Control Chart Methodology for Detecting Under-reported Emissions

I. Introduction

Control charts are commonly used to measure the extent that a process is operating under control (i.e., stable with variability due only to common sources of variation). If a control chart indicates that a process is not in control, operators can analyze the chart and additional data to identify the source of the variation. The EPA uses the control chart methodology described in this paper to analyze data recorded by gas concentration continuous emission monitoring systems (CEMS) at coal-fired units with certified Part 75 CO_2 monitors. The purpose of the data analysis, which is performed quarterly, is to identify possible sampling system air in-leakage (i.e., probe leaks) or other monitoring system issues that can result in under-reporting of emissions. In the absence of analyses such as the one described in this paper, such problems are often detected only by performing relative accuracy test audits (RATAs) that are typically conducted only once or twice per year.

Control charts are an effective tool for identifying unusual variation in data. A typical control chart, represented in figure 1, below, consists of:

1) Data points representing measurements taken over time;
2) A center line, representing the overall mean of the data set; and
3) Upper and lower control limits representing the outer bounds of statistically-likely data.

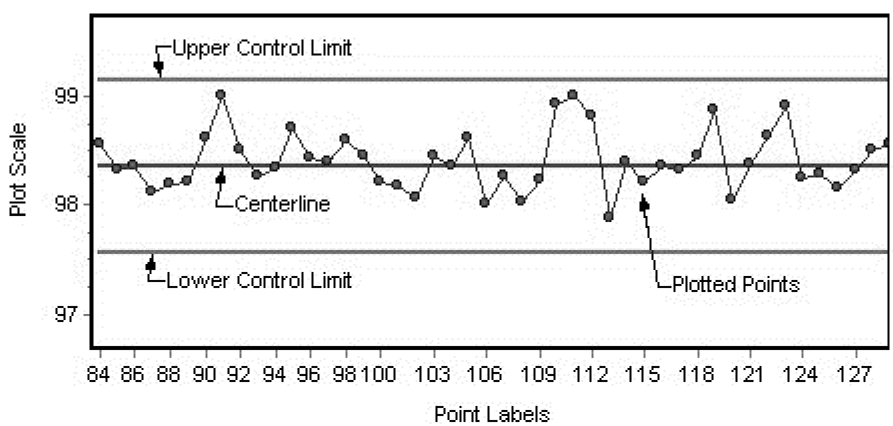

Figure 1. Example control chart

This paper describes how the EPA uses the control chart methodology to identify data below the lower control limit which may be abnormal or suspect and could indicate under reporting of emissions.

II. Data needed for the analysis

- Hourly CO_2 concentration
- Hourly load bin
- Hourly method of determination code (MODC) for CO_2 concentration
- CO_2 relative accuracy test audit (RATA) completion date
- Hourly stack gas flow rate
- Hourly MODC for stack gas flow rate
- Hourly heat input

III. Procedure

Step1: Identify an appropriate load bin for analysis

It is necessary to evaluate data from a narrow operating band to remove (or at least to reduce) the effects that varying operating conditions have on the CO_2 concentration data. The EPA therefore performs a preliminary load bin evaluation in order to identify the most frequently used load bin in the calendar quarter(s) under consideration. The EPA then applies the control chart methodology to the CO_2 concentration data in that load bin, unless the most frequently-used load bin is bin 1 or 2 where startup and shutdown data will invariably be found in these two load bins and unit operation is often unstable (i.e., ramping up or down).

Figure 2, below, demonstrates a typical load bin evaluation. Based on this evaluation, the EPA would apply the control chart methodology to the CO_2 concentration data in load bin 5.

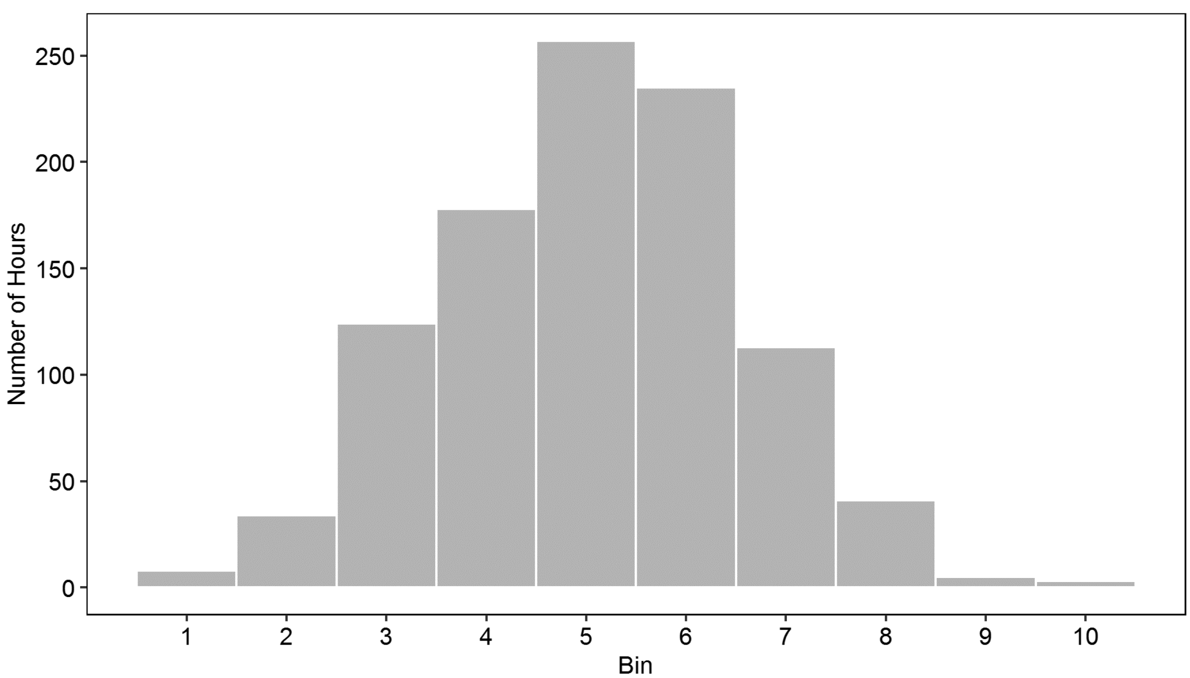

Figure 2. Load bin evaluation

Step 2: Create a daily CO_2 concentration data set

The EPA creates a data set for the control chart analysis using the CO_2 concentration data and MODC for all hourly measurements in the most frequent load bin determined by step 1 above. For each day that there are at least six valid hourly CO_2 concentration measurements[1] in the most frequently used load bin, the EPA calculates the daily average CO_2 concentration using equation A, below. The daily average CO_2 concentration is used instead of the hourly average CO_2 concentration because there is less variation in the daily averages and, therefore, greater certainty that measurements below the lower control limit represents out-of-control operation.

$$\overline{C}_d = \frac{\sum_{1}^{h} C_h}{h} \qquad \textbf{(Equation A)}$$

Where:

$\overline{C}_d$ = Daily average CO_2 concentration for the most frequently used load bin (% CO_2)

C_h = Quality-assured hourly CO_2 concentration in the most frequently used load bin with an MODC of 1 (% CO_2)

h = Number of quality-assured hourly CO_2 concentrations with an MODC of 1 (h must be ≥ 6)

EPA believes that constructing the data set with days where six or more valid hourly CO_2 concentration measurements have been recorded by the primary monitoring system ensures that the daily averages will be meaningful.

Step 3: Determining the baseline parameters

The EPA calculates a performance baseline using the daily average CO_2 concentration data calculated in step 2 above. The baseline is calculated from the first 30 calendar days following each CO_2 concentration RATA. The EPA uses the completion of the CO_2 concentration RATA for the start of the baseline period because any in-leakage or other problems with the CEMS should be identified and corrected during the RATA. Therefore, the data following the CO_2 concentration RATA should represent in-control operation. Note, however, that if there are not 15 or more daily averages within the first 30 calendar days following the CO_2 concentration RATA, the EPA does not perform the control chart analysis.

In figure 3 below, the CO_2 concentration RATA was performed in June. The data points represent the daily average CO_2 concentrations and the shaded area indicates the daily average CO_2 concentration baseline data immediately following the CO_2 concentration RATA. Notice that starting in October the daily averages show a significant decrease in CO_2 concentration.

---

[1] The EPA considers any hourly value with an MODC of 1 to be a valid measurement. This ensures no substitute data are included in the calculation.

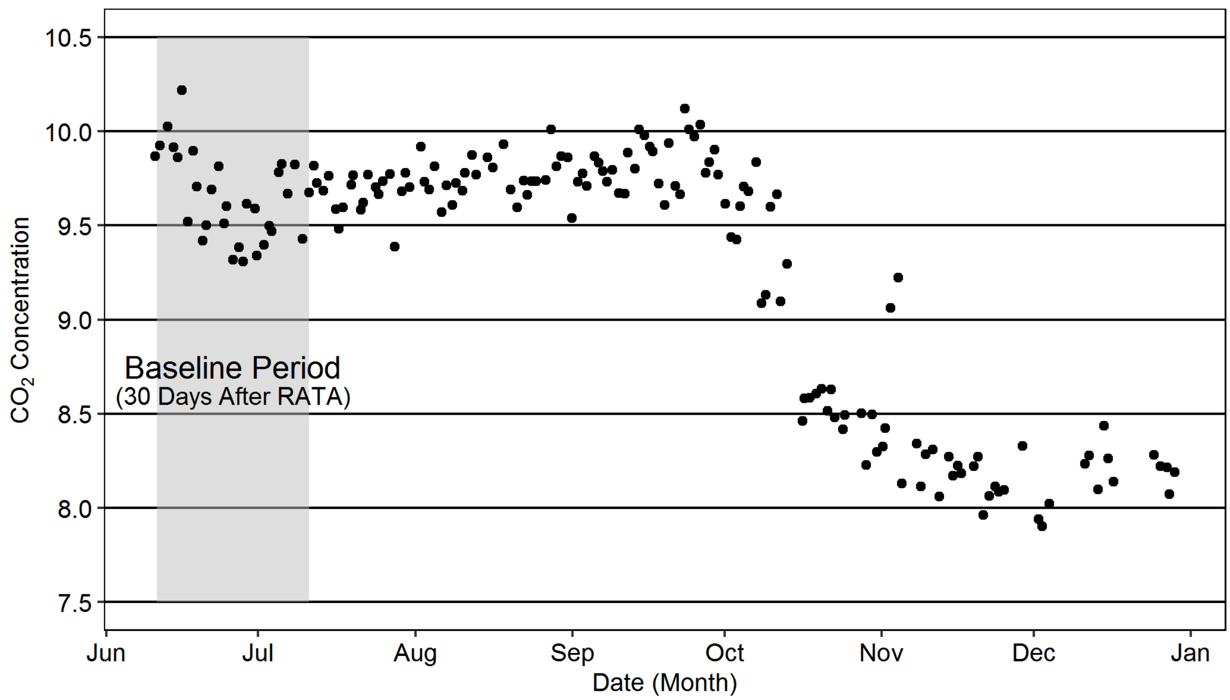

Figure 3. Daily CO_2 averages

The EPA calculates the baseline arithmetic mean and standard deviations using equations B and C, respectively.

$$\overline{C}_B = \frac{\sum_{1}^{d} \overline{C}_d}{d} \qquad \textbf{(Equation B)}$$

Where:

$\overline{C}_B$ = Baseline mean: arithmetic average of the daily average CO_2 concentrations during the baseline period (% CO_2)

$\overline{C}_d$ = Daily average CO_2 concentrations from equation A during the baseline period (% CO_2)

d = Number of days in the baseline period for which a daily average CO_2 concentration was calculated (d must be ≥ 15)

$$\sigma_B = \sqrt{\frac{\sum_1^d \left(\overline{C}_d - \overline{C}_B\right)^2}{(d-1)}} \quad \textbf{(Equation C)}$$

Where:

σ_B = Standard deviation of the daily average CO_2 concentration values during the baseline period

$\overline{C}_d$ = Daily average CO_2 concentrations from equation A during the baseline period (% CO_2)

$\overline{C}_B$ = Baseline mean from equation B (% CO2)

d = Number of days in the baseline period for which the daily average CO2 concentration was calculated (d must be ≥ 15)

Step 4: Calculating the control limits

The EPA calculates control limits to identify data that are outside the expected bounds of in-control operation. The control limits are calculated by adding and subtracting multiples of the standard deviation to or from the baseline mean as shown in equations D-1 and D-2.

$$UCL = \overline{C}_B + 3\sigma_B \quad \textbf{(Equation D-1)}$$

$$LCL = \overline{C}_B - 3\sigma_B \quad \textbf{(Equation D-2)}$$

Where:

UCL = Upper control limit

LCL = Lower control limit

$\overline{C}_B$ = Baseline mean from equation B (%CO_2)

σ_B = Standard deviation from equation C

EPA uses ± 3σ (i.e., three times the standard deviation) for purposes of the control chart analysis to identify daily average CO_2 concentration values that might represent unusual CEMS measurements. This level corresponds to a 99.7% certainty that all normal data fall within the range. Note, however, that if the standard deviation from equation C is very small (less than 0.16% CO_2), the EPA replaces the standard deviation with a default value of 0.16% CO_2 to mitigate potential false positives.

The EPA recommends that sources electing to use the control chart methodology described in this paper to monitor their CEMS operation also establish warning control limits equal to ± 2σ (two times the standard deviation) as shown in equations D-3 and D-4 and take investigative action whenever the daily CO_2 concentration data are outside these warning limits.

$$UWL = \overline{C}_B + 2\sigma_B \quad \textbf{(Equation D-3)}$$

$$LWL = \overline{C}_B - 2\sigma_B \quad \textbf{(Equation D-4)}$$

Where:

UWL= Upper warning limit
LWL = Lower warning limit
$\overline{C}_B$ = Baseline mean from equation B (%CO_2)
σ_B = Standard deviation from equation C

Figure 4, below, shows the baseline mean CO_2 concentration (solid middle line), the upper and lower control limits (dashed lines), and the upper and lower warning limits (dash-dot lines).

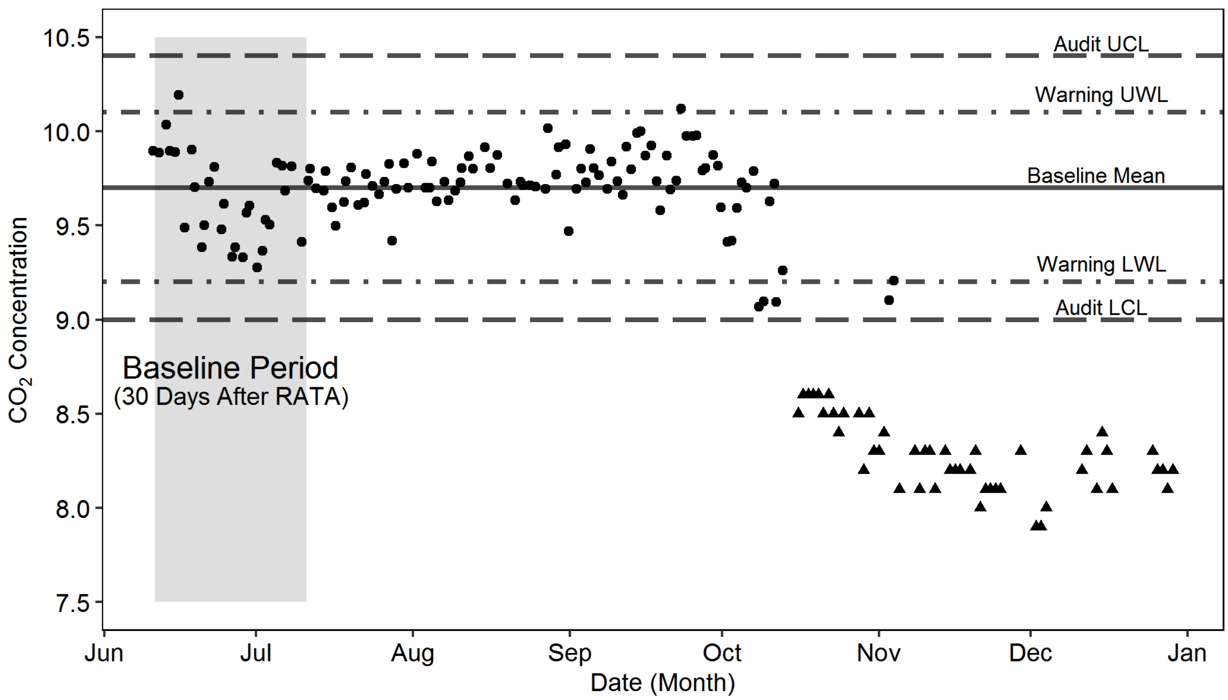

Figure 4. Typical CO_2 control chart

Step 5: Analyzing subsequent daily average CO_2 concentration data

Once the baseline parameters are established, the EPA analyzes the daily average CO_2 concentration for each subsequent operating day. Each daily average CO_2 concentration value is compared to the baseline average and upper and lower control limits from step 4. The EPA identifies all daily average CO_2 concentrations that are outside the lower control limit (triangular markers below the lower dashed line in Figure 4, above), and whenever seven or more consecutive daily average values fall below the lower control limit (i.e., the values are greater than 3σ below the baseline mean) those data are considered to be suspect.

Based on this analysis, the EPA may investigate the suspect time periods by analyzing additional data reported by the source (e.g., supplemental control charts) or by contacting the source for additional information.

IV. Supplemental control charts

In addition to the CO_2 control chart methodology described in this paper, the EPA uses the methodology to construct control charts for stack gas flow rate and heat input for the same operating days that are included in the CO_2 data analysis. As with the CO_2 control chart methodology, at least six quality-assured hourly stack gas flow rates from the same load bin as the CO_2 concentration data are required to determine a daily average. If fewer than six hours of quality-assured flow rate data are found on a given day, that day is not included in the baseline data sets for flow rate, CO_2, or heat input.[2]

These additional control charts are useful in that they can identify changes in operation or detect operational problems such as air in-leakage in the ductwork that, based solely on the CO_2 concentration control chart, could be mistaken for CEMS probe leaks.

V. Recommendations

For on-site quality control, EPA recommends that owners or operators of coal-fired units[3] that are subject to Part 75 implement the above described control chart methodology or a more stringent control chart methodology (e.g., additional load bins, tighter control limits, less than seven consecutive out of bounds days) on a regular basis.[4] Constructing control chart graphs, such as the one represented in figure 4, provides a simple, yet effective way to track monitoring system performance. Data that fall below the lower 3σ control limit (LCL) are of greatest concern, as they may indicate the presence of a probe leak or other monitoring system problem which may cause underreporting of CO_2, nitrogen oxides, and sulfur dioxide emission concentration data.

---

[2] To provide the most meaningful comparison between the CO_2 chart and the supplemental charts, only those days with at least six quality-assured CO_2 concentrations *and* six quality-assured flow rates in the load bin of interest are included in the baseline data sets.

[3] Except for circulating fluidized bed (CFB) boilers that produce significant quantities of non-combustion CO_2 from limestone injection

[4] The EPA has learned that some data acquisition and handling system vendors provide control chart functionality for their customers.