Examination of United States Carbon Dioxide Emission Databases

R. Paul Borthwick, James Whetstone, Jiann C.Yang, and Antonio Possolo
National Institute of Standards and Technology
Gaithersburg, MD 20899, U.S.A.

Introduction

Recent concerns with long term climate change effects due to carbon emissions from human activity point to a need for reliable and accurate CO_2 emission measurement techniques. To appreciate the magnitude of the problem, during 2007 approximately 6.7 billion tons of carbon dioxide was emitted as a result of human activity in the US alone [1]. There are numerous discussions on the need to limit, ways to limit, and incentives to encourage voluntary reductions to carbon dioxide emissions. For any discussions to be meaningful, effective and consistent methods for measuring carbon dioxide emissions must be available.

While present methods for measuring or estimating carbon dioxide emissions may be suitable for estimation or academic discussion, there are also proposals that carbon dioxide emissions be limited though use of either a trading or a taxation system. While the details of such systems are outside the scope of this report, the end result is that carbon dioxide emissions would become a traded or taxed commodity. In order for either type of system to work, measurement uncertainty far better than the 6 % -10 % estimated by an early carbon dioxide emissions study [2] or 5 % from recent Department of Energy documentation [3] will be required.

There exist two fundamental ways to estimate carbon dioxide emissions from an industrial facility. The first method is to estimate the amount of carbon introduced into the process by accounting for the carbon in the fuel or process feedstock and assume that for combustion processes all or a large fixed percentage of the carbon introduced to the process is emitted from the plant. The second method is to measure the carbon dioxide concentration and the volume of exhaust gases being emitted from the plant and calculate the total mass of carbon dioxide within the emissions.

Measurements of the carbon dioxide emissions, either derived from the composition of fuels and the assumption of almost complete combustion, or made directly on the products of combustion, have associated uncertainties that can be substantial. For carbon measurement prior to combustion, there can be variations in the amount of carbon present from one lot of fuel to the next. For solid fuels, such as coal, varying amounts of moisture can influence the heating value of the fuel affecting some methods of carbon emissions calculation. Although losses are likely small, there is also no guarantee that all coal on a conveyor or oil or gas in a pipe makes it to the combustor. For post combustion event systems that rely on carbon dioxide concentration and gas volume flow measurements, there is concern that present measurement techniques have too large an uncertainty in the gas volume flow measurement to produce suitably accurate results. The main source for gas volume measurement uncertainty is likely variations and asymmetries in the velocity flow field and turbulence in the measurement section of the stack.

This paper will compare carbon dioxide emissions from plants obtained in two ways, first from the US Department of Energy, Energy Information Administration EIA-767 database [4] and

EIA-1605 Greenhouse Gas Reporting Program guidelines [5] and second from the U.S. Environmental Protection Agency (EPA) eGRID (Emissions & Generation Resource Integrated Database) database [6].

The EIA-767 database is a report covering the plant, boilers, stacks, emission control system and most important for this discussion, all fuels and the quantities consumed in the generation of steam for electric power. EIA-767 data is collected for all steam-electric power plant operations 10 MW or greater [7]. The comparison in this paper focuses on the year 2005. Administratively, 2005 was the last year for the EIA-767 report. For 2006 the Energy Information Administration did not collect the information requested by form EIA-767 and starting in 2007 much of the information was covered in either form EIA-860 or EIA-923.

The EIA-767 database collects information regarding the major physical and operating attributes of steam-electric power plants. This also includes the steam-electric portion of electric power generating facilities where the majority of the electricity may be generated by non-steam cycle means. EIA-767 collects information starting with plant name and location then moving to plant specifics such as, number of boilers, generators, stacks, cooling system(s), emissions control equipment and most important for this study: type, amount and heating value of all fuels consumed, tabulated on a monthly basis.

EIA-1605 stems from a voluntary reporting of Greenhouse Gasses initiative and is an outgrowth of the Energy Policy Act of 1992. The primary use for the EIA-1605 report [5] and instructions is a comprehensive appendix equating energy release with carbon dioxide emission by fuel type. While there are several sources for energy to carbon dioxide information, the EIA-1605 appendix was chosen for emission factors as it uses the same set of fuel type identifiers as the EIA-767 report, eliminating ambiguity in applying carbon dioxide emission factors to the fuels when calculating carbon dioxide emissions. Carbon emission factors are scalar quantities assigned to the fuel types and are defined in units of carbon or carbon dioxide released per either mass of fuel consumed or amount of energy released from the fuel. The EIA-1605 [5] Appendix H lists Fuel Emission Factors in units of kg CO_2/MMBtu (millions of British Thermal Units) for 29 different fuels. Other documents use different combinations of units and may cover fewer or more fuels.

The eGRID database, as noted, is published by the U.S. Environmental Protection Agency. It contains minor information regarding the physical attributes of a given power plant or boiler system but concentrates on measured emissions from each boiler. The eGRID database captures information regarding carbon dioxide, nitrous oxides, sulfur dioxide, methane and mercury emissions. Not only does the database tabulate total emissions, there is an effort to rate the emissions per unit of electrical or thermal output. An important aspect, and limitation, of the eGRID database is that it lists only the primary fuel consumed by a specific boiler, not all fuels consumed in that boiler.

The U.S. EPA describes the eGRID database as “a comprehensive inventory of environmental attributes of electric power systems” [8]. The eGRID database is specifically intended to report emissions data while the EIA-767 database collects only physical attributes of the plant and details on the types, amount and qualities of the fuel. The eGRID database records very minimal information on the fuel(s) used and concentrates on total emissions. The eGRID database also records how the emission data is obtained, either by continuous emission monitoring (CEM) measurement or from calculation based on carbon emission factors for materials entering the plant. Entries in the EIA-767 database and the eGRID database can be matched and compared as both eGRID and EIA databases use the same plant and boiler identifiers. The use of the same plant and boiler identifiers between the databases has been utilized by earlier studies comparing carbon dioxide emission values derived from the two different data sets [9].

Analysis

The objective of this analysis is to compare carbon dioxide emissions calculations based on fuel types and heat input, referred to as EIA carbon dioxide calculation, to carbon dioxide emissions reported as continuous emission monitoring (CEM) measurements. The analysis is expected to reveal two things, the global systematic error between measurements obtained by the different methods and an element of random error between these measurements.

The eGRID database reports total measured carbon dioxide emissions. However, it is impossible to calculate potential carbon dioxide emission from a given boiler by examination of the eGRID database. The eGRID database lists only the primary fuel consumed and the quantity of heat input into the boiler. Primary fuel is determined as the fuel that was used for the majority of heat input for the survey year. By definition, there is no guarantee that all of the heat input into the boiler was from the primary fuel source. If the primary fuel represents close to 100 % of all fuel consumed by a specific boiler, an appropriate carbon dioxide emission factor could be applied, and there would be little analytical error introduced in the comparison of the calculated versus measured carbon dioxide emissions. For the eGRID database there is no assurance that the primary fuel reported provided any more than a majority, however small, of the heat input from all fuels to the boiler for the reporting year. This aspect can leave a large percentage of fuel input to a boiler unreported in the eGRID database.

Comparison of boilers identified in the eGRID database to the boilers listed in the EIA-767 database is essential to a full accounting of all fuels consumed by a given boiler. The information utilized in the analysis comes from the eGRID file "eGRID2007V1_1_year05_plant.xls" under the tab "BLR05" and the EIA-767 file "F767_BOILER_FUEL.xls". The EIA-767 spreadsheet contains only one tab "F767_BOILER_FUEL". Both databases report on the year 2005, the last year for which the data analyzed was reported on the cited forms.

While the eGRID database lists each boiler only once, the EIA-767 Boiler Fuel database will list many boilers twice and occasionally three or four times. Each listing of a boiler in the EIA-767 database will provide the detail on a different fuel used in that boiler. The details in the EIA-767 database include each fuel type utilized, quantity consumed by month and heating value, also by month. Therefore, the EIA-767 database provides detailed fuel and quantity information that can be used to calculate the carbon dioxide emission from any listed boiler. The amount of carbon dioxide emitted that is calculated from EIA-767 data can then be compared directly with the annual carbon dioxide emissions measured by a continuous emission monitoring technique used for reporting in the eGRID database.

To calculate carbon dioxide emissions from the EIA-767 database, carbon dioxide emission factors for the fuels must also be used. The carbon dioxide emission factors, which equate carbon dioxide emission for each type of fuel used based on quantity of heat release, were obtained from Appendix H in the Instructions for Form EIA-1605 [5]. With these additional factors a monthly carbon dioxide emissions number for each subject boiler can be calculated.

The general procedure for identifying plants where all necessary information from both databases is available to compare carbon emissions by both calculated and measured methods is shown by the flowchart in figure 1.

As shown in the flowchart there are several conditions that must be met to have a specific boiler be useful for this analysis. For cases where all conditions are met, a valid comparison between calculated and measured carbon dioxide emissions can be made. Within the eGRID database the first criteria to be met was a primary fuel of either bituminous coal (BIT), sub-bituminous coal(SUB), synthetic coal (SC) or residual fuel oil (RFO) and that carbon dioxide emissions were reported by CEM measurement. For the EIA-767 database the list of fuels was expanded to 11 fossil fuels with the critical criteria being that both heating values and heat based carbon dioxide emission factors were available. Table 1 shows the fuel code definitions and for which database each code was an accepted search criteria.

Table 1. Fuel codes used in EIA documentation and where in the database reduction each fuel was accepted search criteria.

Fuel Code	Description	eGRID Search Criteria	EIA-767 Search Criteria
BIT	Bituminous Coal	*	*
LIG	Lignite Coal		*
SC	Syncoal	*	*
SUB	Sub-bituminous Coal	*	*
DFO	Distillate Fuel Oil		*
JF	Jet Fuel		*
KER	Kerosene		*
PC	Petroleum Coke		*
RFO	Residual Fuel Oil	*	*
NG	Natural Gas		*
PG	Propane Gas		*

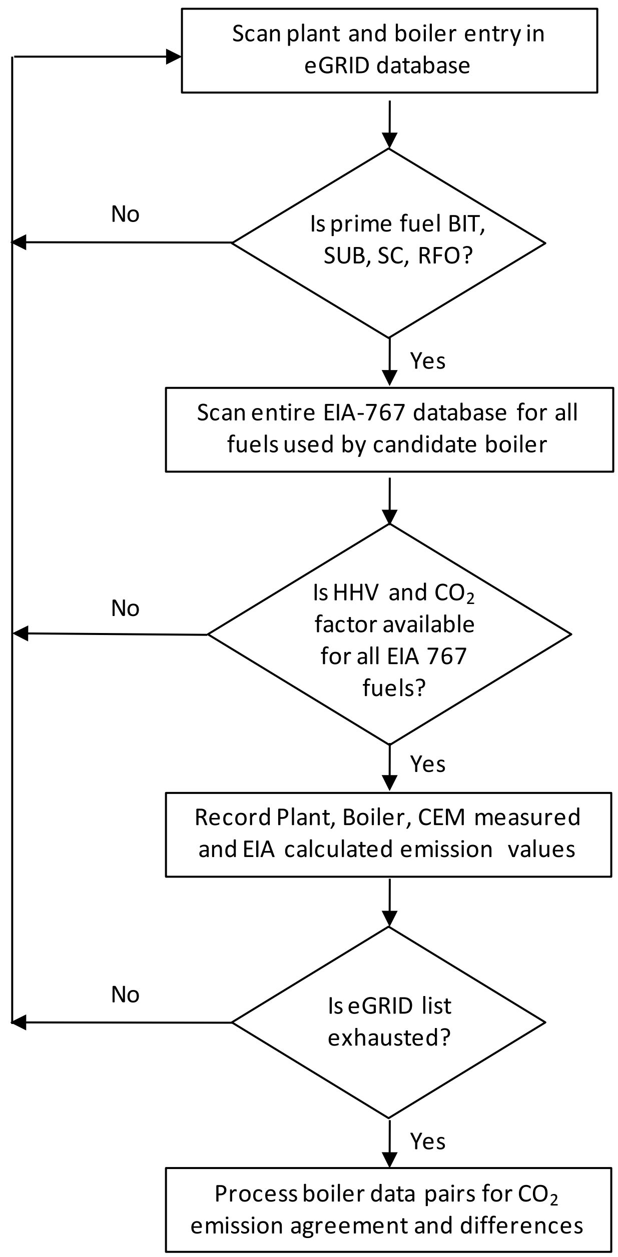

Figure 1. Flowchart for eGRID and EIA database comparison and parsing. Fuel codes are covered in the text while HHV is for Higher Heating Value.

Results and Discussion

The eGRID file lists 4866 boilers, of those 4866 there are 1664 that meet the eGRID database valid fuel search criteria of the primary fuel being BIT, SUB, SC or RFO. Continuing the database reduction, the second criteria from the eGRID database is the use of CEM for carbon dioxide measurements. Following discovery of a valid boiler in the eGRID database the EIA-767 database is searched for all occurrences of that specific boiler. Of the 1664 qualified boilers from the eGRID database, 1066 boilers listed in both the eGRID and EIA-767 database met all the criteria for comparison in this study: The final goal being that CEM reported carbon dioxide emissions could be compared to EIA calculated carbon dioxide emissions. The plot of this comparison, CEM reported carbon dioxide emissions versus EIA calculated carbon dioxide emissions, is shown in Figure 2.

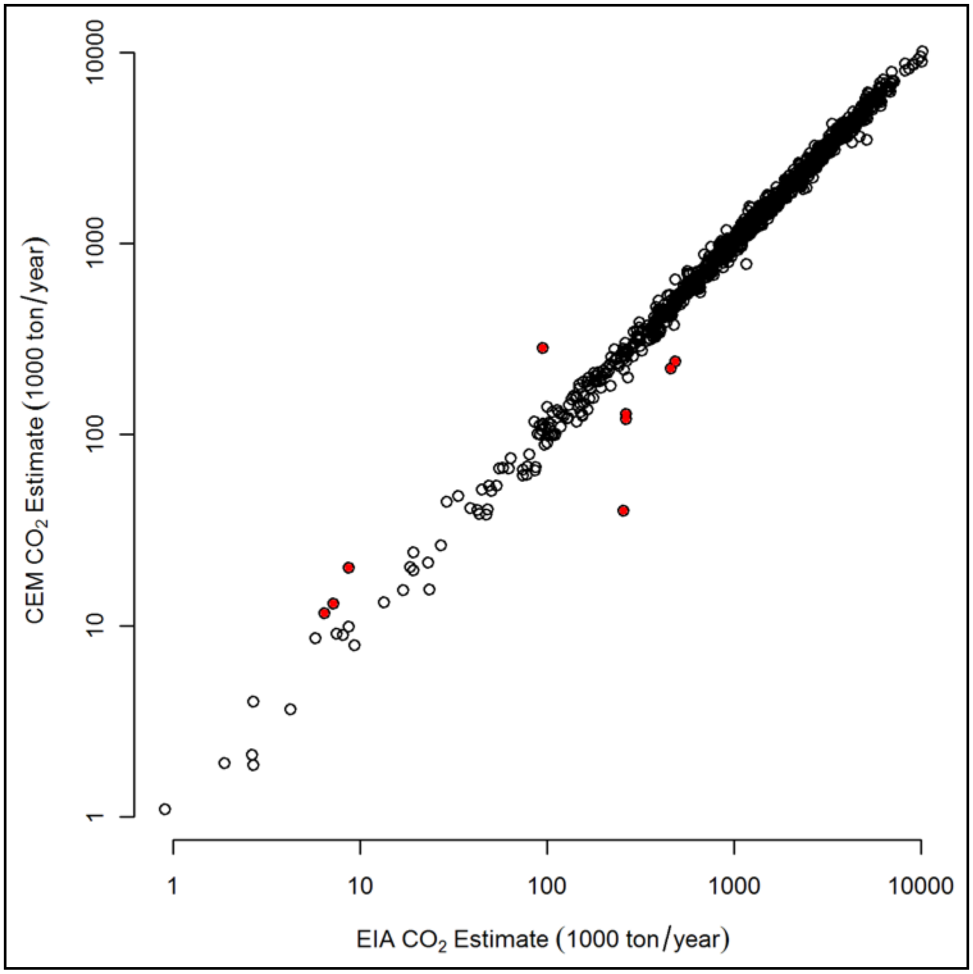

Figure 2. Log-log plot of CEM estimated carbon dioxide measurement versus EIA calculated measurements for 2005. Datapoints shown in red are statistical outliers.

The CEM measured carbon dioxide versus EIA calculated carbon dioxide show a very strong correlation. While the general agreement between the values obtained by the two methods is good, statistical comparison of the values in each dataset is more important.

In the statistical analysis, there is no predisposed notion that either the EIA calculated amount or CEM measured amount of carbon dioxide is more accurate. For the baseline comparisons, all percentage differences are reported in terms of the CEM measured values compared to the EIA calculated values. The selection of the EIA calculated values as the baseline for the comparison is arbitrary. Ultimately, the most important aspect of the statistical analysis will not be the average deviation which represents a systematic offset, but the spread of individual differences which is a measure of the uncertainty in the comparison between the eGRID CEM measured carbon dioxide output and an EIA carbon emission factor calculated carbon dioxide emission.

In an effort to avoid possible measurement or tabulation errors from degrading the statistical analysis, the data was first examined for statistical outliers using advanced outlier detection [10]. This eliminated nine datapoints, less than one percent of all datapoints, from the analysis.

Prior to outlier elimination, analysis of all 1066 boilers shows the sum of all carbon dioxide based on EIA emissions calculations accounts for 1.982×10^9 tons of emitted carbon dioxide. It is further noted that with the elimination of the nine outlying data points the total carbon dioxide emissions considered are reduced to 1.981×10^9 tons or a reduction of slightly more than 0.05 % from the original total carbon dioxide emissions.

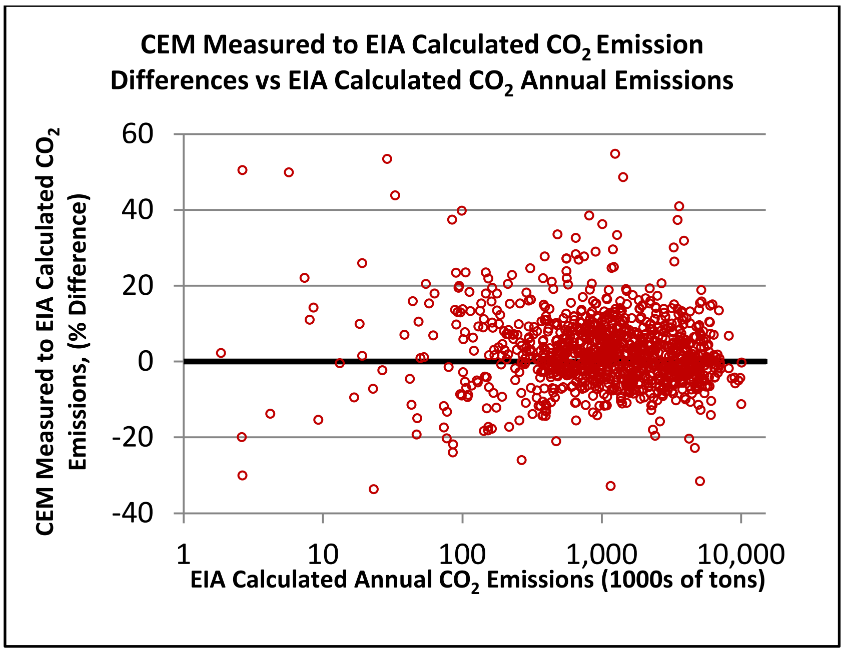

Figure 3. The percentage difference between CEM Measured and EIA Calculated CO_2 emissions plotted against the EIA Calculated annual emissions. Outlying data points have been removed.

Examining the differences between the CEM carbon dioxide measurements compared to EIA calculated values shows that the CEM measurements are on average 1.5 % larger than the EIA values. Considering the number of boilers and millions of tons of carbon dioxide involved, this is a surprisingly small offset and it provides an indication as to the potential effectiveness of the two different carbon dioxide evaluation techniques on a global scale. Related to the average CEM to EIA difference, figure 3 shows a plot of all datapoints less the outliers. While figure 3 cannot be used as a reliable means for determining the average difference, it does display a general trend for the carbon dioxide emission quantities of the steam-electric powerplants in this study and some trends for carbon dioxide measurement to calculation differences. Figure 3 may show a predisposition for the smaller emitters to produce larger relative differences in carbon dioxide measurements based on different evaluation techniques than the larger boilers, that specific investigation is outside the scope of this article. The systematic offset between the two measurements could likely be easily corrected. For this reason, the offset value is noted but will not be discussed in depth.

The more interesting property of these data is the distribution of the differences between the CEM measurements and EIA predictions. Figure 4 shows a histogram plot of the CEM measurement versus EIA calculated annual carbon dioxide emissions. Overlaid on the histogram is a Gaussian curve. Due to the heavy tails of the histogram, well in excess of the values expected by the Gaussian curve, it can be stated that the distribution of differences is markedly non-Gaussian; this is confirmed by a kurtosis of 2.8 (which should be 0 for a Gaussian distribution). The conclusion of the non-Gaussian nature of the histogram was also confirmed by the Anderson-Darling test of normality [11]. Despite the non-Gaussian nature of this distribution, it can be established that the individual CEM measurement to EIA calculation differences have a standard deviation[1] of about 8 %.

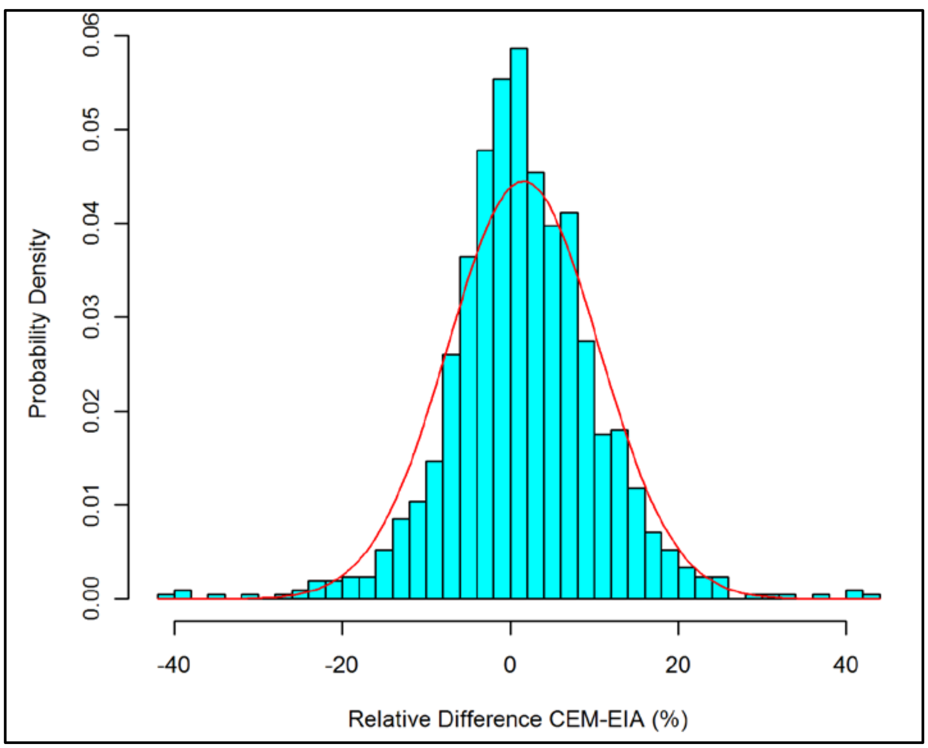

Figure 4. Histogram of percentage differences of CEM vs. EIA CO_2 emissions.

[1] The standard deviation is an indicator of the uncertainty of the differences. Due to the non-Gaussian nature of the distribution, the probabilities associated with standard deviations of Gaussian distributions do not apply.

Conclusions

Since quantitative measurement uncertainties are not available for either of the methods discussed here, one can only comment upon the characteristics of the population of differences between the two methods for the same boiler. For the approximately 1,000 U.S. boilers considered here, the population mean of 1.5 % is much smaller than its width. Consider that a minor revision of EIA published carbon emission coefficients or CEM system calibration could effectively relieve the systematic offset, represented by the average difference of 1.5 % to effectively zero. Identification of causes of the width of the distribution of differences between fuel calculation methodology and direct emissions measurements and its non-Gaussian nature require information not available from these databases. Clearly these differences are significant and one would observe that the approximate 20 % distribution half-width is near the 20 % relative accuracy limit set by EPA in the 1970's [12,13]. Although significant technological advances have occurred in the intervening time, they appear to have had little benefit in reducing differences between the two methods.

Future Studies

To resolve the sources of the differences between CEMs and fuel calculation method-based carbon dioxide emissions, direct assessment of both methodologies including analyses of the contributions made by their individual components to the total uncertainty is required. Neither database contains ancillary or metadata information sufficient to gain insight into potential causes of these differences; therefore, potential remedies are elusive. Such reduction will likely require improvements to the fundamentals of both methodologies. Future databases could strive to fully describe the complete system upon which boiler emissions determinations are made to more firmly place them upon a scientific basis. Should explicit economic value be placed on carbon emissions, one anticipates the need to improve the accuracy of emissions quantification methods for carbon dioxide and other greenhouse gases beyond current practice to support approaches based on purely regulatory, market based, or other approaches. In addition, the need for transparency in the description of emissions determination methodologies and their acceptance in the international arena are quite likely.

Areas of future research investigation include:

- Improved measurement capabilities for average velocity determination in the highly turbulent flows found in emission stacks with particular attention to flow velocity asymmetries and three dimensionality of these flow fields and their effects on measuring instruments, both those used in future CEM systems and those used for auditing purposes;
- Reduction in the variability of gas concentration determinations (one notes that current regulations allow ± 2 percent uncertainty for EPA protocol gases);
- Methodologies needed to improve the accuracy of fuel calculation methodologies, e.g., greater scientific rigor may be required to determine carbon content whether from direct measurement methods or reference data for fuel types. For example, these include improving determination of fuel mass arriving at the boiler combustion zone and fuel calorific value or carbon content.

References

[1] United States Department of State, *U.S. Climate Action Report 2010,*Washington: Global Publishing Services, June 2010

[2] Marland, G.; Rotty ,R.M. *Carbon Dioxide emissions from fossil-fuels: A procedure for estimation and results for 1950-1982*. Tellus, Ser B: 1984, 36, 232-261

[3] U.S. Energy Information Administration. Documentation for Emissions of Greenhouse Gasses in the United States 2005; EIA: Washington D.C., 2007

[4] United States Department of Energy, Energy Information Administration, electronic file *f76705.zip*, contains EIA-767 database

[5] United States Department of Energy, Energy Information Administration, *Form EIA-1605 (2007), Voluntary Reporting of Greenhouse Gases*, Form Approved OMB No. 1905-0194, Expires 07/31/2010

[6] United States Environmental Protection Agency, electronic file *eGRID2007_Version1-1.zip*, contains eGRID 2007 database

[7] United States Department of Energy, Energy Information Administration, *Form EIA-767 (2005), Steam-Electric Plant Operation and Design Report*, Form Approved OMB No. 1905-0129, Expires 11/30/2007

[8] United States Environmental Protection Agency, *The Emissions & Generation Resource Integrated Database for 2007, (eGRID2007) Technical Support Document*, E. H. Pechan, Springfield VA, Sept. 2008

[9] Ackerman, K.; Sundquist, E., *Comparison of Two U.S. Power-Plant Carbon Dioxide Emissions Data Sets*, Environ. Sci. Technol. 2008, 42, 5688-5693

[10] J. Fox. *Applied Regression, Linear Models, and Related Methods*. Sage Publications, Inc., Thousand Oaks, California, 1997.

[11] M. A. Stephens. Tests based on EDF statistics. In R. B. D'Agostino and M. A.Stephens, editors, *Goodness-of-fit-techniques*, volume 68 of *Statistics: a Series of Textbooks and Monographs*, pages 97–194. Marcel Dekker, New York, NY, 1986.

[12] 40CFR60

[13] 40CFR63