

**Results of the 2020-2022 USEPA
Burn Wise Residential Wood
Heater Testing Laboratory
Proficiency Test**

**Stack Test Solutions, LLC
August 30, 2022**

**John F. Buresh
Principal Research Scientist**

Table of Contents	Page 1
Introduction	Page 2
Methods and Materials	Page 2
Discussion	Page 4
Conclusions and Suggestions	Page 6
Table 1 Results	Page 11
Attachment 1-3 Dixon Tests	Page 13

Introduction

Stack test Solutions (STS), sole provider of the 2020 - 2022 USEPA Burn Wise Proficiency Test Program, began with proficiency testing at the first laboratory in February of 2020. This round of the proficiency testing was completed with the final report submitted from the last laboratory in May of 2022. In all, eight laboratories participated in the program to remain on the USEPA list of accredited Residential Wood Heater Testing Laboratories in the Burn Wise program. Those eight include (in alphabetical order):

ClearStak

Danish Technical Institute

Intertek

OMNI-Test Laboratories

PFS Teco

Poly-Tests Services

Research Institute of Sweden

Strojirensky Zkusebni Ustav

It should be pointed out that the proficiency test was performed at one lab for each of these companies. If any of these companies have more than one lab performing wood stove certifications, STS cannot verify nor ascertain the same techniques or lab setup or testing equipment at any satellite laboratories of these companies. STS submits this final report to satisfy the requirements of the USEPA Burn Wise Proficiency Test Provider requirements as described in the USEPA protocols.

Materials and Methods

Special acknowledgement should be given to Indeck Energy, who provided the pellets for both the conditioning burn and the pellets for the test burn. The pellets were predominantly from oak trees with some maple and possibly small amounts of birch. These made for consistent pellets and a pellet that held up well. The pellets were ¼ inches in diameter and between ½ inches to ¾ inches long. All pellets were taken within a half an hour of each other from the pelletizing and bagging line. While the conditioning pellets were shipped “as is”, care was taken with the test burn pellets to store them in nominal 5 pound hermitically sealed, evacuated storage bags to ensure there was no moisture or oxidation degradation between pellets burned in February of

2020 and March of 2022. Pellets were shipped near the testing date to ensure pellets would not have time to degrade even if the vacuum seal was lost due to shipping.

All single use audit samples (filters) were procured from ERA-QC in Colorado. ERA-QC is a well-known provider of environmental sampling audit materials. Filter audit samples were 47mm glass fiber filters with a known quantity of a white dry material on the surface. These were sent in advance of the proficiency test so STS could observe the final weight on site.

Mr. John F. Buresh of STS arrived at the laboratories on Monday mornings, and after introductions, provided the laboratory staff with the operating parameters provided by the USEPA, and began the proficiency test. The Proficiency test included all activities described in the USEPA Protocols: Observation of laboratory technique, equipment set up, cleaning activities, sample recovery activities and inspection of all equipment associated with the testing as it pertained to ASTM 2515 and ASTM 2779. STS remained at the testing location during all of the testing periods, and followed the sample throughout the stages of recovery until the final deposition in the desiccating trays.

Upon completion of the final test run, STS affixed seals on the stoves and provided the laboratories with a final review of observations and allowed time for any follow-up questions. STS collected the final results of the calculations sheets, and documented the final analysis of the audit samples. Several weeks later, the laboratories provided STS with draft reports that were finalized shortly thereafter.

Upon receiving the final report, STS calculated the Dixon outlier test on the gram per kilogram fuel combusted emissions of each individual test run, and found that no individual run was an outlier from the data set of the 24 runs (Attachment 1). For Room Air and Sample Train blank data, STS calculated The Dixon Outlier test on these (Attachments 2 & 3). The 8 data set is the minimum that can be run with the Dixon outlier test.

Discussion

The Covid Pandemic started shortly after we had begun our 2020-2021 Round Robin Proficiency Testing Program. STS and the laboratories had to work around this, changing our schedules as international travel was suspended to half the labs. Some labs had to reschedule due to Covid with the staff. STS staff was not immune to the pandemic; coming down with Covid during a trip to Europe servicing labs there. Due to the severity of Covid, the USEPA granted an extra year for the labs and STS to complete this round of testing.

STS attended all activities of the USEPA protocols in person with the exception of three audit filter weighings. Two Laboratories in question did not pass the initial probe audit filter portion of the test and were required to redo that portion of the proficiency test. A third laboratory found difficulty with the local customs and postal authorities and the filter did not arrive at the laboratory until after STS had left the country. All activities of the redo, including the opening, initial, intermediate and final weighings were observed via a Microsoft Teams connection.

One laboratory had the unfortunate experience of losing their sample stove. It was explained to STS that stoves undergoing or having undergone certification have identifiers to keep them stored indefinitely. Other stoves without those identifiers are removed on a regular basis. Unfortunately, their proficiency test stove did not have the proper identifiers and was removed and sent for disposal. This was discovered shortly before STS was to arrive and a new stove could not be procured in time. Another laboratory shipped their stove before STS arrived and STS was able to find the security labels intact and removed them in time for the test.

The Laboratories are de-identified by a color code. STS retains the actual data under the laboratory name for records kept at STS company offices. The color-coded final results can be found in Table 1.

At all of the participating Laboratories, STS inspected the wood stoves, the mixing and sampling ductwork, the external sensors, the sampling trains, and the recovery areas to ensure they met the standards in ASTM 2515 and ASTM 2779. STS utilized a checklist that was developed from requirements found in ASTM 2515 and ASTM 2779.

STS observed minor discrepancies from the ASTM methods from lab to lab, but observed nothing that we believe would invalidate the results of the testing, or overly bias the results. STS found all labs and staff were capable of performing the testing.

No laboratory was without findings or deviations from the written Method. Some findings we were able to correct immediately. Corrections such as locations of thermocouples, or pitot markings were corrected on the spot. Several labs had inappropriate transition elbows between the mixing duct and the sampling duct. They agreed to correct that in the near term and be ready for inspection at the next round of proficiency testing. Several labs utilized Method 5 sample trains, which are not nominally designed to sample at rates near 5-10 liters per minute (lpm), and calibrations were not appropriate for the sampling range.

We found one laboratory in the Sample Train blank did not meet the 90% outlier requirement set by the USEPA in the 2020 Protocols. In reviewing the data, it was insisted by our statistician that this test is inappropriate for such a small sample when operating at such low values near or below the practical quantification limits of the methods. Perhaps under several rounds of blank tests could an actual outlier be found, when the data set reaches more than 20. Conversely, having each lab perform several blanks to achieve a statistically valid number would be another option. Either way we look to this in the future current data does not justify identifying any laboratory as outside the round robin under the original scheme.

Other examples of shortcomings and corrects this round included:

1. *Train not leak-checking for the entire 60 seconds:* I informed the lab that the method describes a 60 second leak check even though the DGM was not moving at a pace that would indicate a leak check failure. The next leak check(S) was/were performed properly.
2. *Hood conical area not meeting 4X diameter of chimney requirement.* I ensured through detailed observation that ALL chimney emissions into the hood were captured.
3. *Anemometers not scaled low enough to meet Method specifications.* There was not much more that could be done with this other than procure a new anemometer which was not possible in our time available at the lab(s). The air was not moving in the sample location in my observation with the anemometer available indicated no movement of air.

4. *Sample probe location inaccurate.* There are two factors to use when calculating the sample locations and the lab(s) missed the second requirement. Upon presenting the secondary consideration, the lab(s) made the necessary corrections on the probe and sampled correctly
5. *Filter exposed for longer than 2 minutes during recovery.* We reviewed the method language and discussed how handling changes could be made to minimize exposure of the filters.
6. *Various deficient laboratory techniques.* We discussed how standard laboratory practices could be implemented to minimize risk of losing sample or data.
7. *Duct lengths not meeting method specifications.* We presented our the measurements and had them double check our measurements and review the test methodology language. It was long by around 12". Again, testing equipment that would change the outcome are not what a lab would want during a round robin test. It appeared they put things together and did not measure the final result. The Lab indicated they would correct the length for future testing.
8. *Using gloves when handling filters, probes...* a couple of labs did not use gloves for handling the filters and the probes. I explained that for a round robin test, it would be in their interests to utilize gloves as the majority of labs do. They found gloves for subsequent sample handling.
9. *No permanent (machine ink) identification marks on filters.* The lab used regular pen to identify the filters. They felt this was adequate, but indicated they would investigate finding machine ink for labeling filters.

All of these findings were documented and reviewed with the laboratory managers in their respective labs.

Conclusions and Suggestions

The 2020-2022 USEPA Burn Wise Proficiency Testing Program was the second association between STS and the Laboratories, and the level of comfort and confidence with STS attendance was more relaxed. The laboratories allowed me to review technique and equipment, and as some of this could be considered proprietary, STS again made all efforts to avoid any documentation that might identify the individual laboratory. STS concludes that all the results are accurate and represent actual testing and procedures of the individual laboratories.

After the inaugural proficiency test STS made several suggestions and have listed how these suggestions affected the results of this round of testing.

1. *The Dixon Test used to identify outliers suggests that 8 samples be a minimum number to use in the test, as it is a significance test, not a confidence test. The USEPA decided to consider each test run to be an independent variable rather than the average of the three runs from each laboratory to improve the statistical strength of the analysis. While this is not an unreasonable decision, it presumes each run is independent of the prior run. Observations in the field indicated that since the burn pot in the stove was not cleaned after each run, the prior run possibly influenced the proceeding emissions value. Field observations recorded one combustion pot was so fouled the stove could not re-light until the pot was agitated with pellets in the ash to allow for initial start-up. Be that as it was, the Dixon test did not find any of the 24 runs outliers due to the wide variability of the test runs. The mean was 2.00 grams per kilogram with a standard deviation of 0.727 grams per kilogram. **If the next protocols are using this scheme to identify outliers, STS suggests that both the combustion pot as well as the duct be cleaned prior each run to reduce this variable.***

The USEPA protocols were changed to require the combustion pot to be cleaned prior to each test run. The duct cleaning was NOT incorporated. As you can see in the data from each lab, the between-run variability of each lab was reduced as well as the actual measured emissions. It is the opinion of STS that this one change provided a much more accurate assessment of the laboratory staff's internal quality assurance skills in sampling.

The USEPA did not incorporate the suggestion to clean the chimney prior to each run, but clean it prior to the beginning of the first run only.

2. *For the Probe analysis of the Protocols, a 2% error limit was overly liberal, considering the probes were 39-45 grams in mass. For the next proficiency test, STS suggests using a Dixon test on three separate probe challenges to each lab. STS will present the probes on day one and the labs will have three days to achieve final weights. If final weight cannot be met while STS is on the premises, they can be finished as per protocols approved by the EPA for remote viewing of laboratory practices. **STS suggests calculating the outlier based on those 24 independent measurements.***

The USEPA dropped the probe audit portion of the proficiency test. This suggestion became moot.

3. *The calculations data set proved problematic for many of the laboratories. Their spread sheets were not designed to take single points. Some found the only way to calculate the results was by hand instead of the spread sheets they normally use. The rounding conventions and carrying of significant figures in the answer*

sheet did not seem to follow USEPA conventions. Perhaps STS can work with the USEPA developing the answer key in the next calculations sheet to ensure we understand them well enough to provide guidance to the laboratories. **The USEPA might want to consider either creating a data set of 60 data points for an hour of simulated testing of the data that a laboratory must collect on a run, and challenge the laboratory in that manner, or possibly drop this portion of the test.**

The USEPA dropped this portion of the proficiency test.

4. *STS will work with audit sample providers to find a proper audit sample with suspended particulate in acetone for the labs that recover the sample equipment with solvent for gravimetric analysis.*

The USEPA dropped this portion of the proficiency test

5. *STS recognized room air balance and combustion air might be mitigating factors in combustion efficiency for these small stoves. To eliminate this, **STS suggests USEPA to consider requiring these stoves be attached to an unobstructed outside air source.***

The USEPA did not include this as a requirement in the most recent proficiency test.

6. *STS suggests the USEPA requires **a flow to be performed prior to each run.***

This was included in the most recent Protocols.

7. *STS requests guidance whether the leak check should occur with the flow meter (rotameter) or the dry gas meter.*

STS again requests guidance.

8. *Examining the first runs from each laboratory, it appears two or possibly three laboratories would not pass the Dixon test and would have been identified as outliers. Allowing for three runs protects the labs from a very unforgiving statistical analysis. **STS suggests the USEPA considers maintaining the 3-run course.***

The USEPA maintained the three-run course.

9. *Three-hour test runs allow the Laboratories to complete testing in two days. If the EPA wants greater mass collected on the filters, they could consider going up*

to four-hour test runs. Five-hour test runs would require the laboratories to have three days of testing.

The USEPA maintained the three- hour run.

- 10. The back filter never collected any measurable particulate, and in many instances actually subtracted from the total catch. I presume this requirement is for wetter wood testing when there is a greater chance of condensable material captured. For this testing, USEPA could consider either dropping the need for the back filter, or only using the data when the mass is a positive value.*

The back filter was maintained in the test.

- 11. STS recognized some laboratories chose to induce draft in the chimney to a number just below the ASTM limit of 1.25 Pa (0.005 inches of water). STS is not certain what that does for combustion, but it probably has an effect. **STS suggests the USEPA considers dropping that limit to 0.25 Pa (0.001 inches of water) to eliminate that effect.***

The ASTM limit was used.

- 12. The stove has a 1-9 setting with one being lowest and 9 being highest. The laboratories operated at #4 setting for the tests. There was some variability that might be innate or due to some other lab parameter, possibly room air balance or draft induction. **STS suggest EPA selects a number (1-9) for the proceeding rounds of testing.***

The USEPA selected the #7 operation setting for the testing.

- 13. Required sample flow rate was an issue last year, as the EPA requested one at 10 lpm, when the method does not allow greater than 7 (LPM). There were some labs that did not have the proper equipment to reach the 10 LPM rate. **STS suggests the EPA provides STS with ample time to review protocols in advance of the proficiency tests to insure appropriateness of the test parameters.***

The USEPA reduced the sample rate to 5 lpm.

Below I have provided ideas and suggestions for the USEPA to consider for future proficiency testing under the Protocols the EPA might consider when developing the protocols for the next round (2023-2024):

1. The air flow issue should be addressed. We had the fortune to see one stove operated by two labs. While neither lab had an air source pulling in outside air, it is very possible the ventilation system allowed for more air and better burn than the other one. The particulate results suggest that. As I examined the rooms where testing was occurring, it was very apparent that the combustion air could be affected by multiple fans pulling air out of the room. Although it may be difficult for some labs to accommodate this, the lack of an unobstructed outside combustion air source may very well be the reason for differences between labs seen in this round of testing. The manufacturer is very specific in this requirement.
STS suggests again that the manufacturers requirement for outside air be heeded.

2. The 5 lpm was still too high for some labs.
STS suggests going to a longer run and dropping the sample rate to 3 or 4 lpm.

3. The USEPA has indicated it plans to reduce the testing to 1 run per lab for the 2023-2024 proficiency test. That may well reduce the time labs have to put aside for testing, and I understand that burden. However, in reviewing the data and performing the Dixon test on one sample for each lab, moving only one sample result towards the mean (not an outlier) caused laboratories to fail the 95% significance test on both the high and low side. 8 samples are the absolute minimum one can use on the Dixon test. And the statistical treatment is very rigid and unforgiving when $n=8$. **STS does not make a suggestion on this decision, but wishes to provide this warning.**

4. As can be seen in the data, the labs nearer the low end of the particulate loading demonstrated greater variability in the results, which may be due to sampling near the limit of quantification (LOQ). If a requirement for outside air combustion source reduces all the stoves emissions, we may see LOQ issues driving variability in the program. That would undercut our purpose to evaluate proficiency and begin to put chance as a greater part of the differences between laboratories. **STS suggests sample volume of air be considered in the next proficiency test.**

Table 1

		Red			
Parameter	Run 1	Run 2	Run3	Mean	Values
grams/hour	2.24	2.30	2.06	2.20	
grams/kg	1.68	1.71	1.48	1.61	
Sample Blank Value					0.00 mg
Room Blank Value					0.1 mg
Filter Error					-1.1 mg

		Pink			
Parameter	Run 1	Run 2	Run3	Mean	Values
grams/hour	1.92	2.18	1.99	2.03	
grams/kg	1.07	1.20	1.06	1.11	
Sample Blank Value					0.00017 mg
Room Blank Value					0.00008 mg
Filter Error					-0.93 mg

		White			
Parameter	Run 1	Run 2	Run3	Mean	Values
grams/hour	1.94	1.99	1.93	1.96	
grams/kg	1.02	1.06	1.05	1.04	
Sample Blank Value					0.0 mg
Room Blank error					0.0 mg
Filter Error					-0.7 mg

		Black			
Parameter	Run 1	Run 2	Run3	Mean	Values
grams/hour	2.59	2.42	2.50	2.50	
grams/kg	1.34	1.30	1.36	1.33	
Sample Blank Value					0.0000 mg
Room Blank Value					0.0000 mg
Filter Error					-1.3 mg

Table 1 (cont.)

		Orange			
Parameter	Run 1	Run 2	Run3	Mean	Values
grams/hour	2.101	1.977	1.895	1.991	
grams/kg	1.070	1.030	0.977	1.026	
Sample Blank Value					0.0000 mg
Room Blank Value					0.0000 mg
Filter Error					-1.0 mg

		Blonde			
Parameter	Run 1	Run 2	Run3	Mean	Values
grams/hour	2.46	2.39	2.50	2.45	
grams/kg	1.35	1.29	1.37	1.34	
Sample Blank Value					0.0 mg
Room Blank Value					-0.20 mg
Filter Error					-2.8 mg

		Blue			
Parameter	Run 1	Run 2	Run3	Mean	Values
grams/hour	2.93	2.84	2.73	2.83	
grams/kg	1.60	1.54	1.57	1.57	
Sample Blank Value					0.2 mg
Room Blank Value					0.00 mg
Filter Error	-0.5				-0.7 mg

		Brown			
Parameter	Run 1	Run 2	Run3	Mean	Values
grams/hour	2.675	2.445	2.480	2.53	
grams/kg	1.390	1.318	1.337	1.348	
Sample Blank Value					0.008 mg
Room Blank Value					-0.19 mg
Filter Error					-1.01 mg

Attachment 1

Dixon's Outlier Test ASTM 2515 Emissions Testing

Number of Observations = 24

10% critical value: 0.367

5% critical value: 0.413

1% critical value: 0.497

1. Observation Value 1.71 is a Potential Outlier (Upper Tail)?

Test Statistic: 0.162

For 10% significance level, 1.71 is not an outlier.

For 5% significance level, 1.71 is not an outlier.

For 1% significance level, 1.71 is not an outlier.

2. Observation Value 0.997 is a Potential Outlier (Lower Tail)?

Test Statistic: 0.055

For 10% significance level, 0.997 is not an outlier.

For 5% significance level, 0.997 is not an outlier.

For 1% significance level, 0.997 is not an outlier.

Attachment 2

Dixon's Outlier Test for Sample Train Blank

Number of Observations = 8

10% critical value: 0.479

5% critical value: 0.554

1% critical value: 0.683

1. Observation Value 0.2 is a Potential Outlier (Upper Tail)?

Test Statistic: 0.960

For 10% significance level, 0.2 is an outlier.

For 5% significance level, 0.2 is an outlier.

For 1% significance level, 0.2 is an outlier.

2. Observation Value 0 is a Potential Outlier (Lower Tail)?

Test Statistic: 0.000

For 10% significance level, 0 is not an outlier.

For 5% significance level, 0 is not an outlier.

For 1% significance level, 0 is not an outlier.

Attachment 3

Dixon's Outlier Test for Room Air Blank

Number of Observations = 8

10% critical value: 0.479

5% critical value: 0.554

1% critical value: 0.683

1. Observation Value 0.1 is a Potential Outlier (Upper Tail)?

Test Statistic: 0.345

For 10% significance level, 0.1 is not an outlier.

For 5% significance level, 0.1 is not an outlier.

For 1% significance level, 0.1 is not an outlier.

2. Observation Value -0.2 is a Potential Outlier (Lower Tail)?

Test Statistic: 0.050

For 10% significance level, -0.2 is not an outlier.

For 5% significance level, -0.2 is not an outlier.

For 1% significance level, -0.2 is not an outlier.