# EPA ULSD Round Robin Data Analysis Results

*Robert L. Mason, Ph.D.*
*Janet P. Buckingham*

*Southwest Research Institute*
*Fuels & Lubricants Research Division*

*November 1, 2005*

# *Program Overview*

- EPA issued a Work Assignment to Southwest Research Institute in late March 2005.
- The objective was to review EPA's test plan and data processing for its 2005 ULSD Round-Robin test program, and, upon receipt of the sulfur results from the participating labs, to perform statistical analyses using the recommended procedures identified after the analysis review.
- The project began on April 4, 2005, and extends until the end of November.

# ULSD Round-Robin Test Design

- Test fuel samples=10
  – Sulfur range is 7-21 ppm.
- Test methods=6
  – D5453 D2622 D7039 EDXRF D3120 D7041
  – Composite=All except D7041
- Calibration curves=2
  – In-House, NIST
- Outlier detection methods=2
  – Robust, Gravimetric
- Analysis methods=2
  – ASTM Crosscheck, ANOVA (analysis of variance)

# Outlier Deletion Methods

- Two methods of outlier deletion were used in this program: robust outlier deletion and gravimetric outlier deletion.

- The robust outlier deletion method is the one currently used in the ASTM inter-laboratory crosscheck program (ILCP).
  – It does not require known fuel sulfur values for any of the sample fuels.

- The gravimetric outlier deletion method is an alternative approach proposed by EPA.
  – It is a possible approach when known gravimetric fuel standards exist – as in this case.
  – It can be seen as a surrogate to a calibration check standard.

# Robust Outlier Deletion Method

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

$$|Y - RM| > 3 * RSD.$$

# *Gravimetric Outlier Deletion Method*

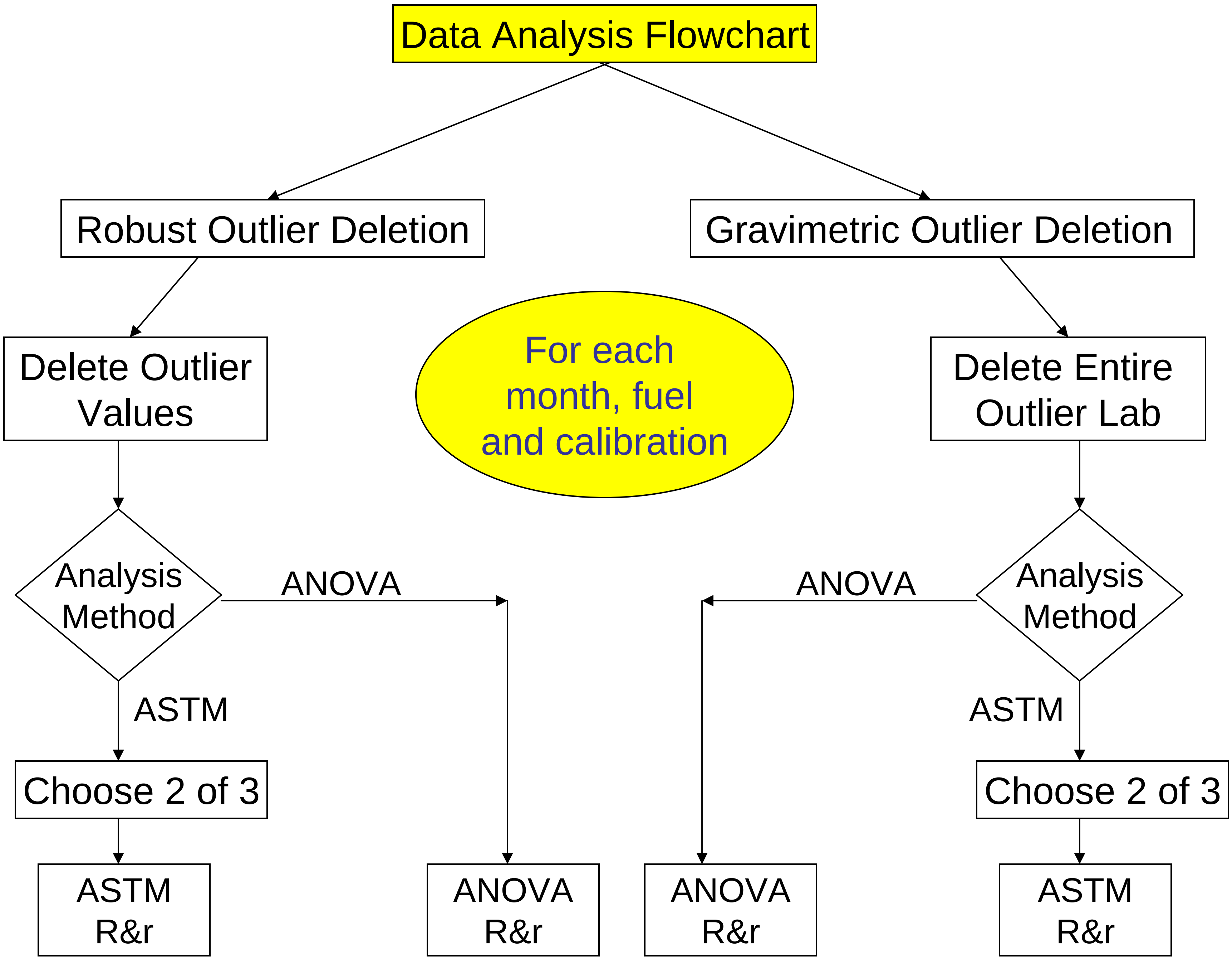
- Compute the average, AVG, of the three repeat tests taken on the gravimetric standard fuel for a given month by a given lab.
  - Use either July fuel #4 or August fuel #4.
- Obtain the accepted reference value (ARV) of the standard fuel.
  - ARV=8.41 ppm in this study.
- Classify the data collected on all five sample fuels for a given month by a given lab as outliers and delete the entire set of lab data if

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

# *R&r Analysis Methods*

- ASTM Crosscheck Method
  - Requires further data deletion (beyond outlier deletion) to reduce repeat count from 3 to 2 per fuel sample, when it occurs.
  - Uses the ILCP iterative algorithm to compute robust mean and robust standard deviation with remaining data.
  - Computes reproducibility (R) and repeatability (r) using the above robust estimates.

- ANOVA Method
  - Uses all available data after outlier deletion.
  - Applies one-factor analysis of variance technique using lab as the single factor.
  - Estimates R and r using variance component estimates of lab (reproducibility) and of error (repeatability).


Data Analysis Flowchart
Robust Outlier Deletion
Gravimetric Outlier Deletion
Delete Outlier Values
For each month, fuel and calibration
Delete Entire Outlier Lab
Analysis Method
ANOVA
ANOVA
Analysis Method
ASTM
ASTM
Choose 2 of 3
Choose 2 of 3
ASTM R&r
ANOVA R&r
ANOVA R&r
ASTM R&r

# *Instrument/Lab Counts*

| Test Method | July 2005 | August 2005 |
|---|---|---|
| **D5453** | 98 | 93 |
| **D2622** | 25 | 24 |
| **D7039** | 16 | 16 |
| **EDXRF** | 6 | 6 |
| **D3120*** | 3 | 3 |
| **D7041**** | 1 | 1 |
| **Total Instruments** | 149 | 143 |
| **Total Labs** | 129 | 125 |

* Only used in the Composite calculations.

** Not used in any of the analyses.

# *Comparisons Before and After Outlier Deletion*

- Means are computed and compared for the following cases:
  - The mean of all the sulfur data in the specified data set is computed.
  - The robust mean of all the sulfur data in the specified data set is computed.
  - The robust mean of the sulfur data remaining after outlier deletion is computed.
- Standard deviations are computed and compared for the following cases:
  - The std. dev. of all the sulfur data in the specified data set is computed.
  - The robust std. dev. of all the sulfur data in the specified data set is computed.
  - The robust std. dev. of the data remaining after outlier deletion is computed.

# *Composite Mean Comparisons Robust Deletion Method*

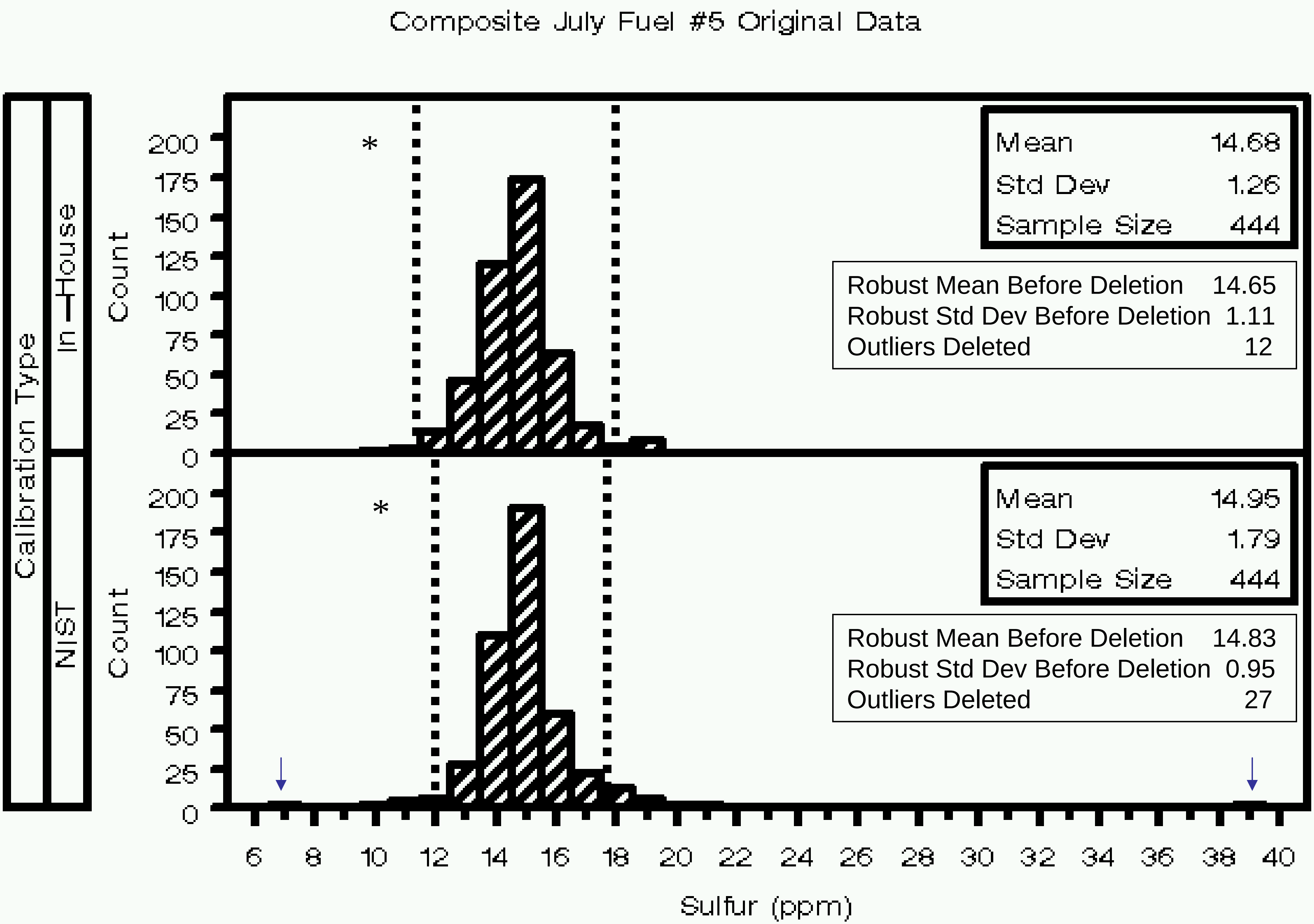
| | Fuel No. | Mean of All Data | Robust Mean Before Outlier Deletion | Robust Mean After Outlier Deletion | No. of Outliers Deleted |
|---|---|---|---|---|---|
| **July In-House** | **1** | 7.26 | 7.25 | 7.24 | 9 |
| | **2** | 10.66 | 10.66 | 10.65 | 8 |
| | **3** | 20.77 | 20.80 | 20.80 | 9 |
| | **4** | 8.29 | 8.26 | 8.25 | 7 |
| | **5** | 14.68 | 14.65 | 14.63 | 12 |

| | Fuel No. | Mean of All Data | Robust Mean Before Outlier Deletion | Robust Mean After Outlier Deletion | No. of Outliers Deleted |
|---|---|---|---|---|---|
| **July NIST** | **1** | 7.44 | 7.38 | 7.36 | 13 |
| | **2** | 10.89 | 10.83 | 10.80 | 18 |
| | **3** | 21.15 | 20.97 | 20.90 | 24 |
| | **4** | 8.50 | 8.42 | 8.38 | 31 |
| | **5** | 14.95 | 14.83 | 14.79 | 27 |

# Composite Std. Dev. Comparisons Robust Deletion Method

| | Fuel No. | Std. Dev. Of All Data | Robust Std.Dev. Before Outlier Deletion | Robust Std.Dev. After Outlier Deletion | No. of Outliers Deleted |
|---|---|---|---|---|---|
| July In-House | 1 | 0.87 | 0.73 | 0.71 | 9 |
| | 2 | 1.01 | 0.89 | 0.86 | 8 |
| | 3 | 1.64 | 1.41 | 1.36 | 9 |
| | 4 | 0.75 | 0.69 | 0.67 | 7 |
| | 5 | 1.26 | 1.11 | 1.05 | 12 |

| | Fuel No. | Std. Dev. Of All Data | Robust Std.Dev. Before Outlier Deletion | Robust Std.Dev. After Outlier Deletion | No. of Outliers Deleted |
|---|---|---|---|---|---|
| July NIST | 1 | 0.98 | 0.72 | 0.66 | 13 |
| | 2 | 1.03 | 0.80 | 0.73 | 18 |
| | 3 | 1.68 | 1.05 | 0.95 | 24 |
| | 4 | 0.92 | 0.62 | 0.52 | 31 |
| | 5 | 1.79 | 0.95 | 0.83 | 27 |




*The dashed lines refer to the limits: (robust mean before deletion) +/- 3*(robust std. dev. before deletion).

# *Summary of Robust Deletion Comparisons*

- The robust means after outlier deletion are slightly reduced from the corresponding means of all the data in all but one case for the July fuels regardless of calibration method. The exception is for July In-House fuel #3 where there was a slight increase.

- The robust std. devs. after outlier deletion are reduced from the corresponding std. devs. of all the data for all the July fuels regardless of calibration, and the NIST data has smaller robust std. devs. after outlier deletion than the In-House data.

- In the histogram of the Composite data for the July fuel #5 (with a target of 15 ppm), the std. dev. for all the data was larger for the NIST calibration, but the robust std. deviation before outlier deletion was smaller for the NIST calibration.
  - This result created narrower NIST limits and thus more outlier deletions for the NIST data.
  - It appears that a few labs had difficulty calibrating their instruments with the NIST SRMs.

# *Composite Mean Comparisons Gravimetric Deletion Method*

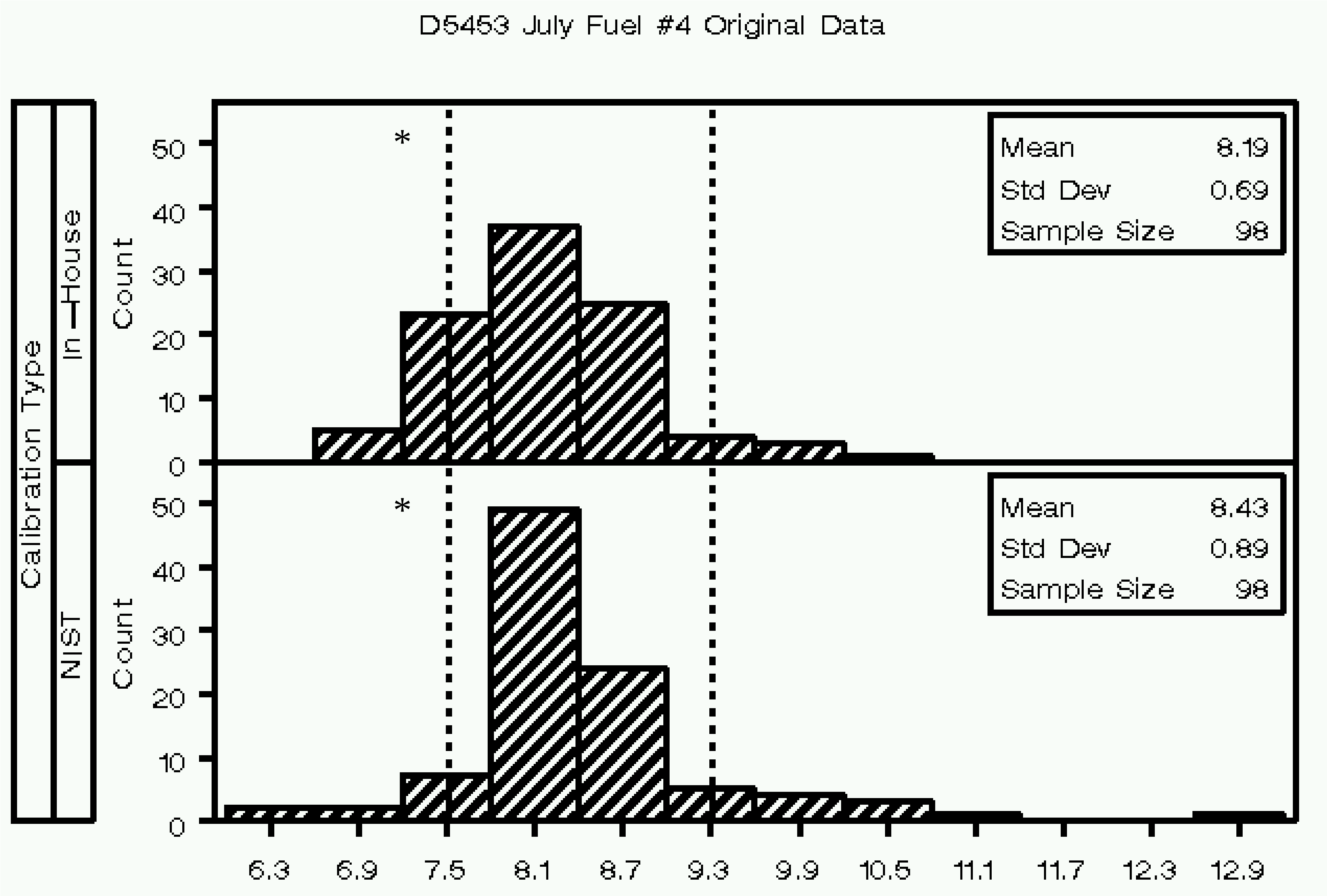
| | Fuel No. | Mean of All Data | Robust Mean Before Outlier Deletion | Robust Mean After Outlier Deletion | No. of Outliers (Labs) Deleted |
|---|---|---|---|---|---|
| **July In-House** | **1** | 7.26 | 7.25 | 7.29 | 96 (32) |
| | **2** | 10.66 | 10.66 | 10.75 | 96 (32) |
| | **3** | 20.77 | 20.80 | 20.91 | 96 (32) |
| | **4** | 8.29 | 8.26 | 8.33 | 96 (32) |
| | **5** | 14.68 | 14.65 | 14.75 | 96 (32) |

| | Fuel No. | Mean of All Data | Robust Mean Before Outlier Deletion | Robust Mean After Outlier Deletion | No. of Outliers (Labs) Deleted |
|---|---|---|---|---|---|
| **July NIST** | **1** | 7.44 | 7.38 | 7.30 | 81 (27) |
| | **2** | 10.89 | 10.83 | 10.77 | 81 (27) |
| | **3** | 21.15 | 20.97 | 20.89 | 81 (27) |
| | **4** | 8.50 | 8.42 | 8.34 | 81 (27) |
| | **5** | 14.95 | 14.83 | 14.78 | 81 (27) |

# Composite Std. Dev. Comparisons Gravimetric Deletion Method

| | Fuel No. | Std. Dev. Of All Data | Robust Std.Dev. Before Outlier Deletion | Robust Std.Dev. After Outlier Deletion | No. of Outliers (Labs) Deleted |
|---|---|---|---|---|---|
| July In-House | 1 | 0.87 | 0.73 | 0.63 | 96 (32) |
| | 2 | 1.01 | 0.89 | 0.72 | 96 (32) |
| | 3 | 1.64 | 1.41 | 1.07 | 96 (32) |
| | 4 | 0.75 | 0.69 | 0.48 | 96 (32) |
| | 5 | 1.26 | 1.11 | 0.83 | 96 (32) |

| | Fuel No. | Std. Dev. Of All Data | Robust Std.Dev. Before Outlier Deletion | Robust Std.Dev. After Outlier Deletion | No. of Outliers (Labs) Deleted |
|---|---|---|---|---|---|
| July NIST | 1 | 0.98 | 0.72 | 0.51 | 81 (27) |
| | 2 | 1.03 | 0.80 | 0.62 | 81 (27) |
| | 3 | 1.68 | 1.05 | 0.86 | 81 (27) |
| | 4 | 0.92 | 0.62 | 0.42 | 81 (27) |
| | 5 | 1.79 | 0.95 | 0.72 | 81 (27) |



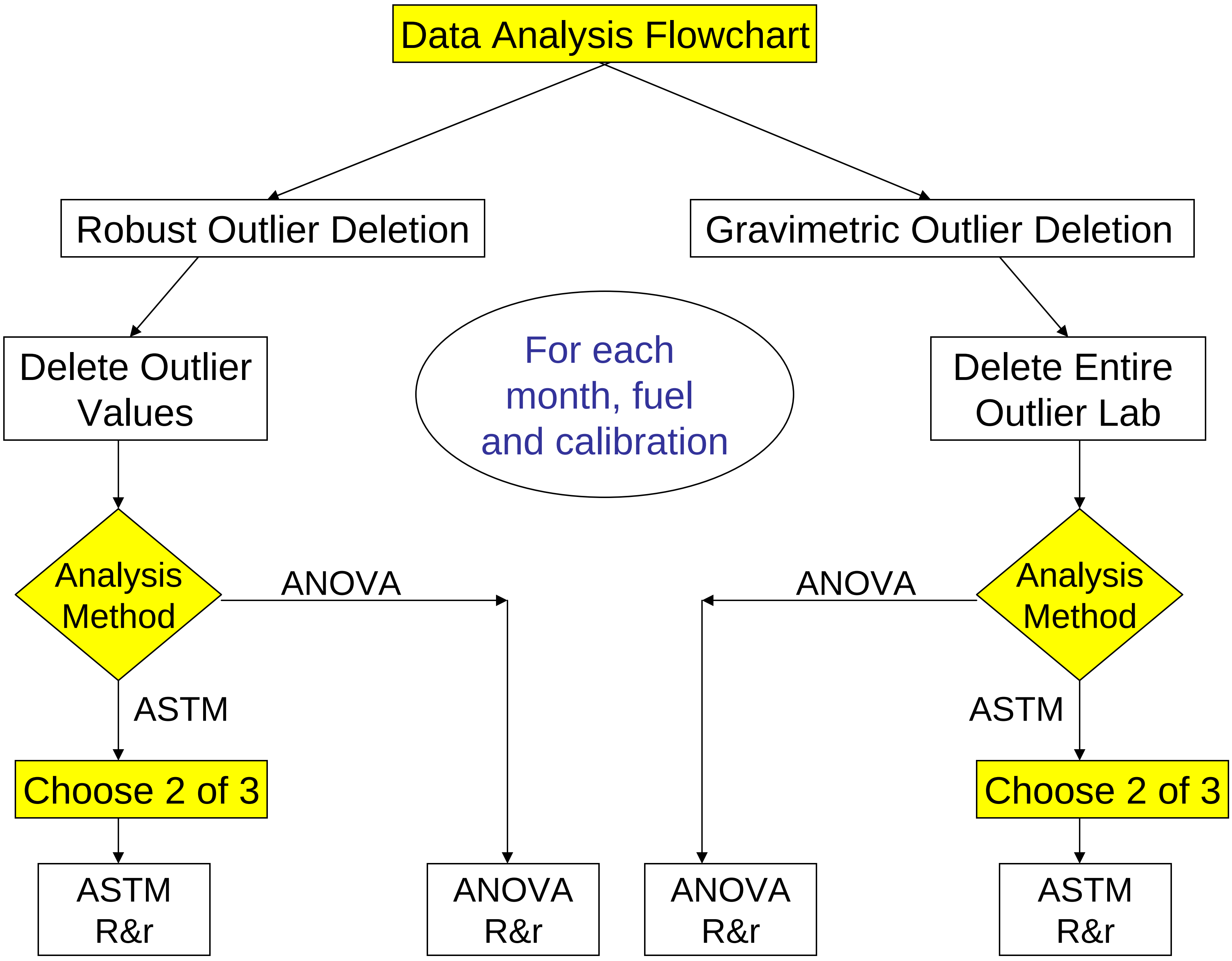

* The two dashed lines refer to the limits set by using 8.41 +/- 0.90.

# *Summary of Gravimetric Deletion Comparisons*

- The robust means after outlier deletion are slightly larger than the means for all the data for the 5 July fuels with the In-House calibration, and slightly smaller with the NIST calibration.

- The robust std. devs. after outlier deletion are reduced from the std. devs. for all the data, for all 5 July fuels regardless of calibration. The NIST data has smaller robust std. devs. after outlier deletion than the In-House data.

- In the histogram of the means of the D5453 lab repeats for the July fuel #4 (with a target of 8.41 ppm) data, the deletion limits are the same for both sets of calibration data. Since the mean of all the data is shifted to the left of the standard value of 8.41 for the In-House data and is nearly identical to 8.41 for the NIST data, the number of outlier deletions is increased when using the In-House method.
  - This result causes more outlier deletions for the In-House data despite its smaller std. dev.


Data Analysis Flowchart
Robust Outlier Deletion
Gravimetric Outlier Deletion
Delete Outlier Values
For each month, fuel and calibration
Delete Entire Outlier Lab
Analysis Method
ANOVA
ANOVA
Analysis Method
ASTM
ASTM
Choose 2 of 3
Choose 2 of 3
ASTM R&r
ANOVA R&r
ANOVA R&r
ASTM R&r

# *Outlier and Repeat Deletion Counts July Fuel #5*

| | Test Method | Total Count | In-House | | NIST | |
|---|---|---|---|---|---|---|
| | | | Outliers Deleted | Repeats Deleted | Outliers Deleted | Repeats Deleted |
| Robust Deletion | **D5453** | 294 | 10 | 94 | 29 | 87 |
| | **D2622** | 75 | 0 | 25 | 1 | 24 |
| | **D7039** | 48 | 0 | 16 | 2 | 14 |
| | **EDXRF** | 18 | 0 | 6 | 0 | 6 |
| | **Composite** | 444 | 12 | 143 | 27 | 141 |

| | Test Method | Total Count | In-House | | NIST | |
|---|---|---|---|---|---|---|
| | | | Outliers Deleted | Repeats Deleted | Outliers Deleted | Repeats Deleted |
| Gravimetric Deletion | | | | 75 | | |
| | | | | 20 | | |
| | | | | 14 | | |
| | | | | 6 | 3 | |
| | | | | 116 | 81 | |



* July fuel #5 had a 15 ppm sulfur target

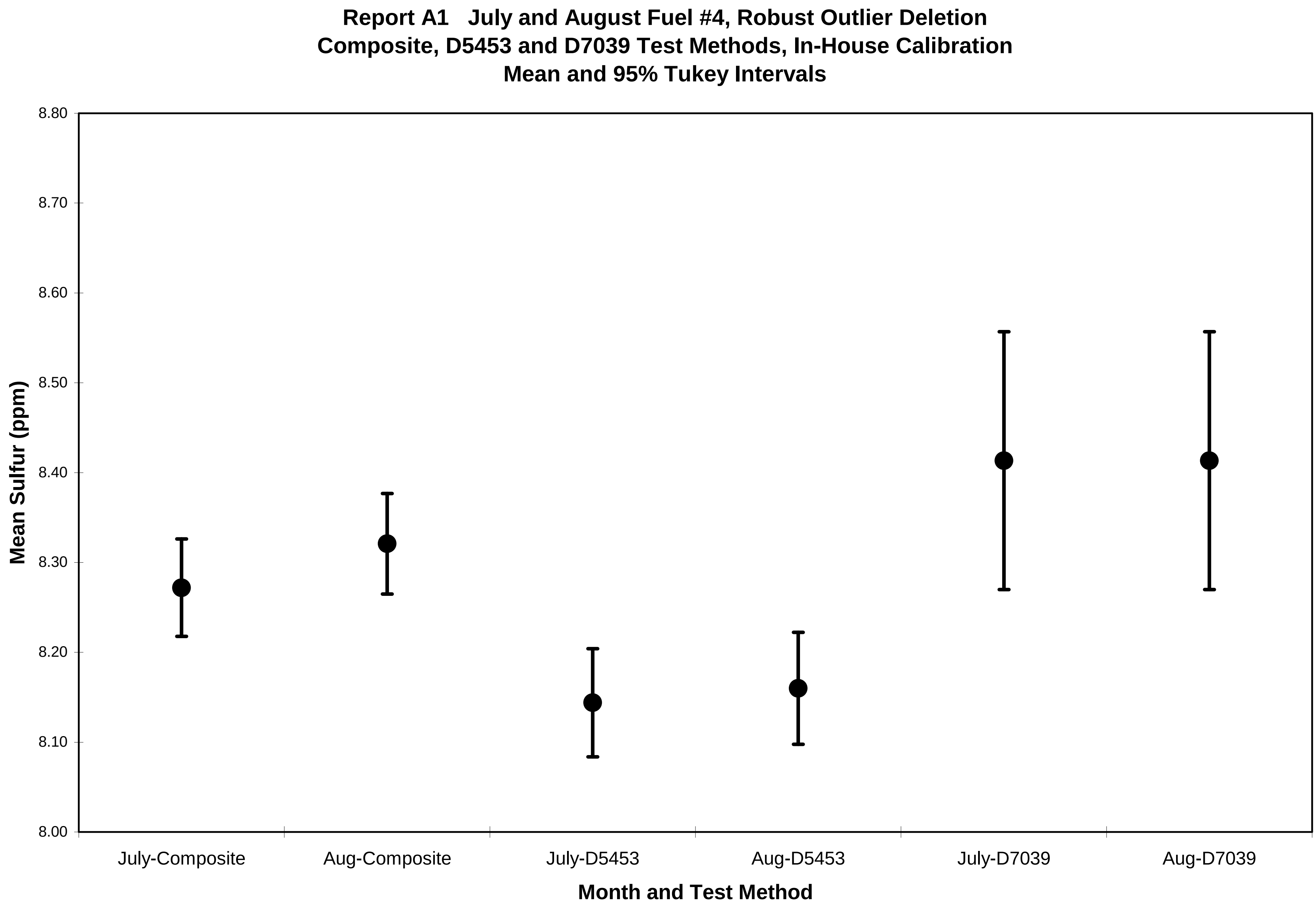
# *Summary of Outlier and Repeat Deletion Counts – July Fuel #5*

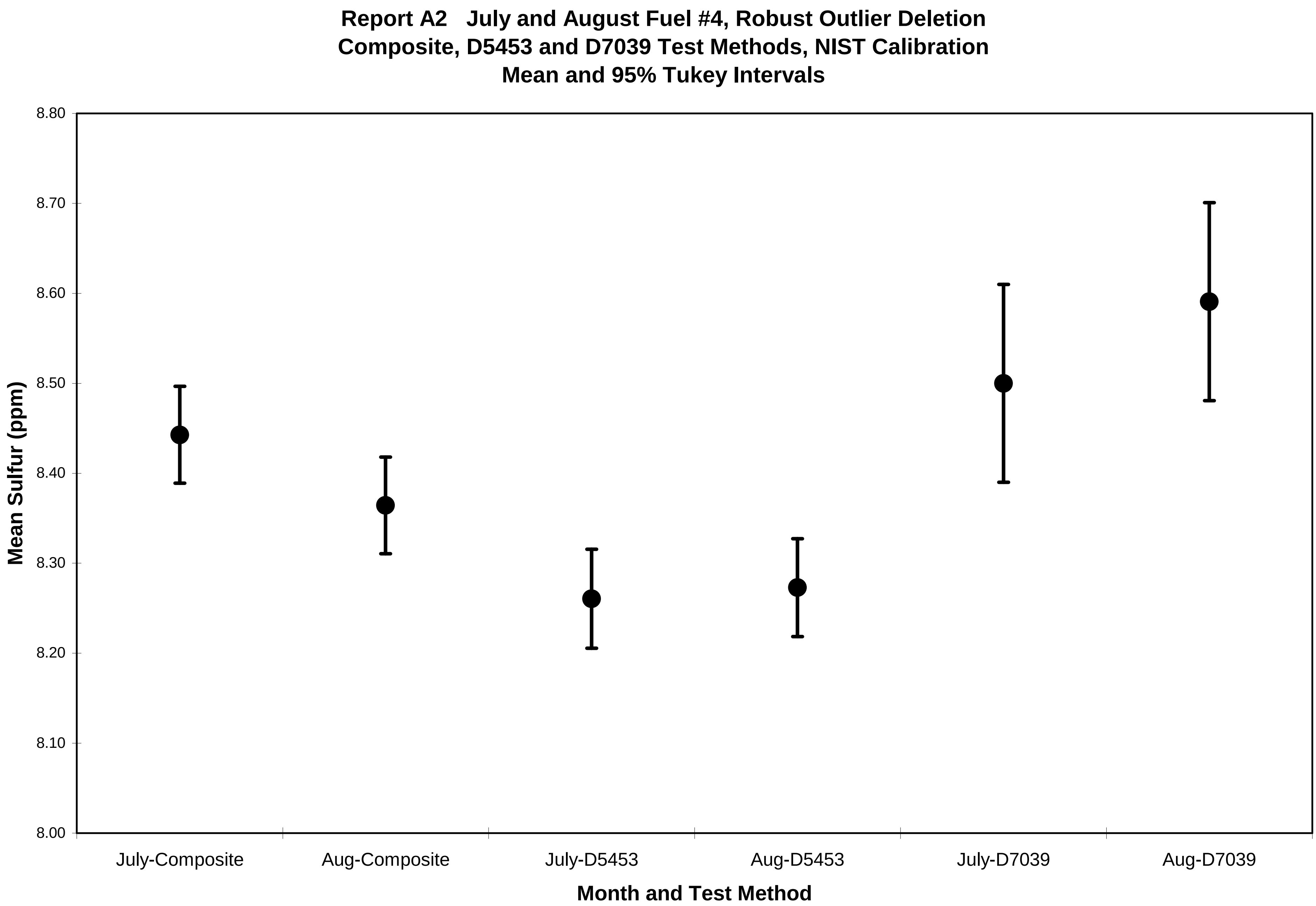
- Using robust deletion, more outliers were removed with the NIST data than with the In-House data, though a similar number of repeat runs were deleted for the two calibration methods. This is a result of narrower limits and a smaller robust standard deviation after deletion for the NIST data.
- More values were deleted using the gravimetric deletion method versus using the robust deletion method. This is a result of the entire set of lab data for a given month being deleted if an outlier occurs in the gravimetric standard fuel values.
- Using gravimetric deletion, more outliers were removed with the In-House data than with the NIST data, though a similar number of repeat runs were deleted for the two calibration methods. This is a result of the increased bias in the mean of the In-House data and the use of the same deletion limits for both sets of data.

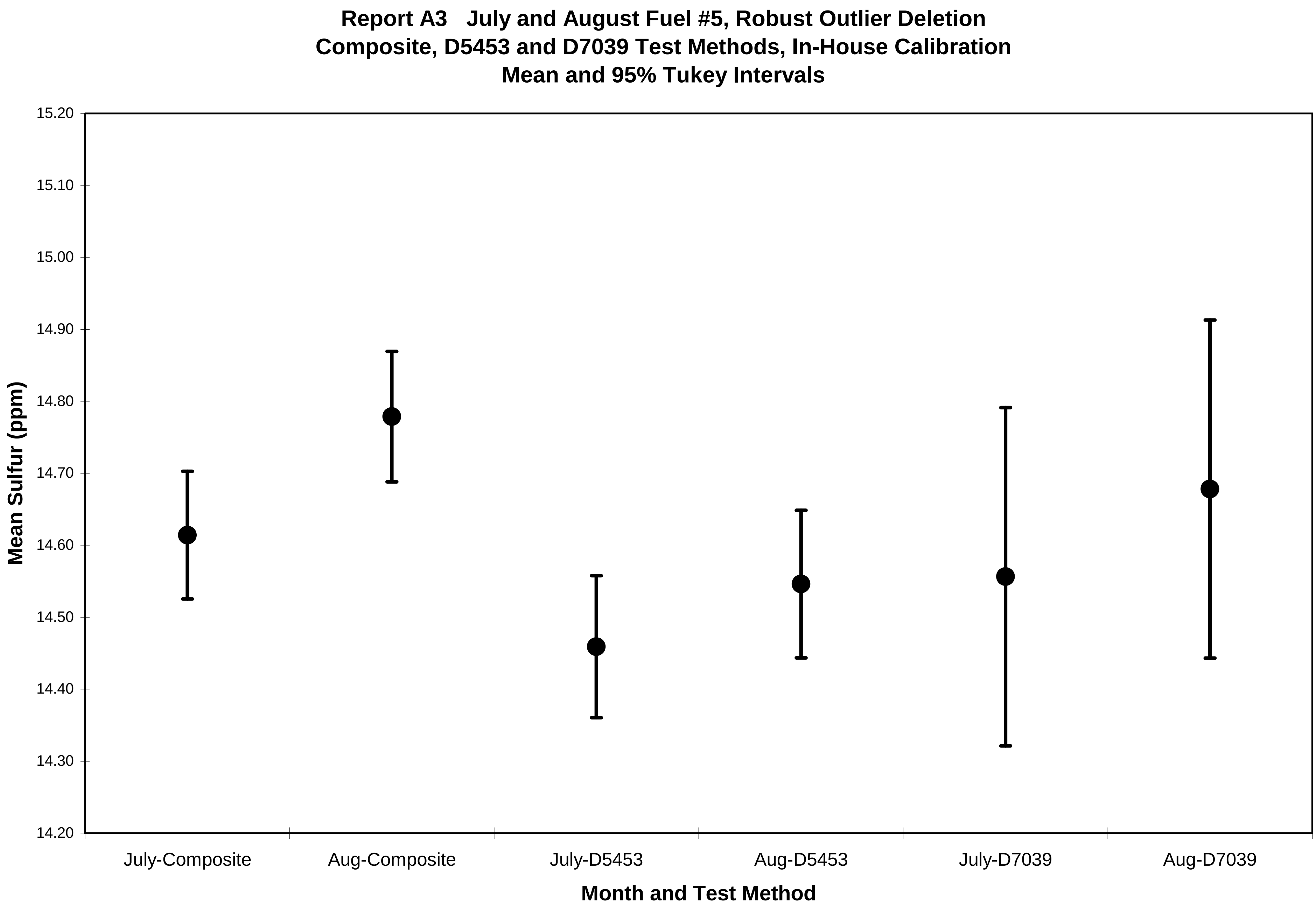
# *Comparisons of Means After Outlier and Repeat Deletions*

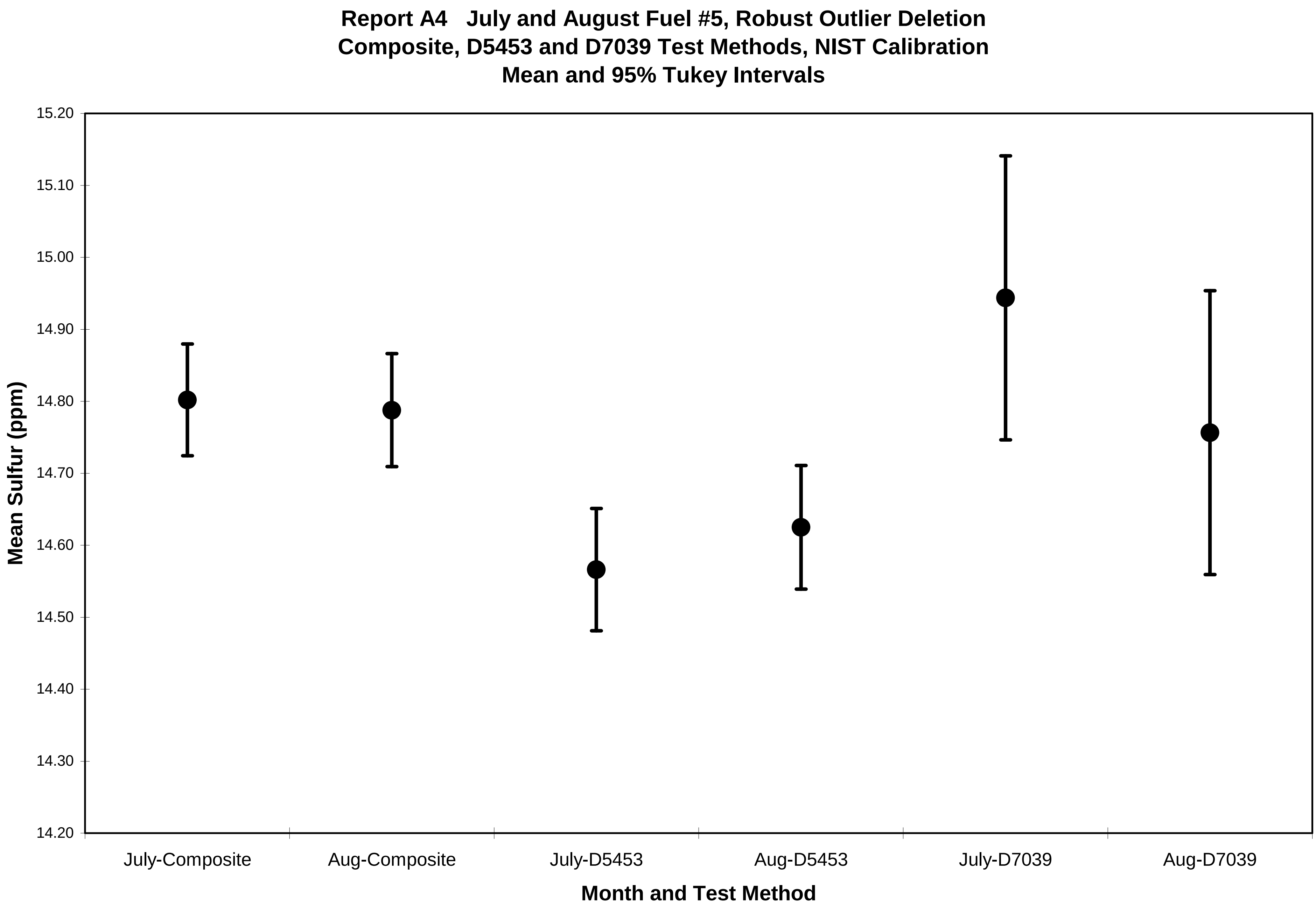
- Compares means of sulfur data remaining after deleting outliers and extra repeats.
  - July and August fuel #4 data, and July and August fuel #5 data are compared.
  - July fuel #4 matches with August fuel #4 and has a sulfur target value of 8.41 ppm.
  - July fuel #5 matches with August fuel #5 and has a sulfur target value of 15 ppm.
  - Composite, D5453 and D7039 test methods are used.
  - In-House and NIST calibrations are included.
  - Robust outlier deletion and extra repeat deletion is used.

- Means and 95% Tukey intervals are plotted.
  - Compares sulfur means for July and August for the same fuel and the same test method.
  - Intervals are constructed in such a way that two overlapping intervals indicate no significant difference between the two means being compared.


Report A1 July and August Fuel #4, Robust Outlier Deletion
Composite, D5453 and D7039 Test Methods, In-House Calibration
Mean and 95% Tukey Intervals
Mean Sulfur (ppm)
8.80
8.70
8.60
8.50
8.40
8.30
8.20
8.10
8.00
July-Composite
Aug-Composite
July-D5453
Aug-D5453
July-D7039
Aug-D7039
Month and Test Method


Report A2 July and August Fuel #4, Robust Outlier Deletion
Composite, D5453 and D7039 Test Methods, NIST Calibration
Mean and 95% Tukey Intervals
Mean Sulfur (ppm)
8.80
8.70
8.60
8.50
8.40
8.30
8.20
8.10
8.00
July-Composite
Aug-Composite
July-D5453
Aug-D5453
July-D7039
Aug-D7039
Month and Test Method


Report A3 July and August Fuel #5, Robust Outlier Deletion
Composite, D5453 and D7039 Test Methods, In-House Calibration
Mean and 95% Tukey Intervals
15.20
15.10
15.00
14.90
14.80
14.70
14.60
14.50
14.40
14.30
14.20
Mean Sulfur (ppm)
July-Composite
Aug-Composite
July-D5453
Aug-D5453
July-D7039
Aug-D7039
Month and Test Method

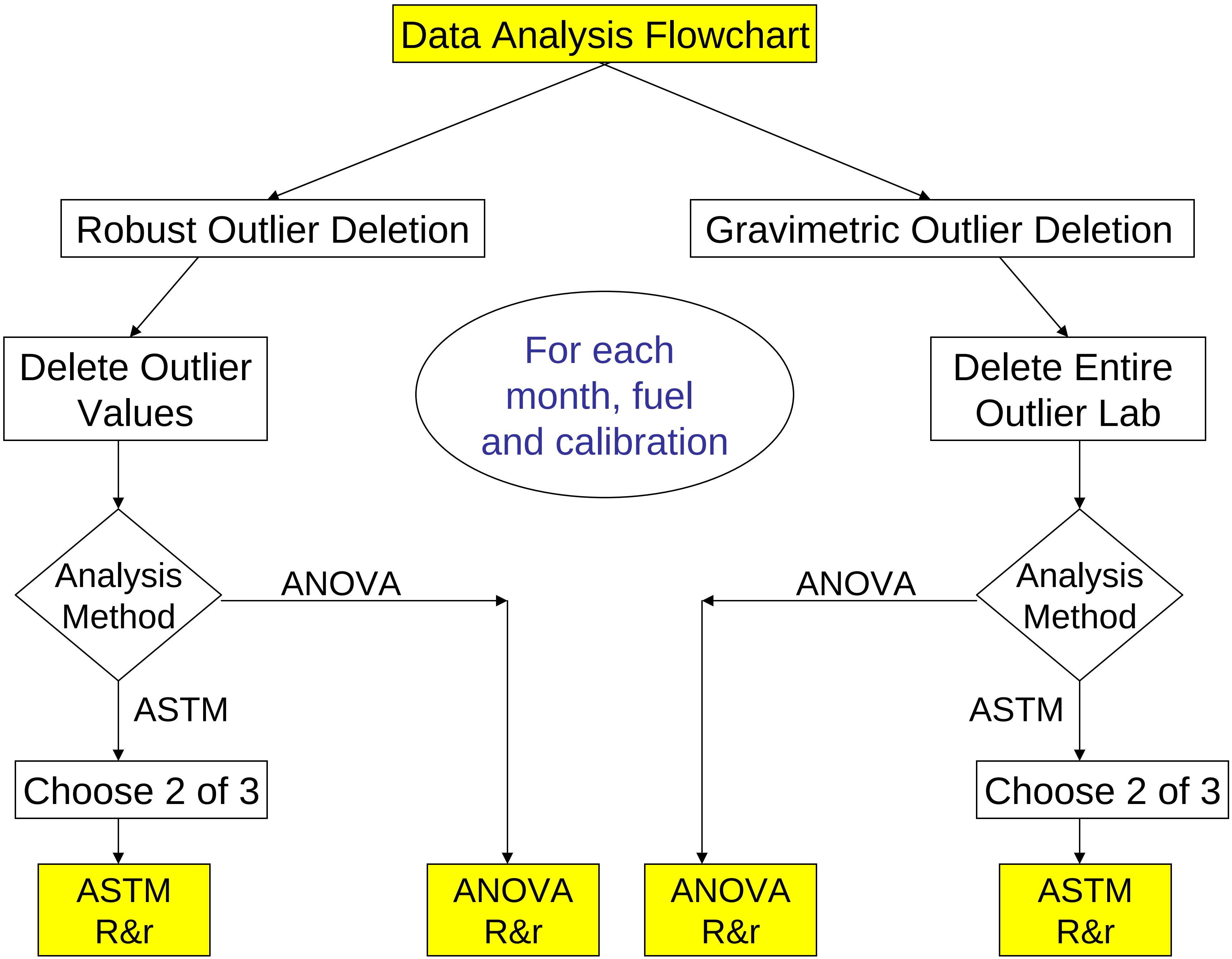

Report A4 July and August Fuel #5, Robust Outlier Deletion
Composite, D5453 and D7039 Test Methods, NIST Calibration
Mean and 95% Tukey Intervals
Mean Sulfur (ppm)
15.20
15.10
15.00
14.90
14.80
14.70
14.60
14.50
14.40
14.30
14.20
July-Composite
Aug-Composite
July-D5453
Aug-D5453
July-D7039
Aug-D7039
Month and Test Method

# *Summary of Mean Comparisons After Outlier and Repeat Deletions*

- The means of the sulfur data, after all deletions, for the July fuel #4 and the August fuel #4 were not significantly different using a 95% confidence interval, either calibration method, the robust deletion method, and the extra repeat deletion.

- Same result holds for the mean comparisons for the July fuel #5 and August fuel #5 data.

- This was demonstrated separately for the composite, and for the D5453 and D7039 test methods.


Data Analysis Flowchart
Robust Outlier Deletion
Gravimetric Outlier Deletion
Delete Outlier Values
For each month, fuel and calibration
Delete Entire Outlier Lab
Analysis Method
ANOVA
ANOVA
Analysis Method
ASTM
ASTM
Choose 2 of 3
Choose 2 of 3
ASTM R&r
ANOVA R&r
ANOVA R&r
ASTM R&r

# *Lab ASTM Reproducibility Comparisons*

- Compares ASTM lab reproducibility values for the following conditions:
  - July fuel #4 and August fuel #5 data
  - Composite test method
  - In-House and NIST calibrations
  - Robust outlier deletion method

- Mean of lab sulfur repeat values (with range) versus lab codes are plotted.
  - Includes limits at (robust mean) +/- 3 *(robust standard deviation)
  - Compares the deviation of the lab mean from the robust mean to the upper and lower limits.
  - Values that exceed limits are noted.
  - Codes of the deleted labs are listed at the right of the figure title.

**Figure D-25 July Fuel #4, Robust Outlier Deletion**
**Composite Test Methods, In-House Calibration**
**ASTM Analysis, Lab Mean and Range**

Codes of deleted labs = 78

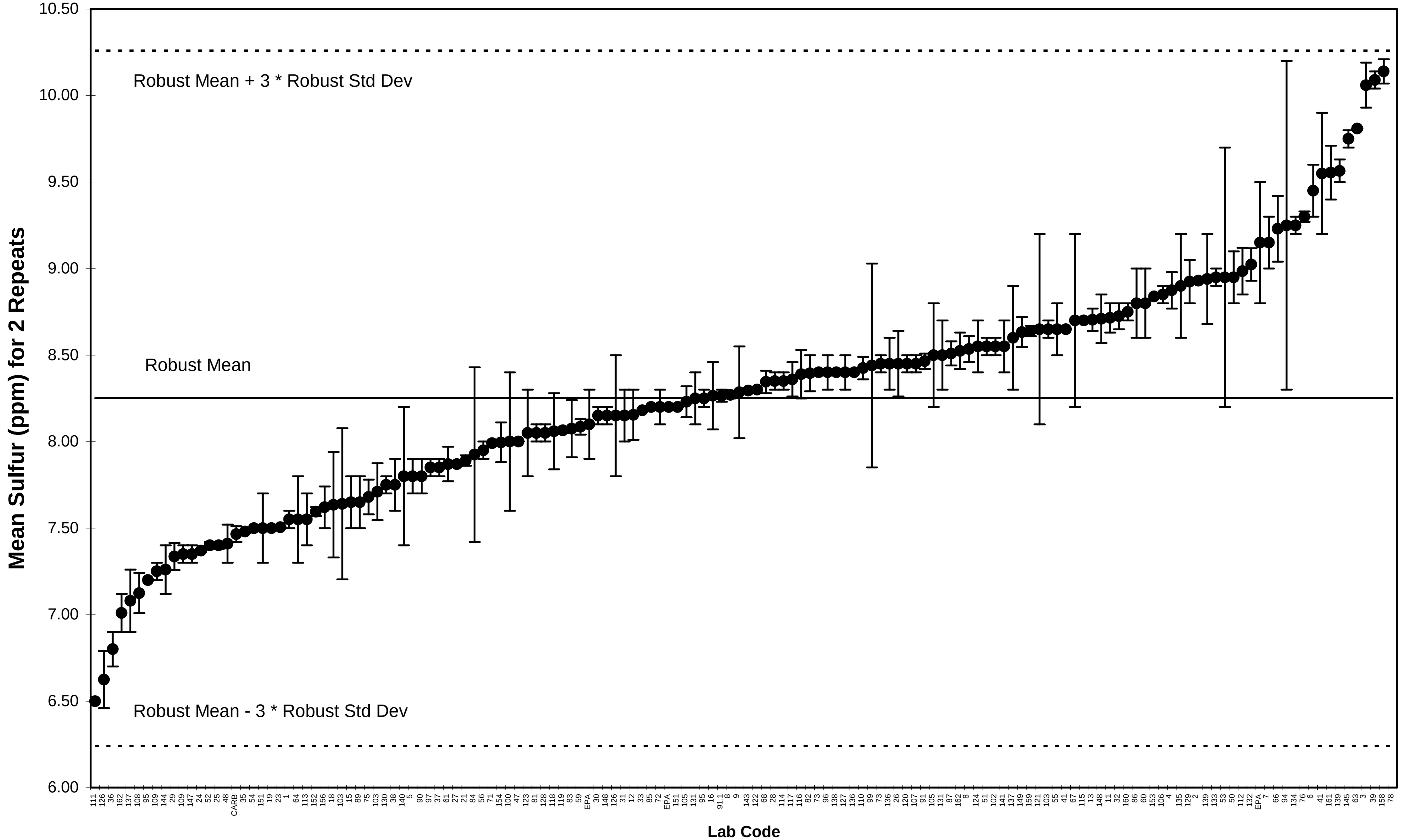


**Figure D-25a July Fuel #4, Robust Outlier Deletion**
**Composite Test Methods, In-House Calibration**
**ASTM Analysis, Lab Mean and Range**

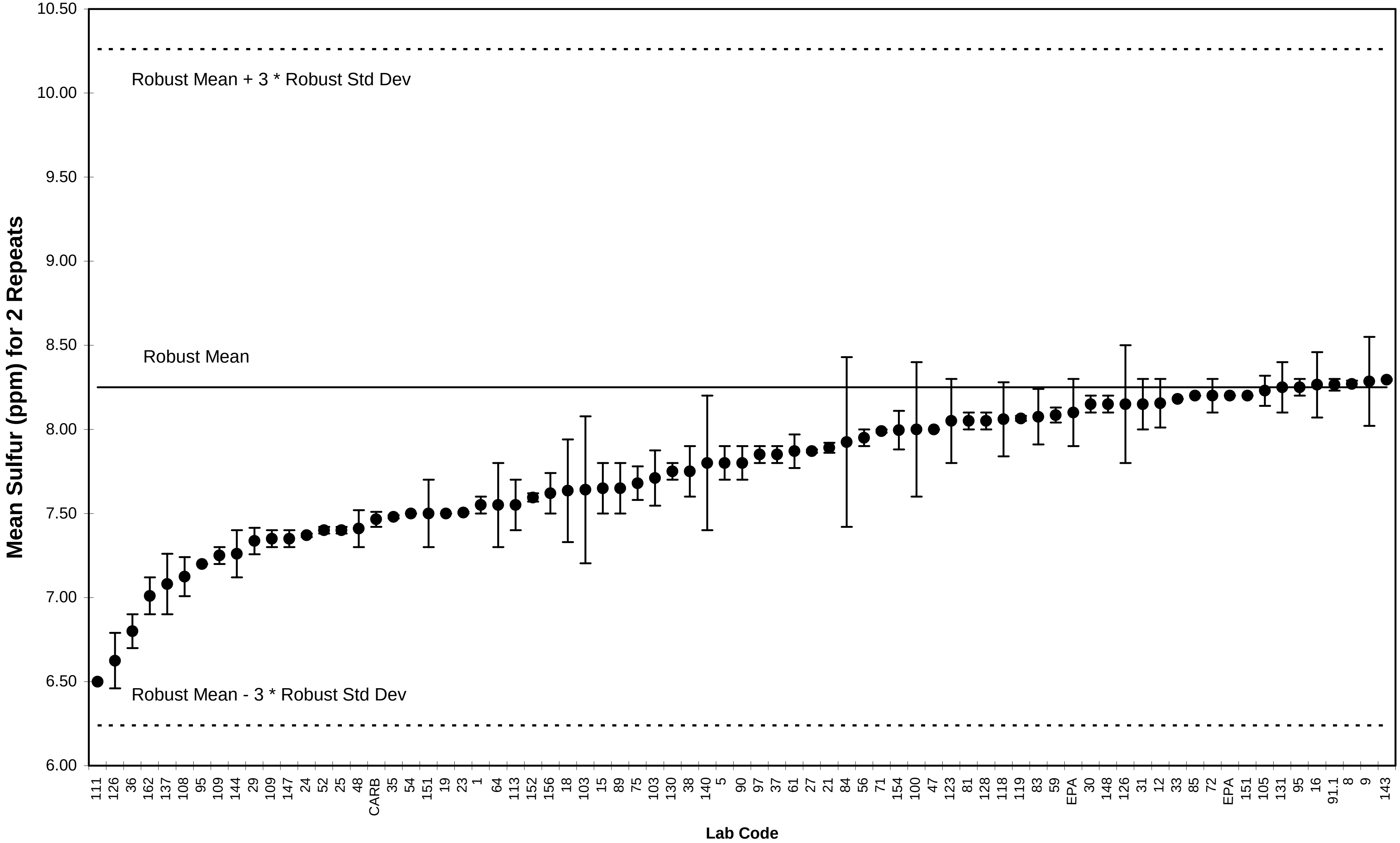


**Figure D-25b July Fuel #4, Robust Outlier Deletion**
**Composite Test Methods, In-House Calibration**
**ASTM Analysis, Lab Mean and Range**

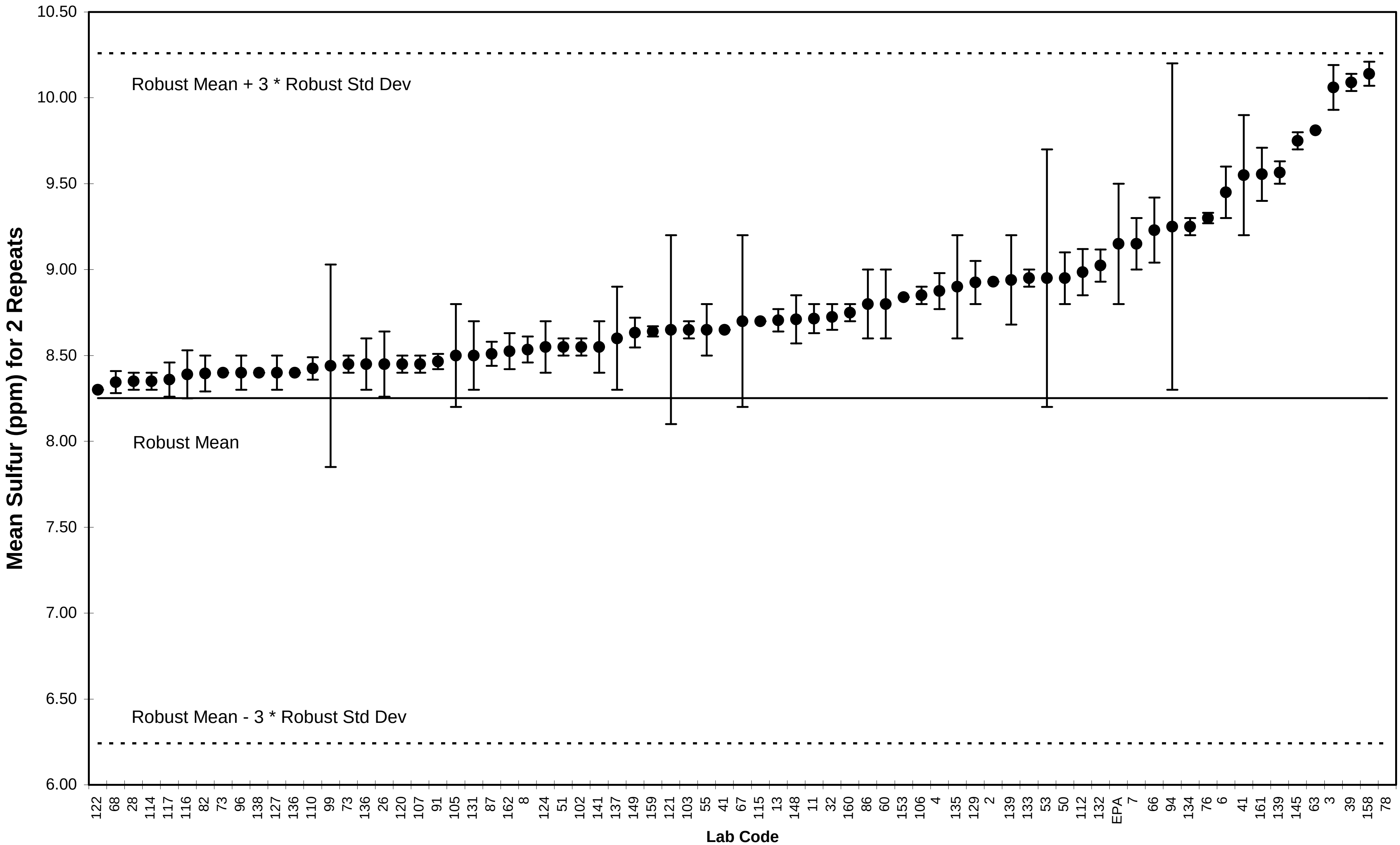


**Figure D-27 July Fuel #4, Robust Outlier Deletion**
**Composite Test Methods, NIST Calibration**
**ASTM Analysis, Lab Mean and Range**

Codes of deleted labs =
33, 78, 84, 94, 100, 103, 139

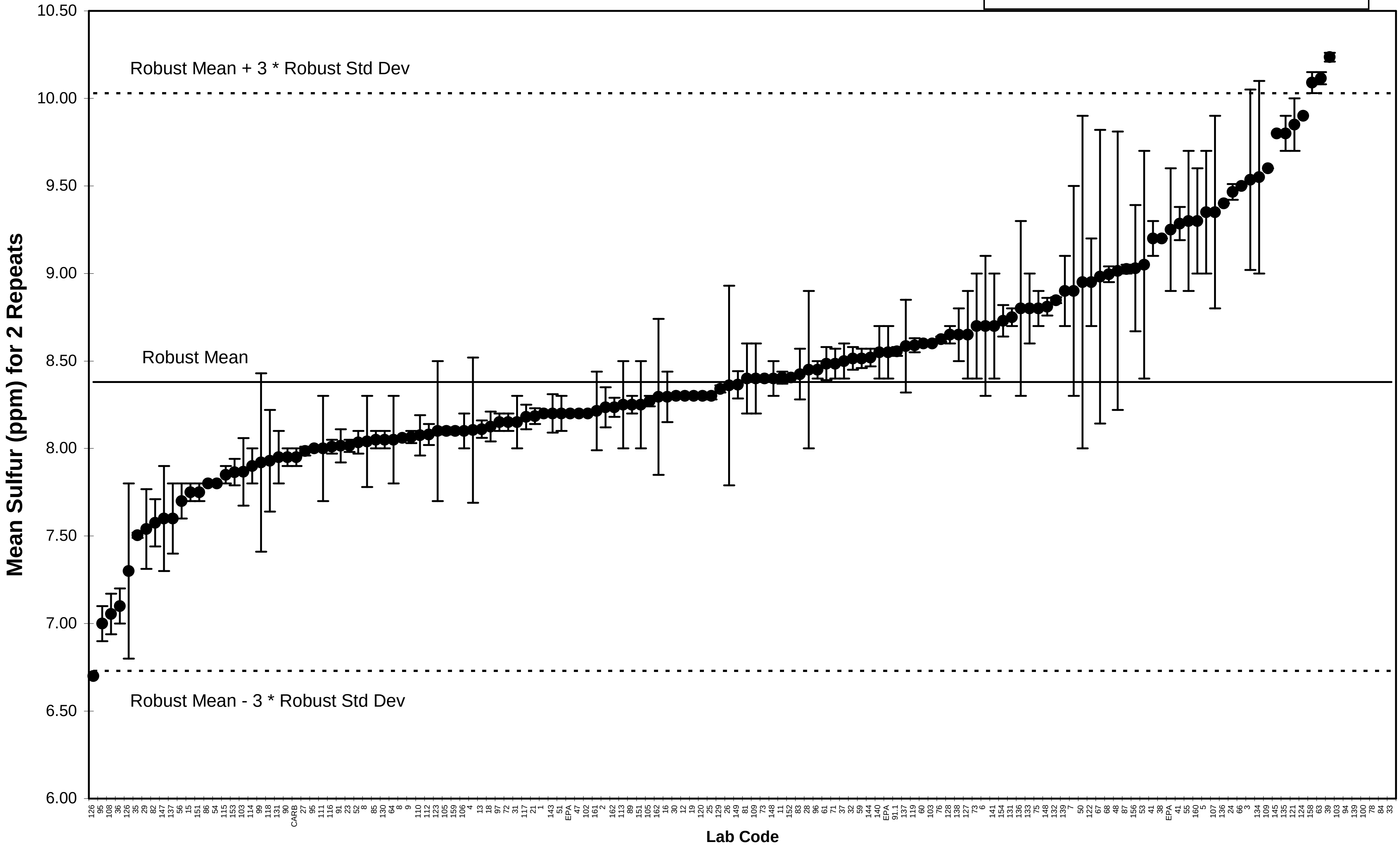
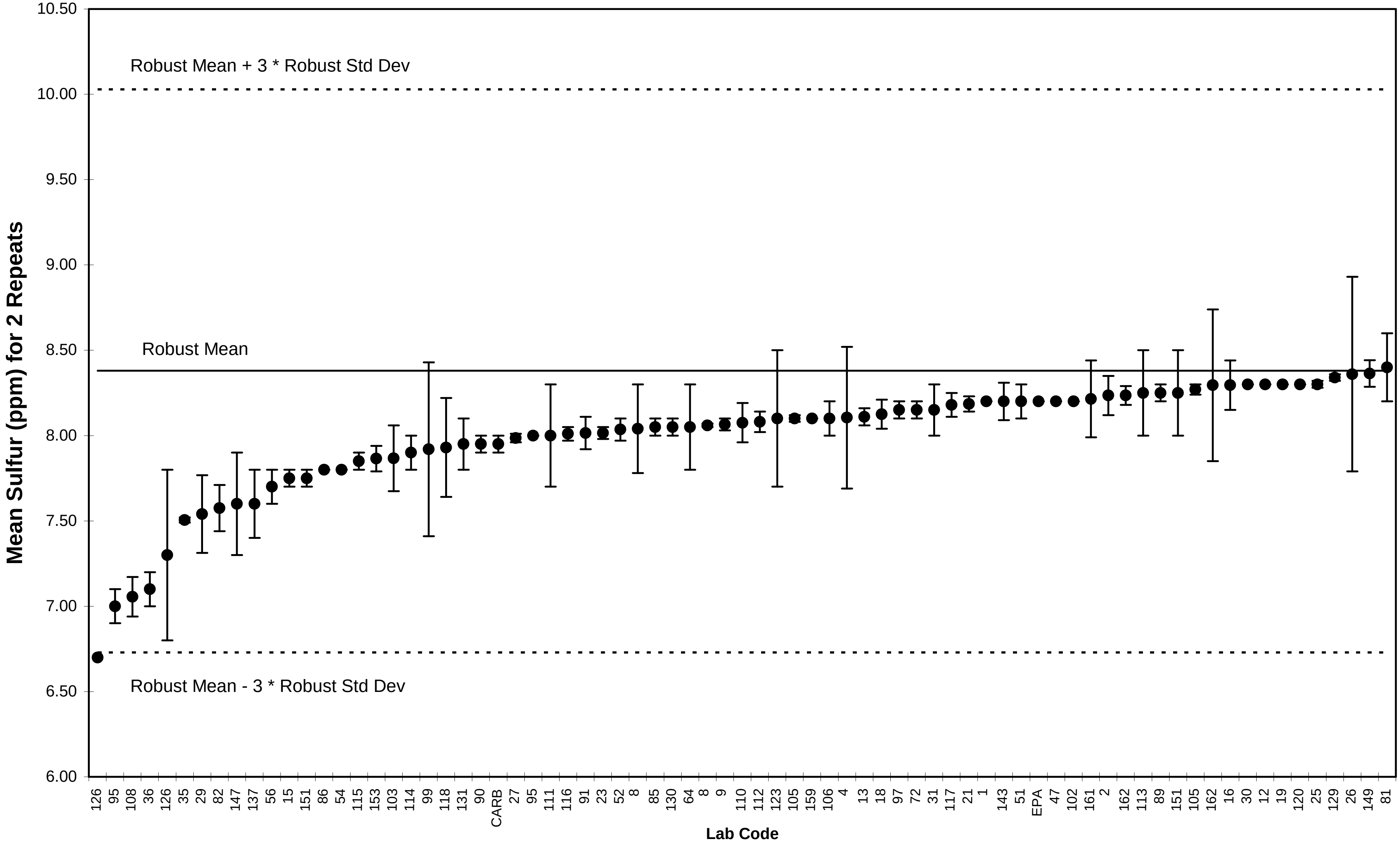


Figure D-27a  July Fuel #4, Robust Outlier Deletion
Composite Test Methods, NIST Calibration
ASTM Analysis, Lab Mean and Range



## Figure D-27b   July Fuel #4, Robust Outlier Deletion
## Composite Test Methods, NIST Calibration
## ASTM Analysis, Lab Mean and Range

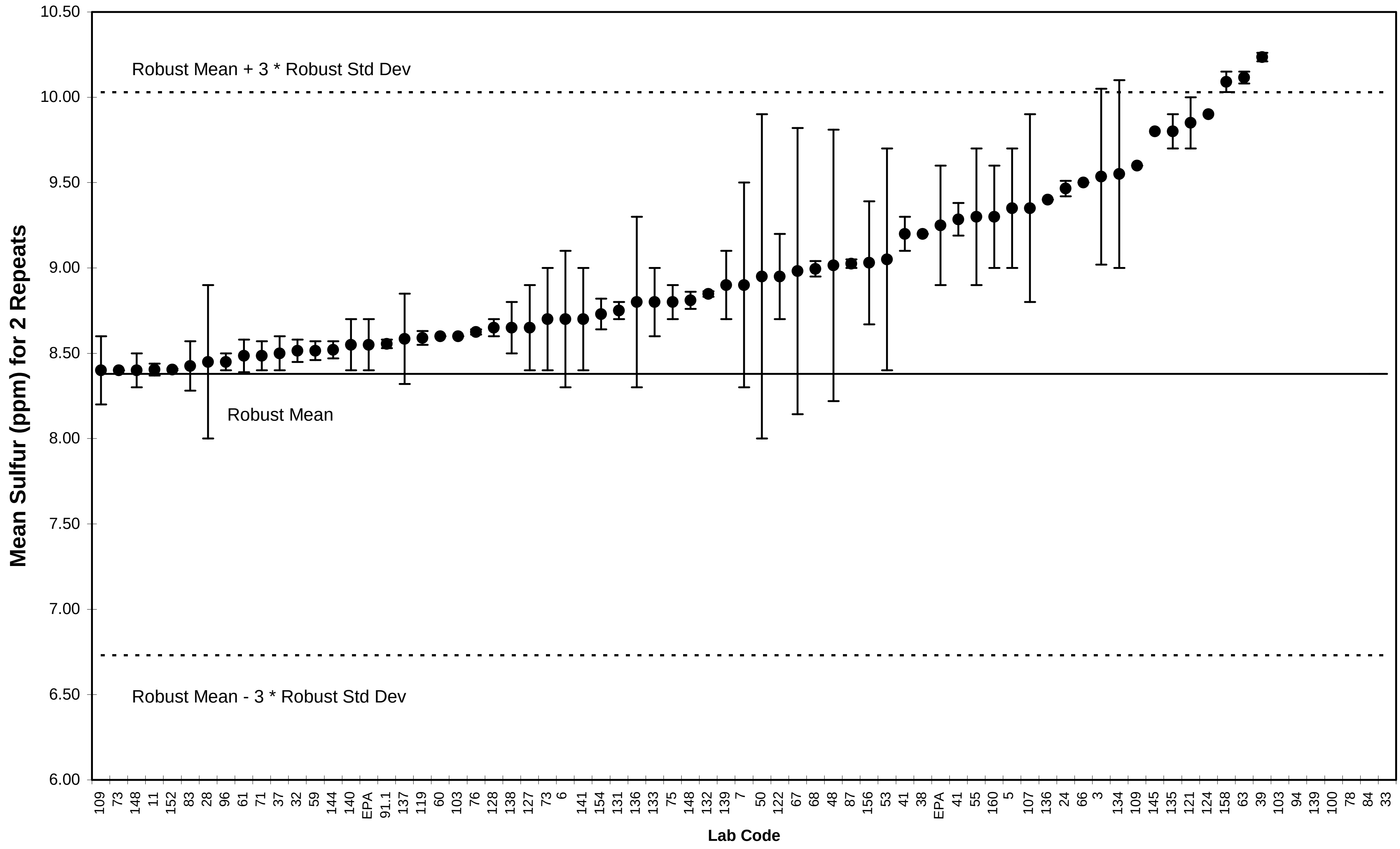


**Figure D-73 August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods, In-House Calibration**
**ASTM Analysis, Lab Mean and Range**

Codes of deleted labs =
68, 140, 145, 158

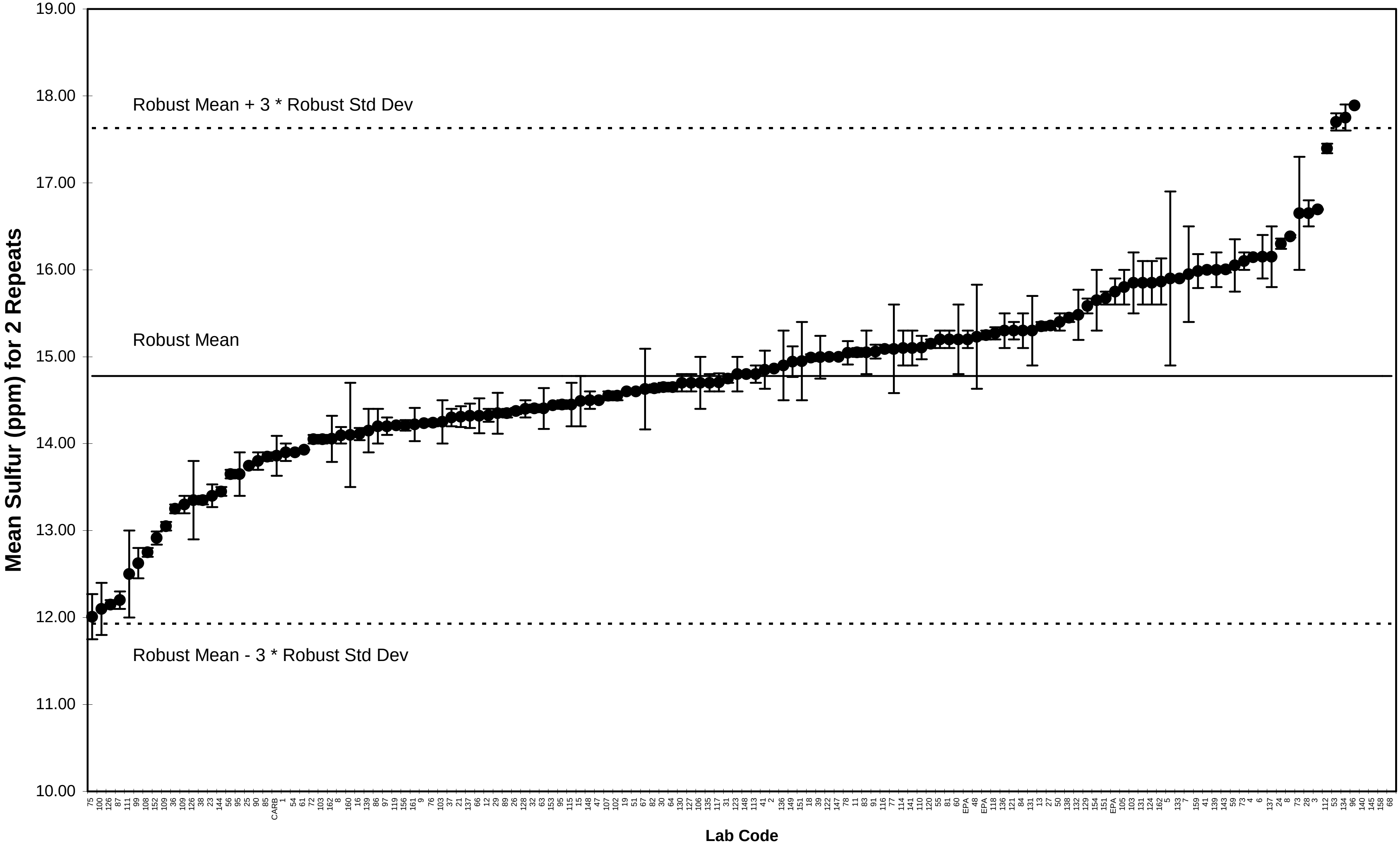


**Figure D-73a   August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods, In-House Calibration**
**ASTM Analysis, Lab Mean and Range**

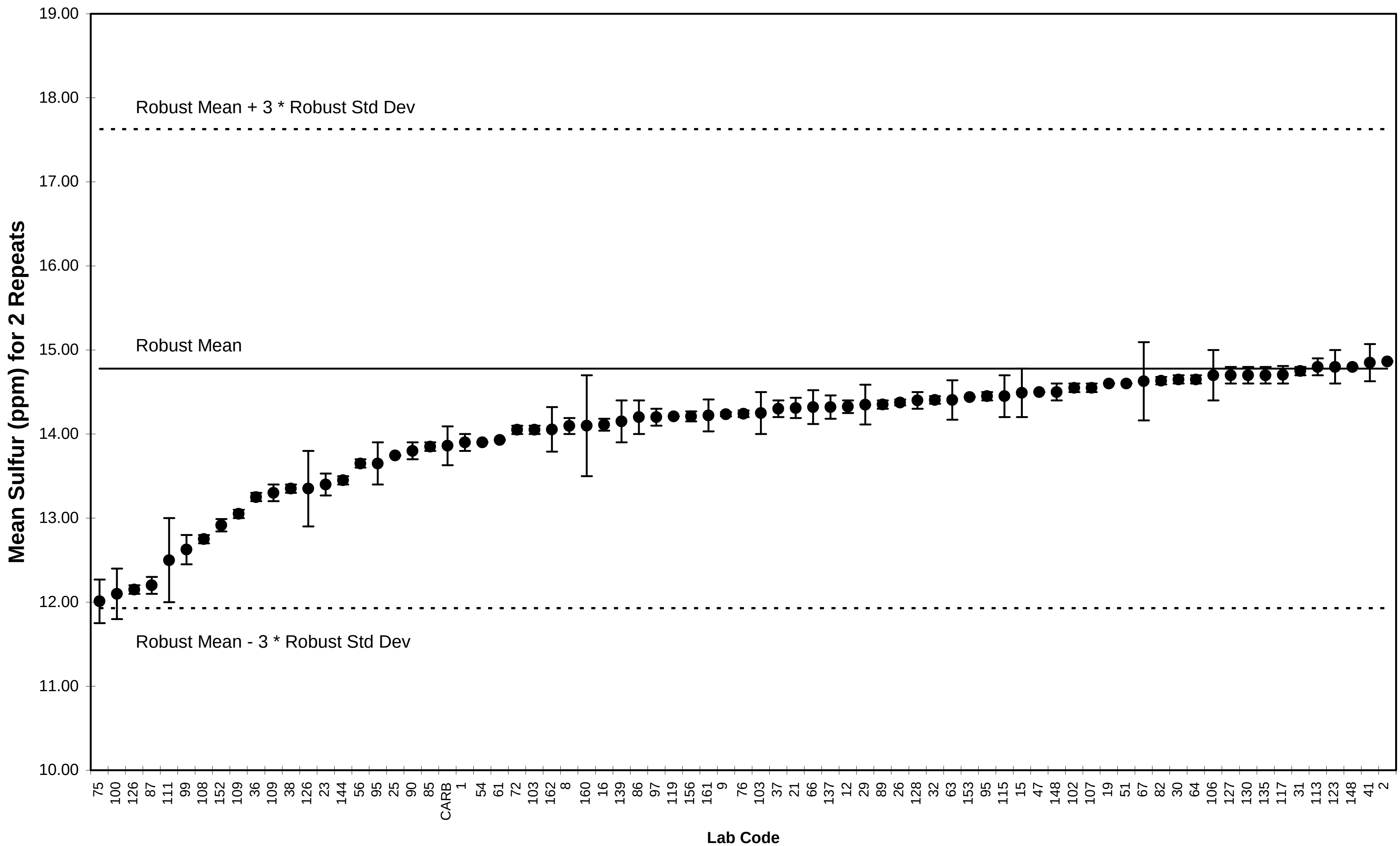


**Figure D-73b August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods, In-House Calibration**
**ASTM Analysis, Lab Mean and Range**

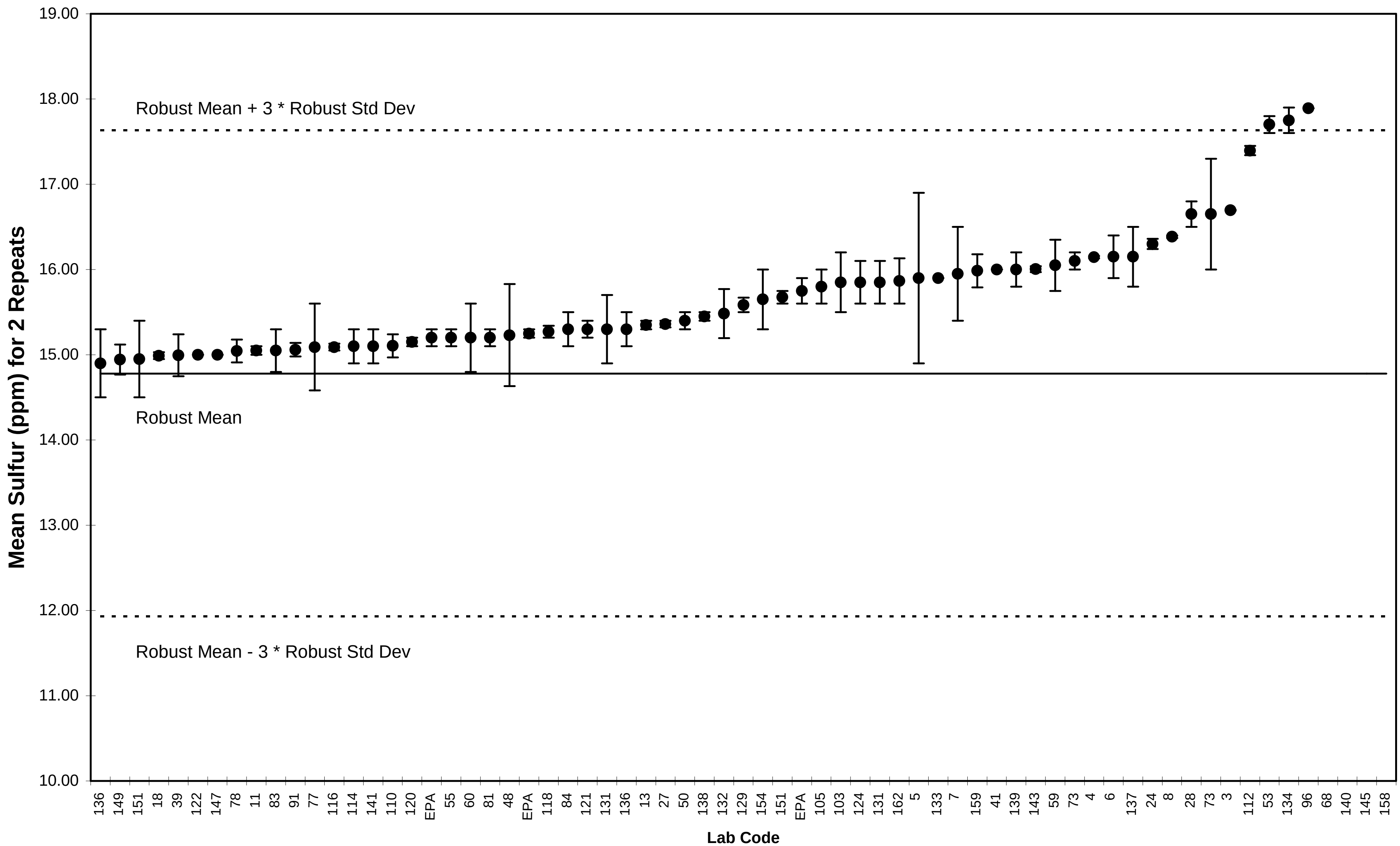


**Figure D-75 August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods, NIST Calibration**
**ASTM Analysis, Lab Mean and Range**

Codes of deleted labs =
51, 68, 87, 96, 145, 158

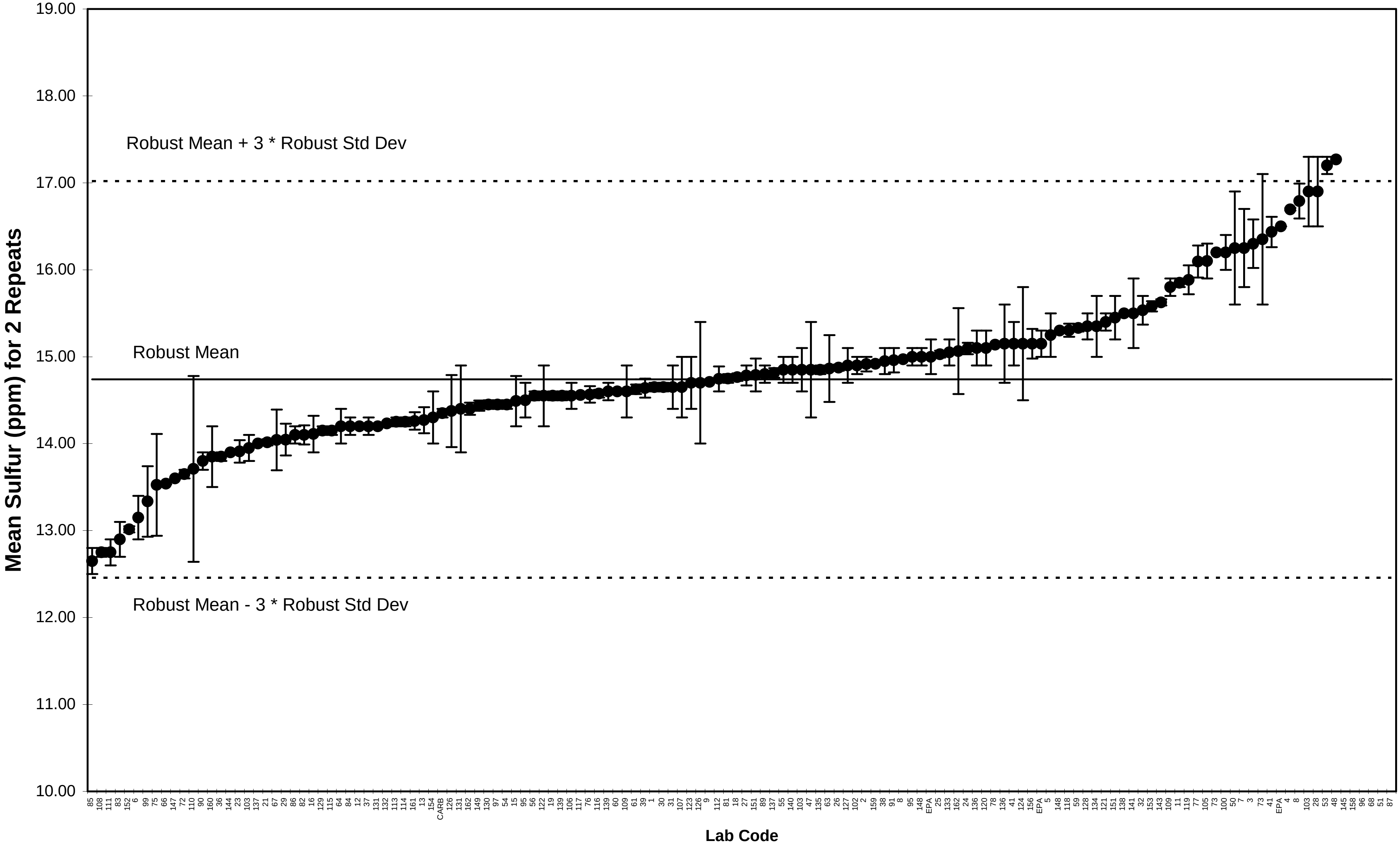


**Figure D-75a   August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods, NIST Calibration**
**ASTM Analysis, Lab Mean and Range**

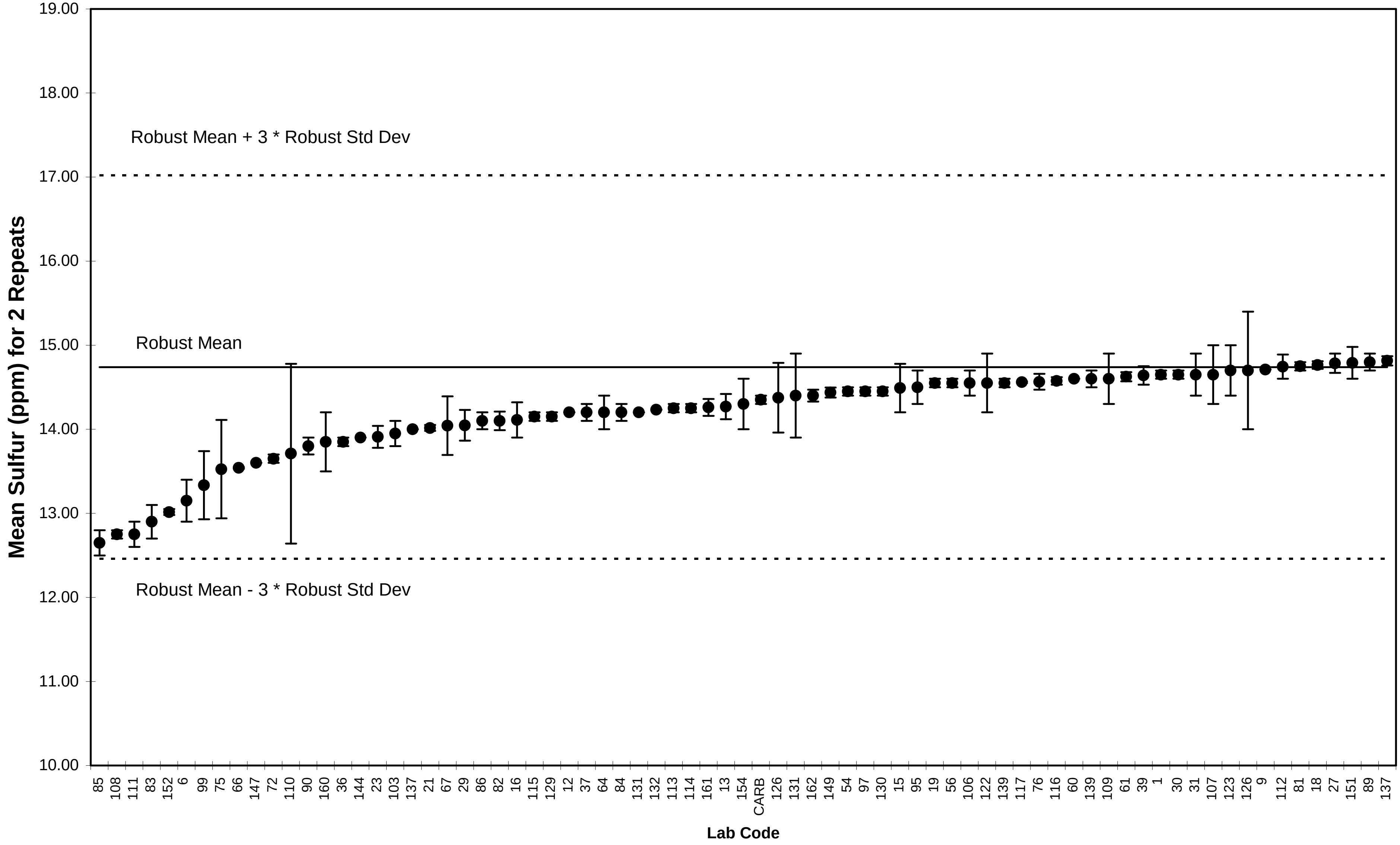
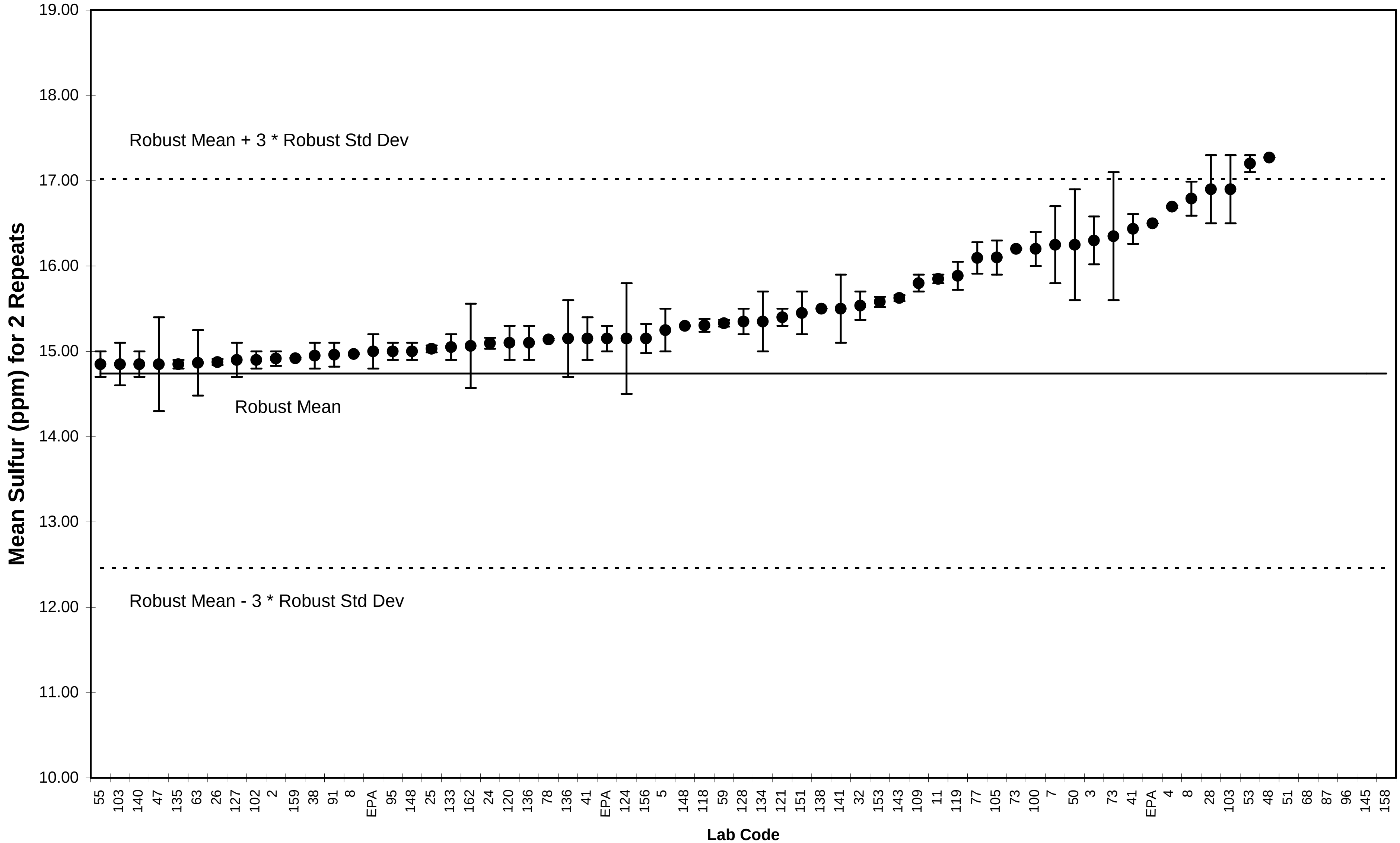


**Figure D-75b August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods, NIST Calibration**
**ASTM Analysis, Lab Mean and Range**



# Summary of Lab ASTM Reproducibility Comparisons

- For the July fuel #4 data, no labs had means of their two sulfur repeat values that exceeded the robust mean limits using the In-House calibration. This compares to 4 labs exceeding the limits when using the NIST calibration.
  - The July fuel #4 data for 1 lab using the In-House method and 7 labs using the NIST method were excluded due to having all 3 repeat values meet the outlier deletion criteria.

- For the August fuel #5 data, 3 labs had means of their two sulfur repeat values that exceeded the robust mean limits using the In-House calibration. This compares to 2 labs exceeding the limits when using the NIST calibration.
  - The August fuel #5 data for 4 labs using the In-House method and 6 labs using the NIST method were excluded due to having all 3 repeat values meet the outlier deletion criteria.

# *Overall Reproducibility Comparisons: Robust vs. Gravimetric Deletion*

- Compares R-values for robust outlier deletion method and gravimetric outlier deletion method for all ten fuels.
  - For D5453, D7039, D2622, EDXRF and Composite.
  - For NIST calibration only
- Computes R-values using ASTM analysis.

**Figure D-15   D5453 Test Method, Robust and Gravimetric Outlier, ASTM Analysis NIST Calibration**

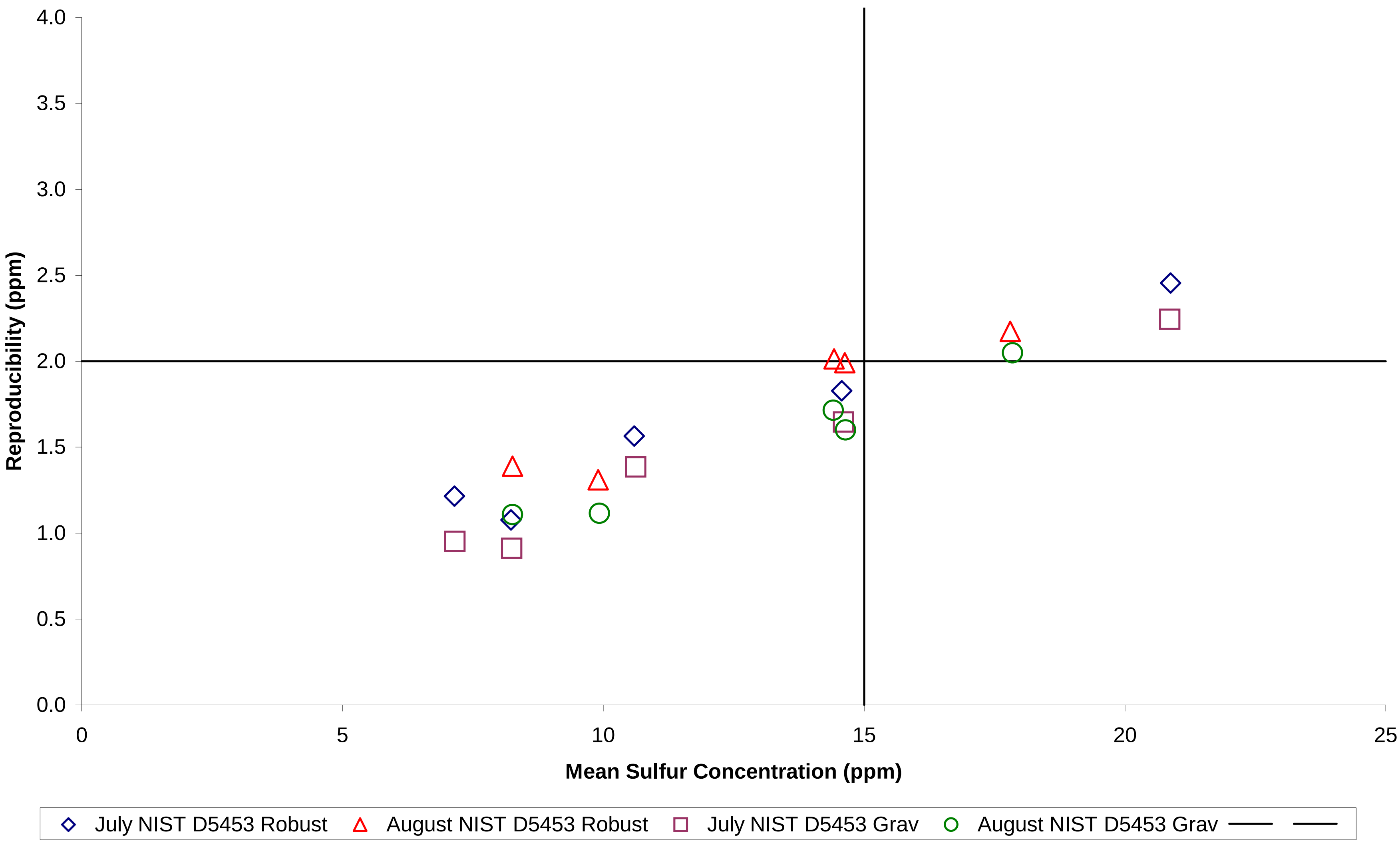
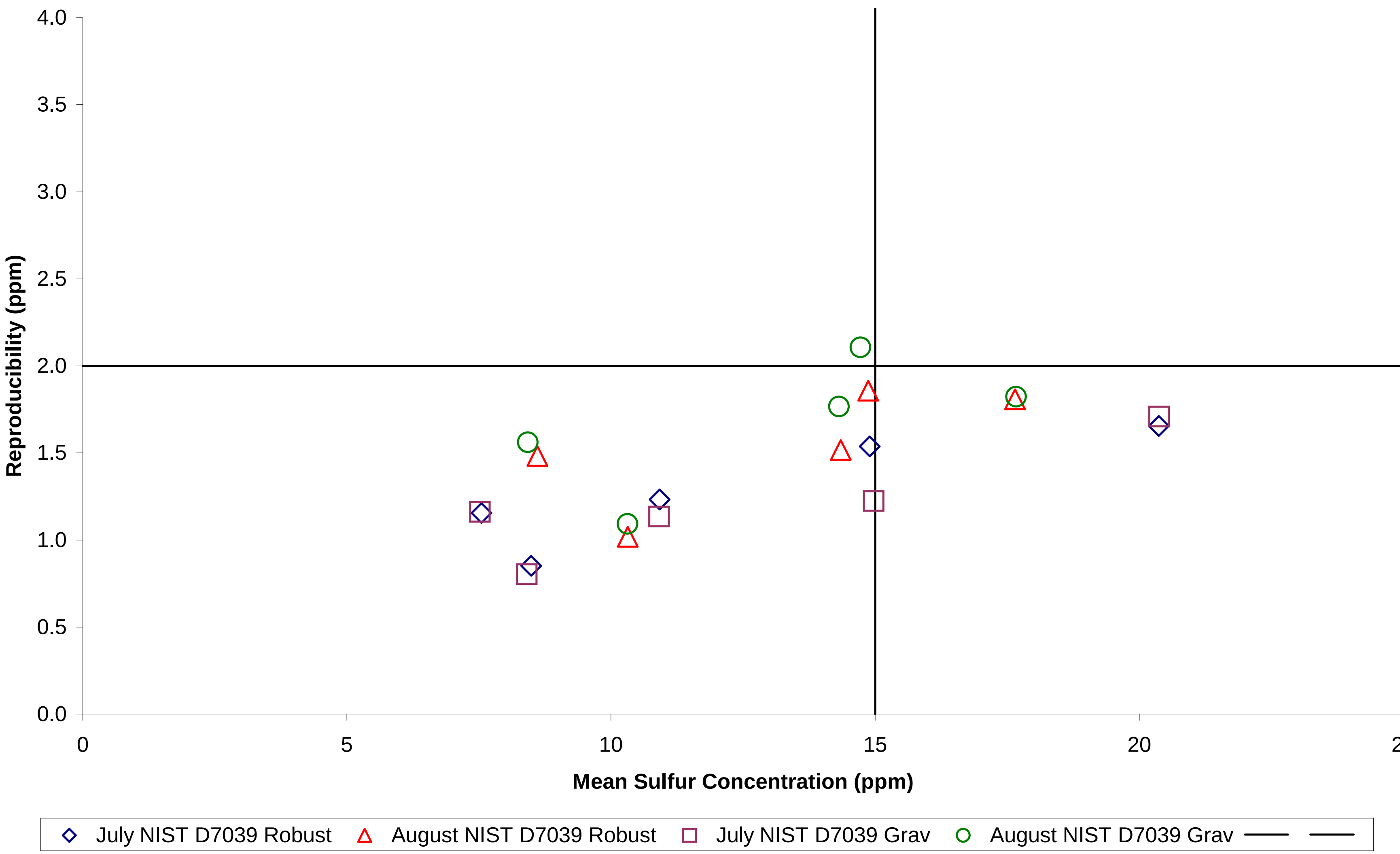


**Figure D-39 D7039 Test Method, Robust and Gravimetric Outlier, ASTM Analysis NIST Calibration**



## Figure D-27 D2622 Test Method, Robust and Gravimetric Outlier, ASTM Analysis NIST Calibration

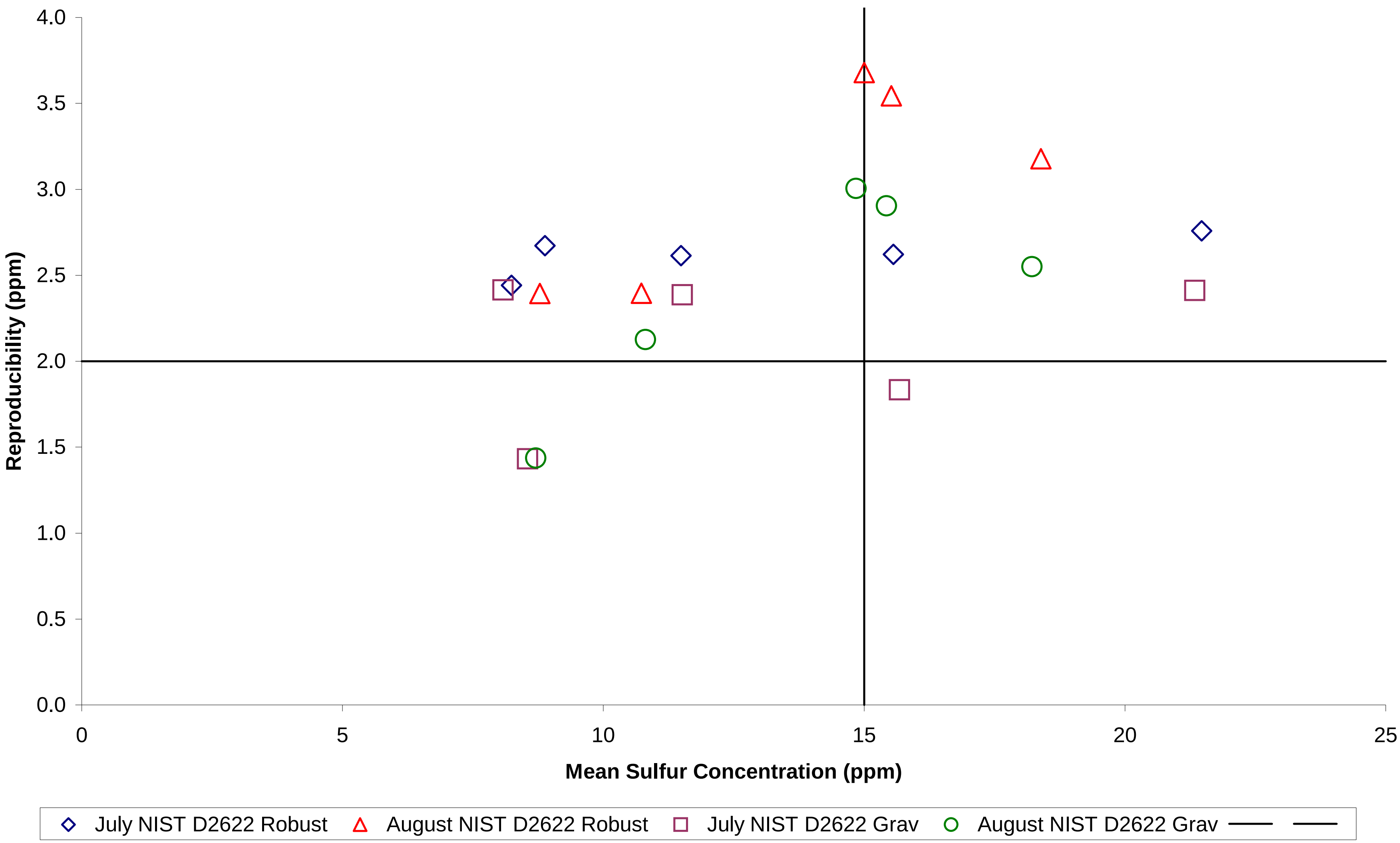


Figure D-51 EDXRF Test Method, Robust and Gravimetric Outlier, ASTM Analysis
NIST Calibration

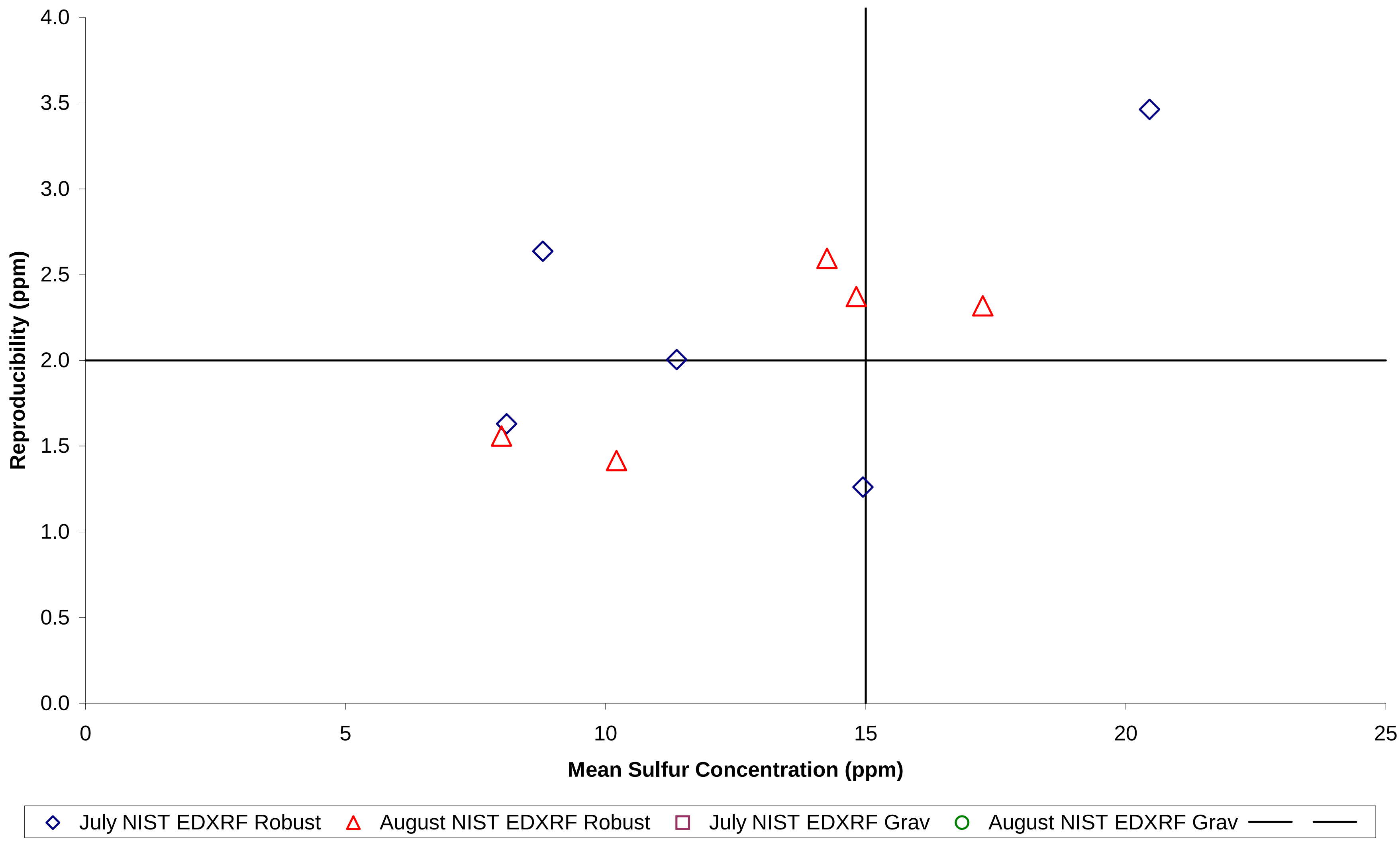
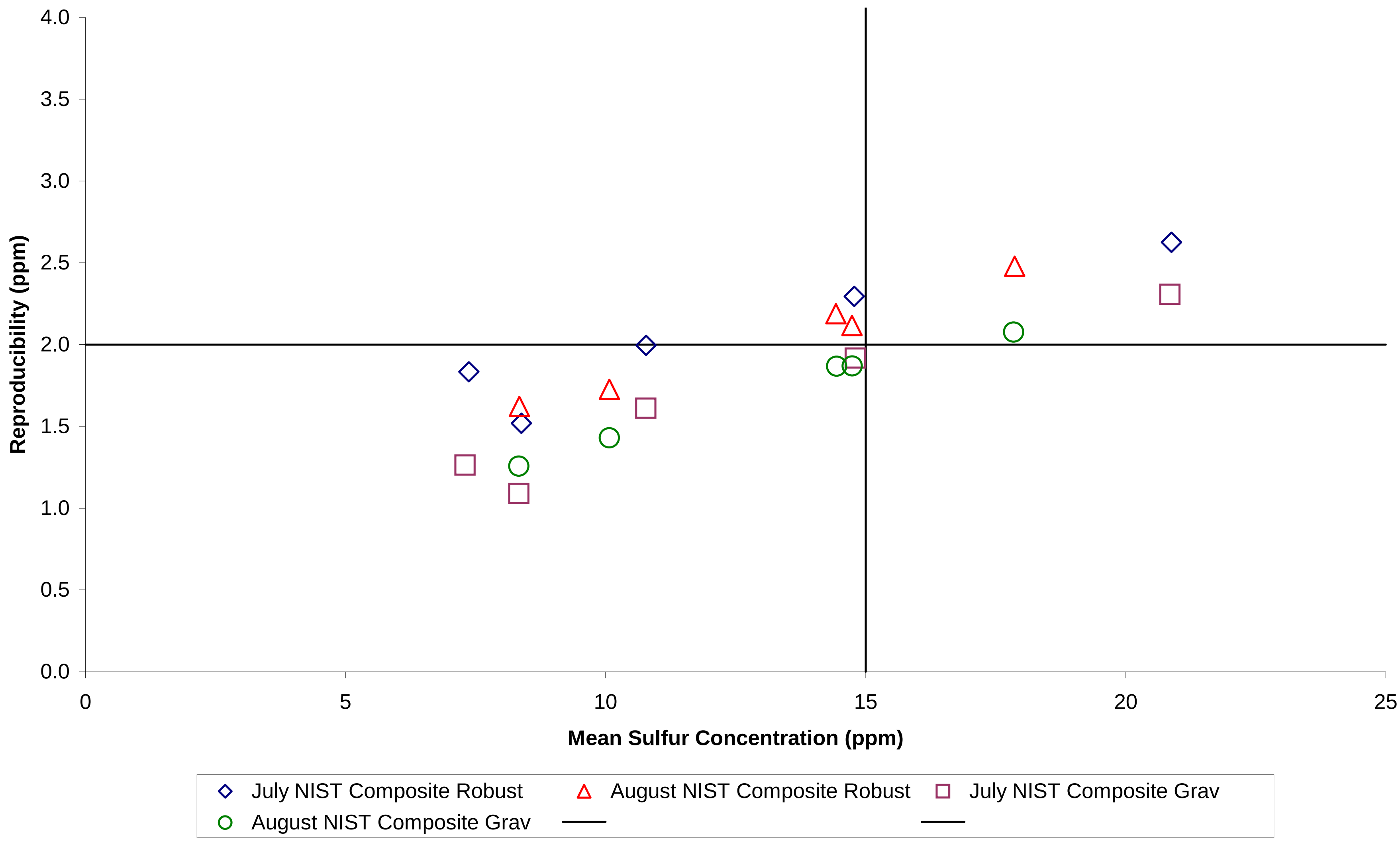


**Figure D-3 Composite Test Methods, Robust and Gravimetric Outlier, ASTM Analysis NIST Calibration**



# *Summary of R Comparisons: Robust vs. Gravimetric Deletion*

- For both D5453 and D7039, the R-values for gravimetric deletion are smaller than those for robust deletion with the NIST calibration.
- The R-values for D2622 and EDXRF have more scatter and no clear trends.

# *Overall Reproducibility Comparisons: ASTM vs. ANOVA Using NIST*

- Compares R-values for ASTM and ANOVA analyses for all ten fuels.
  - For D5453, D7039, D2622, EDXRF and Composite.
- Uses robust outlier deletion method and the NIST calibration.

**Figure D-23 D5453 Test Method, Robust Outlier, NIST Calibration Comparing ASTM and ANOVA Analyses**

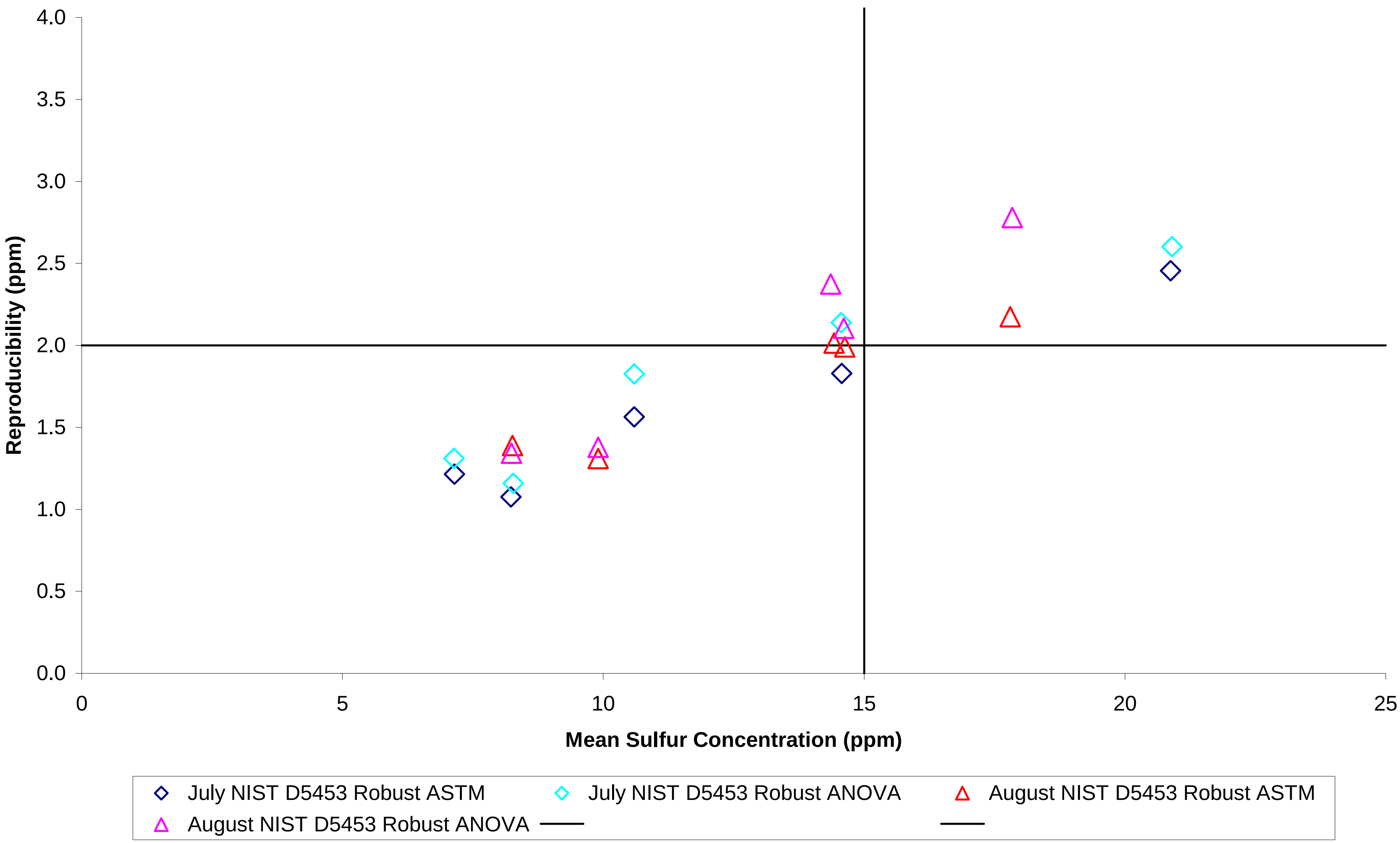


**Figure D-47 D7039 Test Method, Robust Outlier, NIST Calibration Comparing ASTM and ANOVA Analyses**

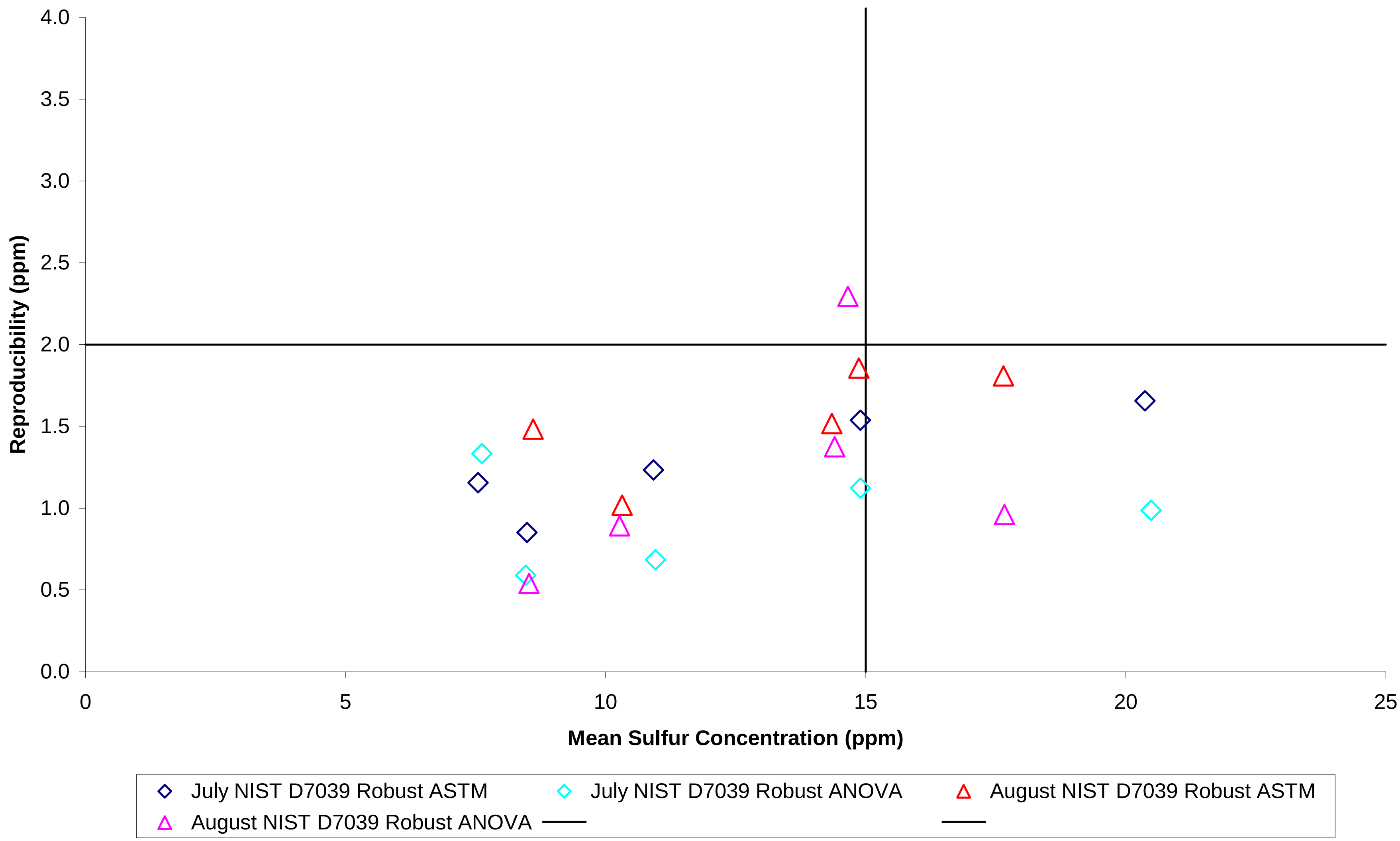


**Figure D-35 D2622 Test Method, Robust Outlier, NIST Calibration Comparing ASTM and ANOVA Analyses**

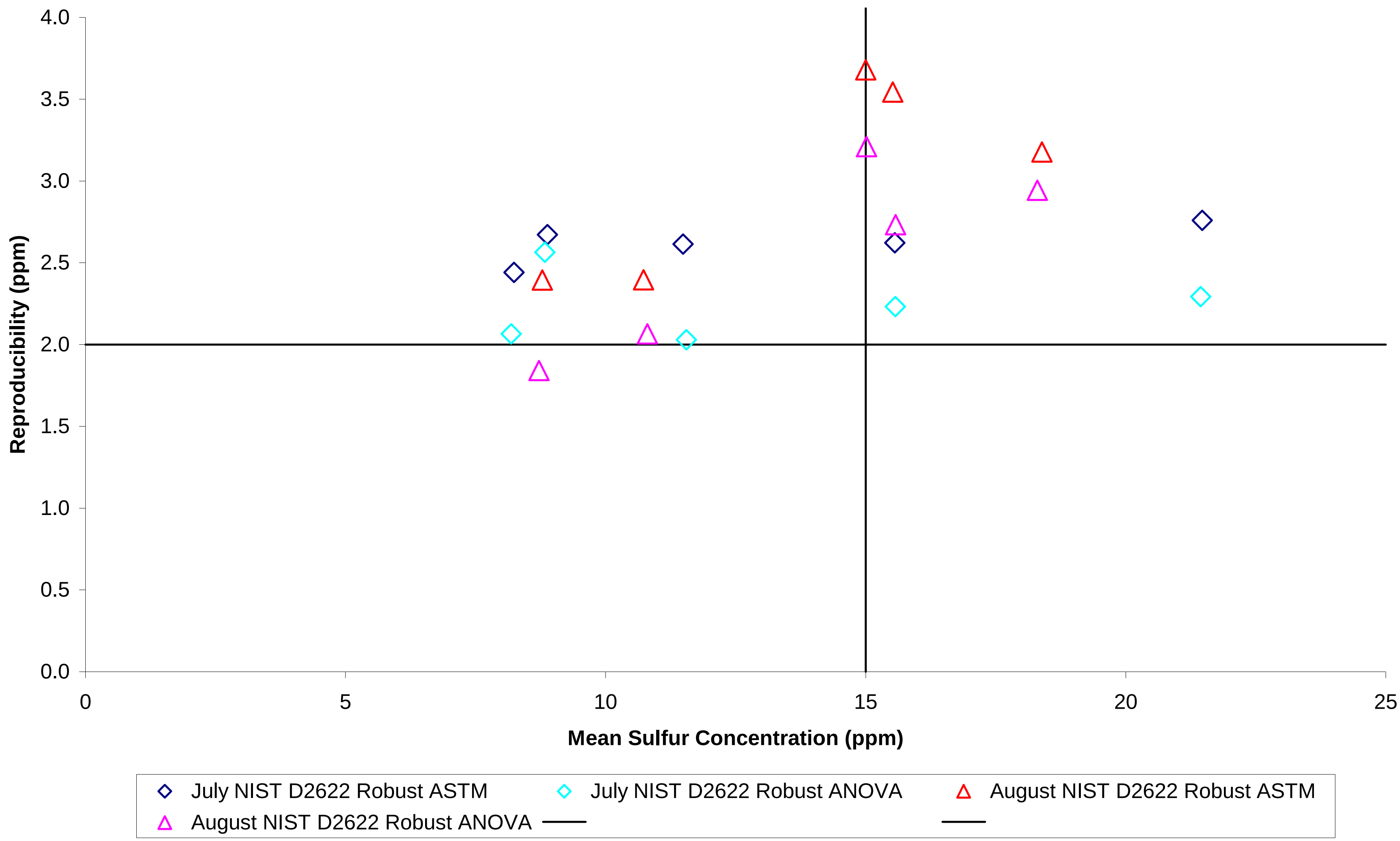


**Figure D-59 EDXRF Test Method, Robust Outlier, NIST Calibration**
**Comparing ASTM and ANOVA Analyses**

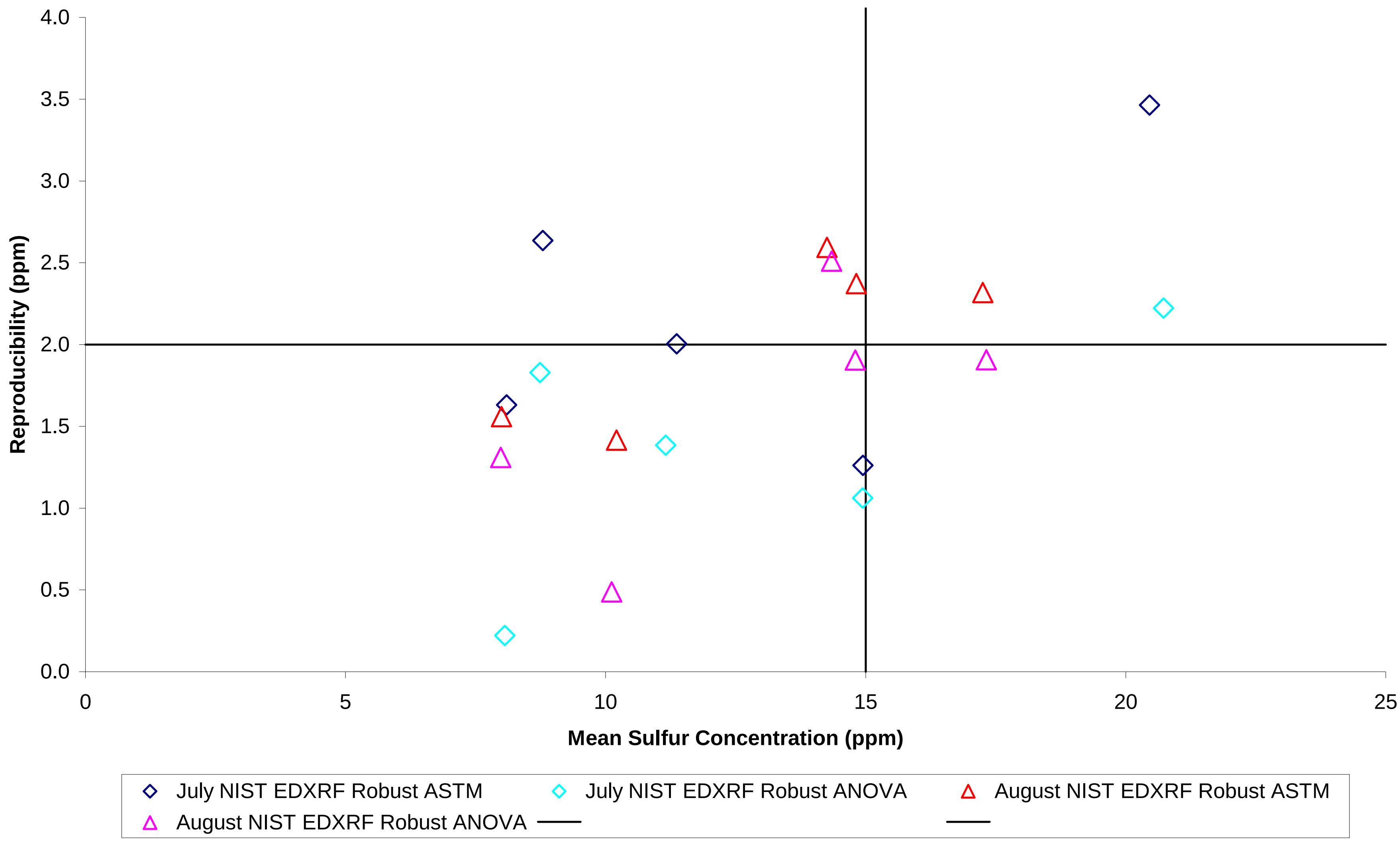
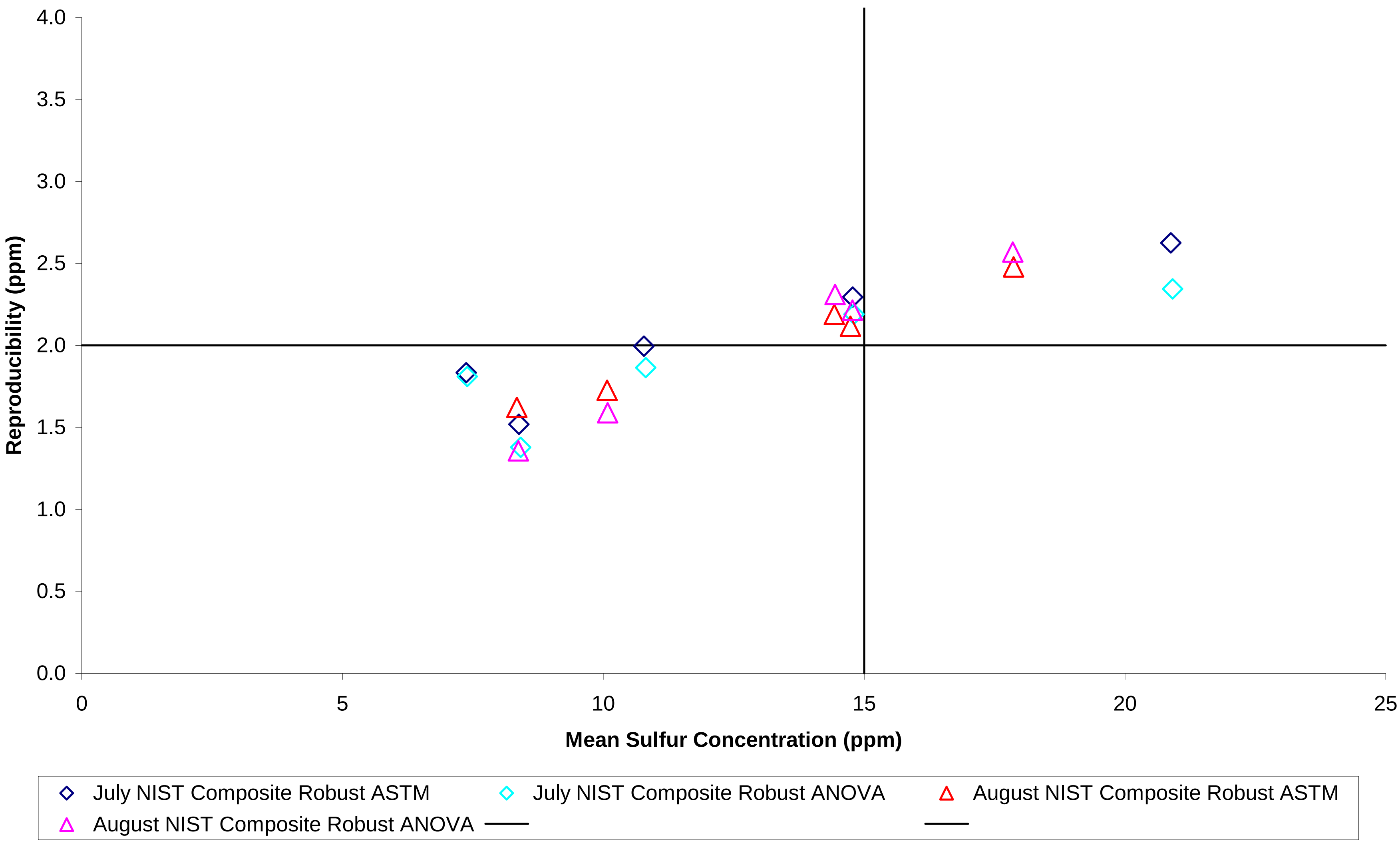


**Figure D-11 Composite Test Methods, Robust Outlier, NIST Calibration Comparing ASTM and ANOVA Analyses**



# *Summary of R Comparisons: ASTM vs. ANOVA Using NIST*

- For D5453, the R-values for the ASTM method are equal to or smaller than the R-values for the ANOVA method with the NIST calibration.

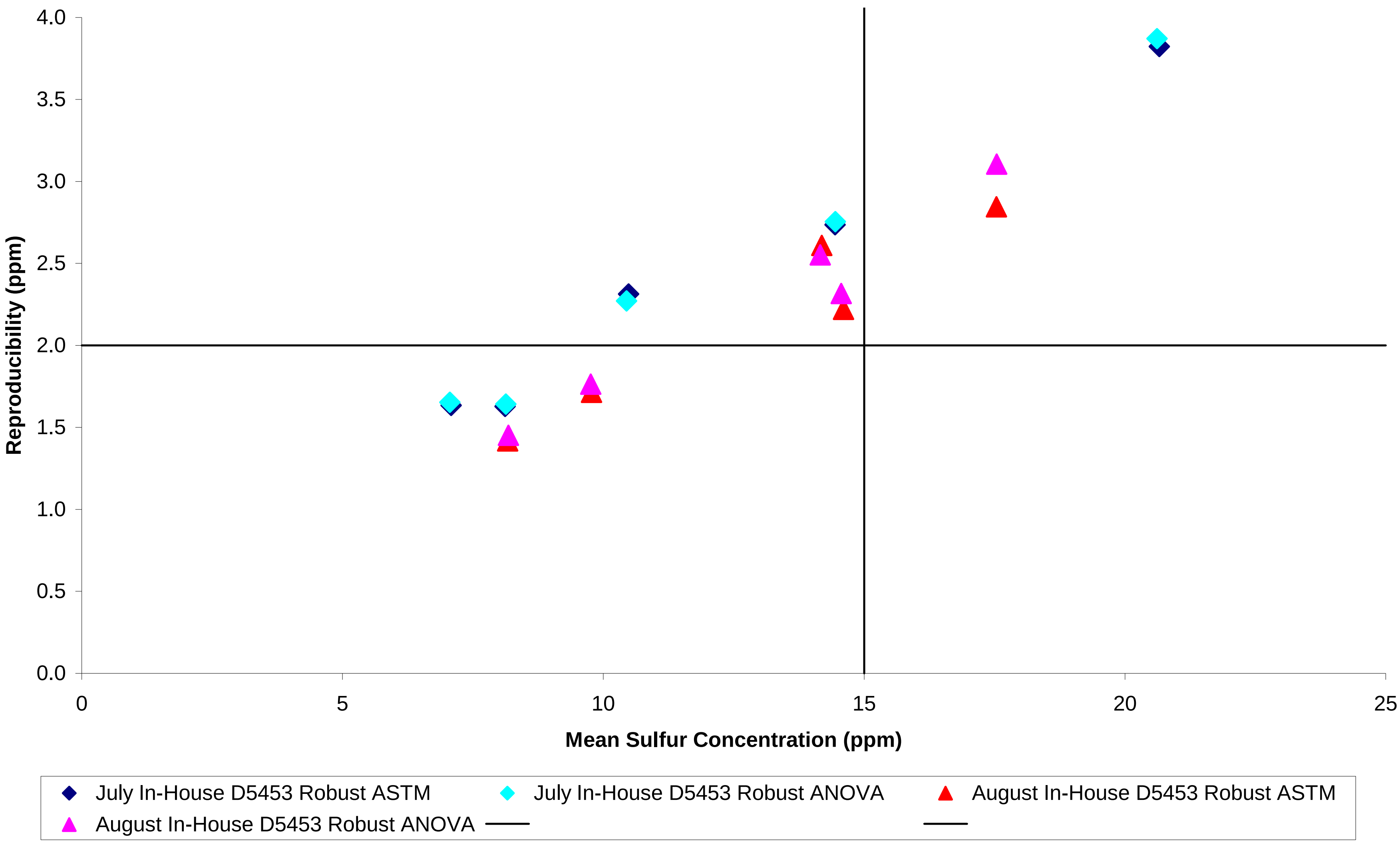
- In contrast, for D7039, D2622, and EDXRF the R-values for the ANOVA method are generally smaller than the R-values for the ASTM method with the NIST calibration.

# *Overall Reproducibility Comparisons: ASTM vs. ANOVA Using In-House*

- Compares R-values for ASTM and ANOVA analyses for all ten fuels.
  – For D5453, D7039, D2622, EDXRF and Composite.

- Uses robust outlier deletion method and the In-House calibration.

## Figure D-21 D5453 Test Method, Robust Outlier, In-House Calibration Comparing ASTM and ANOVA Analyses



## Figure D-45 D7039 Test Method, Robust Outlier, In-House Calibration Comparing ASTM and ANOVA Analyses

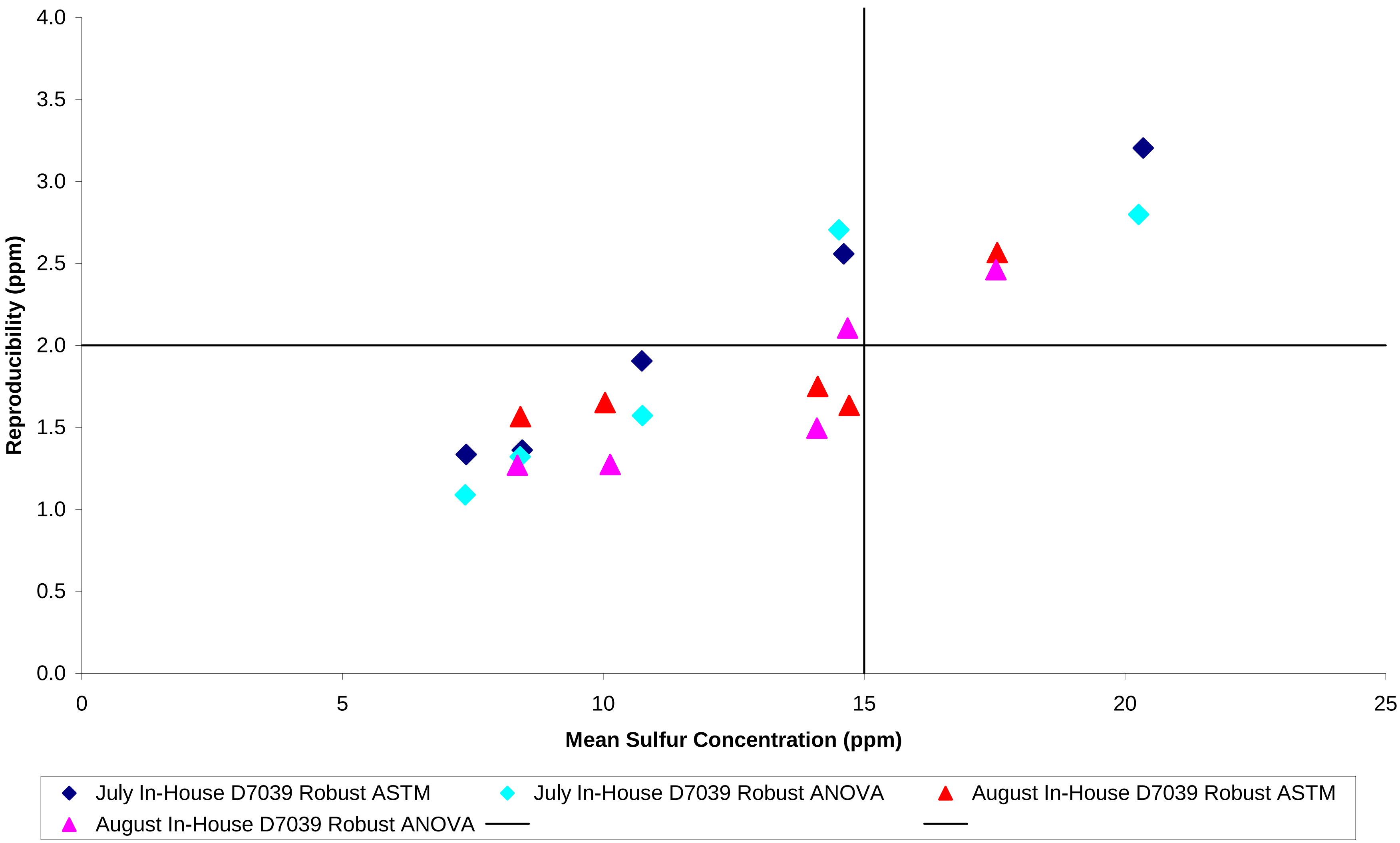


**Figure D-33 D2622 Test Method, Robust Outlier, In-House Calibration Comparing ASTM and ANOVA Analyses**

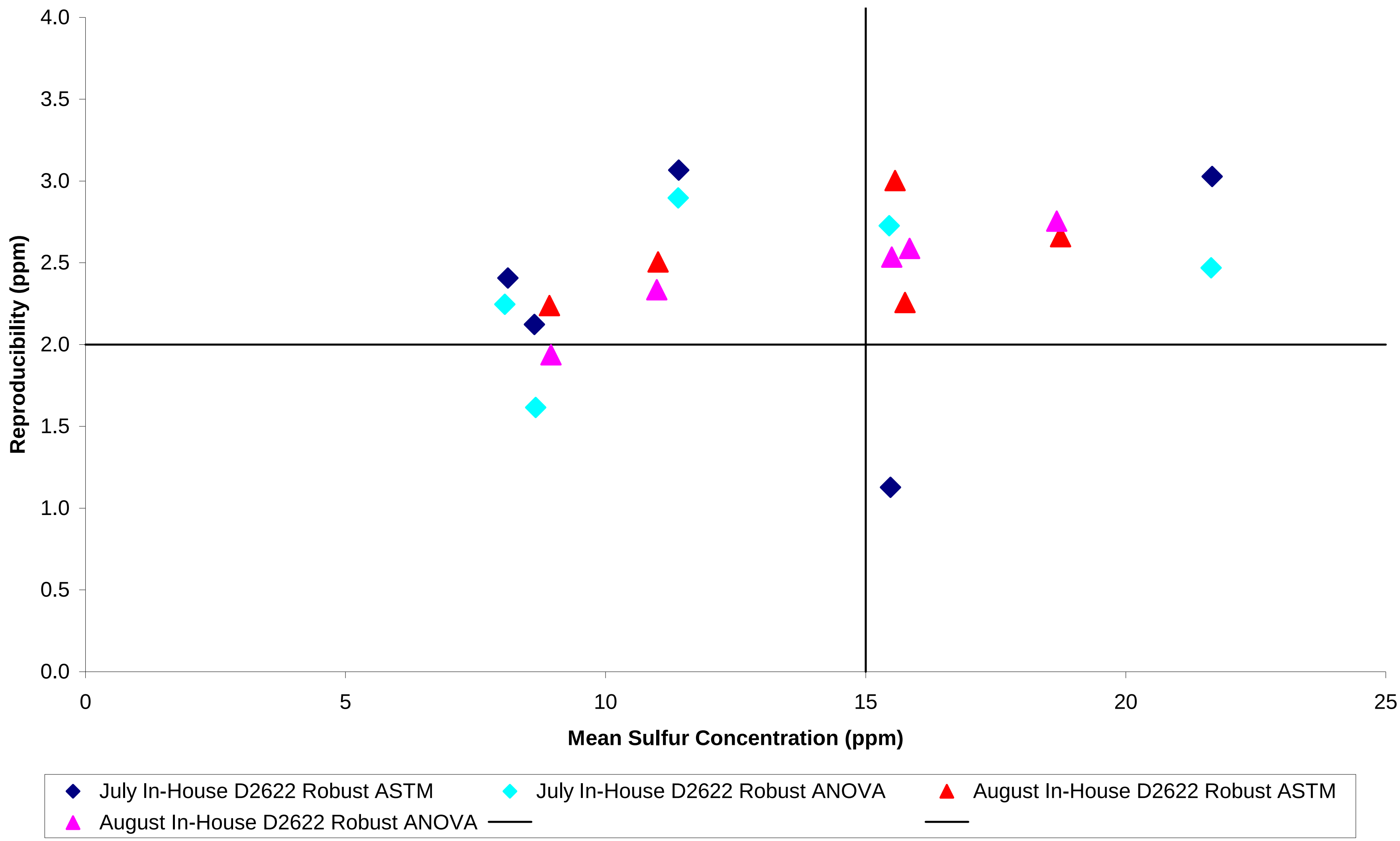
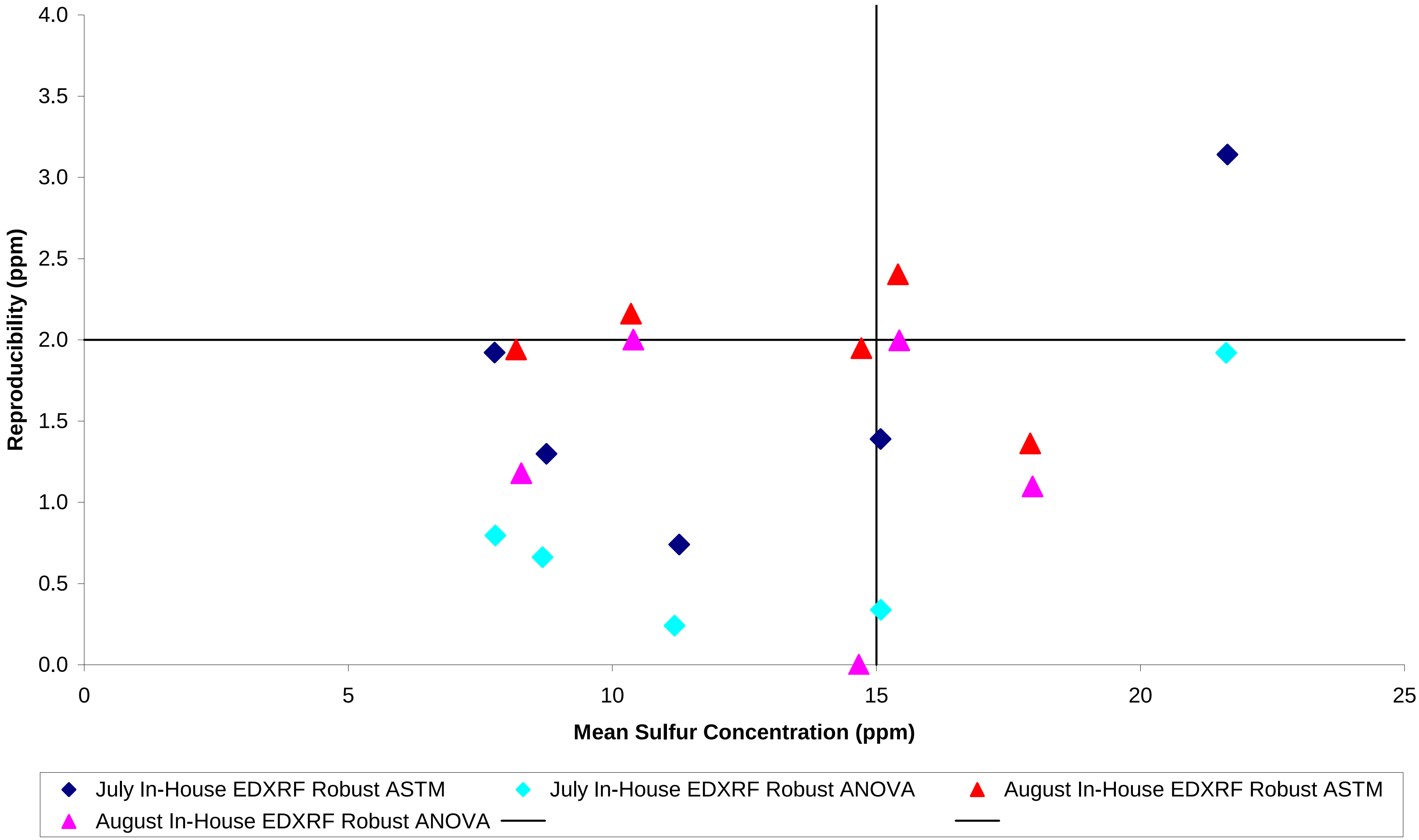
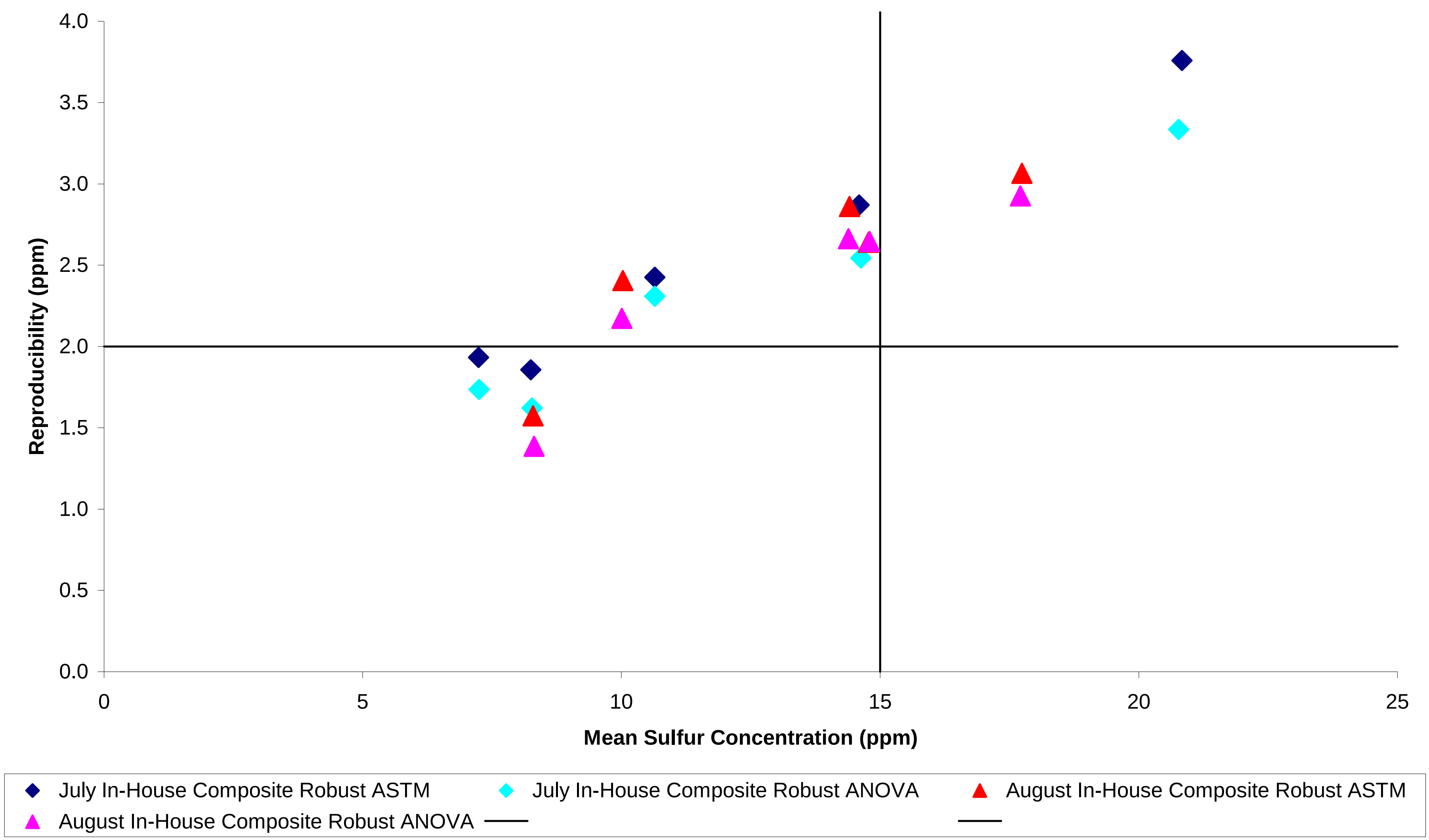


**Figure D-57 EDXRF Test Method, Robust Outlier, In-House Calibration**
**Comparing ASTM and ANOVA Analyses**



## Figure D-9 Composite Test Methods, Robust Outlier, In-House Calibration Comparing ASTM and ANOVA Analyses



# *Summary of R Comparisons: ASTM vs. ANOVA Using In-House*

- For D5453, there is no clear advantage between the R-values for the ASTM and ANOVA methods with the In-house calibration.
- In contrast, for D7039, D2622, and EDXRF the R-values for the ANOVA method are generally smaller than the R-values for the ASTM method and the In-House calibration.

# *Overall Reproducibility Comparisons: In-House vs. NIST*

- Compares R-values for In-House and NIST calibrations for all ten fuels.
  - For D5453, D7039, D2622, EDXRF and Composite.
- Uses robust outlier deletion method and the ASTM analysis.
- Also compares R-values to the 2004-05 ASTM Crosscheck data
  - For D5453, D7039, and D2622.

**Figure D-13   D5453 Test Method, Robust Outlier, ASTM Analysis**
**Comparing In-House and NIST Calibrations**

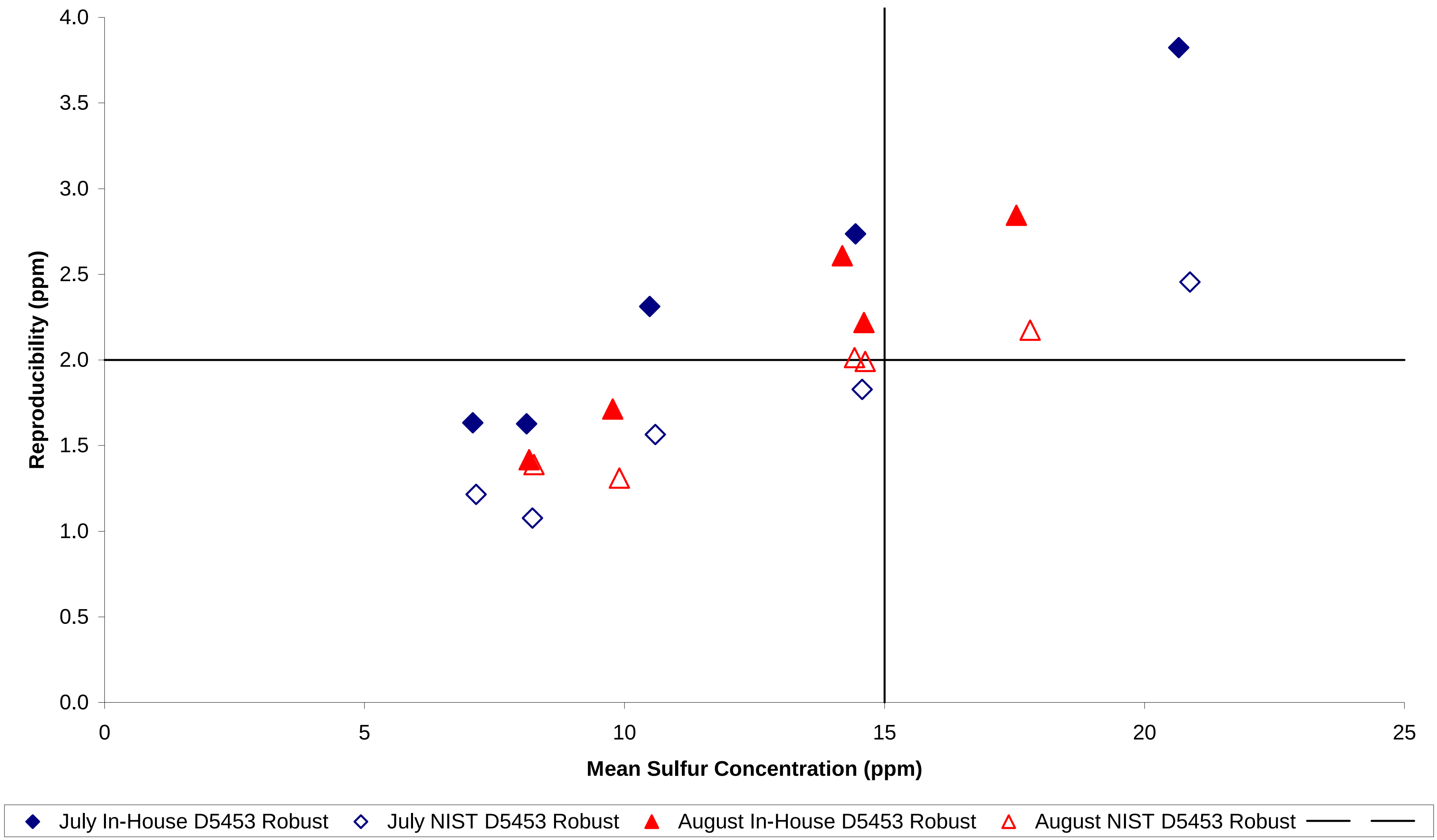


**Figure D-37 D7039 Test Method, Robust Outlier, ASTM Analysis**
**Comparing In-House and NIST Calibrations**

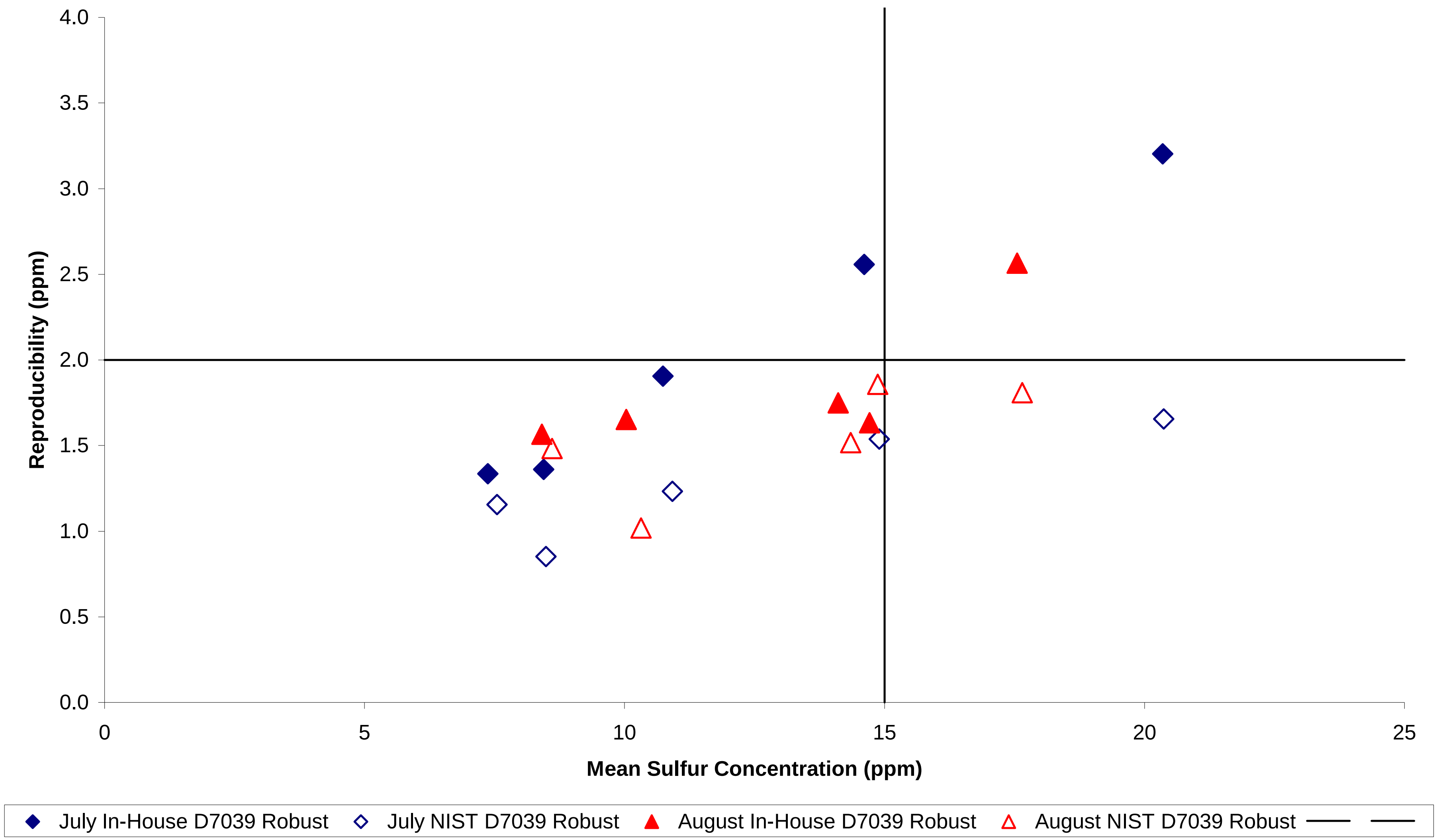


**Figure D-25 D2622 Test Method, Robust Outlier, ASTM Analysis**
**Comparing In-House and NIST Calibrations**

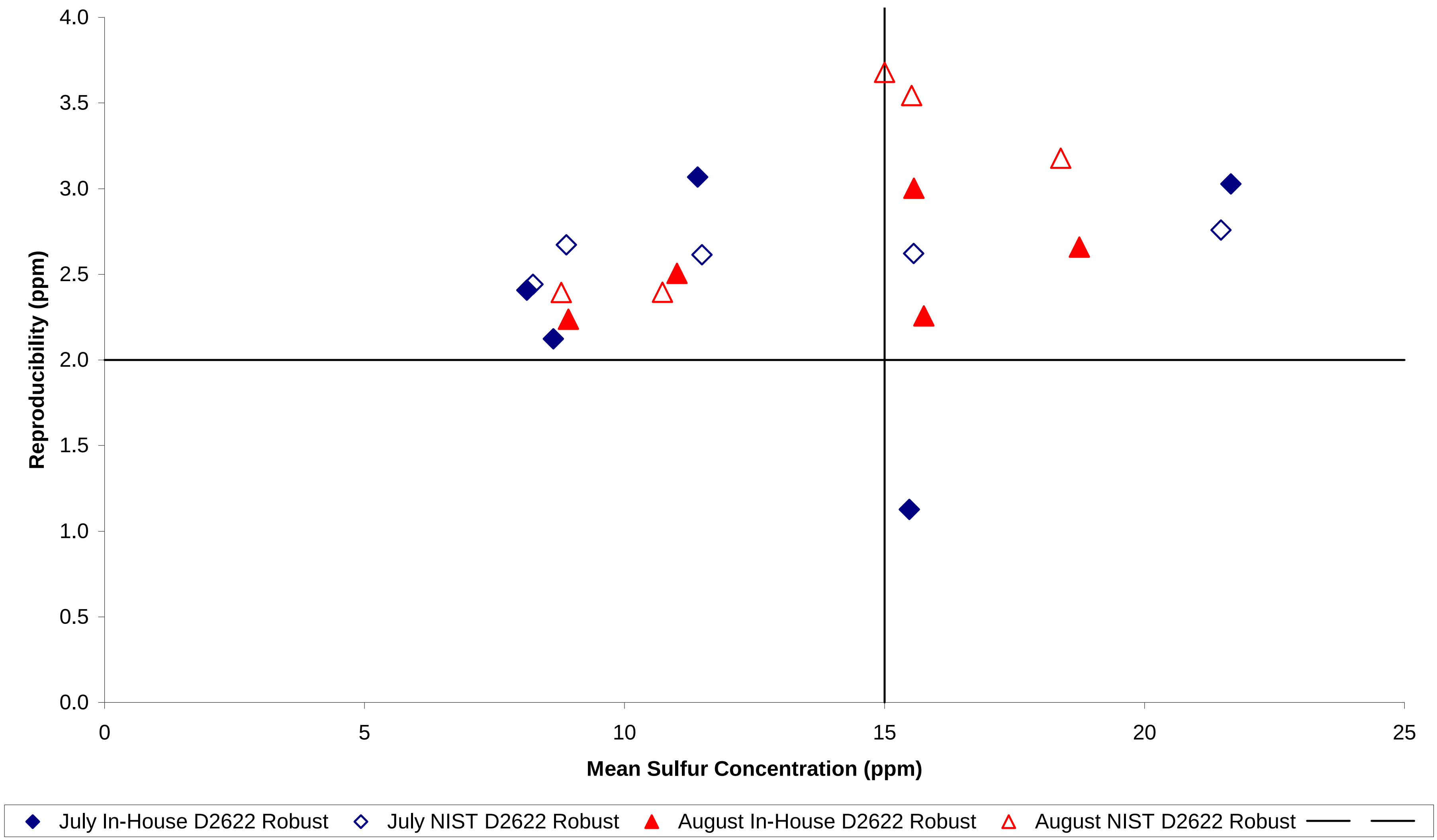


**Figure D-49 EDXRF Test Method, Robust Outlier, ASTM Analysis**
**Comparing In-House and NIST Calibrations**

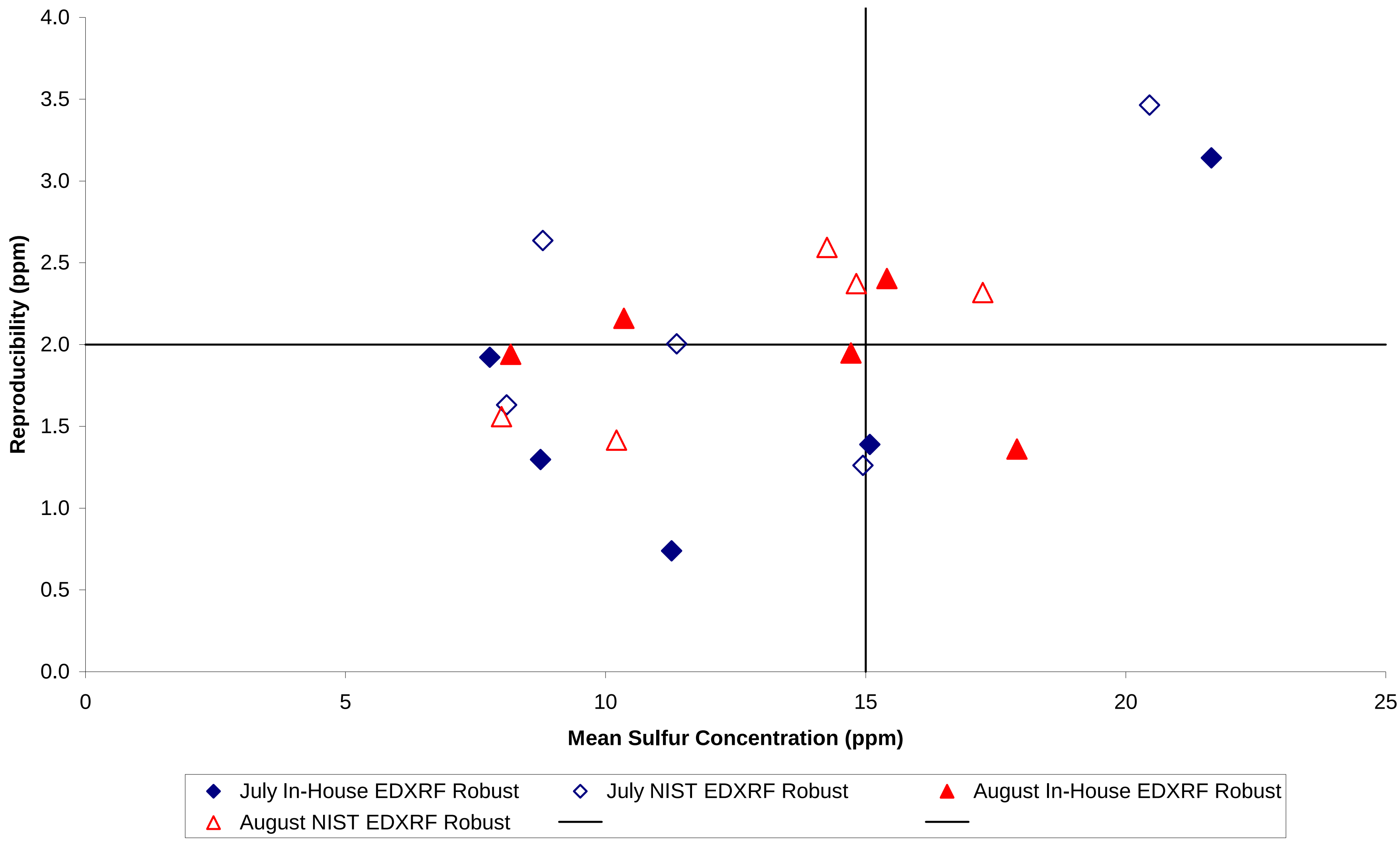


**Figure D-1 Composite Test Methods, Robust Outlier, ASTM Analysis
Comparing In-House and NIST Calibrations**

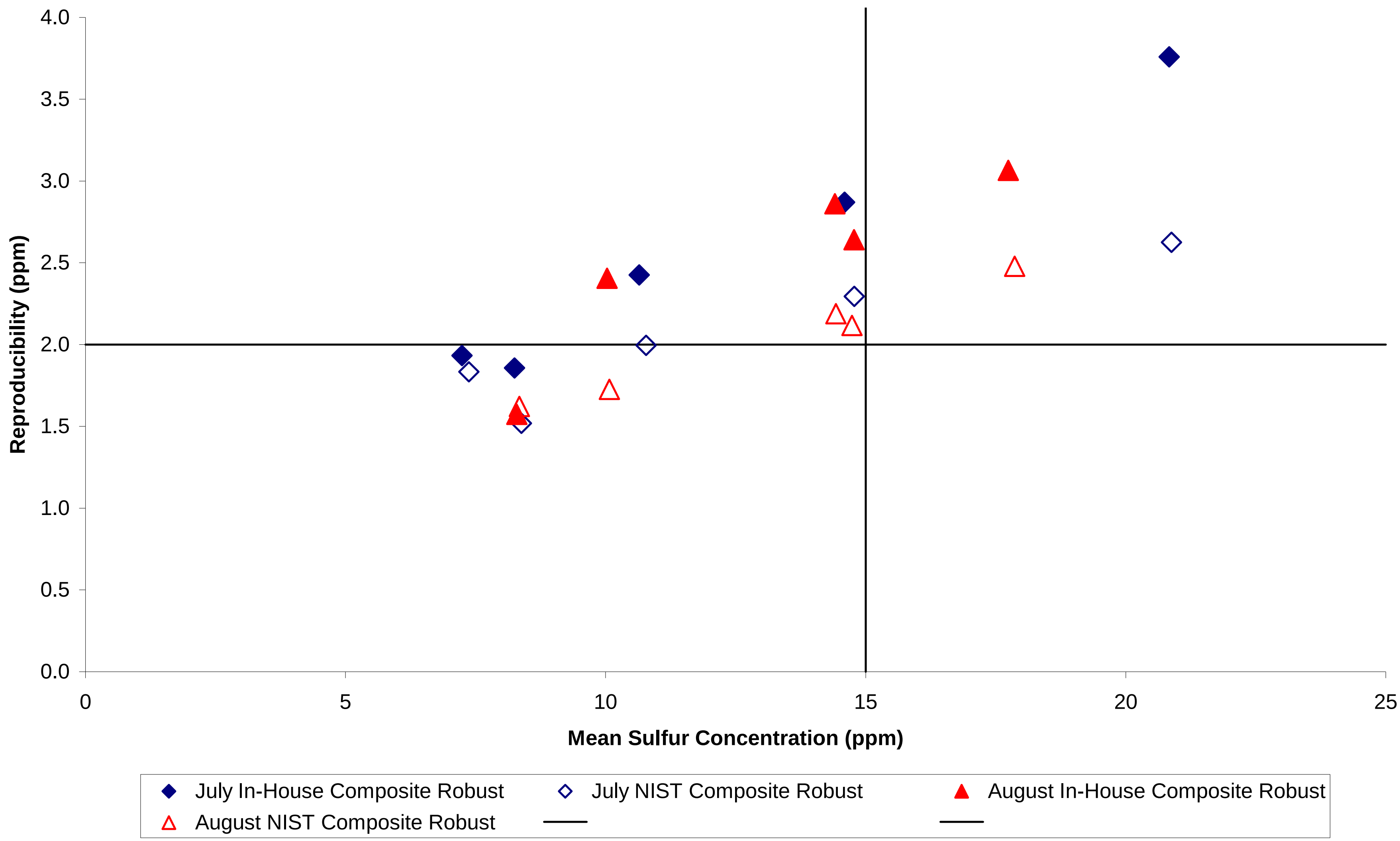


**Figure D-61 D5453 Test Method, Robust Outlier, ASTM Analysis**
**Comparing In-House and NIST Calibrations, Added ASTM 2004-05 Crosscheck Data**

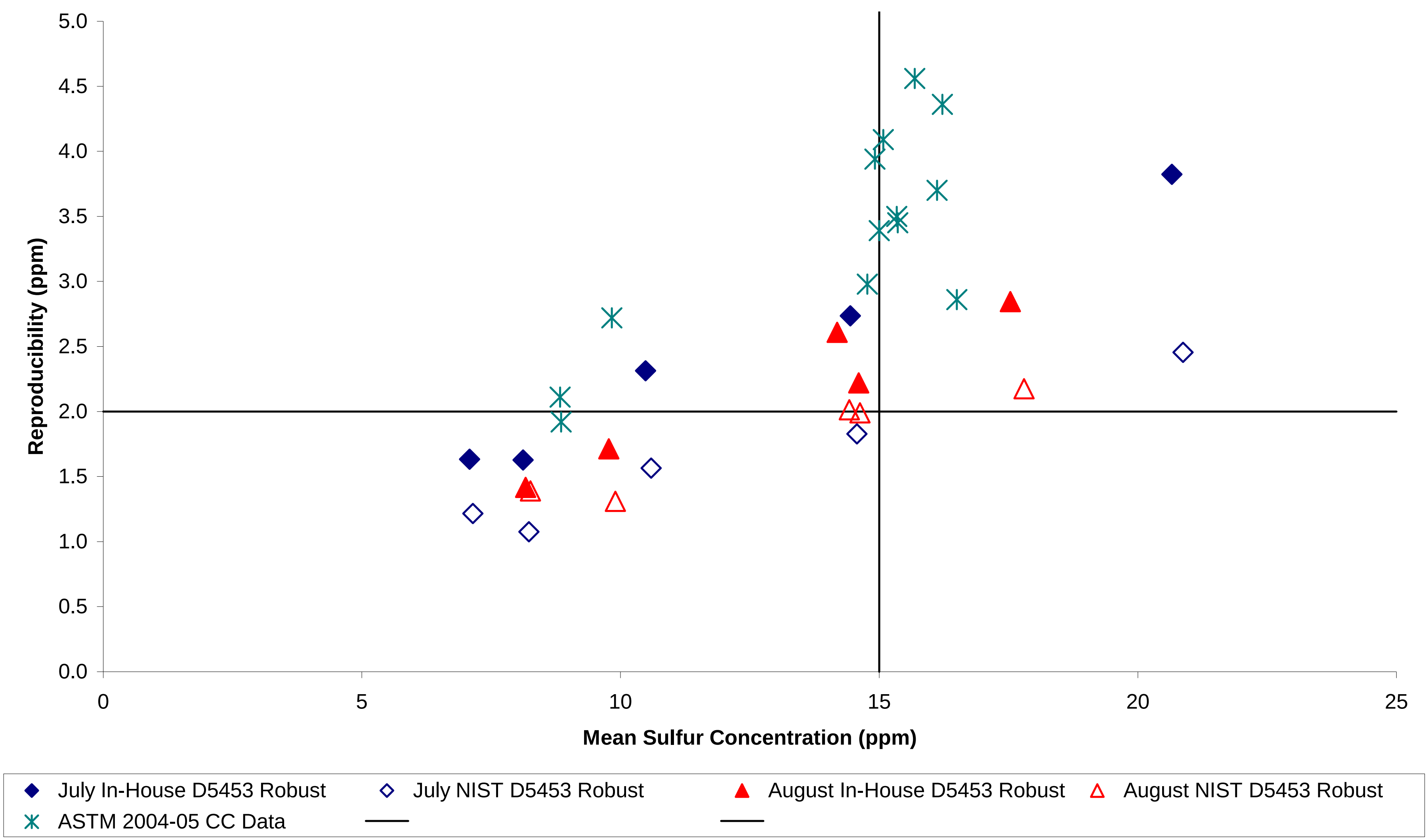


**Figure D-63 D7039 Test Method, Robust Outlier, ASTM Analysis**
**Comparing In-House and NIST Calibrations, Added ASTM 2004-05 Crosscheck Data**

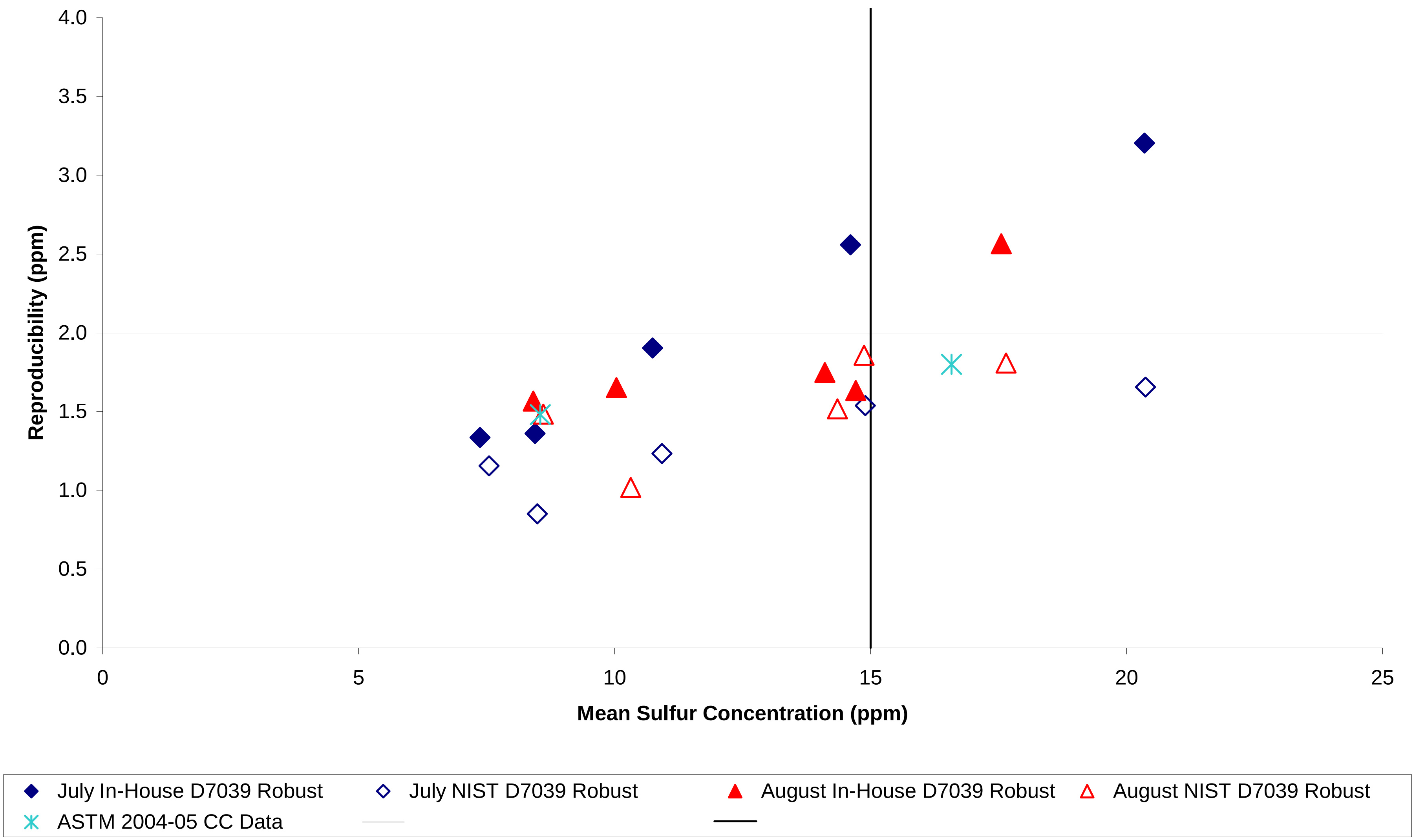
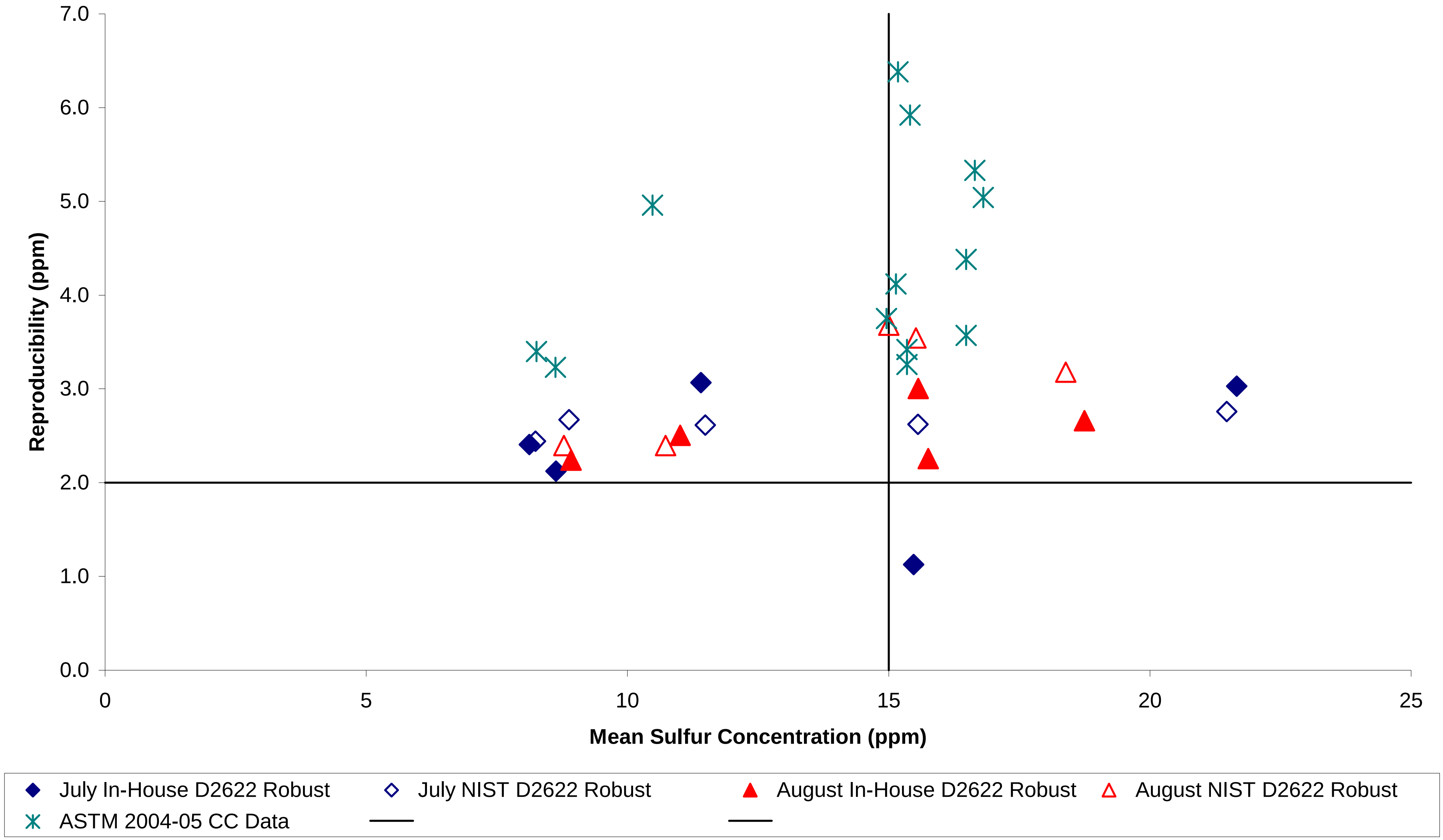


**Figure D-62 D2622 Test Method, Robust Outlier, ASTM Analysis**
**Comparing In-House and NIST Calibrations, Added ASTM 2004-05 Crosscheck Data**



# *Summary of R Comparisons: In-House vs. NIST*

- D7039 has the lowest R-values followed closely by D5453. Both have lower R-values with NIST than with In-House.
- The R-values for D2622 are the largest and have much scatter with neither calibration method being better.
- The R-values for EDXRF have the largest scatter and no clear preferences for a calibration method.
- R-values are smaller relative to the 2004-05 Crosscheck data for D5453 and D2622 data, but similar for D7039 data, and this holds true for both calibration methods.

# *Summary of R Comparisons: Overall Trends*

- The gravimetric deletion method produces lower R-values than the robust deletion method.
- The R-values for NIST are lower than the R-values for In-House.
- The R-values for D5453 and D7039 are superior to those for D2622 and EDXRF, though the EDXRF values may be affected by the small sample size.
- The ANOVA method more often produces similar or slightly lower R-values than the ASTM method for D7039, D2622, and EDXRF, regardless of the calibration, but the results are mixed for D5453 with the ASTM method superior with the NIST calibration method.

# *ASTM Reproducibility Comparisons*

- Compares Robust Means and ASTM R-values, after Robust and Gravimetric outlier deletions and after repeat deletions, for In-House and NIST calibrations for all ten fuels.
  – For D5453, D7039, D2622, EDXRF and Composite.
- Data is presented in rank order within each month of the amount of sulfur in the test fuels.

# D5453 Robust Mean Summary

| Month | Fuel No. | Robust Deletion | | Gravimetric Deletion | |
|---|---|---|---|---|---|
| | | In-House | NIST | In-House | NIST |
| July | 1 | 7.08 | 7.15 | 7.14 | 7.15 |
| | 4 | 8.12 | 8.23 | 8.22 | 8.24 |
| | 2 | 10.48 | 10.59 | 10.57 | 10.62 |
| | 5 | 14.44 | 14.57 | 14.58 | 14.61 |
| | 3 | 20.66 | 20.87 | 20.84 | 20.86 |
| August | 4 | 8.16 | 8.26 | 8.26 | 8.26 |
| | 1 | 9.77 | 9.90 | 9.87 | 9.93 |
| | 2 | 14.19 | 14.43 | 14.33 | 14.41 |
| | 5 | 14.60 | 14.63 | 14.67 | 14.64 |
| | 3 | 17.54 | 17.60 | 17.77 | 17.84 |

# *D5453 ASTM Reproducibility Summary*

| Month | Fuel No. | Robust Deletion | | Gravimetric Deletion | |
|---|---|---|---|---|---|
| | | In-House | NIST | In-House | NIST |
| **July** | 1 | 1.63 | 1.22 | 1.33 | 0.95 |
| | 4 | 1.63 | 1.08 | 1.26 | 0.91 |
| | 2 | 2.31 | 1.56 | 1.61 | 1.38 |
| | 5 | 2.74 | 1.83 | 1.97 | 1.65 |
| | 3 | 3.82 | 2.46 | 3.05 | 2.24 |
| **August** | 4 | 1.42 | 1.39 | 1.25 | 1.11 |
| | 1 | 1.71 | 1.31 | 1.51 | 1.12 |
| | 2 | 2.61 | 2.01 | 1.96 | 1.71 |
| | 5 | 2.22 | 1.99 | 1.86 | 1.60 |
| | 3 | 2.85 | 2.17 | 2.29 | 2.05 |

# D7039 Robust Mean Summary

| Month | Fuel No. | Robust Deletion | | Gravimetric Deletion | |
|---|---|---|---|---|---|
| | | In-House | NIST | In-House | NIST |
| July | 1 | 7.37 | 7.55 | 7.39 | 7.52 |
| | 4 | 8.44 | 8.49 | 8.48 | 8.40 |
| | 2 | 10.74 | 10.92 | 10.82 | 10.91 |
| | 5 | 14.61 | 14.90 | 14.78 | 14.97 |
| | 3 | 20.35 | 20.37 | 20.47 | 20.37 |
| August | 4 | 8.41 | 8.61 | 8.30 | 8.43 |
| | 1 | 10.03 | 10.32 | 10.25 | 10.31 |
| | 2 | 14.11 | 14.35 | 14.31 | 14.32 |
| | 5 | 14.71 | 14.87 | 14.82 | 14.72 |
| | 3 | 17.55 | 17.65 | 17.77 | 17.67 |

# D7039 ASTM Reproducibility Summary

| Month | Fuel No. | Robust Deletion | | Gravimetric Deletion | |
|---|---|---|---|---|---|
| | | In-House | NIST | In-House | NIST |
| **July** | 1 | 1.34 | 1.16 | 1.22 | 1.16 |
| | 4 | 1.36 | 0.85 | 1.22 | 0.80 |
| | 2 | 1.90 | 1.23 | 1.39 | 1.13 |
| | 5 | 2.56 | 1.54 | 1.92 | 1.22 |
| | 3 | 3.20 | 1.66 | 2.50 | 1.71 |
| **August** | 4 | 1.57 | 1.48 | 0.86 | 1.56 |
| | 1 | 1.65 | 1.02 | 1.95 | 1.09 |
| | 2 | 1.75 | 1.52 | 1.53 | 1.77 |
| | 5 | 1.63 | 1.86 | 1.58 | 2.11 |
| | 3 | 2.57 | 1.81 | 1.97 | 1.82 |

# *D2622 Robust Mean Summary*

| Month | Fuel No. | Robust Deletion | | Gravimetric Deletion | |
|---|---|---|---|---|---|
| | | In-House | NIST | In-House | NIST |
| July | 1 | 8.12 | 8.24 | 8.05 | 8.08 |
| | 4 | 8.63 | 8.88 | 8.53 | 8.55 |
| | 2 | 11.40 | 11.49 | 11.44 | 11.51 |
| | 5 | 15.48 | 15.56 | 15.51 | 15.68 |
| | 3 | 21.66 | 21.47 | 21.40 | 21.34 |
| August | 4 | 8.92 | 8.78 | 8.57 | 8.70 |
| | 1 | 11.01 | 10.73 | 10.80 | 10.81 |
| | 2 | 15.57 | 15.00 | 15.29 | 14.85 |
| | 5 | 15.76 | 15.52 | 15.72 | 15.43 |
| | 3 | 18.75 | 18.39 | 18.58 | 18.22 |

# *D2622 ASTM Reproducibility Summary*

| Month | Fuel No. | Robust Deletion | | Gravimetric Deletion | |
|---|---|---|---|---|---|
| | | In-House | NIST | In-House | NIST |
| **July** | 1 | 2.41 | 2.44 | 2.34 | 2.41 |
| | 4 | 2.12 | 2.67 | 1.72 | 1.43 |
| | 2 | 3.07 | 2.61 | 3.24 | 2.39 |
| | 5 | 3.15 | 2.62 | 2.66 | 1.83 |
| | 3 | 3.03 | 2.76 | 2.26 | 2.41 |
| **August** | 4 | 2.24 | 2.39 | 1.01 | 1.44 |
| | 1 | 2.51 | 2.40 | 2.60 | 2.13 |
| | 2 | 3.00 | 3.68 | 2.71 | 3.00 |
| | 5 | 2.26 | 3.54 | 2.15 | 2.90 |
| | 3 | 2.66 | 3.18 | 2.41 | 2.55 |

# *EDXRF Robust Mean Summary*

| Month | Fuel No. | Robust Deletion | | Gravimetric Deletion | |
|---|---|---|---|---|---|
| | | In-House | NIST | In-House | NIST |
| July | 1 | 7.77 | 8.10 | 7.61 | 8.16* |
| | 4 | 8.75 | 8.79 | 8.68 | 9.10* |
| | 2 | 11.27 | 11.36 | 11.26* | 11.35* |
| | 5 | 15.08 | 14.95 | 15.13 | 15.09* |
| | 3 | 21.65 | 20.46 | 21.67 | 20.84* |
| August | 4 | 8.18 | 8.00 | 8.17 | 8.36* |
| | 1 | 10.35 | 10.21 | 10.40 | 10.21* |
| | 2 | 14.72 | 14.26 | 14.65 | 14.89* |
| | 5 | 15.41 | 14.82 | 15.47 | 15.30* |
| | 3 | 17.91 | 17.25 | 17.98 | 17.75* |

* Not enough labs (<6) to run valid reproducibility study.

# *EDXRF ASTM Reproducibility Summary*

| Month | Fuel No. | Robust Deletion | | Gravimetric Deletion | |
|---|---|---|---|---|---|
| | | In-House | NIST | In-House | NIST |
| July | 1 | 1.92 | 1.63 | 1.72 | 1.11* |
| | 4 | 1.30 | 2.64 | 1.39 | 1.59* |
| | 2 | 0.74 | 2.00 | 0.59* | 2.32* |
| | 5 | 1.39 | 1.26 | 2.11 | 1.49* |
| | 3 | 3.14 | 3.46 | 1.88 | 2.59* |
| August | 4 | 1.94 | 1.56 | 2.26 | 0.18* |
| | 1 | 2.16 | 1.42 | 2.42 | 1.51* |
| | 2 | 1.95 | 2.59 | 1.93 | 1.64* |
| | 5 | 2.40 | 2.37 | 2.42 | 1.48* |
| | 3 | 1.36 | 2.32 | 1.49 | 0.89* |

* Not enough labs (<6) to run valid reproducibility study.

# *Summary of ASTM Reproducibility Comparisons*

- There are slight variations in the robust means for D5453 and D7039 across the two deletion methods and two calibration methods. More mean variation is present when using D2622 and EDXRF.
- The smallest R-values occur for NIST with gravimetric deletion.
- The R-values for D5453 are always less for NIST over In-House regardless of the deletion method.
- The R-values for D7039 are always less for NIST for the July fuels regardless of the deletion method, and for 4 of 5 July fuels the gravimetric deletion and NIST results yield the smallest R-values. The results are mixed for the August fuels.
- The R-value results for D2622 and EDXRF are mixed.

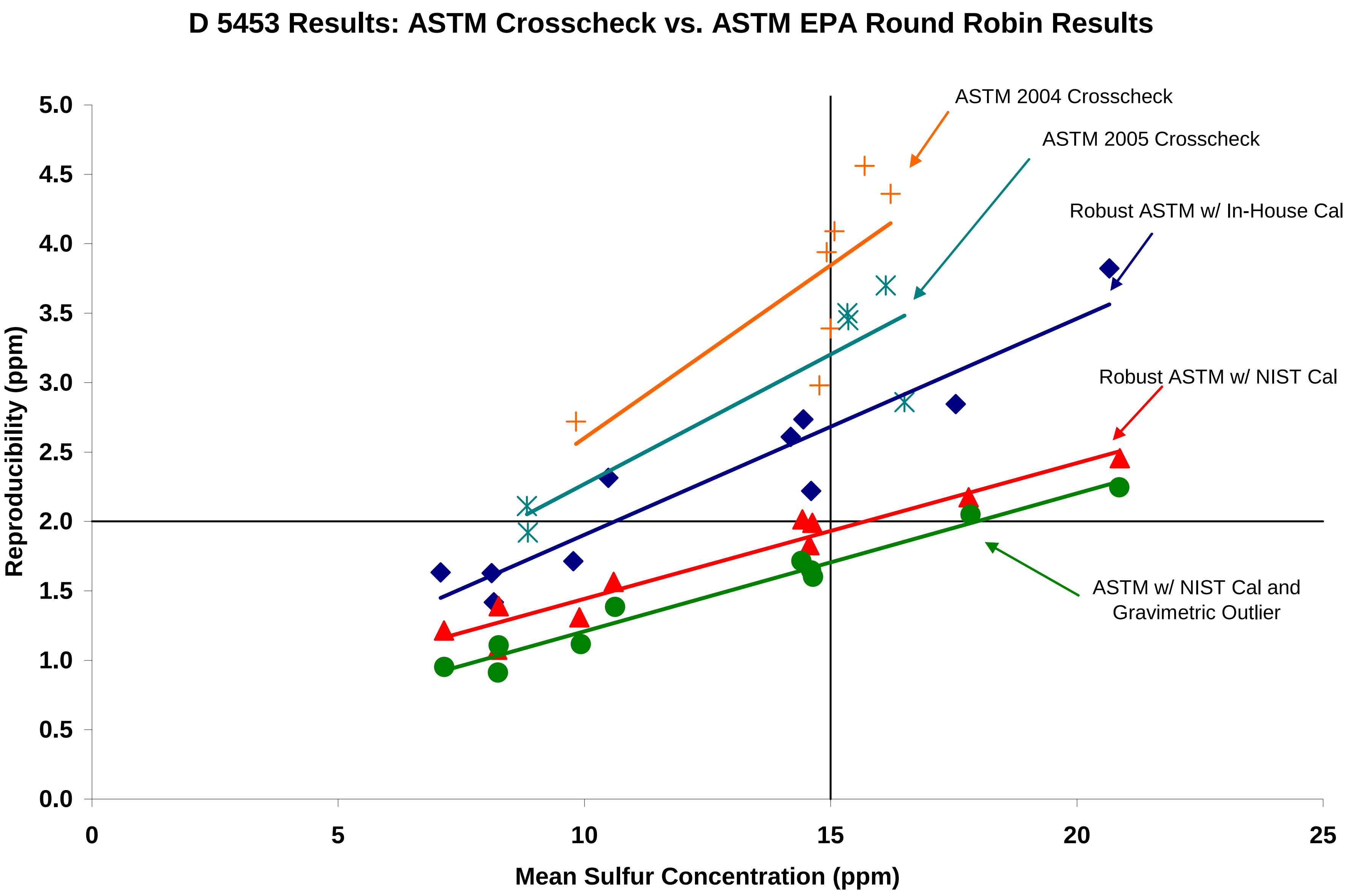
# *Regression Fits to ASTM Reproducibility*

- ASTM Crosscheck data are being compared to the ASTM and ANOVA EPA Round Robin results.
  - The ASTM Crosscheck data are for sulfur concentrations > 8 ppm.
  - The ASTM Crosscheck R-values were computed using the ASTM method.

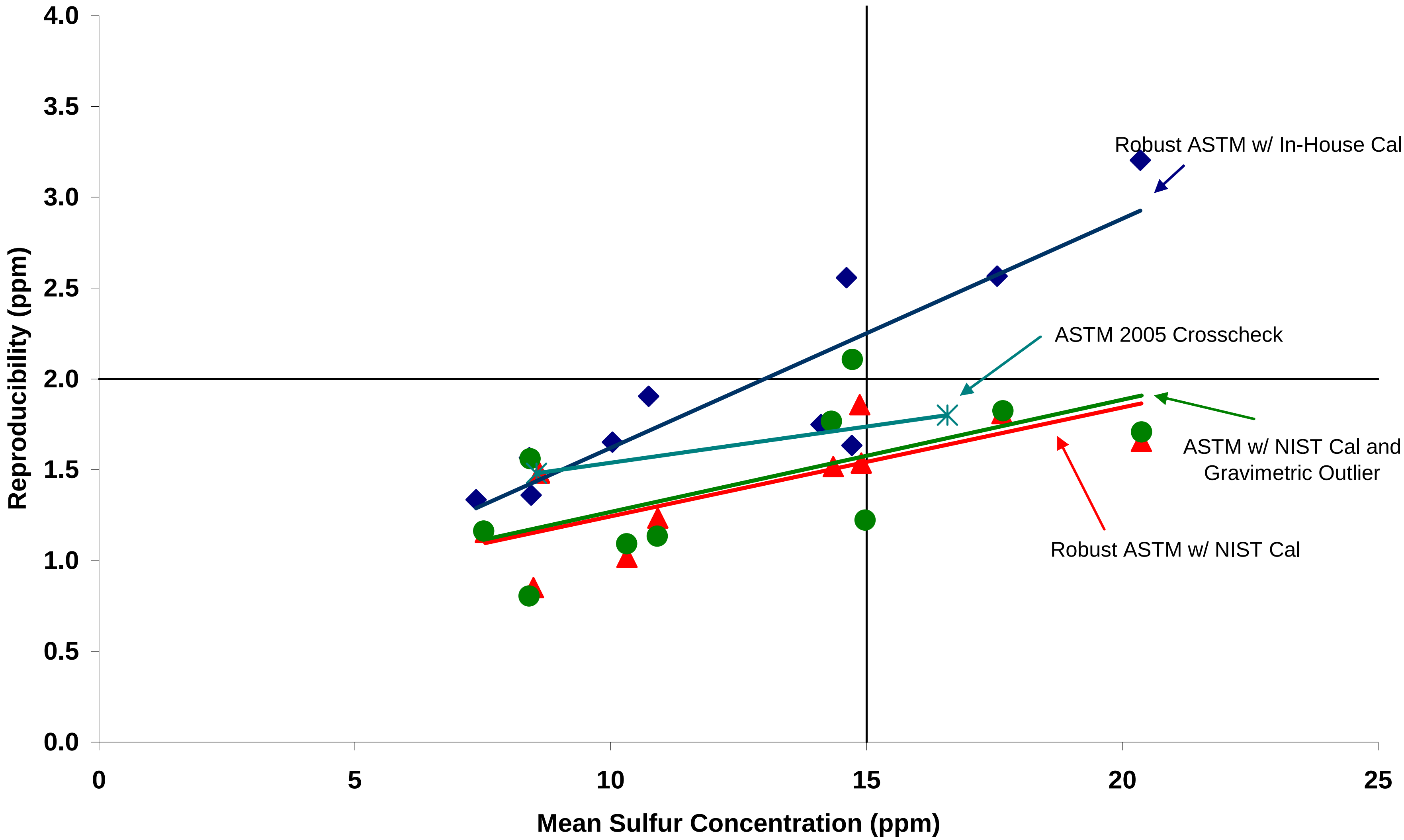
- Straight-line regression models are fit using the form:

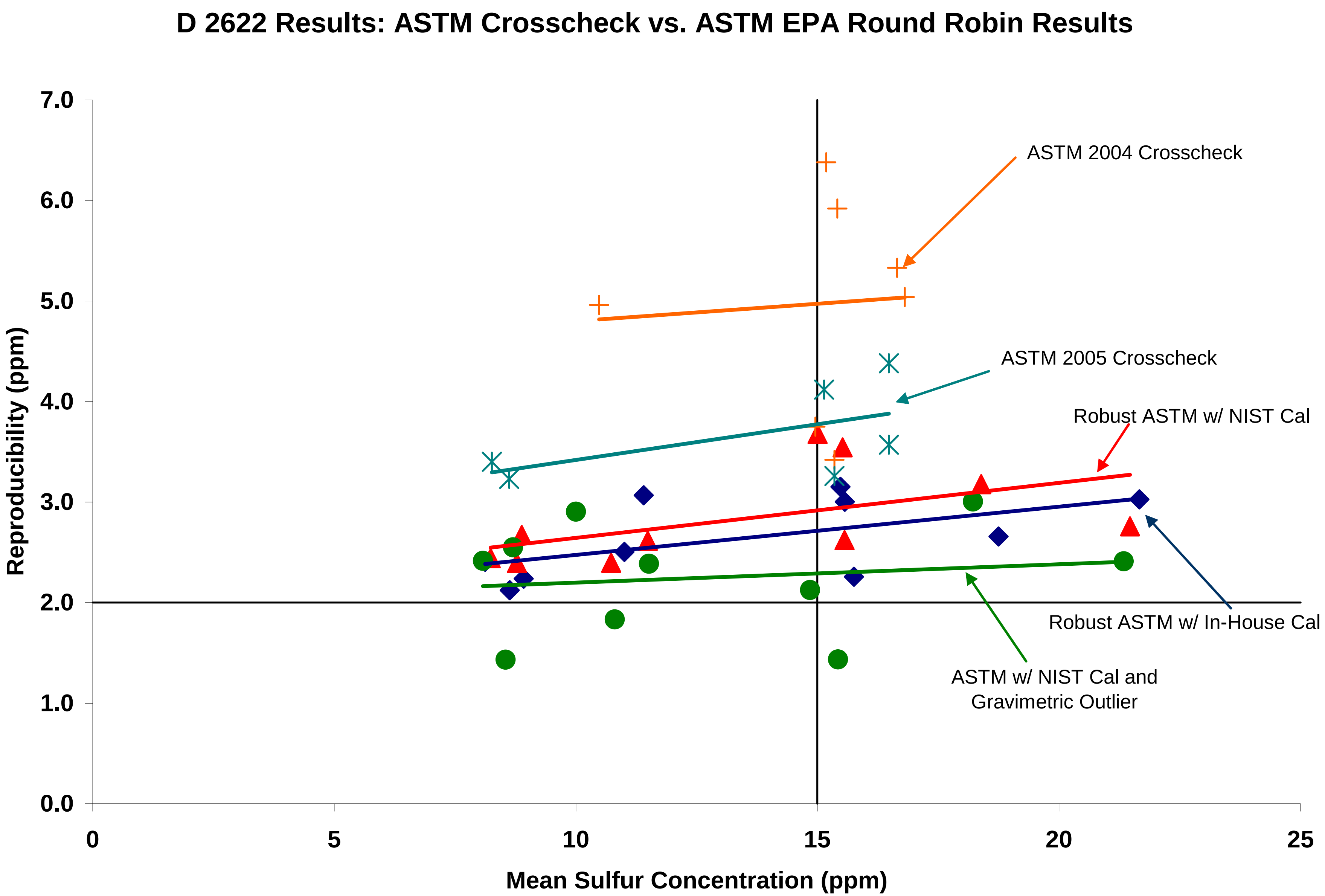
  Reproducibility = Intercept + Coeff.*(Mean Sulfur).

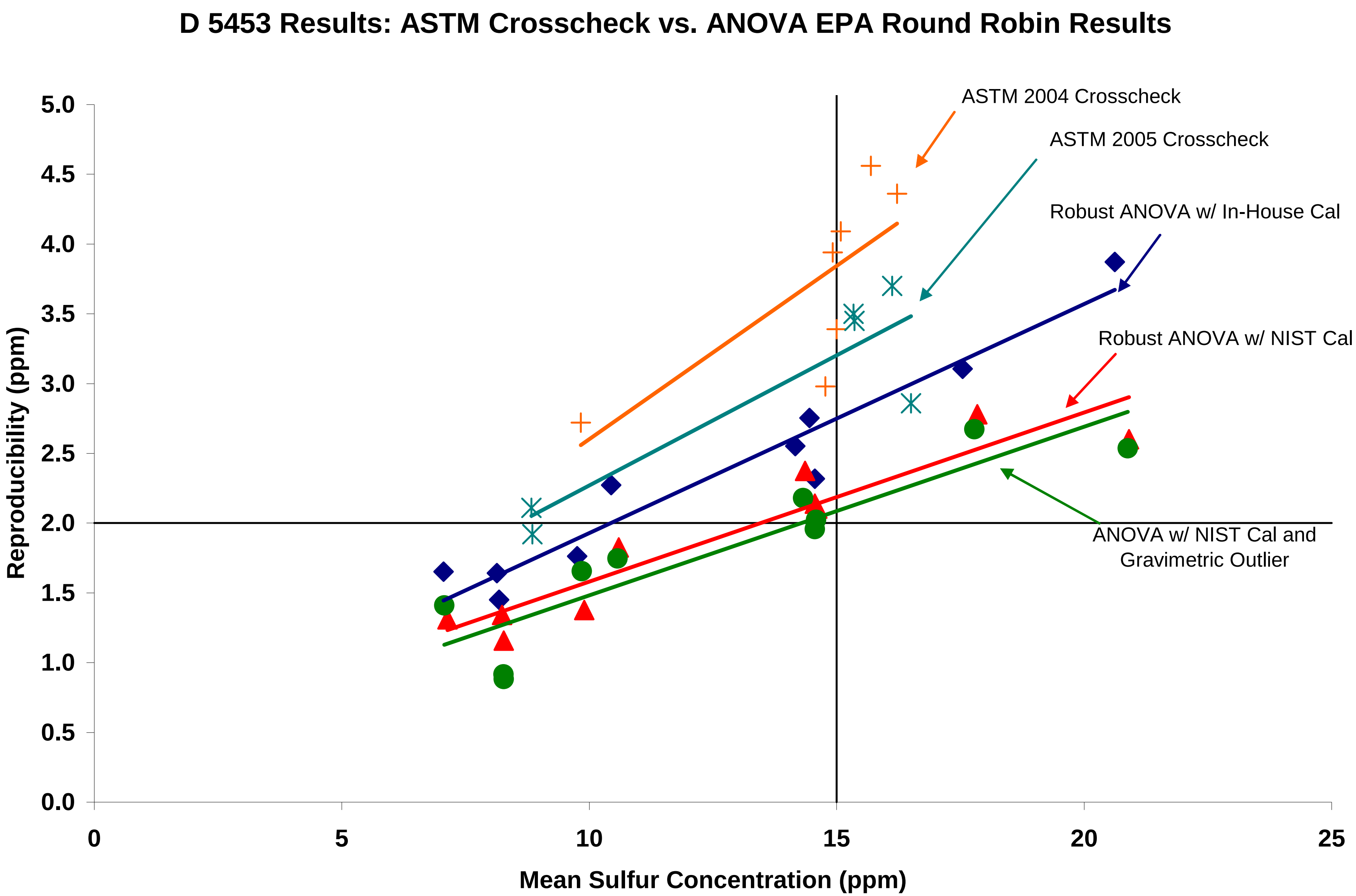
- The regression models are used to predict and compare the R-values when the mean sulfur = 15 ppm.

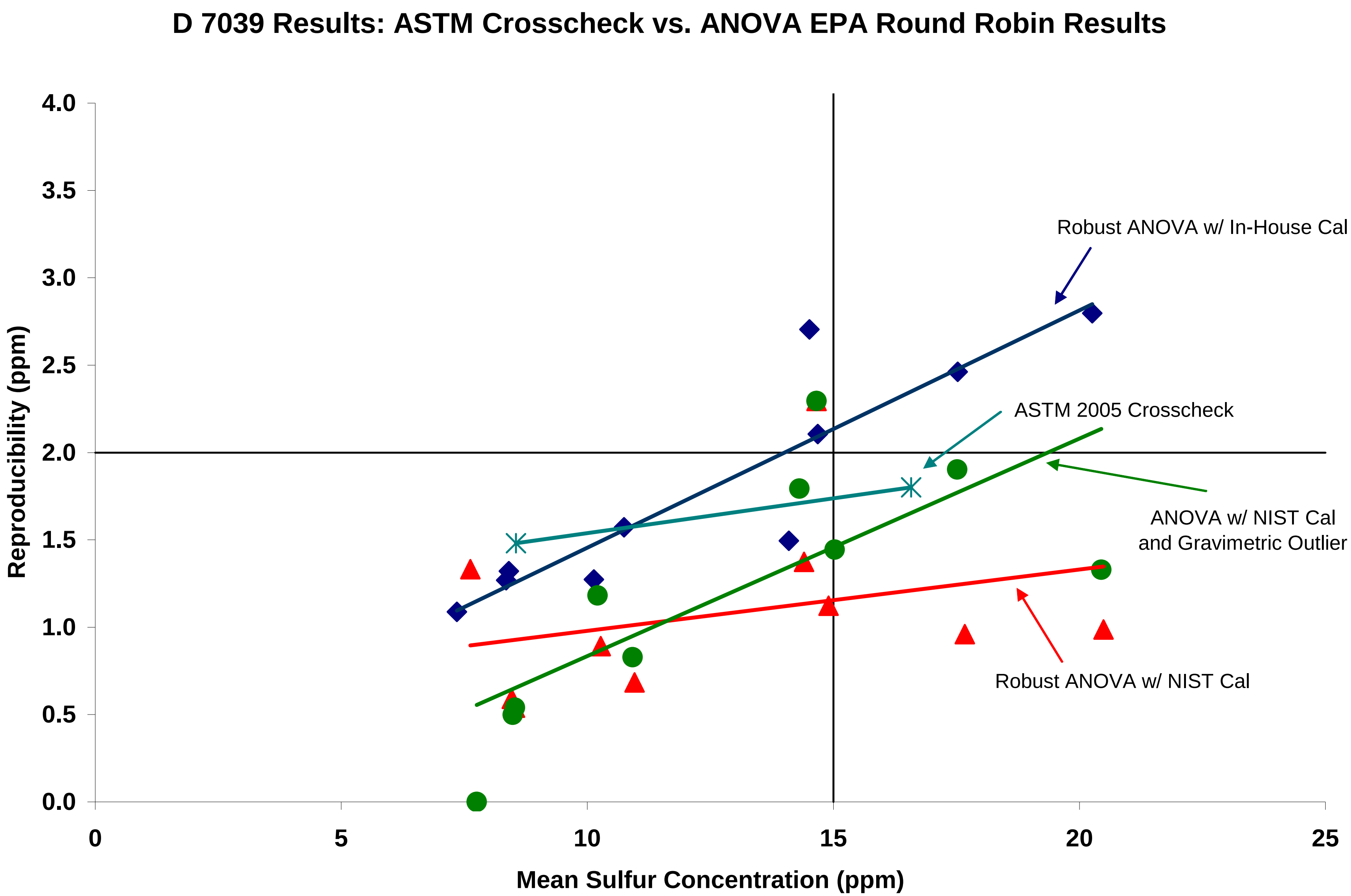

D 5453 Results: ASTM Crosscheck vs. ASTM EPA Round Robin Results
Reproducibility (ppm)
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
ASTM 2004 Crosscheck
ASTM 2005 Crosscheck
Robust ASTM w/ In-House Cal
Robust ASTM w/ NIST Cal
ASTM w/ NIST Cal and Gravimetric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

## D 7039 Results: ASTM Crosscheck vs. ASTM EPA Round Robin Results

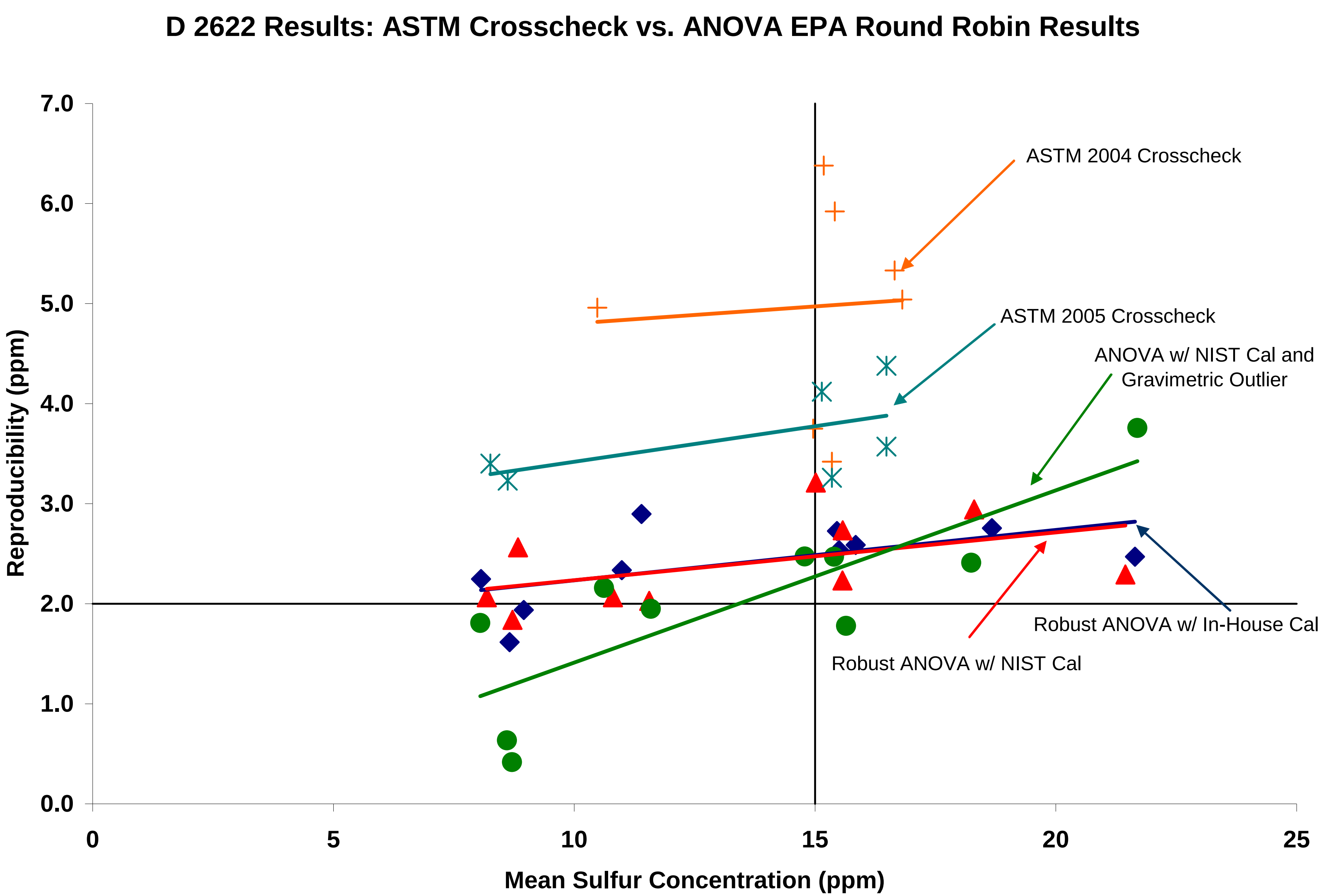



D 2622 Results: ASTM Crosscheck vs. ASTM EPA Round Robin Results
Reproducibility (ppm)
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
ASTM 2004 Crosscheck
ASTM 2005 Crosscheck
Robust ASTM w/ NIST Cal
Robust ASTM w/ In-House Cal
ASTM w/ NIST Cal and
Gravimetric Outlier


D 5453 Results: ASTM Crosscheck vs. ANOVA EPA Round Robin Results
ASTM 2004 Crosscheck
ASTM 2005 Crosscheck
Robust ANOVA w/ In-House Cal
Robust ANOVA w/ NIST Cal
ANOVA w/ NIST Cal and Gravimetric Outlier
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



D 7039 Results: ASTM Crosscheck vs. ANOVA 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)
Robust ANOVA w/ In-House Cal
ASTM 2005 Crosscheck
ANOVA w/ NIST Cal
and Gravimetric Outlier
Robust ANOVA w/ NIST Cal


D 2622 Results: ASTM Crosscheck vs. ANOVA 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
ANOVA w/ NIST Cal and
Gravimetric Outlier
Robust ANOVA w/ In-House Cal
Robust ANOVA w/ NIST Cal

# *Predicted Reproducibility at 15 ppm*

| Approach | Method | ASTM R | ANOVA R |
|---|---|---|---|
| ASTM 2004 CC | D 2622 | 4.97 | |
| | D 5453 | 3.84 | |
| ASTM 2005 CC | D 2622 | 3.78 | |
| | D 5453 | 3.20 | |
| | D 7039 | 1.74 | |
| NIST Calibration Gravimetric Deletion | Composite | 1.87 | 1.96 |
| | D 2622 | 2.29 | 2.27 |
| | D 5453 | 1.71 | 2.09 |
| | D 7039 | 1.58 | 1.46 |
| NIST Calibration Robust Deletion | Composite | 2.21 | 2.15 |
| | D 2622 | 2.91 | 2.47 |
| | D 5453 | 1.93 | 2.19 |
| | D 7039 | 1.54 | 1.15 |
| In-House Calibration Robust Deletion | Composite | 2.86 | 2.64 |
| | D 2622 | 2.71 | 2.49 |
| | D 5453 | 2.68 | 2.75 |
| | D 7039 | 2.25 | 2.13 |

# *Summary of Predicted Reproducibility at 15 ppm*

- 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.
  - EPA RR predicted values are lower for D5453 and D2622.
  - EPA and ASTM predicted values are comparable for D7039.
- 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.

# *Summary of Predicted Reproducibility at 15 ppm*

- The data support the conclusion that using the identical NIST calibration curves across the participating labs reduces curve bias contributions to reproducibility.
  - A reduction in predicted R (ASTM, at 15 ppm) over the predicted R values obtained using the 2004 and 2005 CC data were apparent in all cases when using the NIST calibration curves.
  - The magnitude of the reduction in predicted R (ASTM R, at 15 ppm) from In-House to NIST under robust deletion was 0.73 ppm on average for D5453 and D7039 (excluding D 2622).
- 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).

# Repeatability Results

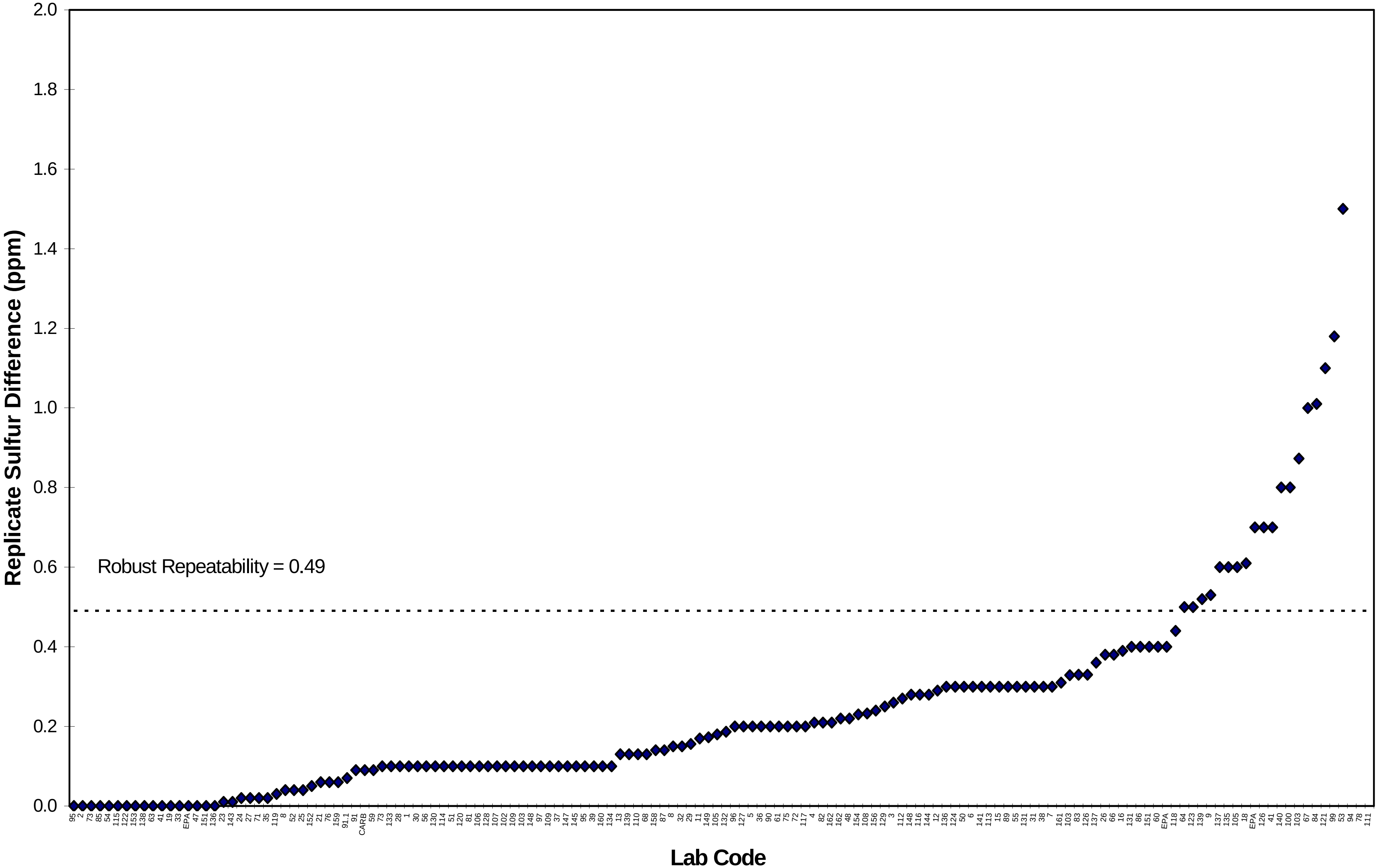
# *Lab ASTM Repeatability Comparisons*

- Compares ASTM lab repeatability values for the following conditions:
  - July fuel #4 and August fuel #5 data
  - Composite test method
  - In-House and NIST calibrations
  - Robust outlier deletion method

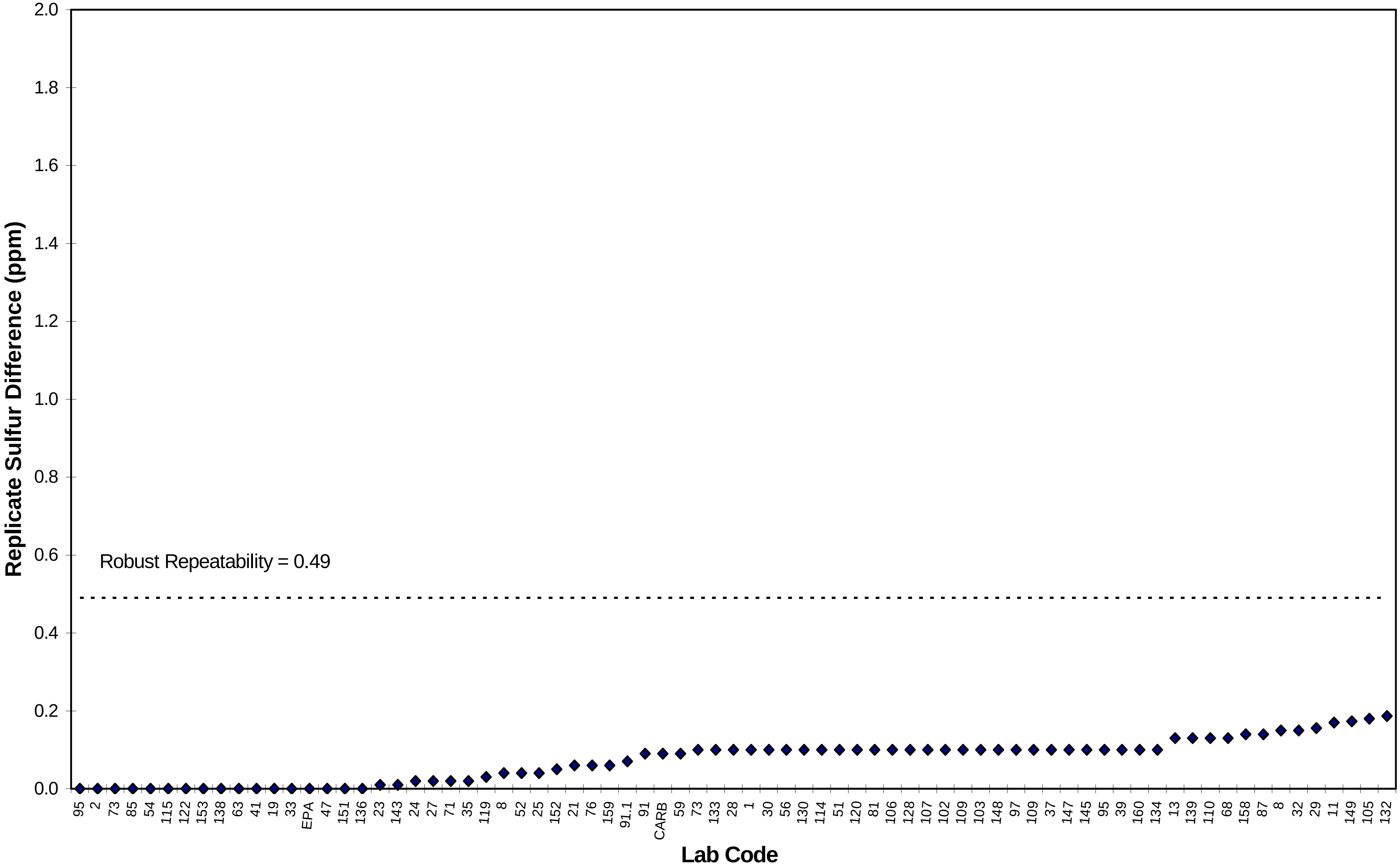
- Lab sulfur differences between repeats versus lab codes are plotted.
  - Compares repeat difference to composite lab repeatability.
  - Differences that exceed repeatability limit are noted.
  - Codes of the deleted labs are listed at the right of the figure title.

**Figure A-1 July Fuel #4, Robust Outlier Deletion**
**Composite Test Methods In-House Calibration**
**ASTM Analysis, Lab Repeatability**

Codes of deleted labs = 78, 94, 111



## Figure A-1a July Fuel #4, Robust Outlier Deletion
## Composite Test Methods In-House Calibration
## ASTM Analysis, Lab Repeatability



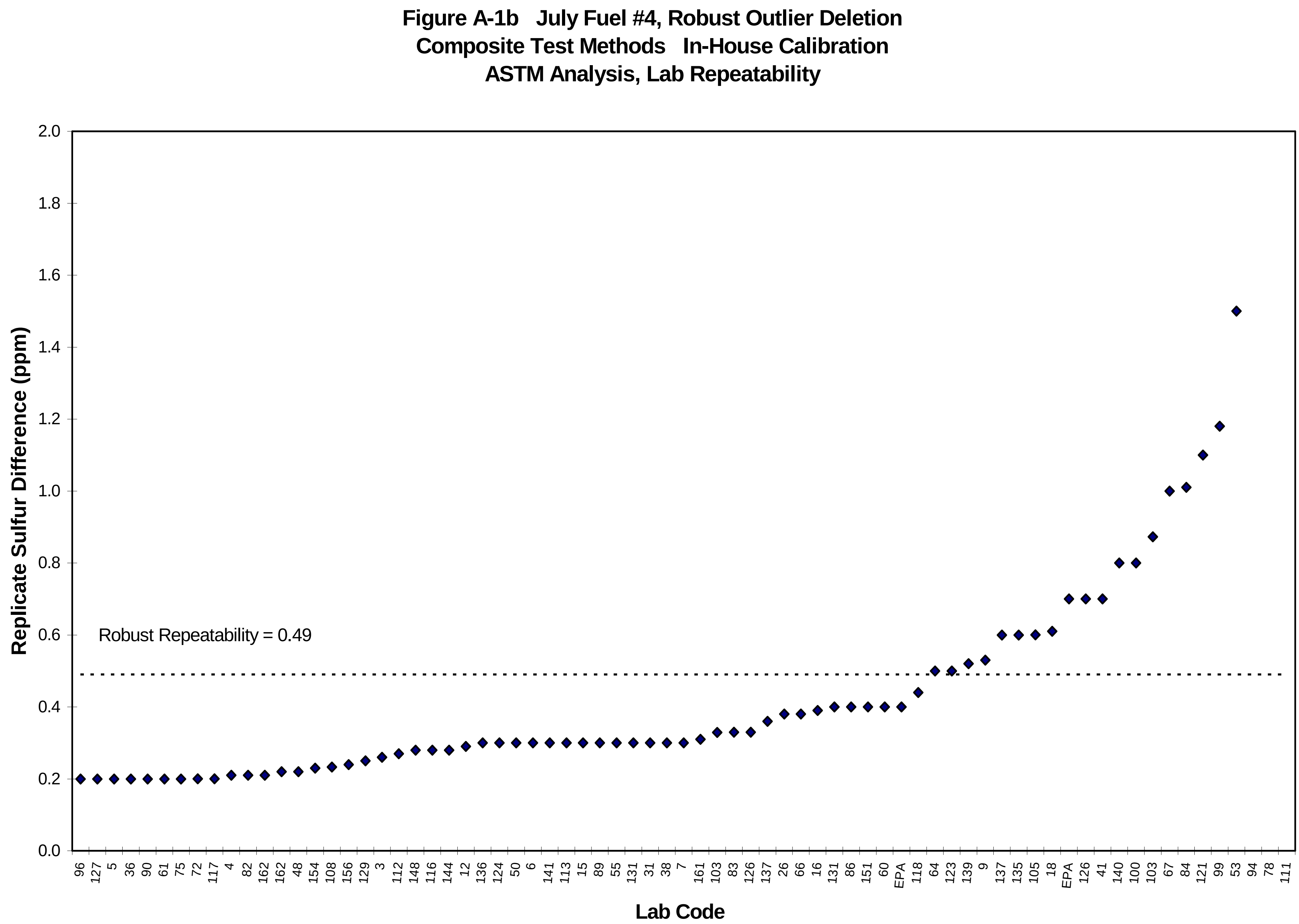


Figure A-1b   July Fuel #4, Robust Outlier Deletion
Composite Test Methods   In-House Calibration
ASTM Analysis, Lab Repeatability
Replicate Sulfur Difference (ppm)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Robust Repeatability = 0.49
96 127 5 36 90 61 75 72 117 4 82 162 162 48 154 108 156 129 3 112 148 116 144 12 136 124 50 6 141 113 15 89 55 131 31 38 7 161 103 83 126 137 26 66 16 131 86 151 60 EPA 118 64 123 139 9 137 135 105 18 EPA 126 41 140 100 103 67 84 121 99 53 94 78 111
Lab Code

**Figure A-2 July Fuel #4, Robust Outlier Deletion**
**Composite Test Methods NIST Calibration**
**ASTM Analysis, Lab Repeatability**

Codes of deleted labs =
33, 78, 84, 94, 100, 103, 124, 126, 139, 145

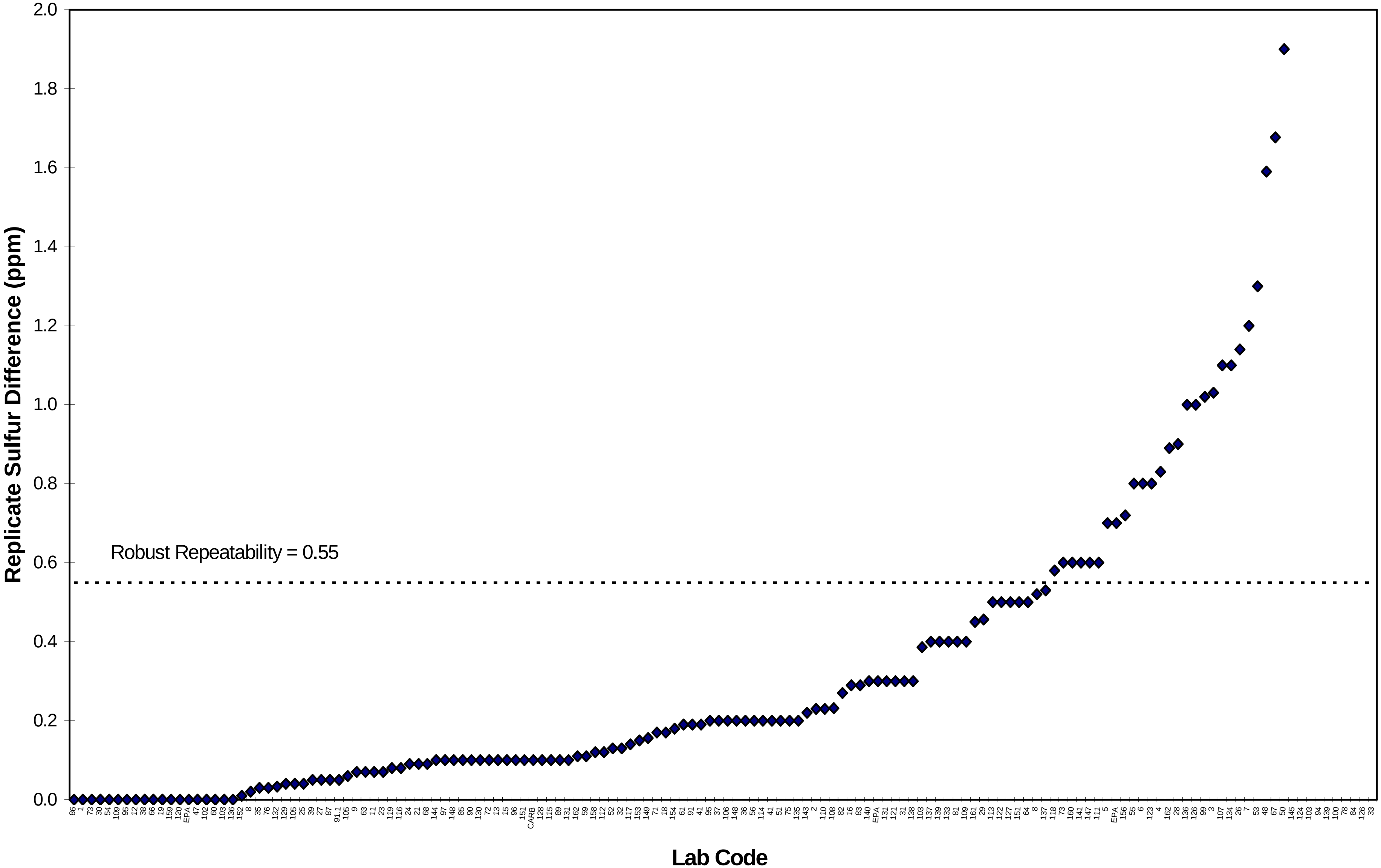


Figure A-2a July Fuel #4, Robust Outlier Deletion
Composite Test Methods NIST Calibration
ASTM Analysis, Lab Repeatability

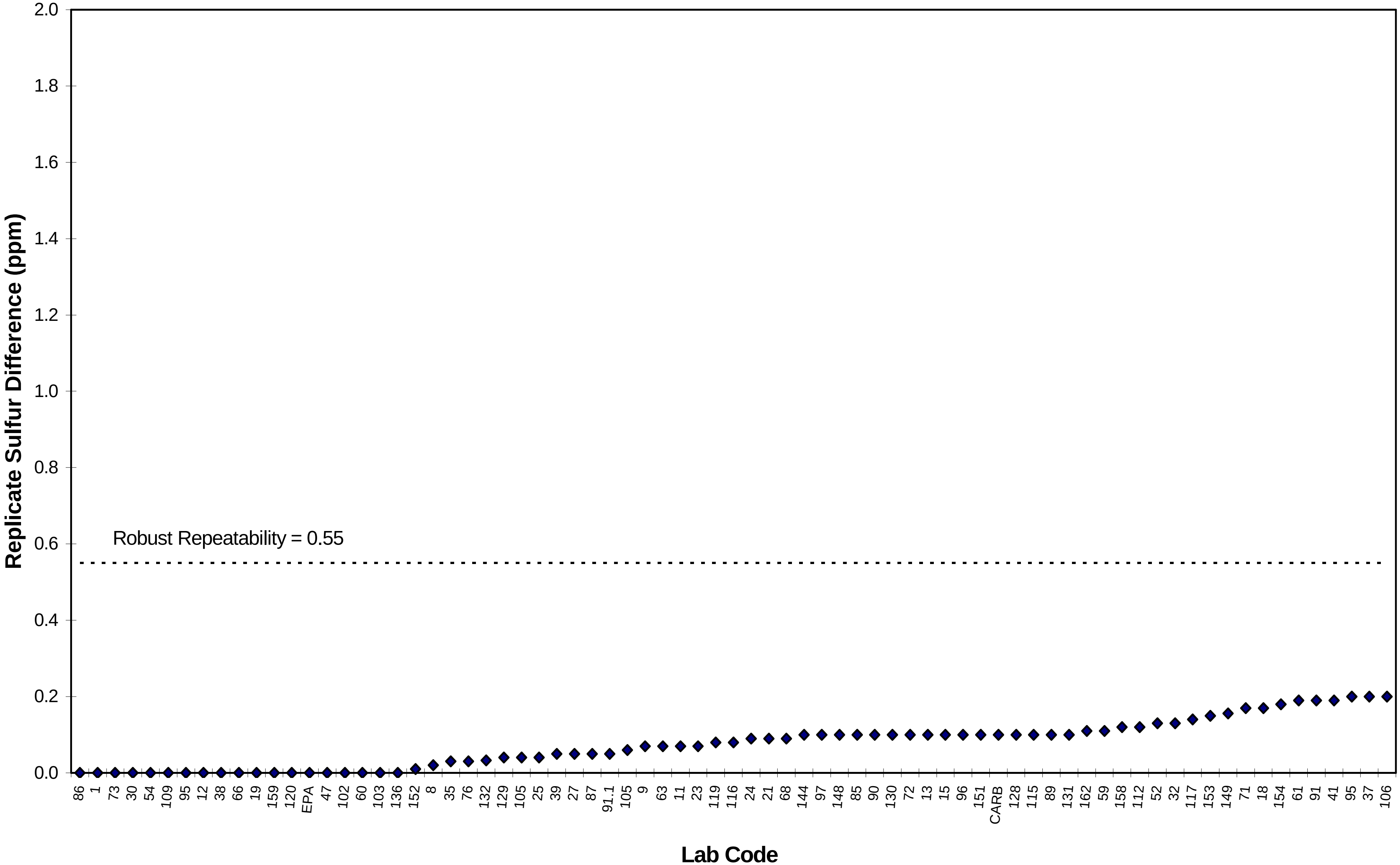


**Figure A-2b July Fuel #4, Robust Outlier Deletion**
**Composite Test Methods NIST Calibration**
**ASTM Analysis, Lab Repeatability**

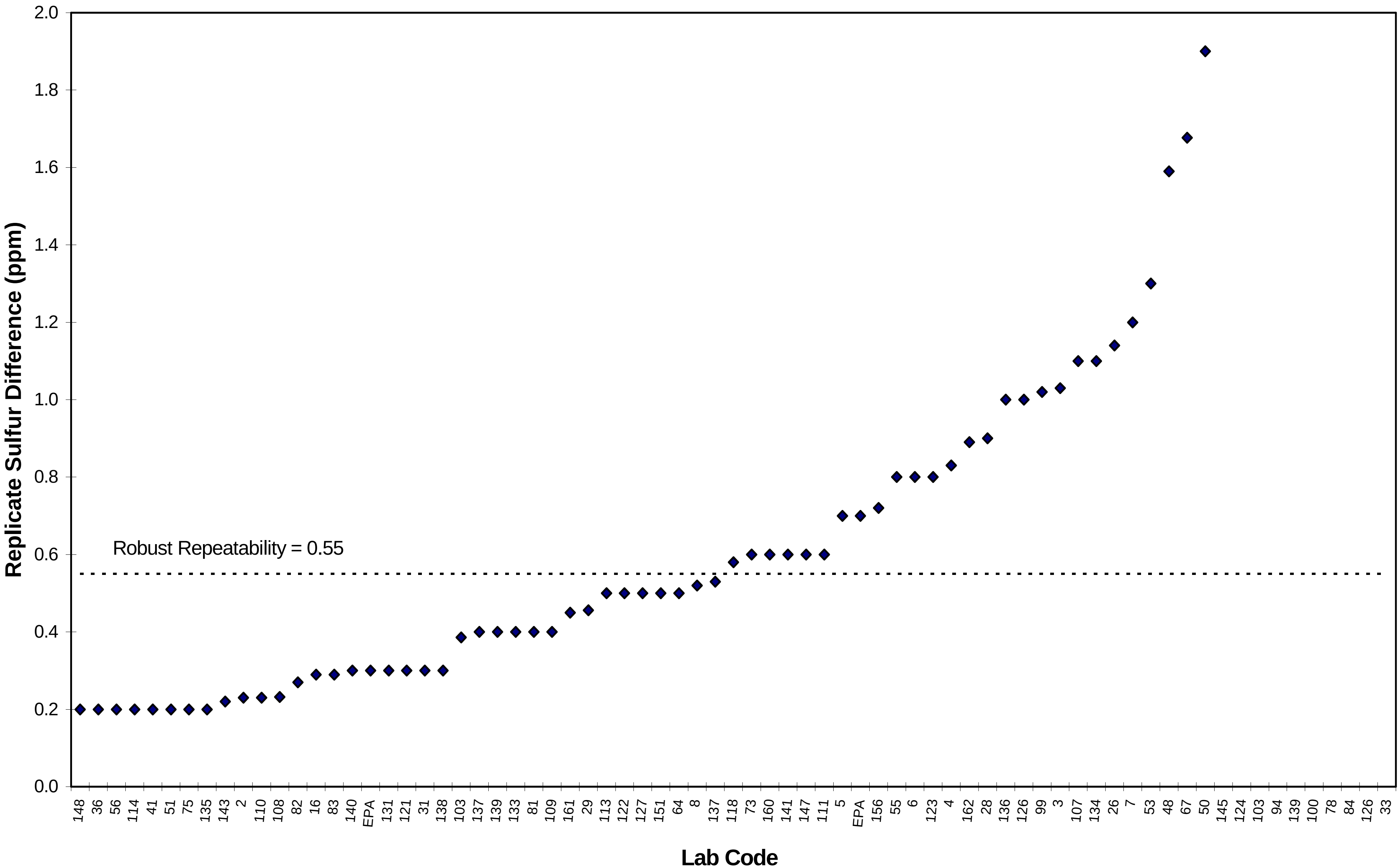


**Figure A-21 August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods In-House Calibration**
**ASTM Analysis, Lab Repeatability**

Codes of deleted labs =
68, 96, 126, 140, 145, 158

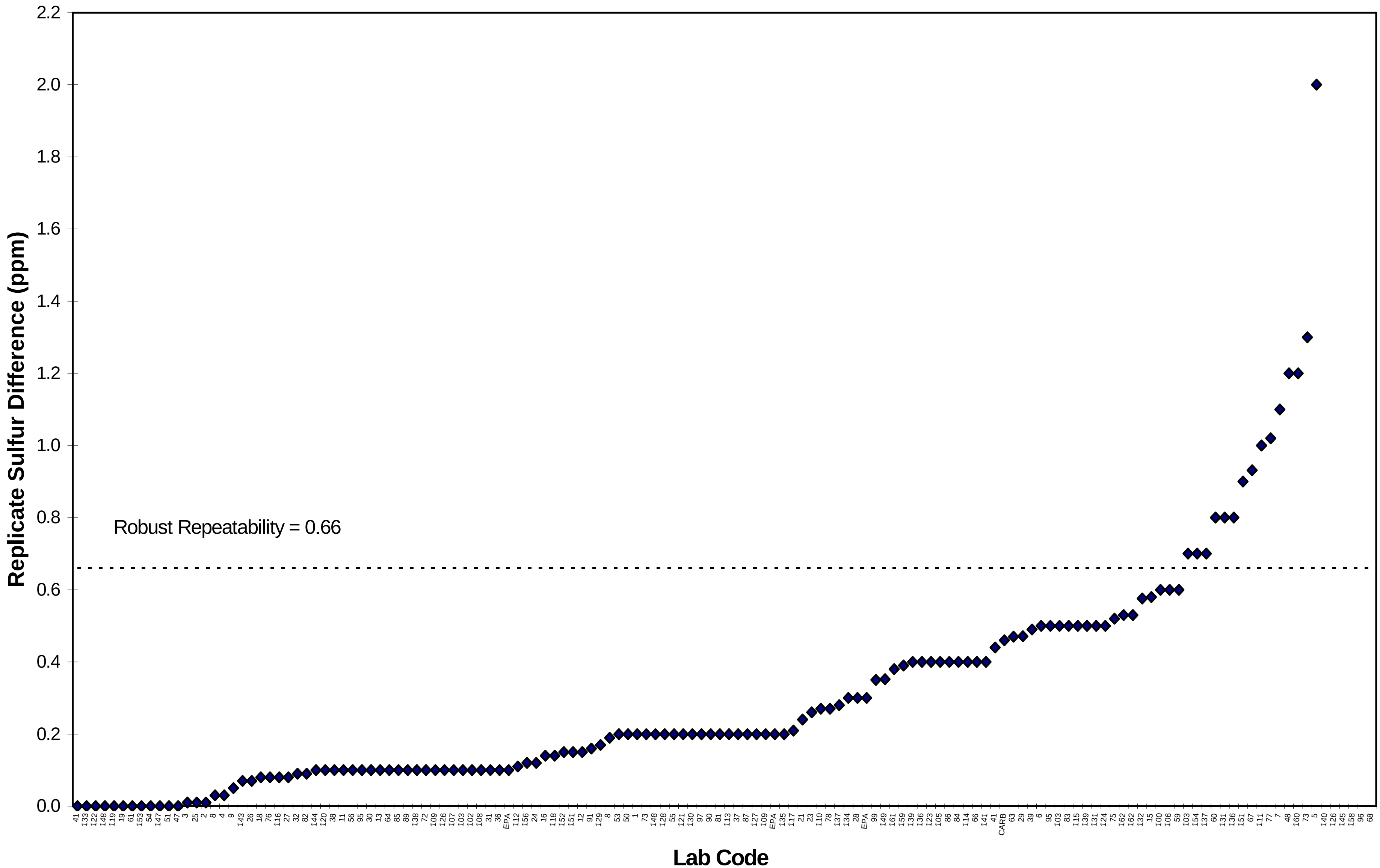


**Figure A-21a August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods In-House Calibration**
**ASTM Analysis, Lab Repeatability**

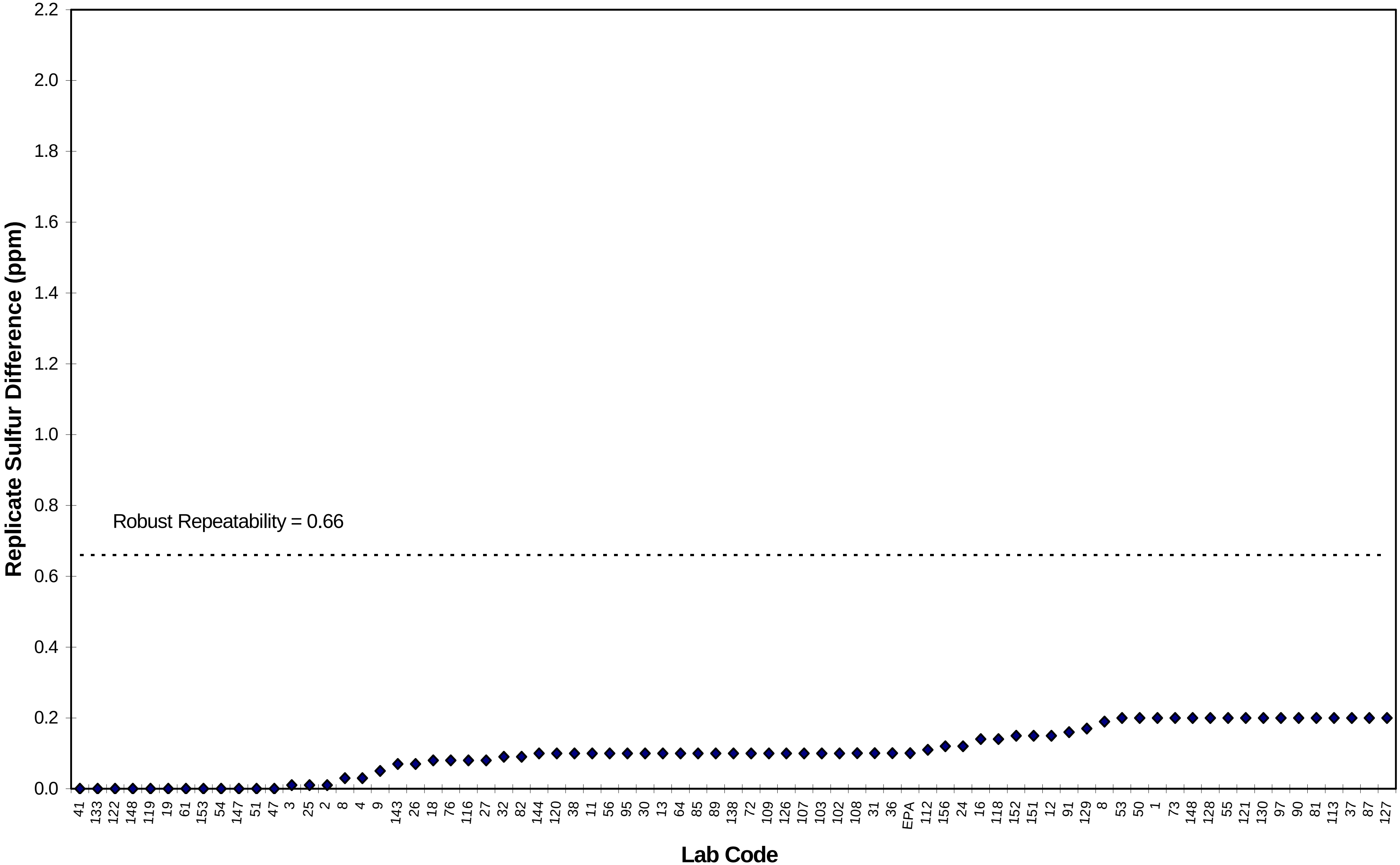


**Figure A-21b   August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods   In-House Calibration**
**ASTM Analysis, Lab Repeatability**

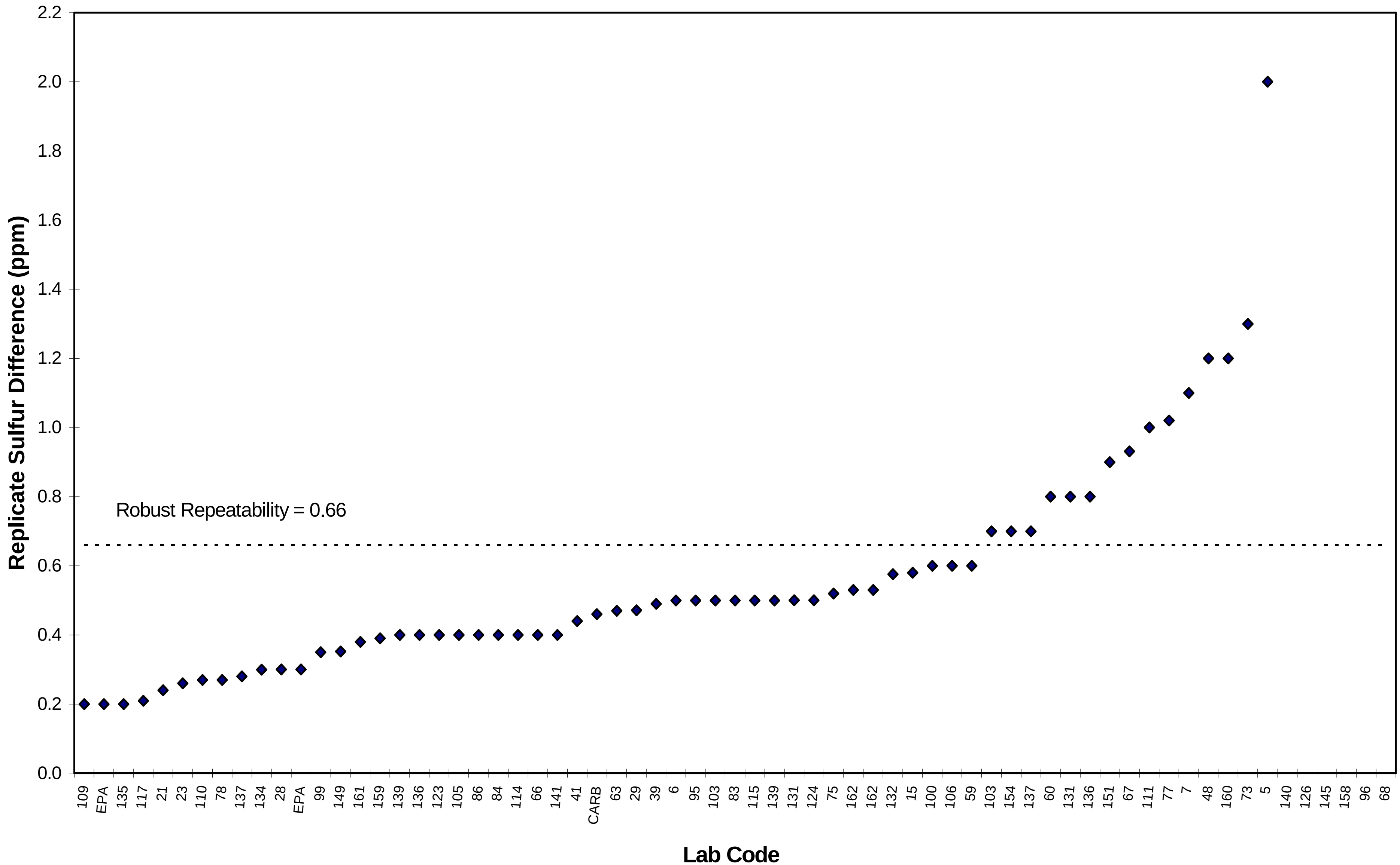


**Figure A-22 August Fuel #5, Robust Outlier Deletion**
**Composite Test Methods NIST Calibration**
**ASTM Analysis, Lab Repeatability**

Codes of deleted labs =
48, 51, 68, 87, 96, 145, 158

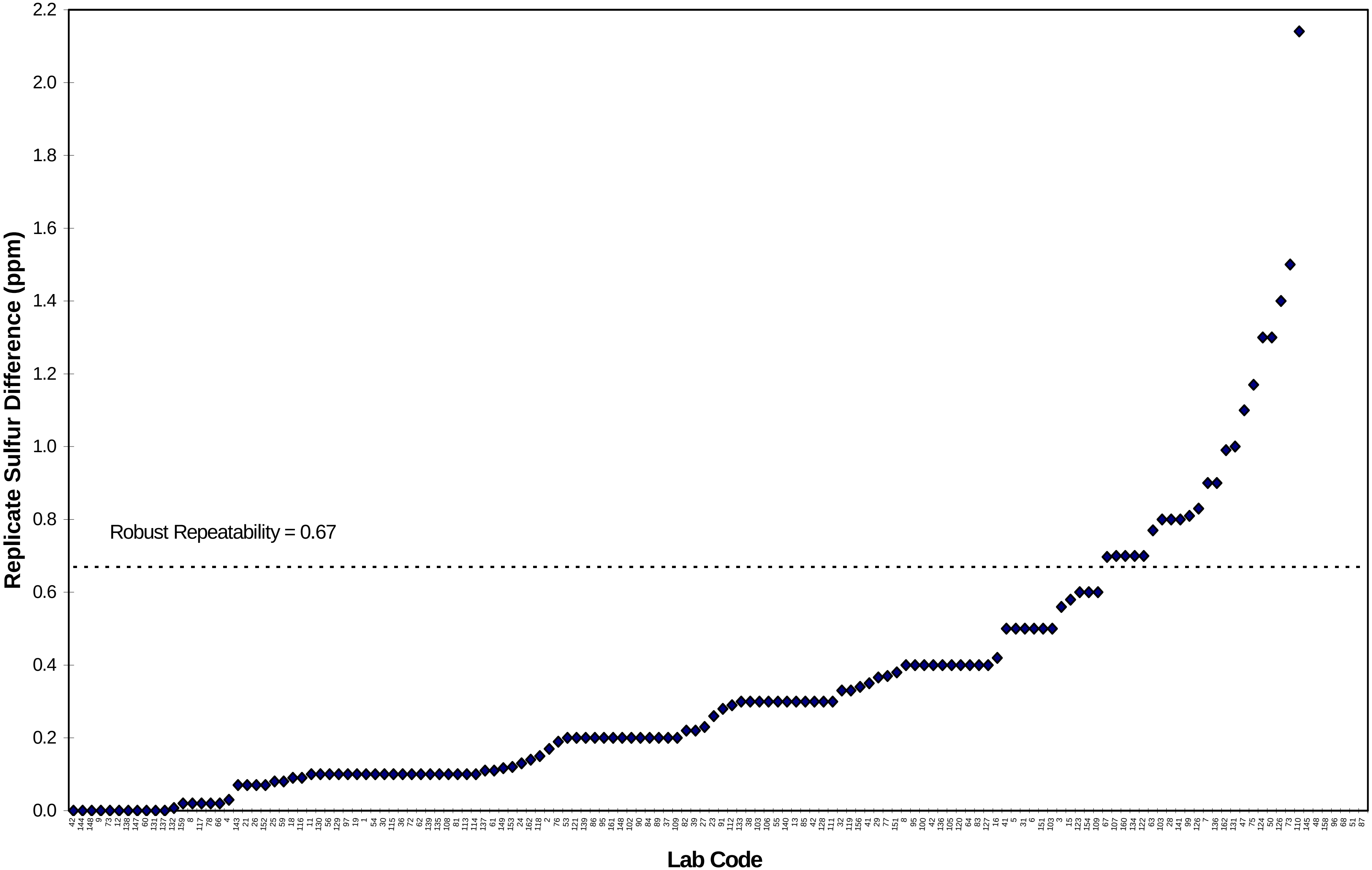
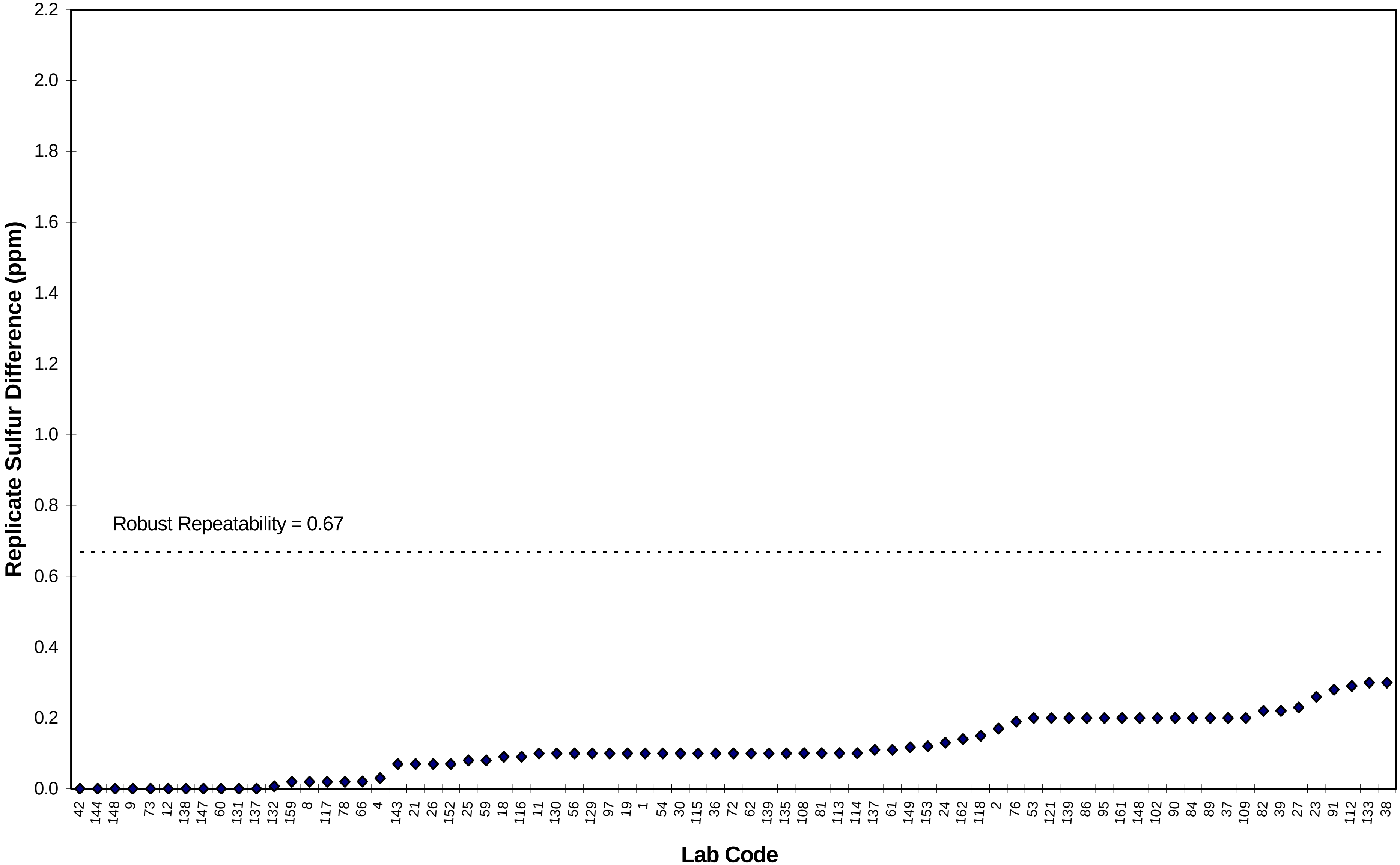
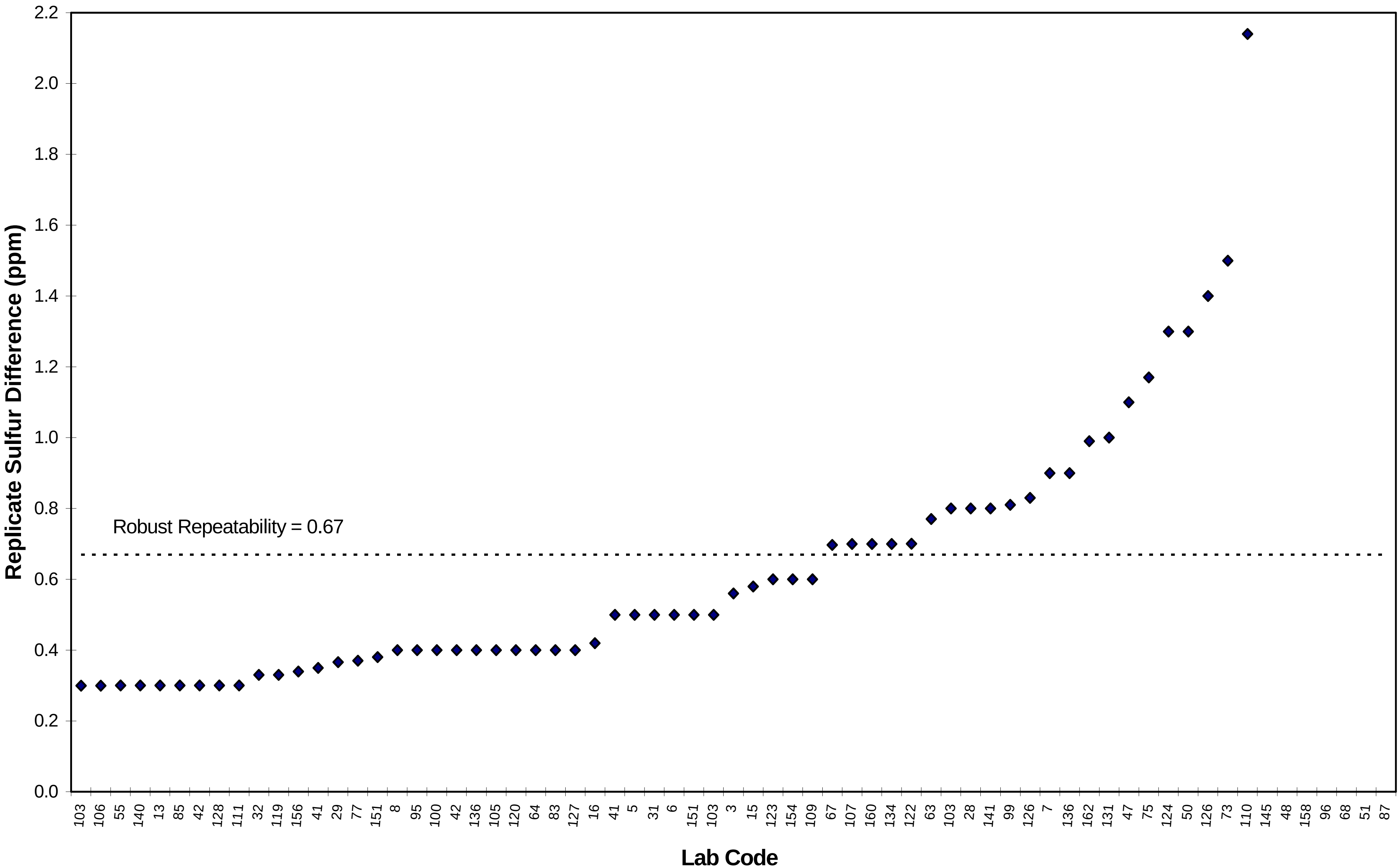


Figure A-22a   August Fuel #5, Robust Outlier Deletion
Composite Test Methods   NIST Calibration
ASTM Analysis, Lab Repeatability



## Figure A-22b August Fuel #5, Robust Outlier Deletion
## Composite Test Methods NIST Calibration
## ASTM Analysis, Lab Repeatability



# *Summary of Lab ASTM Repeatability Comparisons*

- For the July fuel #4 data, 19 labs had differences in their two sulfur repeat values that exceeded the composite ASTM repeatability values using the In-House calibration. This compares to 27 labs when using the NIST calibration.
  - The July fuel #4 data for 3 labs using the In-House method and 10 labs using the NIST method were excluded due to having all 3 repeat values meet the outlier deletion criteria.
- For the August fuel #5, 15 labs had differences in their two sulfur repeat values that exceeded the composite ASTM repeatability values using the In-House calibration. This compares to 22 labs when using the NIST calibration.
  - The August fuel #5 data for 6 labs using the In-House method and 7 labs using the NIST method were excluded due to having all 3 repeat values meet the outlier deletion criteria.

# *Overall Repeatability Comparisons: In-House vs. NIST*

- Compares In-House and NIST calibrations.
  - For D5453, D7039, D2622, EDXRF, and Composite.
  - Uses robust outlier deletion method and ASTM analysis.

- Also compares above results to the 2004-05 ASTM Crosscheck data.
  - For D5453, D7039 and D2622.

**Figure C-6 D5453 Test Method, Robust Outlier, ASTM Analysis Comparing In-House and NIST Calibrations**

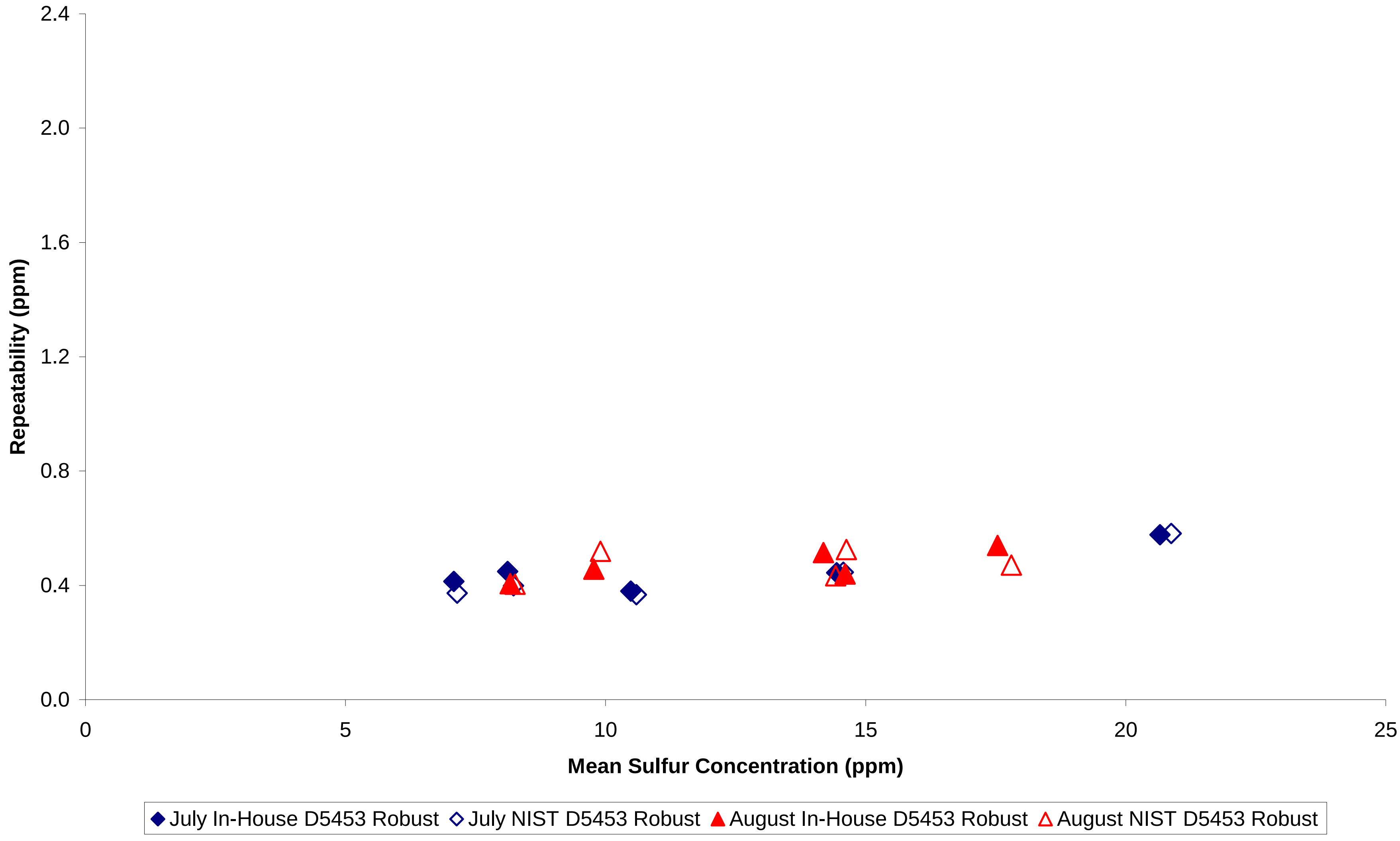


**Figure C-16 D7039 Test Method, Robust Outlier, ASTM Analysis
Comparing In-House and NIST Calibrations**

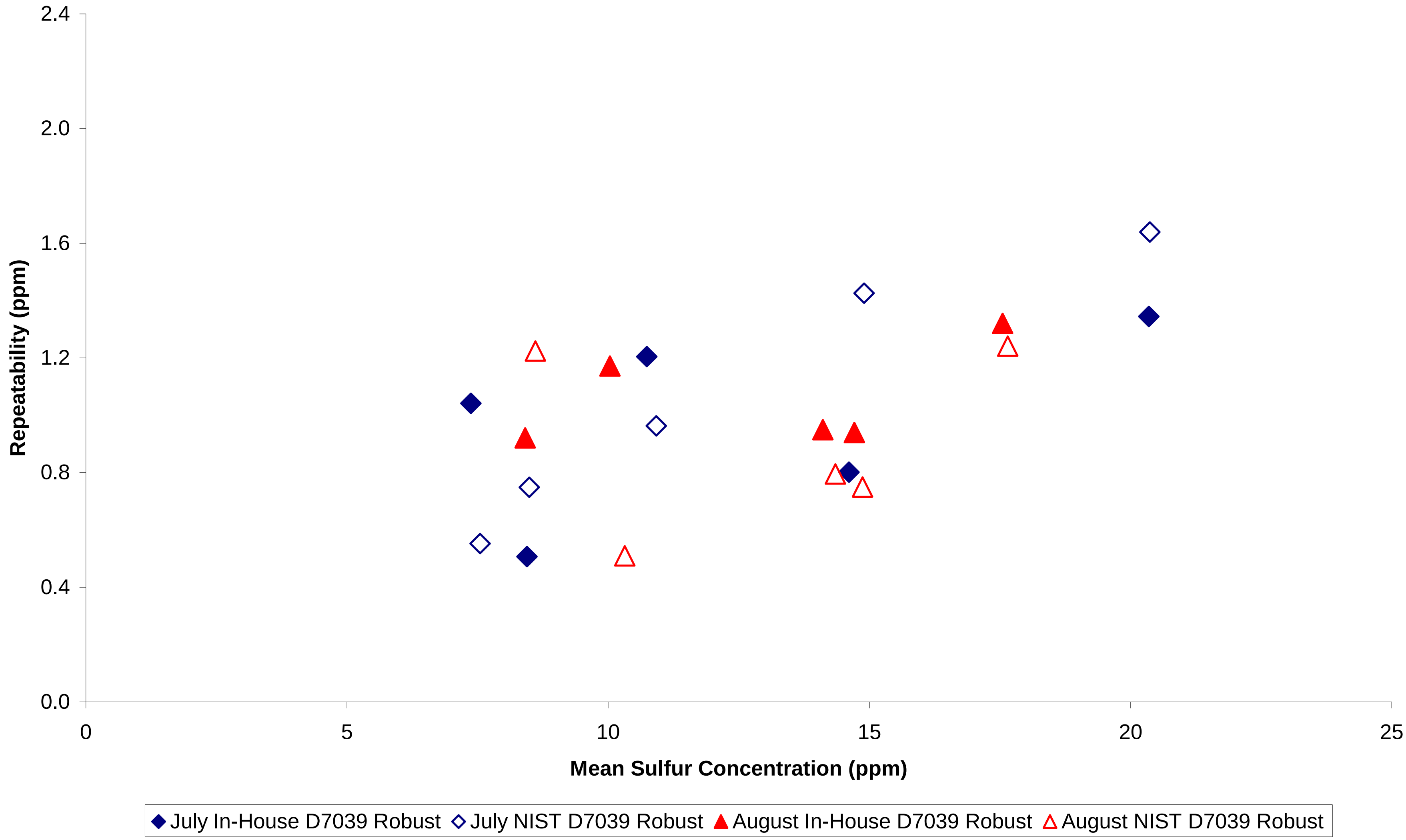


**Figure C-11 D2622 Test Method, Robust Outlier, ASTM Analysis Comparing In-House and NIST Calibrations**

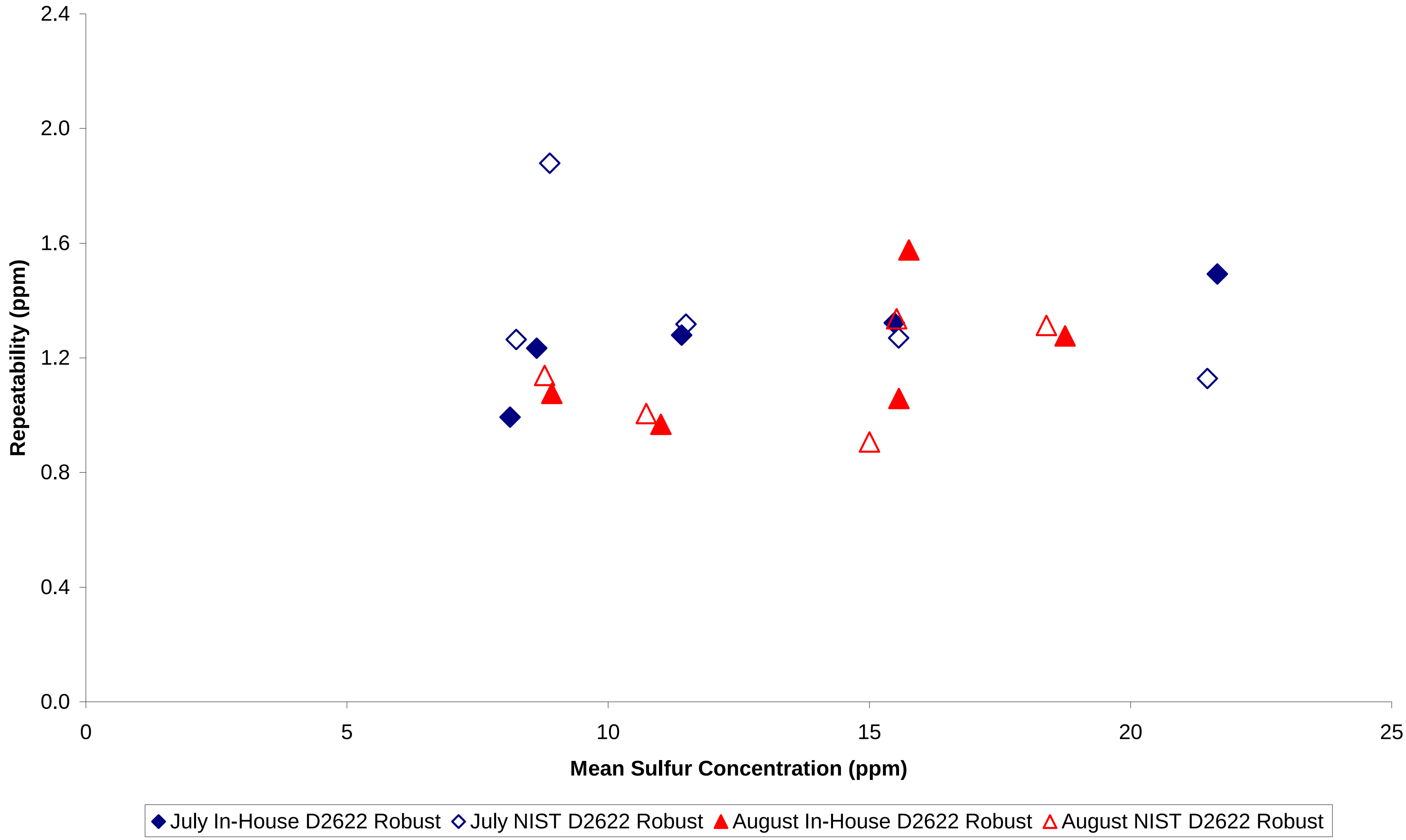
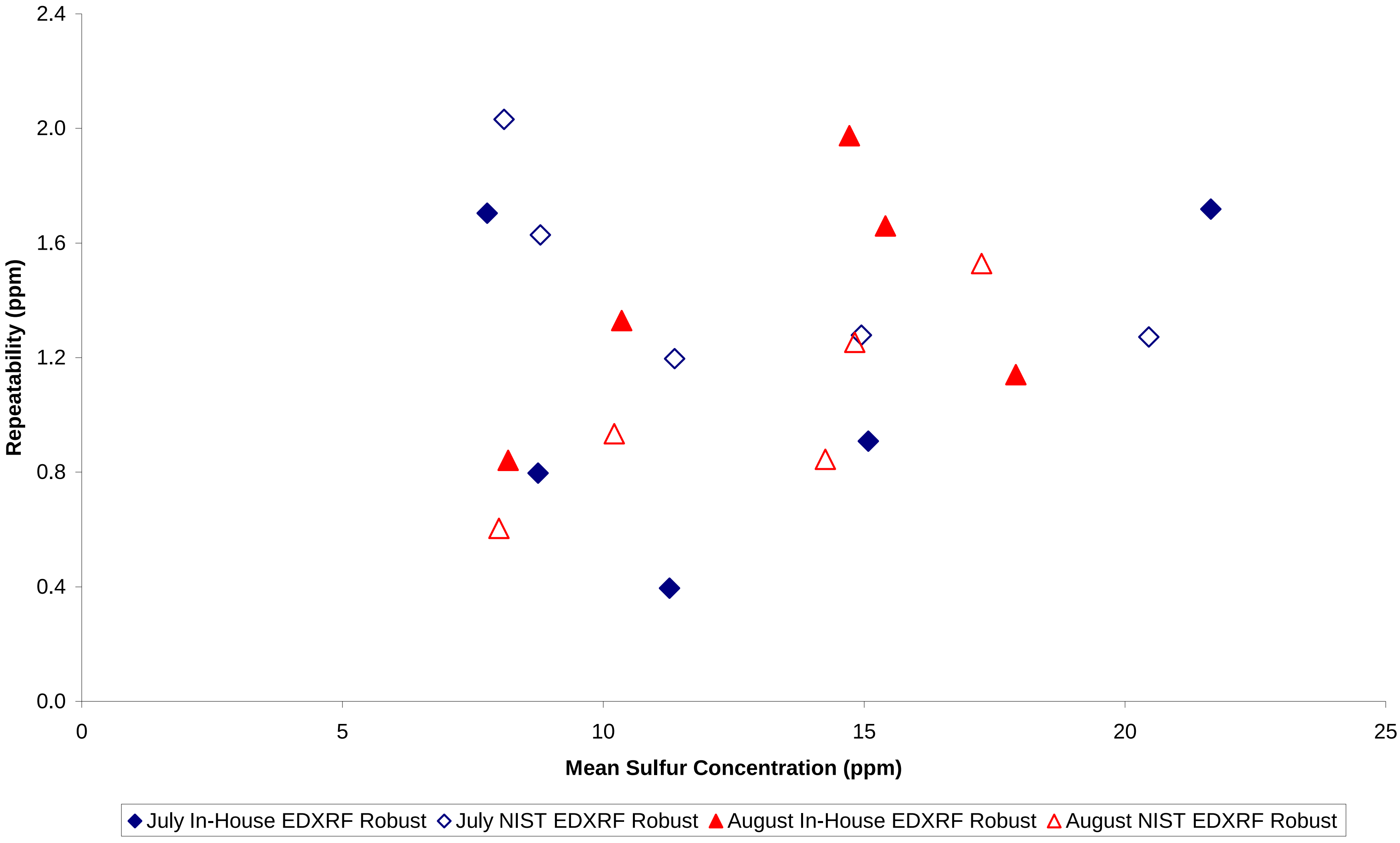
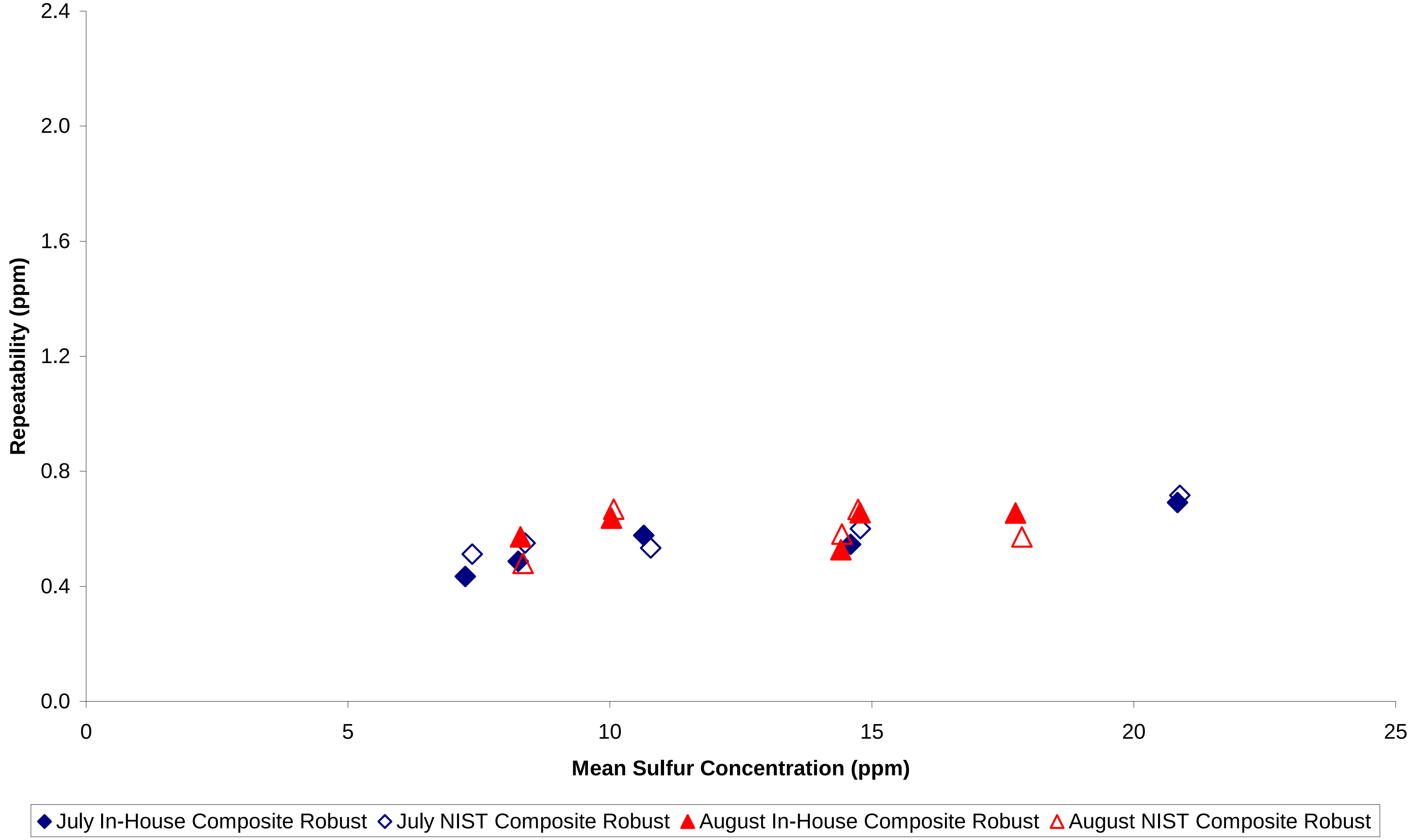


**Figure C-21 EDXRF Test Method, Robust Outlier, ASTM Analysis Comparing In-House and NIST Calibrations**



## Figure C-1 Composite Test Methods, Robust Outlier, ASTM Analysis Comparing In-House and NIST Calibrations



## Figure C-26 D5453 Test Method, Robust Outlier, ASTM Analysis Comparing In-House and NIST Calibrations, Added ASTM 2004-05 Crosscheck Data

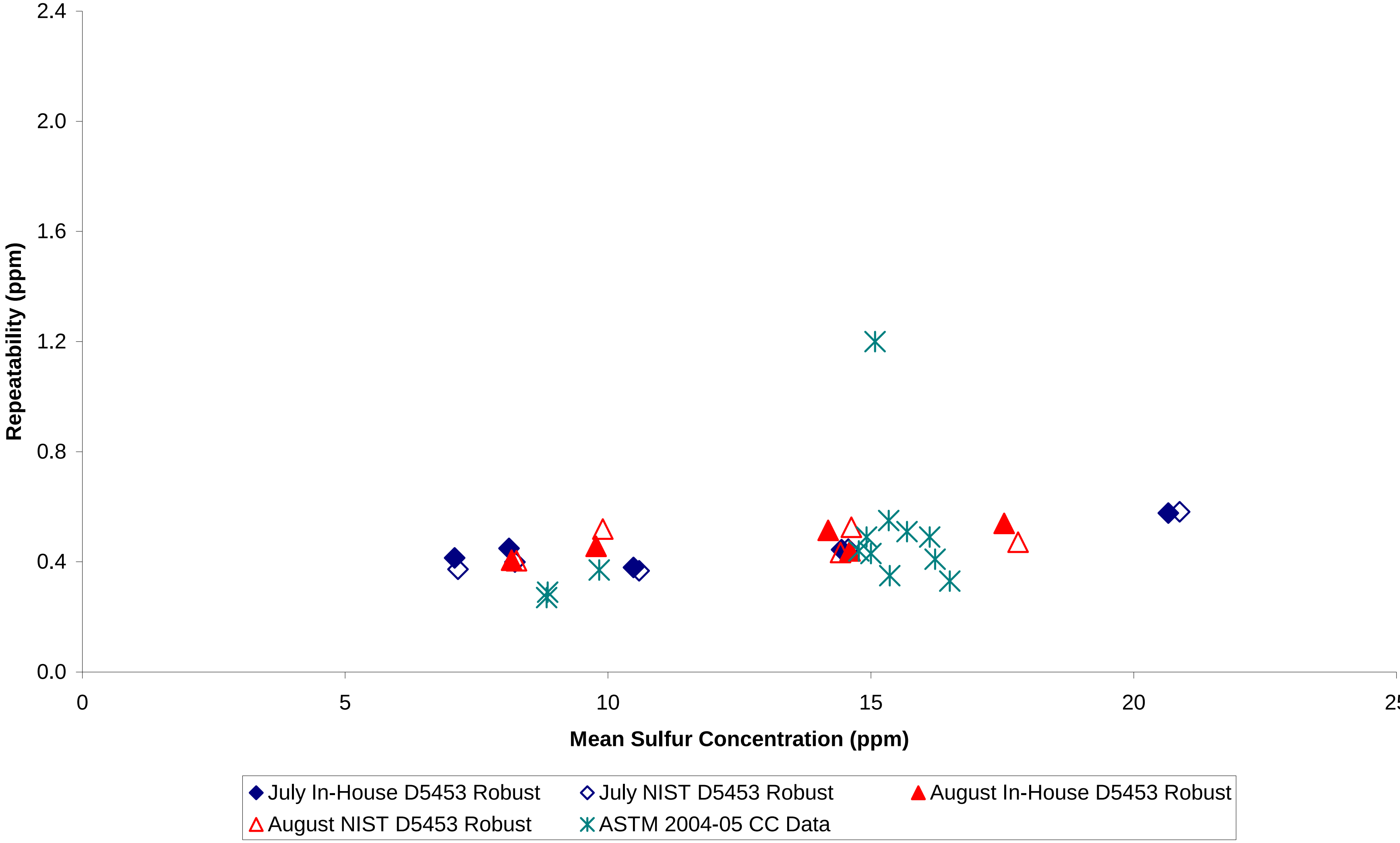
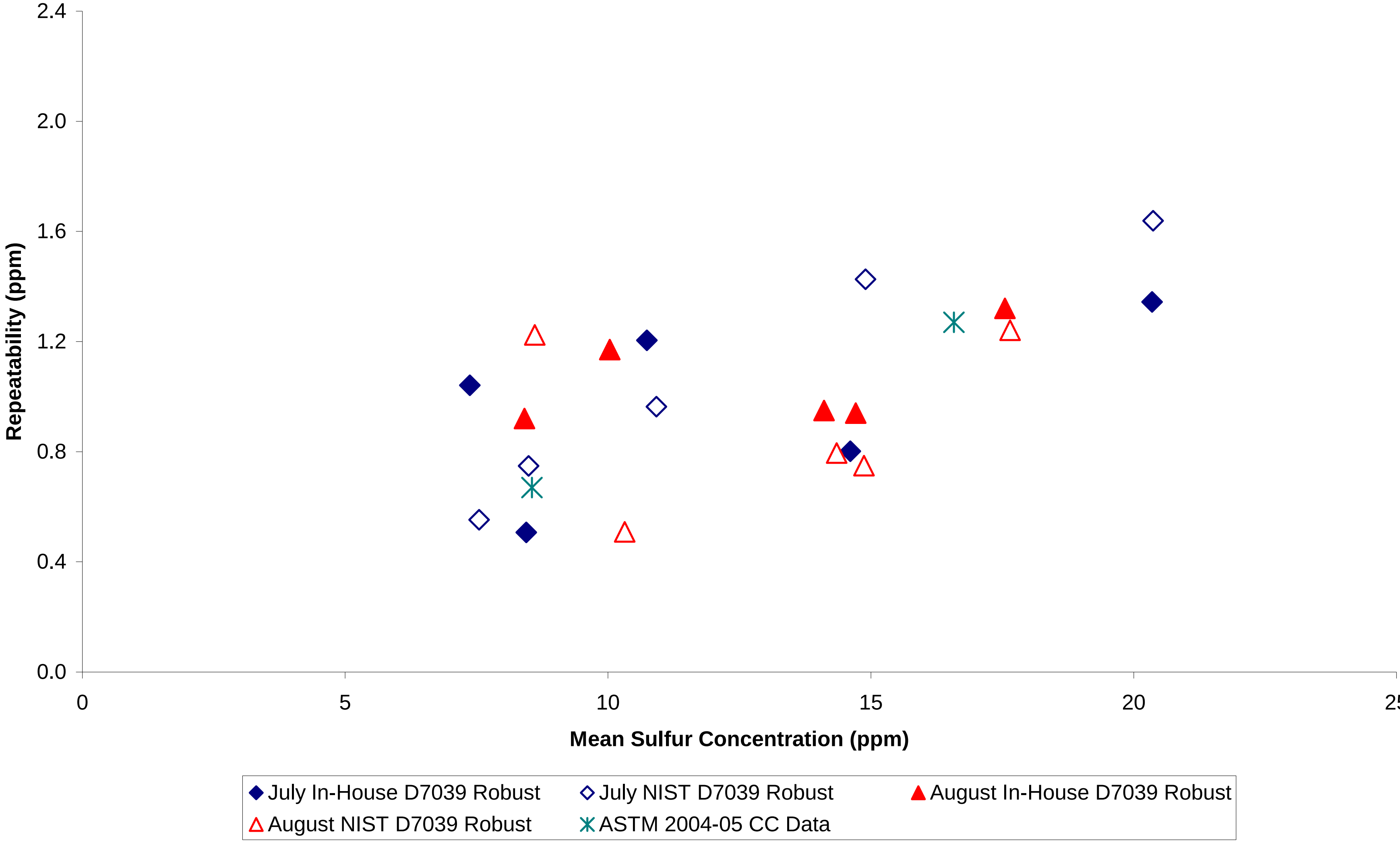


**Figure C-28   D7039 Test Method, Robust Outlier, ASTM Analysis
Comparing In-House and NIST Calibrations, Added ASTM 2004-05 Crosscheck Data**



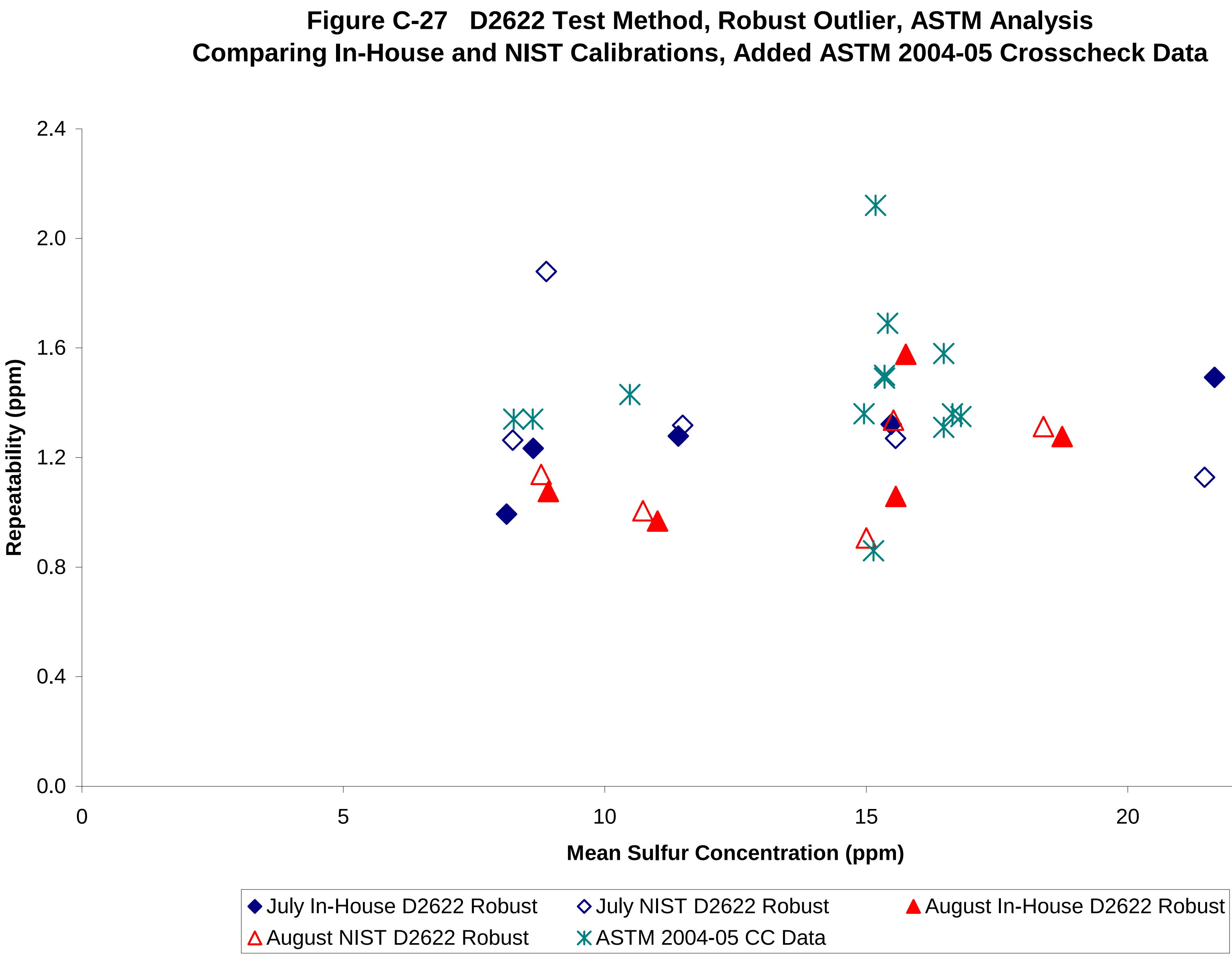


Figure C-27 D2622 Test Method, Robust Outlier, ASTM Analysis
Comparing In-House and NIST Calibrations, Added ASTM 2004-05 Crosscheck Data
Repeatability (ppm)
2.4
2.0
1.6
1.2
0.8
0.4
0.0
0
5
10
15
20
25
Mean Sulfur Concentration (ppm)
July In-House D2622 Robust
July NIST D2622 Robust
August In-House D2622 Robust
August NIST D2622 Robust
ASTM 2004-05 CC Data

# *Summary of Overall Repeatability Comparisons: In-House vs. NIST*

- There was no noticeable difference in the repeatability values across the two calibration methods.
- The smallest r-values and the smallest variation occurred using D5453 and these were slightly smaller than the Composite values. The D7039 and D2622 values were in the middle with the D7039 values being slightly smaller and less variable than the D2622 values but larger than the D5453 values.
- The largest r-values and largest variation occurred using EDXRF. This variation is possibly a result of the small number of labs, 6, using this procedure.
- The 2004-05 Crosscheck r-values were generally in agreement with these results except for one larger CC value that occurred in comparisons using D5453 and in using D2622.