

Test Group Summary Report

C31B Petroleum Hydrocarbons in Soil

January 2026 PT Round

Issued: March 2, 2026

Table of Contents

1.0	The Proficiency Testing Report	1
2.0	Definitions	1
3.0	Scoring System	1
3.1	Homogeneity and Stability Assessment	2
3.2	The z score	2
3.2	Composite (PT) Score	2
3.3	Identifying Bias	2
3.4	Deviations from Evaluation Procedure	3
4.0	PT Round Specific Data Summary	3
4.1	Summary statistics	3
4.2	z - Score Plots	3
4.3	kernel density plots.....	3
4.4	stability and homogeneity Plots.....	3
4.5	Box-and-Whisker Plots	3
4.6	Historic Comparison Plot	3
	Annex A Summary by Analyte	4

1.0 The Proficiency Testing Report

The Proficiency Testing Report consists of two parts.

- *PTC Proficiency Testing Report*: This report contains participant-specific data and other confidential information. This report is emailed to participants at the end of the PT round.
- *Test Group Summary Report*: A Test Group Summary Report is created for each quantified test group at the end of the PT round. These reports contain more detailed information on the round than is found in the participant-specific PTC Proficiency Testing Report. These reports do not contain any confidential information and are made available on the PTC web site.

2.0 Definitions

The participant-specific PTC Proficiency Testing Report contains some terms that new participants may not be familiar with.

<i>Code</i> :	The registration code that is unique to each analyte that a participant is registered for.
<i>App</i> :	If a participant is accredited by CALA, this three-digit number is the appendix number that the accredited method is assigned to.
<i>N</i> :	The number of participants results that were used to calculate the summary statistics. This excludes qualified data (e.g., <) and any results that were flagged as outliers.
<i>Assigned</i> :	The Assigned Value is the robust mean of the reported results, outliers excluded. This is often referred to as the “target” value.
<i>±u</i> :	The uncertainty of the assigned value.
<i>Reported</i> :	The result reported by the participant.
<i>s</i> :	The Standard Deviation of Proficiency Assessment (SDPA). This value is used to determine the acceptance limits for the PT evaluation.
<i>z-Score</i> :	A value assigned to each reported result that is a measure of the degree to which it deviates from the Assigned Value.
<i>Score</i> :	The composite score of the four results reported for each analyte. It is normalized to a score out of 100.
<i>Bias</i> :	A flag assigned if bias is detected using the re-scaled z-score procedure.

3.0 Scoring System

Participant performance is evaluated for each proficiency testing sample by a quantitative method that is consistent with ISO/IEC 17043 – *Conformity assessment- General requirements for the competence of proficiency testing providers*, the *International Harmonized Protocol for Proficiency Testing of (Chemical) Analytical Laboratories* (2006), and ISO 13528:2015 *Statistical methods for use in proficiency testing by interlaboratory comparisons*.

The following is a brief description of the evaluation procedure used by PTC. The detailed evaluation procedure is described in PROC09 – *PT Evaluation Procedure*, which is available on the PTC website (www.PTCCanada.org).

3.1 HOMOGENEITY AND STABILITY ASSESSMENT

Homogeneity and stability are assessed using participant data. Regression analysis is performed on reported result against order of sample production (Homogeneity) and reported result against date of analysis (Stability). If the slope is significantly different than zero for either then the Standard Deviation of Proficiency Assessment (s) is increased to minimize the impact.

3.2 THE Z SCORE

A "z-score" is calculated for each reported result as follows:

$$z - Score = \frac{(x - \bar{X})}{SDPA} \quad \text{where: } \begin{array}{l} x = \text{participant result;} \\ \bar{X} = \text{the Assigned Value;} \\ SDPA = \text{the Standard Deviation for Proficiency Assessment.} \end{array}$$

The assigned value \bar{X} is generally estimated from the inter-laboratory Robust mean after outliers due to obvious gross errors (e.g., reported in wrong units) have been removed.

The Standard Deviation for Proficiency Assessment, s, is determined as follows:

- The inter-laboratory Robust standard deviation ($Stdev_{rob}$) is calculated using reported results, obvious outliers removed;
- The regression equation standard deviation ($Stdev_{reg}$) is estimated from regression equations derived from previous studies (see PROC11- *PT Regression Equations* for details);
- The SDPA is the higher of $Stdev_{rob}$ and $Stdev_{reg}$;
- When a laboratory reports its detection limit, s will be estimated using a pooled variance procedure that uses both the inter-laboratory data and the reported detection limit.

3.2 COMPOSITE (PT) SCORE

Since each PT round involves four or two separate samples of distinct concentration for each test, it is necessary to calculate a composite PT score for each test to determine overall performance. The composite score is calculated by first averaging the absolute z-scores for the four results and then calculating a final score as $100 + (-15 \times \text{avg } |z|)$.

Acceptable PT Scores equal or exceed 70.

3.3 IDENTIFYING BIAS

The proficiency testing report provides flags for bias. These are determined using the re-scaled z-score procedure.

$$RSZ = \frac{\sum z}{\sqrt{N}} \quad \text{where } \begin{array}{l} z = \text{the z- score} \\ N = \text{the number of samples} \end{array}$$

Flags are assigned for each test group/parameter combination as follows:

$$RSZ \geq -2 \text{ and } \leq 2 \text{ no flag assigned}$$

RSZ > 2	H (High)
RSZ > 3	VH (Very High)
RSZ < -2	L (LOW)
RSZ < -3	VL (Very Low)

3.4 DEVIATIONS FROM EVALUATION PROCEDURE

Other than changes to the Standard Deviation of Proficiency Assessment due to homogeneity or stability flags, any deviation from the published evaluation procedure is described on the cover page(s) of the final *PTC Proficiency Testing Report*.

4.0 PT Round Specific Data Summary

The following pages provide more detailed information about the PT round indicated in the cover page of this report than is found in the participant-specific PTC Proficiency Testing Report. The graphical representations and the statistical summaries are based upon the data after outliers have been removed.

4.1 SUMMARY STATISTICS

In addition to some of the statistics found in the customer reports, this table includes additional summary statistics such as Median, different measures of dispersion, the number of outliers removed, the number of results in the Questionable range ($|z|$ between 2 and 3) and the Unacceptable range ($z > 3$), and whether a data set was flagged for Homogeneity or Stability. This section also includes sorted scatter plots of the data for each sample.

4.2 z - SCORE PLOTS

The z-scores for each sample are ranked in increasing order and plotted. When the data is normally distributed, the plot should show a slight sigmoidal curve, with an equal number of points above zero as below. Each bar in these plots is colour-coded to indicate the analytical method used by the participant.

4.3 KERNEL DENSITY PLOTS

Kernel density plots are generated for each data set. These plots are a graphical way to represent the overall data distribution and are used to visualize possible deviations from normality and unimodality.

4.4 STABILITY AND HOMOGENEITY PLOTS

Plots of reported result against analysis date, and reported result against order of bottling are displayed, along with the regression line. These regression analyses are used to determine if the SDPA should be adjusted due to homogeneity or stability.

4.5 BOX-AND-WHISKER PLOTS

Box-and-Whisker plots are another way to display the distribution of the data. The box denotes the first and third quartile and the whiskers are the 5th and 95th percentile.

4.6 HISTORIC COMPARISON PLOT

The Historic Comparison Plot is a plot of robust mean against robust standard deviation for the previous ten PT rounds as well as the current PT round. This plot can be used to identify possible changes in the sample formulation.

F2: C10-C16

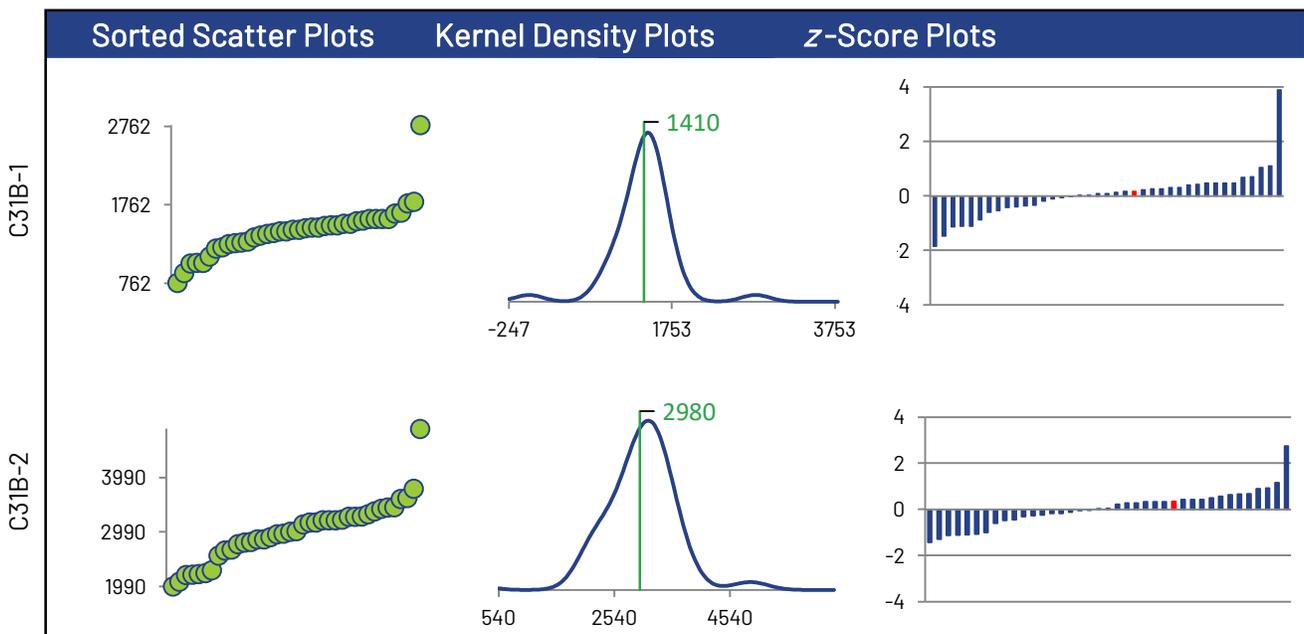
Summary Statistics

Statistic	C31B-1	C31B-2	C31B-3	C31B-4
No. of results used for summary statistics	39	39	39	39
Median mg/kg	1440	3010	3760	403
Robust Mean mg/kg	1410	2980	3810	396
Uncertainty of the Assigned Value mg/kg	45.0	104	103	15.3
Robust Standard Deviation mg/kg	225	521	516	76.2
Regression Standard Deviation mg/kg	351	696	879	127
Stability Flag				
Homogeneity Flag				
Standard Deviation Used (SDPA) mg/kg	351	696	879	127
Flagged Outliers	0	0	0	0
$ z > 3.0$	1	0	0	1
$2 < z < 3$	0	1	1	2
Failure Rate %*	2.6			

Methods Used

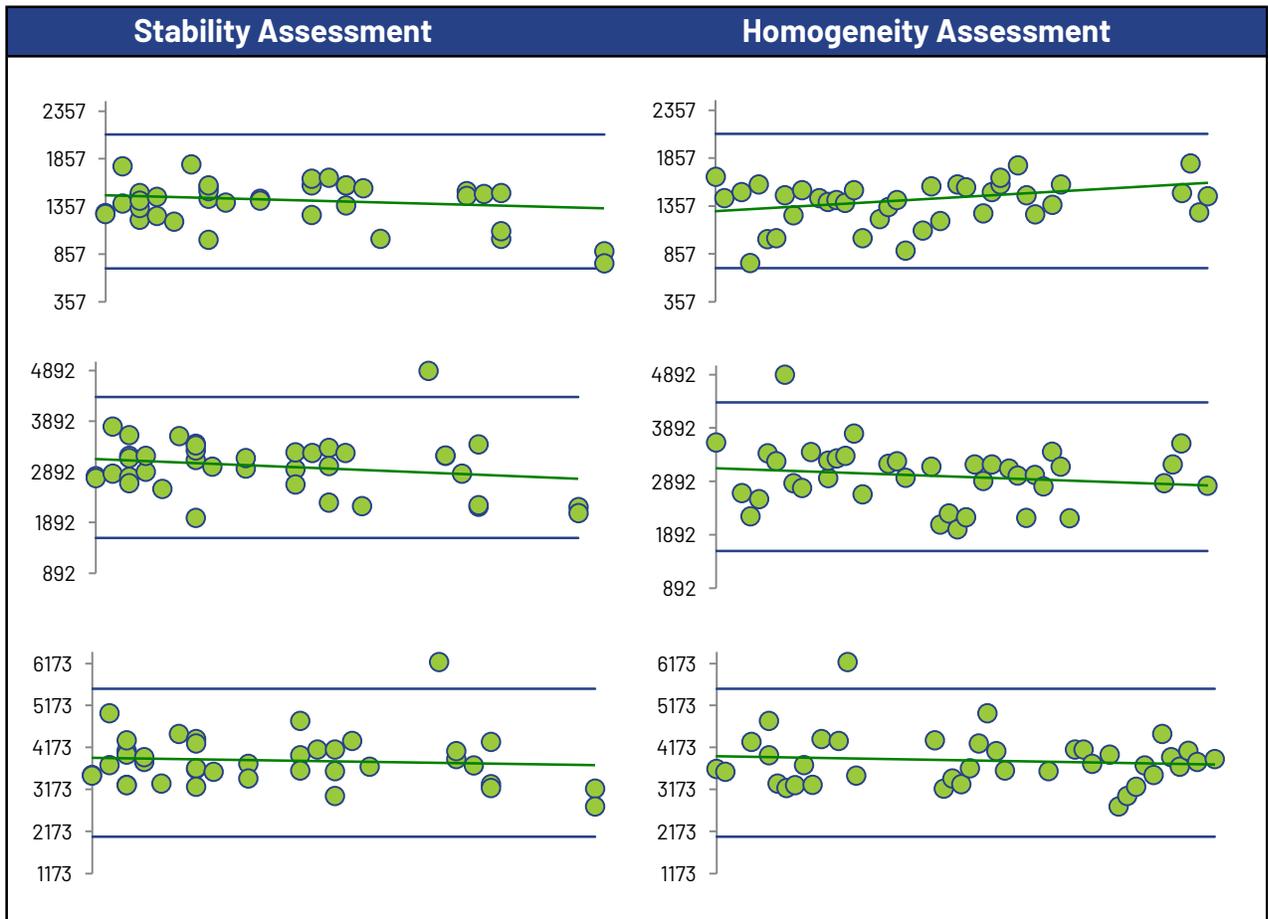
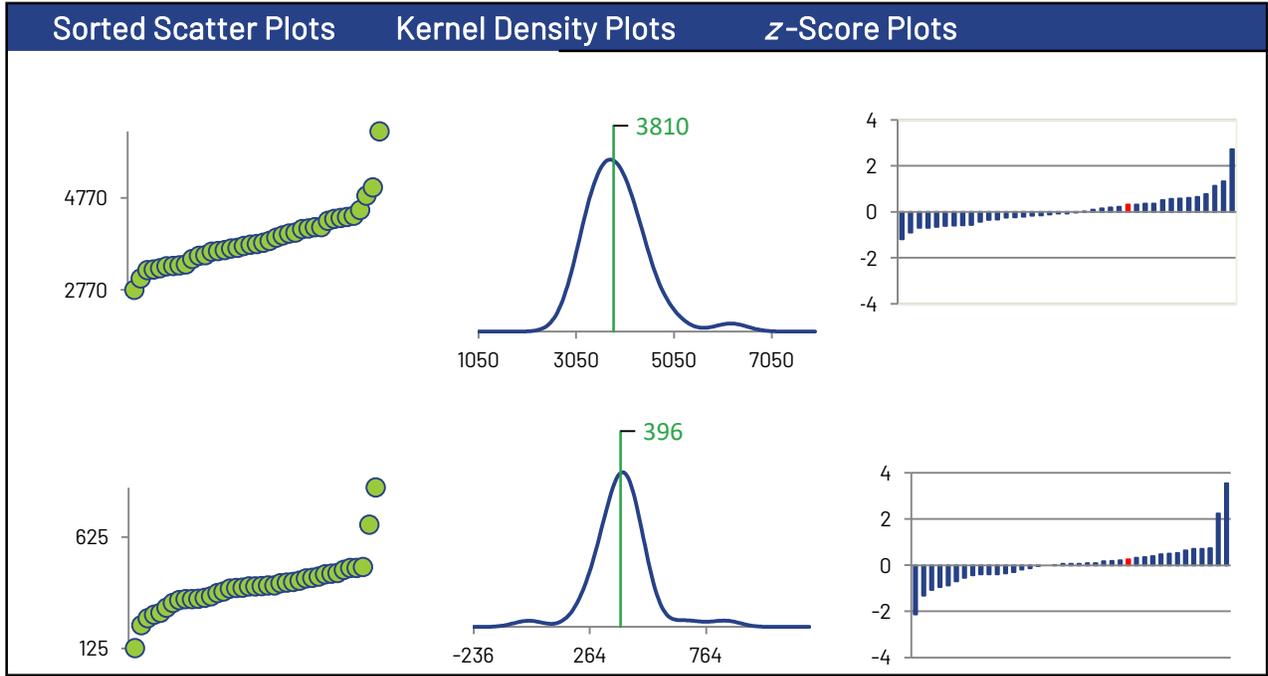
Method	C31B-1	C31B-2	C31B-3	C31B-4
GC/FID (Blue)	38	38	38	38
GC/MS (Red)	1	1	1	1

All summary stats, except the failure rate, and the plots below are based on the data excluding any flagged outliers

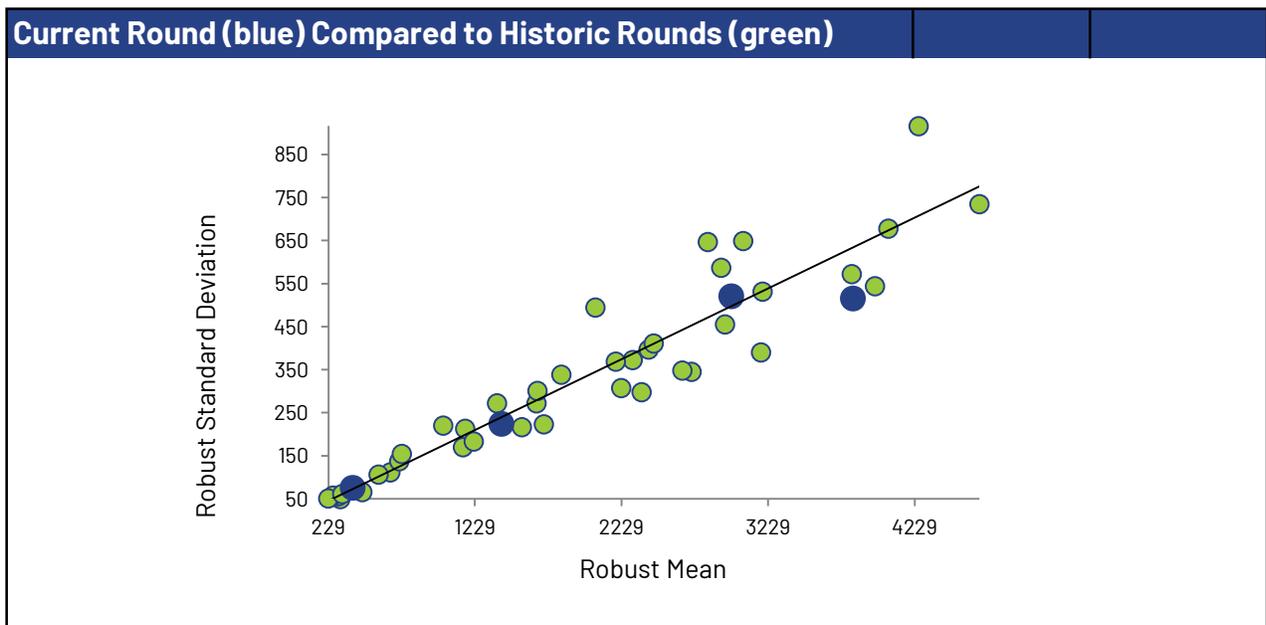
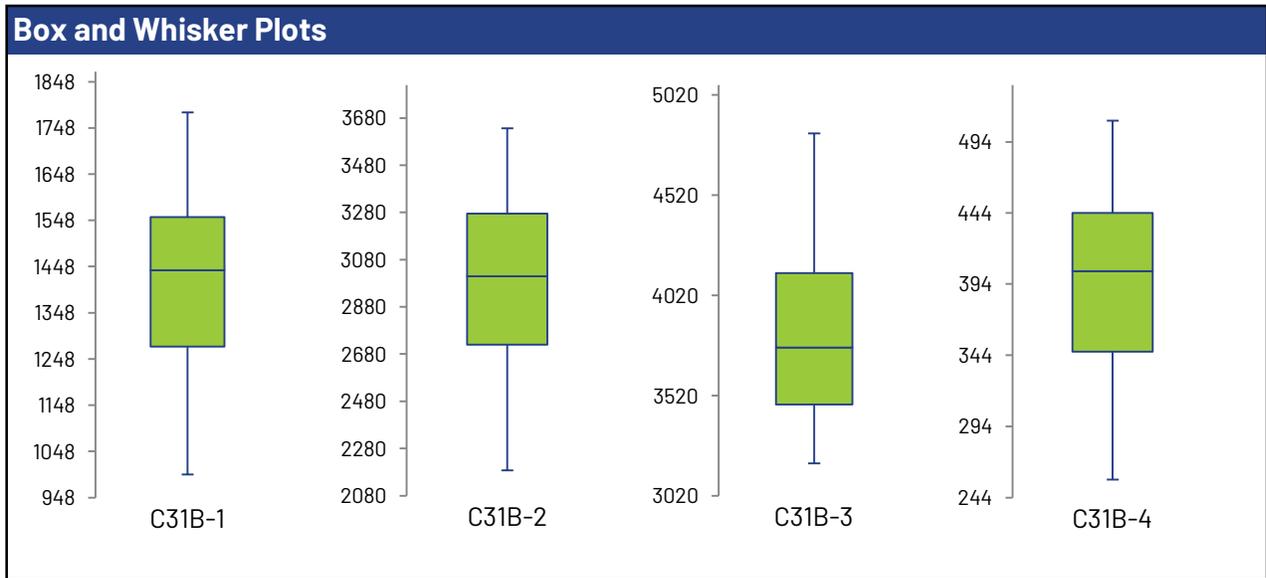
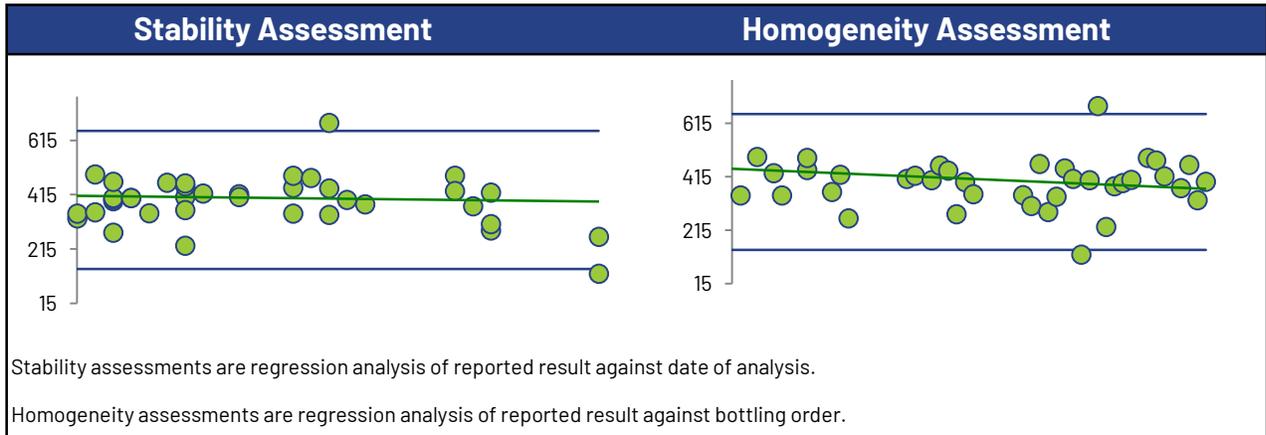


* Displayed failure rate does not account for the RDL modified z-Scores or inaccurately reported non-detects.

F2: C10-C16



F2: C10-C16



F3: C16-C34

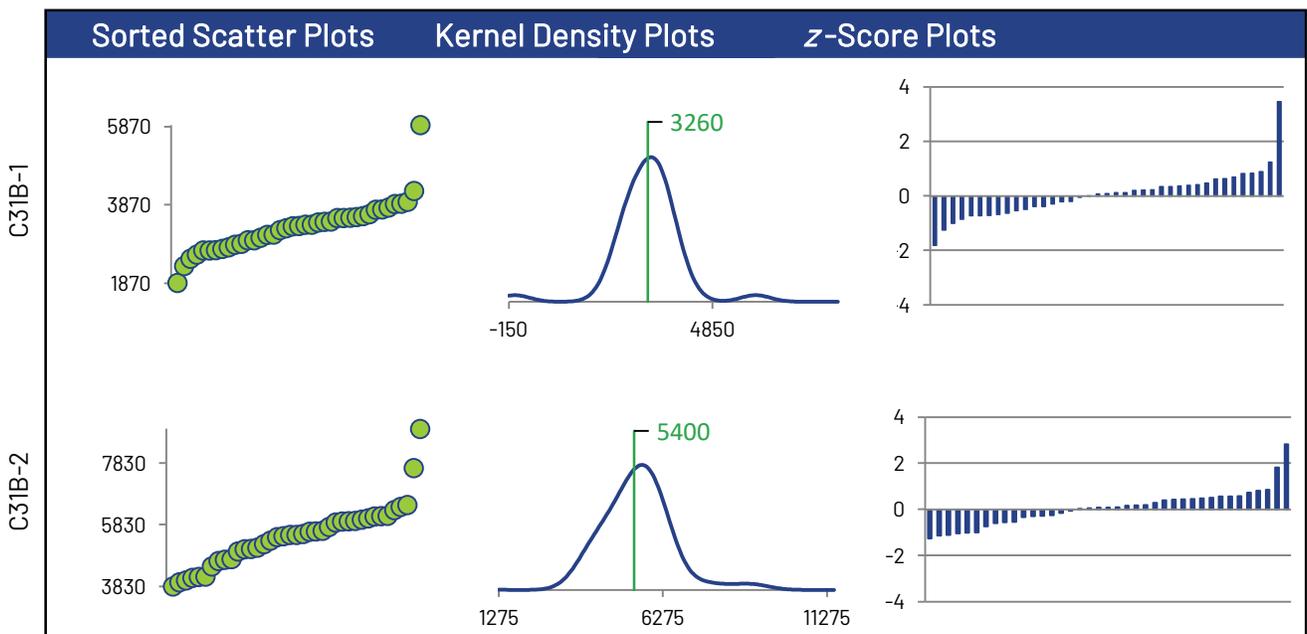
Summary Statistics

Statistic	C31B-1	C31B-2	C31B-3	C31B-4
No. of results used for summary statistics	39	39	39	39
Median mg/kg	3330	5500	7470	1250
Robust Mean mg/kg	3260	5400	7320	1270
Uncertainty of the Assigned Value mg/kg	106	178	198	49.8
Robust Standard Deviation mg/kg	530	890	988	249
Regression Standard Deviation mg/kg	766	1250	1690	315
Stability Flag				
Homogeneity Flag				
Standard Deviation Used (SDPA) mg/kg	766	1250	1690	315
Flagged Outliers	0	0	0	0
$ z > 3.0$	1	0	0	0
$2 < z < 3$	0	1	0	0
Failure Rate %*	2.6			

Methods Used

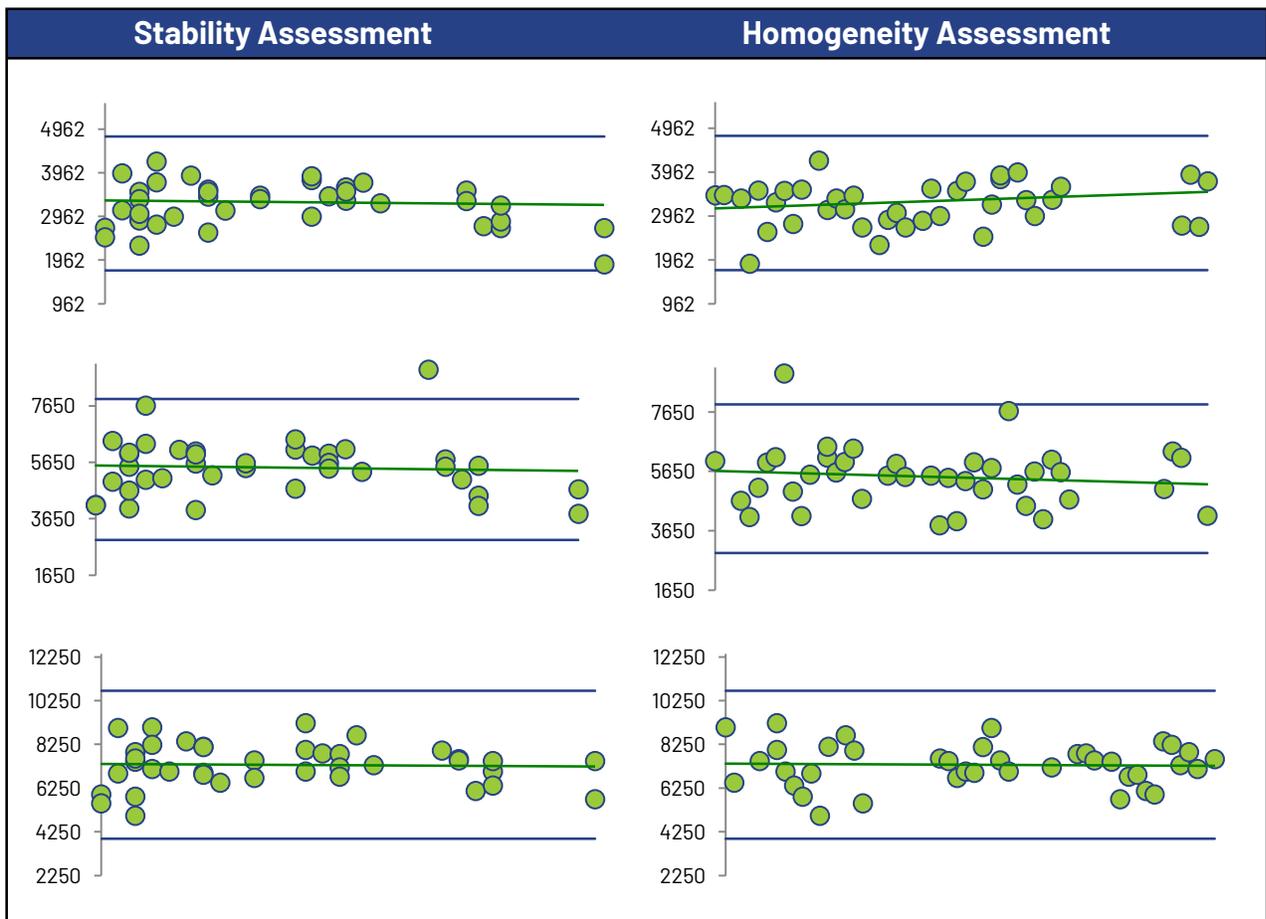
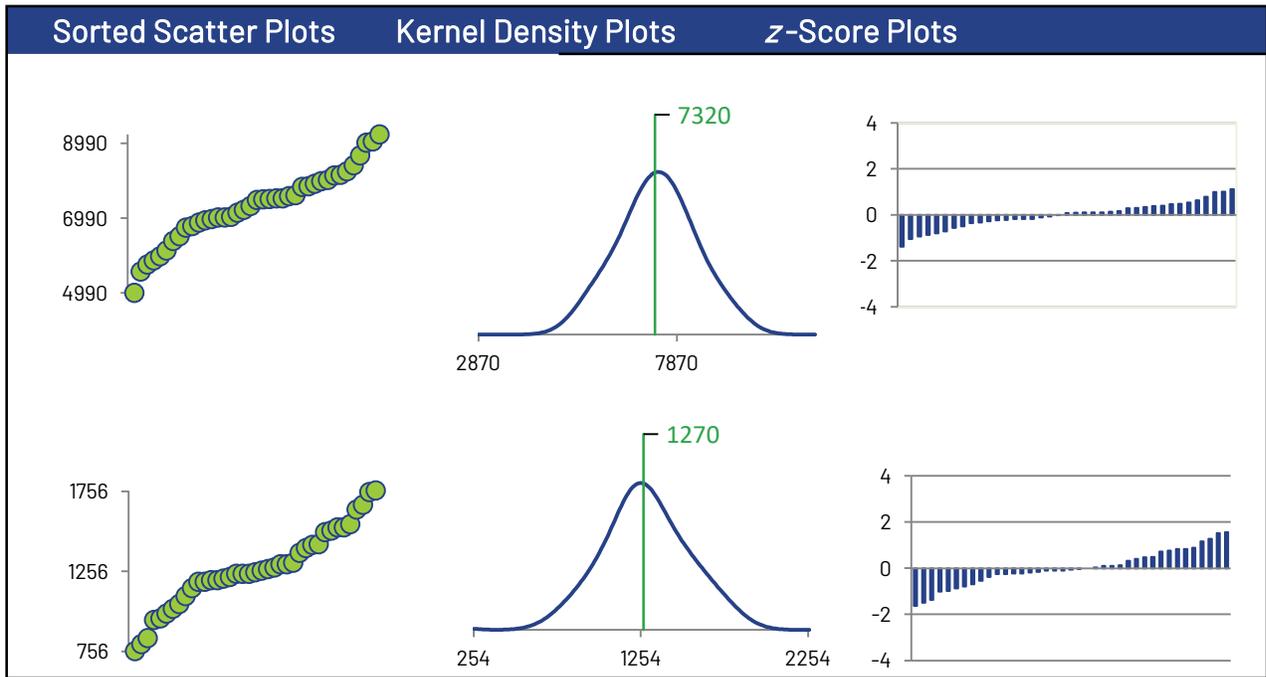
Method	C31B-1	C31B-2	C31B-3	C31B-4
GC/FID (Blue)	39	39	39	39

All summary stats, except the failure rate, and the plots below are based on the data excluding any flagged outliers

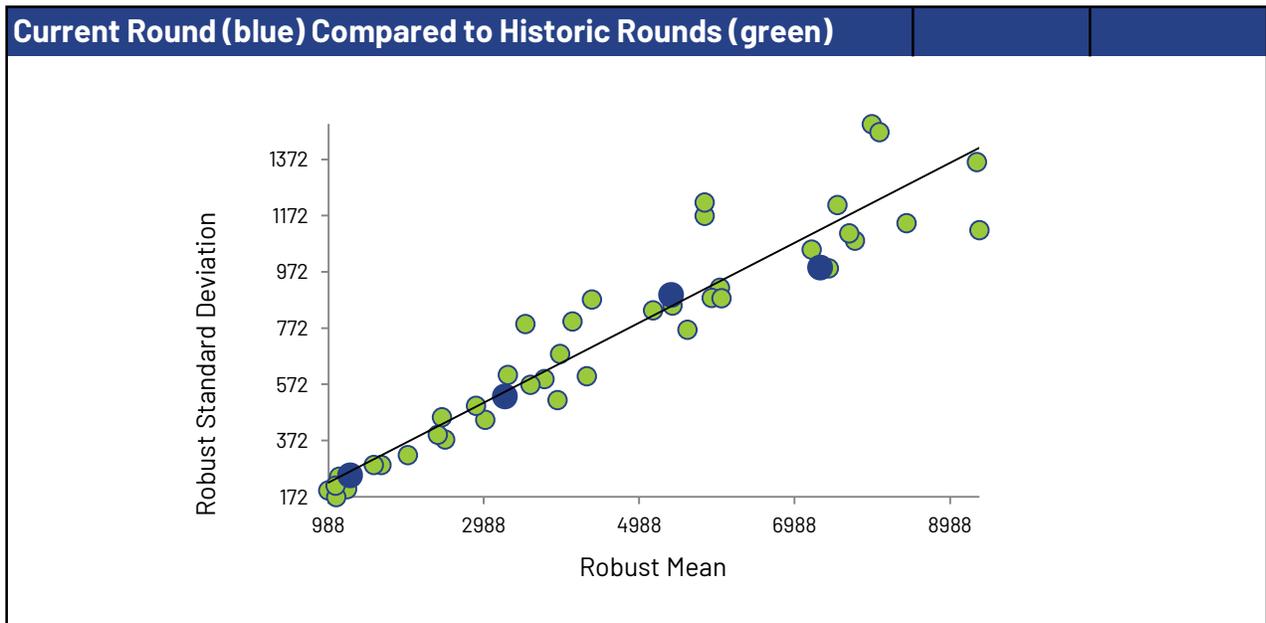
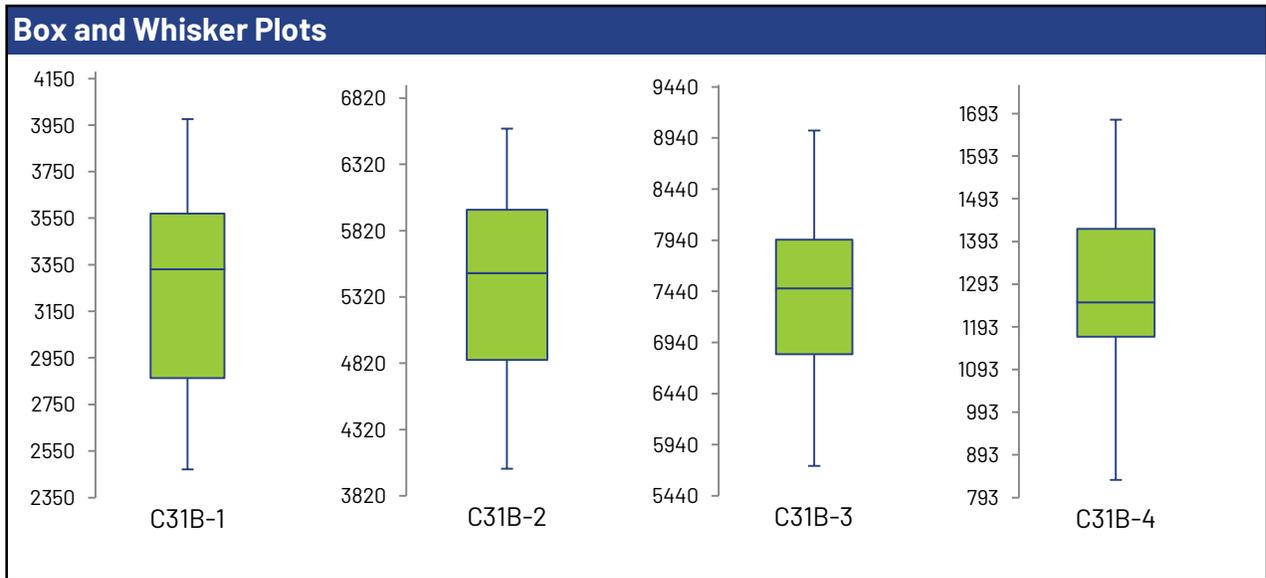
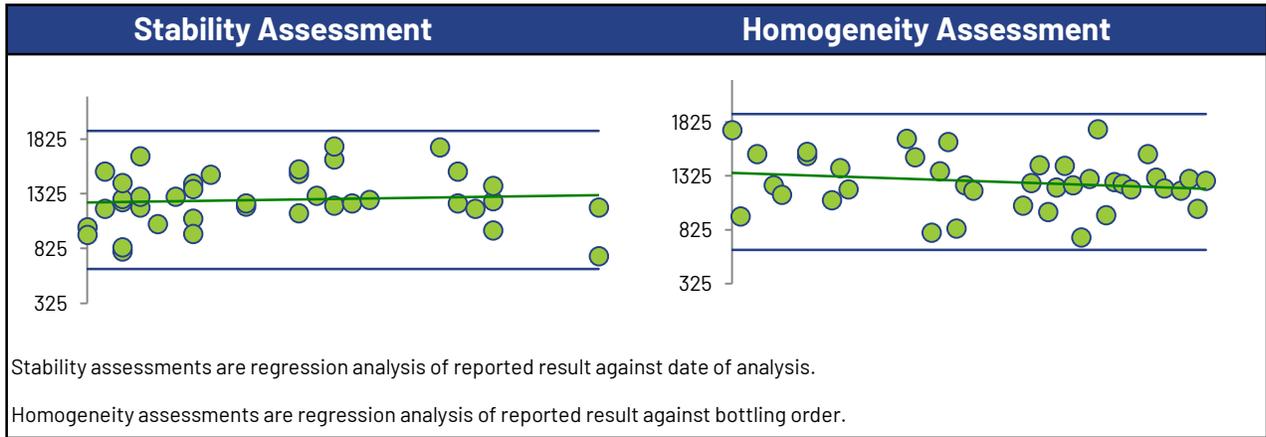


* Displayed failure rate does not account for the RDL modified z-Scores or inaccurately reported non-detects.

F3: C16-C34



F3: C16-C34



F4: C34-C50

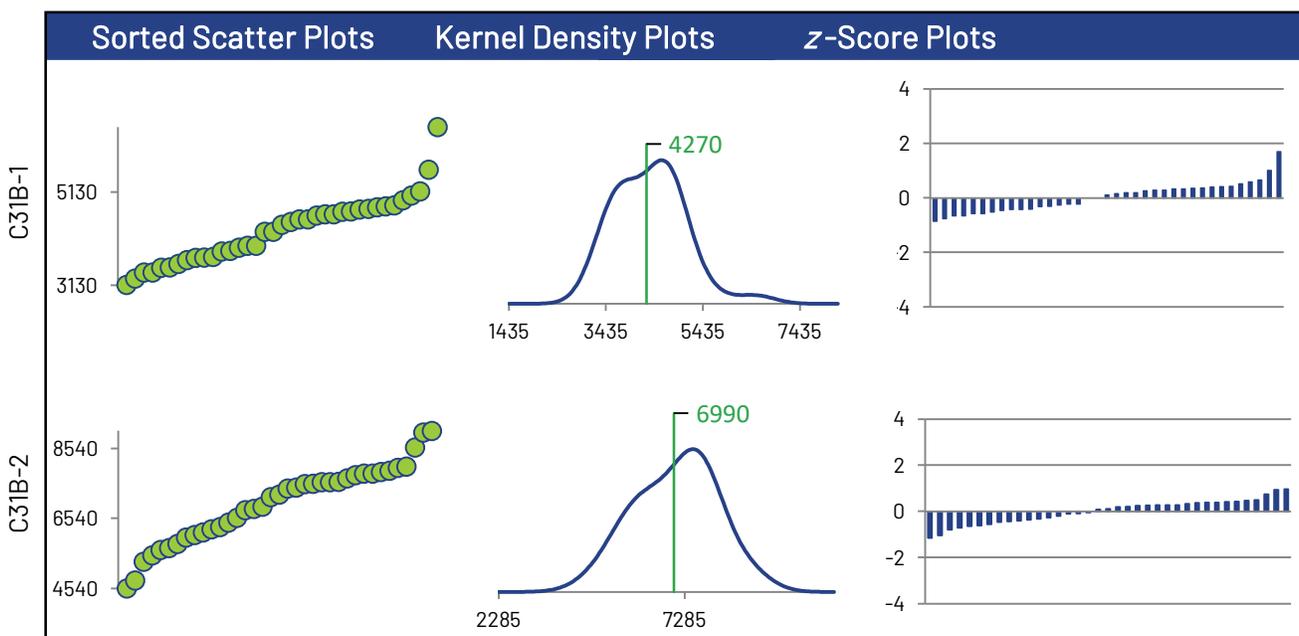
Summary Statistics

Statistic	C31B-1	C31B-2	C31B-3	C31B-4
No. of results used for summary statistics	37	37	37	37
Median mg/kg	4420	7230	9550	1500
Robust Mean mg/kg	4270	6990	9430	1500
Uncertainty of the Assigned Value mg/kg	143	236	279	44.8
Robust Standard Deviation mg/kg	695	1150	1360	218
Regression Standard Deviation mg/kg	1320	2130	2870	484
Stability Flag				
Homogeneity Flag				
Standard Deviation Used (SDPA) mg/kg	1320	2130	2870	484
Flagged Outliers	0	0	0	0
$ z > 3.0$	0	0	0	0
$2 < z < 3$	0	0	0	0
Failure Rate %*	0.0			

Methods Used

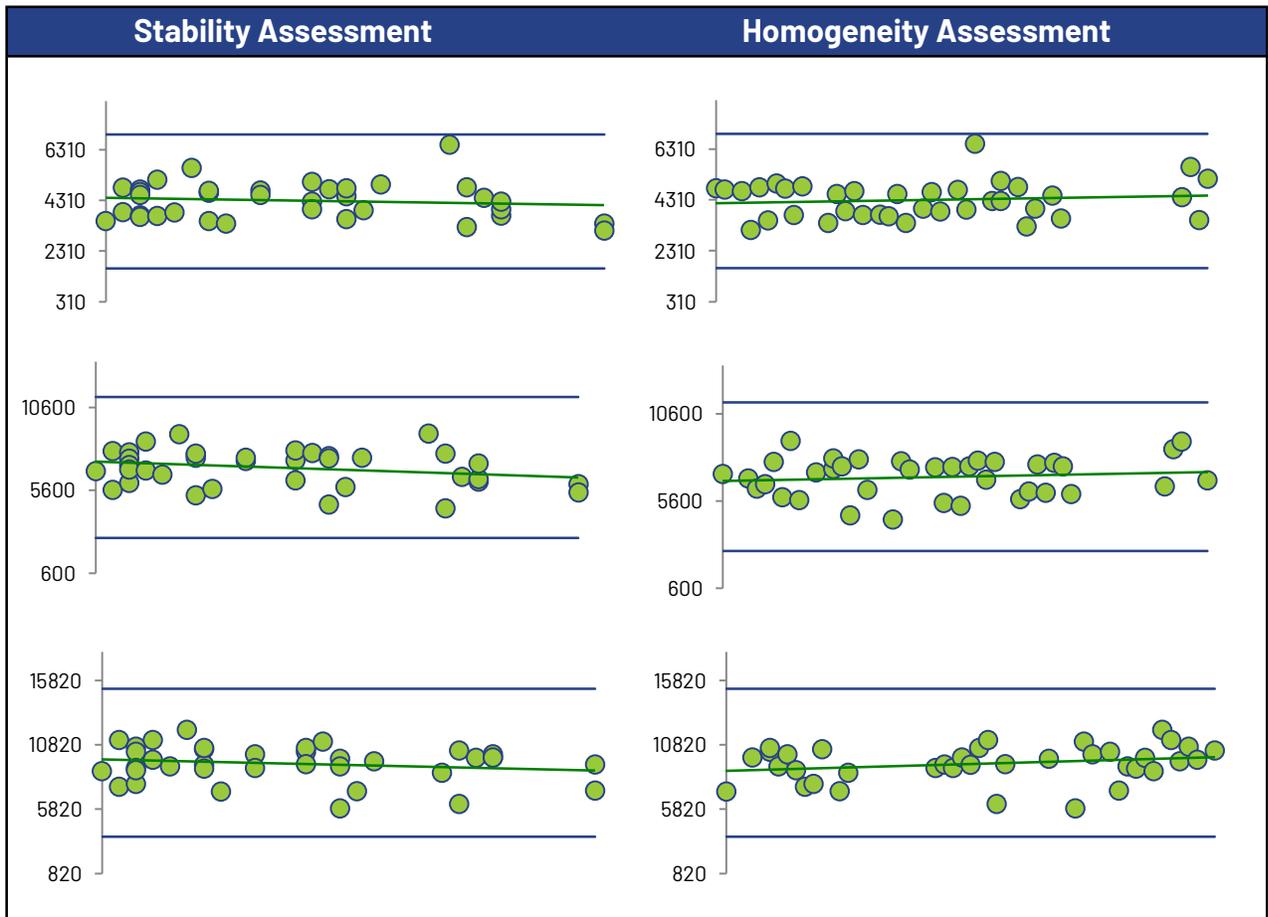
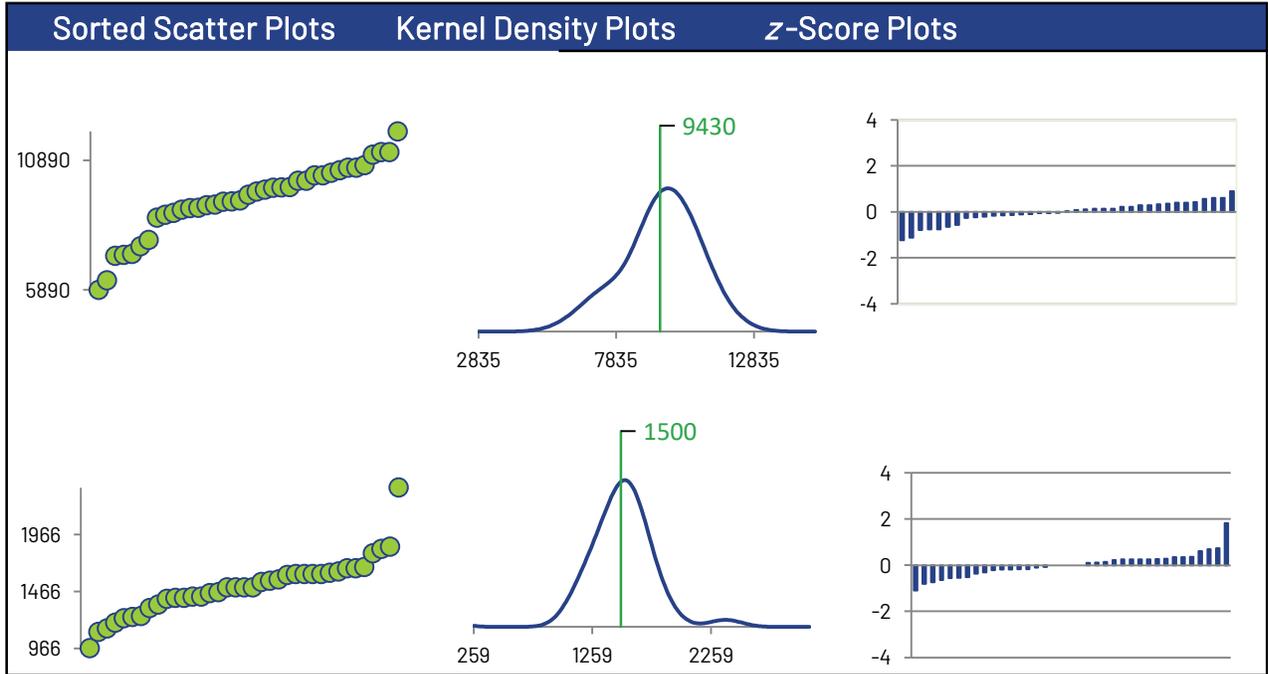
Method	C31B-1	C31B-2	C31B-3	C31B-4
GC/FID (Blue)	37	37	37	37

All summary stats, except the failure rate, and the plots below are based on the data excluding any flagged outliers



* Displayed failure rate does not account for the RDL modified z-Scores or inaccurately reported non-detects.

F4: C34-C50



F4: Gravimetric

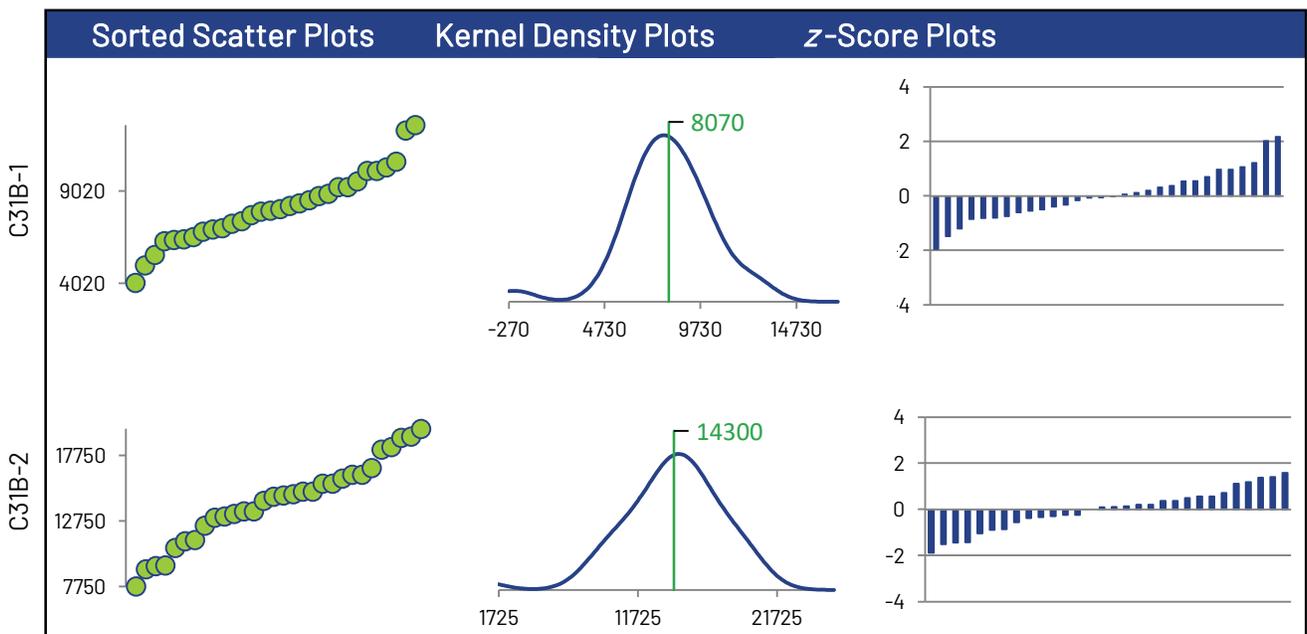
Summary Statistics

Statistic	C31B-1	C31B-2	C31B-3	C31B-4
No. of results used for summary statistics	30	30	30	30
Median mg/kg	8000	14700	19500	3280
Robust Mean mg/kg	8070	14300	18700	3350
Uncertainty of the Assigned Value mg/kg	434	794	981	244
Robust Standard Deviation mg/kg	1900	3480	4300	1070
Regression Standard Deviation mg/kg	2080	3240	4070	1200
Stability Flag				
Homogeneity Flag				
Standard Deviation Used (SDPA) mg/kg	2080	3480	4300	1200
Flagged Outliers	0	0	0	0
$ z > 3.0$	0	0	0	2
$2 < z < 3$	2	0	2	0
Failure Rate %*	0.0			

Methods Used

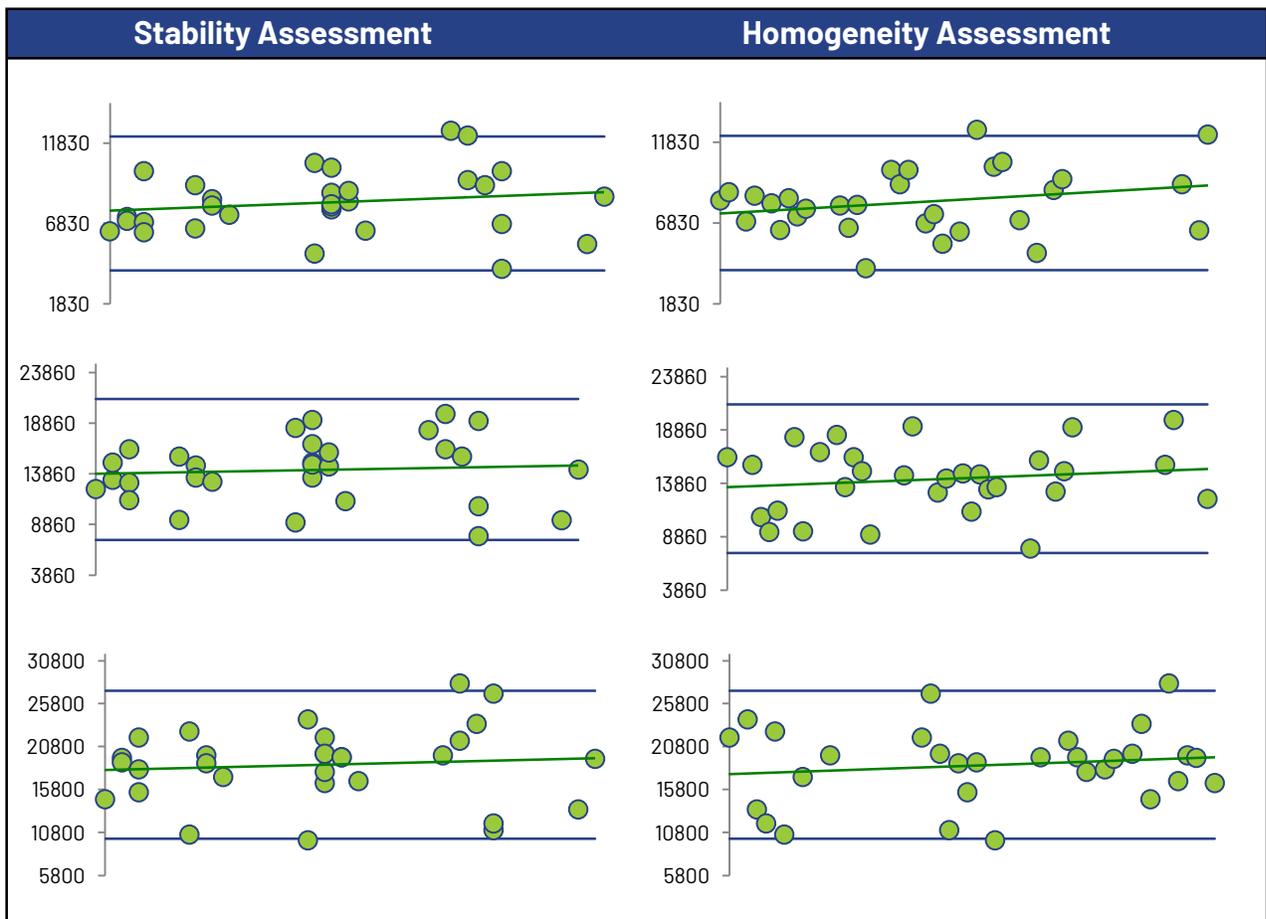
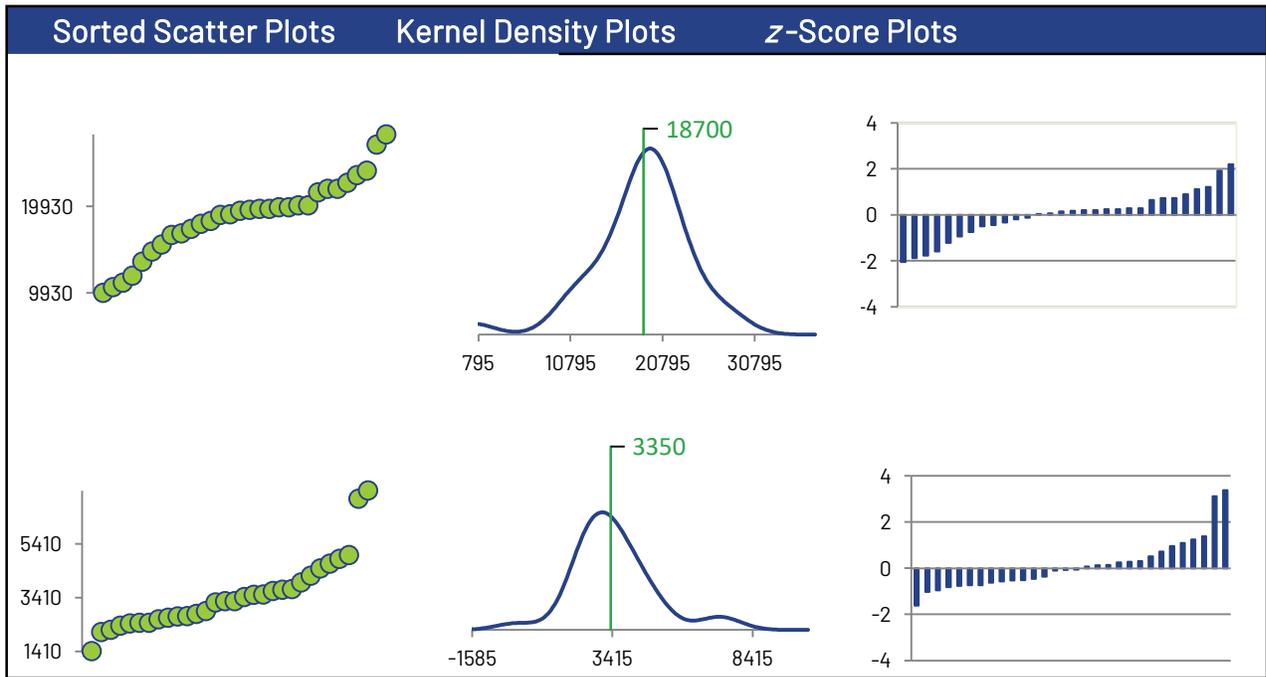
Method	C31B-1	C31B-2	C31B-3	C31B-4
GRAVIMETRIC (Blue)	30	30	30	30

All summary stats, except the failure rate, and the plots below are based on the data excluding any flagged outliers



* Displayed failure rate does not account for the RDL modified z-Scores or inaccurately reported non-detects.

F4: Gravimetric



F4: Gravimetric

