Docket ID No. EPA-HQ-OAR-2018-0195

ATTACHMENT J TO HEARTH, PATIO & BARBECUE ASSOCIATION COMMENTS

Letter to Stephen D. Page, Director, OAQPS/USEPA from Craig T. Kenworthy, Executive Director, Puget Sound Clean Air Agency with attached "Preliminary Review and Critique of Analyses of NSPS Test Method Variability (Curkeet, 2010) and the Relationship of EPA Certified Values to 'In-Home Use' (Houck, 2012)" (December 5, 2012), *previously docketed as* EPA-HQ-OAR-2009-0734-0249

December 5, 2012

Stephen D. Page
Director
Office of Air Quality Planning and Standards
U.S. Environmental Protection Agency
109 T.W. Alexander Drive
Research Triangle Park, NC 27705

Dear Mr. Page,

We urge you to revise the outdated Standards of Performance for New Residential Wood Heaters (NSPS) to include lower, more health-protective emission standards and improved test procedures and fueling methods.

The health impacts from fine particles emitted from residential wood heaters are well documented. Impacts include premature death in people with heart or lung disease, heart attacks, aggravated asthma, decreased lung function and other effects.

In many areas of the country, fine particles (PM2.5) from residential wood burning are the largest contributor to violations of the National Ambient Air Quality Standards (NAAQS). While states and local agencies can take action to reduce PM2.5, we need every possible tool to meet the NAAQS and show we can sustain healthy levels over the long term.

A much stricter standard is both needed and justified based on current knowledge and technologies. The Clean Air Act directs EPA to consider emission limitations achieved in practice when revising the NSPS. It also defines “standard of performance” as a standard that reflects the “best system of emission reduction” that has been adequately demonstrated. Multiple manufacturers of wood stoves have achieved in practice emissions well below both the federal and Washington particulate standards, based on the large volume of test data submitted to the Washington Department of Ecology.

With *existing* technology, about fifty percent of all wood stove manufacturers currently produce a stove that can meet an emission limit of 3.35 grams per hour and the average performance of the top 50% is 2.5 grams per hour. There are twenty-two (almost 17%) catalytic and non-catalytic cord wood stoves currently certified that meet an emission limit of less than 2 grams per hour. This demonstrates that these levels are achievable and likely represent the “best system of emission reduction” and cannot be ignored in setting the standard.

Phone 206.343.8800 or 1.800.552.3565 **Fax** 206.343.7522

1904 Third Avenue - Suite 105 Seattle, WA 98101-3317

Washington State has had stricter wood stove standards (4.5 grams/hour for non-catalytic and 2.5 grams/hour for catalytic) in place for 17 years, along with burn bans and restrictions on sales of uncertified stoves. Despite this suite of measures, the Tacoma area was designated a PM2.5 nonattainment area due to winter wood smoke pollution. Almost half of wood stoves in this nonattainment area are EPA certified, and of those, we estimate that most meet the stricter Washington standard. What's more, we have run aggressive wood stove change-out programs in the nonattainment area for years in which we replaced wood stoves not meeting the Washington standards with those that did or with non-wood burning heating systems.

In spite of having more restrictive emission standards in place for almost 20 years, we were required to take more aggressive actions to ensure the Tacoma area will attain and one of which was requiring the removal or rendering inoperable of all uncertified woodstoves by September 2015. This problem might have been avoided or reduced in scope with a tighter NSPS standard in place. We are "living proof" that to avoid violating the PM2.5 NAAQS and to achieve a workable maintenance plan, the standard needs to be strengthened beyond the current Washington State standard.

In light of the existing data and the direction given to EPA in the Clean Air Act, we strongly support standards more stringent than the current Washington standards.

Wood stove changeout programs alone will neither prevent nor solve many PM2.5 nonattainment areas. Some have suggested that the replacement of older, uncertified stoves with certified stoves is enough to eliminate air quality problems and there is no need to make the emission standard more restrictive. In most wood smoke affected nonattainment areas and areas that will be nonattainment if the PM2.5 NAAQS is lowered, changing out uncertified stoves is helpful, but it will not solve the problem by itself. Changeout programs are difficult to implement and are very expensive. It is a challenge to get people to participate, particularly when many wood stoves have a useful life span of 40 or more years. Over the long term, urban airsheds with high levels of wood smoke and many wood burning devices can't support an unlimited number of new wood stoves unless they are much cleaner than the Washington standards.

As the number of woodstoves in a constrained urban area like Tacoma grows, it is crucial that each new woodstove be as clean as possible. With anticipated population growth, even if <u>every</u> uncertified wood stove were replaced with a stove that meets the more stringent Washington standards, and all new stoves meet the Washington standard, there could be future violations of the NAAQS. A federal standard lower than the Washington standard is needed to help prevent this backsliding.

In addition to a particulate matter emission limitation, EPA should also consider requiring a minimum thermal efficiency for new wood stoves in the NSPS along with requiring proper sizing of the wood stove for the home. Inefficient and over-sized wood stoves contribute unnecessarily to the PM2.5 generated by the stoves. Because these factors are directly related to the emissions from the stoves, EPA should consider them as part of the "best system of emission reduction" as required by the Clean Air Act.

When considering economic feasibility, current knowledge and the unique nature of wood heaters must be taken into account. In addition to directing the EPA to consider emission limitations achieved in practice when revising the NSPS, the Clean Air Act also directs EPA to take into account the cost of achieving the best system of emission reductions. When looking at the cost of achieving the reductions and the incremental costs for the additional reductions from a more stringent standard, EPA should consider the costs over a realistic lifetime of the equipment. Wood stoves last at least 20 years and when properly cared for, can last 40 or more years.

The inhalation intake fraction (the mass of pollution inhaled divided by the mass emitted) is much higher for residential wood heaters compared to stationary sources due to proximity of people to the emission source. Because of this, reducing a ton of fine particle pollution from a wood stove provides greater public health benefits than a similar amount of reduction from a stationary source. EPA should take this important health benefit into consideration when determining what a reasonable cost of achieving the best system of emission reductions would be.

When assessing the cost of developing lower emitting stoves, EPA should spread the cost of development over the full range of devices that will use the newly developed technology. Key components or technologies that are developed can be transferred to the manufacturer's other model lines, thereby reducing the cost of developing additional low-emitting stoves.

As the standard is strengthened, the test methods must also be improved. The NSPS methods and procedures used to certify wood heaters need to be improved to prepare for tighter emission standards. The parallel improvement of test methods and procedures along with tightening a standard is an expected part of continuing to protect public health in the face of a growing population and increases in polluting activities.

Contrary to other analyses EPA has seen, our analysis of the available EPA laboratory proficiency test data shows the current procedures can and do differentiate between good, better, and best stoves. The uncertainty in the test results for an individual stove, in combination with the form of the standard, makes the uncertainty for the best stoves (< 3.5 g/hour) greater than is ideal (see the attachment to this letter for a summary of our analysis). We believe that the test method can be substantially improved and the uncertainty sufficiently reduced to keep pace with a lowered standard.

We suggest EPA analyze the available laboratory proficiency and in-use data more thoroughly to understand the strengths and weaknesses of the existing procedures and not base regulatory decisions solely on one or two analyses performed by others who don't have the <u>express</u> responsibility to protect public health. EPA can then use all available information, including their own analysis, to guide improvement in the procedures where needed.

We have several specific recommendations for EPA action:

- Conduct a significantly more comprehensive and statistically robust analysis of existing testing laboratory proficiency data.
- Develop significantly more comprehensive and statistically robust test lab procedures.
- Ensure that quality assurance and auditing of accredited testing labs are a regular, robust part of the program. The requirements of 40 CFR 60.533 *Compliance and Certification* should be reviewed and evaluated to ensure that they are enforced and this section continues to serve the function for which it was initially written.
- Include provisions in the revised NSPS to gather data for future standards. Require data collection and record keeping to create a data repository to inform the next revision of the NSPS.
- Develop a more sophisticated model of independent variables, interactions, and emissions (e.g., understand better and account for bi-modal emissions behavior of some stoves.)
- Evaluate test methods and procedures used in Europe for possible use here.
- Evaluate moving the form of the standard to an emission factor (lb/mmBTU), or allow either an emission factor or an emission rate to qualify with an upper limit for the other. For regulatory and air quality management purposes an emission factor is more direct, and likely more accurate for modeling and inventories. An emission factor also reduces the complexity of the calculation and could reduce uncertainties introduced by the burn rate measurement.

It is likely that EPA accredited wood stove testing laboratories have performed emission tests that were not submitted to EPA. EPA should request this data from all accredited laboratories using the authority in Clean Air Act Section 114. Analysis of these test results would provide additional information that could be used to improve the test procedures. The revised standard should require that all information from all wood heater tests be submitted to EPA.

Significant changes in the test procedures could yield different results than the existing procedures, which must be taken into consideration when evaluating revised emission standards. EPA should thoughtfully sequence the changes to the standard and test methods to improve comparability and understanding of historic and future results.

With a potentially strengthened PM2.5 NAAQS on the horizon, we need this long overdue revision more than ever. The NSPS is long overdue for revision. In the time since the existing rule was promulgated, the entire landscapes of residential wood heating technology and knowledge of health impacts from fine particle pollution have completely changed. We strongly encourage EPA to take decisive action to propose a meaningful, health-protective standard for new residential wood heaters. This standard should reflect the best emission reduction systems that have been achieved in practice. Anything less will fail to provide states the tools that they need to meet the NAAQS and will fail to protect the health of millions of people who live in areas with extensive wood burning, including residents who do not burn wood, but who are exposed to high levels of fine particle pollution.

Sincerely,

Craig T. Kenworthy
Executive Director

Cc: Gil Wood, EPA
Greg Green, EPA

Cvp

Attachment

Preliminary Review and Critique of Analyses of NSPS Test Method Variability (Curkeet,2010) and the Relationship of EPA Certified Values to “In-Home Use” (Houck, 2012)

Summary

Discussion of the NSPS for New Residential Wood Heaters frequently cites and relies on two analyses: an analysis of the NSPS test method (EPA Wood Heater Test Method Variability Study Analysis of Uncertainty, Repeatability and Reproducibility based on the EPA Accredited Laboratory Proficiency Test Database, Curkeet, 2010) and an analysis of EPA certified emission rates verses “real-world” emission rates (A Comparison of Particulate Emission Rates from the In-Home Use of Certified Wood Stove Models with U.S. EPA Certification Emission Values and A Comparison between In-Home Uncertified and Certified Wood Stove Particulate Emissions, Houck, 2012).

Both of these analyses have substantially limited treatment of the data, ignore key issues, and disregard basic questions of data quality and representativeness. Their conclusions cannot be supported with the analyses that were conducted.

NSPS Test Method

The analyses of the NSPS test method variability (Curkeet, 2010) focuses on a “reproducibility” calculation and ignores the key questions, which are the accuracy of the EPA certification value, and the skill that accredited laboratories have in reproducing it. Curkeet also reports subsets of values as being representative of the overall data set, overlooks basic statistical properties of the data, and ignores the appropriateness of the methods used.

A first basic consideration is that the data are not normally distributed and there are considerable uncertainties in the EPA certification values, therefore “normal” statistics (which would be reasonable for evaluating the performance of many other measurement methods such as the trace pollutant concentrations in drinking water) are likely not appropriate. One of the most important analyses, which was not included in the Curkeet analysis, is a comparison of the laboratory emission rate with EPA certified rate. This comparison is also an example of when normal statistics (based on a normal distribution) are not appropriate. Because the EPA Certified Rate has uncertainty associated with it, an ordinary least squares regression (Pearson correlation) is not appropriate and a robust regression technique should be used.

Figure 1 shows the results of a standard robust regression technique, iteratively reweighted bisquare regression, of the mean laboratory result and the EPA certified value for the five test stoves that have circulated in the round robin tests since 1988. In this figure, the size of the dot and its color correspond to the number of labs which tested the stove. The robust regression technique, without knowing this information, picked out that there was an outlier. The outlier that was identified was the final stove with the fewest labs testing (only 2) and which occurred with a 5 year gap from the previous stoves (the small dot in the upper left). The other stoves were tested either in consecutive years or with only one or two year gaps. The fit shows a good linear relationship between the laboratory mean emission rate and the EPA certified rate down through the lowest stove, which was about 3 g/hour.

Technical Note: Identifying and dealing with outliers is a necessary step in robust statistical analysis. Outliers are data which are significantly far or different from the most of the data and generally do not represent the process or quantity of interest because of errors which are not part of the normal process. These errors can include procedural mistakes, data recording mistakes, equipment malfunction which was not otherwise identified, or other unknown problems. It is generally better to remove outliers from further analysis because they tend to have undue influence on the results.

The “good” stove which had an EPA rating of 7.5 g/hour had an average test lab performance of 14 g/hour. The “better” stoves which had EPA ratings of 4.5 g/hour and 3.5 g/hour had lab average performance of about 9 g/hour and 6 g/hour respectively, and the “best” stove which had an EPA rating of about 3 g/hour had an average lab performance of about 4 g/hour. The results of the four main stoves tested were in the correct rank and had a strong linear relation to the EPA certification value.

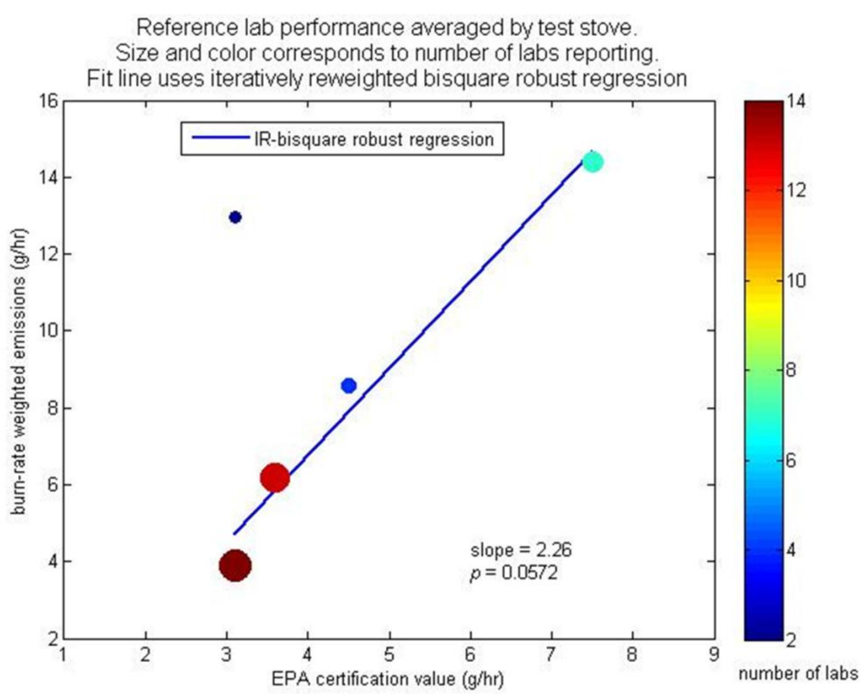


Figure 1. Robust regression of round robin testing data from EPA certified reference stoves. The size and color of the dots indicates the number of lab reports.

Another key piece of information that should be considered in this analysis is how individual tests deviated from the mean performance and therefore from the linear fit line (relation to the EPA certified values). **Figure 2** shows the residual, or difference, of each individual test for the three stoves with the most tests done (the two lowest and the one highest in **Figure 1**). From this figure it can be seen that the individual tests differ from the mean (i.e. have errors) in a well behaved manner (appear to be random) which was consistent across the stoves and years. A series of two-way Kolmogorov–Smirnov tests for difference among these three populations of residuals showed no significant difference (*p>0.38*). This is further evidence that the test procedure is an unbiased, well-behaved system, is under statistical control, and is correlated to the EPA certified values. The standard deviation of the errors is about 1.8 g/hour, which is higher than would be desirable for stoves with values below about 3.5 g/hour, but is more than adequate to distinguish a 3.5 g/hour stove from a 7.5 g/hour stove.

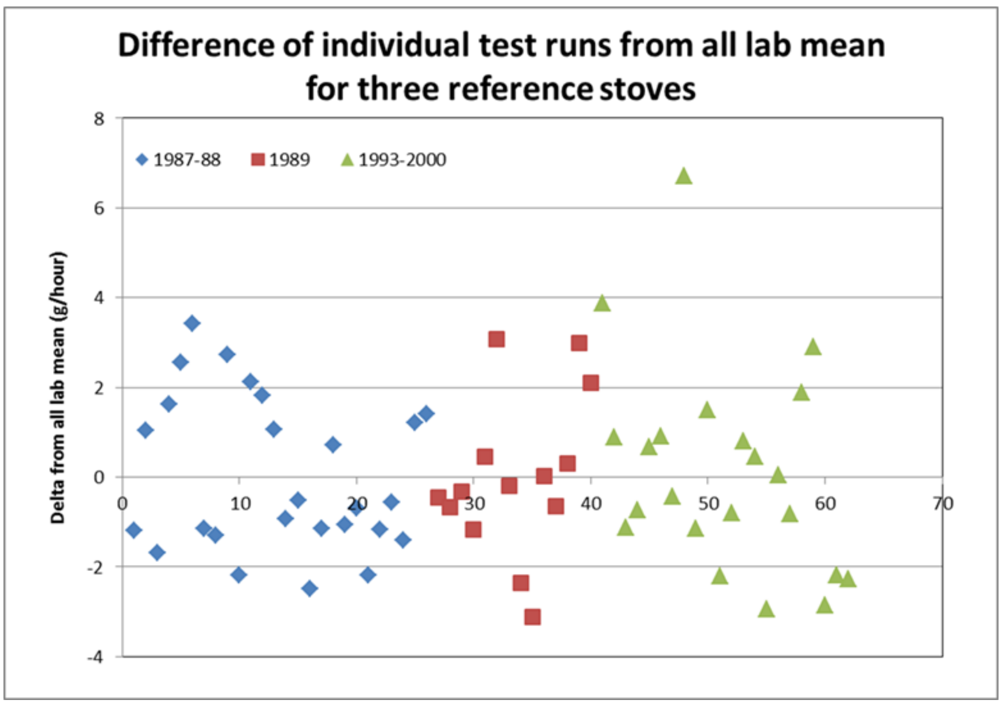


Figure 2. Residuals of the individual tests runs from the mean of all tests for that individual stove. The standard deviation is 1.8 g/hour. Three two-way Kolmogorov-Smirnov tests showed no significant difference between these sets of residuals (*p>0.38*).

A second concern is that Curkeet (2010) divides up the full data set, which allows larger values to be presented out of context as (incorrectly) representing the full data set. The values reported (+/- 4.6 to 9.8 g/hour) are not representative of the collective data set. A range of values is reported (and thus the highest value in the range, 9.8 g/hour) by looking at single stoves which are a subset of the entire data set. By looking at a single stove, the values are necessarily less representative of the entire data set and will have their own, much higher uncertainty. These values should be pooled together to reflect the entire population of data and reported as single value. Looking at the data stove by stove allows particularly extreme values to be highlighted and taken out of context. When in reality, the results obtained from one stove (a subset) - represents only a subset of the entire data set. The applicability of a subset to the overall population, or any new stove, is governed by the statistics of the entire data set and how that stove fits into the entire data set. The largest values (e.g. differences of 9.8 g/hour) would be substantially less likely to occur than smaller values and their frequency is predicted and described by the statistics of entire data set. An analogy could be made to motor vehicle emissions and motor vehicle fleets. It would similarly inappropriate to draw attention to the emissions test from the vehicle type with the highest pollution (say a large truck) and then select the large truck test which happened to have the greatest variability and then present this as being equally representative as the median result from the most common midsize car. The variability seen in the worst performing vehicle does not represent the entire fleet and it's likelihood of occurring is governed by how it fits into the larger fleet.

Curkeet also focuses on a statistic (ASTM "d2s" called "reproducibility") which incorrectly conflates the compounded uncertainty of successive measurements with a confidence interval about a mean, which is the more critical statistic. By using a difference statistics "d2s" – but implying and interchanging it with a confidence interval – the standard deviation is multiplied by a larger factor and thus the uncertainty appears to be much larger than a typically described 1-sigma or 2-sigma confidence interval. The "d2s" term describes the typical **absolute** difference (the gap) that would be expected from **two** tests, and is irrespective of which is higher or lower. This statistic compounds the uncertainty of two measurements. (If $\boldsymbol{\sigma_A}$ is the uncertainty or standard deviation related to measurement *A*, the uncertainty of a difference of two measurements is: $\sigma_{A-B} = \sqrt{\sigma_A^2 + \sigma_B^2}$ and if $\sigma_A = \sigma_B$ this reduces to $\sigma_{A-B} = \sqrt{2\sigma_A^2} = \sqrt{2}\sigma_A$, and a 95% confidence level multiplies this by 1.96 yielding: $1.96 * \sqrt{2}\sigma_A = 2.77\sigma_A$, assuming the data are normal). E.g. quoted from page 8 (bolding added):

> *"Using 2.8 times the standard deviation to estimate the **potential range of results that could be expected with a 95% confidence level** [Ref.: ASTM E177, 28.1], it is clear that the for any given test series the reproducibility is on the order of ± 4.9 – 9.8 grams per hour."*

and

> *"Since CV is calculated from one standard deviation, the implied variability in the actual data is really about 2.8 times the CV at a 95% confidence level. For example, a CV of 40% implies that the results of any one test could vary by **+/-** 112% **from a population mean** with a probability of 5% or less of being farther away."*

Both of these applications and interpretations are incorrect. "Reproducibility" and the "d2s" statistic means that two measurements will differ by that amount, **not** that the first value is a reference point or **population mean** and variations would always randomly occur in either direction + or - from the first value as occurs for a proper confidence interval. For example, if measurement *B* was "2.8 times the standard deviation" greater than *A* (i.e. *B*= *A*+2.8σ), then a third measurement, *C*, could **not** just as likely be "2.8 times the standard deviation" smaller than *A* (*A*-2.8σ). *A* and *B* are already at the ends of the

95% probability range, *A* was at the low end *B* was at the high end. So 95% of successive measurements will fall between *A* and *B* and the probability of having a test result as extreme as *A*-2.8σ would be extremely small (about 1 in 10,000) . A more conventional manner for reporting this uncertainty, and more conventional use of the term "confidence interval" is that the 95% confidence interval of the true (or reference) value is **+/- 1.96σ**, not "± 2.8 times the standard deviation" as Curkeet does. This is also consistent with his description of one standard deviation as a 68% confidence level, which he does on page 8 (bolding added):

> *"Even at **one standard deviation (68% confidence level)**, the reproducibility interval is about 1.7 to 3.5 grams per hour, i.e., there is a 32% chance that a new test result would deviate by more than this amount."*

The combination of these three incorrect analytic steps: applying normal statistics to non-normal data, dividing up the underlying data and implying a subset could represent the population, and using a difference statistic (absolute difference) and conflating it with a confidence interval, produces an uncertainty estimate which is likely unrepresentative and biased high, and disagrees with the clearly observable relationship in **Figure 1**, and the well behaved residuals in **Figure 2**. It is likely that a value of 1.5-2 g/hour is a more accurate representation of the uncertainty of mean test results. If this value can be decreased to about 1 g/hour, then with only four repetitions, a stove that performs at a mean of 2.5 g/hour can be said to be below a standard of 3.5 g/hour with about a 95% level of confidence.

Comparison of EPA Certified Values with "Real-World" Emissions

The analysis of Houck (2012) presents "real-world" data and compares it with EPA certified values. He concludes that there is little relationship between EPA certified values and "real-world" values. The analysis uses inconsistent and unrepresentative data sets, which themselves contain outliers and spurious stoves, and it overlooks a number of important basic properties of the data that are critical for the analysis to be robust. The conclusions cannot be supported by the analysis and substantial further analysis needed. In contrast, when even a few of these statistical issues are addressed, emission rates and emission factors show a relationship with certified values.

The major issues include:

- Data are from several studies that use different sampling methods, locations, testing protocols, testing locations, span many years, and do not include data from a balanced or representative cross spectrum of the stoves and independent variables. In particular, data from "simulated real-world" testing is not equally distributed throughout the range of stoves considered and throughout the independent variables (fuel type, fuel moisture, burn rate, stove EPA rating, etc).

One stove from this set has dramatically larger values than from two other data sets that studied the same stove.

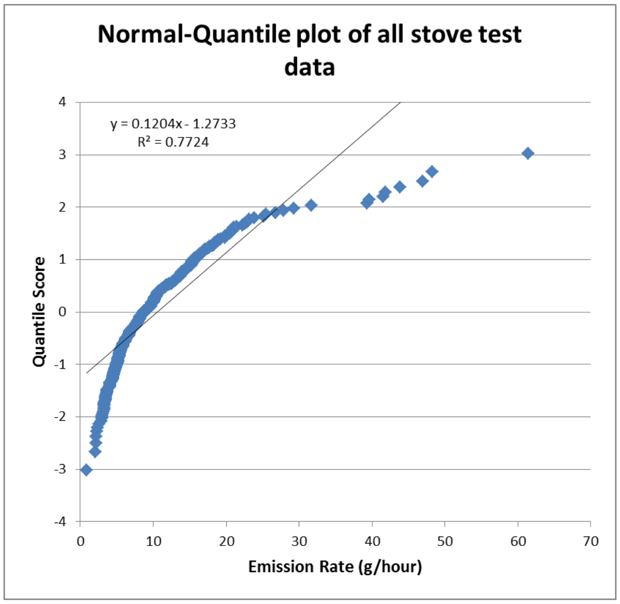
- The underlying data are not normally distributed, but the distributions, appropriateness, and sensitivity of the statistical techniques are not evaluated or discussed. Apart from the representativeness of the overall datasets, there are likely outliers and other non-normal characteristics that are not adequately handled. See **Figure 3a and 3b**.

- A few stoves have a bi-modal behavior that is substantially different from the rest of the stoves individually, and substantially different from the rest of the stoves taken collectively. This behavior consisted of a wide range of emissions, including much higher emissions, at a lower and fairly stable burn rate. Stoves with this behavior are a minority, yet these devices have the most extreme values and appear to inappropriately bias the analysis. This behavior needs to be better understood and accounted for in the testing methods and certification procedure. Normal or basic statistics and testing protocols cannot handle this sort of behavior in an accurate manner. See **Figure 4**.

- The burn rates used were substantially different from the burn rates specified by EPA certification Method 28 and were inconsistent across the bins. Referring to the bins used by Houck, the lowest bin averaged 1.22 kg/hour, the middle bin averaged 1.08 kg/hour, and the highest bin had an average 1.02 kg/hour. Considering the bi-modal behavior of some stoves, and the great sensitivity of their emissions rates on the burn rate, this burn rate inconsistency is likely to have introduced, or reflect a substantial bias. See **Figure 5 and Figure 6**.

- Compositing (binning) of the “real-world” data can introduce bias and be misleading if the data that make up the bins are not representative (e.g. if most of the data in the 5-7.5 g/hour bin are from stoves closer to 5, or closer to 7.5 g/hour). If the conclusion is robust, it should be consistent with different binning approaches, and other calculations (medians, emission factors, etc) See **Figure 7**.

- Reanalysis of the “real-world” data in finer bins (as opposed to the three bins used by Houck, 2012) shows consistency with EPA certification values even without correcting for unrepresentative data sets and inconsistent burn rates. Considering the underlying data are likely not representative, the agreement is good. Median emission rates (which should be a more accurate of the population when the data aren’t normal), and emission factors (which should be consistent with the emission rate data) both show the expected behavior of lower emissions (rate and factor) with a lower EPA certification value. See **Figure 7**.



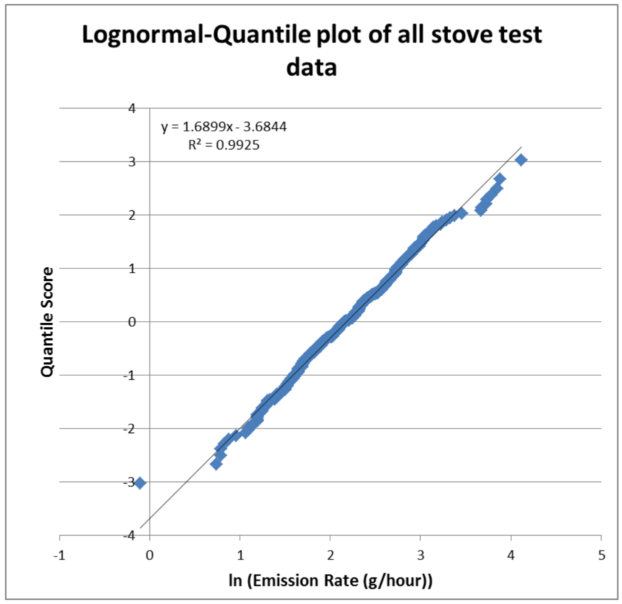


Figure 3a (left) and 3b (right). Quantile plots of all of the individual stove test data. The horizontal axis is the emission rate in g/hour on the left and ln(g/hour) on the right. The vertical axis is the quantile score associated with the rank of the value. If the underlying data are normally distributed and have no outliers, all of the points will lie along the straight line. On the left, it can be seen that there is substantial curvature from the straight line at the lowest and highest emission rates, with a large number of values far from the line. This is how log-normally distributed data appear on a normal-quantile plot. On the right is the same plot with the natural log (ln) of the emission rates. This shows a much better fit and is strong evidence that the underlying data are log-normally distributed. Note there that there may still be some outliers, even after a log-transform.

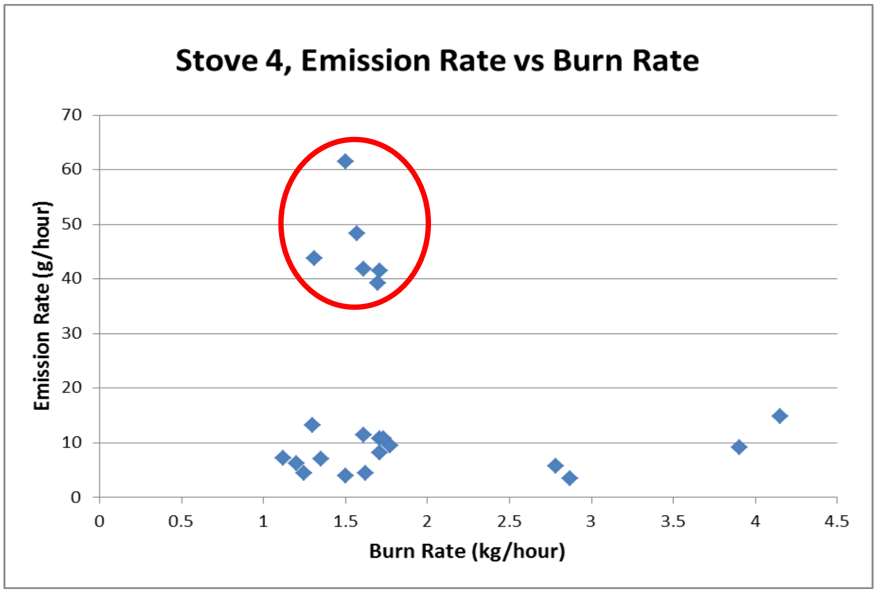


Figure 4. Emission rate (g/hour) vs burn rate (kg/hour) for stove #4. Example of a stove that exhibited a bi-modal behavior, which cannot be handled adequately with normal statistics and simple testing protocols. In this behavior, a higher burn rate can occur with little change in the emission rate (data at the bottom all around 10 g/hour), while a second mode of behavior occurs where the emission rate varies dramatically even though the burn rate remains close to 1.5 kg/hour. The data from the high cluster (circled in red) come exclusively from one study which “simulated in-home use” in the laboratory. The “laboratory simulated in-home use” data from this stoves alone comprise 6 of 9 values above 30 g/hour, with 2 additional values greater than 30 g/hour coming from Stove #11 “laboratory simulated in-home use” which were also inconsistent (substantially greater) than the rest of the data from that stove. Additionally, data from the “laboratory simulated in-home use” test were not distributed evenly across the range of certification values analyzed in the report.

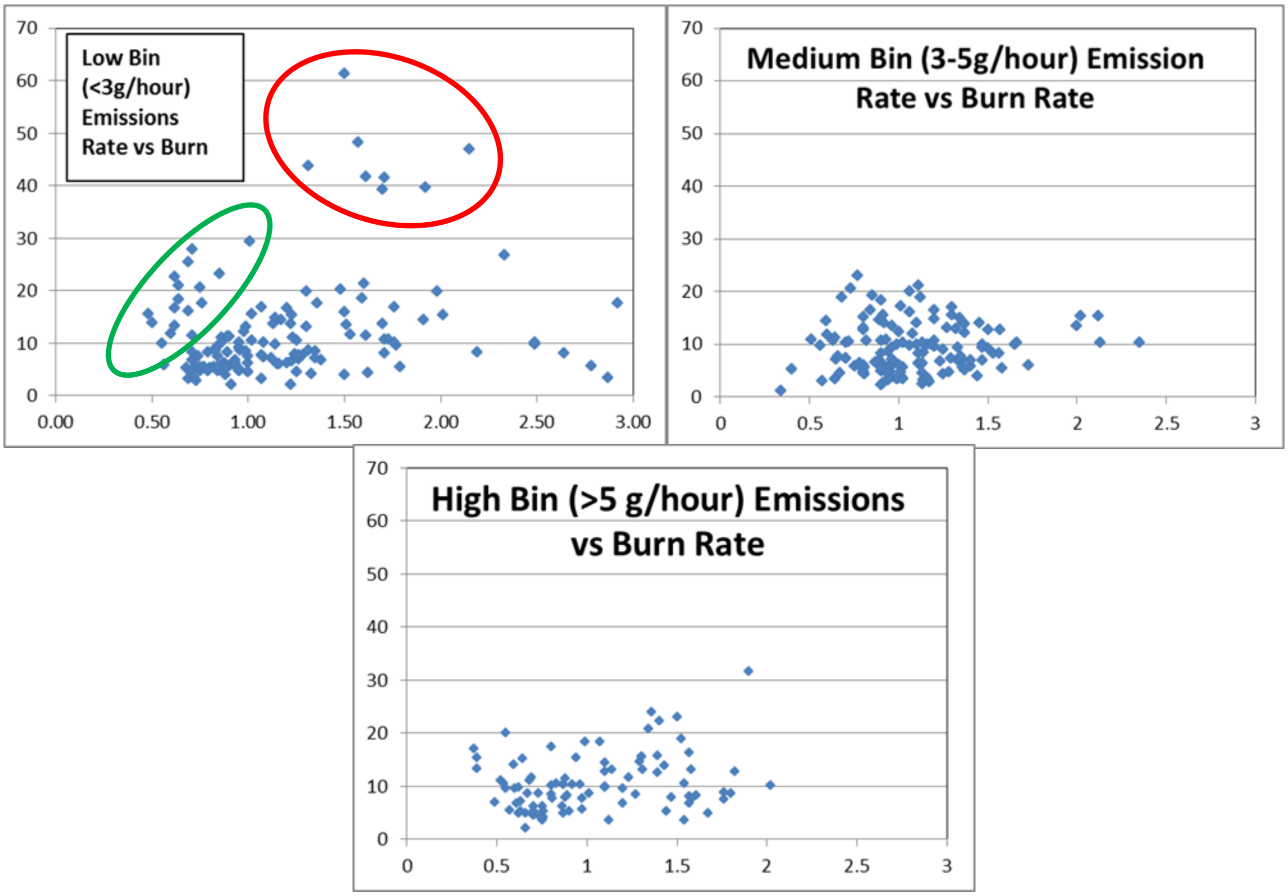


Figure 5. Emission rates on vertical axis (g/hour) vs burn rate on horizontal axis (kg/hour), all plotted on the same scale. The data are binned identically to Houck (2012). The Medium and High Bins data appear to be reasonably distributed and are consistent with each other. Only the Low Bin, however, has a cluster of values much higher than most values (circled in red). Also see **Figures 3 and 4**. The Low Bin also has a cluster of high emission rates at low burn rates that results from bi-modal behavior of a few stoves (circled in green). Also see Figure 4.

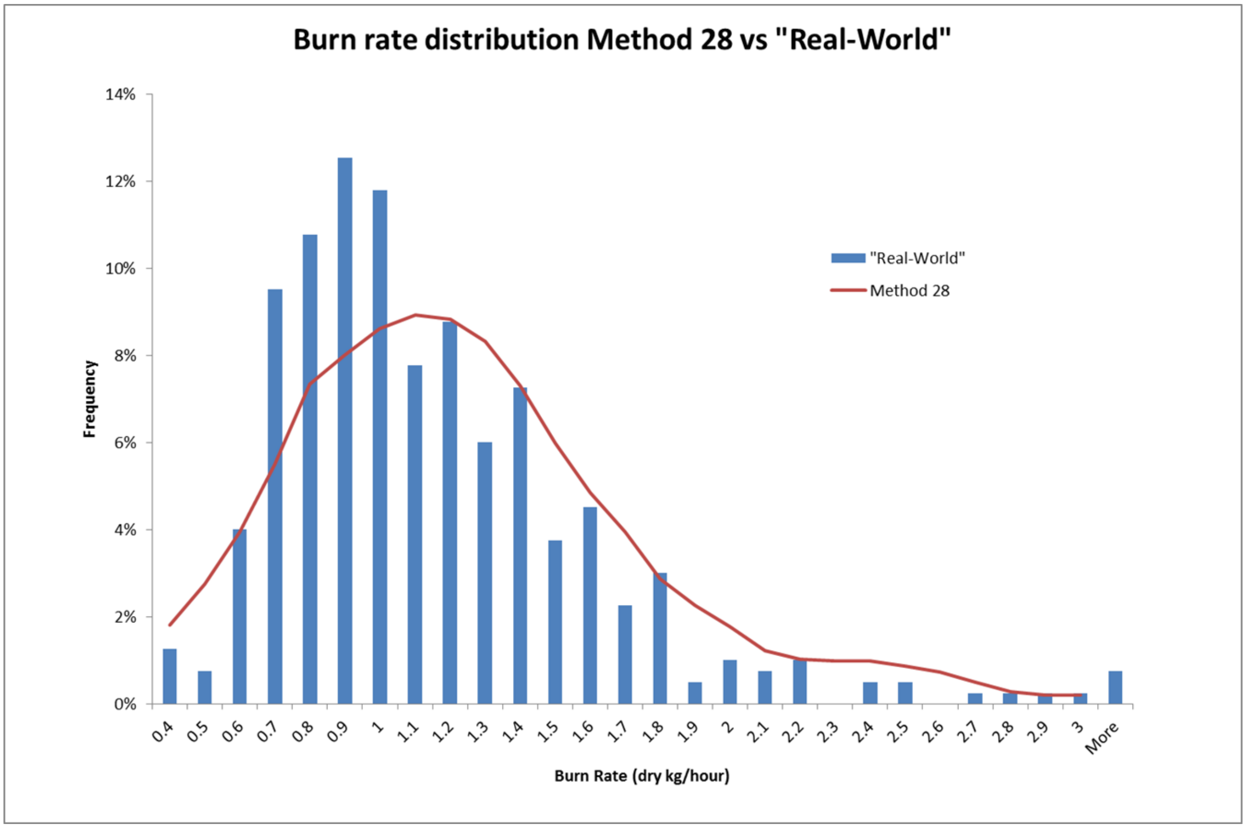
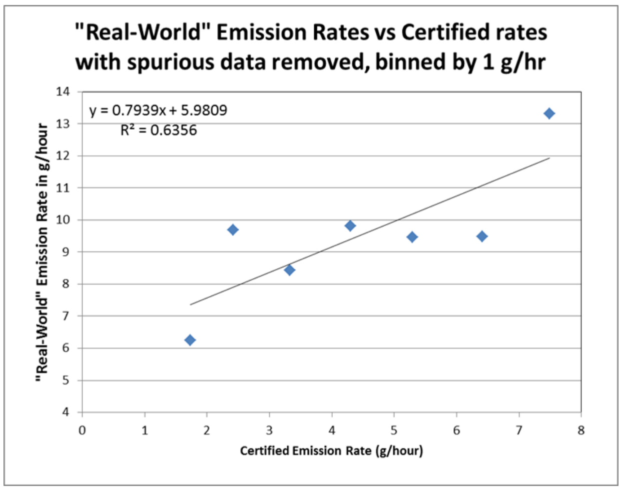


Figure 6. Burn rate histogram of all test data (bars) and the distribution specified by Method 28. The "real-world" data shows substantially more data collected from stoves operating at burn rates of 0.7-1.0 kg/hour than is specified by EPA certification Method 28. Considering the bi-modal behavior of some stoves, and the great sensitivity of their emissions rates on the burn rate, this burn rate inconsistency is likely to have introduced, or reflect a substantial bias



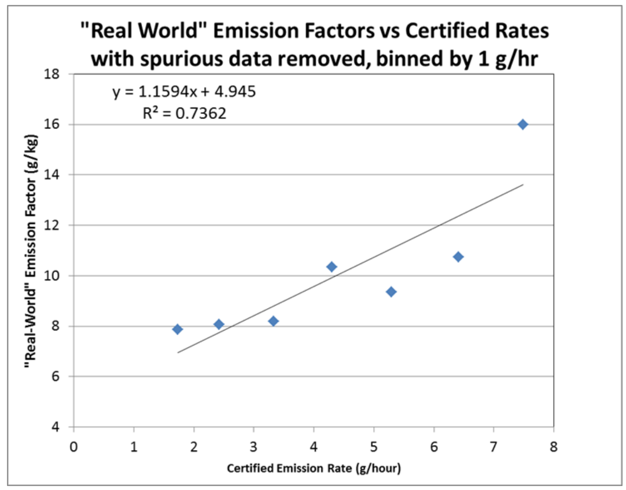


Figure 7a (left) and 7b (right). "Real-World" emission rates (left) and emission factors (right) from the data used by Houck (2012) binned in 1 g/hour EPA certification bins. In the lowest bin (1-2 g/hour) the sharply higher values from the bi-modal behavior of one stove was removed, and in the 2-3 g/hour bin, sharply higher values from the bi-modal behavior of two stoves was removed, including the outlier described in Figure 3 and 4. These figures show general consistency of "real-world" results with EPA certification values **even without correcting for unrepresentative data sets and inconsistent burn rates**. Note that if the burn rates are standardized, the emission factor and emission rates should display very similar behavior. The agreement of these different emissions estimate supports the conclusion that there is general consistency between the "real-world" emissions and the EPA certification values.