

# Test Group Summary Report

## C31B Petroleum Hydrocarbons in Soil

### June 2025 PT Round

---

**Issued: August 25, 2025**

## 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 PTC Proficiency Testing Reports contain some terms that new participants may not be familiar with:

<i>PT Code</i> :	The PT Code is unique to each analyte that a participant is registered for.
<i>Lab Info</i> :	If a participant is accredited by CALA, this three-digit number in the Laboratory Information 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 Value</i> :	The Assigned Value is the Robust Mean of the reported results, outliers excluded. This is often referred to as the “target” value.
<i>Reported Value</i> :	The result reported by the participant.
<i>SDPTA</i> :	The Standard Deviation of Proficiency Testing Assessment; this value is used to determine the acceptance limits for the PT evaluation. The calculations examine the Robust Standard Deviation (from consensus STD) and Regression Standard Deviation (STD from historical studies or TNI limits); and, if any, Homogeneity and Stability flags to determine the SDPTA (whichever is higher).
<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>PT 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.
<i>U</i> :	The Uncertainty of the assigned value.

## 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 *Statistical methods for use in proficiency testing by interlaboratory comparison*.

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.PTCanada.org](http://www.PTCanada.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 5<sup>th</sup> and 95<sup>th</sup> 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

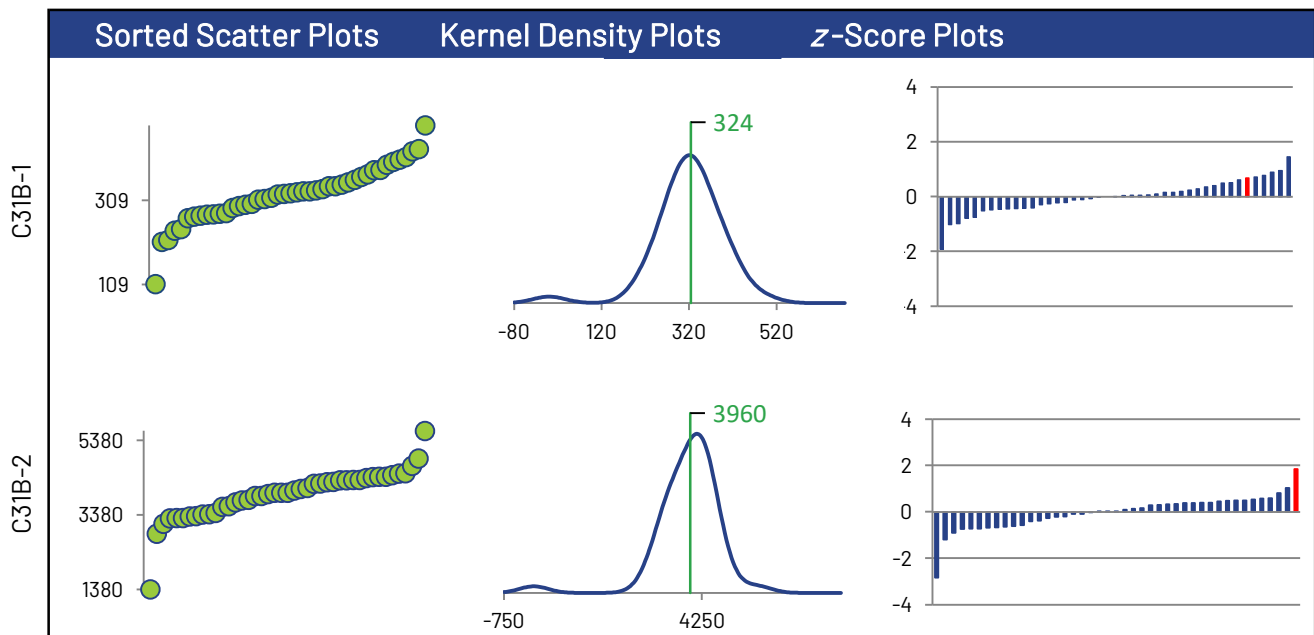
## Summary Statistics

Statistic	C31B-1	C31B-2	C31B-3	C31B-4
No. of results used for summary statistics	43	43	43	43
Median mg/kg	325	3980	2490	2190
Robust Mean mg/kg	324	3960	2450	2190
Uncertainty of the Assigned Value mg/kg	11.9	104	78.5	70.5
Robust Standard Deviation mg/kg	62.4	545	412	370
Regression Standard Deviation mg/kg	112	911	580	523
Stability Flag				
Homogeneity Flag				
Standard Deviation Used (SDPA) mg/kg	112	911	580	523
Flagged Outliers	0	0	0	0
$ z  > 3.0$	0	0	0	0
$2 <  z  < 3$	0	1	1	3
Failure Rate %*	2.3			

## Methods Used

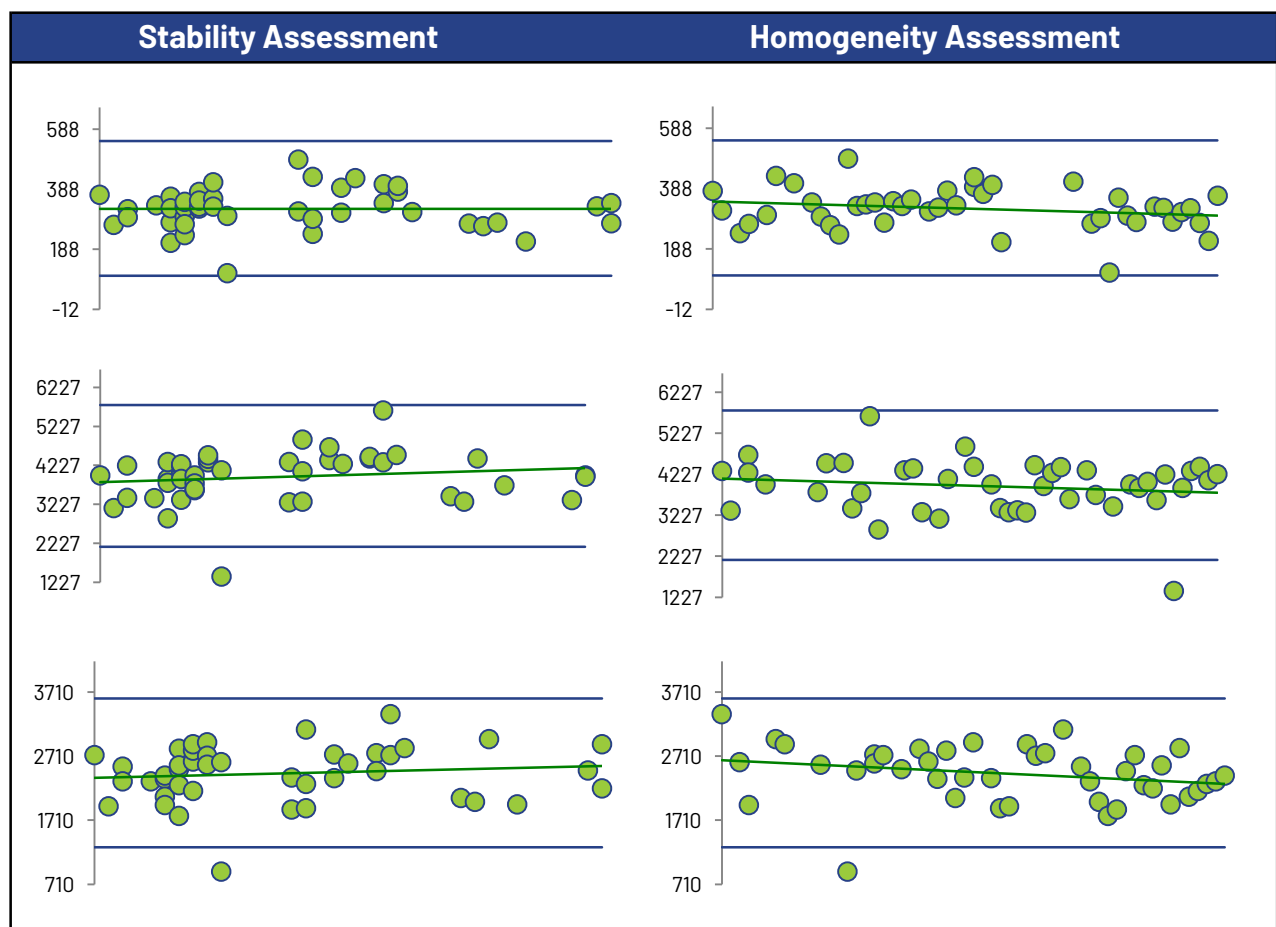
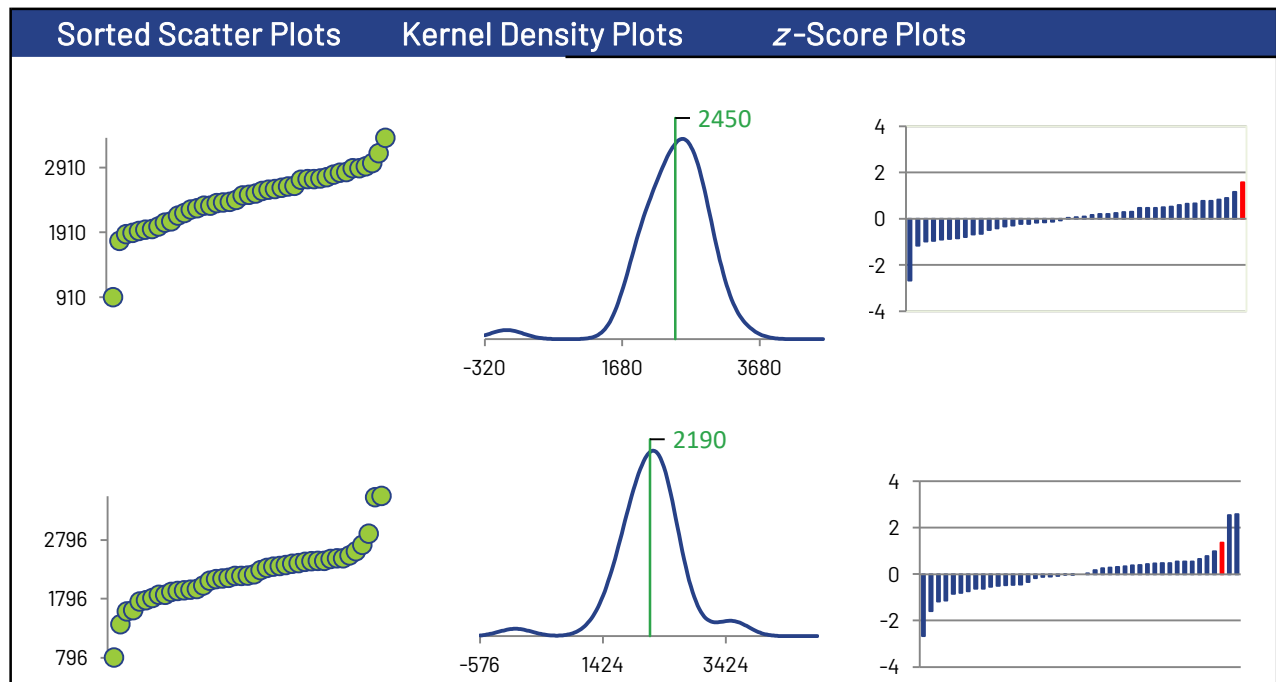
Method	C31B-1	C31B-2	C31B-3	C31B-4
GC/FID (Blue)	42	42	42	42
GC/MS (Red)	1	1	1	1

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

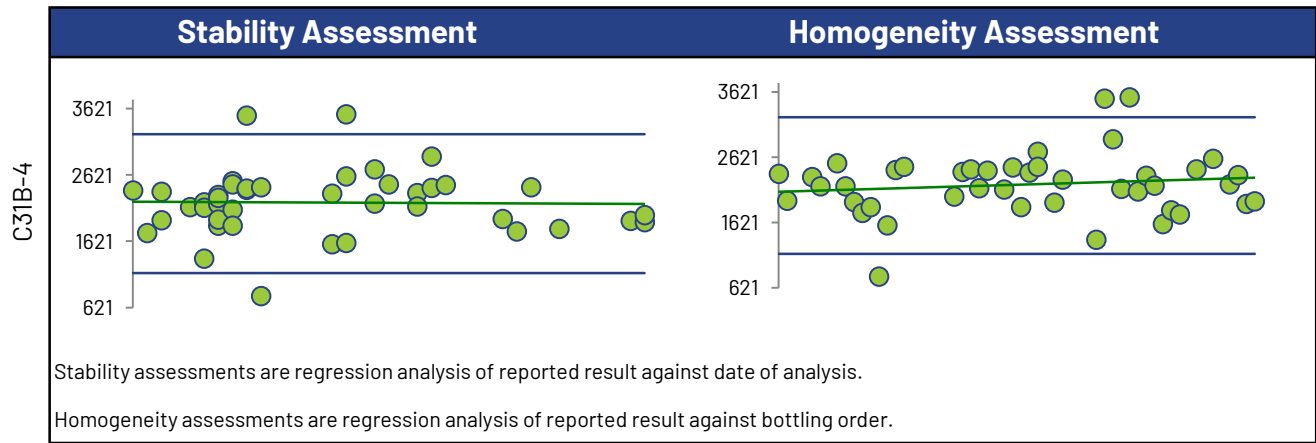


\* Displayed failure rate does not account for the RDL modified z-Scores or inaccurately reported non-detects.

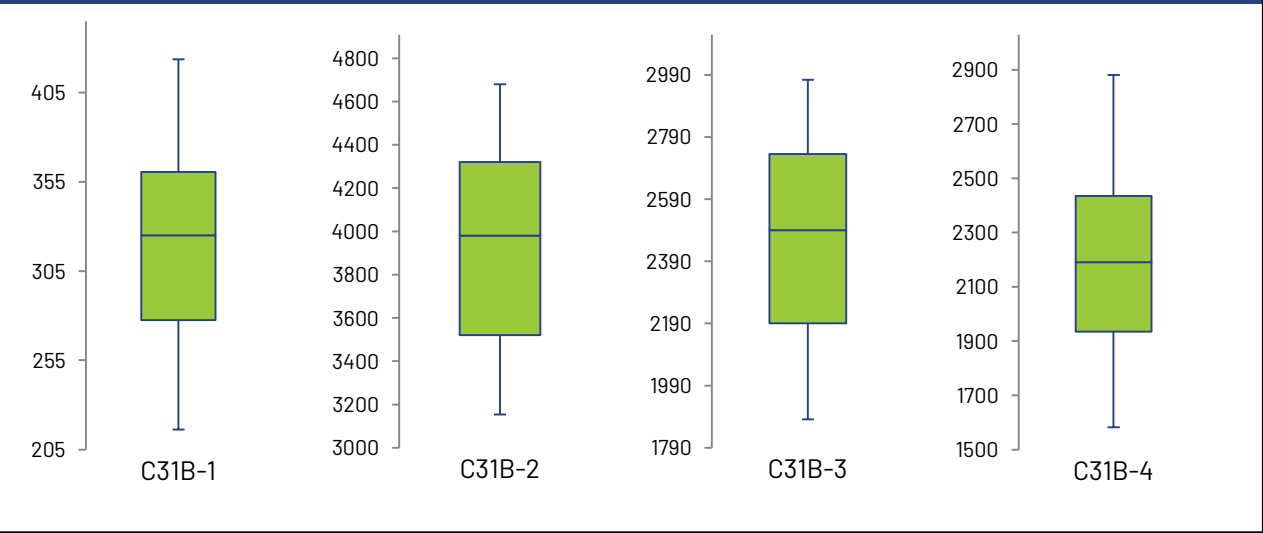
## F2: C10-C16



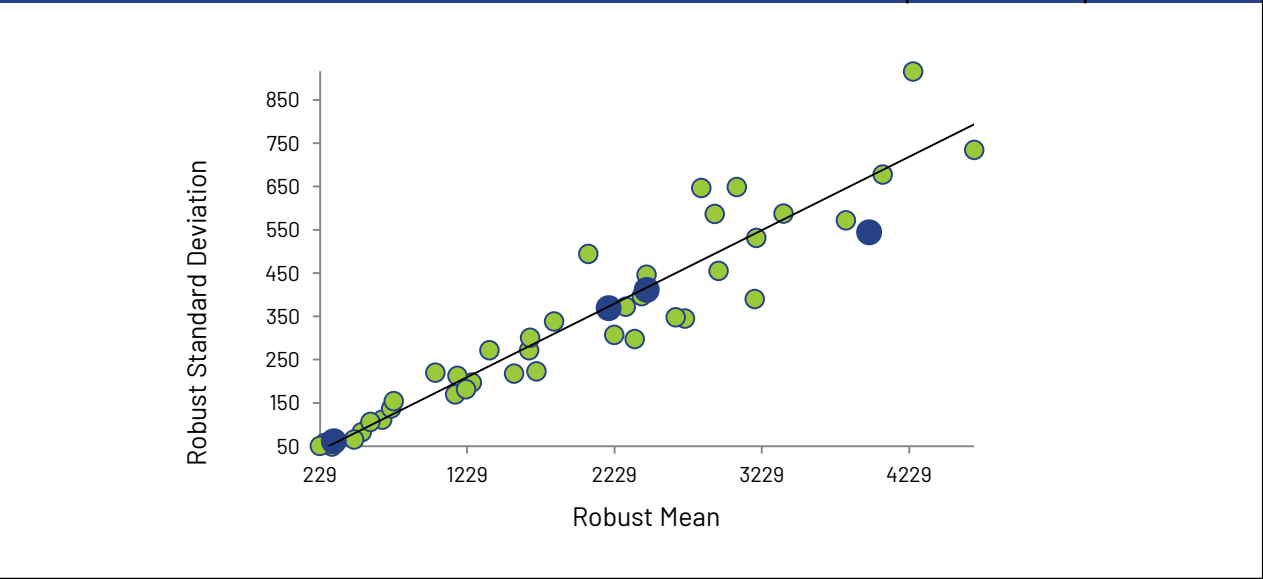
F2: C10-C16



Box and Whisker Plots



Current Round (blue) Compared to Historic Rounds (green)





## F3: C16-C34

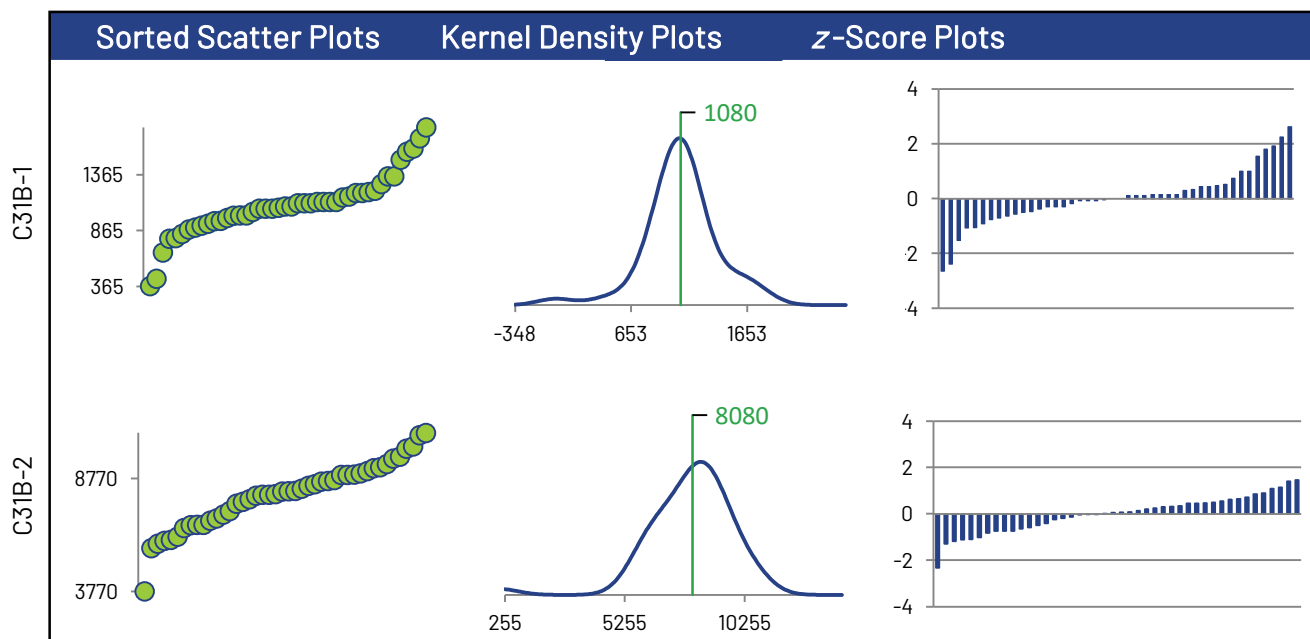
## Summary Statistics

Statistic	C31B-1	C31B-2	C31B-3	C31B-4
No. of results used for summary statistics	44	44	44	44
Median mg/kg	1080	8210	5940	4320
Robust Mean mg/kg	1080	8080	5830	4380
Uncertainty of the Assigned Value mg/kg	40.1	277	230	165
Robust Standard Deviation mg/kg	213	1470	1220	874
Regression Standard Deviation mg/kg	271	1860	1350	1020
Stability Flag				
Homogeneity Flag				
Standard Deviation Used (SDPA) mg/kg	271	1860	1350	1020
Flagged Outliers	0	0	0	0
$ z  > 3.0$	0	0	0	0
$2 <  z  < 3$	4	1	1	3
Failure Rate %*	2.3			

## Methods Used

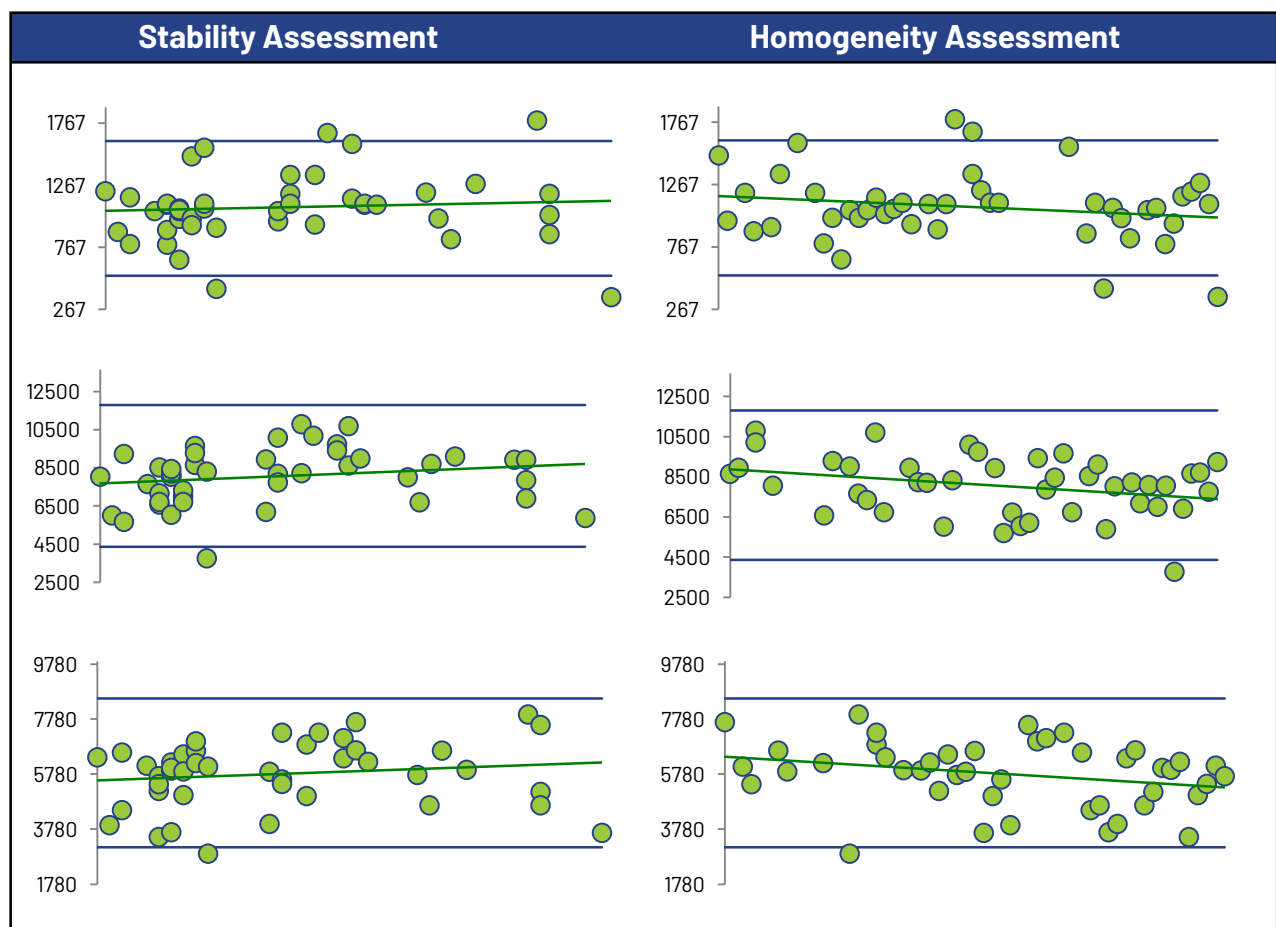
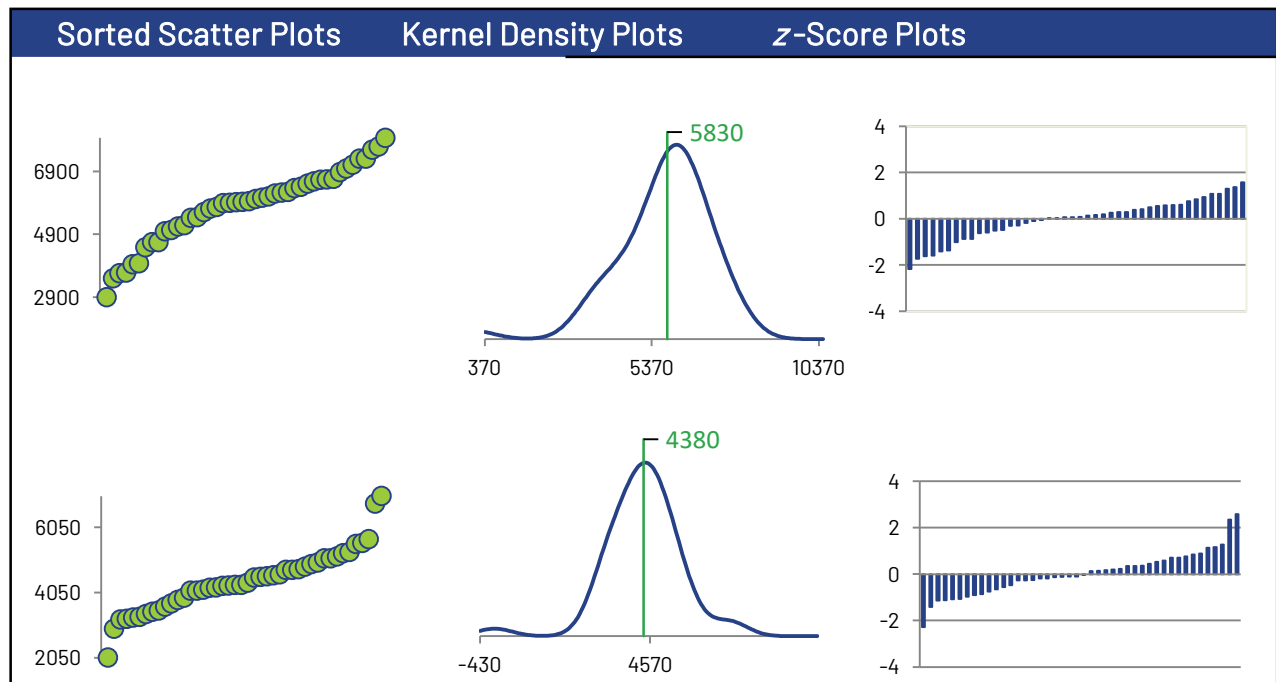
Method	C31B-1	C31B-2	C31B-3	C31B-4
GC/FID (Blue)	44	44	44	44

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

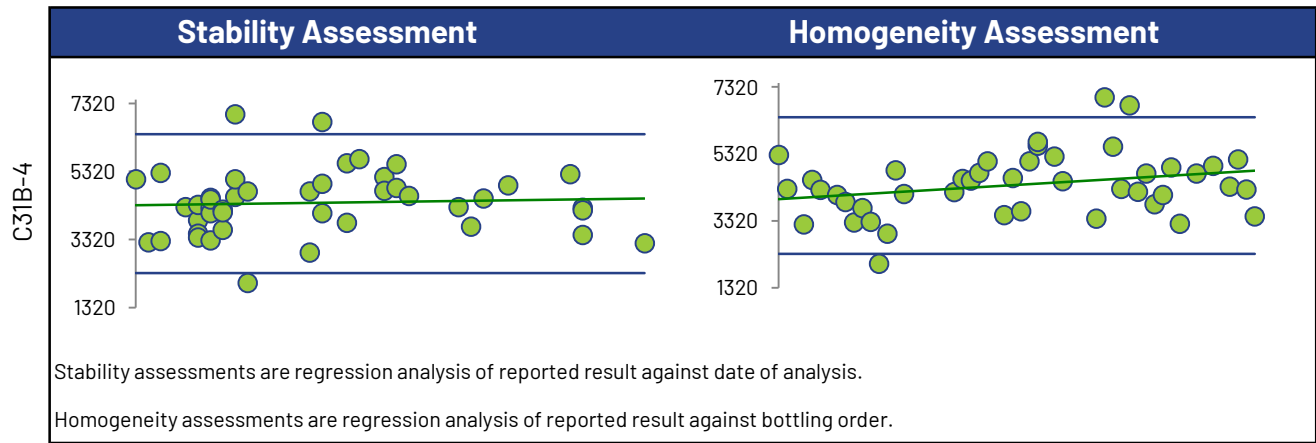


\* Displayed failure rate does not account for the RDL modified z-Scores or inaccurately reported non-detects.

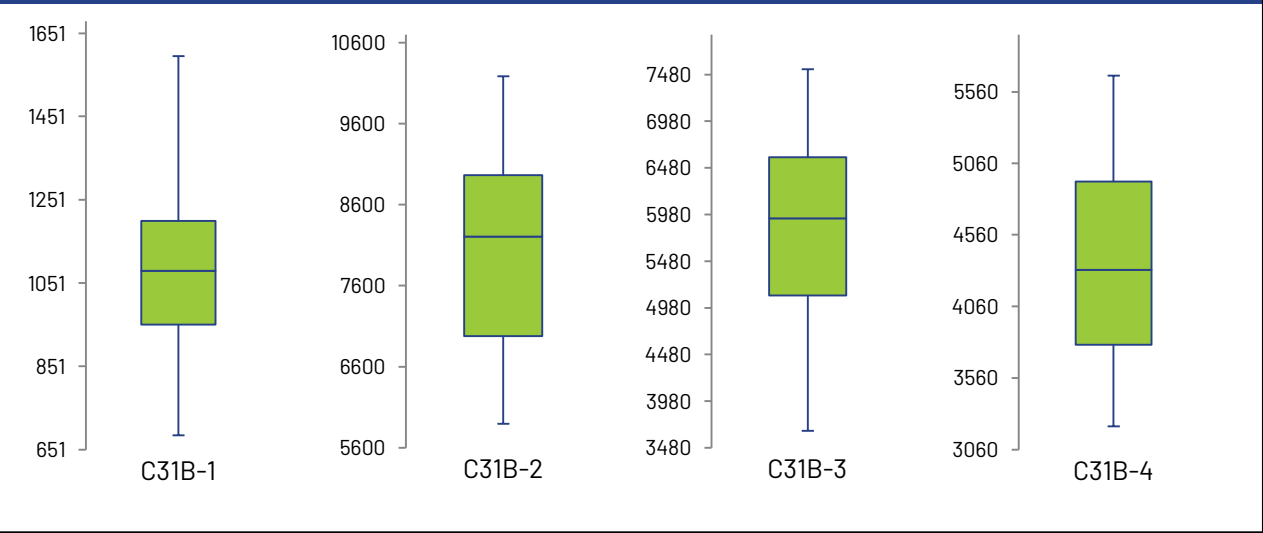
### F3: C16-C34



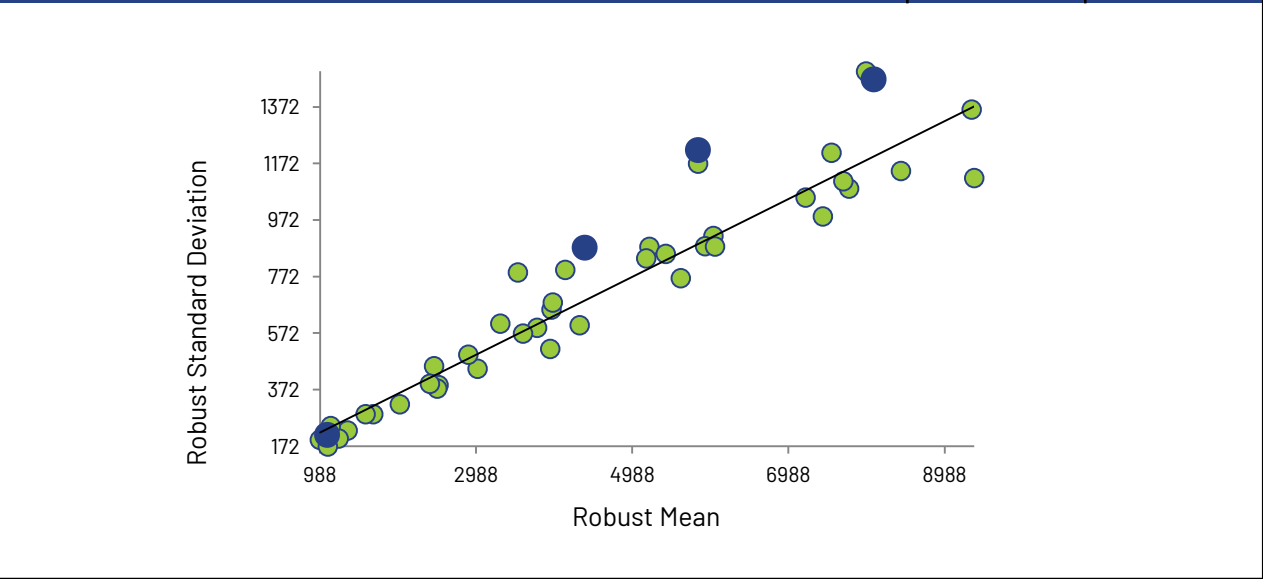
F3: C16-C34



Box and Whisker Plots



Current Round (blue) Compared to Historic Rounds (green)



## F4: C34-C50

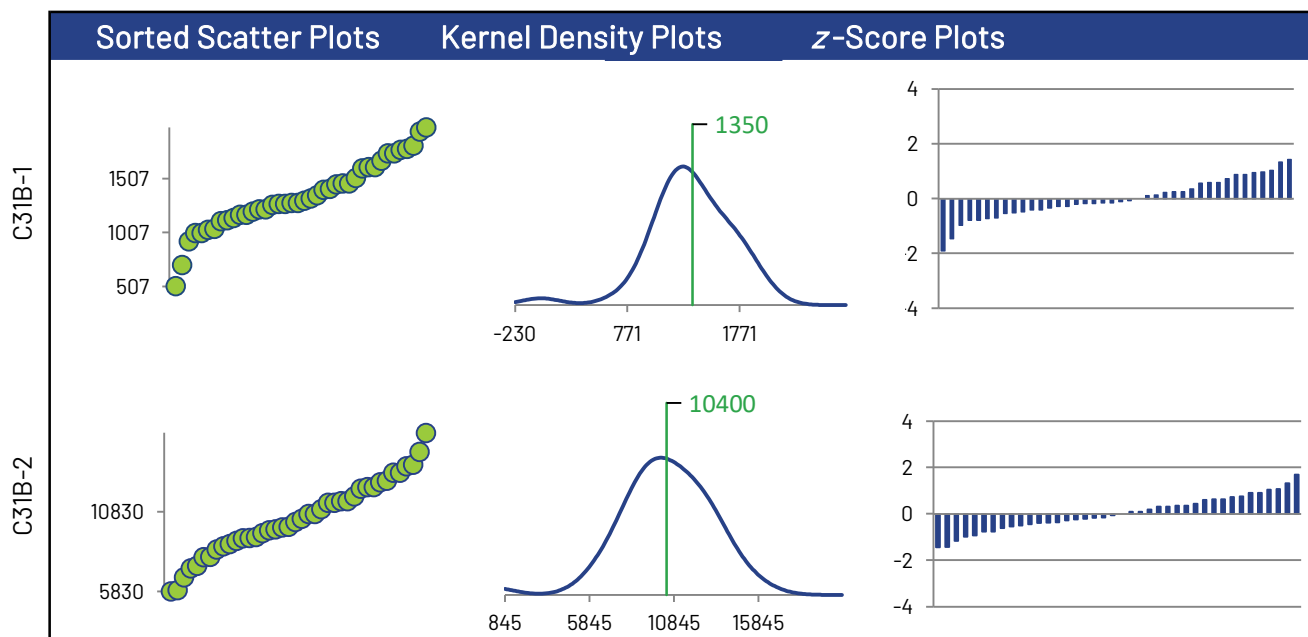
## Summary Statistics

Statistic	C31B-1	C31B-2	C31B-3	C31B-4
No. of results used for summary statistics	40	40	40	40
Median mg/kg	1290	10300	8110	5540
Robust Mean mg/kg	1350	10400	8100	5670
Uncertainty of the Assigned Value mg/kg	63.2	488	342	253
Robust Standard Deviation mg/kg	320	2470	1730	1280
Regression Standard Deviation mg/kg	441	3170	2460	1740
Stability Flag				
Homogeneity Flag				
Standard Deviation Used (SDPA) mg/kg	441	3170	2460	1740
Flagged Outliers	0	0	0	0
$ z  > 3.0$	0	0	0	0
$2 <  z  < 3$	0	0	0	1
Failure Rate %*	0.0			

## Methods Used

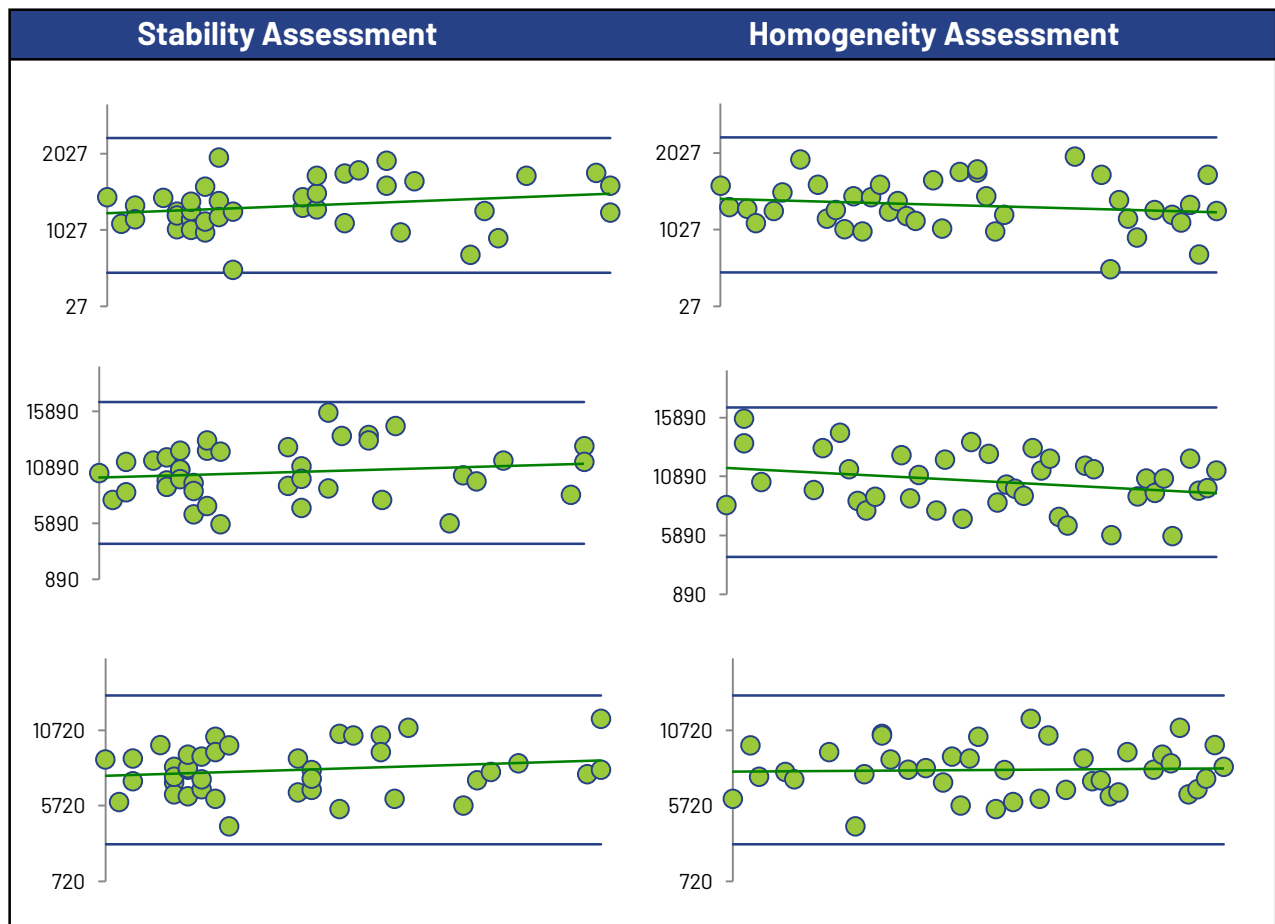
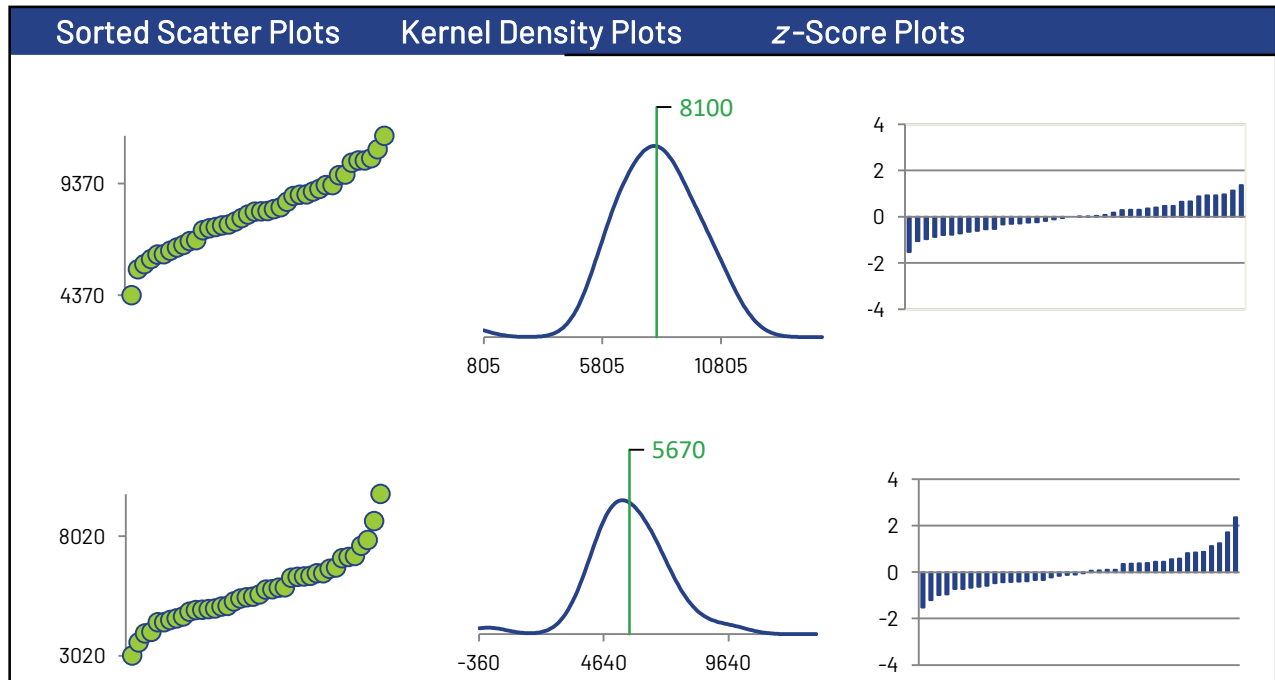
Method	C31B-1	C31B-2	C31B-3	C31B-4
GC/FID (Blue)	40	40	40	40

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

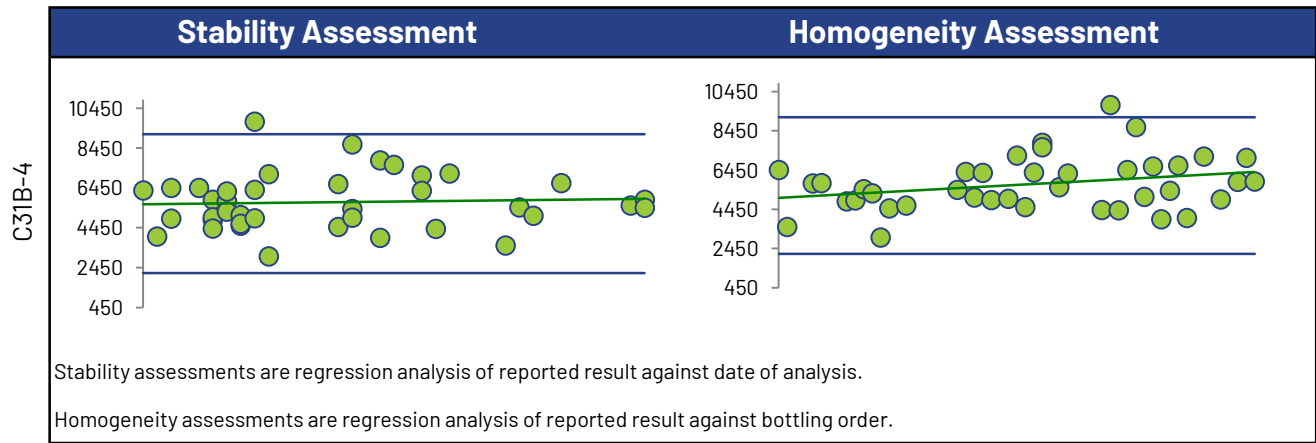


\* Displayed failure rate does not account for the RDL modified z-Scores or inaccurately reported non-detects.

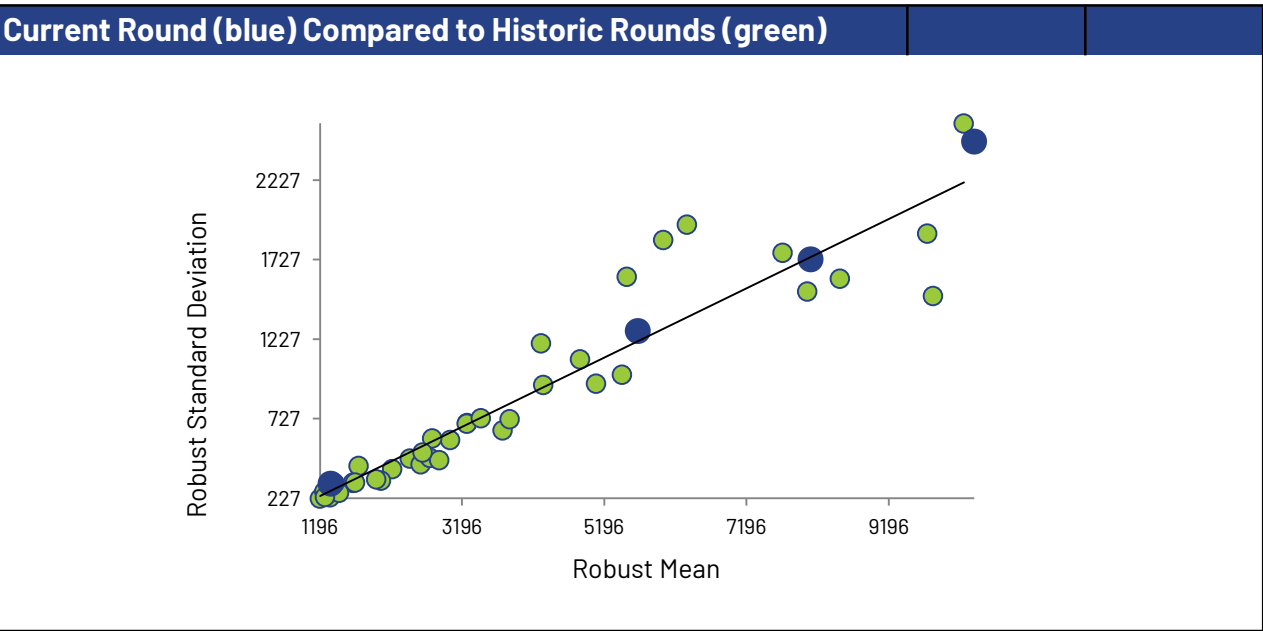
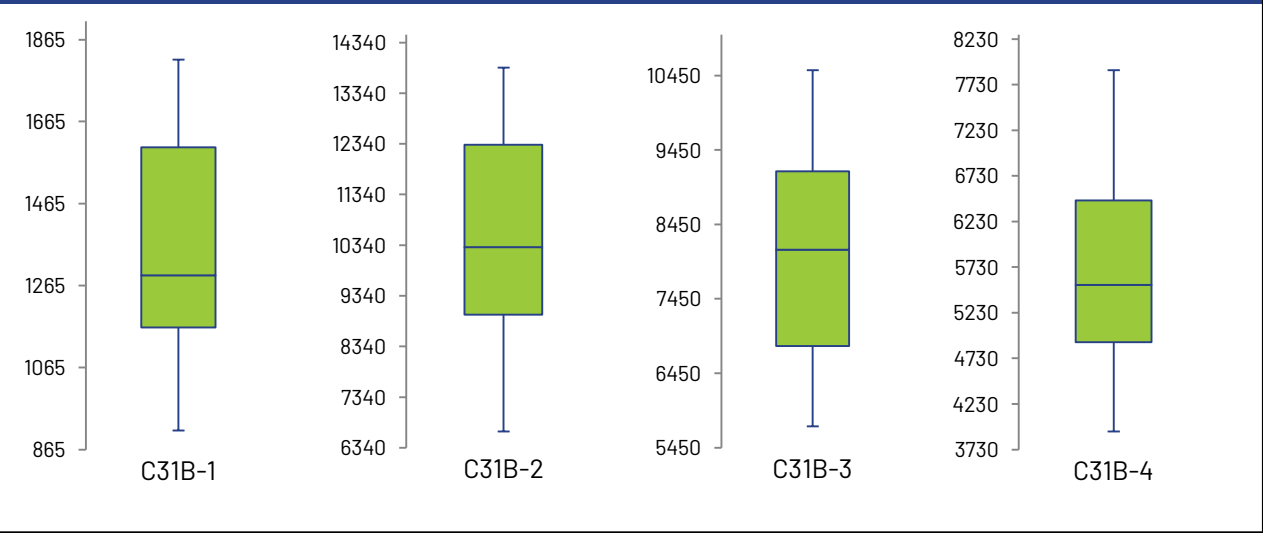
## F4: C34-C50



F4: C34-C50



Box and Whisker Plots



## F4: Gravimetric

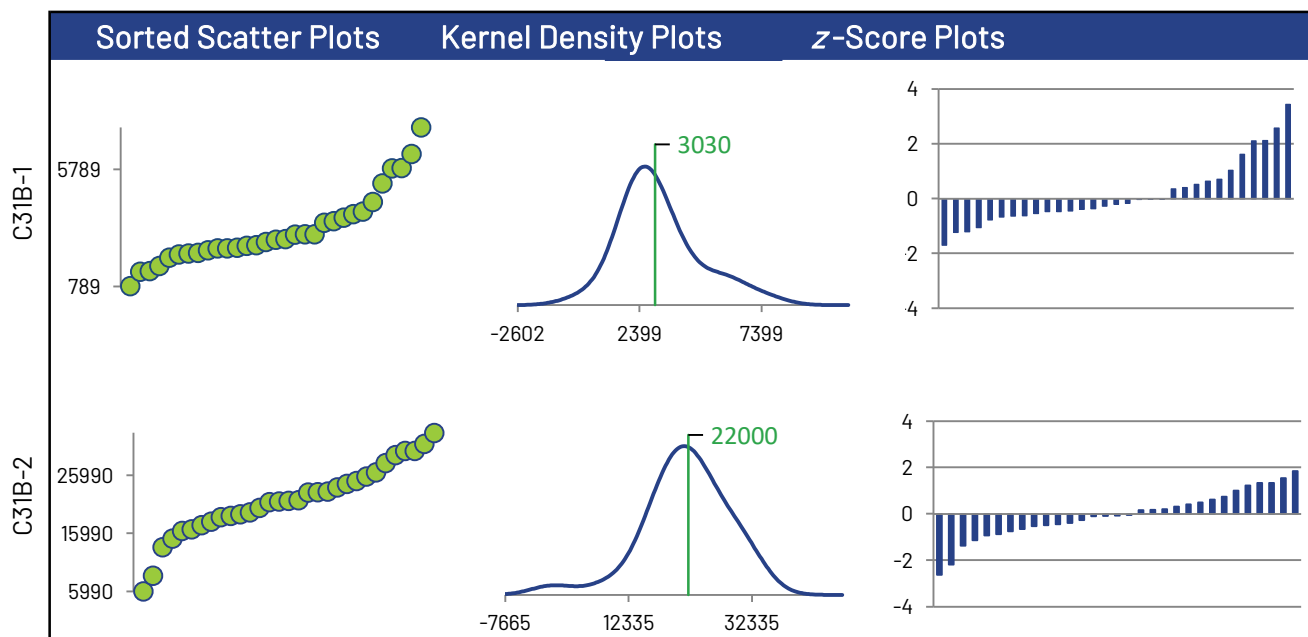
## Summary Statistics

Statistic	C31B-1	C31B-2	C31B-3	C31B-4
No. of results used for summary statistics	31	31	31	31
Median mg/kg	2770	21600	15800	10900
Robust Mean mg/kg	3030	22000	16100	11400
Uncertainty of the Assigned Value mg/kg	296	1370	905	741
Robust Standard Deviation mg/kg	1320	6110	4030	3300
Regression Standard Deviation mg/kg	1140	4680	3590	2700
Stability Flag				
Homogeneity Flag				
Standard Deviation Used (SDPA) mg/kg	1320	6110	4030	3300
Flagged Outliers	0	0	0	0
$ z  > 3.0$	1	0	0	0
$2 <  z  < 3$	3	2	2	2
Failure Rate %*	9.7			

## Methods Used

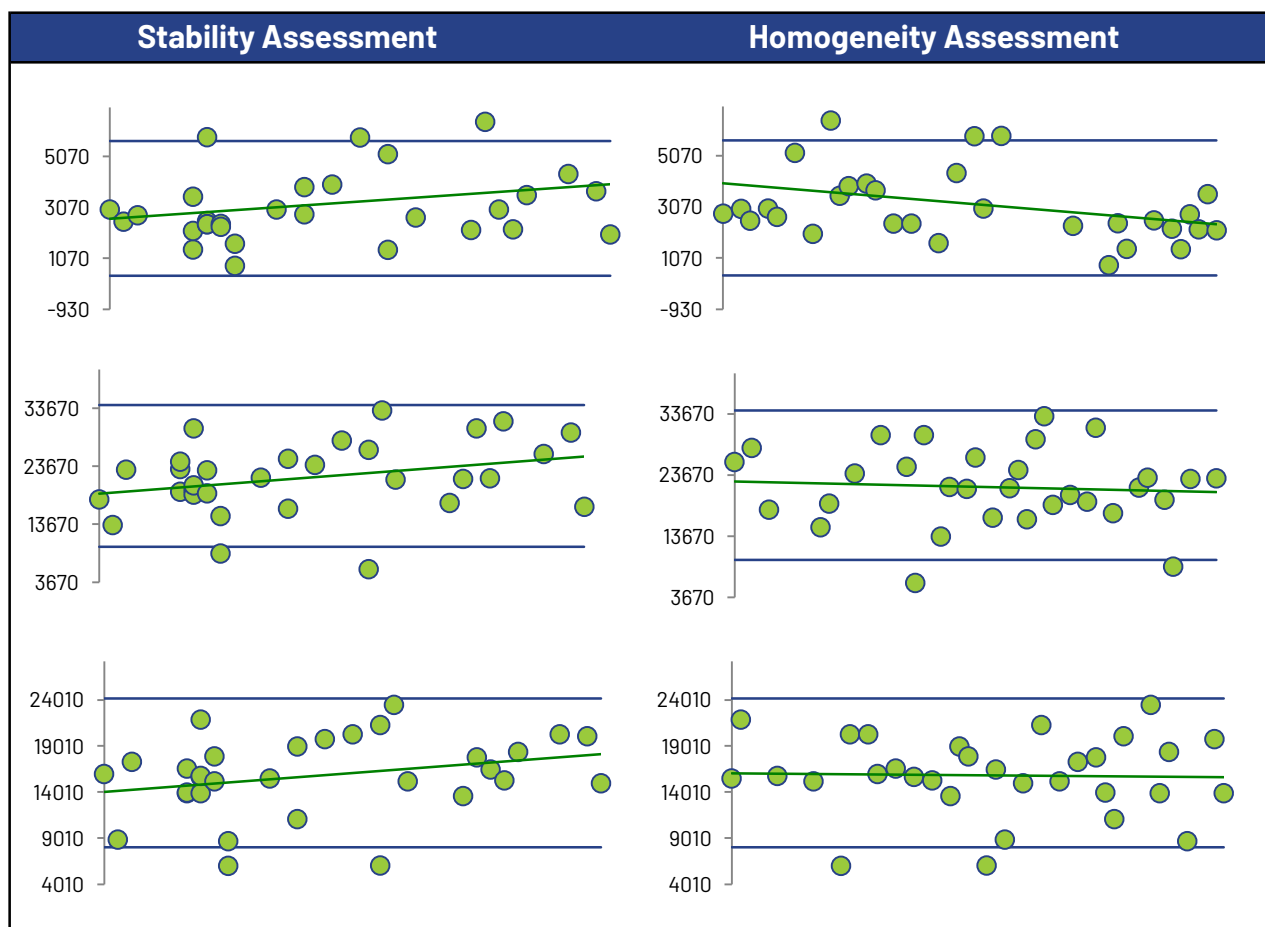
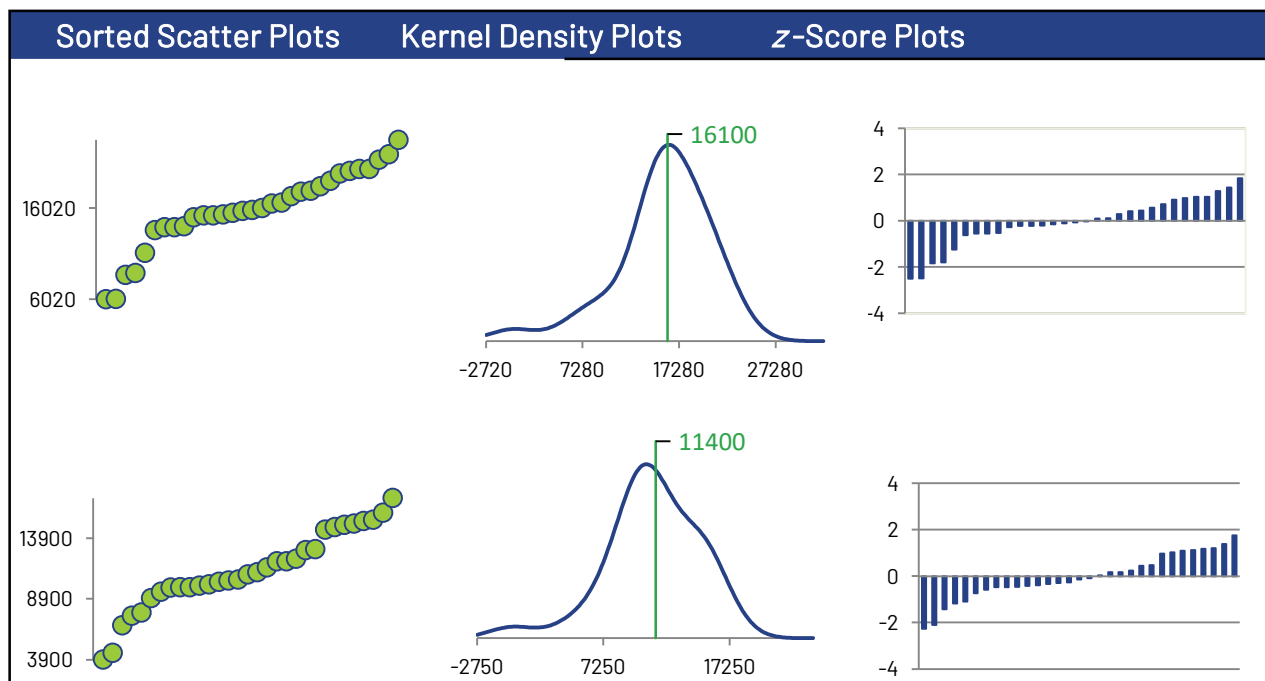
Method	C31B-1	C31B-2	C31B-3	C31B-4
GRAVIMETRIC (Blue)	31	31	31	31

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



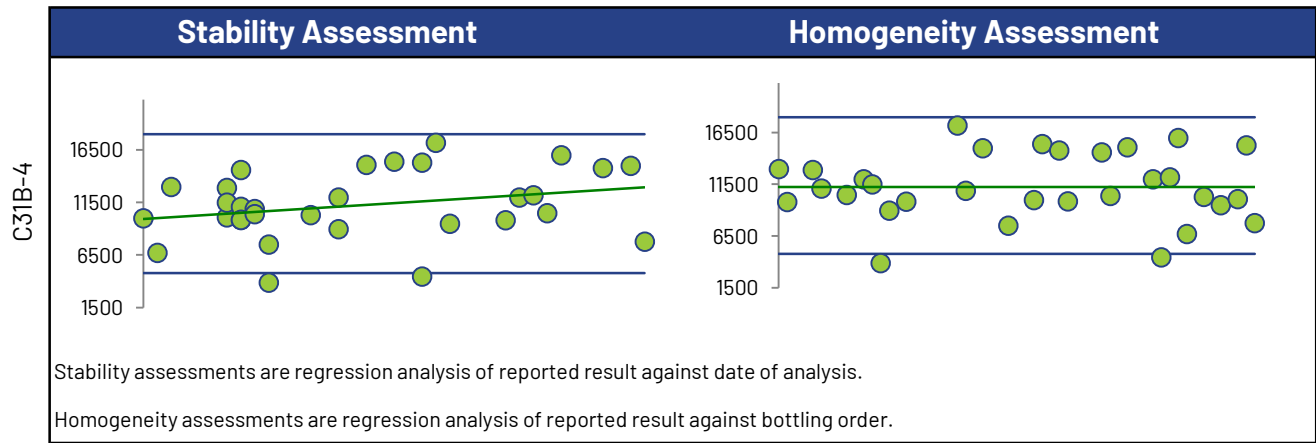
\* Displayed failure rate does not account for the RDL modified z-Scores or inaccurately reported non-detects.

## F4: Gravimetric





F4: Gravimetric



Box and Whisker Plots

