

EPA Wood Heater Test Method Variability Study

Analysis of Uncertainty, Repeatability and Reproducibility based on the
EPA Accredited Laboratory Proficiency Test Database

Rick Curkeet, PE
Chief Engineer – Hearth Products
Intertek Testing Services

Robert Ferguson
Ferguson, Andors & Company

October 6, 2010

I. Introduction

The precision of any test method is an extremely important tool in understanding the limitations of the data generated from the method. No rigorous assessment of the precision of the test methods contained in the woodstove NSPS and required to be used to certify model lines under that NSPS has ever been performed, despite the fact that the NSPS was promulgated over 20 years ago, more than 700 woodstoves have been tested and certified under it, and a large data base of proficiency test data from accredited laboratories has been assembled. The objective of this paper is to perform that assessment, using the EPA proficiency test data base and broadly accepted tools for assessing repeatability and reproducibility.

The paper consists of four major sections. Section I is this introduction. Section II provides back ground information that will help frame the issues associated with test method precision. Section III is the assessment of test method precision and is comprised of several sub-sections that address the EPA proficiency test data, and analyze some of the sources of variability. The last section (IV) provides conclusions about variability that are supported by the preceding in-depth analyses.

II. Regulatory Background.

The NSPS was proposed on February 18, 1987 (52 Fed. Reg. 4994) and promulgated on February 26, 1988 (53 Fed. Reg. 5860). It is codified at 40 CFR Part 60, Subpart AAA – *Standards of Performance for New Residential Wood Heaters*. The regulation includes the following EPA test methods:

EPA Method 28 – This method prescribes the fueling and operation procedures for emission testing. Among other parameters, it specifies the test fuel properties, test fuel load configuration, pre-test operating conditions, charcoal bed weight at the start of the test run, loading and start-up time, allowable air supply adjustments, fuel adjustments, end of test run determination and allowable heater body temperature differential from start to end of test, and other parameters needed to conduct a wood heater emission test.

EPA Method 5H - This particulate measurement procedure uses an EPA modified Method 5 sampling train which draws a flue gas sample from the wood heater stack and collects particulate in the sampling probe, on a heated 110 mm filter and in a series of ice water chilled impingers. The sampling rate for the sampling train is to be maintained at a constant proportion of the stack gas flow rate and includes a measurement protocol for determining the stack flow at equal time increments. The particulate mass collected in the probe, on the filters, and in the impingers is determined gravimetrically after the removal of uncombined water.

EPA Method 5G-1 – This method and all of the “5G” methods use a dilution tunnel which collects all the effluent from the appliance chimney and draws it, with ambient dilution air, through a straight duct at a uniform velocity. The velocity is high enough to allow accurate flow rate measurement using a Standard or S-Type pitot tube. Method 5G-1 uses a sample train similar to the Method 5 train but

uses two 110 mm filters in series that are unheated and does not involve recovery of material past the second filter. The particulate mass collected in the probe and on the filters is determined gravimetrically after the removal of uncombined water.

EPA Method 5G-2 – This method utilizes the Method 5H sampling train but the sample is taken from the dilution tunnel (as described in Method 5G-1 above) rather than the wood heater stack. Sample recovery is the same as Method 5H.

EPA Method 5G-3 – This method used two 47 mm filters in series and is otherwise similar to Method 5G-1. However the method requires that two identical sample trains be used simultaneously and that the results from each agree within 7.5% of the mean for the result to be valid.

At the time the NSPS was promulgated, there had been no rigorous assessment of the precision of the wood heater test methods. In the preamble accompanying the proposal, EPA had this to say about the inter-laboratory and intra-laboratory precision of Oregon Method 7, which became EPA Method 5H in the regulation:

[T]he apparent heater-to-heater differences in the data base reflect not only true differences in performance, but also reflect test method precision. Although data are limited, data obtained by Oregon DEQ suggest that the interlab ... (sic) four-run weighted average precision at the level of the standards is not greater than ± 1 g/hr.

In contrast, the database upon which the standards are based does not include individual wood heaters tested at more than one laboratory. Therefore, it was agreed that overall and inter-laboratory component of precision should be determined before enforcement tests are performed at laboratories other than the laboratory that initially certified the wood heater. Further, it was agreed that if the overall four-run weighted average precision exceeds ± 1 g/hr, then the interlab component of the precision would be added to the standard when . . . [enforcement] tests are conducted at other than the original certifying laboratory.

The EPA will, by July 1, 1990, either publish in the Federal Register a determination that the inter-laboratory precision cannot be determined, or promulgate revisions reflecting what that precision has been found to be. 52 Fed. Reg. 5010-5011 (February 18, 1987).

As noted above, however, no determination of either intra-lab or inter-lab precision has ever been made.

Section 60.535 of the regulation provides for accreditation for wood heater emission test laboratories. The conditions of accreditation require that both initial (see §60.535(b)(5)) and annual (see §60.535(b)(7)) proficiency testing be conducted by each accredited laboratory. The proficiency test requirements include conducting at least eight test runs (two in each of the burn rate categories described in EPA Method 28) on a wood heater identified (or actually provided) by EPA. The tests are conducted using all EPA wood heater emission measurement methods for which the particular test laboratory is seeking

initial accreditation, or seeking to maintain accreditation. For example, a laboratory that was accredited for Methods 5H and 5G-1 would run both methods during their eight proficiency test runs.

III. Test Method Variability

A. Basic Science

All measurement processes have an inherent element of variability in the measurement result if the process has sufficiently fine resolution. This variability can be evaluated empirically by performing multiple measurements of a single artifact and evaluating the mean and standard deviation of the test results. The relative magnitude of the standard deviation to the mean is a measure of the dispersion (variability) of the measurement process. This measure of dispersion includes variability that originates both from the measurement process and from changes in the artifact being measured, i.e., the inherent variability of the phenomenon being measured. When the measurement involves performance of a process which might be affected by uncontrolled variables or by the ranges allowed for operational parameters within the process, the variability associated with the process is usually much greater than that that arises from the uncertainty of the actual measurements made.

The “rifle” example is a common way of explaining these issues: If a rifle is repeatedly fired at a target, the impacts of variability can be visibly demonstrated. If the rifle is locked into a firing stand and carefully prepared cartridges (bullet plus powder load) are used and if the rounds are fired over a short period of time where the effects of variations in wind, temperature, humidity and barometric pressure conditions are minimized, the cluster of hits on the target should be very tight. The dispersion that is seen can be attributed to the minute differences in the rounds, variations in the rifle barrel as it heats and accumulates residue as well as the other unpreventable small shifts in ambient conditions. If the same experiment is repeated under a wider range of ambient conditions, say on two different days with different wind conditions, one might expect a wider dispersion of results. If the test is repeated with the rifle fired hand-held by the shooter, the dispersion will almost certainly be wider yet. If two different shooters fire the same rifle, the dispersion will widen further. If the shooters use two different rifles of the same model, again the results will almost certainly spread further. And so on and so on. In other words, as the opportunity for variation in the test conditions expands, so does the likelihood that the variation in the results will also increase.

Understanding and quantifying the variability in measurement systems is a scientific discipline. That discipline uses standardized terminology to categorize and organize the various sources of variability. That terminology is as follows:

Any test procedure that results in a numerical measurement of a product or material attribute is subject to a quantifiable value of “**measurement uncertainty**”¹. Each

¹ Two other concepts need brief mention: “accuracy,” and “bias”. “Accuracy” is the closeness of agreement between a test result and an accepted reference value. Ref.: ASTM E177. “Bias” is the total

measurement that is made as part of the test procedure, and ultimately combined mathematically to produce a result, contributes to the overall uncertainty of that result. The more measurements made, the greater the overall uncertainty.

The closeness of agreement between independent test results obtained under stipulated conditions is the “**precision**” of the measurement. Ref.: ASTM E177. Precision is usually broken down further into:

Repeatability – The closeness of multiple measurements of the same artifact under the same conditions with the same equipment and operator. The symbol “r” is often used to refer to repeatability. Repeatability can be evaluated by replicate tests of the same product at the same laboratory following the defined procedure.

Reproducibility – The closeness of measurements made on the same or presumed to be identical artifacts by different laboratories, equipment and operators. The symbol “R” is often used for reproducibility. Reproducibility can be evaluated for any quantitative measurement system through an Interlaboratory Study or round robin test program.

B. Quantifying Test Method Precision

1. The Available Data

The primary data set available to evaluate test result variability is that generated through the EPA required Accredited Laboratory Proficiency Test Program. These data supplied by EPA provides a substantial data set which allows assessment of the intra-lab repeatability (r), and inter-lab reproducibility (R) for emissions testing using the NSPS specified test methods.

The laboratories that participated in the proficiency testing are:

Apex Environmental	EEMC
EESPC	Engineering Sciences
Intertek-Middleton	Intertek-Montreal
Lokee Testing	Myren Consulting
Northwest Testing	OMNI-Oregon
OMNI-RTP	PFS
Shelton Research	Underwriters Laboratories

Note: For the purposes of this paper, EEMC, EESPC and Lokee have been grouped together as one entity based on the understanding the test equipment remained the same throughout the changes of ownership.

systematic error in a measurement, as contrasted to random error. Ref.: ASTM E177. Quantification of both accuracy and bias requires a reference standard or material with known properties and known uncertainty of those properties. Since no wood stove exists that can consistently produce a specific emissions rate, there is no way to evaluate the accuracy or bias of the test procedures. Accordingly, “accuracy” and “bias” will not be discussed further in this paper.

In this program each of these accredited laboratories was supplied with a test appliance which EPA modified in a manner designed to assure that it would be robust enough to operate the same way after being shipped from lab to lab.

Each laboratory was directed to conduct two full series of four test runs (two test runs in each of the four burn rate categories) in each round of proficiency testing. In addition, some laboratories that were accredited for more than one of the EPA methods were required to run both types of sampling equipment during their test runs and submit two sets of data. For example, a lab that was accredited for method 5H and 5G-1 would run both during their eight test runs. All Method 5G-1 and 5G-3 data were “adjusted to Method 5H equivalent²” using the formula $E_{\text{Hequiv.}} = 1.82 \times E^{0.83}$. Ref.: 40 CFR Part 60, Subpart AAA, Appendix A, Section 6.6, Eq. 5G-5.

In 1987 and 1988, EPA was able to send the first proficiency stove to each accredited laboratory in both years and obtain complete data sets. In 1989 a second proficiency stove was tested at each accredited laboratory, but several laboratories dropped their accreditation in the first two years of the program. From 1990 on, there was a further reduction in the number of laboratories participating and the schedule for proficiency testing was less rigorous. However, one test stove was used from 1993 - 2000. So there are three stoves with multiple test data sets from multiple laboratories. The data from these tests show no pattern of change over time that would indicate deterioration of the test appliance’s performance.

Table 1 is a compilation of weighted average emission rates using all eight runs (or more if the lab ran more than 2 in a burn rate category). The weighted average emission rates were computed from individual test run data that were provided to the primary author by EPA upon request in 2006. These data are included in their entirety in Appendix A. The primary author of this paper performed a limited quality check on the data, by comparing a sample of the data to the original test reports submitted by his laboratory and others, and determined that the check sample and the original submissions were identical. Table 1 utilizes all proficiency test program data provided by EPA, and is grouped for each stove tested as well as being categorized by test laboratory and test year.

² Although this conversion of data can be thought of as adding another layer of uncertainty to the results, especially since the adjustment equation is non-linear and it was itself based on a “best judgment” curve-fit to experimental data (with its own undetermined uncertainty), including the data adjustment is appropriate because it results in an analysis of the precision of the EPA methods as written and used.

Table 1

EPA Proficiency Test Data Summary - EPA Weighted Average Emissions - g/hr

Lab Code	Method	YEAR										
		1987	1988	1989	1990	1993	1995	1996	1997	1999	2000	2005
A	5G	2.71										
B	5G	5.96	3.03									
C	5H	2.21	1.51	14.46	6.51							
D	5G	5.72	2.77		7.32	6.24					5.89	15.60
E	5G	6.16	2.69	14.42	13.79		9.40			6.39		10.32
E1	5H	6.09		13.66								
F	5G		4.12	12.53	6.69							
F1	5H			12.40								
G	5G	2.95										
H	5G	19.10										
H1	5H	12.45	6.22	15.39		4.90			4.12	2.88		
I	5G		2.42									
J	5G			17.57		4.86						
K	5G						6.43				13.82	
L	5G							6.06	4.81	4.90		
Stove		Catalytic 1		Non-Cat 1	Non-Cat 2	Non-Cat 3						Non-Cat 4
EPA Cert.		3.1		7.5	4.5	3.6						3.1

2. Data Analysis Options

Since there is no published EPA method for determining test method precision, two separate data analysis methodologies were employed.

- “Macro” Analysis Using Standard Statistical Tools

In Table 2, we present the results of a “macro” analysis of the data using standard statistical tools. Since there is a wide range in the weighted average emissions performance for each stove in the database, both the standard deviation and the coefficient of variation (CV)³ were computed for each stove to avoid any potentially misleading assessments of variability that might result from the wide range in the calculated means for the individual test stoves.

³ The coefficient of variation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other. The coefficient of variation is useful because the standard deviation of data must always be understood in the context of the mean of the data. The coefficient of variation is a dimensionless number. So when comparing between data sets with different units or widely different means, one should also determine the coefficient of variation for comparison instead of the standard deviation alone. In this case there is a wide range in the average emission performance of the various test stoves and assessing the variability of the results using just the standard deviation without also assessing the CV could be misleading.

Table 2.

EPA Proficiency Test Data Summary - EPA Weighted Average Emissions - g/hr

Lab Code	Method	YEAR										
		1987	1988	1989	1990	1993	1995	1996	1997	1999	2000	2005
A	5G	2.71										
B	5G	5.96	3.03									
C	5H	2.21	1.51	14.46	6.51							
D	5G	5.72	2.77		7.32	6.24					5.89	15.60
E	5G	6.16	2.69	14.42	13.79		9.40			6.39		10.32
E1	5H	6.09		13.66								
F	5G		4.12	12.53	6.69							
F1	5H			12.40								
G	5G	2.95										
H	5G	19.10*										
H1	5H	12.45*	6.22	15.39		4.90			4.12	2.88		
I	5G		2.42									
J	5G			17.57		4.86						
K	5G						6.43				13.82	
L	5G							6.06	4.81	4.90		
Stove		Catalytic 1		Non-Cat 1	Non-Cat 2	Non-Cat 3					Non-Cat 4	
EPA Cert.		3.1		7.5	4.5	3.6					3.1	
Mean		3.90		14.35	8.58	6.21					12.96	
SD		1.74		1.78	3.49	2.75					3.73	
CV		44.7%		12.4%	40.7%	44.4%					28.8%	
N		14 (Lab H/H1 1987 Excluded)		7	4	13					2	

* Data excluded as outliers based on Grubb's test for outliers.

From the Table 2 data it is apparent that the repeatability and reproducibility of the test procedures are not very good.

Without applying any statistical tools, one can simply look at the results of testing the same stove in the same laboratory over two or more years to see the range of intra-laboratory results.

And looking at the results from all the laboratories for all the stoves over all the years of the program, it is possible to assess the inter-laboratory precision. Using 2.8 times the standard deviation to estimate the potential range of results that could be expected with a 95% confidence level [Ref.: ASTM E177, 28.1], it is clear that the for any given test series the reproducibility is on the order of $\pm 4.9 - 9.8$ grams per hour.

This means that for any emissions rate measured using the EPA test methods, the result could be 4.9 to 9.8 grams per hour higher or lower if the appliance were tested again at a different laboratory. Even at one standard deviation (68% confidence level), the reproducibility interval is about 1.7 to 3.5 grams per hour, i.e., there is a 32% chance that a new test result would deviate by more than this amount.

The CV calculations further confirm that variability is high (>40 % CV) for the three test stoves with significant data sets but it is not consistent for all units. The 2005 data can be discounted since two data points are too few to draw any valid conclusions. Since CV is calculated from one standard deviation, the implied variability in the actual data is really about 2.8 times the CV at a 95% confidence level. For example, a CV of 40% implies that the results of any one test could vary by +/- 112% from a population mean with a probability of 5% or less of being farther away.

- Precision Analysis Using ASTM E691

Consistent with the National Technology Transfer and Advancement Act, which creates a presumption in favor of using consensus standards, we determined that ASTM E691 – *Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method* was appropriate for use here.

ASTM E691 provides procedures for planning and conducting studies of the precision of test methods, and also provides statistical techniques for performing precision analyses. Since the EPA proficiency test program has been on-going for more than 20 years, the test program design components {Sections 6 – 14) of ASTM E691 are not germane here. However, the analytical procedures can be applied to the EPA proficiency test data for the purposes of determining precision (both *r* & *R*). In the following subsections, these procedures are first described generally and then specifically as they were applied to the EPA proficiency test data.

- ASTM E691 Analytical Procedure and Terminology

The analysis utilizes tabular, graphical, and statistical diagnostic tools for evaluating the consistency of the data so that unusual values may be detected and investigated, and also includes the calculation of the numerical measures of precision of the test method pertaining to both within-laboratory repeatability and between-laboratory reproducibility (*r* & *R*).

Table of Test Results—The test results received from the laboratories are arranged in rows and columns. Each column contains the data obtained from all laboratories for one test sample, and each row contains the data from one laboratory for all test samples. The test results, *x*, from one laboratory on one test sample constitute a cell. The results of the following calculations for that test sample are also entered on the table.

Cell Statistics:

*Cell Average*⁴, \bar{x} —This statistic is the cell average for each laboratory.

$$\bar{x} = \sum_I^n x / n \quad \text{where } n = \text{the number of test results per cell}$$

Average of the Cell Averages, $\bar{\bar{x}}$ —This statistic is the average of all the cell averages for the one material and all laboratories.

$$\bar{\bar{x}} = \sum_I^p \bar{x} / p \quad \text{where } p = \text{the number of laboratories}$$

Cell Standard Deviation, *s* —This statistic is the standard deviation of the test results in each cell and is calculated as the square root of the sum of the squares of the cell

⁴ ASTM E691 refers to the term “average” when defining statistical terminology. “Average” is equivalent to the “arithmetic mean”. The term “mean”, when used in this paper should be considered interchangeable with “average” as defined in ASTM E691.

value minus the cell average divided by one less than the number of test results in the cell.

$$s = \sqrt{\sum_I^n (x - \bar{x})^2 / (n - 1)}$$

Cell Deviation, d — The cell deviation is calculated by subtracting the average of the cell averages for all laboratories from the cell average for each laboratory.

$$d = \bar{x} - \bar{\bar{x}}$$

Standard Deviation of the Cell Averages, $s_{\bar{x}}$ —This statistic is calculated as the square root of the sum of the squares of the cell deviation divided by one less than the number of laboratories.

$$s_{\bar{x}} = \sqrt{\sum_I^n d^2 / (p - 1)}$$

Precision Statistics

While there are other precision statistics, the fundamental precision statistics of the Interlaboratory Study are the repeatability standard deviation and the reproducibility standard deviation. Other statistics are calculated from these standard deviations.

Repeatability Standard Deviation, s_r —This statistic is calculated as the square root of the sum of the squares of the cell standard deviation divided by the number of laboratories.

$$s_r = \sqrt{\sum_I^p s^2 / p}$$

Reproducibility Standard Deviation, s_R — This statistic is calculated as the square root of the square of the standard deviation of the cell average plus the square of the repeatability standard deviation (first adjusted for the number of test results) or is equal to s_r , if s_r is larger than s_R .

$$s_R = \sqrt{(s_{\bar{x}})^2 + (s_r)^2 (n - 1) / n}$$

Consistency Statistics

The statistical analysis of the data for estimates of the precision statistics is simply a one-way analysis of variance (within- and between-laboratories) carried out separately for each material. Since such an analysis can be invalidated by the presence of severe outliers, it is necessary to first examine the consistency of the data.

Between-laboratory Consistency Statistic, h — This statistic is calculated as the cell deviation divided by the standard deviation of the cell averages.

$$h = d / s_{\bar{x}}$$

Within-laboratory Consistency Statistic, k — This statistic is calculated as the cell standard deviation from one laboratory divided by the repeatability standard deviation of the material.

$$k = s / s_r$$

Critical Values of the Consistency Statistics—A table⁵ list critical values of the h and k consistency statistics at the 0.5 % significance level. The critical values for h depend on the number of laboratories, p , and the critical values for k depend both on the number of laboratories, p , and on the number of replicate test results, n , per laboratory per material. When cell values approach or exceed the critical values for h and k , those cells or laboratories should be investigated for data problems.

- ASTM E691 Precision Analysis

The proficiency test data was then analyzed to determine test method precision using the procedures outlined in ASTM E691 as described above.⁶

Separate tables (Tables 3a-3c) were created for each of the different stoves that were tested over the years the proficiency test program was conducted where there is sufficient data to apply the procedures specified in ASTM E691⁷. Each participating test lab has a row in each table. Since the proficiency test program required two emission tests in each Method 28 burn rate category, it was possible to calculate two EPA weighted average emission results for each proficiency test series. To do this, the first runs conducted in each of the four burn rate categories were grouped together for the purposes of determining the first weighted average emissions. Likewise, the second run in each burn rate category produces the second weighted average emissions result. This methodology was employed since the burn rate category tests were generally not conducted in any specific order and grouping the first runs in each category most closely simulates an actual emission certification test. The numbered columns in each table represent the weighted average emissions results for the grouped data for the different test series on the particular test stove. For example, in Table 3a, Lab A only participated in one year of testing on Catalytic Stove 1. By grouping the eight individual test runs as described above, two weighted average emissions values result for that lab for that stove and are shown in the columns labeled 1 and 2. Lab E participated in both test years and ran two test methods simultaneously in both years. The results from the second method are shown in the row labeled E1. The data grouping methodology results in four separate EPA weighted average emission rates for

⁵ ASTM E691 Table 5

⁶ In some cases, inadequate data precluded using all data in the analyses. In 2005, only two test laboratories conducted proficiency testing and Non-Cat 4 was not tested in any other year. ASTM E691 requires a minimum of three laboratories for a single sample so the 2005 data was not used.

⁷ Ideally the data set used for this analysis should contain a minimum of 3 replicate tests from each of at least 6 laboratories. In Tables 3a-3c, only data from the proficiency test stoves that have sufficient data for a statistically meaningful evaluation are presented.

this lab for Catalytic Stove 1 for the 5G method as shown in the columns labeled 1 through 4 and rows labeled E and four additional results for the 5H method in row E1.

The next column shows \bar{x} , the mean of all test series results on the tested stove for each lab. In this case, this includes the two test series for each proficiency test and multiple proficiency tests results on the same test stove whenever available. The columns to the right of \bar{x} include the statistical information about the results from each lab for all test series for each sample tested per ASTM E691, as described above. This statistical information includes s , the standard deviation for individual lab results, d , the cell deviation (how much the cell mean deviates from the overall mean), h , a “between laboratory consistency statistic” and k , a “within laboratory consistency statistic”. These latter two are measures of whether data included in the analysis should be investigated for problems. Three weighted average emissions data points were excluded from the analysis based on this test. They are marked with * in the tables that follow. Also included in the table are $\bar{\bar{x}}$, the overall mean for all test series from all labs, $S_{\bar{x}}$, the standard deviation of the overall mean, S_r , the repeatability standard deviation as well as S_R , the reproducibility standard deviation. Finally, values for repeatability (r) and reproducibility (R) are shown to provide the true assessment of the precision demonstrated by the EPA proficiency test data.

Table 3a

1987-1988 Catalytic Stove 1 - ASTM E691 Analysis												
Lab	x				\bar{x}	s	d	h	k	Critical Values		
	1	2	3	4						$h-Crit.$	$k-Crit.$	
A	2.77	2.67			2.72	0.07	-1.19	-0.95	0.06	2.15	2.36	
B	5	6.69	3.44	3.28	4.6	1.59	0.7	0.56	1.27	2.15	1.92	
C	2.28	1.78	1.48	1.78	1.83	0.33	-2.08	-1.66	0.26	2.15	1.92	
D	5.6	6.09	2.83	2.8	4.33	1.76	0.42	0.34	1.4	2.15	1.92	
E	6.53	5.78	4.69	3.4	5.1	1.36	1.19	0.95	1.08	2.15	1.92	
E1	7.39	5.02	2.91	2.55	4.47	2.23	0.56	0.45	1.78	2.15	1.92	
F	2.81	3.04			2.93	0.16	-0.98	-0.78	0.13	2.15	2.36	
G	22.44*	11.19*	5.19	5.37	5.28	0.13	1.37	1.1	0.1	2.15	2.36	
					$\bar{\bar{x}}$	3.91						
					s_x	1.25	95% Confidence (\pm)					
					s_r	1.26	Repeatability (r)	3.52 g/hr				
					s_R	1.62	Reproducibility (R)	4.53 g/hr				

* Data excluded as outliers.

Table 3b

1989 Non-Catalytic Stove 1 - ASTM E691 Analysis											
Lab	x				\bar{x}	s	d	h	k	Critical Values	
	1	2	3	4						h-Crit.	k-Crit.
A	13.55	11.65			12.6	1.34	-1.41	-0.1	1.29	2.05	2.30
A1	13.34	10.9			12.12	1.73	-1.89	-0.13	1.65	2.05	2.30
B	13.68	14.03			13.86	0.25	-0.15	-0.01	0.24	2.05	2.30
B1	12.84	13.35			13.1	0.36	-0.91	-0.07	0.35	2.05	2.30
C	14.47	14.31			14.39	0.11	0.38	0.03	0.11	2.05	2.30
D	17.08	16.99			17.04	0.06	3.03	0.22	0.06	2.05	2.30
E	13.81	16.11			14.96	1.63	0.95	0.07	1.56	2.05	2.30
					$\bar{\bar{x}}$	14.01					
					s_x	1.66	95% Confidence (\pm)				
					s_r	1.04	Repeatability (r)		2.92 g/hr		
					s_R	1.82	Reproducibility (R)		5.1 g/hr		

Table 3c

1993-2000 Non-Catalytic Stove 3 - ASTM E691 Analysis													
Lab	x						\bar{x}	s	d	h	k	Critical Values	
	1	2	3	4	5	6						h-Crit.	k-Crit.
A	10.22	5.93	7.15	5.54	18.52*	7.61	7.29	1.85	-1.41	-0.1	1.29	1.92	1.75
B	7.25	13.06	6.8	8.24			8.84	2.88	-1.89	-0.13	1.65	1.92	1.84
C	5.23	5.2					5.22	0.02	-0.15	-0.01	0.24	1.92	2.22
D	5.61	7.84	3.41	9.25	4.16	5.27	5.92	2.22	-0.91	-0.07	0.35	1.92	1.68
E	7.02	4.15	6.39	3.5	4.08	1.5	4.44	2.01	0.38	0.03	0.11	1.92	1.68
F	7.26	5.56					6.41	1.2	3.03	0.22	0.06	1.92	2.22
						$\bar{\bar{x}}$	6.35						
						s_x	1.56	95% Confidence (\pm)					
						s_r	1.92	Repeatability (r)		5.38 g/hr			
						s_R	2.28	Reproducibility (R)		6.39 g/hr			

*Data excluded as outlier.

Abbreviations:

s = standard deviation for individual lab results

d = cell deviation = $\bar{x} - \bar{\bar{x}}$

s_x = standard deviation of \bar{x}

s_r = repeatability of standard deviation

s_R = reproducibility of standard deviation

h = between laboratory consistency statistic

k = within laboratory consistency statistic

- Conclusions from ASTM E691 Analyses

The Table 3a, 3b and 3c data shows that the repeatability (within lab), r , of the weighted average emissions rate determined by the EPA test methods at the 95% confidence level is at best ± 2.9 grams per hour and typically about ± 3.5 to ± 5.4 grams per hour. The reproducibility (between labs), R , at the 95% confidence level is ± 4.5 to ± 6.4 grams per hour.

3. Sources of Variability in EPA Wood Heater Testing

Given that 12 labs (five are still accredited) and all four particulate measurement methods are represented, it is important to try to understand if the repeatability and reproducibility issues are laboratory and method related or arise from a different source – that is, variable performance of the appliance itself.

To evaluate that question, we next investigated the potential sources of variability in the test methods. We started with the emissions measurement methods (5G-1, 5G-2, 5G-3 and 5H), and then addressed Method 28, which specifies how the appliance is to be operated during emission testing.

- Emission Measurement Methods

Method 5G-1

Potential sources of variability in Method 5G-1 include the following:

- Accuracy of dilution tunnel gas flow and sample flow measurement and the resulting consistency of proportionality.
- Recovery of particulates from the probe and front half of filter holder using a solvent cleaner.
- Weighing errors for tare and final weights of filters and evaporation containers used to collect and weigh the catch.

These sources combine to produce an estimated measurement uncertainty of about ± 2.7 to 3% of the emission value measured for a typical passing wood stove test. See Appendix C for the derivation of this estimated uncertainty range.

Method 5G-2

Potential sources of variability in Method 5G-2 include the following:

- Accuracy of dilution tunnel gas flow and sample flow measurement and the resulting consistency of proportionality.
- Recovery of particulates from the probe and sampling line using a solvent cleaner.
- Recovery of particulate collected in the impingers and connecting glassware.
- Weighing errors for tare and final weights of filters and evaporation containers used to collect and weigh the catch.

These sources combine to produce an estimated measurement uncertainty of about $\pm 3 - 3.5\%$ of the emission value measured for a typical passing wood stove test. See Appendix C for the derivation of this estimated uncertainty range.

Method 5G-3

Potential sources of variability in Method 5G-3 include the following:

- Accuracy of dilution tunnel gas flow and sample flow measurement and the resulting consistency of proportionality.
- Weighing errors for weights of filters, probes including front filter housings and filter seals.

The measurement uncertainty of method 5G-3 has been determined to be approximately $\pm 2.5\%$ of the emission value measured for a typical passing wood stove test. See Appendix C for the derivation of this estimated uncertainty range.

Method 5H

Potential sources of variability in Method 5H include the following:

- Accuracy of flue gas flow and sample flow measurement and the resulting consistency of proportionality.
- Recovery of particulates from the probe and sampling line using a solvent cleaner.
- Recovery of particulate collected in the impingers and connecting glassware.
- Weighing errors for tare and final weights of filters and evaporation containers used to collect and weigh the catch.

These sources combine to produce an estimated measurement uncertainty of about ± 20 to 30% of the total particulate mass determination. See Appendix C for the derivation of this estimated uncertainty range.

- Conclusions Regarding the Contribution of Emissions Measurement Method Uncertainty to Overall Method Variability

It is apparent from the foregoing analysis that the emission measurement methods explain only a small part of the overall variability that has been discussed and quantified in Section III. It is clear, however, from this analysis that some emission measurement methods contribute less to variability than others with Method 5G-3 the best performer in this regard.⁸

⁸ With the objective of improving the wood heater test methods based on more than twenty years of experience with the current EPA methods, ASTM E06.54 sub-committee was charged with creating new methods. The first step was to specify only one particulate measurement procedure to help reduce the measurement uncertainty associated with multiple measurement methods. This resulted in the subcommittee's development of ASTM E2515, which specifies a procedure that corresponds very closely to EPA Method 5G-3 but has provisions that allow appliance types other than just wood and pellet heaters to be tested. This procedure was selected because it has the smallest measurement uncertainty of the four current EPA methods and contains the dual train measurement which allows for ready detection of invalid results.

- Operating Protocol for the Appliance (EPA Method 28)

Potential Sources of Variability in Method 28 include the following:

- Fuel density variation from approximately 30 to 40 lb/ft³ (dry weight basis).
- Fuel moisture content variation from 19 to 25% dry basis (varies in uniformity as well as average).
- Fuel load configuration details.
- Coal bed size (20-25% of fuel load weight) and pre-burn temperature conditions.
- Loading time and start-up procedure.
- Ambient temperature, barometric pressure and humidity.
- Variations in control settings and resulting burn rates.
- Random uncontrollable variables such as when and how the fuel load settles, falls and collapses.

Using the EPA proficiency test data, it is possible to examine several of the various Method 28 parameters for their impact on precision.

The first is an analysis of the impacts of burn rate variations. It is known that emissions performance is related to burn rate to some degree with higher burn rates generally producing more complete combustion.

The proficiency test data does not lend itself well to determining burn rate variability in that each lab must adjust air controls to a setting that will produce a burn rate in one of the required categories. The exception is the high burn rate (Category 4) which is always run with the air controls fully open. Table 4 is a compilation of the Category 4 burn rates from the proficiency test program. Note that the specific labs designated by the lab codes were different from year to year.

Table 4

EPA Proficiency Test Data - High Burn Rates - Dry kg/hr

Stove	Year	Lab							Mean	SD	CV
		1	2	3	4	5	6	7			
Catalytic Stove 1	1987	2.16	2.66	1.74	2.13	2.06	1.69	2.99			
		2.24	2.90	1.89	2.36	2.06	1.74	4.27*			
	1988	1.96	2.08	1.41	2.13	2.27	2.21	1.57			
		2.11	2.13	2.09	2.27	2.92	2.23	2.04	2.15	0.38	17.8%
Non-Cat 1	1989	1.77	2.06	2.50	1.95	1.85					
		1.83	2.32	2.60	2.02	1.91			2.08	0.29	14.0%
Non-Cat 2	1990	4.39	3.05	2.26							
		4.84	3.14	3.49					3.53	0.94	26.7%
Non-Cat 3	1993	3.20	2.91	4.76							
		3.23	3.75	5.18							
	1995	4.08	1.38								
		4.73	1.40								
	1996	3.10									
		3.70									
	1997	3.17	3.24								
		3.29	3.27								
	1999	4.18	3.80	3.73							
		4.49	4.57	4.67							
2000	4.71	2.90									
	7.09	3.97						3.79	1.15	30.4%	
Non-Cat 4	2005	2.54	2.82								
		2.57	3.20						2.78	0.31	11.0%

* Data not included in statistical analysis.

The high burn rate variability is substantial, and it is certain that this variability is a contributing factor in the emissions variability, but it is not likely that it is the most important factor. There are many instances in the proficiency test data where two runs in one lab at very similar burn rates produced substantially different emissions rates. See Appendix B for burn rate versus emissions plots of all the individual run data.

Charts 1 and 2 show the relationship between emissions and fuel moisture content and emissions and load weight (assumed to relate to fuel density) for the 121 runs conducted on the catalytic proficiency test stove in 1987 and 1988. Using the “Corner Score” statistical test where a score of >11 indicates potential dependency between variables [Ref.: Mark’s Standard Handbook for Mechanical Engineers, 17-22], it is clear that there is no relationship.

Chart 1

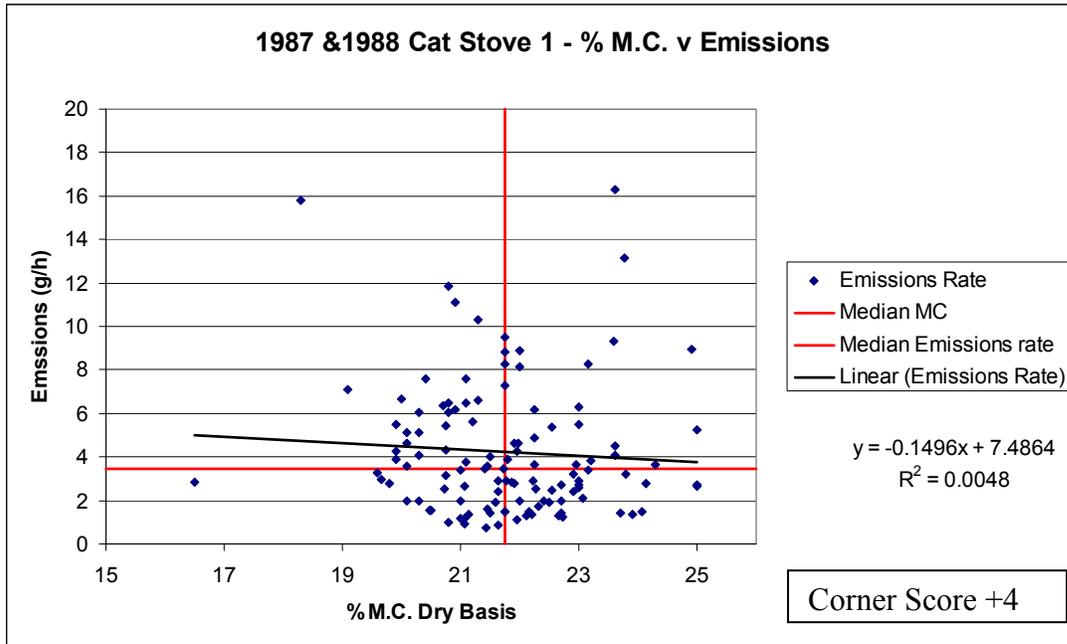
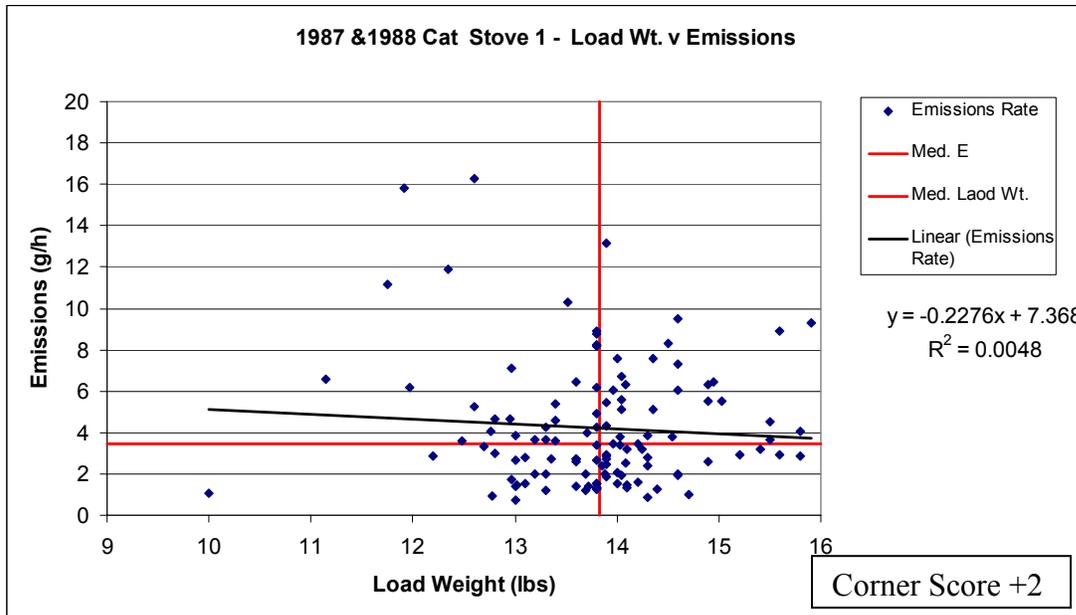


Chart 2



By practical necessity, EPA Method 28 contains tolerances on operational and fueling parameters. These include fuel density, moisture content, laboratory ambient temperature, coal bed size, fuel load weight and several others. While it might be supposed that the method variability could be lessened by even tighter specifications within the test method relating to fuel and operating conditions, the data does not support this proposition.

Without these tolerances or specified ranges wood heater emission testing would become prohibitively expensive with significant additional costs incurred while attempting to control these variables within a significantly narrower range than specified. As demonstrated, these additional costs would not significantly reduce variability.

IV. Final Conclusions

The repeatability and reproducibility of wood heater emission testing as demonstrated by the EPA accredited laboratory proficiency test data is quite poor. At the 95% confidence level, repeatability for the EPA weighted average emission rate is at best ± 2.9 g/hour and ranged as high as ± 5.4 g/hour. The reproducibility was no better than ± 4.5 g/hour and ranged as high as ± 6.4 g/hour.

Further analyses of the potential sources of variability have shown that the emission measurement test methods (EPA Methods 5G-1, 5G-2, 5G-3 and 5H) are not major contributors to the high overall variability being demonstrated by the data. However, Method 5G-3 does demonstrate the lowest uncertainty of the four methods currently specified in the NSPS.

Analyses of the variability inherent in EPA Method 28 also indicate that the operational and fuel parameter tolerance ranges specified in that method are not major contributors to the high variability. Tightening these parameters to improve test precision would simply increase costs and the data show that such tightening would not significantly improve precision.

This leaves the only logical conclusion. Variability in wood heater emission testing results for any given appliance is most likely a function of the random nature of burning wood, no matter how tightly you try to control the process. Many relatively small, uncontrollable variables that are inherent in the wood combustion process can combine to significantly affect the outcome of any given test.

These conclusions should not be taken to mean that the current test methodology is inadequate to characterize an appliance's emissions performance. There is no question that the currently certified low emissions wood heaters produce far less particulate emissions than their pre-regulation predecessors which produced emissions rates of 50 to 100 grams per hour or more. Our primary conclusion is that the current testing process simply cannot consistently distinguish emissions performance differences of less than 3 to 6 grams per hour. The process is certainly capable of reliably distinguishing between good and bad performance, but it cannot reliably distinguish between "good, better and best" performance.⁹

⁹ To return to the "rifle" analogy, wood stoves are not modern high powered rifles in the hands of an expert marksman who can place a dozen shots in a 2 inch bull's-eye at 100 yards. They are rather more like a smooth bore flintlock pistol where just "hitting the paper" at 50 feet is an accomplishment.

Appendix A. – Raw Data

Note: The laboratory codes indicated in this appendix are not those used in the tables in the body of the paper since EPA assigned the lab codes to different labs from year to year. Labs have not been identified by name to maintain the anonymity of the participants.

Year		1987			
Stove		Catalytic -1			
Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M WET	Emissions (g/h)
5G	1	1.15	13.90	18.40	2.48
	2	2.16	13.30	18.00	4.25
	3	1.15	13.00	17.40	2.69
	4	1.35	13.30	18.20	3.67
	5	2.24	13.80	16.60	4.25
	6	0.79	13.30	17.40	1.18
	7	1.74	13.00	16.60	3.88
	8	0.73	10.00	18.00	1.09
Lab B	Run			DRY	
5G	0	1.10	12.69	19.60	3.30
	1	2.90	12.76	18.90	28.11
	2	2.16	11.92	18.30	15.79
	3	0.56	12.19	16.50	2.83
	4	1.10	12.96	19.10	7.10
	5	0.99	12.76	20.30	4.07
	6	2.66	11.75	20.90	11.13
	7	2.03	11.97	20.90	6.18
	8	2.10	12.34	20.80	11.87
9	1.69	11.15	21.30	6.60	
Lab C	Run			DRY	
5H	1	1.42	14.10	22.90	3.20
	2	1.89	13.20	24.29	3.63
	3	1.74	14.30	24.13	2.79
	4	1.71	14.00	23.07	2.08
	5	1.12	14.60	21.60	1.91
	6	1.21	13.80	22.20	1.35
	7	0.87	13.80	22.67	1.27
	8	0.97	14.30	21.63	0.89
	9	0.61	15.40	20.75	3.16

Lab D	Run			DRY	
5G	1	2.13	15.03	19.90	5.50
	2	2.36	14.05	20.00	6.68
	3	1.36	13.97	20.30	6.07
	4	0.83	14.05	21.20	5.60
	5	0.57	14.03	21.00	3.40
	6	1.43	14.01	21.08	7.60
	7	1.10	14.05	20.10	5.14
	8	1.06	14.08	20.70	6.34

Lab E	Run			WET	
5G	1	2.06	14.90	18.70	6.30
	2	2.06	13.80	18.03	8.15
	3	1.56	14.60	17.86	7.30
	4	0.67	13.80	17.86	1.50
	5	1.34	13.80	18.20	6.15
	6	0.59	13.70	17.36	2.00
	7	0.91	13.80	17.86	8.25
	8	0.83	13.90	17.18	5.45

Lab E1	Run			WET	
5H	1	2.06	14.90	18.70	5.50
	2	2.06	13.80	18.03	8.90
	3	1.56	14.60	17.86	9.50
	4	0.67	13.80	17.86	1.50
	5	1.34	13.80	18.20	4.90
	6	0.59	13.70	17.36	1.20
	7	0.91	13.80	17.86	8.80
	8	0.83	13.90	17.18	4.35

Lab F	Run			DRY	
5G	1	1.69	13.52	21.30	10.30
	2	1.63	14.03	23.20	3.80
	3	0.99	13.36	23.00	2.70
	4	0.93	13.86	22.90	2.40
	5	0.93	13.88	22.70	2.00
	6	0.49	13.72	23.70	1.40
	7	1.74	14.25	23.80	3.20
	8	1.62	13.71	21.50	4.00

Lab G	Run			WET	
5G	1	4.27	14.50	18.80	8.29
	2	1.02	14.10	19.20	35.98
	3	0.72	14.30	19.00	19.49
	4	0.82	13.90	19.10	17.01
	5	0.77	14.10	19.30	1.36
	6	1.53	12.60	19.10	16.29
	7	1.46	13.90	19.20	13.16
	8	2.99	13.90	18.70	2.91

Lab H	Run				
5H	1	1.68	n/a	n/a	5.537
	2	1.12			5.713
	3	1.72			6.62
	4	1.01			7.531
	5	2.22			8.08
	6	2.24			19.018
	7	0.7			7.213
	8	0.44			2.386

Year 1988
Stove Catalytic-1

Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M WET	Emissions (g/h)
5H	1	1.51	15.20	17.88	2.91
	2	1.96	15.50	18.67	3.66
	3	0.60	15.60	18.19	2.89
	4	0.65	15.90	19.09	9.32
	5	1.45	15.50	19.11	4.49
	6	1.20	15.60	19.94	8.94
	7	2.11	15.80	19.11	4.08
	8	1.22	15.80	17.94	2.85

Lab B	Run			WET	
5G	1	1.47	14.90	18.70	2.62
	2	1.03	13.60	20.00	2.72
	3	1.92	13.80	20.00	2.63
	4	2.08	12.60	20.00	5.25
	5	0.64	14.10	19.40	1.47

6	0.61	13.10	17.00	1.54
7	1.03	13.20	18.30	1.98
8	2.13	13.40	18.40	5.38

Lab C	Run			DRY	
5G	1	1.11	13.90	21.63	2.88
	2	0.87	13.60	20.73	2.56
	3	1.41	13.80	23.17	3.38
	4	0.75	13.80	21.13	1.34
	5	2.12	12.80	20.08	4.66
	6	0.59	13.00	22.71	1.40
	7	2.09	12.80	19.67	2.96
	8	1.32	13.10	21.92	2.80

Lab D	Run			DRY	
5H	1	0.45	13.00	21.43	0.74
	2	2.13	14.30	21.64	2.39
	3	2.27	14.20	21.42	3.45
	4	1.82	13.90	22.50	1.89
	5	1.38	13.60	21.50	1.39
	6	0.76	14.40	22.73	1.23
	7	0.90	13.80	22.12	1.29
	8	1.18	14.20	21.46	1.61

Lab E	Run			DRY	
5G	1	1.82	13.97	21.73	3.48
	2	2.92	12.96	22.32	1.71
	3	0.65	13.02	22.17	1.47
	4	1.83	12.95	21.97	4.64
	5	1.18	12.48	21.46	3.57
	6	1.14	14.09	22.28	2.55
	7	0.47	12.77	21.06	0.93
	8	2.27	12.62	22.10	5.06

Lab F	Run			DRY	
5G	1	2.23	14.35	20.30	5.13
	2	1.85	14.95	20.80	6.47
	3	0.96	14.60	20.10	1.98
	4	1.47	14.55	21.10	3.77
	5	2.21	14.35	20.40	7.58

6	0.49	14.70	20.80	0.99
7	0.76	14.00	20.50	1.52
8	1.03	14.05	20.30	1.96

Lab G	Run			DRY	
5G	1	1.43	14.60	20.80	6.03
	2	2.04	13.60	21.10	6.47
	3	1.03	14.30	21.80	3.86
	4	0.75	13.90	22.70	2.73
	5	0.71	13.30	22.00	1.98
	6	0.93	13.60	19.80	2.75
	7	1.57	13.40	21.90	4.60
	8	1.17	13.40	20.10	3.61

Year 1989
Stove Non-Catalytic-1

Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M DRY	Emissions (g/h)
5G	1	1.53	11.50	21.98	15.80
	2	0.89	9.80	22.38	10.95
	3	1.18	10.00	20.33	9.70
	4	0.58	10.40	20.56	20.10
	5	1.83	10.00	20.83	6.60
	6	0.91	9.80	20.56	11.35
	7	0.62	9.70	20.44	18.00
	8	1.77	10.40	20.46	9.85

Lab A1	Run			DRY	
5H	1	1.53	11.50	21.98	16.10
	2	0.89	9.80	22.38	9.80
	3	1.18	10.00	20.33	9.35
	4	0.58	10.40	20.56	21.00
	5	1.83	10.00	20.83	9.85
	6	0.91	9.80	20.56	10.25
	7	0.62	9.70	20.44	17.25
	8	1.77	10.40	20.46	9.10

Lab B	Run			DRY	
5G	1	0.80	10.90	21.63	22.18
	2	1.04	11.20	20.85	13.82

3	0.48	11.20	20.19	18.07
4	0.98	11.40	21.98	14.19
5	1.29	10.70	19.93	11.71
6	1.58	11.40	23.48	13.69
7	2.32	11.50	22.47	7.55
8	2.06	11.60	23.05	7.46

Lab B1	Run			DRY	
5G	1	0.80	10.90	21.63	22.05
	2	1.04	11.20	20.85	12.04
	3	0.48	11.20	20.19	18.39
	4	0.98	11.40	21.98	13.60
	5	1.29	10.70	19.93	10.93
	6	1.58	11.40	23.48	12.24
	7	2.32	11.50	22.47	6.83
	8	2.06	11.60	23.05	6.80

Lab C	Run			DRY	
5H	1	0.51	10.90	22.71	21.08
	2	0.79	10.90	22.15	14.41
	3	1.03	11.50	21.83	15.55
	4	0.92	11.50	21.66	16.28
	5	1.33	11.20	25.13	12.69
	6	1.47	11.20	21.88	15.05
	7	2.50	11.10	21.13	8.28
	8	2.60	11.60	21.73	8.54

Lab D	Run			DRY	
5G	1	1.53	10.90	21.88	17.75
	2	1.39	11.10	23.82	11.20
	3	1.95	11.40	23.34	14.02
	4	0.64	11.30	21.28	23.31
	5	1.03	11.30	21.74	19.69
	6	0.79	11.40	21.16	18.38
	7	2.02	10.50	21.75	4.93
	8	1.24	10.40	23.17	20.14

Lab E	Run			DRY	
5H	1	INVALID			
	2	1.7112	11.5	18.223	5.413

3	0.8368	11.5	18.534	15.492
4	1.3117	11.5	18.351	19.298
5	1.8494	11.6	18.066	4.558
6	1.1465	11.3	18.066	20.368
7	1.1703	10.5	18.172	11.303
8	0.8284	11.5	18.027	22.939
9	1.9069	9.7	17.012	8.424

Year 1990
Stove Non-Catalytic-2

Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M DRY	Emissions (g/h)
5H	1	1.70	17.69	21.49	3.07
	2	1.54	16.84	20.25	2.71
	3	1.02	17.36	21.00	5.68
	4	0.63	18.35	22.48	24.09
	5	4.39	19.03	23.51	8.92
	6	1.10	17.50	21.21	4.91
	7	4.84	18.58	22.82	11.23
	8	0.78	18.01	21.97	2.97

Lab B	Run			DRY	
5G	1	1.91	10.40	22.34	17.40
	2	0.64	9.50	22.33	8.80
	3	1.29	9.50	22.81	7.82
	4	1.09	10.00	22.33	5.04
	5	3.14	9.90	22.60	12.47
	6	0.99	9.60	23.48	7.82
	7	3.05	9.50	22.77	11.98
	8	0.74	10.70	21.60	38.04

Lab C	Run			DRY	
5G	1	2.18	10.40	22.63	13.36
	2	0.67	9.90	20.37	5.14
	3	3.49	10.80	21.95	14.89
	4	1.24	10.30	22.00	2.64
	5	1.24	10.40	21.23	2.63
	6	2.26	10.60	21.68	6.27
	7	1.28	9.80	22.23	5.66
	8	0.82	9.50	21.23	5.52

Lab D	Run			DRY	
5G3	1	4.375	9.36	21.33	10.26
	2	0.708	10.41	20.59	6.28
	3	0.639	10.93	19.89	7.24
	4	1.675	9.64	20.42	6.90
	5	1.665	10.45	20.42	9.58
	6	1.478	10.65	21.25	6.90
	7	1.387	10.34	21.65	7.50
	8	4.147	10.69	23.23	9.73

Year 1993
Stove Non-Catalytic-3

Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M WET	Emissions (g/h)
5H	1	0.83	10.10	18.06	10.98
	2	3.23	10.10	18.06	13.02
	3	1.22	10.40	18.41	3.34
	4	1.41	10.50	18.39	1.62
	5	1.74	10.60	18.01	2.32
	6	3.20	10.00	17.69	8.20
	7	1.14	10.20	17.93	2.90
	8	0.88	10.20	17.56	6.30

Lab B	Run			DRY	
5G	1	3.75	14.40	23.65	9.31
	2	0.67	10.20	23.20	7.13
	3	2.91	10.30	22.57	5.12
	4	0.81	11.60	22.73	7.79
	5	1.20	10.90	22.48	3.72
	6	1.20	9.90	23.20	5.16
	7	1.36	12.30	21.87	1.54
	8	1.29	12.50	22.23	2.01

Lab C	Run			WET	
5G	1	0.94	9.56	18.82	7.02
	2	0.95	9.64	18.91	7.20
	3	1.07	10.17	19.10	5.79
	4	1.18	9.89	19.61	5.28
	5	1.39	10.04	18.81	5.35
	6	1.47	10.17	17.98	1.66

7	5.18	9.99	18.17	28.47
8	4.76	10.38	19.13	11.89

Year 1995
Stove Non-Catalytic-3

Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M WET	Emissions (g/h)
5G	1	1.24	10.10	16.76	3.70
	2	1.07	10.20	18.82	12.00
	3	1.70	10.20	18.52	2.85
	4	4.08	11.10	19.03	11.26
	5	1.20	11.40	19.76	19.90
	6	4.73	11.60	17.57	14.44
	7	1.51	11.50	16.06	8.29
	8	0.95	10.50	18.69	11.00

Lab B	Run			DRY	
5G	1	0.70	10.29	21.90	14.46
	2	0.74	9.66	19.80	10.64
	3	0.87	10.83	19.90	7.49
	4	1.12	9.55	18.50	6.17
	5	1.38	9.80	19.10	4.26
	6	1.40	10.70	20.10	1.26
	7				
	8				

Year 1996
Stove Non-Catalytic-3

Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M WET	Emissions (g/h)
5G	1	1.24	9.70	17.79	4.12
	2	0.94	9.30	18.04	9.45
	3	3.10	9.80	18.56	4.02
	4	1.59	10.30	17.96	2.84
	5	0.93	9.90	17.56	10.19
	6	3.70	9.80	16.81	11.27
	7	1.36	10.20	16.84	3.22
	8	1.13	10.60	17.79	8.07

Year 1997

Stove Non-Catalytic-3

Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M WET	Emissions (g/h)
5G	1	0.94	10.00	17.48	3.59
	2	3.29	10.70	17.46	3.01
	3	1.42	10.10	17.65	4.96
	4	0.98	10.70	17.76	6.62
	5	1.13	10.60	17.84	4.70
	6	1.16	10.80	17.51	3.58
	7	3.17	10.30	14.51	6.66
	8	1.29	11.00	17.76	3.72

Lab B	Run			DRY	
5H	1	0.85	10.70	16.96	12.56
	2	0.79	10.20	17.62	3.78
	3	3.27	10.40	19.02	8.85
	4	1.22	10.60	17.75	0.82
	5	3.24	10.80	17.32	5.21
	6	1.32	10.50	16.96	1.18
	7	1.46	11.20	18.45	2.45
	8	1.06	11.10	17.75	3.54

Year 1999

Stove Non-Catalytic-3

Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M WET	Emissions (g/h)
5H	1	1.24	10.80	17.84	2.50
	2	4.18	11.30	19.39	5.53
	3	1.44	10.00	17.89	3.09
	4	4.49	10.20	19.10	4.78
	5	1.20	10.00	18.23	0.87
	6	1.61	10.20	19.02	0.96
	7	0.91	9.80	16.41	5.24
	8	0.98	10.00	17.27	1.28

Lab B	Run			DRY	
5G	1	1.16	10.60	23.87	6.06

2	3.80	10.40	24.25	9.26
3	1.62	10.30	23.71	4.11
4	1.10	10.50	23.14	6.74
5	0.99	10.20	21.31	9.16
6	4.57	10.20	21.55	15.99
7	1.32	10.60	21.68	3.34
8	0.97	9.70	23.41	8.72

Lab C	Run	WET			
5G	1	0.966	10.50	17.18	2.42
	2	3.836	9.30	17.44	5.27
	3	1.150	9.50	17.69	3.76
	4	4.672	9.60	17.75	17.97
	5	0.617	9.70	17.70	4.35
	6	0.814	9.60	17.41	12.08
	7	1.262	9.40	17.63	2.45
	8	1.268	9.70	17.37	3.32

Year 2000
Stove Non-Catalytic-3

Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M DRY	Emissions (g/h)
5G	1	1.26	10.08	22.51	6.10
	2	0.91	10.16	20.97	5.62
	3	0.79	10.04	21.65	7.42
	4	1.17	10.26	20.36	5.34
	5	1.11	10.19	23.05	4.72
	6	1.43	10.19	22.47	1.74
	7	4.71	10.16	22.37	11.40
	8	7.09	10.31	23.68	11.06

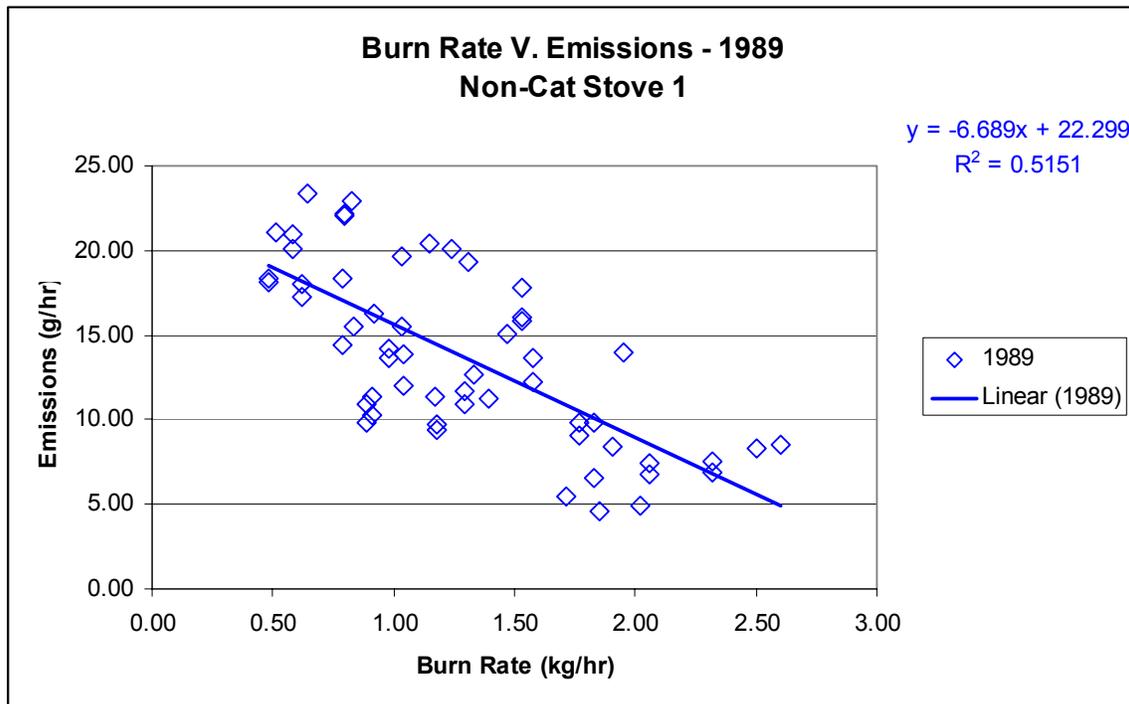
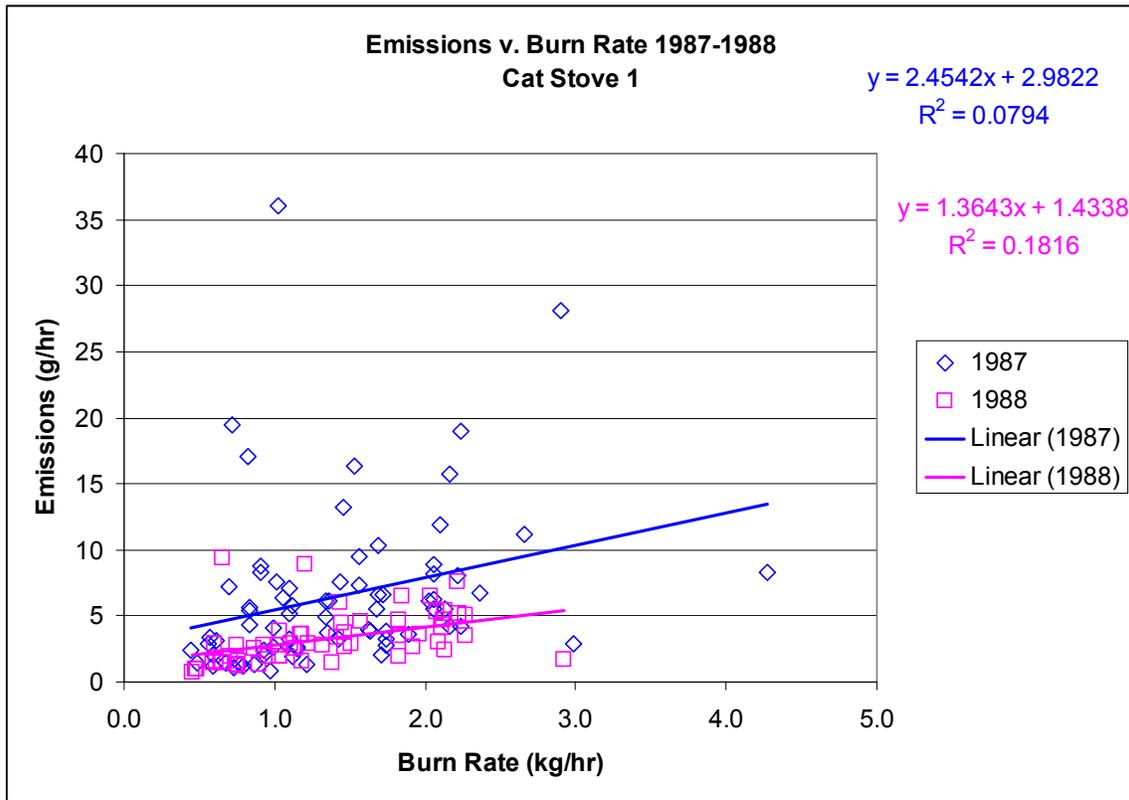
Lab B	Run	DRY			
5G	1	1.287	10.88	20.20	4.64
	2	0.699	10.33	19.70	59.60
	3	0.983	11.05	20.80	7.09
	4	3.967	10.76	21.20	19.82
	5	0.955	9.78	20.80	7.54
	6	1.428	9.62	21.30	4.74
	7	0.803	9.85	21.80	7.75
	8	2.901	10.67	22.00	14.42

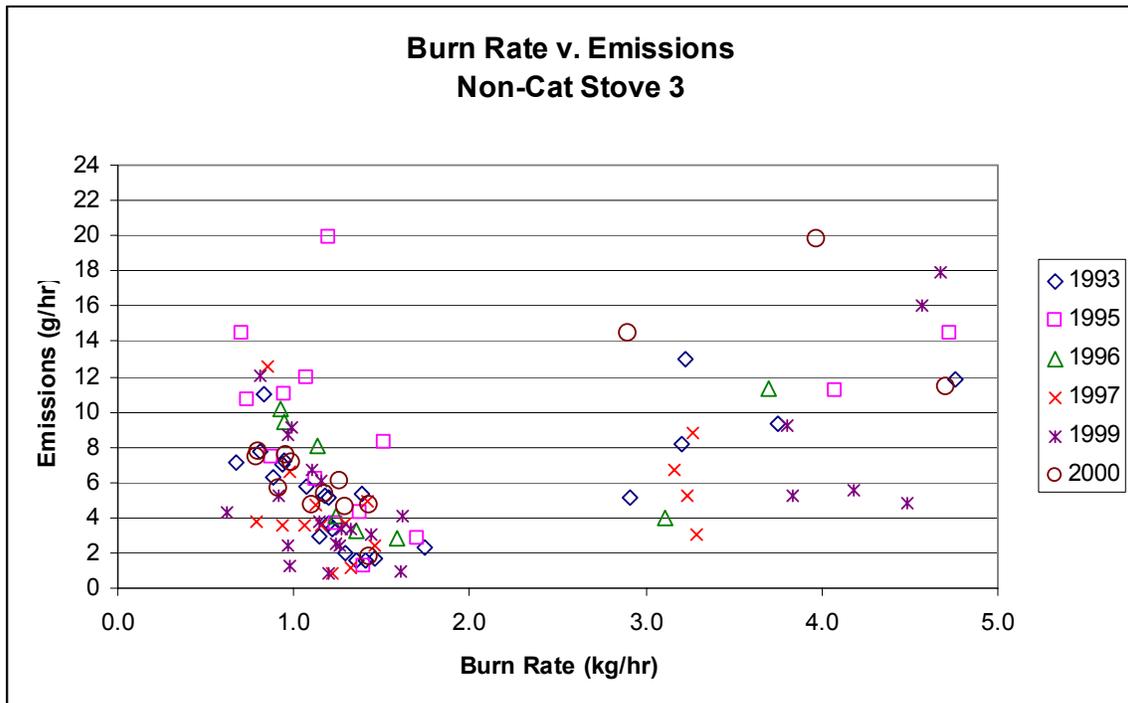
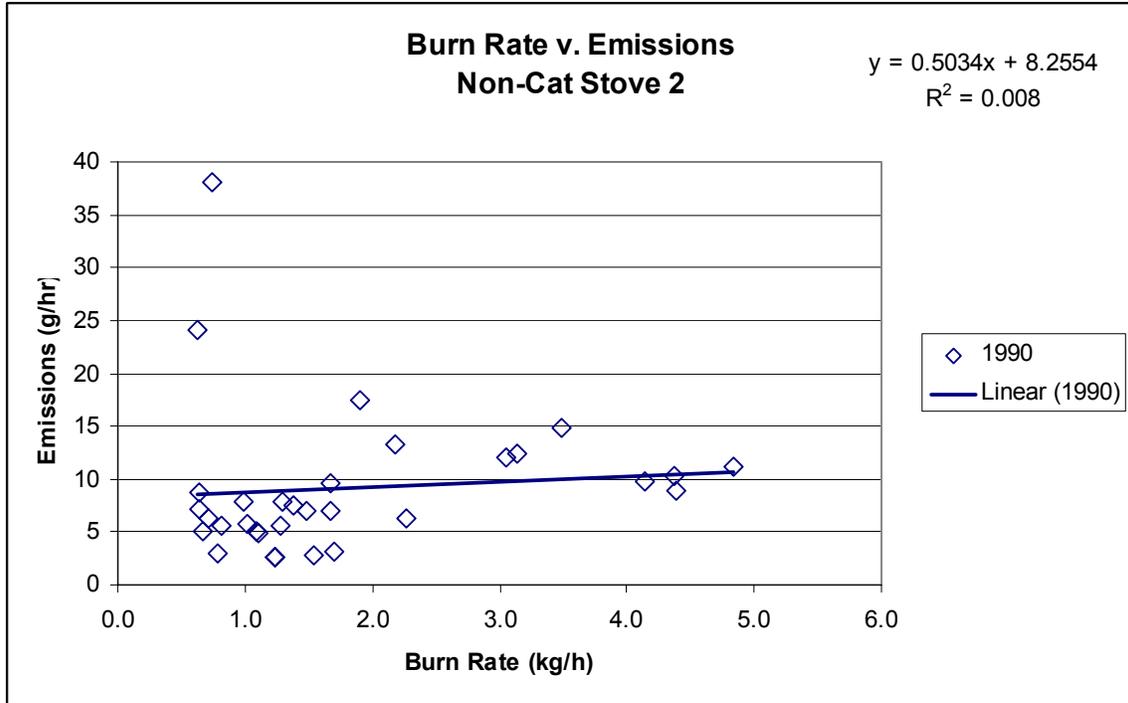
Year 2005
Stove Non-Catalytic-4

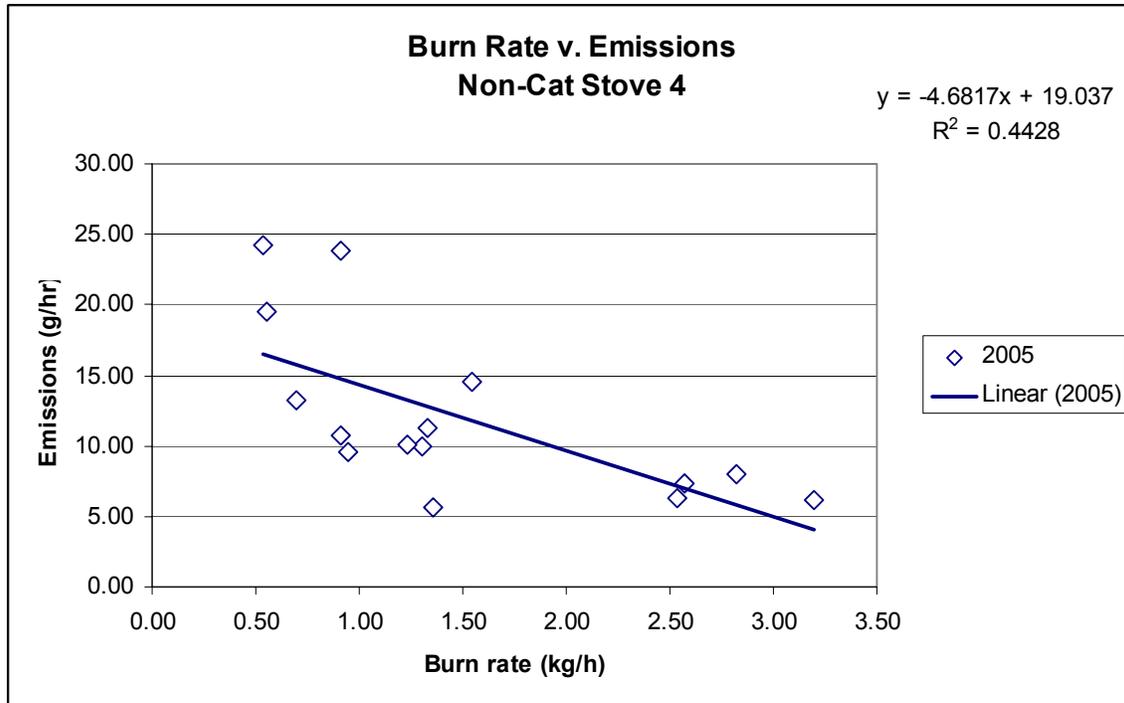
Lab A	Run	Burn Rate (Kg/h)	Test Load Wt. (lb)	% M DRY	Emissions (g/h)
5G	1	2.54	9.10	22.00	6.30
	2	1.33	8.20	19.50	11.30
	3	0.54	9.00	19.20	24.30
	4	0.95	9.20	19.40	9.50
	5	0.91	8.40	20.10	10.80
	6	0.70	8.40	21.50	13.20
	7	1.36	8.00	23.20	5.60
	8	2.57	9.10	20.30	7.40

Lab B	Run			DRY	
5G	1	INVALID	8.89	21.79	
	2	0.91	8.34	20.85	23.82
	3	1.55	8.23	20.76	14.59
	4	3.20	8.13	21.51	6.10
	5	1.31	8.05	19.74	9.94
	6	0.55	8.05	20.44	19.55
	7	1.23	8.22	20.87	10.04
	8	2.83	8.23	20.14	7.93

Appendix B – Burn Rate versus Emissions Rate Charts







Appendix C - Measurement Uncertainty Analyses for EPA Wood Heater Test Methods

Every numerical physical measurement process is subject to a quantifiable level of uncertainty. This uncertainty is determined in the process of calibrating the measurement instrument. Therefore, the measurement uncertainty for each direct measurement required in a test method can be determined or estimated. When multiple measurements of specific quantities are combined into a final numerical measurement result, the combined uncertainty can be calculated by application of well defined and accepted procedures. These procedures are outlined fully in ISO “Guide to the Expression of Uncertainty in Measurement.”

The ISO Guide defines two types of evaluation of measurement uncertainty. Type A is defined as “method of evaluation of uncertainty by statistical analysis of series of observations.” This is a quantitative evaluation. Type B is defined as a “method of evaluation of uncertainty by means other than the statistical analysis of series of observations.” This describes a more qualitative evaluation. The ISO Guide further elaborates that a “Type B evaluation of standard uncertainty is usually based on scientific judgment using all the relevant information available, which may include:

- previous measurement data,
- experience with, or general knowledge of, the behavior and property of relevant materials and instruments,
- manufacturer’s specifications,
- data provided in calibration and other reports, and
- uncertainties assigned to reference data taken from handbooks.”

Since there is no such thing as a wood heater test method calibration standard – that is, a wood heater with a “known” emissions rate – the uncertainty calculations and estimates that follow below are based on a hypothetical emissions value and evaluate only the particulate sampling method uncertainty relative to the hypothetical value. Variability in measured emissions due to the combination of the fueling and operation method and general wood burning variability are not components of measurement uncertainty. Measurement uncertainty quantifies only the potential difference between the actual measured result and the “true” value for one specific test run. Variation in results between different test runs includes both the effects of measurement uncertainty and many other variables that influence the actual performance of the product.

An ISO Guide Type A measurement uncertainty analysis is included in this appendix for Method 5G-3 as an example of the way measurement uncertainty is determined using the ISO Guide procedures. Detailed measurement uncertainty calculations are not included for the other methods. However, estimates of the measurement uncertainties for the other 5G methods along with the underlying assumptions used to achieve those estimates are provided. In the case of Method 5H, an estimate of overall method measurement uncertainty¹⁰ is provided. The estimate is based on both ISO Guide Type A and Type B evaluations.

¹⁰ This estimate includes both measurement uncertainties that can be analyzed using ISO Guide Type A evaluations and other uncertainties that implicate Type B analyses. The use in this method of one approach for determining stack gas flow rate for setting the proportional sampling rate and another procedure for

EPA Methods 5G-3, 5G-1 and 5G-2

Analysis of the measurement uncertainty for the total particulate emissions in these dilution tunnel test methods requires combining the uncertainty of the following individual measurements.

- Dilution tunnel volumetric flow rate.
- Sampling system volumetric flow rate.
- Filter and sample probe particulate gravimetric catch.

Component Measurement Uncertainties:

- Dilution Tunnel Flow Rate—The dilution tunnel flow is measured by a standard Pitot tube and a differential pressure gauge. The direct measurement is of flow velocity at the operating temperature and pressure. Pitot tube measurements are considered a primary reference method and therefore the primary component of uncertainty in the velocity measurement is that of the differential pressure gauge. Conversion of the velocity measurement to flow rate at standard temperature and pressure conditions requires additional measurements of the tunnel cross sectional area, the temperature and the absolute pressure. An analysis of the overall uncertainty of the tunnel flow rate indicates that the measurements of tunnel cross-sectional area, temperature and pressure have a relatively minor affect. The procedure in this standard has been specified such that the overall uncertainty of this parameter is about ± 2 % of the flow rate.
- Sampling Volumetric Flow Rate—The equipment available to measure the sample flow rate has a high precision and can be calibrated such that the measurement uncertainty is ± 1 % or better of the actual flow rate. For the purposes of this appendix, a ± 1 % uncertainty will be assumed.
- Filter and Probe Particulate Catch—The analytical balance specified in the method has a resolution of ± 0.0001 g (0.1 mg). For the purposes of measurement uncertainty analysis, we will conservatively assume that the balance resolution equals the measurement uncertainty.

Detailed Measurement Uncertainty (MU) Analysis for Method 5G-3

Weighing Uncertainty

For Method 5G-3, a minimum total of four weighings is needed to determine the net particulate catch (filters and o-rings weighed together and probe weighed separately, before and after the test run). Since each weighing is subject to the same uncertainty, the overall uncertainty of the weighing is:

$$MU_{weighing} = \sqrt{(MU_{Balance})^2 \cdot X}$$

determining stack gas flow rate for the calculation of total emissions is the principle source of measurement error that is assessed using the Type B Analysis. We note, however, that a rigorous analysis of the broader uncertainty implications of this and other Method 5H procedures is outside the scope of this paper.

Where:

$MU_{Balance}$ = Measurement uncertainty of the analytical balance (0.1 mg), and
 X = The total number of weight values (4) actually used to calculate emissions. Note: This does not include weight measurements made during pretest or post-test drying, only the final pretest and post-test weights.

$$MU_{weighing} = \sqrt{0.1^2 \cdot 4} = 0.20$$

For a process that involves two pretest and two post-test weighings, this uncertainty is ± 0.20 mg.

An additional component of the filter catch weight uncertainty is the potential for incomplete recovery of the filter material. There is no objective data to determine the magnitude of this factor, but laboratories familiar with the procedures report that they believe it to be less than 0.1 mg. Combining an additional ± 0.1 mg with the weighing uncertainty results in an overall weighing uncertainty of ± 0.22 mg (0.00022 g).

Combined Measurement Uncertainty

Combined uncertainties are calculated by taking the square root of the sum of squares of the component uncertainties multiplied by a “sensitivity coefficient”. The sensitivity coefficient is the partial derivative of the function used to calculate the result with respect to the specific measurement parameter. The general formula (law of propagation of uncertainty) is:

$$uY = \sqrt{((\delta Y / \delta x_1) \times u_1)^2 + \dots + ((\delta Y / \delta x_n) \times u_n)^2}$$

Where:

$\delta Y / \delta x_i$ = Partial derivative of the combining formula with respect to individual measurement x_i , and
 u_i = the uncertainty associated with that measurement.

The formula to calculate total particulate emissions is:

$$E_T = c_s Q_{std} \theta$$

Where:

c_s = sample filter catch/(sample flow rate x test duration), g/dscf,
 Q_{std} = average dilution tunnel flow rate, dscf/min, and
 θ = sampling time, minutes.

For the sake of example the following values will be used in an MU analysis of E_T .

Measurement	Measured Value	MU	Units
Sample Filter Catch (F_c)	0.0210	± 0.00022	g
Sample Flow Rate (Q_{sample})	0.250	± 0.0025	dscfm
Sampling Duration (θ)	180	± 0.1	minutes
Tunnel Flow Rate (Q_{std})	150	± 3	dscfm

(1) Calculate the MU of c_s :

$$c_s = F_c / (Q_{sample} \theta) = 0.0210 / (0.25 \times 180) = 0.000467 \quad \text{g/dscf}$$

$$\frac{\delta c_s}{\delta F_c} = \frac{1}{Q_{sample} \cdot \theta} = \frac{1}{0.25 \cdot 180} = 0.0222$$

$$\frac{\delta c_s}{\delta Q_{sample}} = \frac{-F_c}{Q_{sample}^2 \cdot \theta} = \frac{-0.021}{0.25^2 \cdot 180} = -0.001867$$

$$\frac{\delta c_s}{\delta \theta} = \frac{-F_c}{Q_{sample} \cdot \theta^2} = \frac{-0.0210}{0.25 \cdot 180^2} = -0.000003$$

$$MUC_s = \sqrt{(0.00022 \cdot 0.0222)^2 + (0.0025 \cdot -0.001867)^2 + (0.1 \cdot -0.000003)^2}$$

$$MUC_s = 0.000007 \text{ g/dscf}$$

Thus, c_s would be 0.467 mg/dscf \pm 0.007 mg/dscf at a 95 % confidence level.

(2) Calculate E_T and MU_{ET}

$$E_T = c_s Q_{std} \theta = .000467 \times 150 \times 180 = 12.60 \text{ g}$$

$$\frac{\delta E_T}{\delta c_s} = Q_{std} \cdot \theta = 150 \cdot 180 = 27,000$$

$$\frac{\delta E_T}{\delta Q_{std}} = c_s \cdot \theta = 0.000467 \cdot 180 = 0.084060$$

$$\frac{\delta E_T}{\delta \theta} = c_s \cdot Q_{std} = 0.000467 \cdot 150 = 0.070050$$

$$MU_{ET} = \sqrt{(27,000 \cdot 0.000007)^2 + (0.084060 \cdot 3)^2 + (0.070050 \cdot 0.1)^2} = 0.315 \text{ g}$$

Thus the result in this example would be:

ET = 12.60 g ± 0.32 g at a 95 % confidence level or stated as MU % = ±2.5%.

Conclusion—This example, which is representative of the measurement method as it is currently applied to woodstoves under the EPA NSPS, indicates that the uncertainty related to the dilution tunnel flow rate measurement and filter catch weights are the primary components of the overall uncertainty of the result. Tunnel flow rates may be much higher than necessary to capture all emissions. This can influence the uncertainty of the test method. For example, increasing the tunnel flow rate from 150 to 600 scfm in the above example increases the MU of the result to ±0.8 g or about ±6.5 % of the measured emissions. Therefore, keeping tunnel flow rates near the minimum necessary to reliably capture the exhaust stream, while keeping the tunnel velocity at a level that can be accurately measured during tests will minimize the uncertainty of the measurement.

Measurement Uncertainty Estimate for EPA Method 5G-1

The major difference between EPA Method 5G-3 and 5G-1 is that 5G-1 uses solvent recovery to determine the particulate weight collected in the probe and front half of the front filter holder. The requirement to weigh a solvent blank results in at least one additional weighing before and after the test when compared to Method 5G-3. It is also common practice to weigh the 100mm filters separately which also results in another weighing before and after the test. These additional weighings add to uncertainty. There is also the likelihood of more substantial fiber loss while recovering adhered filter fibers from the filter gaskets with the 100 mm filters and the possibility of some loss of particulate from the probe during the solvent recovery process.

If the same balance uncertainty of 0.1 mg for six total weighings is combined with a conservative estimate of 0.1% of total catch for the combination of fiber loss from the gaskets and particulate loss during solvent recovery is used, the resultant uncertainty for the filter and probe catch weight is significantly higher than with Method 5G-3.

However, the sample flow rate, and therefore the probe and filter catch, with Method 5G-1 is typically about twice as high as Method 5G-3 and this difference offsets much of the impact of the increase in weighing uncertainty on the overall method measurement uncertainty. Following the same analysis track as shown for Method 5G-3 above, the Method 5G-1 MU is about 10 - 15% higher than the MU for 5G-3. This MU% range of ± 2.7 to 3% is still quite good.

Measurement Uncertainty Estimate for EPA Method 5G-2

Using the same logic applied to the Method 5G-1 measurement uncertainty estimate, the primary difference between 5G-1 and 5G-2 is in the number of filter train components subjected to solvent recovery and the total number of weighings before and after each test run. After consultation with a highly experienced 5G-2/5H test technician, it was determined that there are typically seven individual weight measurements before and after each test run. These include four beakers used for solvent recovery, a beaker for the solvent blank and the front and rear filters. These additional weighings contribute to additional uncertainty.

It would not be unreasonable to assume that some additional fraction of particulate matter could be lost during solvent recovery due to all the additional components that are involved in the recovery. But, even if we continue to assume that a very conservative 0.1% of the total particulate catch is lost from filter gaskets and solvent recovery, the weighing uncertainty increases by about 40% over Method 5G-1. However, since the total catch is expected to be somewhat higher with 5G-2, the larger catch mitigates some of the impact of the increased weighing uncertainty on the overall method measurement uncertainty. We would expect that the measurement uncertainty for Method 5G-2 is in the range of 15 -20% higher than Method 5G-1. The MU% is in the range of $\pm 3 - 3.5\%$ which is also quite good.

Overall Measurement Uncertainty Estimate for EPA Method 5H

The same particulate sampling equipment is used for both Method 5H and Method 5G-2, but the methods are not close to the same in terms of overall measurement uncertainty. If we first examine the sampling protocols for the two methods, we find there are notable differences. These include:

- The particulate concentrations in the 5H sampling stream are much higher than 5G-2 (undiluted stack gas in 5H versus diluted gas in the tunnel with 5G-2).
- Lower sampling rates are typically used for 5H to prevent filter plugging (0.1 – 0.3 cfm for 5H and 0.5 cfm or higher for 5G-2).
- Total particulate catches are higher (by an order of magnitude or more) for 5H compared to 5G-2.

However, despite these differences, the particulate catch measurement uncertainty for Method 5H is not appreciably higher than that for 5G-2. This is because the impacts of the differences noted above cut to some extent in different directions, e.g., the added uncertainty contributed by the low sampling rate is somewhat mitigated by the much higher total particulate catch. Since the measurement uncertainty in the sample flow rate

and the total particulate catch parameters are reasonably quantifiable, we would estimate their combined contribution to the overall measurement uncertainty of Method 5H using an ISO Guide Type A analysis to be no more than a few percentage points. At this level, the sampling-related measurement uncertainty is only a secondary contributor to the overall measurement uncertainty of Method 5H.

The primary contributor to the large increase in overall measurement uncertainty associated with Method 5H when compared to Method 5G-2 (or the other dilution tunnel methods) is the determination of the stack gas flow rate over the course of the test. The associated complexities with this issue make quantitative assessments of the contribution to measurement uncertainty unworkable using ISO Guide Type A evaluations. The more qualitative Type B evaluation is appropriate in this case.

Factors Contributing to Stack Flow Uncertainty

The stack flow must be determined for purposes of knowing where to set the sample flow rate in order to maintain sample flow proportionality over the test run. The procedure specified in Method 5H is a tracer gas method. This method has been generally replaced by adding a dilution tunnel to the system and determining stack flow using the ratio of tunnel CO₂ to stack CO₂ multiplied by the easily measured tunnel flow. At each sampling interval, the sample flow is adjusted based on this stack flow determination. There is additional uncertainty in this procedure if a single gas analyzer is used to determine both the stack and tunnel CO₂ concentrations since the accuracy of these instruments is generally 1% of full scale and the tunnel CO₂ readings are quite low. Some of this can be mitigated by using two analyzers or multi-scale analyzers where the accuracy can better match the concentrations.

The 5H procedures then go on to calculate stack gas flow for the purpose of determining emissions using a mass balance algorithm where carbon from the fuel is balanced against the carbon measured in the stack (CO₂ and CO) and an assumed hydrocarbon value to make the balance work. The stack flow calculated by this means often results in very different incremental stack flow values than the tracer gas methods. The impact of this difference is that the proportionality of the sample gas flow rate to the stack gas flow rate can be much different between the two procedures. Additionally, there is a wider range in variation from proportional sample flow rate allowed in this method than for the dilution tunnel methods. The allowable range is a necessity due to the fact that frequent stack gas flow calculations and sample flow rate adjustments are needed due to the instability in the actual stack gas flow. The impact of proportionality ranges is hard to assess but contribution to method uncertainty is definitely real. Finally, the total stack gas flow over the entire test run is also impacted by the mass balance procedure and this obviously directly impacts the particulate emission determination. The uncertainty of the carbon balance stack flow rate calculation is impacted by the unknown uncertainties of the fuel carbon content, the assumed hydrocarbon concentration and the measurement uncertainty of stack CO and CO₂ measurements.

Overall measurement uncertainty for 5H is commonly estimated in the $\pm 20 - 30\%$ range by those with considerable experience with the method. The increment of measurement uncertainty that can be analyzed using ISO Guide Type A procedures accounts for only a

few percentage points of this range. The remainder (~17 – 27%) can be supported by a qualitative (Type B) analysis using reasonable estimates of the measurement uncertainties associated with the stack flow procedures in Method 5H. This estimate corresponds to the magnitude of the typical difference between stack flow rates determined by the tracer gas and carbon balance procedures for an individual test run.

REFERENCES

1. 40 CFR, Part 60, Subpart AAA – Standards of Performance for New Residential Wood Heaters.
2. EPA Woodstove NSPS Proficiency Test Results – Supplied by Michael Toney – EPA OAQPS, Emissions Measurement Branch.
3. Guide to the Expression of Uncertainty in Measurements, First Edition 1995, International Organization for Standards.
4. ASTM E177-2008 “Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods”, ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.
5. ASTM E691-2009 – “Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method”, ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.
6. CSA B415.1-2010 - Performance Testing of Solid-Fuel-Burning Heating Appliances, CSA International, 5060 Spectrum Way, Suite 100, Mississauga, Ontario, Canada L4W 5N6, March, 2010.
7. ASTM E2515 – 2007 – “Standard Test Method for Determination of Particulate Matter Emissions Collected by a Dilution Tunnel”, ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.
8. ASTM E2618 – 2009 – “Standard Test Method for Measurement of Particulate Emissions and Heating Efficiency of Outdoor Solid Fuel-Fired Hydronic Heating Appliances”, ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.
9. Mark’s Standard Handbook for Mechanical Engineers, Ninth Edition, Eugene A Avallone, Theodore Baumeister III, McGraw Hill Book Co., 1987.