Clarkson University, Potsdam N.Y
26 September 2014

Blowing Smoke: The Curious Case of the Mangled Metric

Why improved cookstove performance tests are mostly hot air.

- Crispin Pemberton-Pigott
  Principal Technical Advisor, Sustainable Energy Technology and Research Center (SeTAR), Research Village, Bunting Road Campus, University of Johannesburg, South Africa.

When fire was invented the workload began. Carrying wood to the fire was a new task that provided humankind great benefits: warmth, light and protection from wild beasts. The smoke gave relief from mosquitoes and blackflies and added flavour to dried fish. Over time roasted vegetables and meat came to supplement and then dominate our diet and the next thing you know, we have stir-fry and Whoppers and caramel lattes. My, how clever we are.

The problem was of course the need for fuel to be added now and then to keep a fire going and before long, opinions formed about how to most efficiently operate the fire so as to reduce the workload of keeping it going.

Inventions regarding fuel efficiency abound here in the NE USA. The bundling board, which was placed between courting couples as they sat huddled in a single bed was invented to save fuel during the bitter cold of winter days. The Franklin stove, patented in the 1740's was a popular stove made from cast iron. It is based on an even earlier 1680's invention of the downdraft stove in England which, so its inventor claimed, was clean-burning to the extent that no odour could be detected when fuelled with 'coal soaked in cat piss' the stinkiest fuel then known to exist.

I would be remiss if I did not mention the invention in this same region of the medical procedure by which a recently drowned corpse was inflated with smoke via a posterior-applied bellows in the belief – attested by the medical community of the time – that it might revived the poor fellow. Specialised equipment was designed to assist. This led to the formation, no doubt, of a manufacturing association, the New England Blowing Up Men Society, or the "New England BUMS". (I have yet to confirm that rumour.) But by 1850 'blowing smoke' had passed into the popular lexicon attracting a meaning equivalent to, 'talking nonsense' and is used in that spirit in the title of this presentation.

The invention of the engineering degree (by engineers, of course) brought standardisation to the measurement of fuel consumption in service of the fuel burning community. Early on, attention focussed principally on determining either (A) how much 'task' could be accomplished with a fixed 'quantity of fuel' or (B) how much 'fuel' was required to accomplish a particular 'task'. These two approaches gave us two metric: 'miles per gallon' and 'litres per hundred kilometers' each based on a different engineering philosophy.

In the cooking stove community 'fuel consumption' has historically been based on the 'litres per hundred kilometers' philosophy: a fixed task is defined and the fuel requirement determined. By

1985 Volunteers in Technical Assistance (VITA – another fine New England organisation) had paid enough attention to the matter to produce a test method called the Water Boiling Test. It included a defined proxy cooking task – boiling and simmering water – and a method for determining a proxy heat transfer efficiency. The heat transfer efficiency was considered at that time to be the same as the fuel consumption. Higher efficiency, lower fuel consumption, just like the big boys in the power industry.

It is important to consider where this notion of efficiency came from. The efficiency of a boiler was determined by measuring the mass of fuel consumed and the amount of steam generated. My 1916 First Edition Engineering Handbook from McGraw-Hill includes a method for making this calculation. The boilers were in constant operation, the fuel feed rate was constant and the net system efficiency including all mechanical and chemical losses was accurately determined. One problem with this approach is that the start-up and shut-down emissions of the boiler are not included. It is a constant 'burn method' and stoves are not used like that. Further, changing the fuel for a boiler changes the emissions and this has historically been attributed to 'inherent combustion properties of the fuel', not an inherent property of the combustor which is a pretty big mistake. Fuels were then classed as 'high' and 'low' quality depending on emissions.

Early on, the error with respect to the ignition cycle was recognized but it was a long time before the idea that "some fuels are inherently clean-burning" was replaced by the realisation that it is the fuel+stove+operating cycle that defines emissions and performance, not the fuel alone. In some quarters the misconception still prevails because we hear, for example, that ethanol is 'clean burning' even if the stove stinks and the fire produces a high level of carbon-monoxide.

The VITA WBT gained some acceptance though Baldwin in 1987 criticised it for certain shortcomings and made recommendations where were not implemented as it did not have an owner and the stove community was a fringe group with clay under their fingernails. The test was replaced at some time by one written by stove enthusiasts which evolved into something called the WBT-3.1. This was very similar to the original VITA version seeking to capture the heat transfer efficiency while boiling and simmering a volume of water. It contained additional conceptual mistakes on which I will now focus.

The initial impulse was, as I said, to determine the thermal efficiency by using a method adopted from boiler testing which is a steady state problem. The cold-start performance of a stove can be affected by the thermal mass of the components. A brick stove can take a lot longer to get going because so much heat is invested in the bricks. Later, that heat might be applied to keeping a pot hot. In order to try to determine the influence of the thermal mass it was decided to start the stove cold, boil a pot of water and then boil a second pot with the stove already hot and then average the performance to reflect the influence of the thermal mass. This was a major step away from a proxy for cooking because almost no one cooks like that. You might cook one pot or two, but you don't cook them with the stove in an average temperature condition. For any defined task, the *difference* in performance between a cold and hot start is informative, but the average is not. In fact the average hides the difference so we can't predict what the stove will do if alternate burn cycle is employed.

Fuel consumption

It was clear enough that the mass of water affects the mass of fuel needed to complete the water heating portion of the cooking simulation so the mass of fuel consumed was divided by the number of litres boiled to give a Specific Fuel Consumption (SFC) in units of grams of dry fuel burned per litre of water boiled. This metric was widely used thereafter. The definitions of the two elements are, however, curious:

‘Fuel consumed’ is the dry mass of fuel burned, with the energy value of unburned ‘charcoal remaining’ converted to a dry fuel equivalent and subtracted from the mass of fuel burned. The metric defines the fuel consumption as the dry fuel equivalent of the energy released during the burn, ignoring chemical losses, treating partly burned sticks as raw fuel, and char remaining as an equivalent unburned mass of raw fuel with the conversion being based on the relative lower heating energy values of the char and the raw wood.

The number of litres boiled was defined as the volume remaining in the pot at the time the water was boiled. During a boil, an amount of water is evaporated and depending on the observer, the mass missing by the time ‘boiling’ is achieved varies from test to test and from observer to observer. There is actually a website called “The Boiling Point Myth” which discusses the difficulty of trying to say when a pot is boiling.

One explanation offered as to why the final mass of water is used instead of the initial mass is that the final mass represents cooked food so they wanted to reflect the amount of cooked food created by the burned fuel. This is a conceptual divergence from the boiler performance test. One does not consider what the steam from a boiler is used for, only the performance raising that steam. The use of the final mass of water creates an interesting opportunity for miscalculation.

Rule 1: If two stoves have different performance, they must receive different performance ratings.

Rule 2: If two stoves have the same performance, they must receive the same performance ratings.

Rule 3: If one stove is comparatively ‘better’ than another, it must receive a ‘better’ performance rating.

Suppose two stoves have exactly the same fire burning in them consuming exactly the same mass of fuel. One has a higher heat transfer efficiency, and both observers subjectively conclude that the time to boil is exactly the same. The stove with the higher heat transfer efficiency will have evaporated more water by that time than the stove with the lower heat transfer efficiency. Calculating the SFC with the final mass of water means the more efficient stove has its fuel mass divided by a smaller number than the less efficient stove. This results in the more efficient stove having a higher fuel consumption rating than the less efficient stove which is the opposite of what was intended.

If the time to boil was slightly shorter for the more efficient stove, the final mass of water might have been slightly higher, the mass of fuel consumed slightly less, and the two stoves would be given the same rating even though their actual performance was different. It is rather obvious that the missing, evaporated water had first to be boiled so a tester could be forgiven for dividing the mass of fuel burned by the initial water mass thus getting the correct answer.

A great debate was started on the discussion list bioenergylists.org hosted by Tom Miles of Portland, Oregon. It centered on the fact that the charcoal remaining from a cooking session was often thrown away. Unless it could be reused in the same stove during the next cooking cycle, new fuel would have to be obtained. This created two schools of thought: those who wanted to know the actual amount of raw fuel required to conduct the test, and those who want to know the dry fuel equivalent of the energy used during the test. (We won't quibble about the combustion efficiency.) These two approaches cannot be reconciled. The stove developer wants to know if they has improved the heat transfer efficiency, and the programme managers want to know how much fuel the stove will use from the available supply. The WBT attempts to provide the former. No one was providing the latter.

After various protests, a technical committee convened by ETHOS[1] in 2008 produced an updated version of the WBT-4 which expressed the same performance by uses MJ instead of fuel mass in order to accommodate fuels with different energy values per kg. The SCF became MJ/litre boiled (still based on the final mass of water) and emissions expressed as PM2.5 mg/litre boiled.

Average performance

Remember that a choice had to be made between the 'Miles per gallon' or 'Litres per hundred kilometers' approach. It is 'tasks per fuel' vs 'fuel per task'. The WBT is fuel per task. The efficiencies both the cold start and hot start are expressed as a percentage with the numerator being the energy absorbed by the water (not including the pot) and the denominator being the net energy in the fuel consumed. The task, the energy used to boil the water, is expected to be the same and can be represented by 'A'. The denominator is now the fuel energy needed to complete the task and can be represented by 'B' for Cold and 'C' for Hot phases.

A/B   A/C

Remember that the original purpose of performing these two burns is because the answers for B and C are expected to be different. The outputs are efficiencies A/B and A/C expressed as %.

The website http://mathhelpforum.com/statistics/168063-ratios-what-do-they-mean-relative-each-other.html has this to say about averaging:

> Do not average averages
> Do not average %'s
> Do not average ratios

Efficiencies are ratios expressed as %. The averaging of two ratios where the task is the same *is* permitted, provided the denominator is constant, which is not the case. We must therefore

---

[1] Engineers in technical and humanitarian opportunities of service, based at Univ of Iowa, holds an annual stove-oriented event in or near Kirkland, WA.

calculate the harmonic mean. This does not confirm whether the cold/hot concept is valid in the first place, but at least it should be calculated correctly.

Low Power (Simmering) performance

The performance during the simmering section of the test is treated in the same manner as the boiling section. The reported metrics include the efficiency based on the amount of water missing, specific fuel consumption is given and emissions are calculated on a per litre basis.

It is at this point that the WBT departs from reality. Something that puzzled both VITA and Baldwin was that the efficiency during the simmering phase varied widely even for carefully performance tests and the number made no meaningfully accurate prediction of performance.

Simmering in WBT-4 is called ‘low power’ to disguise the fact that ‘simmering’, like ‘boiling’, is an undefined and therefore unscientific term. Following criticism, the name was change but the metrics remain. The efficiency of simmering is an oxymoron because simmering is a task that does not contain any requirement to perform work. By ‘simmering’ is meant ‘keep the water hot’. That could be accomplished by placing the hot pot between two pillows in a cardboard box using no fuel at all. The efficiency would be 1 divided by zero.

By applying the same approach used to determine the efficiency of heating and evaporating water, the simmering metric is determined by using the mass of water evaporated while attempting to keep the water in the pot hot. In other words, the attempt is to extract engineering efficiency measurements during a cooking-defined task that does not necessarily require any work to be performed. It requires providing a ratio answer (efficiency) to a yes/no nominative question (did you or did you not simmer the pot: Y/N) in order to provide a rank order rating of product performance (worse, good, better, best) or a performance ranking on a scale that does not include zero (out of 20 stoves tested it scored 4th place).

Here are the metrics:

1. $PM_{2.5}$, mg/litre/minute
2. CO, g/litre/minute
3. Specific fuel consumption/litre simmered (using the final volume of water in the denominator)
4. Efficiency (energy used to evaporate water, factored for water temperature change, divided by the energy in fuel burned, factored for charcoal remaining)

The efficiency number is again a proxy for the heat transfer efficiency but fails on two counts: it treats charcoal remaining as unburned raw fuel even if it is thrown away, and it is applying an energy efficiency metric to a task that has no requirement for a change in enthalpy. Its meaninglessness is demonstrated by the observation that a product that evaporated no water at all but burned some fuel to overcome convective and other losses from the pot. Such a ‘best possible stove’ performing a perfect simmer would be given, but this calculation, an energy efficiency of 0%.

Quite why the PM and CO are expressed ‘per minute’ is not clear as the simmering time is specified at 45 minutes in all cases so it is not a variable. Let’s not quibble.

The most interesting metrics are the emissions of pollutants ‘per litre’. Apart from repeating the water boiling error using the final mass of water as a divisor, the CO emission ‘per litre’ makes the assumption that the fuel needed to keep a pot ‘simmering’ is somehow affected by the mass of boiling water in it. This is not true.

When boiling, there is a clear relationship between the mass of fuel burned and the required change in enthalpy to bring the water to a boil. There is no such requirement when simmering as the water is already hot when the measurement starts. Let me explain.

It is commonly assumed that to keep a hot pot of water hot will require putting in more energy if there is more water in the pot. This is a conceptual error conflating heating water with keeping it hot. The matter was reported by Rani et al 1991 who said that his experiments show a very small increase in the need for heat if the pot was less than half full, but none if it was more. Zhang et al 2014 performed an elegant experiment using an induction cooker and a full pot of water. If the energy needed to keep a pot of water hot required less energy as the water mass dropped, then the evaporation rate would increase as the pot progressively lost water if the input power was held constant. They reported that the evaporation rate from ‘full pot’ to 15% full was a perfectly straight line (4 nines precision).

As the initial mass of water in the pot at the start of simmering varies from test to test, and the final mass varies even more, dividing the mass of CO emitted to keep the pot hot by either initial or final mass is pointless and the resulting numbers, highly variable. The relationship is between the pot and the fire, not between the fire and what is in the pot. The metric isn’t reporting anything. You may as well divide the fuel mass by ‘Tuesday’.

Even though the metric MJ/litre simmering has no physical basis, let us examine the influence of using the final volume of water as the divisor.

Consider the following two stoves. They are identical in every respect save heat transfer efficiency. During the two tests they start simmering with exactly the same mass of water and happen to consume exactly the same mass of fuel. Their total emissions are also identical. The only difference between them is that Stove A has a higher heat transfer efficiency that Stove B. As a result, by the end of the test it has evaporated more water than stove B. When the specific energy consumption is determined, the energy in the mass of fuel consumed will be divided by a smaller water mass than Stove B. Stove A will be rated as having consumed more energy per litre simmered than Stove B, even though it is more efficient and consumed the same amount of fuel. This violates rules 1 and 3 above.

Thus it is demonstrated that the three Low Power metrics for fuel consumption, CO and PM2.5 emissions ‘per litre’ have no physical basis. If you want a better rating, be sure the pot is fuller than the competition. You can then divide your energy consumption number by a larger number of litres!

Prediction

One of the more disturbing aspects of the WBT is its inability to predict performance in use. Stoves are given ratings and when checked against field performance there is basically no correlation between the two data sets. The reasons are:

- The burn cycle is not typical of use in the target community
- The pot and pot loads are not typical of use
- The fuel may be dissimilar to that used in the target community
- The operating method may be different
- The metrics are poorly chosen or baseless and do not reflect current performance, let alone future performance
- There are misconceptions embedded in the protocol such as the assumption that the charcoal remaining from the cold start and hot start will be the same, even though there is an expectation that the mass of fuel burned will be different
- There are systematic errors in the calculations
- There are common metric names used in lab and field tests but the tests define them differently
- There is concatenations of errors in the calculations which result in a higher variability in the performance rating than is present in the raw data from the test.

My conclusion is that the following metrics have no physical basis and are worthless for rating stove performance:

- Specific energy consumption, MJ per litre (simmered)
- PM emitted, mg/litre/minute (simmered)
- CO emitted, g/litre/minute (simmered)

Even if all these protocol problems were removed by developing a culturally relevant, correctly conceived performance test, there remains a significant problem in the measurement of particles emitted from stoves. I wrote to TSI, the manufacturers of the Dusttrak particle counter, asking if we should consider the detection of condensed water droplets in a gas sample as a confounding issue. They replied with a technical note which said that if the humidity is 70%, water droplets suspended in the air (fog) will be detected by the scattering of laser light, and that we should reduce the mass detected by 62%. This confirms that light scattering instruments are recording water droplets as smoke particles and wet smoke particles as large-than-they-are particles.

When conducting an emissions test using a hood and fan system such as EPA Method 5G there is a significant increase in the PM detected by light scattering instruments. Because stoves burning wood create a large number of condensed volatile particles, the errors inherent in the TEOM system caused by the evaporation of condensates means that filter mass is not a perfect option either. If this subject is opaque to you let me briefly explain.

Burning moist wood in stove that has a boiling pot on top on a hot humid day creates conditions in the hood & fan system perfect for condensing water vapour – large amount of water vapour. Because the development of particle measurement from combustors was largely based on dry dust, the protocols and equipment are not well suited to measuring ‘particles of interest’ from damp wood. Boiling water into the same hood makes matters worse. That can be avoided by channelling the steam out of the hood, but the problem remains and is large. The TEOM vibrates a filter and monitors the mass change in real time. The criticism of it is that the accumulated mass can reduce with time as condensed volatiles evaporate from it. The same thing applies to stationary filters but is not noticed because they are weighed afterwards. It happens that with wood smoke, it is the liquid particles which present a great portion of the health risk and they are handled poorly. At least with lasers they are visible and can be counted.

To address this I have developed a system with variable (not fixed) dilution using dry CO_2-free air as a dilutant that dehydrates the sample before the water vapour condenses. It uses the CO_2 from combustion as a tracer gas to determine the level of dilution. This system largely prevents the condensation of water vapour while still condensing most of the volatiles with a boiling point higher than water. This system has been installed in Mongolia and Indonesia. Version without dry CO_2-free air have been in use at the Univ of Johannesburg and LBNL.

Thank you everyone.

Questions.

Abstract

It is rare that at the ANSI or ISO level a performance metric with no physical basis is proposed for rating product performance. The talk explores popular performance tests used in the improved cookstove testing community from the inventor's garage up to the UNFCCC. It will explore the evolution of a reasonable proxy for cooking performance from VITA – the boiling and simmering of a pot of water – as it mutates from a simple assessment of fuel consumption through a series of increasingly complex calculations that attempt to extract engineering measurements from a socially defined task. It is a journey from artisanal knowhow to scientific nowhere.

It is a case study of how valid engineering concepts can become skewed when real world use is modelled, then remodeled again and again. Each calculation step added to an analysis increases the danger that the methodology will lose sight of the real world and what the customer wanted to know. In one of the most popular versions, the compounding of a series of small systematic errors, each based on a conceptual error, each leading the observer further astray, produces misleading results with higher variability than the raw data and virtually no predictive value – the very antithesis of good science.

Some technical issues related to the measurement of particles (which will literally be about blowing smoke into a diluter with dry, CO_2-free air) show that even with valid methods and metrics, formidable challenges remain when measuring smoke emissions.

+++++++

Blowing Smoke: The Curious Case of the Mangled Metric

Why improved cookstove performance tests are mostly hot air.

- Crispin Pemberton-Pigott
    Principal Technical Advisor, Sustainable Energy Technology and Research Center (SeTAR), Research Village, Bunting Road Campus, University of Johannesburg, South Africa.

"All history suggests that neither authority nor the mind is sufficient if it works outside a restraining context of general purpose." Elting Morrison, *From Know-how to Nowhere*, 1977, pp. xii