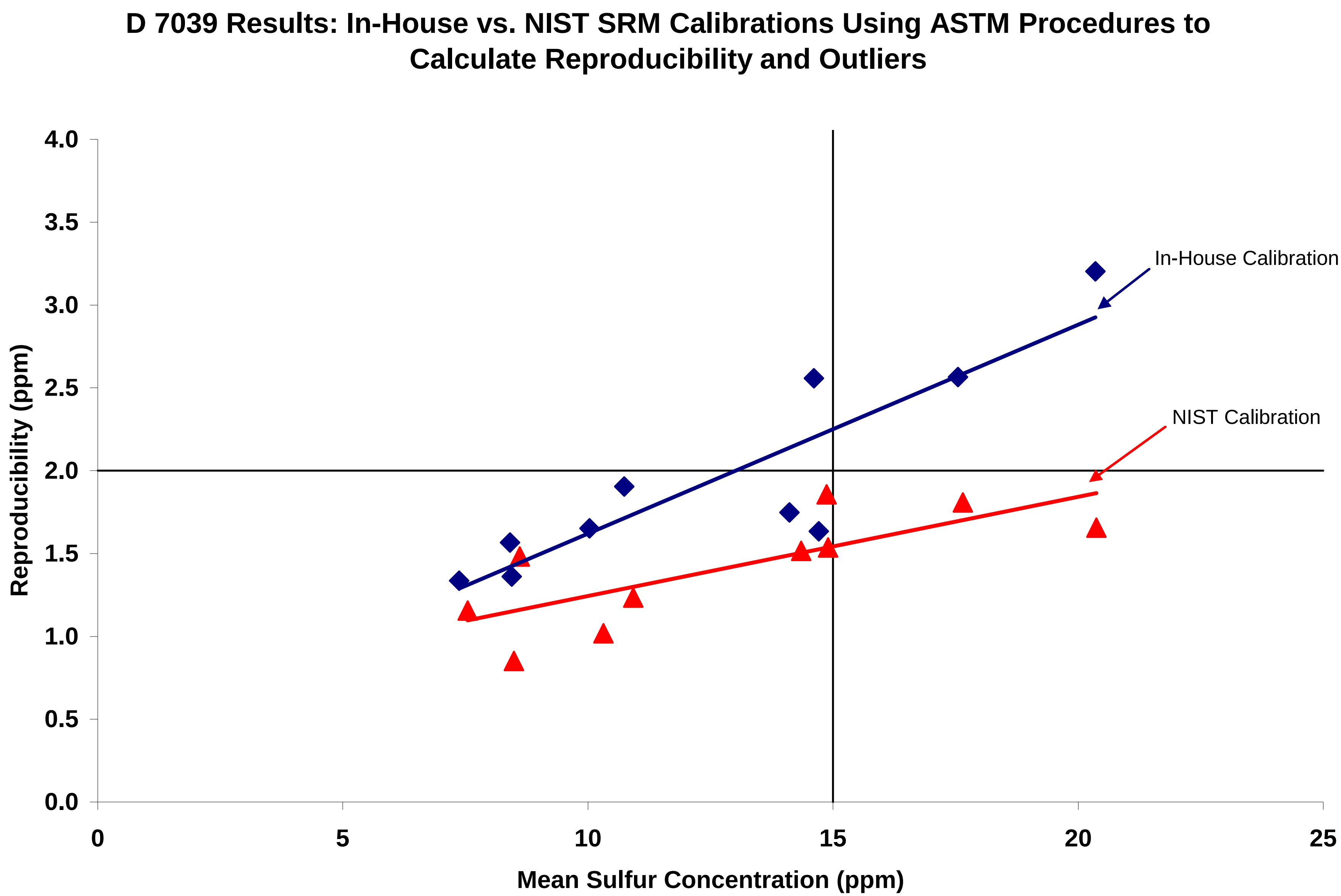
Conclusions

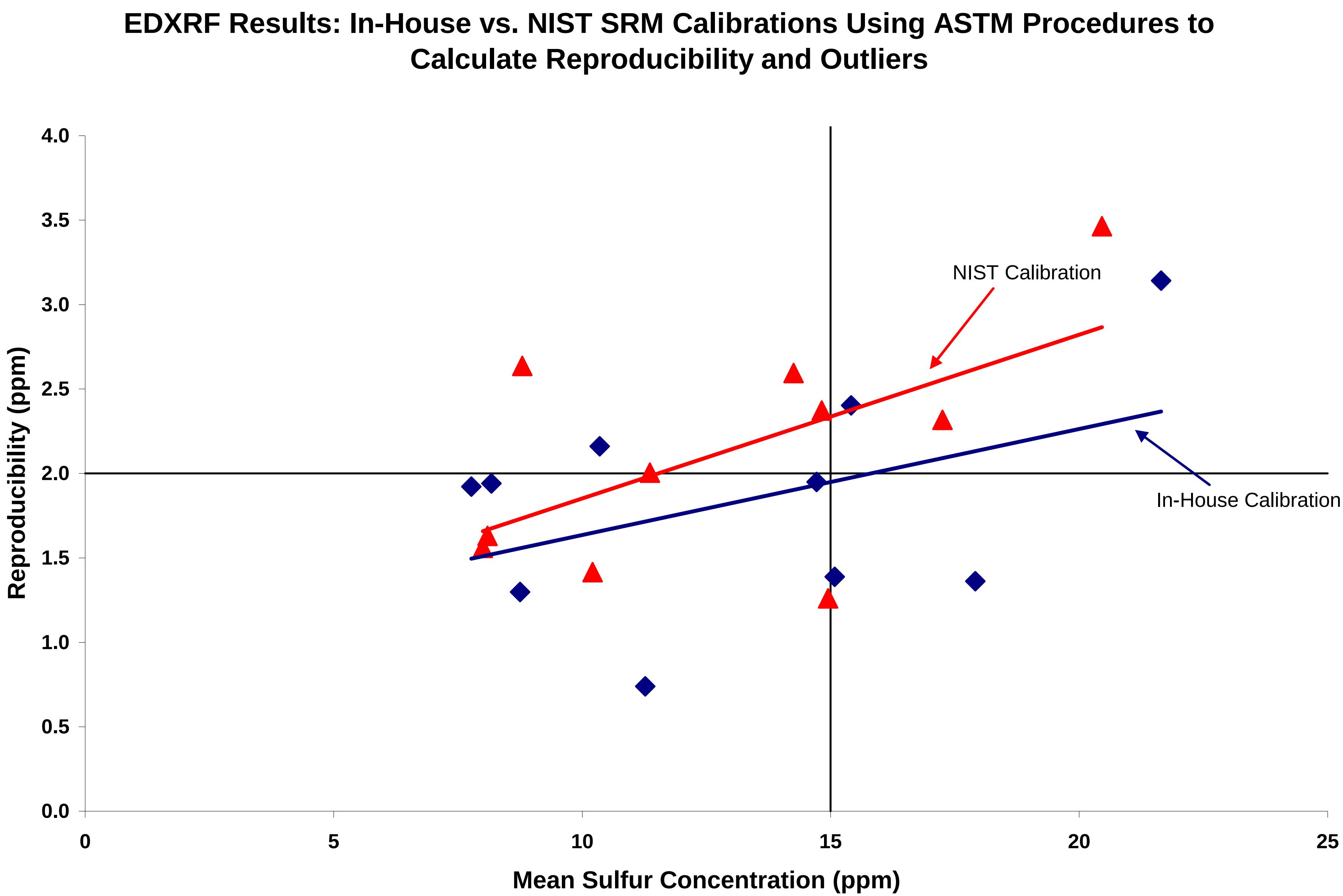
- The gravimetric deletion method produces lower R-values than the ASTM robust deletion method.
- For labs that can pass a calibration check standard, R is well below 2.0 ppm for D 5453 and D 7039.
- Oldest test method (D 2622) apparently not up to the challenge.
  - High R
  - Poor R^2
  - High variability may be due to wide range in instrument ages and capabilities of different instruments being used today.

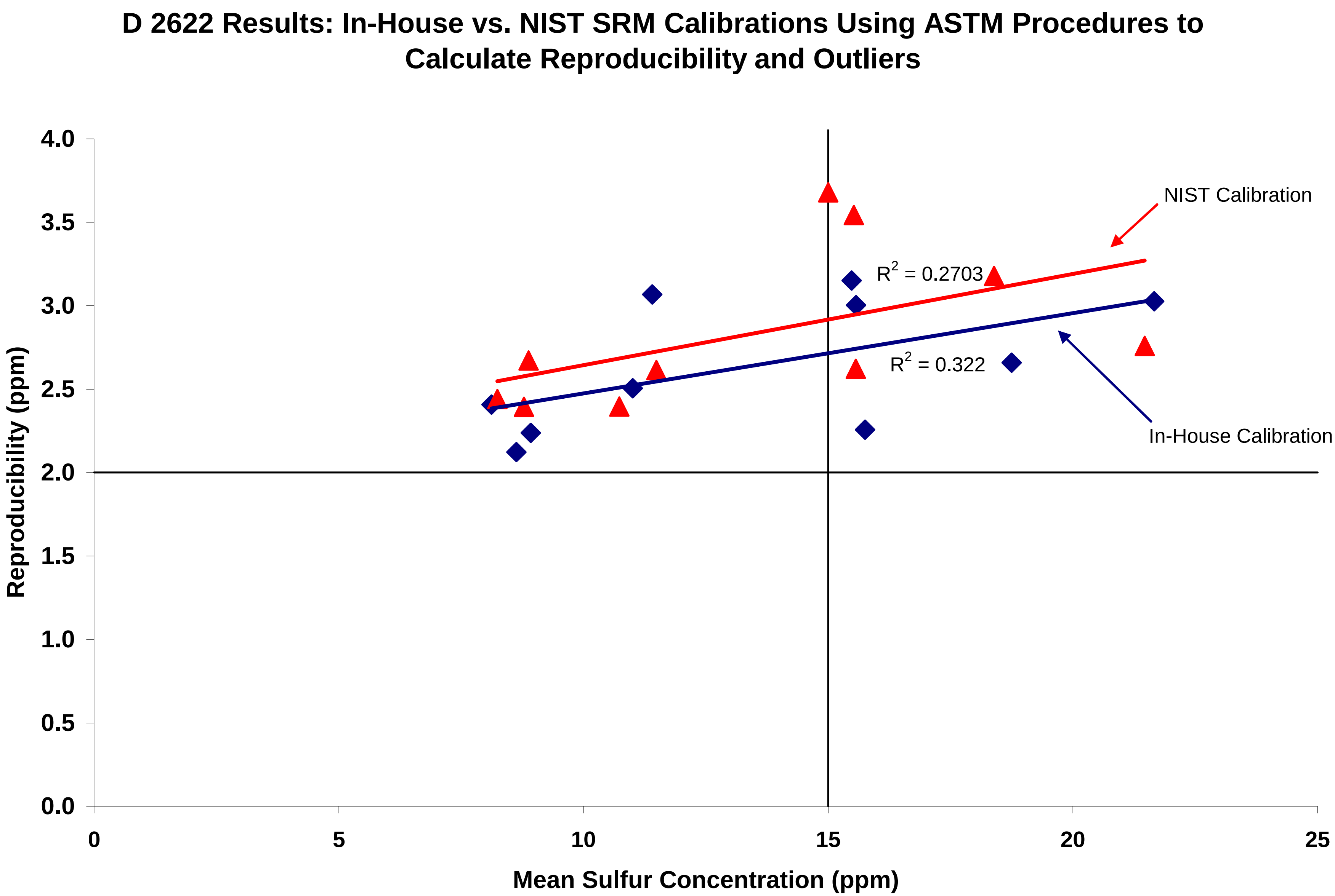
In-House Calibration Curves vs. NIST Calibration Curves – ASTM Reproducibility and ASTM Robust Outlier Determination

D 5453 Results: In-House vs. NIST SRM Calibrations Using ASTM Procedures to Calculate Reproducibility and Outliers




D 7039 Results: In-House vs. NIST SRM Calibrations Using ASTM Procedures to Calculate Reproducibility and Outliers
Reproducibility (ppm)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
In-House Calibration
NIST Calibration


EDXRF Results: In-House vs. NIST SRM Calibrations Using ASTM Procedures to Calculate Reproducibility and Outliers
Reproducibility (ppm)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
NIST Calibration
In-House Calibration

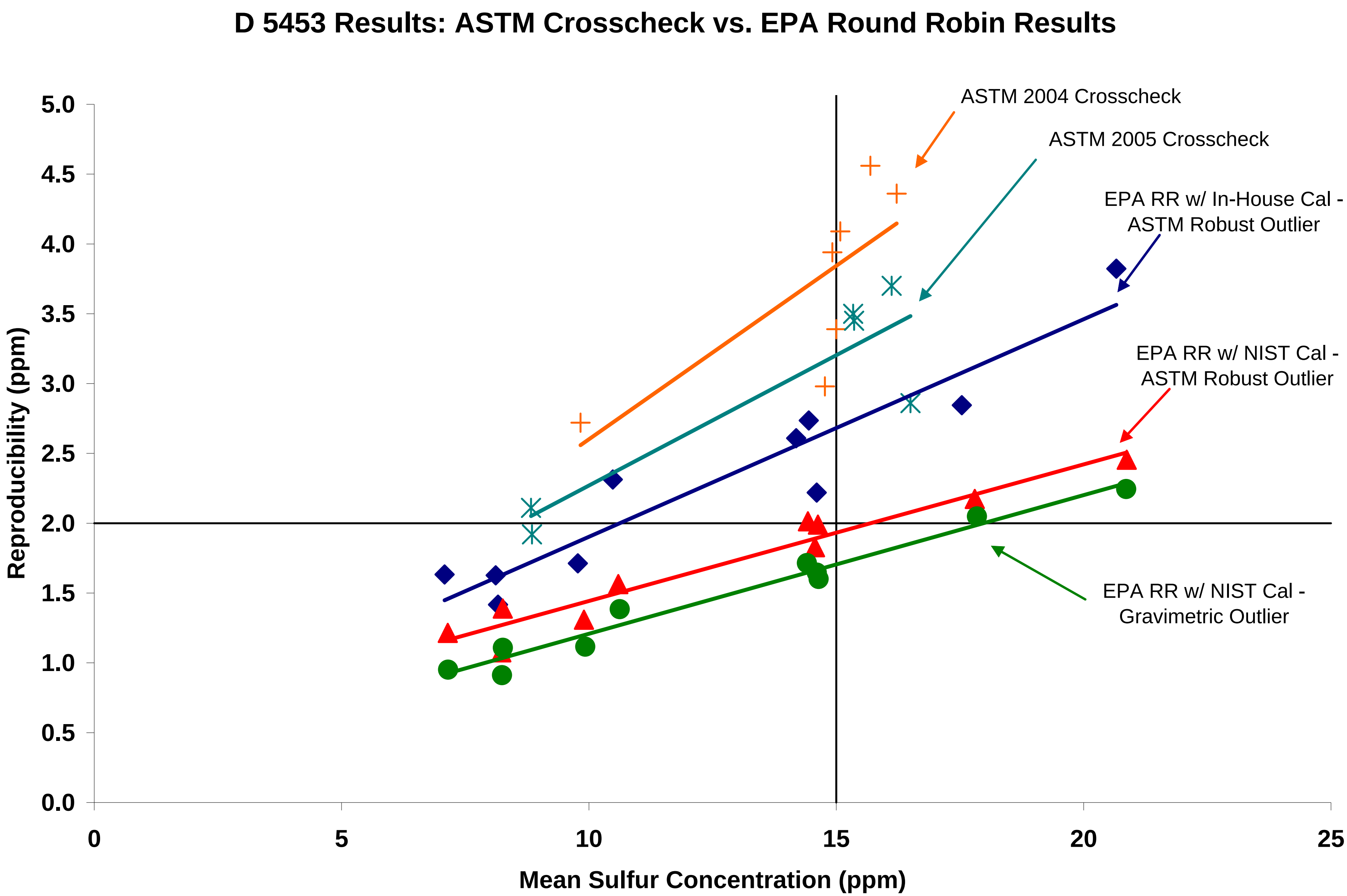

D 2622 Results: In-House vs. NIST SRM Calibrations Using ASTM Procedures to Calculate Reproducibility and Outliers
Reproducibility (ppm)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
NIST Calibration
R^2 = 0.2703
R^2 = 0.322
In-House Calibration

Conclusions

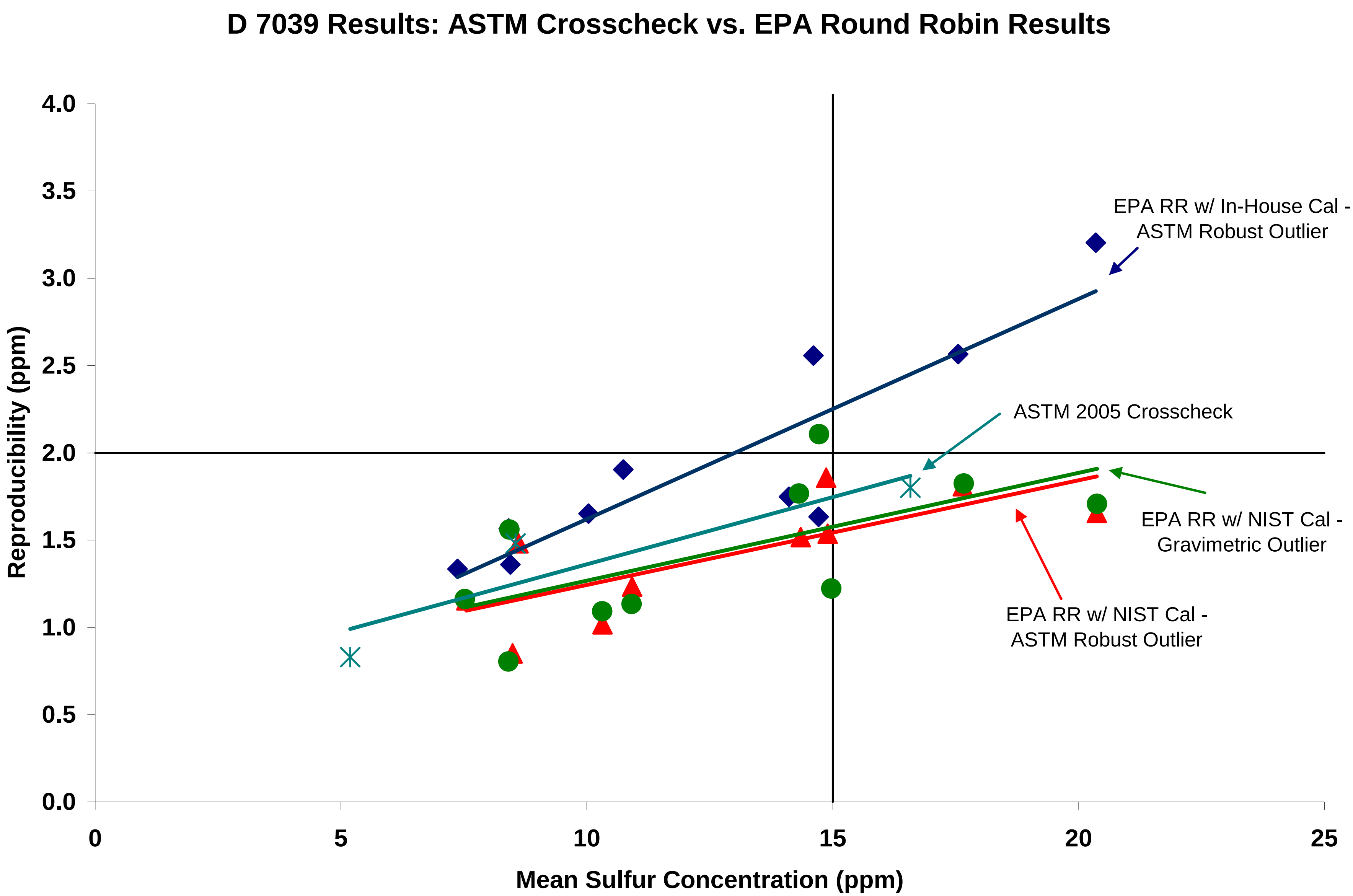
- The R-values for D 5453 and D 7039 are always less using the NIST calibration curves compared to the in-house calibration curves.
- The R-value results for D 2622 and EDXRF are mixed.

2004 and 2005 ASTM ULSD Crosscheck Results Comparison to EPA RR Results

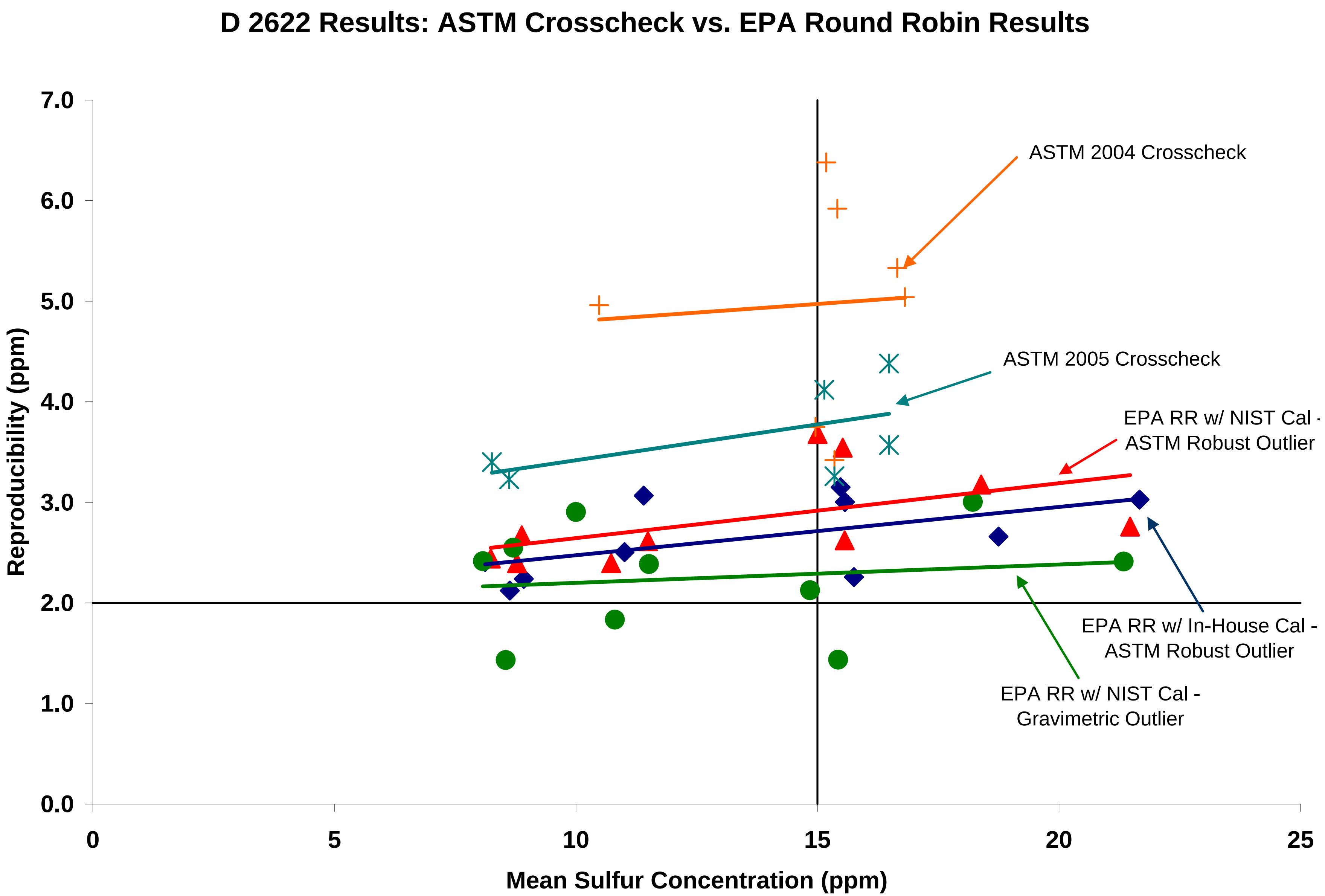
Using ASTM Robust Outlier Determination and Gravimetric Outlier Determination – ASTM Reproducibility Calculation



D 5453 Results: ASTM Crosscheck vs. EPA Round Robin Results
Reproducibility (ppm)
Mean Sulfur Concentration (ppm)
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
ASTM 2004 Crosscheck
ASTM 2005 Crosscheck
EPA RR w/ In-House Cal - ASTM Robust Outlier
EPA RR w/ NIST Cal - ASTM Robust Outlier
EPA RR w/ NIST Cal - Gravimetric Outlier



D 7039 Results: ASTM Crosscheck vs. EPA Round Robin Results
Reproducibility (ppm)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
EPA RR w/ In-House Cal - ASTM Robust Outlier
ASTM 2005 Crosscheck
EPA RR w/ NIST Cal - Gravimetric Outlier
EPA RR w/ NIST Cal - ASTM Robust Outlier


D 2622 Results: ASTM Crosscheck vs. EPA Round Robin Results
Reproducibility (ppm)
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
ASTM 2004 Crosscheck
ASTM 2005 Crosscheck
EPA RR w/ NIST Cal - ASTM Robust Outlier
EPA RR w/ In-House Cal - ASTM Robust Outlier
EPA RR w/ NIST Cal - Gravimetric Outlier

Conclusions

- Qualification process appears to have significantly improved R compared to ASTM crosscheck results.

Predicted Reproducibility at 15 ppm

Approach	Method	ASTM R Calculation
ASTM 2004 ILCP	D 2622	4.97
	D 5453	3.84
ASTM 2005 ILCP	D 2622	3.78
	D 5453	3.20
	D 7039	1.74
EPA RR Results NIST Calibration – Gravimetric Outlier Determination	D 2622	2.29
	D 5453	1.71
	D 7039	1.58
EPA RR Results NIST Calibration – ASTM Robust Outlier Determination	D 2622	2.91
	EDXRF	2.34
	D 5453	1.93
	D 7039	1.54
EPA RR Results In-House Calibration – ASTM Robust Outlier Determination	D 2622	2.71
	EDXRF	1.94
	D 5453	2.68
	D 7039	2.25

Conclusions Summary

- The regression equations produce lower predicted R-values (at 15 ppm) for the EPA RR results relative to the 2004 and 2005 ASTM CC results.
  - The data support the conclusion that limiting the RR participation to labs that have qualified their methods under 40 CFR 80.584 has had a favorable impact on lowering reproducibility.

Conclusions Summary

- The data also support the conclusion that using identical NIST calibration curves across participating labs reduces curve bias contributions to reproducibility.
  - A reduction in predicted R (at 15 ppm) over the predicted R-values obtained using the 2004 and 2005 ILCP data were apparent in all cases when using the NIST calibration curves.
  - The magnitude of the reduction in predicted R (at 15 ppm) from in-house to NIST under ASTM robust deletion was 0.73 ppm on average for D 5453 and D 7039.
- Using gravimetric outlier deletion further improves reproducibility.
  - Use of this method can be analogous to a calibration check standard.
- New test methods are producing results with lower R (D 5453 and especially D 7039).