Ferguson, Andors & Company

Consultants in Product Development and Regulatory Compliance

An Evaluation of EPA's New Wood Heater NSPS

Compliance Determination Concept

Prepared for:

Hearth, Patio & Barbecue Association

Prepared by:

Robert Ferguson
Ferguson, Andors & Company

Technical Advice and Report Review by:

Richard Reiss, Sc.D.
Principal Scientist, Exponent, Inc.

May 1, 2014

Executive Summary

As part of the current proposal to modify 40 CFR 60 Subpart AAA and to add additional subparts for regulation of particulate matter (PM) emissions from the full range of new wood-burning residential heating equipment, EPA is considering changing the way compliance with the emission limits is determined.

Subpart AAA, which has been in effect for 25 years, uses a burn rate based probability distribution data weighting methodology where test runs are conducted in each of four burn rate categories spanning the full range of burn rate capabilities of the heater from lowest to highest. The probability distribution is part of EPA Method 28 and is based on in-situ data gathered from woodstove users in the 1980's. The particulate emission test results from each burn rate category are assigned a fractional weighting in the overall emission average determination based on each discreet burn rate probability. Because the probability "bell curve" is centered at a burn rate of about 1.2 dry kg/h, burn rates in the low and medium low burn rate categories receive much more weighting than data in the higher burn rate categories.

A similar weighting concept is also applied within the current EPA Hydronic Heater Voluntary Program. The results for test runs in four heat output categories are weighted based on a heat output probability distribution that was modeled on heating demands for a typical home in a northeastern United States location. As with woodstoves, emissions for the lower heat output categories get most of the weighting.

EPA's current NSPS proposal includes a two or three step approach where the PM emission limits ratchet down five years (2 Step) or three years and eight years (3 Step) after the effective date of the Step 1 standards. The Step 1 standards continue with the data weighting methodology used in the current Subpart AAA and in the Hydronic Heater Voluntary Program. However, the proposal includes a significant departure from the current data weighting methodologies for determining compliance with the proposed Step 2/3 emission limits. This new procedure (referred to herein as the "new compliance algorithm") will be applied to woodstoves, hydronic heaters and warm air furnaces.

The new compliance algorithm relies on data from only the lowest and highest burn rate (or heat output) categories for the appliance. These are currently referred to as Category 1 and Category 4. Two initial screening test runs are conducted—one in Category 1 and the other in Category 4. Two additional test runs are then conducted in the category with the worst emissions during the screening runs. While the proposal is unclear in this regard, it appears that EPA intends that the average of the three worst case test runs as well as the single "best" screening run must meet the emission limit in order for compliance to be achieved. However, other interpretations of the EPA's language are possible.

Since this new compliance methodology has never been used in wood-burning product regulations, we have to look for available data that can be useful as well as employ other statistical means for evaluating the impacts and risks that might be imposed by this new compliance algorithm concept.

The HPBA Enhanced EPA Certified Woodstove Database allows the results for the Category 1 and Category 4 test runs for a large group of currently produced certified stove models to be examined. These data include only one run per category for the most part but clearly highlight the difference between the current weighted average emissions algorithm and the emission values for individual burn rate categories that comprise those weighted averages. Figs. 3 – 7 show that information graphically.

The EPA Laboratory Proficiency Round Robin Test Program also provides information that is useful. Besides allowing the precision of the test methods to be estimated, the pairs of data in each burn rate category for each of the labs and test stoves for each of the test program years allows a basic assessment of the variability that can be expected. These data for the Category 1 and Category 4 test results are presented in Figs. 8 – 13. But even these datasets do not allow the impacts of three test runs in the same burn category to be evaluated. Since there are no actual data available, a different approach must be used. A statistical simulation sometimes referred to as a Monte Carlo simulation or analysis can be employed to model the probable outcomes of the proposed compliance algorithm and possible variations.

The EPA Laboratory Proficiency Round Robin Test Program Category 1 and Category 4 data can be used to determine the inter- and intra-laboratory variability for those run categories which are used in the proposed compliance algorithm. The variability can be expressed in terms of standard deviations or, because of the wide range of mean emission values in the round robin data, coefficients of variation (CVs) which expresses the variability for each burn rate category as a percentage of mean emission rate.

Since emission test data, regardless of the degree of variability inherent in the test methods and in the random nature of burning wood, can’t have values less than zero, log-normal distribution best models the range of outcomes expected around a true mean emission value. The log-normal distribution constrains values to be positive, consistent with any possible emission value, and includes the right skew of the emissions data distribution.

By inputting values for the estimated inter-and intra-laboratory coefficients of variation for Category 1 and Category 4 burn rates, as well as a range of hypothetical true mean emission performance values for a given stove model at a given burn rate, two-stage log-normal distributions were created. A large number of randomly generated probable outcomes can then be sampled from the modeled distributions. By comparing the probable results of individual test runs, the average of groups of three outcomes or combinations of both with the proposed passing grade, the chance of passing or failing the various scenarios can be predicted. The use of CV values represents a very conservative approach because it assumes that the variability is consistent across all emission values, even at low emission values implicated by the proposed emission limits. The test methods have not actually demonstrated this to be case and due to the inherent variability when burning wood, the test methods have a discrimination threshold for emission determination below which any measured values cannot be assumed to represent the true mean. Sensitivity cases where higher CV values at lower emission rates are modeled can help inform the impact of the discrimination threshold. The Monte Carlo analyses for this new compliance concept point out that it has potentially high risk for manufacturers trying to get new products certified to Step 2/3 emission limits.

The Monte Carlo simulation can also be used to evaluate the impacts of conducting additional test runs when one of the initial test runs is higher than anticipated. This is the "outlier" provision included in the NSPS proposal where the emission results from one or more new test runs can replace emission results from one of the test runs in the initial worst case three run test series.

The proposed compliance algorithm for Step 2/3 appears to present a myriad of issues if adopted. Many EPA certified models have emission performance profiles that are not flat, with better performance generally focused in the range of primary concern under the current compliance algorithms —the lower burn rates. Sacrificing the performance at the highest burn rates has often been the necessary trade-off that stove designers have needed to make to insure the best performance at the heavily-weighted low burn rates while still meeting maximum heat output expectations from consumers. Although it is known that homeowners most often operate their heating appliances at lower burn rates (heat outputs) to match the typical heating demands of their homes, there are times when high heat output is needed. For example, a cold house after a prolonged absence or on the coldest winter nights. In these cases, the stove owner expects the appliance to provide substantially higher heating capacity, even if it is only occasionally or for a short duration. Manufacturers have found that maximum heating capacity is an important specification for stove purchasers. Equalizing the importance of the highest burn rate or heat output emission results to those from the lowest burn rate or heat output, while ignoring the emission performance in between, as EPA is now proposing, is a radical change with many product design implications. And, it obviously penalizes manufacturers for making the design choices clearly implicated by the current compliance algorithms.

The conclusions that can be drawn from the Monte Carlo simulations are quite chilling. For manufacturers to have a high (95%) confidence level that their stove models will meet the either of the proposed emission limits using the new compliance determination algorithm, true means of PM emission performance of that model in both the Category 1 and Category 4 burn rates (or heat output) must be less that 50% of the emission limit, even if the conservative CV values were to actually be achievable by the test methods. If the variability of the test methods is higher than predicted based on assuming that the average CV values determined from the EPA proficiency test data can be applied over all emission rates, the probability of failure increases at the proposed emission limits. As with all testing procedures, the various solid-fuel heater emission test methods each have a lower threshold below which there is no ability to reliably discriminate differences between emission test results. This is the case when using CV values results in predicted standard deviations at low emission rates that are well below the test method discrimination threshold. For example, we would be skeptical of the ±0.3 g/h standard deviation predicted by using an intra- or inter-lab CV value of 30% for a model with an assumed true mean emission rate of 1.0 g/h at a given burn rate. There is simply no evidence to support that this level of precision is achievable under any circumstances. To the contrary, there is significant evidence to dispute this possible level of precision. Additional simulations have been conducted using the mean CVs plus one standard deviation above the means. This shows the impact of higher variability at low emission rates. However, it is likely that even this adjustment to the CV values doesn't adequately address the concerns about the discrimination threshold for the method since the resultant standard deviations are still lower than have ever been demonstrated, and therefore, would still result- in under-estimations of the risk of failing compliance with Step 2/3 emission limits based on the proposed new compliance algorithm.

There is another component of the EPA proposal that also adds uncertainty, and therefore more risk. That is the proposed transition from crib fuel to cordwood fuel for emission testing for all product categories for determining compliance with Step 2/3 emission limits. There is simply no data to inform the impacts on emission outcomes or test method precision of changing test fuel from cribs to cordwood. Cordwood emission performance using standardized test methods is generally unknown for EPA certified stove models that have been designed and tested using crib fuel for the past 25 years. And efforts are still underway to create a new cordwood test method for woodstoves that better reflects homeowner use patterns, so any cordwood data that currently exists would be irrelevant relative to any final cordwood methods, in any event. Moreover, cordwood test method precision can't be determined until any new test method is finalized and even then, only with a properly designed and executed multi-sample, multi-lab test program. It is reasonable to assume that the variability when burning cordwood will not be better than when burning cribs. It is therefore also a reasonable assumption that using the test method precision, standard deviations about a mean emission value and coefficients of variation determined using available crib data provides what can only be considered an absolute best case prediction of what might be expected with cordwood test results with the anticipated new test method. Not whether the risk of failure to achieve compliance is most likely to increase with cordwood testing but by how much the risk increases is the question. And the same or even greater concerns apply to applying this analysis to predict the risks associated with the proposed new compliance algorithm for other appliance categories beyond woodstoves (e.g., pellet stoves, hydronic heaters, and warm air furnaces). The test methods for these categories are new or relatively new, and no comprehensive evaluation of their precision has been performed or is even possible. Beyond that, some of these methods involve additional measurements beyond PM measurements, which are likely to raise their own significant precision issues (e.g., heat output measurements in air plenums for warm air furnaces).

Introduction

These analyses and report were prepared by Robert Ferguson, President of Ferguson, Andors & Company for the Hearth, Patio & Barbecue Association. Assistance with the analyses and review of the report were provided by Dr. Richard Reiss, Principal Scientist, Exponent, Inc. Curriculum Vitae are provided for Mr. Ferguson and Dr. Reiss at the end of this report.

Evaluation of EPA's New Compliance Algorithm Concept

EPA's NSPS proposal includes the use of a new algorithm for determining compliance with the Step 2/3 standards. Although the proposal language is not explicit, EPA appears to intend that the new compliance determination methodology will be applied to woodstoves, hydronic heaters, warm air furnaces, and possibly other appliance categories as well.

Because of the long history of EPA regulation, woodstoves provide the best vehicle to examine the differences between EPA's proposed new compliance algorithm and the current Subpart AAA compliance methodology. In 2010[1], HPBA completed and provided to EPA an enhanced database that includes data from individual EPA certification test runs for a large number of currently produced woodstove models. This significant level-of-effort project was undertaken because EPA's list of certified wood heaters under Subpart AAA includes all heater models certified since the beginning of NSPS certifications in 1988. The list includes many duplicates caused by certification transfers due to acquisitions as well many discontinued models and defunct manufacturers. The cropping of the list was accomplished by surveying certified wood heater producers and obtaining their current model line-ups as well as the detailed certification test data for those current models.

In addition, under the current NSPS, EPA conducted a multi-year accredited laboratory proficiency round-robin test program where a single test stove was shipped from lab to lab. Two complete certification test series were conducted on the sample stove in each round-robin cycle. The test stove model was changed several times over the course of the proficiency test program and has included stoves with catalytic and non-catalytic technology. This proficiency test database – which exists only for woodstoves – provides the only data available that allows test method precision to be determined. In 2010, a rigorous evaluation of test method precision was undertaken using recognized analytical procedures[2]. The results of those analyses have been provided to EPA and otherwise widely distributed to interested parties.[3]

---

[1] As noted, the enhanced database was completed in 2010; an updated enhanced database reflecting the universe of currently (2014) produced certified heater models is not currently available.

[2] ASTM E691 – *Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method*

[3] Curkeet, Rick and Ferguson, Robert *EPA Wood Heater Test Method Variability Study Analysis of Uncertainty, Repeatability and Reproducibility based on the EPA Accredited Laboratory Proficiency Test Database,* Rick

In summary the analyses showed:

- The "within the same lab" precision of the weighted average emissions at the 95% confidence level is at best ~ ±3 grams per hour.
- The "between labs" precision is poorer (at least ±4.5 grams per hour again at 95% confidence).

What does this mean?

- The EPA test methods are certainly capable of reliably distinguishing between good and bad performance, but they cannot reliably distinguish between "good, better and best" performance.
- The EPA test methods cannot reliably distinguish emissions performance differences of less than 3 grams per hour.

Current Woodstove NSPS Testing and Compliance

Under the current NSPS requirements, a minimum of four individual test runs are required spanning the full operating range of the appliance from the lowest to highest achievable burn rates. The burn rate categories are defined in EPA Method 28[4] as:

BURN RATE CATEGORIES

[Average kg/hr (lb/hr), dry basis]

Category 1	Category 2	Category 3	Category 4
< 0.80 (< 1.76)	0.80 to 1.25 (1.76 to 2.76)	1.25 to 1.90 (2.76 to 4.19)	Maximum burn rate

The exception is that stoves that are not capable of achieving a burn rate less than 0.80 kg/h can conduct a test run at less than 1.00 kg/h and still meet the requirements. This exception is most often employed with non-catalytic stove models.

The particulate emission results from the four burn rate categories are averaged using a weighting scheme based on a probability distribution as shown in the Fig 1.

Fig. 1

Curkeet, PE, Chief Engineer – Hearth Products, Intertek Testing Services, Inc. and Robert Ferguson, Ferguson, Andors & Company, October 6, 2010.

[4] EPA Method 28 – *Certification and Auditing of Wood Heaters*

This averaging protocol was based on the burn rate distributions from several in-home studies conducted in the Northeast and Northwest that confirmed that stoves are used most frequently at lower burn rates. . The center of the distribution is at ~ 1.15 kg/hr.[5,6]

For a typical stove model, that means that the lower burn rates receive the most of the weighting when determining compliance with the standards. The high burn category emission results receive less weighting relative to the emission results from the lower burn rate tests.

It is also important to note that the current standards include an individual test run "not-to-exceed cap" that is significantly above the level of the standards. To achieve compliance, the weighted average emissions of the test run series must meet the appropriate emission limit AND all test runs required to be used when determining the weighted average emissions[7] must have emissions levels at or below the cap value. The caps and corresponding passing grades are shown in Fig 2.

[5] The low burn rate requirement was based on burn rate data from in-home studies, without considering the relationship between the laboratory test burn rates determined with Method 28 procedures (including dimensional lumber crib fuel) and the in-home burn rate determinations (using cordwood and a different method for determining burn rate). A recent evaluation of this "apples and oranges" problem by Dr. James Houck (An Evaluation of Method 28 Wood Heater Burn Rates, September 2009) shows a significant disconnection between the two values for the same stove.

[6] In 2009, an ASTM replacement for EPA Method 28, the fueling and operating protocol for woodstoves, was published. The four burn rate probability weighting scheme from M28 was replaced with simplified three burn rate weighting formula where the low and medium burn results each receive 40% of the weighting and the high burn results 20%. This simplification was based on analyses of the enhanced certification database where the simplified weighted scheme resulted in similar average emission rates for most certified stove models. EPA participated in the ASTM process that resulted in the new method and weighting, and raised no objections to it.

[7] The current EPA Method 28 protocol also includes an additional test run ("outlier") provision where only two-thirds of the emission data in any burn rate category are required to be used when determining compliance.

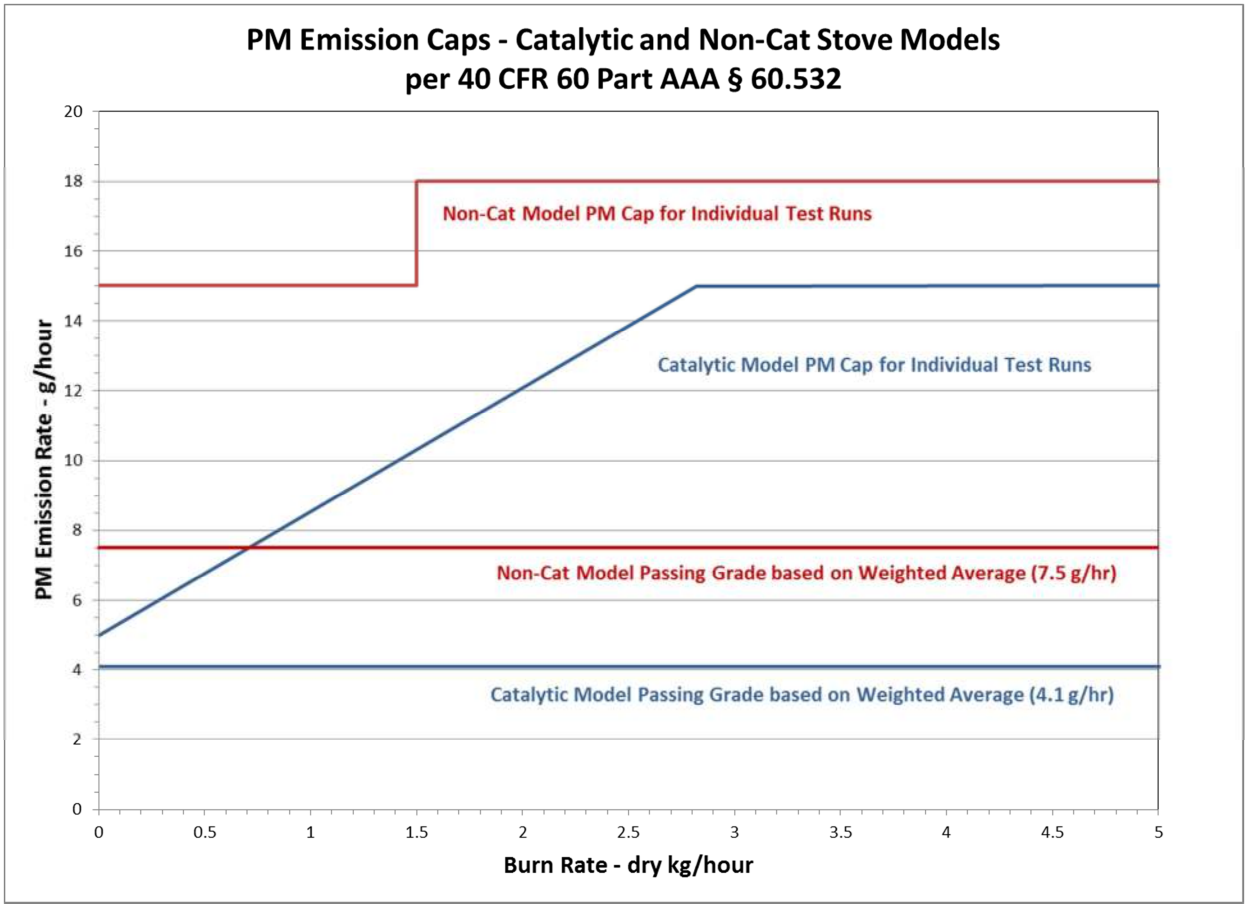

Fig. 2

Stove models often have higher emissions in one burn rate category, but have still achieved compliance because of the distribution of the weighting. This can be seen in the Fig. 3 where some models with low certification values have higher emissions at some burn rates. Two models representing the lowest range of certification values are shown for reference.

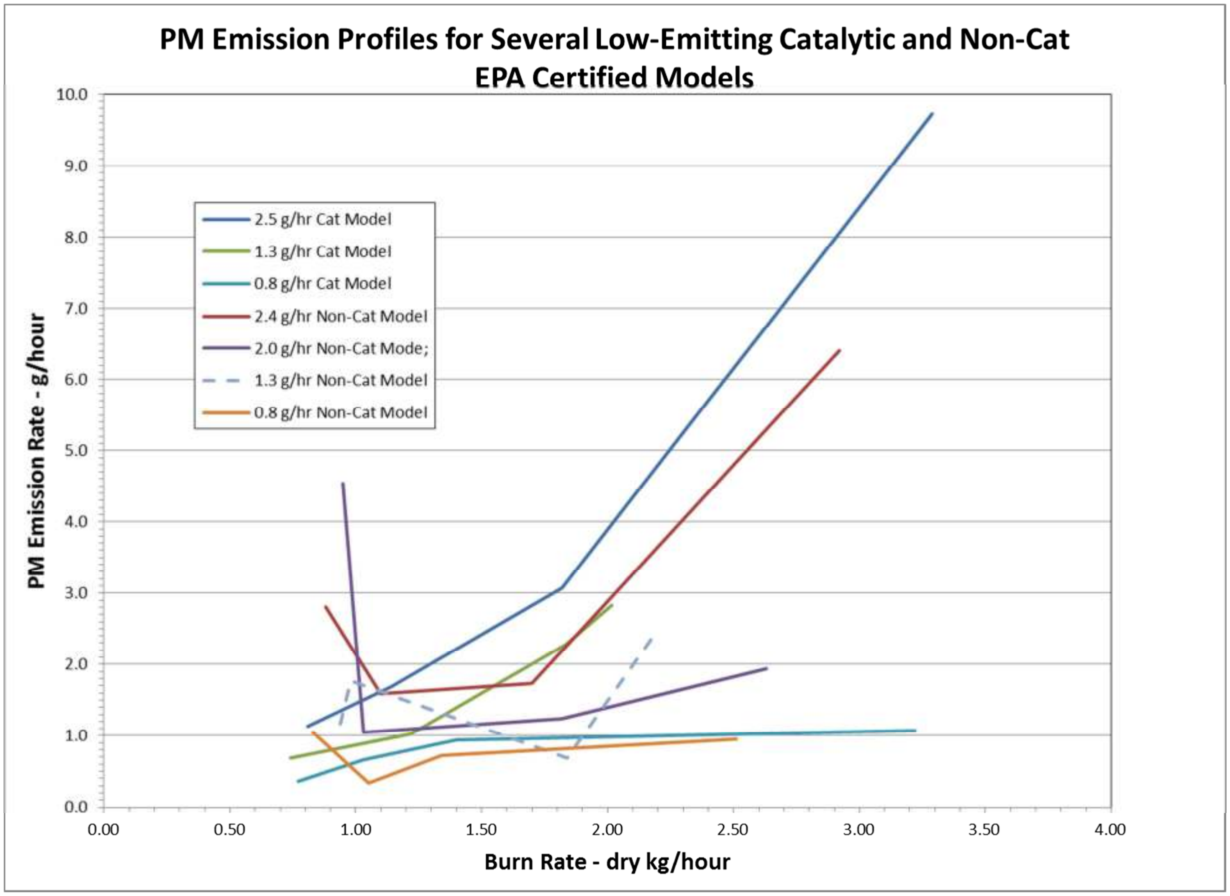

Fig. 3

Proposed New Compliance Algorithm

EPA has proposed abandoning the four burn rate probability distribution weighting scheme in favor of a very different compliance determination algorithm for Step 2/3 NSPS standards.[8,9,10,11] The EPA proposal includes the following methodology for determining compliance for Step 2/3 standards:

"Screening" Test Runs

1. Conduct one emission test run at the Category 4 air setting (maximum combustion air).
2. Conduct one emission test run at the Category 1 (low fire) air setting.
3. Conduct two additional test runs at the combustion air setting with the worst emission results from the first two screening test runs.

[8] **Federal Register** / Vol. 79, No. 22 / Monday, February 3, 2014 / Proposed Rules, p. 6342. "We are also proposing that the burn rates not be weighted at all for the Step 2 standards but rather that the emission limits be separate for Burn Rate Category 1 (lowest burn rate category) and Burn Rate Category 4 (maximum burn rate category) and that compliance for each be shown separately.

[9] **Federal Register** /Vol. 79, No. 22 /Monday, February 3, 2014 / Proposed Rules, p. 6367. "Further, we are proposing new compliance requirements for Step 2 with emissions limits at the lowest burn rate (Category 1) and the maximum burn rate (Category 4), not a weighted average of the four burn rates, as in the current 1988 NSPS."

[10] **Federal Register** / Vol. 79, No. 22 / Monday, February 3, 2014 / Proposed Rules, p. 6380. §60.534(a)(3).

[11] **Federal Register** / Vol. 79, No. 22 / Monday, February 3, 2014 / Proposed Rules, p. 6386. §60.5476(b) and (c).

Determination of compliance with the standards is not clear in the proposal. Statements in the proposal can be interpreted to require:

1. The average of the three "worst case" test runs must be at or below the level of the standard **or** that all worst case runs must be below the level of the standard.
2. The single test run with the lower emission value from the first two screening test runs must be at or below the level of the standard.
3. All test runs must be below the level of the standard.
4. No specific provision for addressing additional test runs for the Category 1/Category 4 compliance algorithm has been proposed, however, nor is there any indication given that any such provision would be conceptually different than the "outlier relief" provisions in the current test methods where "The results from at least two-thirds of the test runs in a burn rate category shall be used in calculating the weighted average emission rate."[12,13,14,15]

To bracket the possible interpretations of how EPA might employ this new compliance determination concept, four analyses have been conducted. These are:

1. Examine the emission performance outcome for Category 1 and Category 4 burn rate data from currently produced certified stove models.
2. Examine the emission profiles for a number of current models with low Method 28 weighted average emission values.
3. Analyze the Category 1 and Category 4 test results from the EPA laboratory proficiency test round robin.
4. Use a statistical simulation to assess the probabilities of compliance using realistic input values for critical modeling parameters.

Category 1/Category 4 Analyses

The weighted average emissions using current EPA Method 28 Probability Distribution data weighting is not a good predictor of compliance with the proposed new compliance algorithm. By comparing the Category 1 and Category 4 emission data from a large subset of currently

---

[12] **Federal Register** / Vol. 79, No. 22 / Monday, February 3, 2014 / Proposed Rules, p. 6394. Test Method 28R for Certification and Auditing of Wood Heaters, Section 2.1

[13] ASTM E2780 - *Standard Test Method for Determining Particulate Matter Emissions from Wood Heaters***.** Section 9.5.13

[14] **Federal Register** / Vol. 79, No. 22 / Monday, February 3, 2014 / Proposed Rules, p. 6398. Test Method 28 WHH for Measurement of Particulate Emissions and Heating Efficiency of Wood-Fired Hydronic Heating Appliances, Section 12.6.

[15] **Federal Register** / Vol. 79, No. 22 / Monday, February 3, 2014 / Proposed Rules, p. 6408. Method 28WHH–PTS A Test Method for Certification of Cord Wood-Fired Hydronic Heating Appliances With Partial Thermal Storage. Section 12.6.

produced stove models[16] with the level of standards expected in the NSPS proposal, a clear trend can nevertheless be shown.

Since the current NSPS proposal imposes a Step 1 weighted average emissions limit of 4.5 g/h for woodstoves (this is the current Washington State limit for non-catalytic stove models) and a Step 2 limit of 2.5 or 1.3 g/h, the available data can be analyzed to see the impacts when only the Category 1 and Category 4 test run data are considered. Of the 96 certified non-catalytic stove models in the database with Method 28 weighted emissions values at or below 4.5 g/h, 48 models have individual Category 1 or Category 4 emission values above 4.5 g/h. Of the 22 non-cat models in the database that meet a 2.5 g/h Method 28 weighted average emission limit, 11 have individual Category 1 or Category 4 emission values above 2.5 g/h. Of the 5 non-cat models in the database that meet a 1.3 g/h Method 28 weighted average emission limit, 2 models have individual Category 1 or Category 4 emission values above 1.3 g/h. This impact is shown graphically in Fig. 4.

Of the 13 certified catalytic stove models in the database with weighted emissions values at or below the current Washington State limit of 2.5 g/h for catalytic models, 8 models have individual Category 1 or Category 4 emission values above 2.5 g/h. Of the 3 catalytic models in the database that meet a 1.3 g/h Method 28 weighted average emission limit, 1 model has a Category 4 emission value above 1.3 g/h. This impact is shown graphically below in Fig. 5.

[16] *HPBA Enhanced EPA Certified Wood Heater Database - Excel Workbook,* Robert Ferguson, Ferguson, Andors & Company, February 25, 2010

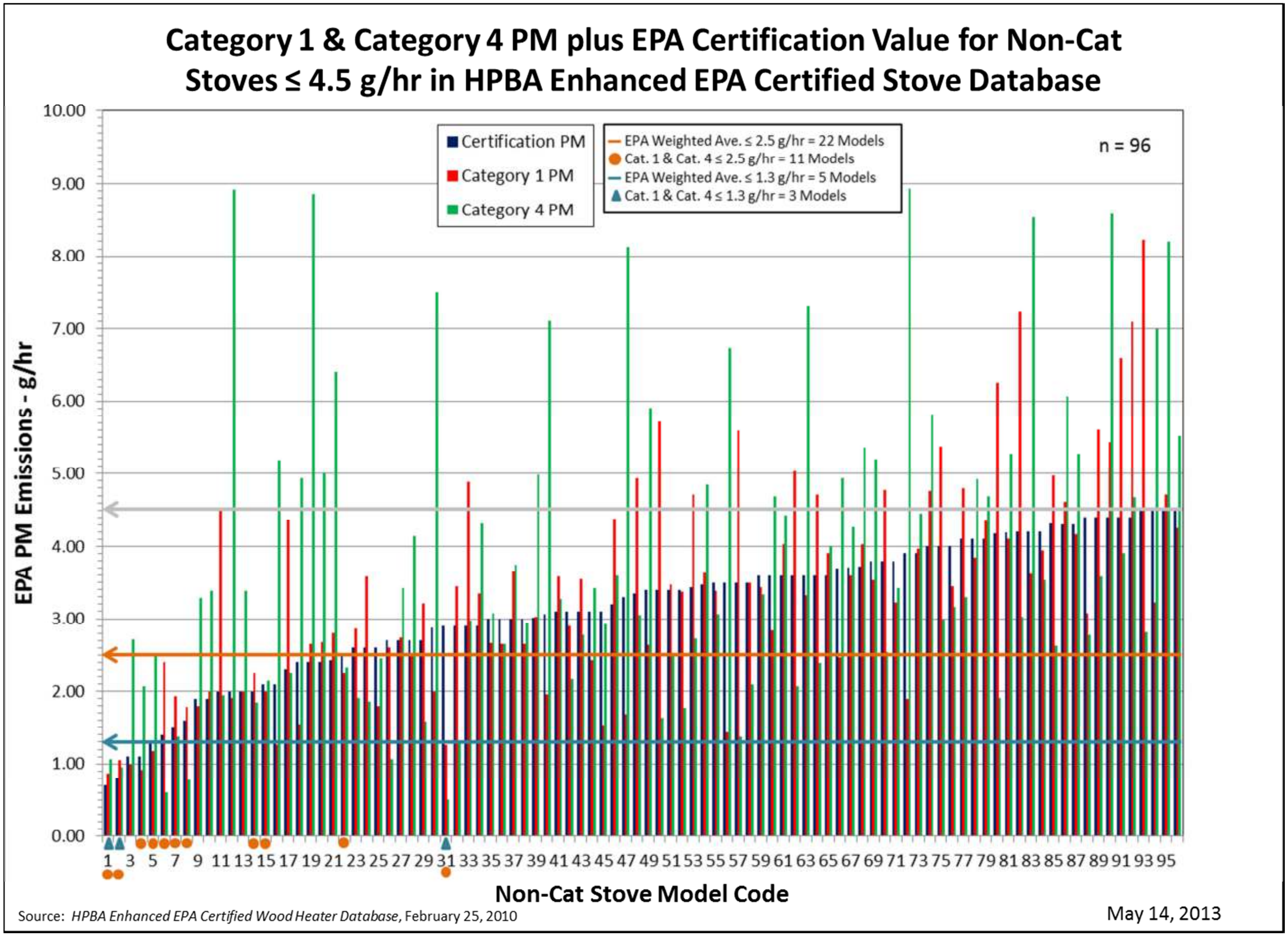
Category 1 & Category 4 PM plus EPA Certification Value for Non-Cat Stoves ≤ 4.5 g/hr in HPBA Enhanced EPA Certified Stove Database
Certification PM
Category 1 PM
Category 4 PM
EPA Weighted Ave. ≤ 2.5 g/hr = 22 Models
Cat. 1 & Cat. 4 ≤ 2.5 g/hr = 11 Models
EPA Weighted Ave. ≤ 1.3 g/hr = 5 Models
Cat. 1 & Cat. 4 ≤ 1.3 g/hr = 3 Models
n = 96
EPA PM Emissions - g/hr
Non-Cat Stove Model Code
Source: HPBA Enhanced EPA Certified Wood Heater Database, February 25, 2010
May 14, 2013

Fig. 4

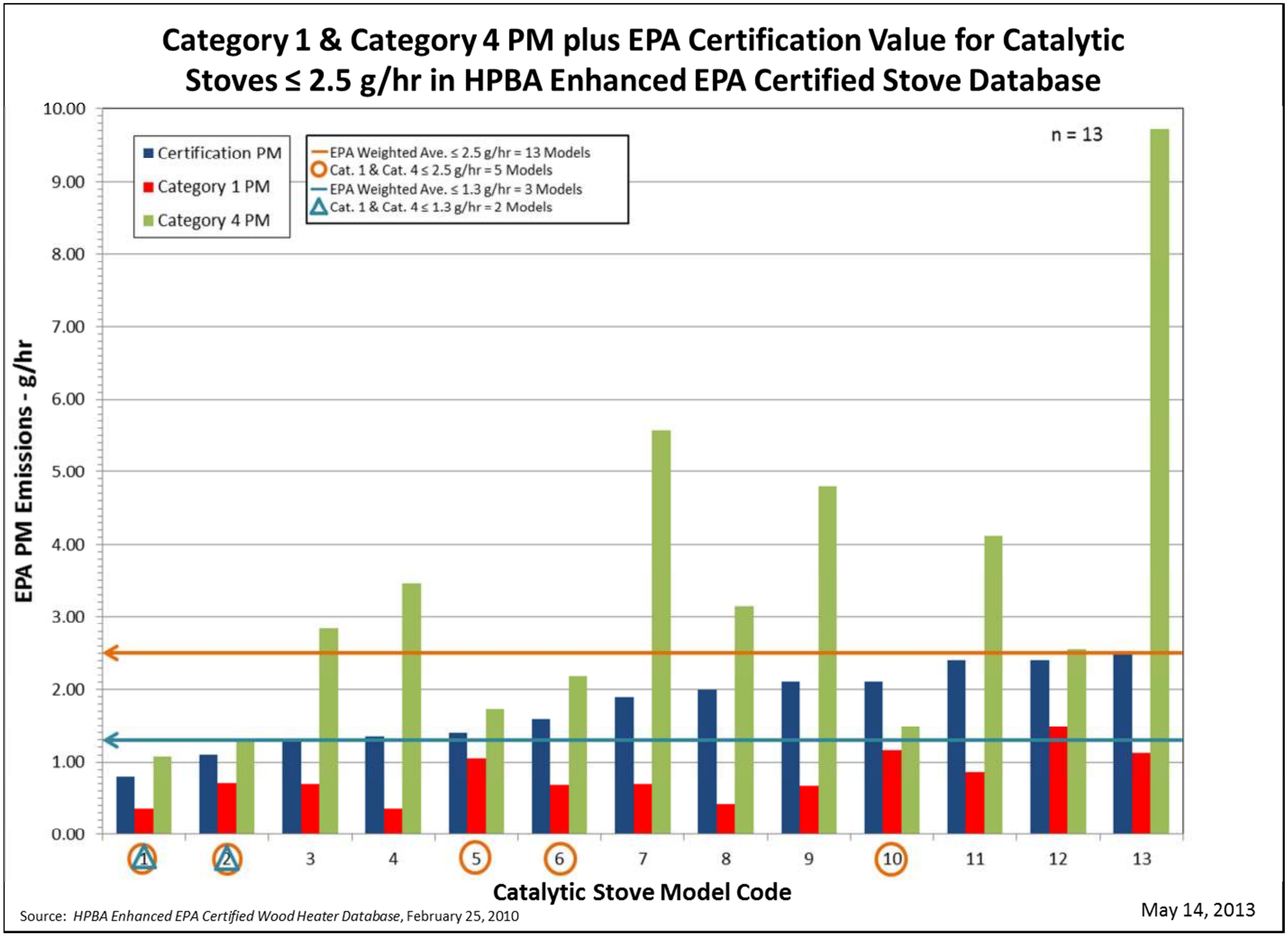
Category 1 & Category 4 PM plus EPA Certification Value for Catalytic Stoves ≤ 2.5 g/hr in HPBA Enhanced EPA Certified Stove Database
Certification PM
Category 1 PM
Category 4 PM
EPA Weighted Ave. ≤ 2.5 g/hr = 13 Models
Cat. 1 & Cat. 4 ≤ 2.5 g/hr = 5 Models
EPA Weighted Ave. ≤ 1.3 g/hr = 3 Models
Cat. 1 & Cat. 4 ≤ 1.3 g/hr = 2 Models
n = 13
EPA PM Emissions - g/hr
10.00
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
Catalytic Stove Model Code
Source: *HPBA Enhanced EPA Certified Wood Heater Database*, February 25, 2010
May 14, 2013

Fig. 5

Even pellet stoves face the same challenge. Fig. 6 shows that the six models in the enhanced database meet 1.3 g/h but five of those models have Category 1 and/or Category 4 emissions above 1.3 g/h.

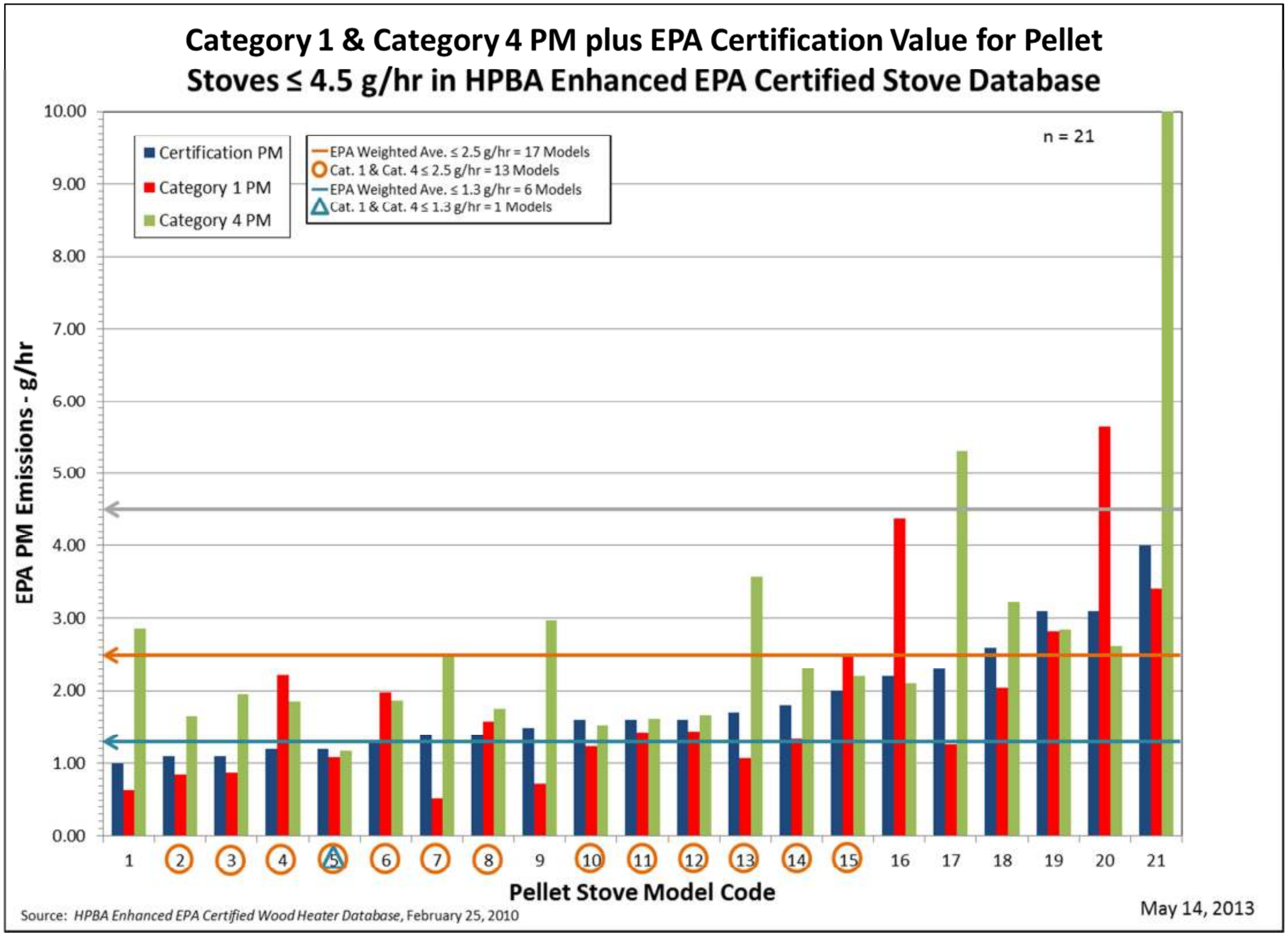

Fig. 6

Emission Profile Analyses

One has only to look again at the PM emissions profiles (Fig. 7) for some of models with the lowest certification values to see the pitfalls of trying to relate the weighted average emissions to the results from individual burn categories. Even with a simple qualitative examination it can be observed that the emissions profiles are typically not particularly flat, except for the two models rated at 0.8 g/h weighted average PM. You can argue that those models represent superior technology OR that they were fortunate to string together four good test runs –a real possibility since results well above and below the true mean of performance are equally probable due to the poor precision of the test method. With the available data, you simply can't be sure which hypothesis is correct. And you definitely cannot draw any conclusions about the impacts of conducting multiple test runs in the same burn rate category using the certification test data.

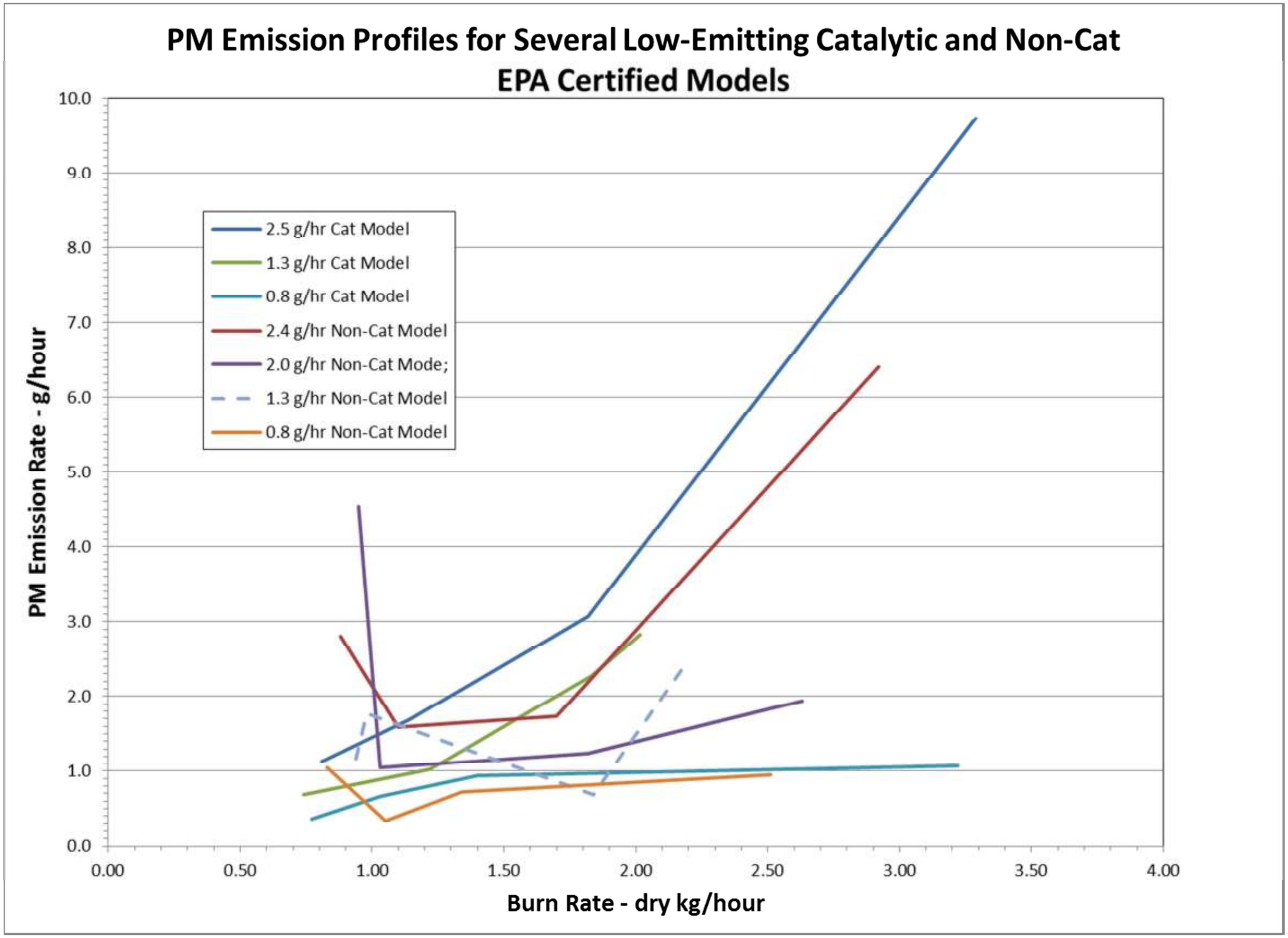
PM Emission Profiles for Several Low-Emitting Catalytic and Non-Cat EPA Certified Models
2.5 g/hr Cat Model
1.3 g/hr Cat Model
0.8 g/hr Cat Model
2.4 g/hr Non-Cat Model
2.0 g/hr Non-Cat Mode;
1.3 g/hr Non-Cat Model
0.8 g/hr Non-Cat Model
PM Emission Rate - g/hour
Burn Rate - dry kg/hour
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
0.00
0.50
1.00
1.50
2.00
2.50
3.00
3.50
4.00

Fig. 7

EPA Accredited Laboratory Proficiency Round Robin Test Data

Looking again for sources of data to inform the issues surrounding EPA's proposed new compliance algorithm, the EPA Laboratory Proficiency Test Data can provide some additional insights beyond its usefulness for determining test method precision. Standard deviations have been estimated using the results from the ASTM E691 analyses of the proficiency test round-robin inter-lab precision[17]. The average standard deviation for inter-lab precision (S_R) which estimates the standard deviation for the same test unit when tested in different labs[18] has been determined. The range of S_R is 1.6 to 2.3 g/h with an average of 1.9 g/h. However, it must be recognized that these values are based on the weighted average emissions for the test stoves for all burn rate categories and not for any individual burn rate category.

Since the participating laboratories were instructed to conduct two test runs in each burn rate category, it is possible to look at the results of sets of data in the minimum and maximum burn rate categories. If the proficiency test data is analyzed looking only at the Category 1 and Category 4 test data for the three test stoves (Catalytic 1, Non-Cat 1 and Non-Cat 3) with adequate datasets, it is possible to establish standard deviation and coefficient of variation values more appropriate for analysis of the proposed new compliance algorithm. The results are presented in Tables 1 and 2. Statistical outliers[19],[20] or suspected outliers were excluded from these analyses. The result of excluding the outliers was lower standard deviations, which is conservative for our purposes.

[17] Op. Cit., Curkeet

[18] Ibid., Tables 3a, 3b, 3c

[19] Possible outliers are first flagged when "h" and "k" critical values in the ASTM E691 analyses are exceeded. If the suspect value was more than three standard deviations from the mean, based on the mean and standard deviation without the suspect value, the suspect value was designated as an outlier and excluded from the analysis.

[20] Op. Cit., Curkeet, p. 11, *Critical Values of the Consistency Statistics*—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.

Table 1 - Summary ASTM E691 Analyses for Category 1 and Category 4 Emission Results for Catalytic 1, Non-Cat 1 and Non-Cat 3 EPA Proficiency Round Robin Test Results

Model	Burn Rate Category	# of Labs	$\bar{\bar{x}}$	$s_{\bar{x}}$	s_r	s_R	r - 95%	R - 95%	Years Tested
Cat 1	1	10	1.82	0.49	0.749	0.783	2.1	2.19	2
Cat 1	4	11	5.56	1.363	2.56	2.56	7.17	7.17	2
NC 1	1	7	19.48	1.018	3.516	3.516	9.84	9.84	1
NC 1	4	6	7.82	1.107	1.475	1.521	4.13	4.26	1
NC 3	1	7	8.22	1.858	2.942	3.153	8.24	8.83	6
NC 3	4	7	11.07	4.28	4.66	5.73	13.06	16.04	6
Averages				1.7	2.7	2.9	7.4	8.1	

$\bar{\bar{x}}$ - 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
s_R - the reproducibility standard deviation
r- repeatability at 95% Confidence
R - reproducibility at 95% Confidence

Table 2 - Summary of Coefficient of Variation (CV) Determinations for Category 1 and Category 4 Emission Results for Catalytic 1, Non-Cat 1 and Non-Cat 3 EPA Proficiency Round Robin Test Results

			Average CV	
Stove	Category	Mean Emission Rate	Intra-Lab	Inter-Lab
Catalytic-1	High BR	5.593	0.354	0.272
	Low BR	2.186	0.353	0.579
Non- Catalytic-1	High BR	7.82	0.132	0.142
	Low BR	19.475	0.171	0.052
Non-Catalytic-3	High BR	11.641	0.397	0.451
	Low BR	8.082	0.331	0.326
		Averages	0.290	0.304
		Standard Deviations	0.110	0.194
		Average + 1 SD	0.400	0.498

Figs. 8 - 13 are plots of PM emissions versus dry burn rate for Catalytic 1, Non-Cat 1 and Non-Cat 3 proficiency round-robin test stoves. As with the ASTM E691 determination of precision and standard deviation, statistically validated outliers have been excluded. The degree of variability for the same stove tested in the same laboratory during the same test year and during different years can easily be observed with even a qualitative assessment. Note: The lab codes are consistent from year to year and stove to stove. For example, Lab D is the same test lab for all models tested but did not participate in the round robin every year it was conducted.

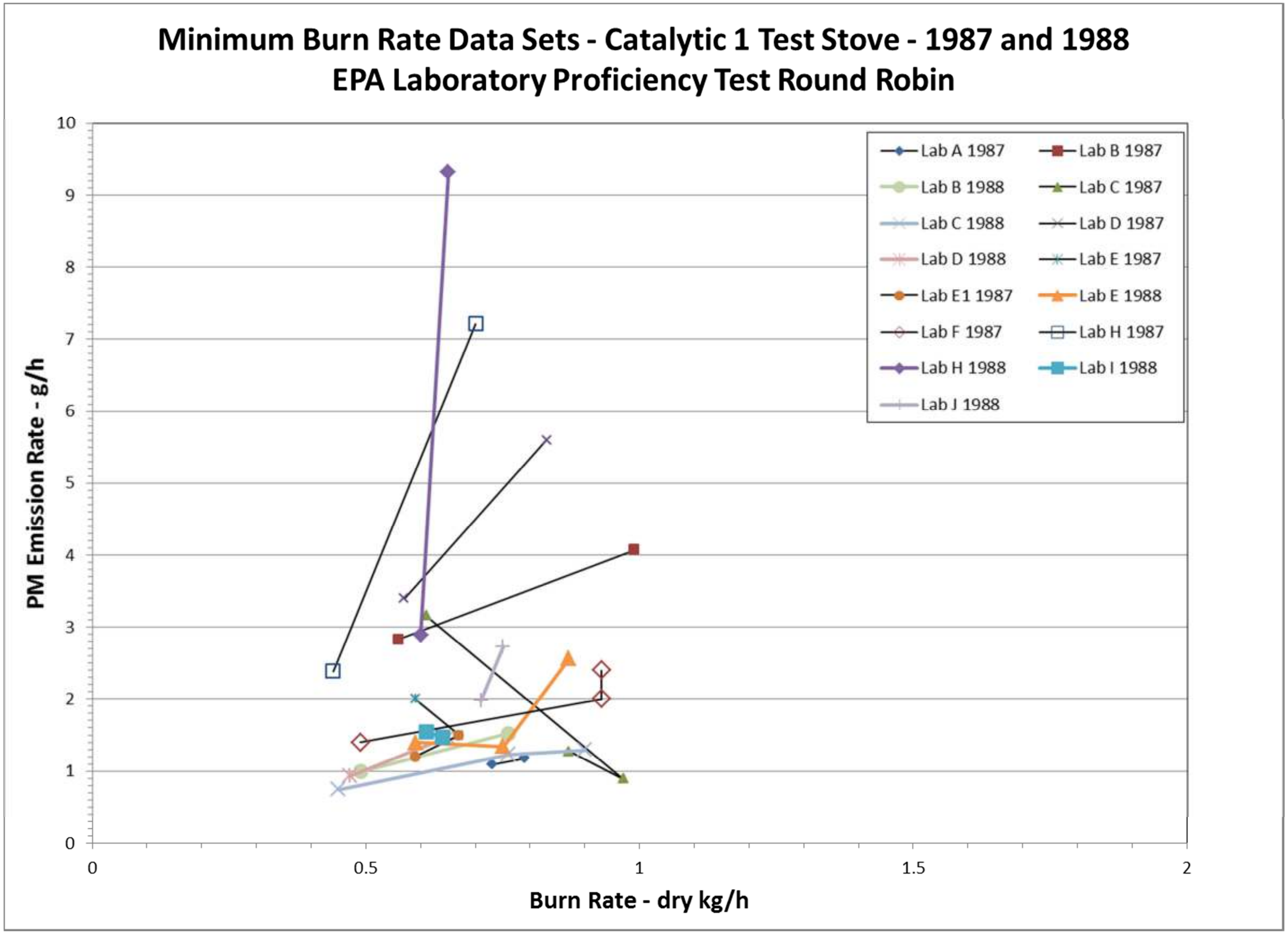
Minimum Burn Rate Data Sets - Catalytic 1 Test Stove - 1987 and 1988
EPA Laboratory Proficiency Test Round Robin
PM Emission Rate - g/h
Burn Rate - dry kg/h
0
1
2
3
4
5
6
7
8
9
10
0
0.5
1
1.5
2
Lab A 1987
Lab B 1987
Lab B 1988
Lab C 1987
Lab C 1988
Lab D 1987
Lab D 1988
Lab E 1987
Lab E1 1987
Lab E 1988
Lab F 1987
Lab H 1987
Lab H 1988
Lab I 1988
Lab J 1988

Fig. 8

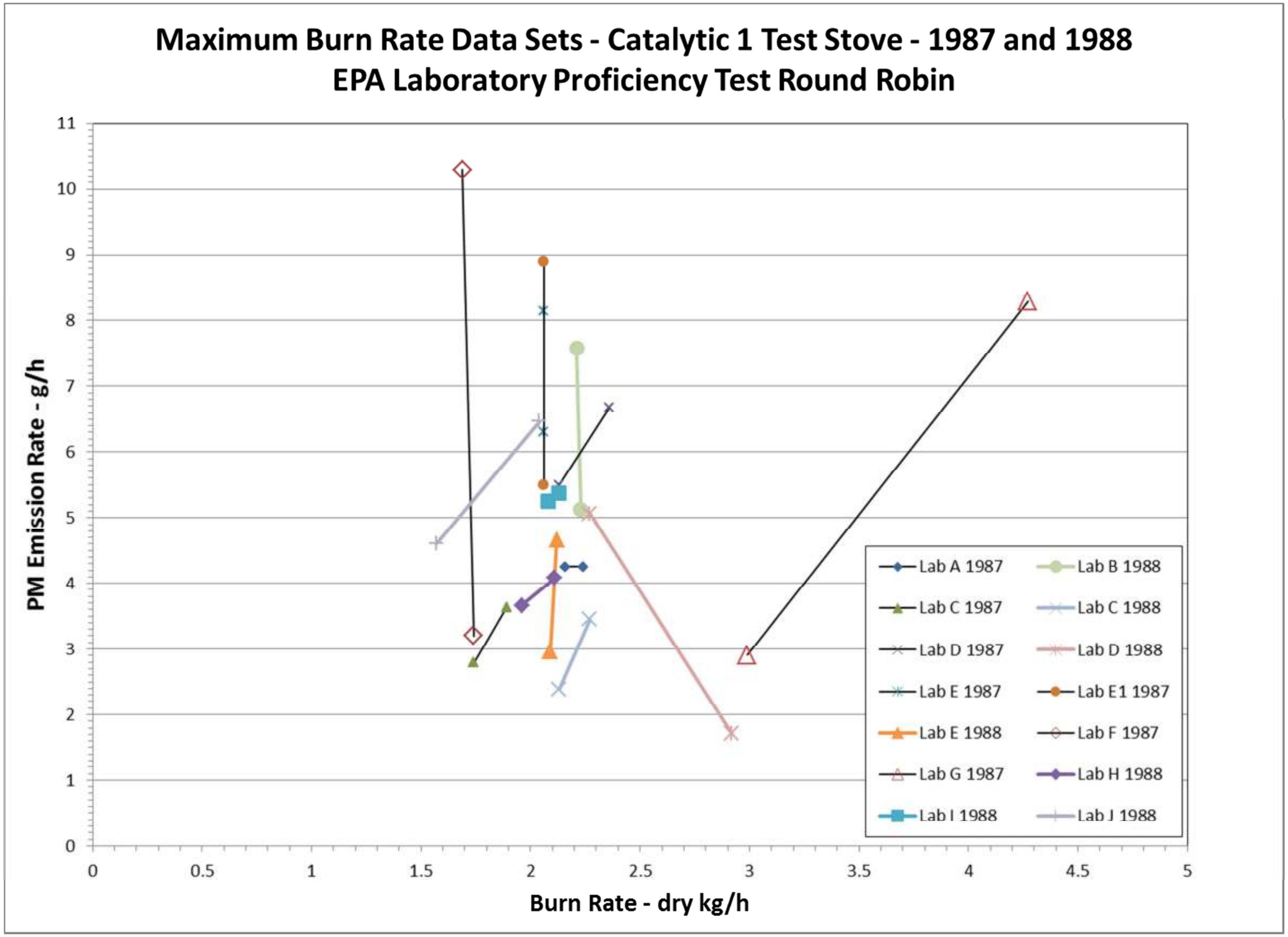
Maximum Burn Rate Data Sets - Catalytic 1 Test Stove - 1987 and 1988
EPA Laboratory Proficiency Test Round Robin
PM Emission Rate - g/h
Burn Rate - dry kg/h
0
1
2
3
4
5
6
7
8
9
10
11
0.5
1.5
2.5
3.5
4.5
Lab A 1987
Lab B 1988
Lab C 1987
Lab C 1988
Lab D 1987
Lab D 1988
Lab E 1987
Lab E1 1987
Lab E 1988
Lab F 1987
Lab G 1987
Lab H 1988
Lab I 1988
Lab J 1988

Fig. 9

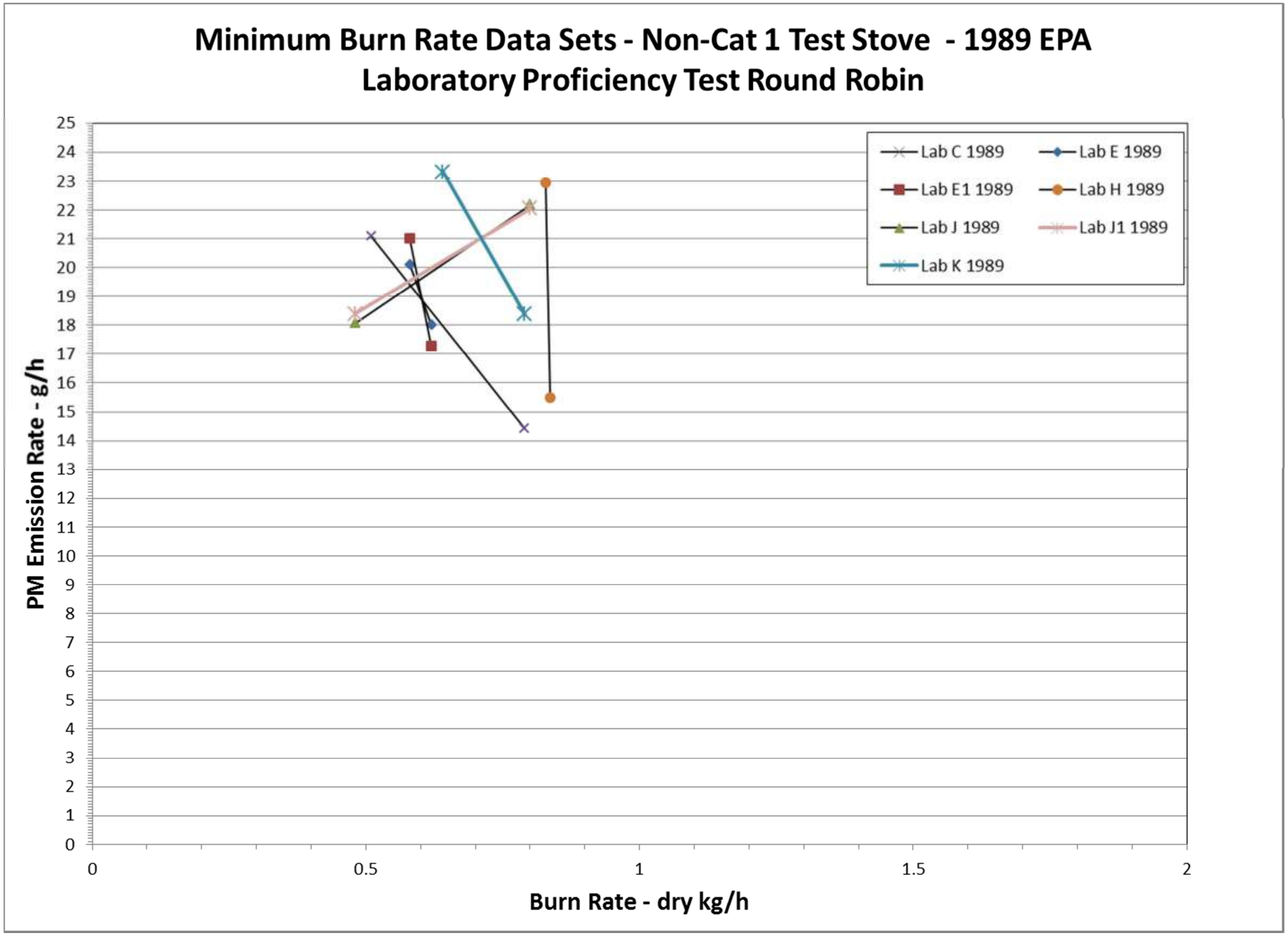
Minimum Burn Rate Data Sets - Non-Cat 1 Test Stove - 1989 EPA Laboratory Proficiency Test Round Robin
PM Emission Rate - g/h
Burn Rate - dry kg/h
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
0
0.5
1
1.5
2
Lab C 1989
Lab E 1989
Lab E1 1989
Lab H 1989
Lab J 1989
Lab J1 1989
Lab K 1989

Fig. 10

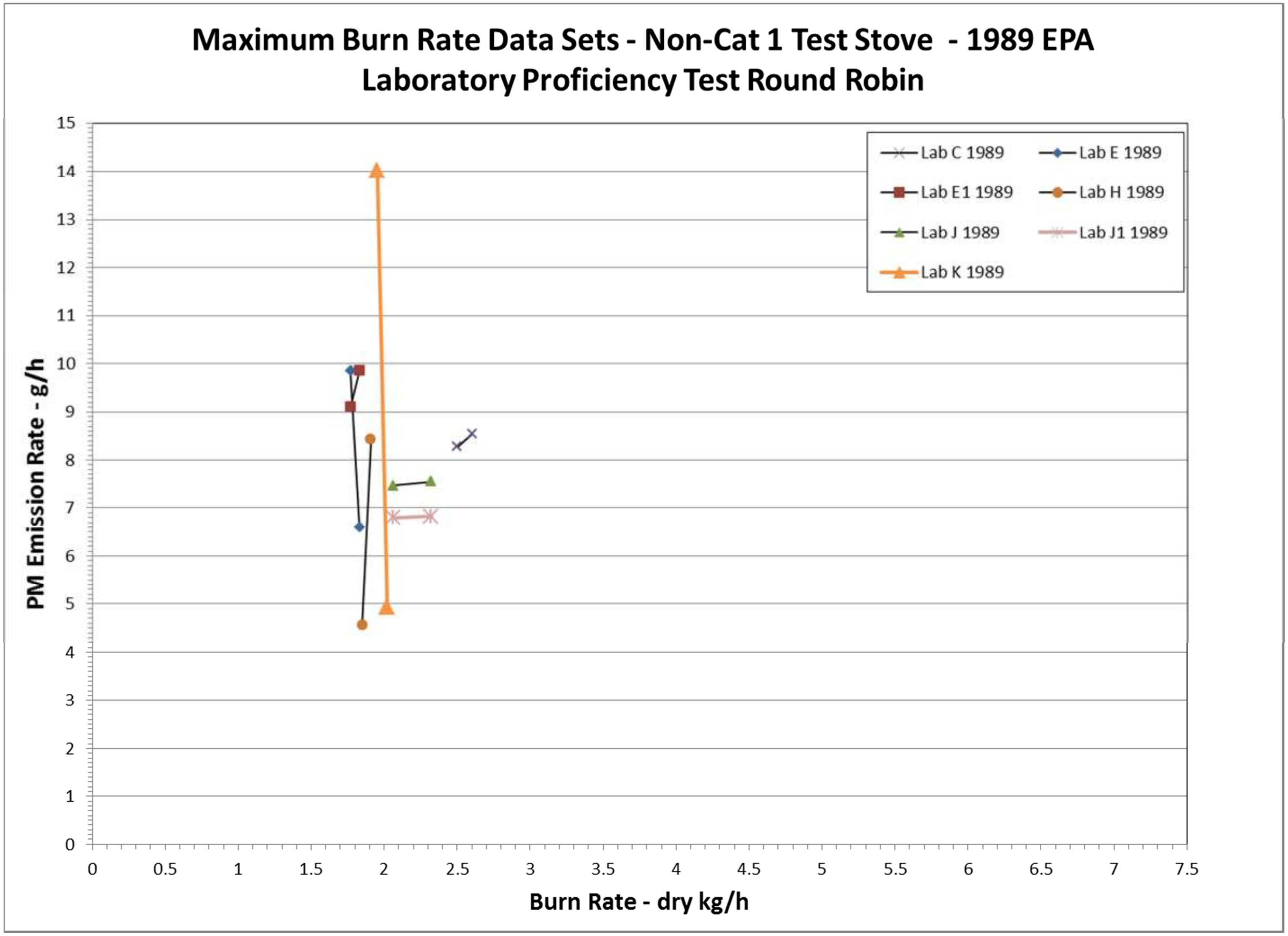
Maximum Burn Rate Data Sets - Non-Cat 1 Test Stove - 1989 EPA Laboratory Proficiency Test Round Robin
PM Emission Rate - g/h
Burn Rate - dry kg/h
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
Lab C 1989
Lab E 1989
Lab E1 1989
Lab H 1989
Lab J 1989
Lab J1 1989
Lab K 1989

Fig. 11

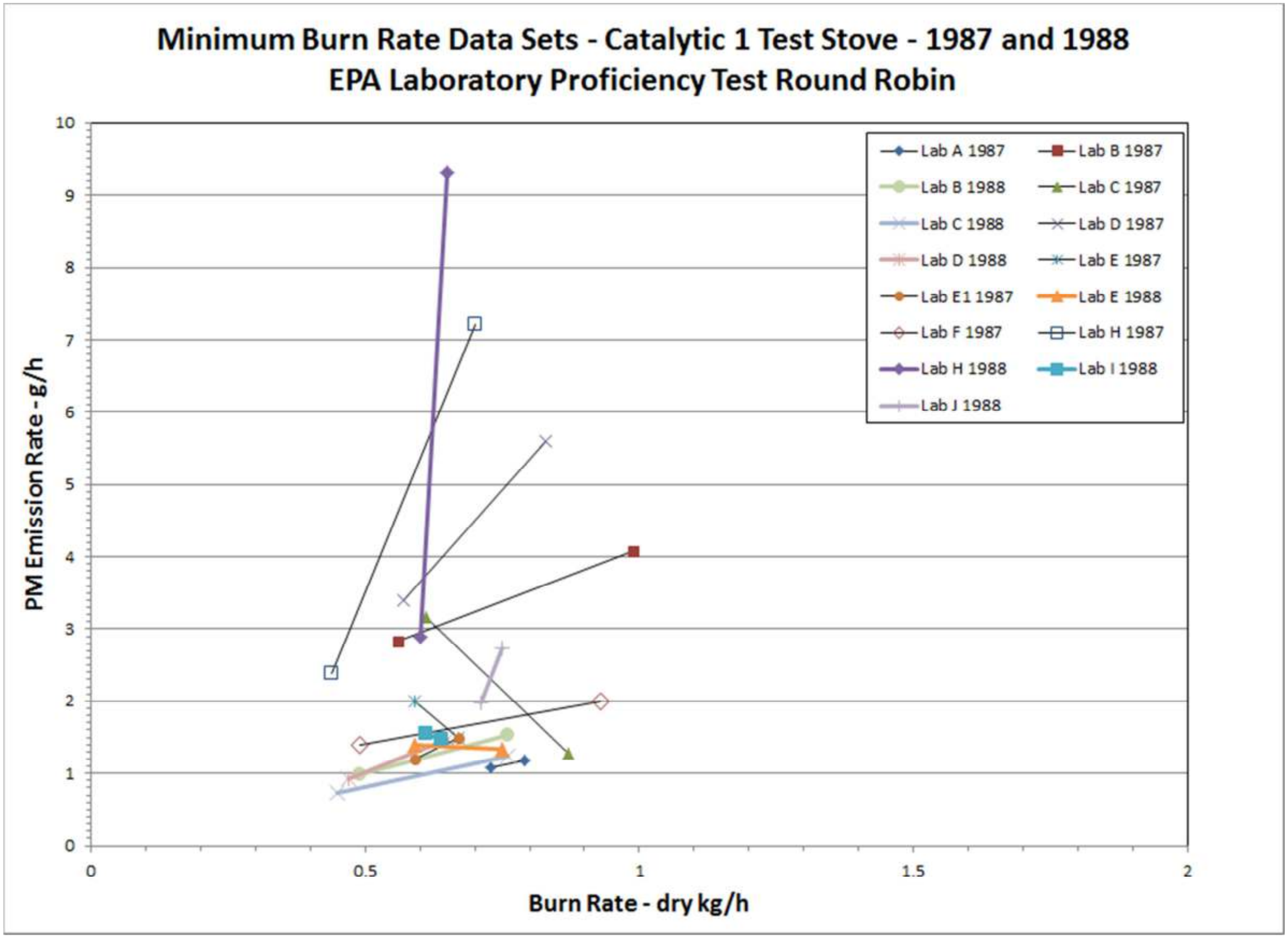
Minimum Burn Rate Data Sets - Catalytic 1 Test Stove - 1987 and 1988
EPA Laboratory Proficiency Test Round Robin
PM Emission Rate - g/h
Burn Rate - dry kg/h
0
1
2
3
4
5
6
7
8
9
10
0
0.5
1
1.5
2
Lab A 1987
Lab B 1987
Lab B 1988
Lab C 1987
Lab C 1988
Lab D 1987
Lab D 1988
Lab E 1987
Lab E1 1987
Lab E 1988
Lab F 1987
Lab H 1987
Lab H 1988
Lab I 1988
Lab J 1988

Fig. 12

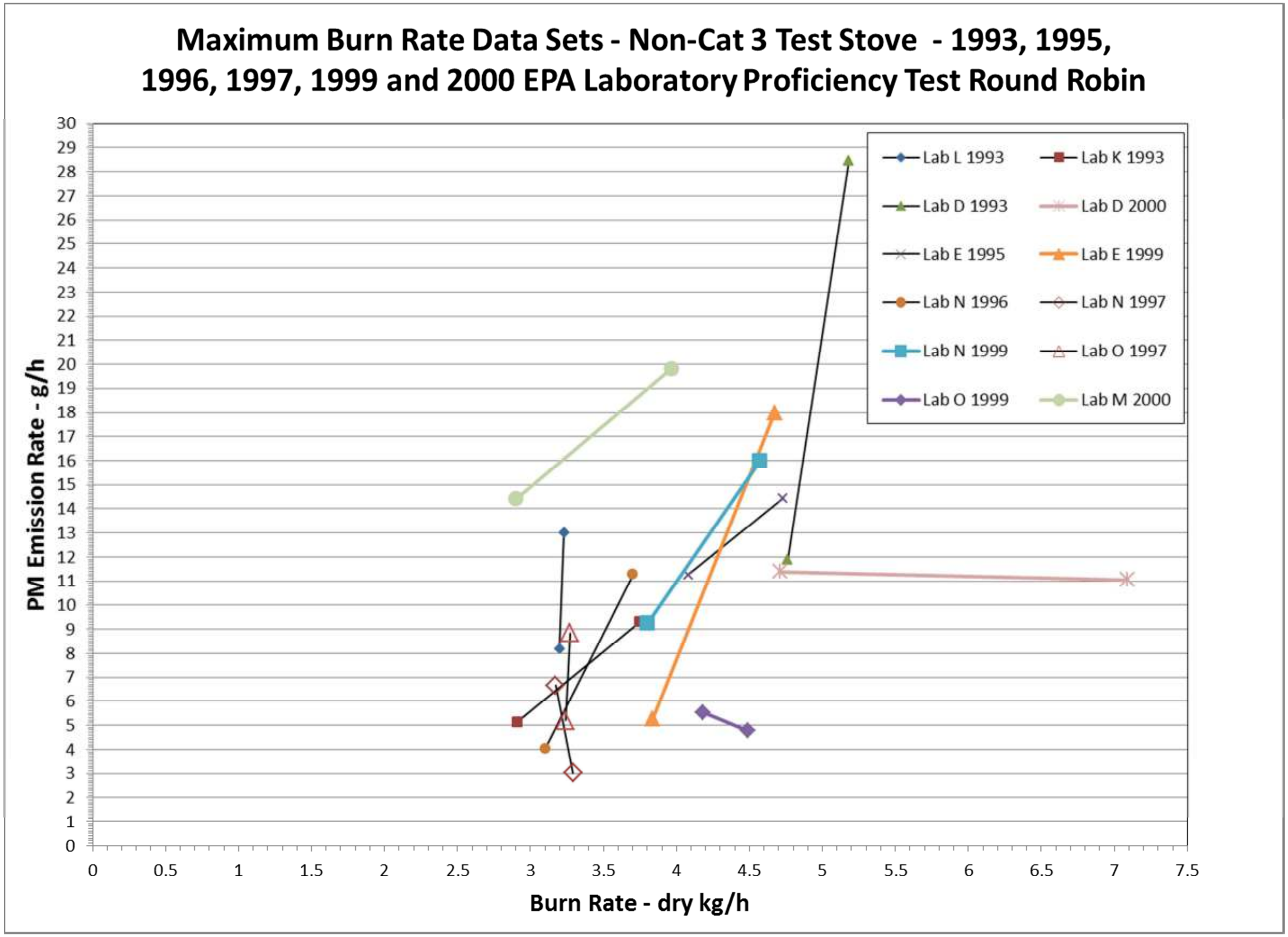
Maximum Burn Rate Data Sets - Non-Cat 3 Test Stove - 1993, 1995, 1996, 1997, 1999 and 2000 EPA Laboratory Proficiency Test Round Robin
PM Emission Rate - g/h
Burn Rate - dry kg/h
Lab L 1993
Lab K 1993
Lab D 1993
Lab D 2000
Lab E 1995
Lab E 1999
Lab N 1996
Lab N 1997
Lab N 1999
Lab O 1997
Lab O 1999
Lab M 2000

Fig. 13

Statistical Modeling

The quantitative and qualitative analyses of the enhanced certified stove database and the EPA laboratory proficiency test data are telling evidence of severe problems with the proposed new compliance algorithm, but still don't adequately capture the impacts of the new compliance algorithm since the certification database only includes one test run in each burn rate category for each certified model and the proficiency test data is limited and includes only two runs in each category. We need to understand the impact of conducting three test runs in the worst case burn category as well as the issues presented by the single run.

To overcome the shortcomings in the available data, a statistical simulation methodology can be employed. This type of analyses is commonly referred to as a Monte Carlo analyses or simulation. This analysis tool has proven to be a very useful method to analyze the behavior of all kinds of complex systems where the number of variables and uncertainties make direct mathematical models too cumbersome or impossible to apply. By inputting key parameters into the simulation software, a log-normal[21] distribution of possible outcomes can be modeled. One can then generate a population of 100's or 1000's of possible results that fall within the distribution. By evaluating a large number of potential outcomes based on that distribution, it is possible to mimic the expected results of doing the actual physical tests. The analysis can be further strengthened by running sensitivity cases where the inputs are varied to cover their possible ranges.

Probabilistic analyses using Monte Carlo methods are widely used by EPA in making regulatory decisions. The dietary risk assessments done for the registration and re-registration of pesticides is a good example of the use of Monte Carlo[22]. The two major variables for dietary exposure are the amount of a food item that people eat and the residue level of pesticide on that food item. EPA establishes distributions for food consumption and residue levels and uses Monte Carlo to estimate the dietary exposures in the population. The Monte Carlo method allows EPA to avoid simply assuming that the person with the highest consumption of a food item also consumes the item with the highest residue. The method also allows EPA to estimate an upper-bound exposure (the 99.9th percentile) for risk assessment so that they can apply a "reasonable certainty of no harm" standard.

EPA also uses Monte Carlo to make decisions about remediating hazardous waste sites under Superfund[23]. In this case, EPA uses Monte Carlo to account for varying levels of a contaminant across a site and differing pathways of exposure to the contaminant (e.g., quantities of incidental soil ingestion or consumption of contaminated water).

[21] A log- normal distribution was considered a better choice than a normal distribution since values in the log-normal distribution can't go below zero.
[22] General principles for performing aggregate exposure and risk assessments. U.S. Environmental Protection Agency Office of Pesticide Programs.
[23] Risk assessment guidance for Superfund, Volume II – Part A, process for conducting probabilistic risk assessment. U.S. Environmental Protection Agency, EPA 540-R-02-002.

In other words, EPA uses Monte Carlo methods for some of its highest profile scientific assessments.

In the case of the proposed NSPS, the Monte Carlo simulation and analysis is used in a somewhat different manner. Here, the Monte Carlo simulation is employed to model the probability of achieving compliance with EPA's proposed new compliance determination methodology which includes both the sampling plan (compliance algorithm) and the acceptance criteria (passing grades). Single and three-run sample sets are drawn at random from a large population of outcomes with a specific probability distribution that is based on supportable inputs including mean emissions levels and coefficients of variation about those means for given hypothetical stove models. The analysis is further strengthened by running sensitivity cases where the inputs are varied to cover their possible ranges. Analyses were also conducted using log-normal distributions.

If we assume that there is a true emission rate for a stove, there are two separate reasons why measurement data could deviate from the true emission rate. First, the ASTM E691 round robin analysis[24] shows that there is a significant inter-laboratory variability, meaning that different laboratories have measured significantly different results for the same stove. Thus, the selection of the laboratory to perform the tests will affect the result. Second, there is also significant intra-laboratory variability, meaning that repeated runs at the same laboratory may generate significantly different results. Ideally, a stove with a true emission rate in compliance with the standard will pass the emission tests, but because of the large inter- and intra-laboratory variability, this is not necessarily the case.

To model the effect of inter- and intra-laboratory variability, a two-stage Monte Carlo analysis was conducted. First, a log-normal distribution is constructed with the assumed true mean for a stove and the coefficient of variance (CV) for the inter-laboratory variability. A log-normal distribution is chosen to constrain emission values to be greater than zero (it is physically impossible to have a negative emission rate) and to model the possibly long tail of the right side of the distribution. The mean and CV are converted into geometric means and standard deviations using standard formulas to parameterize the log-normal distribution. For a given laboratory, a mean value is simulated from the distribution. In the second stage of the Monte Carlo, a new log-normal distribution is constructed using the mean value from the first stage and the CV for intra-laboratory variability. Using this new distribution, emission rates from four test runs are simulated and the results are tested against the compliance standard.

For the purposes of evaluating the impacts of possible emission limits and the new compliance algorithm in the EPA NSPS proposal, several inputs are needed for the Monte Carlo simulation.

1. Emission limit (or passing grade). EPA has proposed a single emission limit (4.5 g/h for Non-Cat and Catalytic models) for Step 1 in the revised NSPS and will maintain the

[24]Op. Cit., Curkeet

current burn rate weighted average approach to determine compliance. EPA has proposed lowering the limit to 2.5 or 1.3 g/h for Step 2, and ultimately to 1.3 g/h if a three step approach is adopted, using a new compliance algorithm as previously described in this paper.

2. <u>The cap for the emission level for any individual test run.</u> The proposal does not include individual test run caps as part of the proposed Step 2/3 compliance algorithm. However, EPA would be setting a defacto cap at the level of the standard for the screening run that is not the worst case run and/or for the three worst case runs if all must meet the level of the standard.
3. <u>Emission data treatment or sampling plan.</u> The worst case condition is first determined by the results of two screening runs (Category 1 and Category 4 Burn Rates). Then, two additional runs are conducted in the category with the worst emissions. The proposal does not specify the way compliance will be determined. Options include:
   Option A. The single screening run AND the average of the worst case runs must meet the limit or,
   Option B. All test runs required to be used to determine compliance must meet the level of the standard.

This means that there are different outcomes that could occur depending on the final requirements.

- i. Both the single run and the three worst case runs are at or below the specified limit –PASS Option A or B.
- ii. Both the single run and the average of the worst case runs are at or below the limit – PASS Option A, May FAIL Option B.
- iii. Both screening runs are above the limit – FAIL Option A or B.
- iv. One screening run is above the limit – FAIL Option B, Decision to continue with two more worst case runs for Option A.
- v. The results from one burn rate category comply and the results from the other does not – FAIL Option A or B.

Assessing these various outcomes requires determining the number of failing values and comparing those to all modeled values for each criteria to determine the percent failure rate.

These failure criteria are as follows:

1. Run 1 exceeds the emission limit
2. Run 2 exceeds the emission limit
3. Run 1 or Run 2 exceeds the emission limit
4. The worst-cast 3-run average exceeds the emission limit
5. Run 1 and the worst-case 3-run both exceed the emission limit
6. At least one of the worst-case 3 runs exceeds the emission limit

4. The true mean emission value for a given stove model. This is the mean value that should be expected if the hypothetical test model was run multiple times under the same test conditions.

5. The coefficients of variation (CV) that one should expect about the true mean of the emission performance of the example stove. This relates to the actual precision of the test methods. Coefficients of variation for Category 1 and Category 4 test results have been estimated using the data from the EPA accredited laboratory proficiency test round-robin. That data can be found in the Curkeet-Ferguson test method variability paper.[25] CV values for both the intra-lab and inter-lab variability where determined. See Appendix A for the determination of CV values.

6. Provisions to Address Additional Test Runs[26]. Although not addressed expressly in the preamble, the EPA NSPS proposal can be interpreted to envision that the final rule will include a provision to address additional test runs that is similar to the "outlier relief" provision in the current NSPS, since provisions for handling additional test runs are included in the test methods in the proposed rule.[27] Currently, the results from all test runs must be reported but only two-thirds of any test runs in a given burn rate category must be used for determining compliance to the standards. The impacts of this provision as applied to the proposed new compliance algorithm can be modeled but that requires certain assumptions about what EPA will allow. We have modeled two scenarios to show the impacts. In the first case, we modeled the situation where a fourth test run replaces the worst of the initial three runs when the initial three run average exceeds the emission limit. This means that 75% (3/4) of the available data is used to determine the average. In the second case, two additional test runs replace the worst of the initial three runs and results in a four run average. This means 80% (4/5) of available data is used to determine the average. This is consistent with the current NSPS concept where a suspected "outlier" may be averaged with an additional test run or can be replaced by two additional test runs.

Procedure

The initial simulation generates a set 1000 PM emission values for each of the two screening test runs and for each of the additional two worst cast test runs. The values are randomly selected from an effectively infinite population of possible emission values that could occur within the two-stage log-normal distribution with the specified mean and coefficients of variations. The

[25] Op. Cit., Curkeet

[26] EPA's provision to address additional test runs for compliance determination is not an "outlier" provision in any statistical sense. A high emission test value may simply be a result that can be anticipated to occur with normally (or log-normally) distributed data. When additional test runs are conducted to average or replace a high emission test result, manufacturers are simply hoping to get a passing result. However, the term "outlier" is used here to mean the situation when additional test run(s) are conducted to try to overcome a poor test run result.

[27] EPA Method 28WHH, Sect. 12.6; EPA M28WHH-PTS, Sect. 12.6; ASTM E2780-10, Sect. 9.5.13.

generated values are then evaluated per the six criteria described in Section 3 above. Compliance Option A was the primary focus of the analyses since it relies on the average of the three worst case emission runs which it is felt was EPA's real intent in the proposal despite a lack of specificity needed to be absolutely certain

Compliance Option B (where all three worst case test runs must be below the emission limit was also considered. Since this option represents an even more stringent requirement than the three-run average option, modeling was limited to that needed to confirm the trend.

For the extra test run ("outlier relief") simulation, two additional test runs were added to the simulation. The average of the values for Runs 2, 3 and 4 the original group was then examined. If the three run average for a particular group was below the applicable emission limit no further action was taken. If the three run average for any group was above the applicable limit for any group, the two "outlier" options were then employed. In the first case, the emission value from the first additional simulated test run was substituted for the highest of the three values in the three run average group, a new three run average was calculated and the results compared to the applicable emission limit. In the second case, the first and second additional emission values were substituted for the worst value of the three run average group a new four run average is calculated and the results compared to the applicable emission limit. A sample of the simulation including inputs and results is provided in Appendix B.

There are several ways to illuminate the implications of the new compliance algorithm. One option that was employed was to look at what the value of the true mean emission performance of a given model would have to be relative to the proposed emission limit in order that the manufacturer would have a 95% confidence of meeting the emission limit when the model undergoes for certification testing. The needed emission performance level can be expressed as a percentage of the emission limit and therefore be universally applied to all product categories and all emission limits even though the modeling is based on woodstoves. Individual examples are also are also modeled based on the proposed emission limits for woodstoves.

As with all testing procedures, the various solid-fuel heater emission test methods each have a lower threshold below which there is no ability to reliably discriminate differences between emission test results. This is the case when the CV values used result in predicted standard deviations at low emission rates that are well below the test method discrimination threshold. For example, we would be skeptical of the ±0.3 g/h standard deviation predicted by using an intra- or inter-lab CV value of 30% for a model with an assumed true mean emission rate of 1.0 g/h at a given burn rate. There is simply no evidence to support that this level of precision is achievable under any circumstances. To the contrary, there is significant evidence to dispute this possible level of precision. Additional simulations have been conducted using the mean CVs plus one standard deviation above the means. This shows the impact of higher variability at low emission rates. However, it is likely that even this adjustment to the CV values doesn't avoid conflict with the discrimination threshold for the method since the resultant standard deviations are still lower than have ever been demonstrated, and therefore, still results in under-estimations of the risk of failing compliance with Step 2/3 emission limits based on the proposed algorithm.

It also must be noted that all of these values are based on crib fuel testing, and EPA has expressed an intention to require cordwood for testing all stick-fired product categories including woodstoves for determining compliance with Step 2/3 emission limits. Although it is reasonable to expect that laboratory testing with cordwood as the test fuel will improve the correlation between laboratory testing results and real-world performance, there is simply no data to inform the impacts on emission outcomes or test method precision of changing test fuel from cribs to cordwood. Cordwood emission performance using standardized test methods is generally unknown for EPA certified stove models that have been designed and tested using crib fuel for the past 25 years. And efforts are still underway to create a new cordwood test method for woodstoves that better reflects homeowner use patterns, so any cordwood data that currently exists would be irrelevant relative to any final cordwood methods, in any event. Moreover, cordwood test method precision can't be determined until any new test method is finalized and even then, only with a properly designed and executed multi-sample, multi-lab test program. It is reasonable to assume that the variability when burning cordwood will not be better than when burning cribs. It is therefore also a reasonable assumption that using the test method precision, standard deviations about a mean emission value and coefficients of variation determined using available crib data provides what can only be considered an absolute best case prediction of what might be expected with cordwood test results with the anticipated new test method. Not whether the risk of failure to achieve compliance is most likely to increase with cordwood testing but by how much the risk increases is the question. And the same or even greater concerns apply to applying this analysis to predict the risks associated with the proposed new compliance algorithm for other appliance categories beyond woodstoves (e.g., pellet stoves, hydronic heaters, and warm air furnaces). The test methods for these categories are new or relatively new, and no comprehensive evaluation of their precision has been performed or is even possible. Beyond that, some of these methods involve additional measurements beyond PM measurements, which are likely to raise their own significant precision issues (e.g., heat output measurements in air plenums for warm air furnaces).

<u>Results</u>

Figures 14 and 15 present the probability of failing one or more of the compliance criteria (average of three worst case runs or an individual test run exceeding the emission limit) for two potential Step 2/3 passing grades, 2.5 and 1.3 g/h. We have looked at a range of true mean emissions and have used the mean intra- and inter-lab CV values per Appendix A.

Fig. 14 shows the impacts of setting the emission limit at 2.5 g/h. For a manufacturer to have a 95% chance of meeting the emission limit for both of the first two screening runs, the true mean emission value for the tested model must be under 1.25 g/h in both the Category 1 and Category 4 burn rates. In other words the true mean must be at about half the emission limit. Since the three-run average mitigates some of the variability, a 95% chance of the three-run average complying can be achieved if the true mean average of the worst cast runs is about 1.5 g/h. To look at this from a different perspective, a stove with true mean emissions of 1.9 g/h (~25%

below the emission limit) in both burn rate categories has almost a 30% chance of failing one of the two screening runs. And even if the true mean of the three worst case runs is 1.9 g/h, the model has more than a 15% chance of failing to achieve compliance. Finally, the upper line in this figure predicts the probability of at least one of the three worst case runs exceeding the emission limit. This relates to possible outcomes if EPA imposes Option B where all test runs must meet the emission requirement. This option obviously increases the risk of failure significantly above Option A.

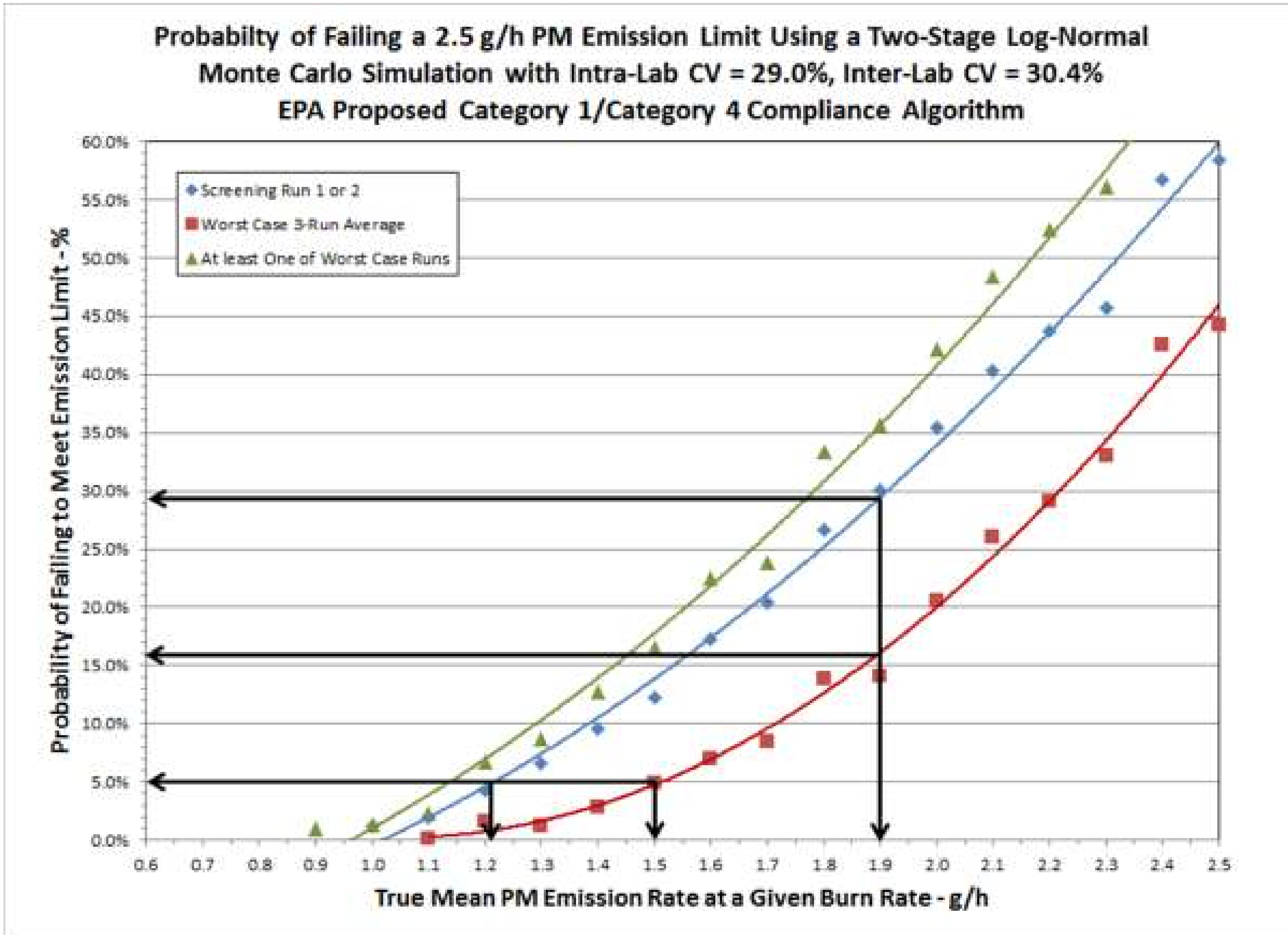
Probabilty of Failing a 2.5 g/h PM Emission Limit Using a Two-Stage Log-Normal
Monte Carlo Simulation with Intra-Lab CV = 29.0%, Inter-Lab CV = 30.4%
EPA Proposed Category 1/Category 4 Compliance Algorithm
Screening Run 1 or 2
Worst Case 3-Run Average
At least One of Worst Case Runs
Probability of Failing to Meet Emission Limit - %
60.0%
55.0%
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3
2.4
2.5
True Mean PM Emission Rate at a Given Burn Rate - g/h

Fig. 14

Fig, 15 shows the impacts of setting the emission limit for woodstoves at 1.3 g/h. The tested model would need to have true mean emission values in both test categories that are less than 50% of the emission limit in both test categories to predict a 95% chance of passing both screening runs. The true mean of the three worst case runs would have to be below 0.8 g/h to have a 95% chance of the three run average meeting 1.3 g/h. A stove model with a true mean emission value of 0.9 g/h(~30% below the emission limit) in both burn rate categories would have a 23% chance of failing one of the screening runs. And, again the upper line in this figure is predictive of the probability of one or more of the worst case runs exceeding 1.3 g/h as part of the Option B consideration.

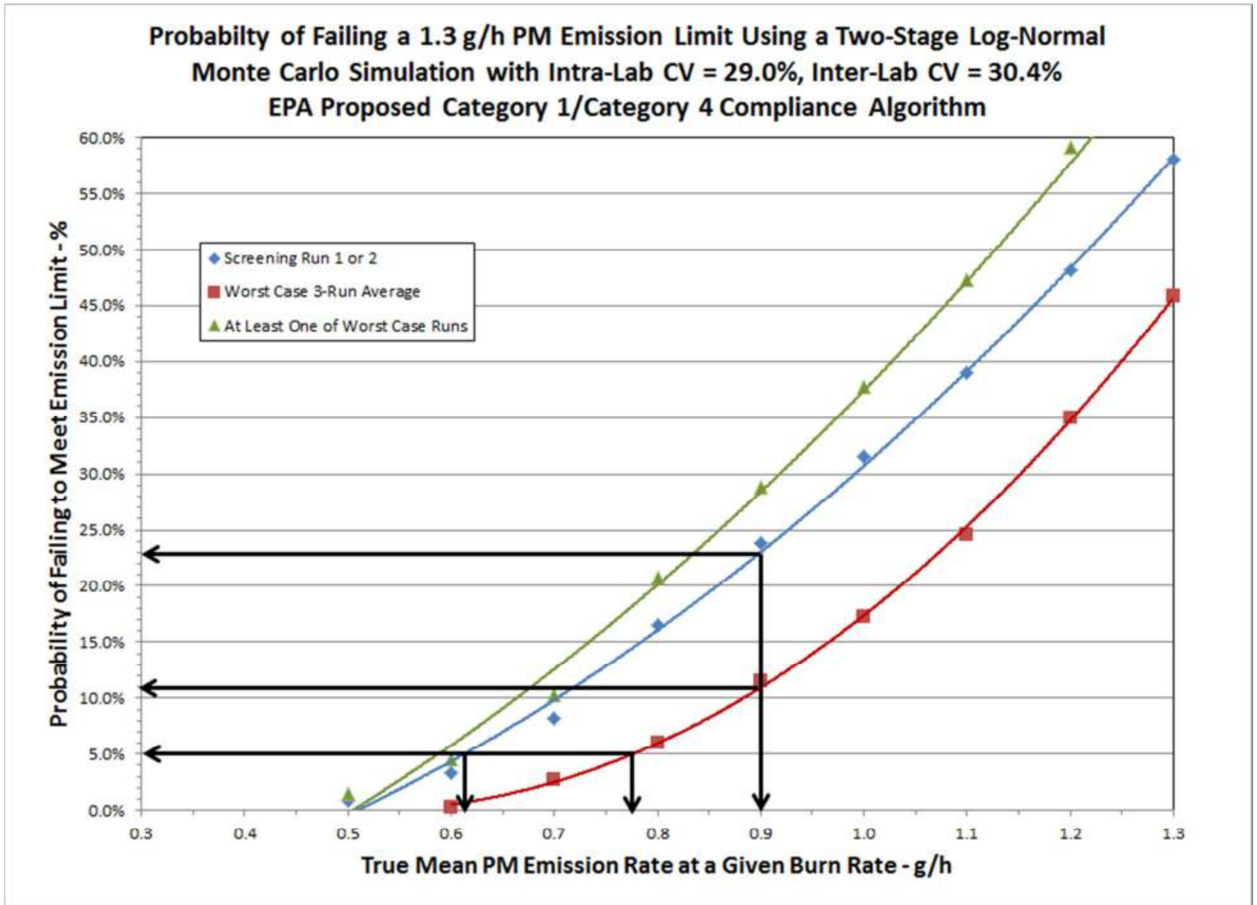
Probabilty of Failing a 1.3 g/h PM Emission Limit Using a Two-Stage Log-Normal Monte Carlo Simulation with Intra-Lab CV = 29.0%, Inter-Lab CV = 30.4%
EPA Proposed Category 1/Category 4 Compliance Algorithm
Screening Run 1 or 2
Worst Case 3-Run Average
At Least One of Worst Case Runs
Probability of Failing to Meet Emission Limit - %
True Mean PM Emission Rate at a Given Burn Rate - g/h
60.0%
55.0%
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3

Fig. 15

Figs. 16 and 17 present the results when the CVs used in the two-stage Monte Carlo simulations are increased by one standard deviation.

Fig. 16 shows the impacts for an emission limit of 2.5 g/h. To achieve a 95% probability of passing the screening tests, the true mean emission rate in both burn rate categories must be about 0.9 g/h. At the 1.2 g/h level, the risk of failing one of the two screening runs is above 11%.

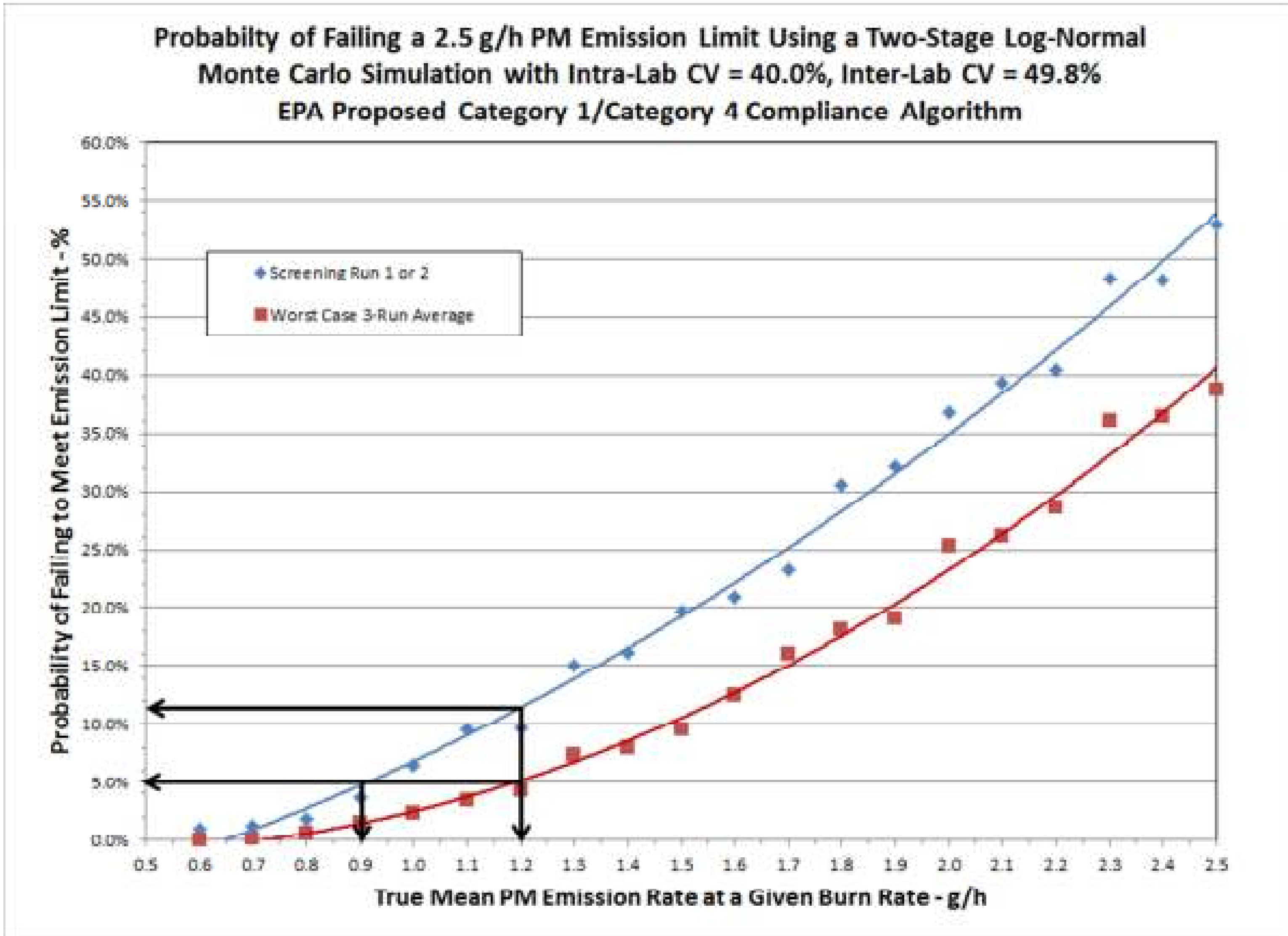
Probabilty of Failing a 2.5 g/h PM Emission Limit Using a Two-Stage Log-Normal
Monte Carlo Simulation with Intra-Lab CV = 40.0%, Inter-Lab CV = 49.8%
EPA Proposed Category 1/Category 4 Compliance Algorithm
Screening Run 1 or 2
Worst Case 3-Run Average
Probability of Failing to Meet Emission Limit - %
True Mean PM Emission Rate at a Given Burn Rate - g/h
60.0%
55.0%
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3
2.4
2.5

Fig. 16

Fig. 17 shows the impacts for an emission limit of 1.3 g/h. To achieve a 95% probability of passing the screening tests, the true mean emission rate in both burn rate categories must be below 0.5 g/h. At the 0.9 g/h level, the risk of failing one of the two screening runs is about 27% while the risk of the three run average failing is 16%.

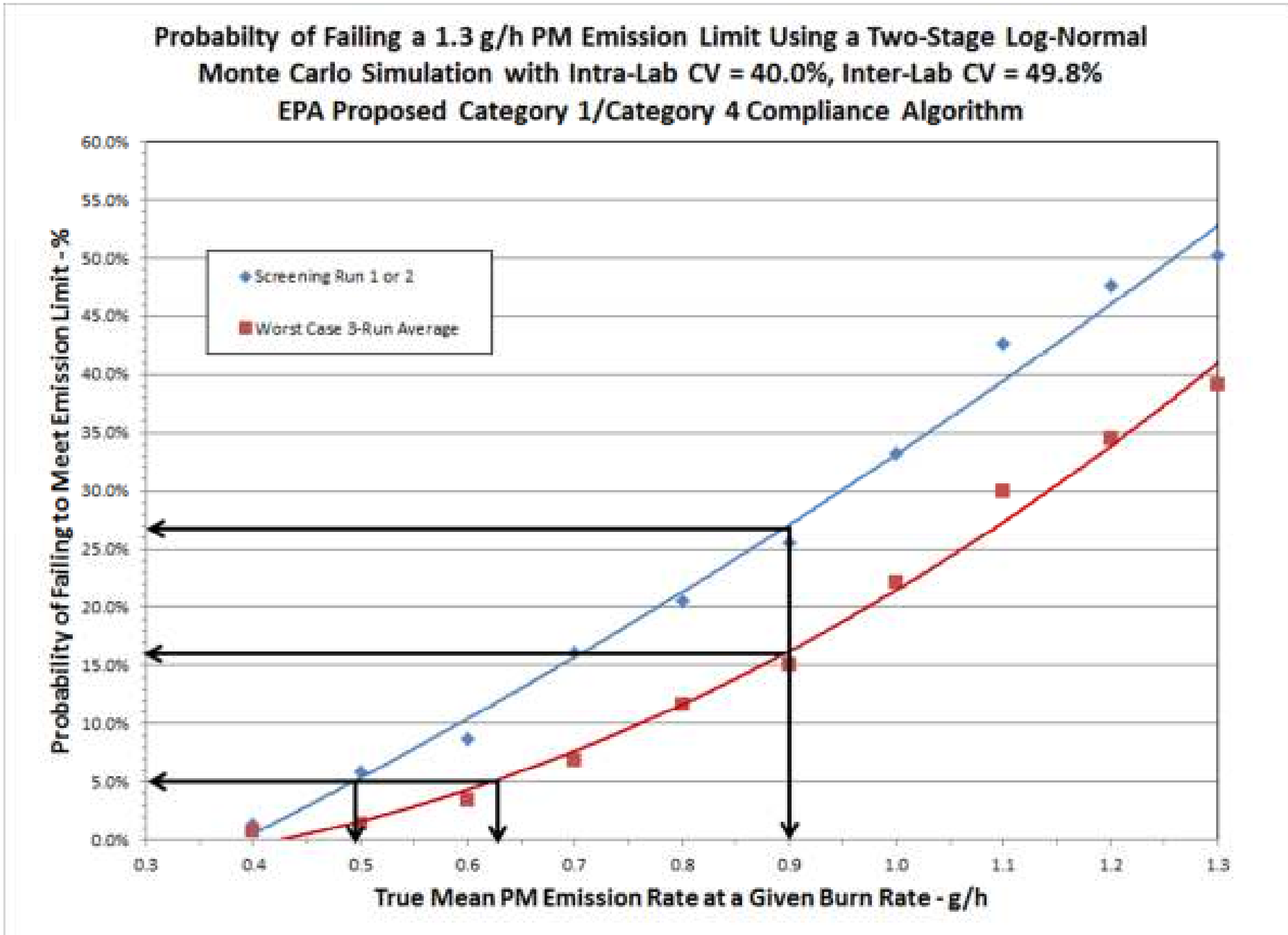
Probabilty of Failing a 1.3 g/h PM Emission Limit Using a Two-Stage Log-Normal
Monte Carlo Simulation with Intra-Lab CV = 40.0%, Inter-Lab CV = 49.8%
EPA Proposed Category 1/Category 4 Compliance Algorithm
Screening Run 1 or 2
Worst Case 3-Run Average
Probability of Failing to Meet Emission Limit - %
60.0%
55.0%
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
True Mean PM Emission Rate at a Given Burn Rate - g/h

Fig. 17

In order to show how the Monte Carlo method can be used to predict the probable outcomes for other emission limits or even for other product categories, Figs. 18 and 19 present the results in terms of the true mean emission performance value as a percentage of the emission limit. Fig. 18 presents the results when the mean CV values are used and Fig. 19 presents the results when the CV values are set at the means plus one standard deviation. If it is assumed that the variability for the test methods for other appliance categories is similar to that for woodstoves, the probability of failure analyses can be applied to hydronic heaters or warm air furnaces to show the levels of true mean emission performance needed to achieve an acceptable probability of meeting the applicable emission limits. Of course, the variability of these other test methods has never been determined so any conclusions drawn must be considered in that light. It is, however, not unreasonable to assume that the other methods do not have better precision values than woodstove testing and therefore any conclusions drawn probably understate the risk by some margin.

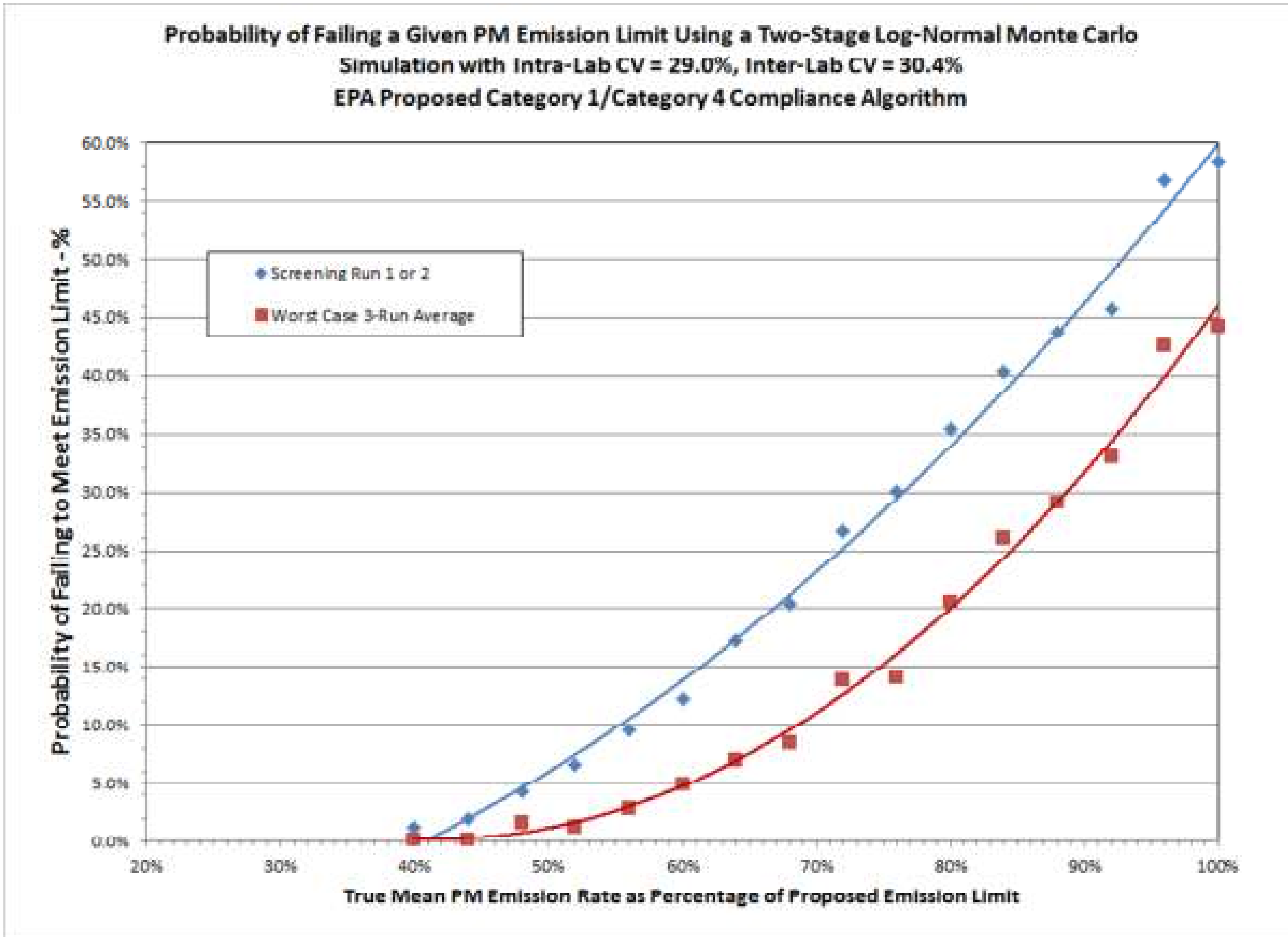
Probability of Failing a Given PM Emission Limit Using a Two-Stage Log-Normal Monte Carlo Simulation with Intra-Lab CV = 29.0%, Inter-Lab CV = 30.4%
EPA Proposed Category 1/Category 4 Compliance Algorithm
Screening Run 1 or 2
Worst Case 3-Run Average
Probability of Failing to Meet Emission Limit - %
True Mean PM Emission Rate as Percentage of Proposed Emission Limit
60.0%
55.0%
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
20%
30%
40%
50%
60%
70%
80%
90%
100%

Fig. 18

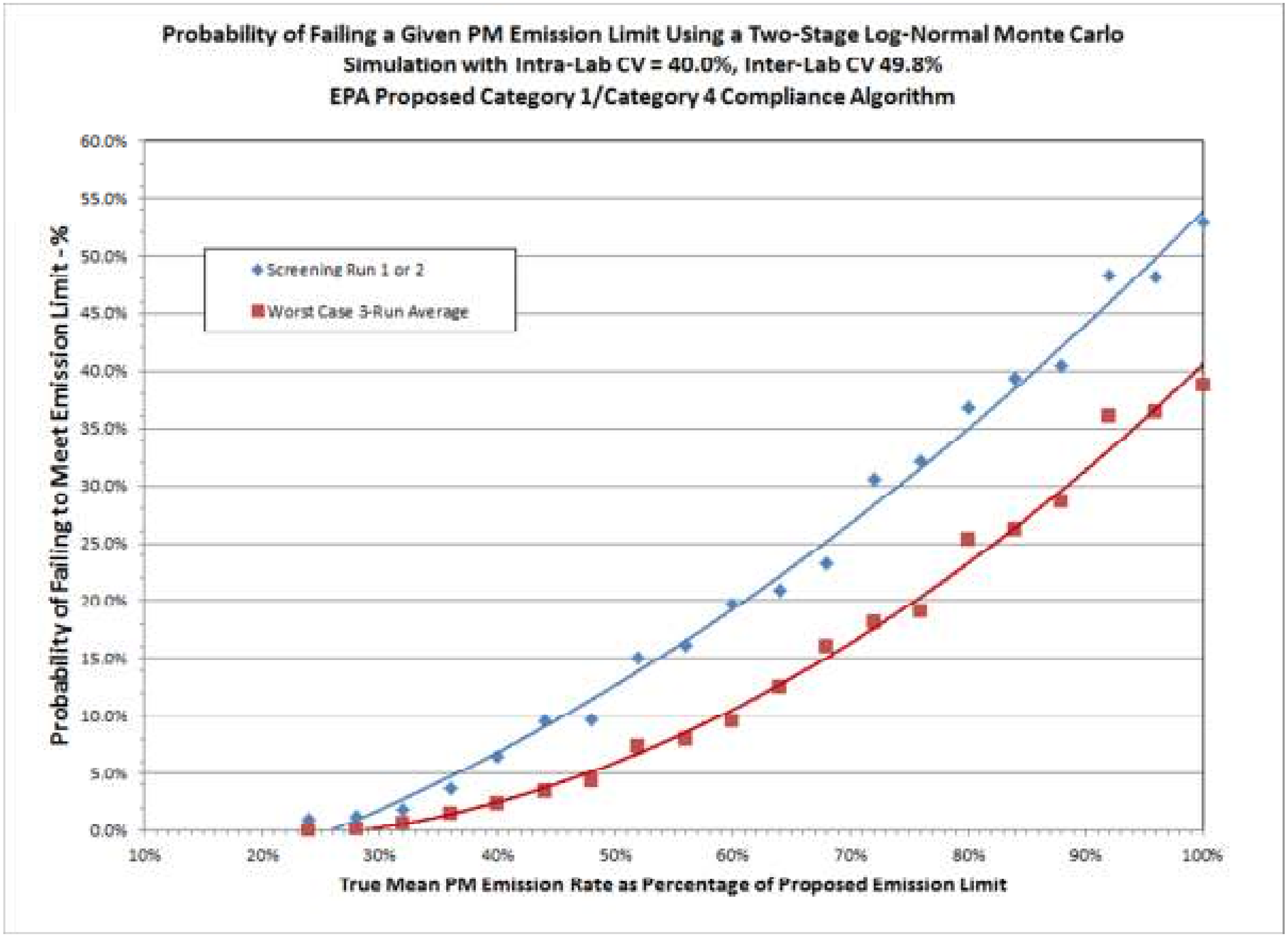
Probability of Failing a Given PM Emission Limit Using a Two-Stage Log-Normal Monte Carlo Simulation with Intra-Lab CV = 40.0%, Inter-Lab CV 49.8%
EPA Proposed Category 1/Category 4 Compliance Algorithm
Probability of Failing to Meet Emission Limit - %
60.0%
55.0%
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
Screening Run 1 or 2
Worst Case 3-Run Average
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
True Mean PM Emission Rate as Percentage of Proposed Emission Limit

Fig. 19

Figs. 20 and 21 present an assessment of the probability of failure when the two additional test run options ("outlier" relief) are applied. In this part of the simulations, the highest value from any failing worst case three run averages is replaced by one or two addition values as previously explained. Fig. 20 represents the results using the mean CV values. Fig. 21 represents the results using the CV values set at the means plus one standard deviation. The figures are presented as true mean emission values expressed as a percentage of the emission limit as a way of showing the trends of substituting one or two additional test runs for the highest value in a failing three run average regardless of the specific emission limit.

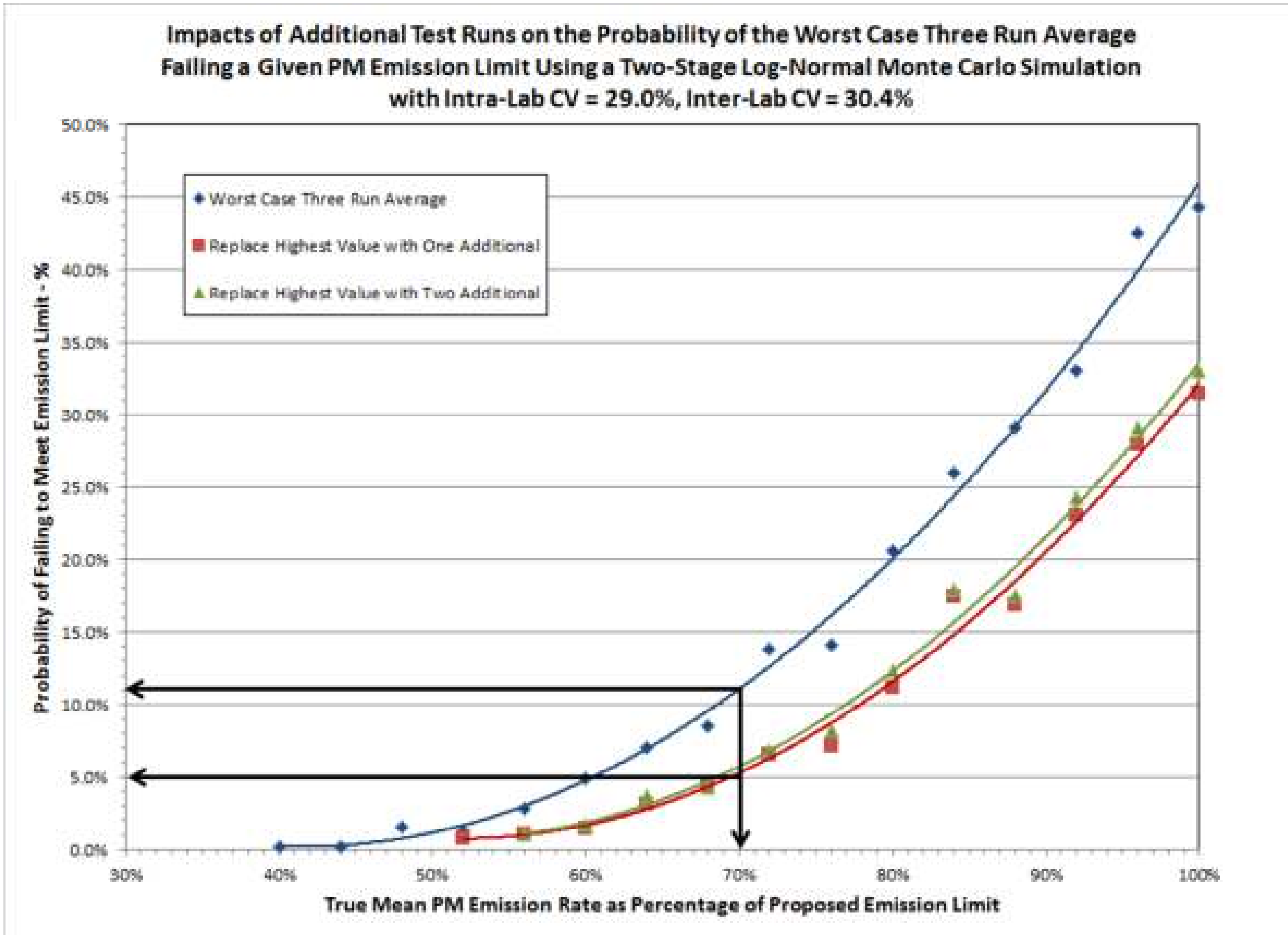
Impacts of Additional Test Runs on the Probability of the Worst Case Three Run Average Failing a Given PM Emission Limit Using a Two-Stage Log-Normal Monte Carlo Simulation with Intra-Lab CV = 29.0%, Inter-Lab CV = 30.4%
Worst Case Three Run Average
Replace Highest Value with One Additional
Replace Highest Value with Two Additional
Probability of Failing to Meet Emission Limit - %
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
30%
40%
50%
60%
70%
80%
90%
100%
True Mean PM Emission Rate as Percentage of Proposed Emission Limit

Fig. 20

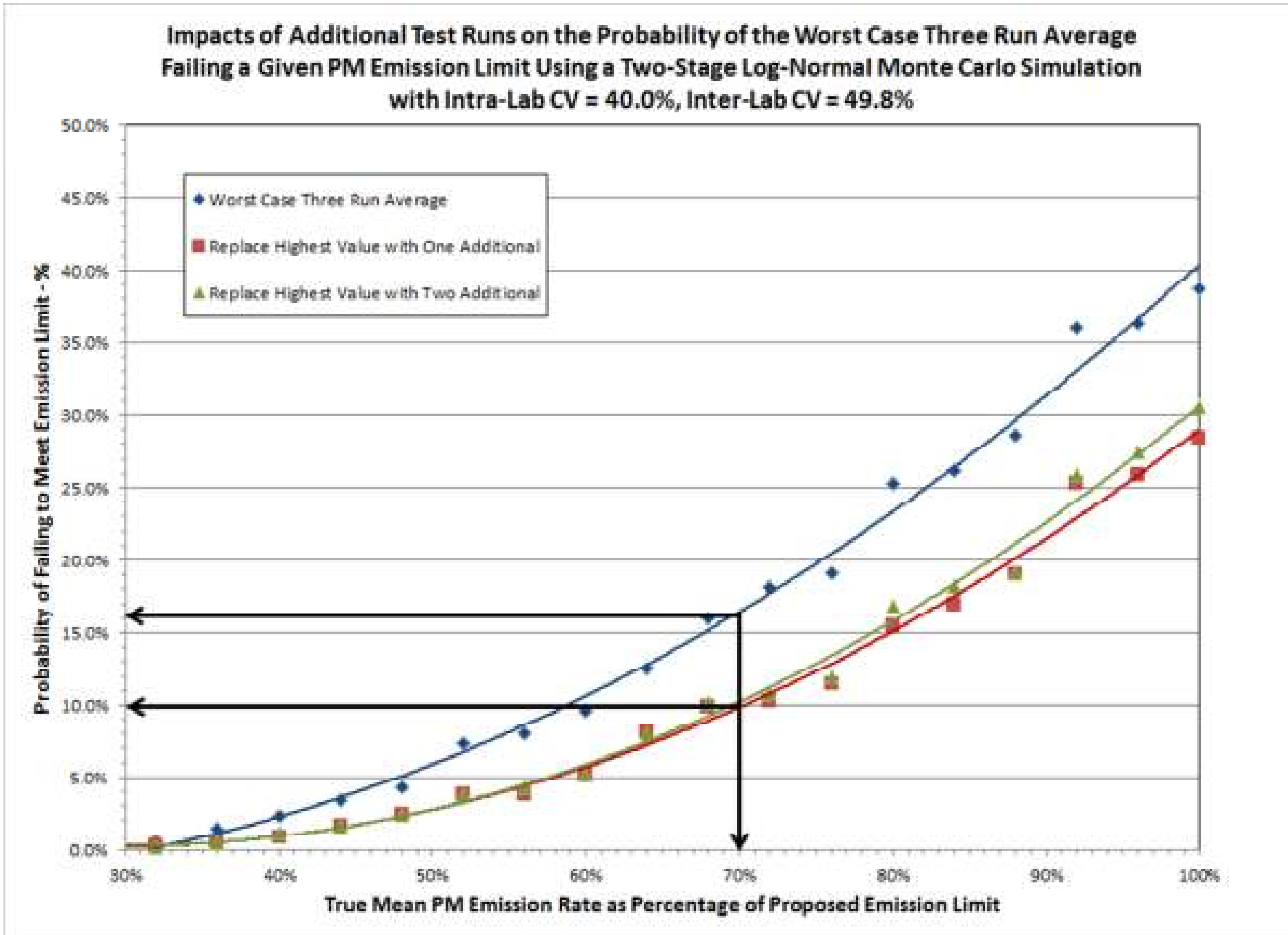
Impacts of Additional Test Runs on the Probability of the Worst Case Three Run Average Failing a Given PM Emission Limit Using a Two-Stage Log-Normal Monte Carlo Simulation with Intra-Lab CV = 40.0%, Inter-Lab CV = 49.8%
Worst Case Three Run Average
Replace Highest Value with One Additional
Replace Highest Value with Two Additional
Probability of Failing to Meet Emission Limit - %
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
30%
40%
50%
60%
70%
80%
90%
100%
True Mean PM Emission Rate as Percentage of Proposed Emission Limit

Fig. 21

In all cases, the probability of failure is reduced using either of the additional run options. The benefit in terms of the magnitude of the improvement by replacing the highest of the three runs in the original average increases as the true mean emission rate increases relative to the emission limit. However, the probability of failing the initial three run average also increases as true mean emission rate increases relative to the emission limit so there is a trade off in that may factor into any decision to add extra test runs. Manufacturers will have to be keenly aware of possible outcomes.

Fig. 20 can be used illustrate the impact. It shows that a stove model with a true worst cast emission value that is at 70% of the emission limit (e.g., ~1.8 g/h versus a 2.5 g/h emission limit) will have an 11% probability of the initial three run average exceeding the emission limit. Replacing the highest emission run with one or two additional test runs improves the probability of failing to ~5% (95% chance of meeting the emission limit), which may indicate that conducting the extra run is a reasonable choice.. However, at true mean emission values above 70% of the emission limit, the improved probability of passing by conducting extra test runs may result in a risk level that still is too high. This could make running additional tests to overcome a high emission test run an unattractive option due to the cost of the additional testing and an unacceptable risk of failing again.

Fig. 21 shows that the when higher test method variability is modeled, the benefit of additional test runs diminishes at true mean emission values that are above 60% of the emission limit. And, interestingly, as with Fig, 20, it can be seen that conducting the second additional test run provides almost no additional improvement in the probability of passing.

<u>Conclusions</u>

The proposed new compliance algorithm where three test runs are conducted in the category with the worst emission performance based on two initial screening test runs (one at the lowest firing rate and the other at the highest) presents a myriad issues if adopted.

With the current data weighting methodology, test run results spanning the full operating range of the stove are used to determine the average emissions. Further, the lower burn rate categories receive most of the weighting based on the typical operating patterns of stove users from in-home studies conducted in the 1980's.

An examination of test run data from currently produced EPA certified models shows that most models have emission performance profiles that are not flat, with better performance most frequently focused in the range of primary concern—the lower burn rates. Sacrificing the performance at the highest burn rates has often been the necessary trade-off that stove designers have needed to insure the best performance at the heavily-weighted low burn rates while still meeting maximum heat output expectation from consumers.

The certified stove data also shows the impact of abandoning the current weighting scheme and looking at only the highest and lowest burn rates for achieving compliance. Of the 96 non-cat models in the database, 22 have weighted average emissions ≤ 2.5 g/h, but only 11 models have emission results ≤ 2.5 g/h in both the category 1 and category 4. Catalytic models show a similar trend. If an emission limit of 1.3 g/h is evaluated, only five models (including non-cat and catalytic) have category 1 and category 4 emission values ≤ 1.3 g/h. Not even pellet stoves are immune from the impacts of the proposed compliance algorithm with only one model out of 21 in the database with both burn rate categories ≤ 1.3 g/h. And because the certified stove database includes only one test run in each category, the impacts of running two additional worst case test runs cannot be evaluated at all with existing data.

The EPA Accredited Laboratory Proficiency Round Robin test data provides the only available data to evaluate test method precision and to determine the average intra- and inter-laboratory standard deviations. A rigorous analysis of that data shows that the test method precision at the typical 95% confidence level is at best ± 3 g/h when testing is conducted in the same laboratory and at best ± 4.5 g/h when different labs test the same stove. The data also show an average inter-lab standard deviation of ± 1.9 g/h but that value is based on weighted average emissions and not on individual test runs.

However, the proficiency round robin data also allows data in individual burn rate categories to be evaluated for precision. The true standard deviation when evaluating category 1 and category 4 test runs is estimated to be at least ± 2.7 g/h on an intra-laboratory basis. And even a qualitative examination of the Category 1 and Category 4 test data for the same test stoves tested in the same lab and in different labs shows the wide variation of emission results that can be anticipated when more than one run is conducted at the same air setting. Since there is a wide range of emission values as part of the proficiency test data, coefficients of variation have also be determined as a way of helping to normalize the intra- and inter-lab precision values. However, the even the intra- and inter-lab CV values have a wide range, another indication that the test methods are not very precise. And since this data consists exclusively of pairs of data points, even it does not allow the impacts of conducting three test runs in the same burn rate category to be adequately evaluated.

This leads to the necessity of using the probability-based assessment commonly referred to as the Monte Carlo analysis to fully understand the potential impacts of the proposed compliance methodology. The conclusions that can be drawn from the first round of Monte Carlo simulations are quite chilling.

To achieve a 95% confidence level that a stove will comply with the proposed emission limits using the proposed compliance algorithm, the true mean emission performance for the stove model in both Category 1 and Category 4 must be significantly below the emission limit. Stoves with true means of PM emission performance that are well below (25% or more) the proposed Step 2/3 emission limits have very high probabilities of failing, either for the three run average of worst case performance and/or for the individual runs themselves, when a realistic and supportable coefficients of variation are used. Since the screening run that is not part of the three run worst case average must also be at or below the level of the standard, with no allowance for test method precision, the probability of failure is unacceptably high even if the true mean emission performance for a given stove approaches half the level of the proposed standard. In the simplest terms, this would impose a level of risk of failure that would effectively drive most manufacturers out of the market. And, predictions of failure are even higher when using CV values that are only one standard deviation above their means, a not unrealistic possibility considering the variability in the CV values determined from the EPA proficiency test data.

The CV value approach in itself ignores the reality that all the test methods for all heater types have a determination threshold, below which the method simply can't reliably distinguish differences in measured emission values. The CV approach that was employed assumes that the average CV values apply to even the lowest emission values implicated by all the proposed Step2/3 emission limits. For a 2.5 g/h emission limit for woodstove, the CV values used in the Monte Carlo analyses result in calculated standard deviation values less than ± 1.25 g/h. For the 1.3 g/h emission limit, the calculated standard deviations are less than ± 0.7. These values are well below any demonstrated test method precision estimates and it is likely that they may ever be achieved in practical or affordable way, if at all. This again emphasizes that use of CV values and the resultant low predicted standard deviations at low emission rates for purposes of the Monte Carlo analyses conducted for this report is a very conservative approach to estimating the probability of failure. The actual failure rates are likely to be higher.

And test method precision has never been evaluated for hydronic heaters or warm air furnaces and no data exists that is adequate for making those determinations. It must be assumed that the variability that can be expected is on the same level as that for woodstoves and perhaps even higher due to the added complexity of the test methods for these product categories.

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. This is a situation that can only be addressed by first recognizing that it exists and the by addressing the random variability by setting standards that account for that inherent uncertainty.

The ideal case for a compliance algorithm is one that has 100% probability of passing a stove that actually complies and 100% chance failing one that doesn't. Obviously, measurement uncertainty and true product variability make this impossible. We are forced to live with test methods that are not very precise and must take into consideration the impacts of that imprecision when standards are set. EPA typically sets standards at a level that represents the 95% confidence interval or in other words at a level that is intended to insure that 95% of products that should comply with the standard will comply.

Manufacturers most certainly want a 95% or better confidence that truly compliant products will pass, but regulators want a process that assures a 95% chance that non-compliant products will fail. Both cannot be achieved when the precision is poor. The risks associated with both types of error – acceptance of an unqualified product and rejection of a qualified one must be considered. One can argue that the first error type is of small consequence to the environment since the appliance will still have to be far cleaner than previously required. The second error type can be financially devastating to a manufacturer given the large investment involved and the inability to make a return on that investment that a failed test represents. Most manufacturers will conduct substantial R&D testing to satisfy themselves that their designs will have a good chance of passing before spending money on certification tests. It is quite unlikely that any manufacturer would submit a product for a certification test that they know is likely to fail and just hope to get lucky and pass. When a certification test subsequently fails, it is a real possibility that the certification test series was simply a poor representation of the product's true performance due to test method variability.

There has to be a sharing of the risk presented by the test method uncertainty and a balancing of the impacts to the environment (relatively low) versus the financial impacts on the manufacturers (relatively high).

A process that makes the determination of compliance primarily a matter of random chance is of no regulatory value and will impose unwarranted risk on manufacturers as they attempt to certify new products. Passing grades AND the compliance algorithm <u>must</u> account for the precision of the measurement and fueling methods.

Appendix A

Determination of Coefficients of Variations Used in the Monte Carlo Simulations

The EPA accredited laboratory proficiency round-robin test data was used as the basis for the determination of coefficients of variation (CVs) for use in the Monte Carlo simulations used to evaluate the impacts presented by the Category 1/Category 4 compliance algorithm included in the February 2014 EPA NSPS proposal. CV's were used because of the large variation in emission test results during the round robin testing including variations from test stove to test stove, laboratory to laboratory and year to year. For our purposes, the CV values help normalize the variability assessment.

The proficiency round robin test data is limited to woodstove testing but is the only data available that allows conclusions to be reached as to solid-fuel emission testing precision.

The proficiency test data has been previously analyzed and overall weighted average intra-laboratory and inter-laboratory test method precision estimated at the 95% confidence level using the ASTM E691 protocol. Two of the test models in the proficiency test data sets did not included adequate data to allow the ASTM E691 analysis to be performed. For the same reasons, we have eliminated those data from the determination of coefficients of variation (CVs) needed for the Monte Carlo simulations. But unlike the overall test method variability study, only Category 1 and Category 4 proficiency test results were included in the CV determination. It should be noted that for our purposes, the terms "low burn rate" and "high burn rate" are used interchangeably with Category 1 and Category 4 respectively.

Data from proficiency test models Catalyic-1, Non-catalytic-1 and Non-catalytic-3 are used. Suspect data has been eliminated from the datasets.

The data are presented in Tables A1 through A3. There is one table for each test stove. The CV's for each test lab are first determined by dividing the individual laboratory standard deviation by the corresponding mean emission rate for all test runs on the given stove model. The average intra-lab CV is then determined by averaging the individual lab CVs. The inter-lab mean emission rate is next determined by averaging the mean emission rates from each of the test labs. The inter-lab standard deviation about that mean is then determined. The inter-lab CV is determined by dividing the inter-lab standard deviation by the inter-lab mean emission rate. This process was conducted separately for both the Category 1 and Category 4 data and for each of the three test stoves. The high and low burn rate CV where then determined. The results are presented in Table A4.

Table A1: EPA Proficiency Round Robin Test Stove "Catalytic-1" Data and CV Determination

	High Burn Rate PM Data - g/h				High Burn Rate			Low Burn Rate PM Data - g/h				Low Burn Rate		
Catalytic-1	Year		Year		Mean PM	Stdev	CV	Year		Year		Mean PM	Stdev	CV
Lab	1987		1988					1987		1988				
A	4.250	4.250			4.250	0.000	0.000	1.180	1.090			1.135	0.064	0.056
B	28.110	11.130	5.130	7.580	7.947	3.017	0.380	2.830	4.070	0.990	1.520	2.353	1.382	0.587
C	3.630	2.790	2.390	3.450	3.065	0.577	0.188	3.160	1.270	0.740	1.230	1.600	1.068	0.667
D	5.500	6.680	1.710	5.060	4.738	2.131	0.450	5.600	3.400	1.470	0.930	2.850	2.118	0.743
E	6.300	8.150	4.660	2.960	5.518	2.222	0.403	1.500	2.000	1.340	1.400	1.560	0.301	0.193
E1	5.500	8.900			7.200	2.404	0.334	1.500	1.200			1.350	0.212	0.157
F	10.300	3.200			6.750	5.020	0.744	2.000	1.400			1.700	0.424	0.250
G	8.290	2.910			5.600	3.804	0.679	19.490	1.360					
H	8.080	19.018	3.660	4.080	5.273	2.440	0.463	7.213	2.386	2.890	9.320	5.452	3.368	0.618
I			5.250	5.380	5.315	0.092	0.017			1.470	1.540	1.505	0.049	0.033
J			6.470	4.600	5.535	1.322	0.239			2.730	1.980	2.355	0.530	0.225
Intra-Lab CV							0.354							0.353
Ave. Intra- Lab CV								0.354						
Averages					5.593	1.520						2.186	1.265	
Inter-Lab CV					0.272							0.579		
Ave. Inter- Lab CV							0.425							

Indicates suspected outliers excluded from analyses

Table A2: EPA Proficiency Round Robin Test Stove "Non-Catalytic-1" Data and CV Determination

Non- Catalytic-1	Year		Mean High BR PM - g/h	Stdev	High BR CV	Year		Mean Low BR PM - g/h	Stdev	Low BR CV
Lab	1989					1989				
C	6.600	9.850	8.225	2.298	0.279	20.100	18.000	19.050	1.485	0.078
E	9.850	9.100	9.475	0.530	0.056	21.000	17.250	19.125	2.652	0.139
E1	7.550	7.460	7.505	0.064	0.008	22.180	18.070	20.125	2.906	0.144
H	6.830	6.800	6.815	0.021	0.003	22.050	18.390	20.220	2.588	0.128
J	8.280	8.540	8.410	0.184	0.022	21.080	14.410	17.745	4.716	0.266
J1	14.020	4.930				23.310	18.380	20.845	3.486	0.167
K	4.558	8.424	6.491	2.734	0.421	15.492	22.939	19.216	5.266	0.274
Intra-Lab CV					0.132					0.171
Ave. Intra- Lab CV						0.151				
Averages			7.820	1.107				19.475	1.018	
Inter-Lab CV			0.142					0.052		
Ave. Inter- Lab CV					0.097					

Indicates suspected outliers excluded from analyses

Table A3.1: EPA Proficiency Round Robin Test Stove "Non-Catalytic-1" High Burn Rate Data and CV Determination

	High Burn Rate PM Data - g/h												High Burn Rate		
Non-Catalytic-3	Year		Year		Year		Year		Year		Year		Mean PM	Stdev	CV
Lab	1993		1995		1996		1997		1999		2000				
L	13.020	8.200											10.610	3.408	0.321
K	9.310	5.120											7.215	2.963	0.411
D	28.470	11.890											20.180	11.724	0.581
E			11.260	14.440					5.270	17.970	11.400	11.060	11.900	4.210	0.354
M											19.820	14.420	17.120	3.818	0.223
N					4.020	11.270	3.010	6.660	9.260	15.990			8.368	4.855	0.580
O							8.850	5.210	5.530	4.780			6.093	1.864	0.306
Intra-Lab CV															0.397
Ave. Intra- Lab CV															
Averages													11.641	5.246	
Inter-Lab CV													0.451		

Indicates suspected outliers excluded from analyses

Table A3.2: EPA Proficiency Round Robin Test Stove "Non-Catalytic-1" Low Burn Rate Data and Average CV Determination

	Low Burn Rate PM Data - g/h												Low Burn Rate		
Non-Catalytic-3	Year		Year		Year		Year		Year		Year		Mean PM	Stdev	CV
Lab	1993		1995		1996		1997		1999		2000				
L	10.980	6.300											8.640	3.309	0.383
K	7.130	7.790											7.460	0.467	0.063
D	7.020	7.200											7.110	0.127	0.018
E			12.000	11.000					4.350	12.080	5.620	7.420	8.745	3.395	0.388
M			14.460	10.640							59.600	7.750	10.950	3.366	0.307
N					9.450	10.190	3.590	6.620	9.160	8.720			7.955	2.454	0.308
O							12.560	3.780	5.240	1.280			5.715	4.847	0.848
Intra-Lab CV															0.331
Ave. Intra- Lab CV															0.364
Averages													8.082	1.629	
Inter-Lab CV													0.202		
Ave. Inter- Lab CV															0.326

Indicates suspected outliers excluded from analyses

Average values for intra-lab and inter-lab CVs where then determined using the results from the three test stoves. Because of the large range of CV's across the test stove models and between low burn rate and high burn rate emission results, the standard deviations for both the average intra- and inter-labs CV's were also determined. The Monte Carlo simulations were conducted using the average CV's and the average CV's plus one standard deviation. The results of are presented in Table A4.

Table A4: Coefficients of Variation Used for Monte Carlo Simulations

			Average CV	
Stove	Category	Mean Emission Rate	Intra-Lab	Inter-Lab
Catalytic-1	High BR	5.593	0.354	0.272
	Low BR	2.186	0.353	0.579
Non- Catalytic-1	High BR	7.82	0.132	0.142
	Low BR	19.475	0.171	0.052
Non-Catalytic-3	High BR	11.641	0.397	0.451
	Low BR	8.082	0.331	0.326
		Averages	0.290	0.304
		Standard Deviations	0.110	0.194
		Average + 1 SD	0.400	0.498

Appendix B

Sample of Two-Stage Monte Carlo Simulation with Log-normal Distribution

Emission Limit	2.5
True Mean	1.8
Interlaboratory CV	30.4%
Interlaboratory Std	0.21
Mean (ln(x))	-0.401
Std (ln(x))	0.297
Intralaboratory CV	29.0%
Intralaboratory Std	0.20

Probability of Failure	Criteria	Description
14.8%	1	Run 1 is over
14.6%	2	Run 2 is over
23.4%	3	Run 1 or Run 2 is over
13.1%	4	3-Run Average is over
6.5%	5	Run 1 and 3-Run Average is over
30.7%	6	One of the Runs in the 3-Run Average is over
6.7%	7	Highest value in failing 3-run average replaced by one additional run
7.4%	8	Highest value in failing 3-run average replaced by two additional runs

Notes: 1. Input Values in Yellow Cells

2. In criteria sections in the results chart, 0 indicates emission limit met, 1 indicates emission limit exceeded.

Sample	Lab-Mean	Intralab Std	Mean(ln(x)-Lab	Std(ln(x)-Lab)	Run 1	Run 2	Run 3	Run 4	Average - Runs 2,3,4	Max - Runs 1 - 4
1	2.179	0.632	0.739	0.284	2.700	1.750	3.354	3.845	2.983	3.845
2	1.242	0.360	0.176	0.284	0.683	1.511	1.178	1.194	1.294	1.511
3	2.370	0.687	0.823	0.284	2.060	3.250	3.226	2.722	3.066	3.250
4	0.889	0.258	-0.158	0.284	0.781	1.022	1.544	0.692	1.086	1.544
5	1.470	0.426	0.345	0.284	0.950	1.345	1.533	1.270	1.383	1.533
6	1.367	0.396	0.272	0.284	1.143	1.754	1.987	1.018	1.587	1.987
7	1.795	0.521	0.545	0.284	1.786	1.967	1.142	2.494	1.868	2.494
8	1.154	0.335	0.103	0.284	1.558	1.115	0.805	0.715	0.878	1.558
9	2.508	0.727	0.879	0.284	3.310	3.285	2.102	3.215	2.867	3.310
10	0.848	0.246	-0.206	0.284	0.895	0.625	0.721	0.761	0.703	0.895
11	2.100	0.609	0.701	0.284	2.357	3.214	2.092	1.841	2.382	3.214
12	2.378	0.690	0.826	0.284	2.510	1.718	1.584	2.001	1.768	2.510
13	1.359	0.394	0.266	0.284	0.929	1.122	1.131	1.410	1.221	1.410
14	2.499	0.725	0.876	0.284	2.622	1.907	1.595	2.698	2.067	2.698
15	1.529	0.443	0.384	0.284	2.950	0.876	2.140	1.731	1.582	2.950
16	1.638	0.475	0.453	0.284	2.099	2.526	1.043	1.612	1.727	2.526
17	1.746	0.506	0.517	0.284	2.240	1.726	1.197	1.804	1.576	2.240
18	1.407	0.408	0.301	0.284	1.773	0.893	2.095	1.239	1.409	2.095
19	0.680	0.197	-0.427	0.284	0.369	0.990	0.543	0.590	0.708	0.990
20	1.486	0.431	0.356	0.284	1.604	0.794	1.174	1.505	1.158	1.604
↓										
999	0.575	0.167	-0.593	0.284	0.697	0.709	0.499	0.658	0.622	0.709
1000	0.693	0.201	-0.406	0.284	0.380	0.567	0.482	0.681	0.576	0.681

Sample	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Criteria 5	Criteria 6	Data Replacement: Run 5	Run 6	Replace Highest w/ Two Runs	Criteria 7	Replace Highest w/ Two Runa	Criteria 8
1	1	0	1	1	1	1	2.790	1.484	2.631	1	2.345	0
2	0	0	0	0	0	0	1.306	1.148	0.000	0	0.000	0
3	0	1	1	1	0	1	2.085	2.294	2.678	1	2.582	1
4	0	0	0	0	0	0	1.017	0.740	0.000	0	0.000	0
5	0	0	0	0	0	0	1.857	0.580	0.000	0	0.000	0
6	0	0	0	0	0	0	0.956	1.114	0.000	0	0.000	0
7	0	0	0	0	0	0	1.257	2.581	0.000	0	0.000	0
8	0	0	0	0	0	0	0.912	1.228	0.000	0	0.000	0
9	1	1	1	1	1	1	2.805	1.817	2.707	1	2.485	0
10	0	0	0	0	0	0	1.134	0.702	0.000	0	0.000	0
11	0	1	1	0	0	1	2.088	2.776	0.000	0	0.000	0
12	1	0	1	0	0	0	1.902	2.979	0.000	0	0.000	0
13	0	0	0	0	0	0	1.224	1.477	0.000	0	0.000	0
14	1	0	1	0	0	1	1.812	2.055	0.000	0	0.000	0
15	1	0	1	0	0	0	1.501	1.292	0.000	0	0.000	0
16	0	1	1	0	0	1	1.285	1.192	0.000	0	0.000	0
17	0	0	0	0	0	0	1.926	1.662	0.000	0	0.000	0
18	0	0	0	0	0	0	2.196	1.178	0.000	0	0.000	0
19	0	0	0	0	0	0	0.923	0.818	0.000	0	0.000	0
20	0	0	0	0	0	0	0.718	1.855	0.000	0	0.000	0
↓												
999	0	0	0	0	0	0	0.619	0.705	0.000	0	0.000	0
1000	0	0	0	0	0	0	0.417	0.679	0.000	0	0.000	0

Name: Robert W. Ferguson

Total Years in the Hearth Products Industry: 33

Companies and Dates of Affiliation:
Vermont Castings 1980-1990
Ferguson, Andors & Company 1991 - Present

Positions Held and Description of Responsibilities:
<u>Vermont Castings</u>

- Director of Research and Development
  - Responsible for all aspects of product development, product performance and product safety.

<u>Ferguson, Andors & Company</u>

- President
  - Founded Ferguson, Andors & Company in 1991, offering a full range of product development consulting and regulatory compliance services focused on the hearth, patio and barbecue industry. Clients include both small and large companies from around the world. Products developed include solid fuel and gas-burning appliances.
  - Providing HPBA with technical consulting services for the NSPS review/revision process that is now in the proposal stage at EPA.

Significant Accomplishments (include US Patents if applicable):

- Co-inventor for a number of patents related to the hearth product performance and combustion technology.

Trade and Professional Group Affiliations and Positions Held:

- Wood Heating Alliance (HPA/HPBA) Board of Directors
- Hearth Education Foundation Board of Directors/Treasurer
- WHA/HPA Government Affairs Committee Chair
- Represented the manufacturers' interests during the Regulatory Negotiations (RegNeg) that resulted in the current EPA New Source Performance Standards for Wood Heaters.
- ASTM Member, Task Group and Working Group Chairs
  - Chaired or acted as facilitator during the development of the ASTM solid fuel particulate measurement, fireplace PM emissions, wood heater PM emissions, pellet heater PM emissions and partial thermal storage hydronic heater PM emissions test methods.
- CSA B365 and B415.1 Technical Committee Member.

Other Relevant Information:

- BS Chemical Engineering, Clarkson University, 1972

Exponent
1800 Diagonal Road
Suite 500
Alexandria, VA 22314

telephone 571-227-7200
facsimile 571-227-7299
www.exponent.com

Richard Reiss, Sc.D.
Principal Scientist

Professional Profile

Dr. Richard Reiss is a Principal Scientist in Exponent's Health Sciences Center for Chemical Regulation and Food Safety. He is an environmental health scientist with expertise in risk assessment, exposure assessment, environmental chemistry and fate, mathematical modeling, and applied statistics. He provides consulting services related to scientific issues associated with numerous environmental statutes, and has expertise in both air quality and chemical risk assessment. He has conducted risk assessments, data analyses, probabilistic exposure modeling and environmental exposure modeling for environmental agents, such as pesticides, industrial chemicals, consumer product chemicals, and asbestos. He has conducted risk assessments for new and existing products.

Dr. Reiss is very active in the application and development of quantitative methods in risk assessment. He is the developer of the Probabilistic Exposure and Risk assessment model for FUMigants (PERFUM), which is an air dispersion model designed to evaluate bystander inhalation exposure following fumigant applications. PERFUM was favorably evaluated by a multidisciplinary expert panel assembled by Environmental Protection Agency (EPA), and is being used by EPA to evaluate the registration of new fumigant active ingredients and the re-registration of existing fumigant products. Generally, he has used a variety of mathematical models in conducting occupational and ecological risk assessments for pesticides and industrial chemicals; and performed statistical analyses, including dose-response modeling to evaluate chemical toxicity.

Dr. Reiss is actively involved in several scientific societies and he is the Past-President of the Society for Risk Analysis, the leading scientific society devoted to the field of risk assessment. Dr. Reiss was the Managing Editor of *Risk Analysis: An International Journal*, the leading scholarly journal for risk analysis, from 2001 through mid–2008. He was the winner of the 2001 Chauncey Starr award from the Society for Risk Analysis. This award recognizes a risk analyst less than 40 years of age that has made major contributions to the field of risk analysis. Dr. Reiss was also a councilor in the Society for Risk Analysis (term 2005–2008). In 2010, he was elected a Fellow of the Society for Risk Analysis.

Academic Credentials and Professional Honors

Sc.D., Environmental Health, Harvard University, School of Public Health, 1994
M.S., Environmental Engineering, Northwestern University, 1991
B.S., Chemical Engineering, University of California, Santa Barbara, 1989

Chauncey Starr Award from the Society for Risk Analysis, 2001, recognizing a scientist under 40 years of age who has made significant contributions to risk analysis; Outstanding Service Award, Society for Risk Analysis, 2009; Leslie Silverman Scholarship, Harvard University, 1991; Walter P. Murphy University Fellowship, Northwestern University, 1989–1990

Publications

Reiss R, Johnston J, Tucker K, DeSesso JM, Keen CL. Estimation of cancer risks and benefits associated with a potential increased consumption of fruits and vegetables. Food Chem Toxicol 2012; 50:4421–4427.

Reiss R, Neal B, Lamb JC, Juberg DR. Acetylcholinesterase dose-response modeling for chlorpyrifos and chlorpyrifos-oxon. Regul Toxicol Pharm 2012; 63:124–131.

Bogen R, Reiss R. Generalized Haber's law with exponential decline, with application to riparian-aquatic pesticide ecotoxicity. Risk Anal 2012; 32:250–258.

Cantor R, Lyman M, Reiss R. Asbestos claims and litigation. The John Liner Review 2009; 23:28–38.

Reiss R, Lewis G, Griffin J. An ecological risk assessment for triclosan in the terrestrial environment. Environ Toxicol Chem 2009, 21:2483–2492.

Levy J, Reiss R. The importance of modeling in exposure and risk assessments. Environmental Manager 2008; 14–17, June.

Reiss R, Anderson EL, Cross CE, Hidy G, Hoel D, McClellan R, Moolgavkar S. Evidence of health impacts of sulfate and nitrate containing particles in ambient air. Inhalat Toxicol 2007; 19:419–449.

Reiss R. Temporal trends and weekend–weekday differences for benzene and 1,3-butadiene in Houston, Texas. Atmos Environ 2006; 40:4711–4724.

Reiss R, Griffin J. A probabilistic model for acute bystander exposure and risk assessment for soil fumigants. Atmos Environ 2006; 40:3548–3560.

Reiss R, Schoenig GP, Wright, GA. Development of factors for estimating swimmer exposures to chemicals in swimming pools. Hum Ecol Risk Assess 2006; 12:139–156.

Reiss R, Gaylor D. Use of benchmark dose and meta-analysis to determine the most sensitive endpoint for risk assessment for dimethoate. Regul Toxicol Pharmacol 2005; 43:55–56.

Reiss R, Anderson EL, Lape J. A framework and case study for exposure assessment in the Voluntary Children's Chemical Evaluation Program. Risk Anal 2003; 23:1069–1084.

Reiss R, MacKay N, Habig C, Griffin, J. An ecological risk assessment for triclosan in lotic systems following discharge from wastewater treatment plants in the U.S. Environ Toxicol Chem 2002; 21:2483–2492.

Wilkinson CF, Christoph GR, Julien E, Kelley JM, Kronenberg J, McCarthy J, Reiss R. Assessing the risks of exposures to multiple chemicals with a common mechanism of toxicity: How to cumulate? Regul Toxicol Pharmacol 2000; 31:30–43.

Allen G, Sioutas C, Koutrakis P, Reiss R, Lurmann FW, Roberts PT, Burton RM. Evaluation of the TEOM method for measurement of ambient particulate mass in urban areas. J Air Waste Manage Assoc 1997; 47:682–689.

Reiss R, Ryan PB, Koutrakis P, Tibbetts S. Ozone reactive chemistry on interior latex paint. Environ Sci Technol 1995; 29:1906–1912.

Reiss R, Ryan PB, Tibbetts S, Koutrakis P. Measurement of organic acids, aldehydes, and ketones in residential environments and their relation to ozone. J Air Waste Manage Assoc 1995; 45:811–822.

Reiss R, Ryan PB, Koutrakis P. Modeling ozone deposition onto indoor residential surfaces, Environ Sci Technol 1994; 28:504–513.

Selected Presentations and Conference Proceedings

Reiss R. Estimation of cancer risks and benefits associated with a potential increased consumption of fruits and vegetables. Invited presentation at the U.S. Department of Agriculture, Washington, DC, 2012.

Reiss R. Measuring risk exposure when using global supplier. Society for Risk Analysis World Congress, Sydney, Australia, 2012.

Reiss R, Johnston J, DeSesso J, Tucker K. Pesticide residues on food: A mountain or a molehill. Society for Risk Analysis, Charleston, SC, 2011.

Reiss R, Bogen K. Modeling risk to aquatic species subject to realistic, dynamic exposures using a generalized form of Haber's law. American Chemical Society, Denver, CO, 2011.

Ma Q, Reiss R, Habig C. Applying the joint probability distribution analysis for Pacific Northwest salmonid risk assessment. American Chemical Society, Denver, CO, 2011.

Li A, Reiss R, Lowe K, McIntosh L, Mink P. Framework for integration of human and animal data for risk assessment. Society of Toxicology, Washington, DC, 2011.

Reiss R. Atmospheric modeling of fumigants. Workshop on methyl bromide alternatives, Kansas State University, Manhattan, KS, May, 2010.

Reiss R. Health risk assessment for fumigants. Keynote address to the annual meeting of the Australia-New Zealand Chapter of the Society for Risk Analysis, Sydney, Australia, September 2010.

Reiss R. Evaluation of water contamination from consumer product uses. Invited presentation to the National Capitol Area Chapter of the Society for Toxicology, Washington, DC, April, 2010.

Reiss R. The evolution of health risk assessment in the United States. Keynote address to the first annual Society for Risk Analysis meeting of the Taiwan SRA chapter, Taichung, Taiwan, January, 2010.

Reiss R. Risk analysis: The evolution of a science. Invited presentation to the Joint IRAC-SRA-CBER-JIFSAN Symposium on New Tools, Methods and Approaches for Risk Assessment, Baltimore, MD, December, 2009.

Reiss R. Exposure analysis: Pathways to refining regulatory risk assessments. Midwest States Risk Assessment Symposium, Indianapolis, IN, November 2009.

Williams P, Reiss R. Modeling the variability in consumer product use patterns. International Society for Exposure Analysis annual meeting, Minneapolis, MN, November 2009.

Cramer S, Poletika N, Everich R, Schocken M, Habig C, Reiss R. Framework for estimating exposure to ESA-listed salmon to pesticides. American Chemical Society semiannual meeting, Washington, DC, August 2009.

Reiss R, Edwards M. Analysis of cholinesterase variability in animals and implications for risk assessment. Society for Risk Analysis Annual Meeting, Boston, MA, December 2008.

Reiss R, Lewis G, Griffin J, Inauen J, Navarro L. Terrestrial risk assessment for triclosan. Poster presentation, Pacific Northwest Organic Residuals Symposium, Davis, CA, October 2008.

Reiss R, Chan R. Estimation of emission rates for building fumigations. Methyl Bromide Alternative Outreach conference, San Diego, CA, October 2007.

Reiss R, Chan R. Impact of estimation methods and tarping methods on flux rates. Methyl Bromide Alternative Outreach conference, San Diego, CA, October 2007.

Reiss R, Anderson E, Turnham P. Exposure and risk assessment for residents and contractors associated with vermiculite attic insulation. International Society for Exposure Analysis. Durham, North Carolina, October 2007.

Reiss R. A critical evaluation of the National Ambient Air Toxics Assessment (NATA) program for benzene. Society for Risk Analysis Annual Meeting, Baltimore, MD, December 2006.

Reiss R, Inauen J, Hoffman-Kamensky M, Capdevielle M. Terrestrial risk assessment for triclosan. Society of Environmental Toxicology and Chemistry Meeting, Montreal, Canada, November 2006.

Reiss R. Near-field air quality impacts from fumigant applications. American Chemical Society Meeting, San Francisco, CA, September 2006.

Reiss R. A probabilistic model for estimating bystander inhalation risks following fumigant applications. American Chemical Society Meeting, San Francisco, CA, September 2006.

Reiss R, Gaylor D. Statistical evaluation to determine the most appropriate endpoint for dimethoate risk assessment. Society for Risk Analysis Annual Meeting, Orlando, FL, December 2005.

Reiss R. Bystander risk assessment for fumigant: an evaluation of current regulatory activity. Society for Risk Analysis Annual Meeting, Orlando, FL, December 2005.

Gibb HJ, Kozlov K, Centeno J, Kolker A, Conko K, Reiss R. Potential health risks from long term mercury exposure in Gorlovka, Ukraine. Society for Risk Analysis Annual Meeting, Orlando, FL, December 2005.

Reiss R. Development of risk-based buffer zones for a fumigant application. Society for Risk Analysis Annual Meeting, Palm Springs, CA, December 2004.

Reiss R. Estimating fumigant buffer zones by air dispersion modeling. Methyl Bromide Alternatives Outreach Conference, Orlando, FL, October 2004.

Reiss R. Air exposure following a fumigant application. International Society of Exposure Analysis Meeting, Philadelphia, PA, October 2004.

Reiss R. Analysis of benzene and 1,3-butadiene emissions in the Houston Ship Channel. Presented at API/EPA Conference on Emissions Uncertainties, Houston, TX, 2003.

Reiss R, Anderson EL. A framework and case study for the Voluntary Children's Chemical Evaluation Program. Presented at the Society for Risk Analysis Annual Meeting, New Orleans, December 2002.

Reiss R. Emerging issues in environmental health for children. Invited lecture given at the Air and Waste Management Association meeting in Baltimore, MD, June 2002.

Reiss R, Griffin, J. A critical review of the National Emissions Inventory for Air Toxics. Presented at the Coordinating Research Council conference on Air Toxics Modeling, Houston, TX, February 2002.

Reiss R, MacKay N, Habig C, Griffin J. A probabilistic ecological risk assessment for Triclosan in lotic systems following discharge from wastewater treatment systems. Presented at the Society of Environmental Toxicology and Chemistry meeting, Baltimore, MD, November 2001.

Reiss R. A review of the National Air Toxics Assessment. Presented at the Mid-Atlantic Section Meeting of the Air and Waste Management Association, Baltimore, MD, December 11, 2000.

Reiss R, Wilkinson CW. Exposure to chemicals with same mechanism of action: How to add the risk? Presented at the Annual Meeting of the American College of Toxicology, McLean, VA, November 9, 1999.

Lurmann FW, Reiss R. Analysis of the first three years of PM2.5 data collected in the Southern California Children's Health Study. Presented at PM2.5 A Fine Particle Standard, Long Beach, CA, sponsored by A&WMA, U.S. Environmental Protection Agency, and the U.S. Department of Energy, January 28–30, 1998.

Reiss R, Chinkin L. Ozone exceedance data analysis: representativeness of the 1995 summer ozone season in the Northeast. Paper presented at the 1st NARSTO Northeast Data Analysis Symposium and Workshop, Norfolk, VA, December 10–12, 1996.

Coe D, Chinkin L, Reiss R, DiSogra C, Hammerstrom K. An emission inventory of agricultural internal combustion engines for California's San Joaquin Valley. Paper presented at the Air & Waste Management Association Emission Inventory: Key to Planning, Permits, Compliance & Reporting Conference, New Orleans, LA, September 4–6, 1996.

Main HH, Roberts PT, Korc ME, Coe DS, Dye TS, Lindsey CG, Reiss R. Analysis of PAMS and NARSTO-Northeast data—Supporting evaluation and design of ozone control strategies: A workshop. Presented at U.S. Environmental Protection Agency, Research Triangle Park, NC by Sonoma Technology, Inc., Santa Rosa, CA, December 11–12, 1995.

Chinkin LR, Ryan PA, Reiss R. A critical evaluation of biogenic emission systems for photochemical grid modeling in California. Paper presented at the Air & Waste Management Association and U.S. Environmental Protection Agency Emissions Inventory Conference, Research Triangle Park, NC, October 11–13, 1995.

Main HH, Roberts PT, Lurmann FW, Wright DB, Reiss R, Hering SV. Measurement of acid gases and PM2.5 in 12 Southern California communities for use in an epidemiologic study. Paper presented at the Air & Waste Management Association and U.S. Environmental Protection Agency Conference on Measurement of Toxic and Related Air Pollutants, Research Triangle Park, NC, May 16–18, 1995.

Reiss R, Lurmann FW, Roberts PT, Schoell BM, Geyh AS, Koutrakis P. A pilot personal ozone study in Southern California for validation of a microenvironmental model. Paper presented at the Air & Waste Management Association and U.S. Environmental Protection Agency

Conference on Measurement of Toxic and Related Air Pollutants, Research Triangle Park, NC, May 16–18, 1995.

Allen G, Koutrakis P, Reiss R, Lurmann F, Roberts PT, Burton R, Wilson W. Evaluation of the TEOM method for measurement of ambient particle mass in urban areas. In: Transactions of the Air & Waste Management Association Conference on Particle Matter: Health and Regulatory Issues, Pittsburgh, PA. Air & Waste Management Association, Pittsburgh, PA, April 4–6, 1995.

Reiss R, Ryan PB, Tibbetts S, Koutrakis P. Ozone reactive chemistry in residential environments. Presented at Air & Waste Management Association Conference, Measurement of Toxic and Related Air Pollutants, Durham, NC, May 1994.

Reiss R, Ryan PB, Koutrakis P, Bamford S. Modeling ozone deposition onto indoor surfaces. Presented at an Air & Waste Management Association Conference, Measurement of Toxic and Related Air Pollutants, Durham, NC, May 1993.

Book Chapters

Ma Q, Reiss R, Habig C, Whatling P. Use of the joint probability distribution analysis for assessment of the potential risks of dimethoate to aquatic endangered species. Chapter 12, pp. 171–181. In: Pesticide Regulation and the Endangered Species Act. ACS Symposium Series, Vol. 1111, American Chemical Society, 2012.

Reiss R. Use of simple stream modeling methods to assess the potential risks of malathion to salmonids. Chapter 11, pp. 159–169. In: Pesticide Regulation and the Endangered Species Act. ACS Symposium Series, Vol. 1111, American Chemical Society, 2012.

Cantor R, Lyman M, Reiss R. Asbestos claims and litigation. In: Product Liability, 2011.

Reiss R. Ozone reactive chemistry on interior surfaces of buildings. In: Encyclopedia of Environmental Analysis and Remediation, 1998.

Prior Experience

Vice President, Sciences International, 2000–2006
Senior Scientist, Quantitative Risk Assessment Expert, Jellinek, Schwartz & Connolly, Inc., 1998–2000
Senior Air Quality Analyst, Sonoma Technology, Inc., 1994–1998
Engineer, Environmental Solutions, Inc., 1990–1991

Editorships

- Managing Editor, *Risk Analysis: An International Journal*, 2000–2008
- Editorial Board, *Risk Analysis: An International Journal*, 2008–Present

Advisory Panels

- Air Quality Public Advisory Panel (AQPAC) for the Metropolitan Washington Council of Governments, Appointment for 2009–2011

Peer Reviewer

- *Risk Analysis: An International Journal*
- *Atmospheric Environment*
- *Environmental Science & Technology*
- *Journal of the Air & Waste Management Association*
- *Journal of Environmental Quality*
- *Regulatory Toxicology and Pharmacology*
- *Ecotoxicology and Environmental Safety*
- *Integrated Environmental Assessment and Management*
- *American Journal of Epidemiology*
- *Neurotoxicity Research*