Control Chart Methodology for Detecting Under-reported Emissions

Introduction:

EPA has begun to apply the control chart methodology described in this paper to electronically identify the potential development of sampling system in-leakage (or other monitoring system problems) that result in the under-measurement of emissions data. Such problems are often only detected by a RATA test which are typically conducted only once per year. Control charts are an effective tool for identifying unusual shifts in data. A typical control chart consists of (1) points representing average measurements taken over an interval of time; (2) a center line, drawn at the overall mean for the data set; and (3) upper and lower control limits that indicate the threshold at which the data is considered statistically unlikely. This paper describes the steps that the Agency intends to use to evaluate CO_2 data for indications of data problems.

Data Needed for Analysis:

- Hourly CO_2 concentration data
- Load Bin data
- MODC data
- Date of completion for the previous CO_2 RATA

Procedure:

Step1: Identify a Load Bin to Evaluate

It is necessary to evaluate data from narrow operating bands to remove (or at least reduce) the affects that operating conditions have on the data variability. EPA automates it's evaluation to focus on the most used load bin to maximize the amount of data available for the evaluation. However, this evaluation can be run at any load bin you desire. Note that, it may be useful to run the evaluation on several load bins to determine if a finding is dependant on load.

Step 2: Creating the Baseline Data Set

Collect 30-days of hourly data from the load bin of interest starting with the day following the completion of the last CO_2 RATA. Use only quality assured data from the primary monitoring system where the MODC for the hour is equal to "1".

Next, count the number of hourly measurements available for each day, discarding the data for any day where the count is less than 6. While this step is optional, EPA believes that doing so will ensure that the daily averages are more meaningful.

Note, EPA recommends that the baseline be based on at least 15 daily averages, so if less that 15 days of data are available, you should use more data until you have at least 15 days worth of data meeting the above criteria.

Step 3: Determining the Baseline Parameters

(A) Calculate the average CO_2 concentration within the load for each day in the baseline period.

$$\overline{C}_d = \frac{\sum_{1}^{h} C_h}{h} \quad \text{Eq. A}$$

Where:

$\overline{C}_d$ = Daily Average CO_2 Concentration
C_h = Hourly CO_2 Concentration from measurement
h = Number of hourly data points in the day

(B) Calculate the baseline mean by averaging each of the daily averages.

$$\overline{C}_B = \frac{\sum_{1}^{d} \overline{C}_d}{d} \quad \text{Eq. B}$$

Where:

$\overline{C}_B$ = Average Daily Average CO_2 Concentration (Baseline Mean)
$\overline{C}_d$ = Daily Average CO_2 Concentration
C_h = Hourly CO_2 Concentration from measurement
d = Number of hourly data points in the day

(C) Calculate the standard deviation of the daily average CO_2 concentration during the baseline period as:

$$\sigma_B = \frac{\sqrt{\sum_{d=1}^{d} \left(\overline{C}_d - \overline{C}_B\right)^2}}{(d-1)} \quad \text{Eq. C}$$

Where:

σ_B = Standard deviation of the daily average CO_2 concentration values during the baseline period.
$\overline{C}_B$ = Average Daily Average CO_2 Concentration (Baseline Mean)
$\overline{C}_d$ = Daily Average CO_2 Concentration
d = Number of hourly data points in the day

Step 4: Calculate Control Limits

Control limits are calculated by adding and subtracting multiples of the standard deviation to or from the baseline mean as follows:

Upper Control Limit:

$$UCL = \overline{C}_B + 3\sigma_B \qquad \text{Eq. D-I}$$

Lower Control Limit:

$$LCL = \overline{C}_B - 3\sigma_B \qquad \text{Eq. D-II}$$

Where:

UCL = Upper Control Limit
LCL = Lower Control Limit
σ_B = Standard deviation of the daily average CO_2 concentration values during the baseline period.
$\overline{C}_B$ = Average Daily Average CO_2 Concentration (Baseline Mean)

Note: EPA will use ±3σ-level (3 times the standard deviation) for audit purposes which correspond to a 99.7% certainty that the data should fall within the range. EPA recommends that sources set a warning control limit at a ±2σ-level (2 times the standard deviation) and take investigative action whenever the data is outside this warning level.

Upper Warning Limit:

$$UWL = \overline{C}_B + 2\sigma_B \qquad \text{Eq. D-III}$$

Lower Warning Limit:

$$LWL = \overline{C}_B - 2\sigma_B \qquad \text{Eq. D-IV}$$

Where:

UWL = Upper Warning Limit
LWL = Lower Warning Limit
σ_B = Standard deviation of the daily average CO_2 concentration values during the baseline period.
$\overline{C}_B$ = Average Daily Average CO_2 Concentration (Baseline Mean)

Step 5: Comparing daily averages to the Baseline

For each day following the end of the baseline period, calculate the daily average CO_2 concentration for the selected load bin using equation A in Step 3(A) above. As described in Step 2, use only quality assured data from the primary monitoring system where the MODC for the hour is equal to "1" for evaluation purposes. Also, if you used the suggested minimum of 6 hourly values to make up a daily averages in the baseline data set; you must continue to discard the data for any day where the count is less than 6.

Compare each daily average CO_2 concentration value to the baseline average and the calculated upper and lower control limits from Step 4. Flag any daily value that is outside the control limits and track High and Low deviations separately.

Whenever 7 or more consecutive daily average values are out-of-bounds, (more than 3σ low as compared to the baseline average), EPA flags that unit as having suspect data and will send a letter to the Designated Representative and copy the monitoring contacts requesting an explanation and possible data substitution. For onsite quality control, EPA suggests that sources investigate and document any actions taken whenever any of the following are observed:

1. Any point falls beyond the ±3σ-level control limits; or
2. 3 of 4 consecutive daily averages fall beyond the ±2σ-level warning control limits; or
3. 8 or more consecutive points fall on one side of the mean

Other Considerations:

Constructing a control chart graph is a simple yet effective way to track monitoring system performance. This can be done manually by plotting the daily averages on a chart where the upper and lower control and warning levels are drawn about the baseline mean; or can be done electronically.

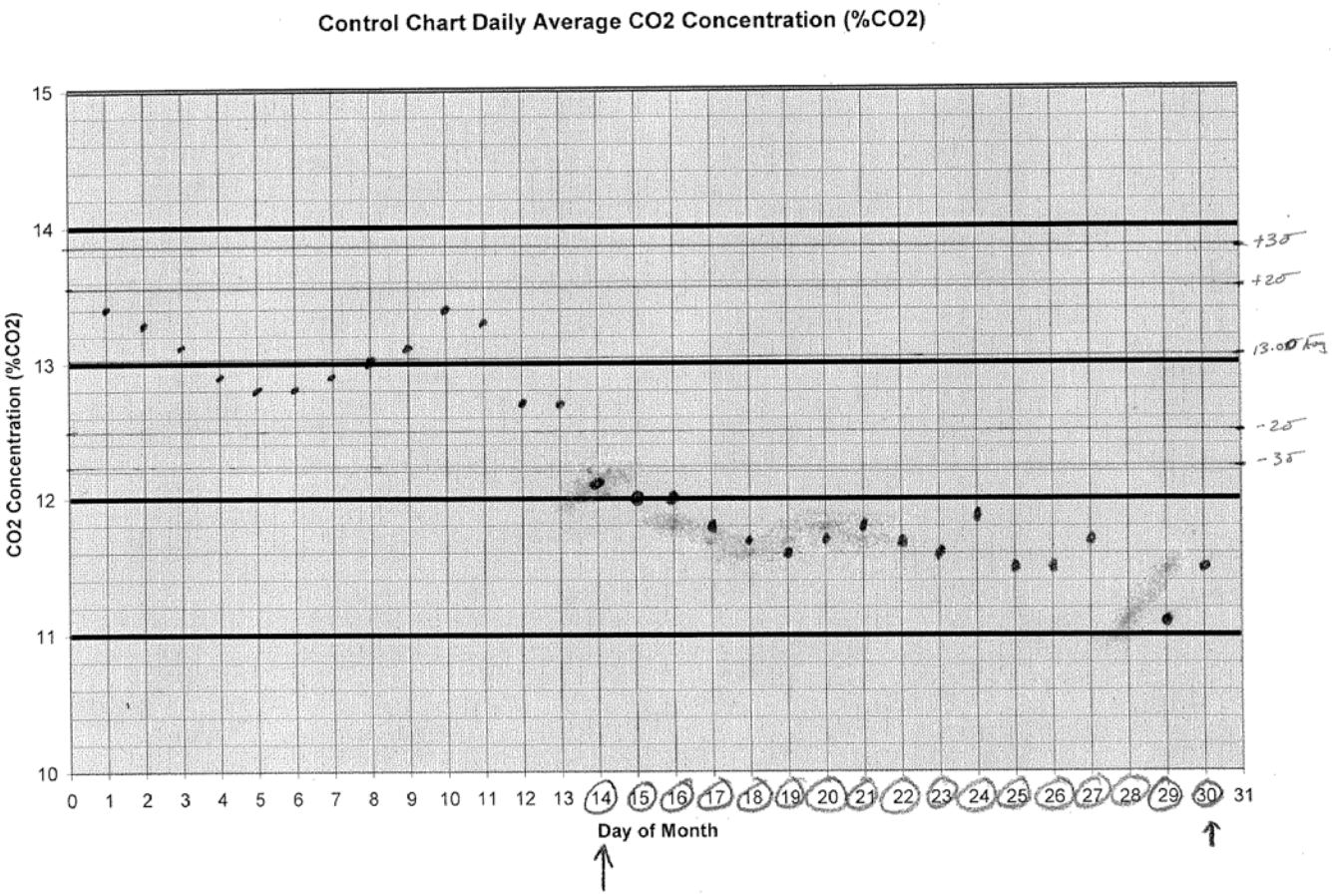
EXAMPLE

Baseline Data (30-days after RATA)

Date	Daily Average CO2 (Baseline Period)
6/23/2006	12.9
6/24/2006	13.2
6/25/2006	13.4
6/26/2006	13.3
6/27/2006	13.1
6/28/2006	13.2
6/29/2006	13.1
6/30/2006	13.0
7/1/2006	12.5
7/2/2006	12.5
7/3/2006	12.7
7/4/2006	12.8
7/5/2006	12.7
7/6/2006	12.9
7/7/2006	13.0
7/8/2006	12.9
7/9/2006	12.8
7/10/2006	12.7
7/11/2006	13.2
7/12/2006	13.2
7/13/2006	13.3
7/14/2006	13.0
7/15/2006	13.0
7/16/2006	13.2
7/17/2006	13.2
7/18/2006	13.4
7/19/2006	13.1
7/20/2006	13.3
7/21/2006	13.4
7/22/2006	13.4
Average	13.0
StDev	0.3
2 sig	0.5
3 sig	0.8
2 sig. Warning Limits	
Upper	13.6
Lower	12.5
3 sig Control Limits	
Upper	13.8
Lower	12.3

November 2006 Data

Date	Daily Average CO2
11/1/2006	13.4
11/2/2006	13.3
11/3/2006	13.1
11/4/2006	12.9
11/5/2006	12.8
11/6/2006	12.8
11/7/2006	12.9
11/8/2006	13.0
11/9/2006	13.1
11/10/2006	13.4
11/11/2006	13.3
11/12/2006	12.7
11/13/2006	12.7
11/14/2006	12.1
11/15/2006	12.0
11/16/2006	12.0
11/17/2006	11.8
11/18/2006	11.7
11/19/2006	11.6
11/20/2006	11.7
11/21/2006	11.8
11/22/2006	11.7
11/23/2006	11.6
11/24/2006	11.9
11/25/2006	11.5
11/26/2006	11.5
11/27/2006	11.7
11/29/2006	11.1
11/30/2006	11.5


Control Chart Daily Average CO2 Concentration (%CO2)
CO2 Concentration (%CO2)
15
14
13
12
11
10
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Day of Month
+3σ
+2σ
13.000 kg
-2σ
-3σ

Control Chart Daily Average CO2 Concentration (%CO2)