Development of the 1-Hour SO_2 Modeling Emission Inventory for the Ohio Valley Electric Company Kyger Creek Plant and AEP Generation Resources General James M. Gavin Plant

Prepared by
David J. Long, PE
Air Quality Services Section
American Electric Power Service Corporation

Prepared for
Ohio Valley Electric Company
and
AEP Generation Resources
for submittal to
Ohio Environmental Protection Agency

September 15, 2015

TABLE OF CONTENTS

INTRODUCTION …… 1
GENERAL ISSUES WITH THE USE OF CAMD DATA FOR AIR QUALITY MODELING ……..……. 1
KYGER CREEK PLANT SPECIFIC INVENTORY ISSUES ……..... 4
KYGER CREEK PLANT INVENTORY DEVELOPMENT ……. 4
GENERAL JAMES M. GAVIN PLANT INVENTORY ISSUES ……. 8
GENERAL JAMES M. GAVIN PLANT FINAL INVENTORY DEVELOPMENT …… 10
CONCLUSION …… 14

LIST OF FIGURES

Figure 1. Gavin Plant Unit 1 Plant Data vs CEMS Heat Input for 2012 to 2014 …….. 11
Figure 2. Gavin Plant Unit 2 Plant Data vs CEMS Heat Input for 2012 to 2014 …….. 12

LIST OF TABLES

Table 1. Summary of modified hours for CS12 during 2012-2015 …… 6
Table 2. Summary of modified hours for CS35 during 2012-2015 …… 8
Table 3. Summary of modified hours for Unit 1 in the final inventory during 2012-2015 … 13
Table 4. Summary of modified hours for Unit 2 in the final inventory during 2012-2015 … 13
Table 5. Unit 1 emissions comparison in tons per year ……... 13
Table 6. Unit 2 emissions comparison in tons per year ……... 13

INTRODUCTION

As part of the development of recommendations for designations to areas containing large emission sources under the 2010 1-Hour SO_2 Standard, the Ohio Environmental Protection Agency (Ohio EPA) requested assistance from the affected sources in developing air quality modeling inventories for use in air quality modeling. As the primary supplier of technical services to both the Ohio Valley Electric Company and AEP Generation Resources, the Air Quality Services Section of the AEP Service Corporation was tasked with developing this information in a form useful to Ohio EPA.

When inventory development began, it was planned to base the Kyger Creek Plant (Kyger Creek) and General James M. Gavin Plant (Gavin Plant) inventories on the data collected under 40 CFR Part 75 and reported to the USEPA Clean Air Markets Division (CAMD). However, based on the requirements of Part 75, this dataset is not a complete dataset for performing one hour air quality modeling without augmentation. Also, as was found in working with the datasets for these two facilities, large amounts of data substitution, acceptable under Part 75, results in an inventory that requires a large amount of manual effort to validate data and remove large portions of the substituted data to develop an inventory that is representative of the actual performance of the units and emission control systems during the time period being examined. The purpose of this document is to identify the problems with the CAMD data both in general and specifically related to the Kyger Creek and Gavin Plants and detail how the hourly inventory was developed during the periods where data was substituted.

GENERAL ISSUES WITH THE USE OF CAMD DATA FOR AIR QUALITY MODELING

As a result of developing a number of modeling inventories based on the CAMD data, a number of shortcomings have been identified that result in potentially erroneous data being used in air quality modeling, if the data is not carefully examined prior to use. The Continuous Emission Monitoring System (CEMS) data reported to CAMD is used to determine the number of allowances under the Acid Rain Program and now under the Cross State Air Pollution Rule that must be surrendered each year to cover SO_2 and NO_x emissions from affected facilities. The data is reported at an hourly level, which gives the appearance that this should be an excellent source of modeling inputs. However, such is not necessarily the case. This is especially true when the "high level" data is all that is extracted for use.

For purposes of this discussion, "high level" data consists of the emission value in lb/hr, the stack flow rate in units of wscfh (wet standard cubic feet per hour), and potentially the unit gross load (generally in gross MW generated during the hour). The individual sensor values in ppm

for emissions are reported, along with a number of data flags that amplify the source and quality of the data, but these items are not generally collected as part of the "high level" data. Further data submitted includes a diluent gas, typically CO_2, bias adjustment factors for each sensor as determined during the most recent RATA Tests, and a calculated heat input value that uses the diluent gas and flow rate along with the bias adjustment factors for each input. It should be noted that the bias adjustment factor is applied at the sensor level and if multiple sensors are used in the equation to develop the "high level" value, multiple bias adjustment factors can come into play. As defined in Part 75, bias adjustment factors also only raise values, as the minimum bias adjustment factor is 1.0. The equations used to develop these values under Part 75 are as follows:

Emission Rate Equation

$$E_h = KC_h(Q_h * BAF) \quad \text{(Eq. F.1)}$$

- E_h = Hourly emissions in lb/hr
- K = 1.660 x 10^{-7} for SO_2
- C_h = SO_2 concentration in ppm
- Q_h = Flow rate in scfh
- BAF = Bias Adjustment Factor >= 1.00

Flow Rate Equation

$$Q_h = \frac{Ts}{Tm}(Q_m * BAF)$$

- Q_h = Hourly flow rate in scfh
- Ts = Standard Temperature (528 R)
- Tm = Measured Stack Temperature (R)
- Q_m = Monitored Flow rate in acfh
- BAF = Bias adjustment factor >= 1.00

Heat Rate Equation

$$HI = (Q_w * BAF)(\frac{1}{F_c})(\frac{CO_{2w}*(BAF)}{100}) \quad \text{(Eq. F-15)}$$

- HI = Hourly Calculated Heat Input
- Q_w = Flow Rate in wscfh
- Fc = Carbon based F Factor
- CO_{2w} = Percent CO_2 Wet
- BAF = Parameter based bias adjustment factor >= 1.00

It should also be noted that temperature data is used in the flow equation, but not mentioned in the description of the "high level" data. This is because temperature data is collected by the monitoring systems to allow the flow data to be converted from actual cubic feet per hour at stack conditions to standard cubic feet per hour for reporting, but the temperature data is not collected by CAMD as part of its reporting requirements.

In addition to these issues, the data used in the equations does not have to be generated using a continuous monitor under some conditions. These cases typically exist during the period when a new emission control device is being started up along with a new stack. This results in no data being initially available from the continuous monitoring system until appropriate testing has been done. This data can be substituted using worst case data from the previous stack and monitoring system or by a series of stack tests.

Data substitution is routinely used for any number of reasons under Part 75. Excessive data substitution or certain types of substitution can result in unreliable "high level" data for air quality modeling purposes. In those cases where the generating unit is operating in a relatively stable configuration, the hour before – hour after technique that may be used under specified conditions for Part 75 generally results in a substituted value that is nearly indistinguishable from the measured data. However, if this technique is not available to the source other techniques will need to be used, up to and including substituting a value for the missing sensor that is twice the maximum design value of the parameter. Under Part 75, the purpose of which is to collect allowances for emissions, data substitutions and adjustments that result in unrealistically high emissions serve to penalize the source in terms of allowance consumption as was the design of the program. However, these substitutions and biases will likely serve to generate unrealistic data for air quality modeling purposes.

As this work was moving forward and the above issues were identified as critical impactors on the development of the modeling inventory, a presentation on the potential impacts of directly using CAMD sourced data was developed and delivered at the USEPA 11th Conference on Air Quality Models as a cautionary statement on these issues[1]. While the inventory developed for Kyger Creek was not significantly impacted by these issues except for the period where the Units 1 and 2 FGD System and new stack were commissioned, the inventory for the General James M. Gavin Plant was significantly impacted by data substitution issues, resulting in the need to perform extensive data substitution and quality control activities outside of the CAMD data.

[1] Long, David J., The Use of CAMD Data for Air Quality Modeling, presented at the 11th USEPA Triennial Modeling Conference, Research Triangle Park, NC, August 12-13, 2015.

KYGER CREEK PLANT SPECIFIC INVENTORY ISSUES

The Kyger Creek Plant, as modeled for the 1-Hour SO_2 Standard, consists of a single 830 foot tall concrete stack with two separate discharge ducts. One duct serves the FGD receiving gas from Units 1 and 2 with a diameter of 7.53 meters and the other ducts serves the FGD receiving gas from Units 3, 4, and 5 and has a diameter of 9.23 meters. Each duct has its own test ports and CEMS.

During the period modeled, January 1, 2012 to December 31, 2014, Units 1 and 2 were initially connected to the FGD stack during the first quarter of 2012 with initial operation of that FGD discharge commencing in February 2012. This startup period resulted in an extensive period of missing, invalid, and substituted data for the Unit 1 and 2 discharge during the first half of 2012. This period of missing data varied by parameter, with the flow data having the most missing hours of data. Beyond this period, the amount of missing, invalid, and substituted data for both discharges was not remarkable when compared to other inventories that have been developed.

The data used for the inventory development consisted of hourly resolution emission data (lb/hr), flow (in wscfh), and gross load (MWe) reported to CAMD and was sourced from a data warehouse maintained by AEPSC for all AEP owned and affiliated stacks. This data warehouse receives a feed from the individual loggers of Part 75 QA data following the routine checking and QA of the data by the individual facility. Temperature data was also sourced from the data warehouse as it is also reported and stored in the data warehouse, along with all logger and data flags and other parameters that are reported to the logger, but may or may not be reported to CAMD in the quarterly reports. Temperature data was available for most of the initial startup period for Units 1 and 2, since that data was QA checked, but not submitted to CAMD.

KYGER CREEK PLANT INVENTORY DEVELOPMENT

The development of the Kyger Creek Plant inventory was begun by sorting the data extracted from the Data Warehouse by stack (discharge duct), parameter, and date and time. To facilitate handling of the data, the data for each discharge duct was loaded into separate Excel Spreadsheets for review and then copied into a third spreadsheet to create the HOUREMIS format data needed for use in AERMOD.

Units 1 and 2

The data for the Units 1 and 2 duct (tagged CS12 in the inventory), was known to have a period in the first and second quarters during which the new FGD system and associated CEMS system was being started up for the first time. This resulted in a period where a large amount of missing

and substituted data was generated due to the CEMS not being certified since this was a new stack and CEMS system. Part of this period was covered by data generated by stack testing that was submitted as part of the quarterly submittal and was present in the data warehouse. However, when stack test data was not available, missing data substitution was used from the old unscrubbed stack that served all five units. These values were very easy to identify due to the large flow values reported and excessively high emission rates in the data. While the CEMS data was invalidated, the temperature data was generally not invalidated and after review was considered generally acceptable for processing. With this in mind, the data for CS12 was extracted from the data warehouse and reviewed for missing data, periods of no operation on the duct so that zero values could be placed into the data set if there were no numeric values present. The Kyger Creek data logger does not report a single temperature value to the data warehouse for either discharge duct. As a result of having multiple temperature values to consider, the temperature data for each duct was initially processed to select the lower of the two temperature measurement values (designated as left and right) being stored in the data warehouse in an effort to give a measure of conservatism to the modeling analysis using the data.

Other than during the period of FGD and CEMS startup, where missing data was found for emissions or flow, the initial substitution was to use a simple hour before hour after average value. However, if the period was typically over four or five hours and there was data suggesting that a change in operations had occurred during the period of missing or invalid data, other techniques which included the following were used:

- a modified hour before hour after approach that allowed the hour before to float after the initial hour of substitution that would allow the substituted data to flow more naturally into the next accurate value.
- for cases where there were multiple changes in operations noted based on unit loading, a binning by load technique was established where each bin had at least 5 hourly values taken from a seasonally representative period around the time of the missing data that were then loaded into the dataset on an hourly basis based on the reported gross load for the hour being substituted. This technique was used mostly on startups where valid data may have been spotty for various reasons.

As previously discussed, the temperature values were developed by selecting the lower of the two sensor values in the database for a given duct. Where the selected value was zero or missing and the stack was in service (one or more units in operation), the first attempt at obtaining a valid temperature value was done by changing the selected value from the minimum to the maximum value. Should the value remain zero after the switch from minimum to maximum was made, then data substitution techniques as described above for the missing and substituted flow and emission data were used.

During the initial startup period for the Unit 1 and 2 FGD and CEMS System, large amounts of CEMS data were deemed missing or invalid for various reasons that included the startup of the system prior to certification of the CEMS under Part 75. For the emissions data used in the modeling inventory, the bulk of this data was able to be substituted using stack test data that served as an excellent source of data that is shown as modified data in Table 1, but is valid data under Part 75. Emission data during the startup period that was not substituted using stack test data, with the remaining hours developed using either hour before hour after or the binning techniques described earlier. The manually corrected hours during the startup period are called out separately from the balance of 2012 in Table 1 through their listing under the heading of 2012 FGD Startup Period. The FGD startup period for the SO_2 monitor ran for 198 hours starting on February 16, 2012 and ended on February 24, 2012, while the flow monitor startup period ran until April 24, 2012.

While correcting the SO_2 emissions data during the startup period was relatively simple due to many of the hours having had stack testing performed, the flow monitor did not receive similar treatment and was not deemed fully operational until April 24, 2012. During this period, the flow was established using binned data across the load range observed during the period based on measured data from after the monitor was declared in service.

Table 1 shows the number of hours on CS12 that were modified in some fashion. The table demonstrates that for CS12, except during the period where the new FGD and CEMS were being started up and various substitution methods were being employed, less than 1% of the hours in any given year for a specific parameter required substitution techniques to be implemented.

Table 1. Summary of modified hours for CS12 during 2012-2015.

Parameter	**2012 FGD Startup Period**	**2012 Non FGD Startup**	**2013**	**2014**	**Total**
Flow Values	**1350**	**2**	**0**	**18**	**1370**
Emission Values	**198***	**53**	**59**	**13**	**323**
Temperature Values - Substituted	**0**	**0**	**3**	**0**	**3**
Temperature Values – Max Used	**335**	**113**	**0**	**11**	**459**

*Includes 30 hours of substituted data and 168 hours of stack test data.

Following the completion of the substitution of missing and Part 75 substituted values, the flow values were converted into velocities by converting the flow from wet standard cubic feet to actual cubic feet using the temperature data for the hour being converted, then dividing the flow by 3600 times the area of the discharge and finally the feet per second value was converted into meters per second to obtain the discharge velocity required for use in AERMOD. Similarly, the emission rate was converted from lb/hr to grams/hour then divided by 3600 to convert the emission rate into grams per second for use in AERMOD. Finally the temperature was converted from °F to °K. Once these conversions were completed, the values were then transferred using an automated program into another spreadsheet that was then formatted to generate an HOUREMIS formatted AERMOD ready input file for the selected emission points.

Units 3 - 5

The data for the Units 3-5 duct (tagged CS35 in the inventory), was reviewed for missing data, periods of no operation on the duct so that zero values could be placed into the data set if there no numeric values present, and the temperature data was initially processed to select the lower of the two temperature values, designated left and right, being stored in the data warehouse.

Where missing data was found for emissions or flow, the initial substitution was to use a simple hour before hour after average value. However, if the period was typically over four or five hours and there was data suggesting that a change in operations had occurred during the period of missing or invalid data, other techniques which included the following were used:

- a modified hour before hour after approach that allowed the hour before to float after the initial hour of substitution that would allow the substituted data to flow more naturally into the next accurate value.
- for cases where there were multiple changes in operations noted based on unit loading, a binning by load technique was established where each bin had at least 5 hourly values taken from a seasonally representative period around the time of the missing data that were then loaded in on an hourly basis based on the reported gross load. This technique was used mostly on startups where flow data may have been spotty for various reasons.

As previously indicated, the temperature values were initially handled by selecting the lower of the two sensor values in the database. Where the selected value was zero or missing and the stack was in service (one or more units in operation), the first attempt at obtaining a value was by changing the selected value from the minimum to the maximum value. Should the value remain zero after the switch from minimum to maximum was made, then the data substitution techniques described above for the missing and substituted flow and emission data were used. Table 2 shows the number of hours on CS35 that were modified in some fashion. The table

demonstrates that for CS35 less than 1% of the hours in any given year for a specific parameter required substitution techniques to be implemented.

Table 2. Summary of modified hours for CS35 during 2012-2015.

Parameter	2012	2013	2014	Total
Flow Values	**87**	**10**	**17**	**114**
Emission Values	**0**	**1**	**4**	**5**
Temperature Values - Substituted	**9**	**37**	**17**	**63**
Temperature Values – Max Used	**62**	**9**	**83**	**154**

Following the completion of this work, the flow values were converted into velocities by converting the flow from wet standard cubic feet to actual cubic feet using the temperature data for the hour being converted, then dividing the flow by 3600 times the area of the discharge and finally the feet per second value was converted into meters per second to obtain the discharge velocity required for use in AERMOD. Similarly, the emission rate was converted from lb/hr to grams/hour then divided by 3600 to convert the emission rate into grams per second for use in AERMOD. Finally the temperature was converted from $^{\circ}$F to $^{\circ}$K. Once these conversions were completed the values were then transferred using an automated program into another spreadsheet that was formatted to generate an HOUREMIS formatted AERMOD ready input file for the selected emission points.

GENERAL JAMES M. GAVIN PLANT INVENTORY ISSUES

The development of the modeling inventory for the General James M. Gavin Plant (Gavin Plant) started in much the same way as did the development of the Kyger Creek Plant inventory by extracting the flow, SO2 emissions, and the two temperature data points captured on Units 1 and 2 from the data warehouse. However, the initial inspection of this data found large amounts of apparently anomalous data in the flow and emission values obtained. After several attempts to understand and correct the data using the techniques described in the Kyger Creek portion of this report, it was determined that this effort was not leading to a high quality modeling inventory and other methods of checking the “high level” CAMD data were pursued.

The first method attempted was to extract the data flags for the “high level” data from the data warehouse. When the flags were inspected, it revealed hundreds of hours of the “high level” data that had potential liabilities due to the substitution of various data elements, but from the “high level” data, it was unclear which element(s) were impacting the values. This data suggested that the sensor level data for the various parameters should be examined in an effort to

determine if the "high level" data was correct as shown or needed to be corrected in some fashion. This step did not result in the development of a new version of the inventory, just the determination of the need to dig deeper into the data to determine what was going on.

The first of these attempts resulted in the extraction of the hourly gross generation data for both units, the hourly CO_2 values, heat input data, and the ppm SO_2 data from the data warehouse to allow a closer look at the CEMS SO_2 values in a more raw form than was reported to CAMD and to allow an independent evaluation of the hourly stack emission rate for comparison to the "high level" values that had large amounts of anomalous data. Hourly emission rates were generated using the wet measured pollutant and diluent gas equations from USEPA Method 19 as follows:

$$E = C_w F_c \frac{100}{\%CO_{2w}}$$

- E = Emission rate in lb/MMBtu
- C_w = Pollutant Concentration in ppm from CEMS
- CF = Conversion Factor from Table 19-1 in Method 19
- F_c = CO2 based F Factor
- CO_{2w} = Percent CO_2 in the gas

While this evaluation confirmed the rejection of a number of hours of elevated "high level" concentrations, it did not offer a clear justification for rejecting a number of hours of data, both high and low, that still looked out of line for the expected performance of the unit. Following a discussion with operational personnel, an effort was made to locate outlier data using an expected operating range of 0.30 to 0.55 lb/MMBtu . While this did allow for much of the remaining elevated SO_2 data to be rejected, it also resulted in the rejection of large amounts of low concentrations since they fell below the 0.30 lb/MMBtu levels that we were advised were the low end of the operating range for the "normal" coal being consumed. Using this range, the CEMS data rejection rate for SO_2 concentration in ppm was in excess of 50% for some months in the 2012 to 2014 period, with the rejections being dominated by low values. Once a set of hourly emission rates were developed, then it was necessary to develop a set of hourly heat inputs to allow the conversion of the data into lb/hr and ultimately g/sec values for insertion into the modeling inventory in place of the "high level" data from CAMD.

Obtaining the heat input data on an hourly basis became a problem, as the data reported to CAMD was found to be of very poor quality at the basic inspection level. Based on the fuel being consumed at Gavin Plant and the design of the units, except for winter periods where the weather was near or below 0 °F , it was clear that much of the heat input data was biased high, sometimes by factors in excess of 2 times the practical maximum heat input for the units, which was established at 13,500 MMBtu/hr, based on operational experience with the 1300 MW series

of units. This can be seen in Figures 1 and 2 which show the plot of the CEMS heat input data reported to CAMD with the final method used to establish the heat input based on available plant data. These plots show how badly biased to the high side the CEMS heat input data is. This issue resulted in the eliminated the CEMS data as a source of heat input data to develop the hourly emissions, so an alternative source of heat input data was sought. The problem with developing the alternative was that all of the heat rate data that was available for the Gavin Units was based on net generation (gross generation minus the internal load of the unit) and the available heat input data from the CEMS System and CAMD data was in terms of gross generation. After several attempts to adjust the readily available heat rate data into a form that would be compatible with gross load data and generate a viable emission inventory, this effort was terminated when the inventory would not pass the Quality Assurance checks.

In addition to the inability of the inventory to pass the Quality Assurance checks, the high data rejection rates for the SO_2 concentration data raised questions over the methodologies being used for the development of the lb/MMBtu data. This concern suggested that a better method of determining the operational characteristics of the unit were needed that included a means to determine the reasonableness of the stack SO_2 ppm data as well as a more realistic means of developing the hourly heat input data. This effort is discussed in the following section.

GENERAL JAMES M. GAVIN PLANT FINAL INVENTORY DEVELOPMENT

Development of the final modeling inventory started by extracting the two FGD inlet SO_2 Sensor ppm values from the data warehouse to allow an evaluation of the inlet SO2 loading and the FGD performance, which should at most times be between 90 and 95 percent. The inlet loading values were averaged where both sensor values were available and used in conjunction with the stack SO_2 value reported to CAMD to generate an approximate hourly FGD efficiency. Unit load was also mated with this data to allow an evaluation of the load conditions on the unit generating the FGD efficiency values. Using these data points, the stack SO_2 values were analyzed to determine if they were reasonable based on the coal being consumed as described by the FGD inlet SO_2 loading values and the load.

Using this data, it allowed many of the low SO_2 ppm values to be accepted since they occurred during periods when the FGD inlet loadings indicated lower sulfur coals were being consumed and/or reduced load conditions existed. It also allowed a number of hours of higher SO_2 ppm values to be retained as they proved to be consistent with the consumption of a higher sulfur coal while obtaining appropriate FGD performance. Even with this new data, there were still a large number of hours where the discharge SO_2 ppm value required adjusting to get it into a reasonable range, suggesting substituted or otherwise invalid data for modeling purposes was involved. Following the examination of the SO_2 data, the CO_2 data was also examined in an effort to

correct any values that were not consistent with the unit operating conditions. This allowed a new set of lb/MMBtu values to be generated, but the heat input data issue had still not been addressed.

Since the problem related to obtaining a matched set of net load and net heat rate values, available plant operating data was examined to see what items might be available to resolve this portion of the issue. Using the Plant Information System, we were able to obtain hourly calculated net heat rate values along with the net generation values. This would allow the calculation of a heat input on a purely net heat rate to net generation basis. Where values were missing, they were back filled using the same techniques described in the Kyger Creek portion of the report for missing and invalid data substitution. This data generated a set of heat input values that were more reasonable than the values generated by the CEMS system. Figures 1 and 2, covering Units 1 and 2 respectively, show plots of both the CEMS and Plant Information System derived hourly heat inputs. What is striking is that the CEMS data has many thousands of hours above the practical maximum heat input level, while the plant derived data has very few hours above that value. It is also noted that on Unit 1, the CEMS data has a number of hours where the heat input went too low relative to the load on the unit at the time. While Unit 2 does exhibit some of this low heat input behavior, it is not observed as extensively as it was on Unit 1.

Figure 1. Gavin Plant Unit 1 Plant Data vs CEMS Heat Input for 2012 to 2014.

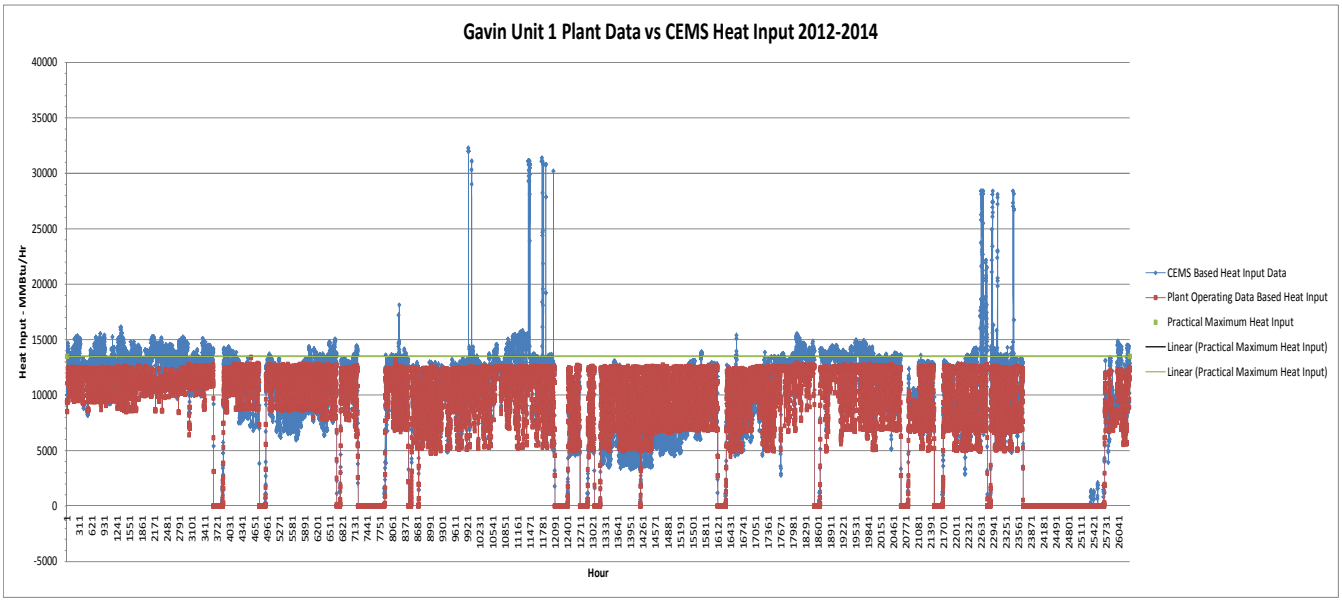
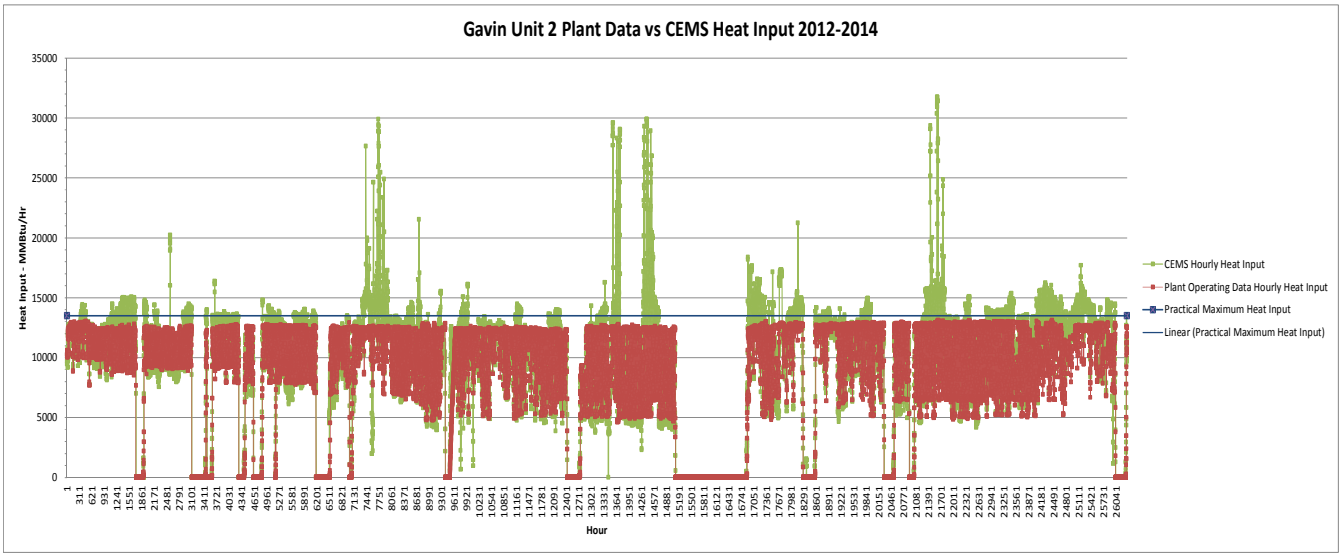


Figure 2. Gavin Plant Unit 2 Plant Data vs CEMS Heat Input for 2012 to 2014.



Once this data was prepared and an hourly heat input was developed, that data was transferred to the spreadsheet where the hourly SO_2 data had been placed and used to generate an hourly emission rate in lb/hr and g/sec for use in the modeling inventory. In the process of transferring the data, color highlights were used to flag the hours where any parameter (ppm SO_2, $\%CO_2$, or heat input) had been substituted or changed as part of data development process for accounting purposes, since there were a number of different spreadsheets used for developing the values from the raw data. The derived "high level" data was then transferred into a formatted data sheet to generate the HOUREMIS file needed for use in AERMOD.

Even with these improved techniques, there were a number of hours where the various modeling parameters were manually changed. Tables 3 and 4, showing Units 1 and 2 respectively, identify the number of hours by parameter where the values were substituted or changed for some reason from the original CEMS based data. It should be noted that all hours of "high level" emission data were discarded and substituted with the generated data described above. The Emission Values row in the table reflects the number of hours that were changed from the raw data components of CEMS SO2 ppm, CEMS %CO2, and missing data from Plant Information System net heat rate and net load. The rows listing Temperature values represent the cases where there was no temperature data available (Substituted) and where only one sensor was operating properly (Max Used).

Table 3. Summary of modified hours for Unit 1 in the final inventory during 2012-2014.

Parameter	2012	2013	2014	Total
Flow Values	**46**	**160**	**122**	**328**
Emission Values	**1067**	**1070**	**1337**	**3474**
Temperature Values - Substituted	**5**	**13**	**233**	**251**
Temperature Values – Max Used	**186**	**2496**	**1523**	**4205**

Table 4. Summary of modified hours for Unit 2 in the final inventory during 2012-2014.

Parameter	2012	2013	2014	Total
Flow Values	**21**	**251**	**219**	**491**
Emission Values	**634**	**1190**	**1137**	**2961**
Temperature Values - Substituted	**10**	**0**	**153**	**163**
Temperature Values – Max Used	**177**	**347**	**1678**	**2202**

With all of the data manipulations that were made to generate the modeling inventory, the amount the emissions portion of the inventory changed should be noted. Tables 5 and 6 show the annual differences between the original CEMS based SO_2 emissions for Units 1 and 2 respectively and the final SO_2 values used in the modeling simulations. The change in annual emissions difference ranges between 7% to 24% in each year on each unit. However, the hourly changes can be much larger and show both increases and decreases depending on the nature of the unit operation in a given hour and any changes made to the data.

Table 5. Unit 1 emissions comparison in tons per year

Year	CEMS Reported	Modeling Inventory
2012	15,992.2	14,220.3
2013	14,721.9	13,755.1
2014	16,682.8	12,699.4

Table 6. Unit 2 emissions comparison in tons per year

Year	CEMS Reported	Modeling Inventory
2012	15,295.3	14,006.8
2013	13,133.9	12,242.8
2014	20,207.7	17,220.9

CONCLUSION

The efforts described in this paper used to develop the hourly modeling inventory for the Gavin and Kyger Creek Plants were successful and demonstrate that the quality of the CAMD Data Sets used by some in air quality modeling are of highly varying quality and not generally suitable for use directly in air quality modeling. With the exception of the startup period for the Unit 1 and 2 FGD System and discharge at Kyger Creek Plant, the Part 75 based data that can be obtained from CAMD required additional quality assurance measures to be used to prepare it for use in AERMOD, but did not require significant modification of the data itself to generate a reasonable representation of hourly operations at Kyger Creek Plant that were suitable for use in an air quality model. However, such was not the case with the Gavin Plant.

Yes, the Gavin Plant CAMD Data Set did pass quality assurance checks under Part 75 and represented a full annual data set for each of the three years examined. However, the amount of substituted data in the data set rendered the emission data completely unusable and the flow data usable with a significant amount of review. Once the data was corrected to eliminate much of the Part 75 substituted data, it was deemed a reasonable and more accurate representation of the Gavin Plant operations during the 2012-2014 period for air quality modeling purposes than the CAMD Data Set that was initially examined.

Finally, the CAMD Data Sets are deficient in that they do not contain the hourly temperature data required as an input to an air quality model and to convert the flow data back into actual cubic feet per minute from wet standard cubic feet as reported to CAMD. This deficiency is due to the fact that CAMD does not collect hourly temperature data as part of the quarterly Part 75 data reporting process, even though it is measured to allow the Part 75 required adjustments to the actual flow values to be made.