POLITECNICO
MILANO 1863

Guidelines for
**REPORTING AND ANALYSING
LABORATORY TEST RESULTS**
for biomass cooking stoves

Guidelines for reporting and analysing laboratory test results for biomass cooking stoves

January 2017 – Version 1.0

Table of contents

Acronyms

AWBT	Adapted Water Boiling Test
BIS	Bureau of Indian Standards
COV	Coefficient of Variation
CS	Chinese Standard
EPTP	Emissions & Performance Test Protocol
HTP	Heterogeneous Testing Procedure
ICS	Improved Cooking Stove
IEA	International Energy Agency
ISO	International Organization for Standardization
GACC	Global Alliance for Clean Cookstoves
WBT	Water Boiling Test
WHT	Water Heating Test

Nomenclature

$\bar{X}$ Average value of a selected parameter for a given set of test replicates

U_e Extended uncertainty referred to a selected parameter

$1-\propto$ Level of statistical confidence (c..l. %)

n Number of test replicates performed

S_n Standard deviation for a sample of n replicates

$t_{(1-\frac{\alpha}{2},n-1)}$ Two-tail coverage factor based on a t-student distribution

Introduction

Energy and sustainable development are deeply linked, especially in low-income countries where energy availability is essential to promote socio-economic growth, to preserve the environment and to foster social inclusion. This is recognised within the 2030 Agenda for Sustainable Development [3] and the Sustainable Development Goals (SDG 7) [4]. Energy demand has grown impressively in many low-income countries over the last 2 decades, but access to modern energy services, though increasing, remains often limited and represents a barrier to development [5,6].

Indeed, according to the most recent estimates by the International Energy Agency (IEA), 1.8 billion people in 2040 will still rely on the so-called "traditional use of biomass" – i.e. the combustion of wood, charcoal, animal dung or crop waste by means of open fires or traditional and inefficient stoves – to satisfy their domestic energy needs [7]. This practice is increasingly recognised as a critical issue, and many efforts worldwide are devoted to better understanding its consequences on health, environment and social development [7–9] as well as to the promotion of technical solutions, including biomass stoves, or Improved Cooking Stoves (ICSs).

In this framework, the evaluation of ICSs performance is particularly critical and different actors worldwide are focusing on this topic. One of the key challenges consists in improving the common practice in reporting and analysing test results: as a matter of fat, a large amount of the reports or studies in the literature provide results that do not allow drawing any statistically significant conclusion – as further discussed in Section 1 [1]. As a consequence, non-technical actors may be led to possible misinterpretations of the results.

The Guidelines are conceived as a support to all the actors involved in the sector of biomass stoves performance evaluation, from the technicians and researchers engaged in laboratory testing, to those who need to better understand and interpret test results in order to select a promising stove model for field trials. The concepts and the methodology here proposed draw upon the most recent studies in the scientific literature on this topic [1,2].

Section 1 provides an overview of the framework of laboratory tests and of the common criticalities in the interpretation and reporting of test results. Section 2 and Section 3 provide a robust methodology for the interpretation and the analysis of test results. In particular, Section 2 discusses the minimum number of test replicates to be performed and how to report the results for a single stove model. Section 3, instead, explains how to deal with comparisons between two stove models and discusses the criteria to statistically infer that a stove model is improved as compared to a baseline reference.

1. The framework of Improved Cooking Stoves laboratory testing

1.1. Testing protocols

Laboratory tests are traditionally conceived as a tool for stove developers to evaluate changes in ICSs performance due to different designs and features or to perform a preliminary technology assessment [10–12]. The most common and widespread laboratory-based testing protocol for ICSs testing is the Water Boiling Test (WBT), first published in 1985 and updated several times; the present version 4.2.3 was released in 2014 and includes contributions from the Global Alliance for Clean Cookstoves (GACC) [10]. However, a number of alternative protocols has been developed and adopted by national authorities or independent organisations over the years. All the officially published protocols are summarised in Table 1.

Protocol name	Water Boiling Test	Emission & Performance Test Protocol	Heterogeneous Testing Procedure	Adapted Water Boiling Test	Chinese Standard	Indian Standard on Solid Biomass Chulha-Specification
Acronym	WBT	EPTP	HTP	AWBT	CS	BIS

Table 1. Summary of all the laboratory-based ICSs testing protocols officially published to date [10,13–17]. Further details about the protocols can be found in their respective documents, available from the GACC website.

Currently, the International Organization for Standardization (ISO) Technical Committee 285 is trying to harmonize the different existing methodologies and the contributions coming from the scientific literature into a new international standard for ICSs performance evaluation [18]. The process is still in progress.

1.2. Common criticalities in reporting and interpreting test results

Regardless of the protocol employed, all ICSs testers have to deal with issues of variability between testing replicates due to the intrinsic variability of the process of biomass combustion [19]. Furthermore, tester's discretion, thermodynamic variabilities and variable ambient conditions may also contribute to increase the variability of the results [19–21]. Accordingly, though some protocols are specifically designed to reduce at least the most easily controllable sources of variability, the accuracy of the results should be always carefully evaluated by the testers. Most recent studies indicate that large sets of replicates might be required to obtain a reliable data bank [2].

The common practice, instead, consists in performing only a very limited number of test replicates (usually three) and in reporting the average of the three as the final result, without even providing any information about the statistical significance of the latter [1]. This way of reporting test results, may lead to misinterpretations [1].

On the one hand, if the results are provided without any confidence interval attached, it is impossible to draw any statistically significant inference about the "improvement" of the ICS as compared to a baseline model. For example, let's consider a comparison between two stoves (Stove A and Stove B) in terms of "Time to boil" (as defined by the WBT [10]). Figure 1-*left side* shows that the average

Time to boil is lower for Stove B. One may **erroneously conclude that** *"Stove B performs better than Stove A in terms of Time to boil"*. Indeed, Figure 1-*right side* shows that, if confidence intervals for the results are provided, there is overlapping between the two. An appropriate statistical test would be needed to verify if there is a significant improvement.
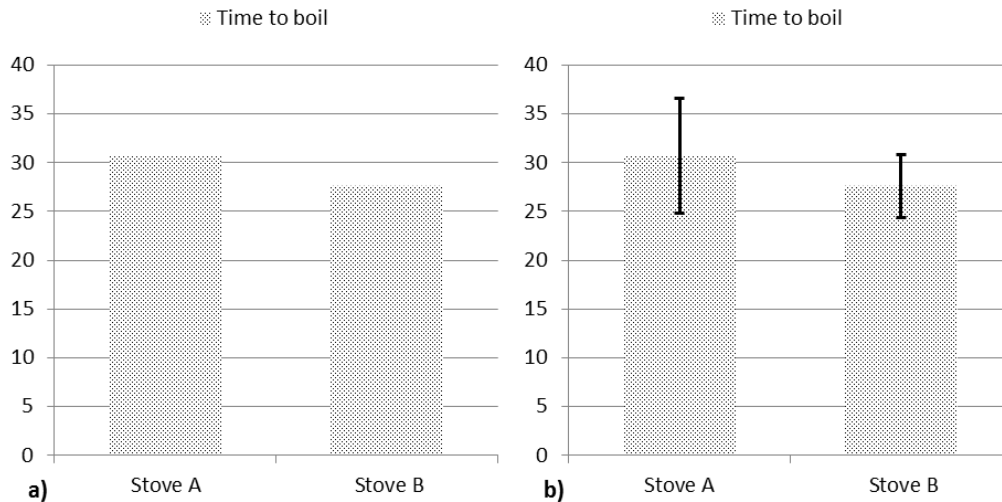

Figure 1. Example of comparison between two stoves in terms of Time to boil.

On the other hand, even when the statistical analysis is included in the final report, statistical inferences based on a three-replicates variance may be strongly biased [1,2]. As a matter of fact, the calculation of the standard deviation is a fundamental prerequisite to perform any statistical analysis, and therefore to draw statistically significant conclusions. However, recent studies show that the standard deviation calculated on a three-replicate basis may be unreliable [2]. For example, it could be relatively small just to increase once more replicates are added. A **minimum number of five test replicates** seem to be necessary to avoid major biases in the estimation of the standard deviation [1,2]. A concrete example is provided in Table 2.

	Time to boil	
Number of replicates	**Average [min]**	**Standard Deviation [min]**
3	18.6	1.4
5	17.6	1.9

Table 2. Average value and standard deviation referred to the WBT metric "time to boil", computed on a three-replicates and on a five-replicates basis. The three-replicates case underestimates the variance compared to the five-replicates case.

The larger the set of replicates, the more reliable will be the data bank. Of course there are often trade-offs between the number of replicates and the duration and cost of the testing campaign. Therefore, performing a large number of replicates (>20), as suggested by some authors [2], may not always be feasible. The present guideline provides a robust methodology to perform the analysis of testing results in case of both large and small sample sizes.

2. A methodology for reporting and analysing test results

2.1. Conceptual scheme

The proposed methodological path is summarised in the logical scheme in Figure 2, which should also serve as a guide for the tester. Each step is discussed in detailed in the following sub-sections.

Figure 2. Logical scheme of the proposed methodology for reporting test results. The final output includes the uncertainty and the level of confidence.

2.2   Number of test replicates and reporting format

At least <u>five test replicates</u>[1] shall be performed. When possible, the tester is encouraged to perform larger sets of replicates ($\geq$ 8) to ensure a greater accuracy in the calculation of the standard deviation.

The results of each test replicate shall be always reported separately and then averaged in the following form:

$$\boxed{\overline{X} \pm U_e \; (c.l.\,\%)}$$

Where:

  - $\overline{X}$ is the average value of a selected parameter for a given set of test replicates;
  - U_e is the extended uncertainty referred to that parameter, to be calculated for 90% and 95% confidence levels (c.l.).

The interval $\left[\overline{X} - U_e ; \overline{X} + U_e\right]$ constitutes the *confidence interval* of the selected parameter, for a given level of statistical confidence.

2.3   Calculation of the uncertainty for a selected level of statistical confidence

The common way to calculate U_e is based on the *t-student* approach. Nevertheless, this method is rigorously valid only for normally distributed data sets (Case A). Different methods are here prescribed for non-normally distributed data sets (Case B). Accordingly, before calculating U_e, the tester shall verify the normality of the data set by means of a normality test. The Shapiro-Wilk normality test is suggested for small sample sizes (less than 20 replicates).

Step 1 – Verifying the normality of the data set

Perform a Shapiro-Wilk test on the selected data set. The normality hypothesis is not rejected if the *p-value* is greater than 0.1; otherwise, there is significant evidence that the distribution is not normal.

> *Statistical software typically includes specific functions to perform the Shapiro-Wilk test automatically. The Excel environment, instead, does not include any specific function, but the required formulas can be implemented into the spreadsheet. Free online tools are available as well (e.g. http://scistatcalc.blogspot.it/2013/10/shapiro-wilk-test-calculator.html). The software choice is left to the tester's discretion.*

Step 2 – Calculating the uncertainty

a) **Case A** – <u>If the normality hypothesis is not rejected</u> (Shapiro-Wilk's p-value > 0.1), the uncertainty shall be calculated as follows:

---

[1] A sample size smaller than 5 may bias further statistical tests for the assessment of data distribution.

$$U_e = t_{(1-\frac{\alpha}{2},n-1)} \cdot \frac{S_n}{\sqrt{n}}$$

Where:

- $(1 - \alpha)$ is the confidence level (c.l.) – e.g. $\alpha = 0.05, \ c.l. = 0.95$;
- n is the largeness of the data set, i.e. the number of replicates performed;
- $t_{(1-\frac{\alpha}{2},n-1)}$ is the two-tail coverage factor based on a t-student distribution, for n replicates and α level of significance;
- S_n is the standard deviation of the sample.

The coverage factor can be derived from common t-student's statistical tables. Alternatively, Excel includes a pre-defined function (CONFIDENCE.T) allowing to immediately calculate U_e for a selected level of confidence, based on the standard deviation of the data set.

Example:

n=6

Example metric	*units*	Value_1	Value_2	Value_3	Value_4	Value_5	Value_6
Thermal Efficiency	%	19,3%	20,0%	20,1%	17,1%	18,8%	18,1%

Step 1

Shapiro-Wilk test for normality	
p-value	0,6492
Reject normality?	no

Step 2 - t-student

Average	SD	U (CI.95%)	U (CI.90%)
18,9%	1,2%	1,2%	1,0%

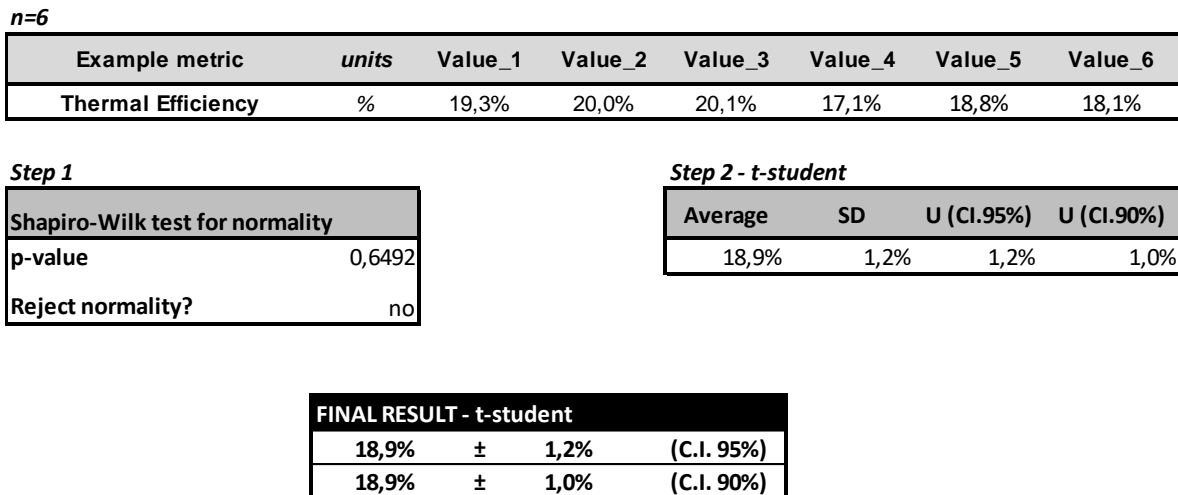| FINAL RESULT - t-student			
18,9%	±	1,2%	(C.I. 95%)
18,9%	±	1,0%	(C.I. 90%)

Figure 3. Example of calculation of the extended uncertainty and final result for the WBT metric "Thermal Efficiency", based on a six-replicates data set. In this case, the Shapiro-Wilk test didn't reject the normality hypothesis and the t-student approach was employed to calculate the extended uncertainty.

b) **Case B** – <u>If the normality hypothesis is rejected</u> (Shapiro-Wilk's p-value ≤ 0.1), the tester shall chose one of the following options:

b.1) Increase the number of replicates until Shapiro-Wilk's p-value is > 0.1 and then come back to **Case A**;

b.2) Rely on a different approach to provide the extended uncertainty, i.e. the more conservative *Chebyshev's inequality*. In this case:

$$U_e = \sqrt{\frac{1}{\alpha}} S_n$$

Where:

- $1 - \alpha$ is the confidence level (c.l.) – e.g. $\alpha = 0.05, \ c.l. = 0.95$;
- S_n is the standard deviation of the sample.

Example:

n=5

Example metric	units	Value_1	Value_2	Value_3	Value_4	Value_5
Time to boil	min	17,8	17,8	17,8	14,3	14,7

Step 1

Shapiro-Wilk test for normality	
p-value	0,0209
Reject normality?	Yes

Step 2 - Chebyshev

Average	SD	U (CI.95%)	U (CI.90%)
16,47	1,82	8,13	5,75

FINAL RESULT - Chebyshev			
16,47	±	8,13	(C.I. 95%)
16,47	±	5,75	(C.I. 90%)

Figure 4. Example of calculation of the extended uncertainty and final result for the WBT metric "time to boil", based on a five-replicates data set. In this case, the Shapiro-Wilk test rejected the normality hypothesis and the Chebyshev's inequality approach was employed to calculate the extended uncertainty.

As shown by the example in Figure 4, the confidence intervals built on the Chebyshev's inequality are considerably larger than those obtainable from the t-student approach. Indeed, this approach provides more conservative results as there is no information about data distribution.

2.4   Results acceptability

The results are considered acceptable **only if** the absolute value of the extended uncertainty is smaller than the average value of the selected metric ($U_e < \bar{X}$). This condition might not be always verified, in particular when applying the Chebyshev's inequality for typical levels of statistical confidence. In general, when the condition does not hold true, the variability of the data set might be too large. Accordingly, the tester shall try to reduce the data variability by increasing the sample size and by checking the presence of experimental errors.

3   A methodology for comparing the performance of different stove models

3.2   Conceptual scheme

In most cases, the tester's goal will be also to compare two test series, i.e. two stove models, to assess which one has a better performance. This should be always the case when the aim of the tester is to assess the "improvement" of a selected stove model as compared to a baseline technology (three-stone fire or others). Figure 5 shows the logical path to be followed in this case.
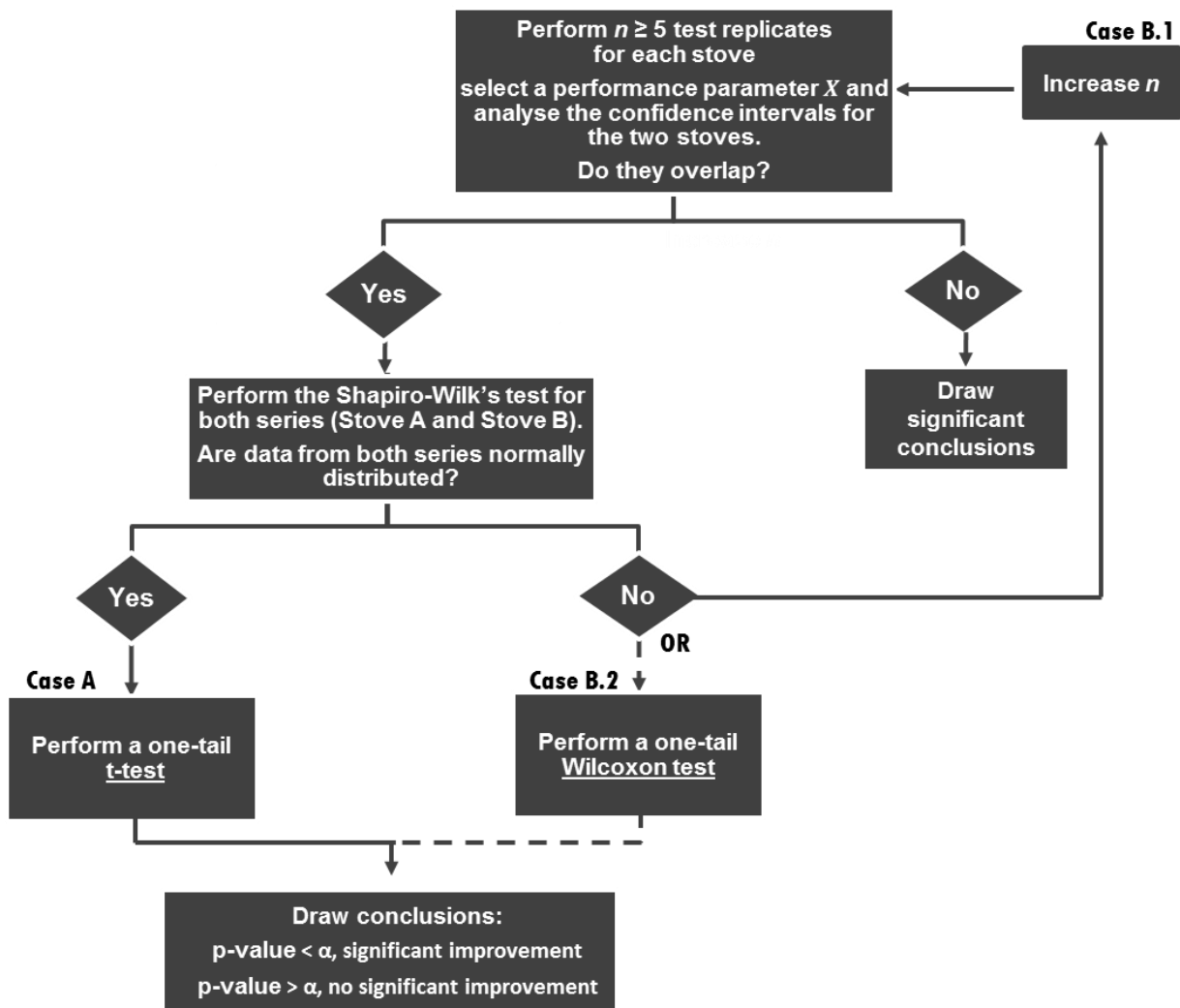

Figure 5. Logical scheme of the proposed methodology for comparing the performance of two different stove models. If there is no overlapping between the confidence intervals, significant conclusions can be easily drawn. Otherwise, a proper statistical test is needed.

3.3 Number of test replicates

An identical number of test replicates <u>shall be conducted for each stove model</u>, following the indications provided in sub-section 3.1 – i.e. at least 5 test replicates for each stove model. When possible, the tester is encouraged to perform larger sets of replicates ($\geq$ 8) to ensure a greater accuracy in the calculation of the standard deviation.

A performance parameter for comparison shall be selected, and the extended uncertainty shall be calculated for each data series (i.e. for each stove model):

$$\overline{X}_A \pm U_{e,A} \ (c.l.\%)$$
$$\overline{X}_B \pm U_{e,B} \ (c.l.\%)$$

Where:

- $\overline{X}_A, \overline{X}_B$ are the average values of a selected parameter for the two series of test replicates (Stove A and Stove B)
- $U_{e,A}, U_{e,B}$ are the extended uncertainties referred to that parameter, to be calculated for 90% and 95% confidence levels (c.l.).

The intervals $\left[\overline{X}_A - U_{e,A}; X_A + U_{e,A}\right]$ and $\left[\overline{X}_B - U_{e,B}; X_B + U_{e,B}\right]$ constitute the *confidence intervals* of the selected parameter, for a given level of statistical confidence.

For the details about how to calculate the extended uncertainty refer to sub-section 2.3.

3.4 Criteria for drawing statistically significant inferences

It is always possible to assert that a stove performs significantly better than another one, in relationship to a selected parameter (e.g. Thermal Efficiency), **if the confidence intervals for the two parameters do not overlap.** For example, the results in Figure 6 satisfy this condition: it is possible to assert that "*the Thermal Efficiency of Stove B is greater than it is for Stove A, with a 95% level of confidence*" (as the bars here represent the confidence intervals for a 95% level of confidence).

In most cases, however, the tester has to deal with **overlapping confidence intervals,** as shown in Figure 7. In this situation, **it is necessary to perform a proper statistical test before drawing any conclusion.** The test may still show that there is a statistically significant difference between the two values, despite the overlapping of confidence intervals. In order to perform a correct statistical analysis, the tester shall follow the logical path proposed in Figure 5 and identify the proper statistical test.
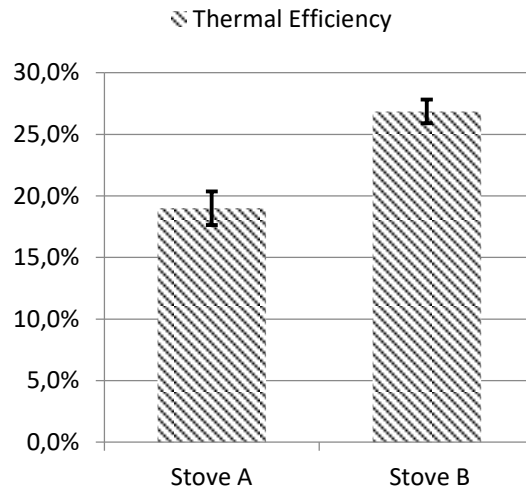
Thermal Efficiency

Figure 6. Thermal efficiency comparison between Stove A and Stove B. Bars represent the confidence interval for a 95% level of confidence. There is no overlapping between the confidence intervals.
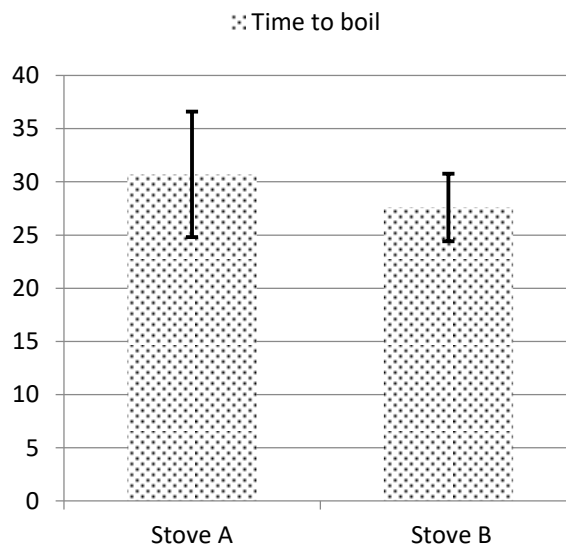
Time to boil

Figure 7. Time to boil comparison between Stove A and Stove B. Bars represent the confidence interval for a 95% level of confidence. The two intervals overlap and is not possible to draw any statistically significant conclusions from the observation of the graph; a statistical test is needed.

3.3.1 Procedure to select and perform a statistical test

Step 1 – Verifying the normality of the data set

A Shapiro-Wilk test should have been already performed on both data series to calculate the confidence intervals. For any details about how to perform the Shapiro-Wilk test refer to sub-section 2.3. The normality hypothesis is not rejected if the *p-value* is greater than 0.1; otherwise, there is significant evidence that the distribution is not normal.

Step 2 – Performing a proper statistical test

a) **Case A** – <u>If the normality hypothesis is not rejected for any of the two series</u> (Shapiro-Wilk's p-value > 0.1), the mean values can be compared relying on a *t-test*.

> *Statistical software typically includes specific functions to perform the t-test automatically. The Excel environment includes the T.TEST function as well as the Data Analysis tool to perform the test. Free online tools are available as well. The software choice is left to the tester's discretion.*

> *In any case, the selected software or tool will require to choose whether to perform a one-tail or a two-tail test (or else it will provide the result for both cases leaving the interpretation to the tester); since the aim is to verify the presence of a significant improvement, rather than just a significant difference, a **one-tail** t-test shall be performed.*

> *If the one-tail t-test's p-value is lower than 0.1, 0.05 or 0.01, it is possible to conclude, respectively, with a 90%, 95% or 99% level of confidence that one stove performs better than the other, for the selected parameter.*

Example (using the Excel function T.TEST to calculate the p-value of the t-test):

n=5

Example metric	*units*	*Stove*	Value_1	Value_2	Value_3	Value_4	Value_5	Average
Time to boil	*min*	*A*	35,5	26,9	30,7	25,1	35,3	**30,7**
		B	25,7	30,9	24,8	29,5	27,1	**27,6**

Step 1

Shapiro-Wilk test for normality		
Stove A	**p-value**	0,236
	Reject normality?	**No**
Stove B	**p-value**	0,694
	Reject normality?	**No**

Step 2

t-test	
p-value	0,122
Time(B) < Time(A)?	**No**
Level of confidence	-

Figure 8. Example of comparison between the time to boil for two stove models (A and B). The confidence intervals overlap, so a proper statistical test is needed to draw any conclusion. First, the normality of the two series is checked by the Shapiro-Wilk's test. Since the normality hypothesis is not rejected, it is possible to perform a t-test, in this case performed through the Excel function "T.TEST". The p-value returned by the function is greater than 0.1, so it is not possible to draw any statistically significant conclusion about which stove has the best performance in terms of Time to boil, despite the difference in the average values.

In the example, the obtained p-value is greater than the minimum acceptable level of significance of the test, i.e. 0.1 (corresponding to a 90% level of confidence). Accordingly, the test was not helpful in solving the issue of the overlapping confidence intervals and it is not possible to draw any statistically significant conclusion. **It would be erroneous to say that** "*Stove B has a statistically significant better performance than Stove A, in terms of Time to boil*".

b) **Case B** – <u>If the normality hypothesis is rejected for at least one series</u> (Shapiro-Wilk's p-value ≤ 0.1), the tester shall chose one of the following options:

> **b.1)** Increase the number of replicates until Shapiro-Wilk's p-value is > 0.1 and then come back to **Case A**;

b.2) Rely on a non-parametrical test, *viz.* the *Wilcoxon rank-sum test* (also known as Wilcoxon–Mann–Whitney test or Mann–Whitney U test).

Statistical software typically includes specific functions to perform the Wilcoxon test automatically. The Excel environment, instead, does not include any specific function, but the required formulas can be implemented into the spreadsheet. The software choice is left to the tester's discretion. Free online tools are available as well (e.g. https://ccb-compute2.cs.uni-saarland.de/wtest).

*In any case, the selected software or tool will require to choose whether to perform a one-tail or a two-tail test (or else it will provide the result for both cases leaving the interpretation to the tester); since the aim is to verify the presence of a significant improvement, rather than just a significant difference, a **one-tail** t-test shall be performed.*

If the one-tail Wilcoxon rank-sum test's p-value is lower than 0.1, 0.05 or 0.01, it is possible to conclude, respectively, with a 90%, 95% or 99% level of confidence that one stove performs better than the other, for the selected parameter.

n=5

Example metric	*units*	*Stove*	Value_1	Value_2	Value_3	Value_4	Value_5	Average
Time to boil	*min*	*A*	35,3	26,6	30,7	32,4	36,3	**32,2**
		B	24,3	22,4	35,7	24,5	25,7	**26,5**

Step 1

Shapiro-Wilk test for normality		
Stove A	**p-value**	0,035
	Reject normality?	**Yes**
Stove B	**p-value**	0,717
	Reject normality?	**No**

Step 2

Wilcoxon test	
p-value	0,047
Time(B) < Time(A)?	**Yes**
Level of confidence	**95%**

Figure 9. Example of comparison between the time to boil for two stove models (A and B). The confidence intervals overlap, so a proper statistical test is needed to draw any conclusion. First, the normality of the two series is checked by the Shapiro-Wilk's test. Since the normality hypothesis is rejected for at least one series, it is not possible to perform a t-test. A Wilcoxon rank-sum test is performed by means of the online tool, and the p-value obtained is lower than 0.05, so it possible to conclude that "*Stove B has a statistically significant better performance than Stove A, in terms of Time to boil, with a level of confidence of 95%*".

In the example, the obtained p-value is lower than the minimum acceptable level of significance of the test, i.e. 0.1 (corresponding to a 90% level of confidence) and it is also lower than 0.05 (corresponding to a 95% level of confidence). Accordingly, the test was helpful in solving the issue of the overlapping confidence intervals and it is possible to draw a statistically significant conclusion. **It is possible to conclude that** "*Stove B has a statistically significant better performance than Stove A, in terms of Time to boil, with a level of confidence of 95%*".

References

[1]   F. Riva, F. Lombardi, C. Pavarini, E. Colombo, Fuzzy interval propagation of uncertainties in experimental analysis for improved and traditional three – Stone fire cookstoves, Sustain. Energy Technol. Assessments. 18 (2016) 59–68. doi:10.1016/j.seta.2016.09.007.

[2]   Y. Wang, M.D. Sohn, Y. Wang, K.M. Lask, T.W. Kirchstetter, A.J. Gadgil, How many replicate tests are needed to test cookstove performance and emissions? — Three is not always adequate, Energy Sustain. Dev. 20 (2014) 21–29. doi:10.1016/j.esd.2014.02.002.

[3]   United Nations, Transforming our world: the 2030 Agenda for Sustainable Development, 2015. doi:10.1007/s13398-014-0173-7.2.

[4]   United Nations, Sustainable Development Goals - Goal 7, 2015. (n.d.). http://www.un.org/sustainabledevelopment/energy/.

[5]   S. Mandelli, J. Barbieri, L. Mattarolo, E. Colombo, Sustainable energy in Africa: A comprehensive data and policies review, Renew. Sustain. Energy Rev. 37 (2014) 656–686. doi:10.1016/j.rser.2014.05.069.

[6]   E. Colombo, S. Bologna, D. Masera, Renewable Energy for Unleashing Sustainable Development, 2013. doi:10.5505/planlama.2013.14633.

[7]   International Energy Agency, Energy and Air Pollution, World Energy Outlook - Spec. Rep. (2016) 266. doi:10.1021/ac00256a010.

[8]   R. Bailis, R. Drigo, A. Ghilardi, O. Masera, The carbon footprint of traditional woodfuels, Nat. Clim. Chang. 5 (2015) 266–272. doi:10.1038/nclimate2491.

[9]   S. Abdelnour, A.M. Saeed, Technologizing humanitarian space: Darfur advocacy and the rape-stove panacea, Int. Polit. Sociol. 8 (2014) 145–163. doi:10.1111/ips.12049.

[10]   The Water Boiling Test 4.2.3, (2014). https://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/399-1.pdf (accessed July 11, 2016).

[11]   K.B. Sutar, S. Kohli, M.R. Ravi, A. Ray, Biomass cookstoves: A review of technical aspects, Renew. Sustain. Energy Rev. 41 (2015) 1128–1166. doi:10.1016/j.rser.2014.09.003.

[12]   P. Arora, S. Jain, A review of chronological development in cookstove assessment methods: Challenges and way forward, Renew. Sustain. Energy Rev. 55 (2016) 203–220. doi:10.1016/j.rser.2015.10.142.

[13]   M. DeFoort, C. L'Orange, C. Kreutzer, Stove Manufacturers Emissions & Performance Test Protocol (EPTP), 2009. http://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/75-1.pdf.

[14]   SeTAR Centre, The Heterogeneous Testing Procedure For Thermal Performance and Trace Gas Emissions, Johannesburg, 2012. http://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/74-1.pdf.

[15]   GERES Cambodia, The Adapted Water Boiling Test, 2010. http://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/75-1.pdf.

[16]   Bureau of Technical Supervision Beijing, General specifications for biomass household stoves-

China, 2008. http://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/77-1.pdf.

[17] Cis 1315 Z, Indian standard on solid biomass chulha specification, (1991) 96–104.

[18] Z. Zhang, Y. Zhang, Y. Zhou, R. Ahmad, Systematic and conceptual errors in standards and protocols for thermal performance of biomass stoves, Renew. Sustain. Energy Rev. (2016) 1–12. doi:10.1016/j.rser.2016.10.037.

[19] C. L'Orange, M. DeFoort, B. Willson, Influence of testing parameters on biomass stove performance and development of an improved testing protocol, Energy Sustain. Dev. 16 (2012) 3–12. doi:10.1016/j.esd.2011.10.008.

[20] S.. Bhattacharya, D.. Albina, A. Myint Khaing, Effects of selected parameters on performance and emission of biomass-fired cookstoves, Biomass and Bioenergy. 23 (2002) 387–395. doi:10.1016/S0961-9534(02)00062-4.

[21] C.M. Quist, R.B. Jones, M.R. Jones, R.S. Lewis, Uncertainty analysis and design guidelines of biomass cookstove thermal effi ciency studies, Energy Sustain. Dev. 34 (2016) 54–61. doi:10.1016/j.esd.2016.07.001.