

New Software for the Statistical Analysis and Qualification of NIST Charpy Verification Specimen Lots

Enrico Lucon



This publication is available free of charge from:
<https://doi.org/10.6028/NIST.IR.8211-upd>
Includes updates as of November 2021, see Appendix D

NISTIR 8211

New Software for the Statistical Analysis and Qualification of NIST Charpy Verification Specimen Lots

Enrico Lucon
*Applied Chemicals and Materials Division
Material Measurement Laboratory*

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.IR.8211-upd>
Includes updates as of November 2021, see Appendix D

April 2018



U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Under Secretary of Commerce for Standards and Technology

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

National Institute of Standards and Technology Internal Report 8211
Natl. Inst. Stand. Technol. Int. Report 8211, 25 pages (April 2018)

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.IR.8211-upd>
Includes updates as of November 2021, see Appendix D

Abstract

New software, in the form of Microsoft Excel 2016^a spreadsheets, has been created for the analysis and qualification of Charpy Verification Lots, to be used by personnel of the Charpy Verification Program at NIST in Boulder, Colorado. The new spreadsheets replace the previously used statistical software (in SYSTAT and R-code), adding a more detailed outlier analysis.

Of the two spreadsheet templates, the first one assesses the acceptability of a Charpy pilot or production lot by calculating its sample size and identifying statistical outliers. The second spreadsheet establishes the reference value of absorbed energy and the associated uncertainty for the whole Charpy Verification Lot.

Both spreadsheets have been checked and validated against calculations performed on several lots by means of the previously used R-language codes, which had been prepared by Jolene Splett of the Statistical Engineering Division of NIST.

Keywords

Analysis software, Charpy Verification Lots, Charpy Verification Program, outlier analysis, pilot lot, production lot.

^a Commercial names are identified in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the NIST, nor does it imply that they are necessarily the best available for the purpose.

Table of Contents

1.	Introduction.....	1
2.	Qualification of a Pilot or Production Lot	2
3.	Establishment of Certified Absorbed Energy (KV_{ref}) and Associated Uncertainties	4
4.	Software Used at NIST for the Analysis and Qualification of Charpy Verification Lots ...	6
	References.....	12
	Appendix A – Charpy ASCII Data File.....	13
	Appendix B – Printed Output of Template <i>Charpy lot statistical analyses</i>	14
	Appendix C – Printed Output of Template <i>Reference & Uncertainties Charpy Lots.xltn</i>	18
	Appendix D – Change Log	20

1. Introduction

According to both ASTM E23 [1] and ISO 148-2 [2] standards, the verification of an impact machine used to perform Charpy tests consists of a physical part (direct verification) and an engineering/experimental part (indirect verification).

The direct verification of a Charpy machine entails performing a detailed evaluation of machine dimensions, perpendicularity/parallelism/alignment of individual machine components (to be periodically verified or just upon machine installation), etc., as well as determining physical machine parameters such as potential energy, impact velocity, and friction losses.

The indirect verification of a Charpy machine entails testing reference specimens with certified absorbed energy values. According to ASTM E23, such certified energy values can only be established “*on the three reference machines owned, maintained, and operated by NIST in Boulder, CO.*”

Starting in the late 1960s, the U.S. Army (Watertown Arsenal, AMMRC) started producing and distributing Charpy specimens for the indirect verification of impact machines in the United States, in accordance with ASTM E23 [3]. In 1989, the Army decided to discontinue the program, which was taken over by NIST. Army personnel helped to transfer the three original “master machines” (currently “reference machines”) and their evaluation procedures to Boulder, Colorado. Two of the original Army machines (identified herein as “TK” and “TO”) are still being used today to certify Charpy Verification Lots at NIST, whereas the third one was replaced by a higher-capacity machine from the same manufacturer in the early 2000s. It is herein identified as “SI”.

Currently, the NIST Charpy Verification Program in Boulder evaluates verification test results for more than 1000 industrial machines located in more than 60 countries across the world. The criterion that has to be fulfilled for a machine to be in compliance with ASTM E23 is based upon the agreement between the customer’s average results and NIST-certified absorbed energy values within the lesser of 1.4 J or 5 %. ISO 148-2 has different, and more lenient, acceptability criteria, which are based upon both the difference between customer’s average and reference energy (*bias*) and customer’s test result scatter (*repeatability*).

For a Verification Lot of certified Charpy specimens (between 1600 and 2000 samples of a specific energy level) to be acceptable for production and distribution by NIST, specific requirements on machining, dimensional tolerances, and most importantly experimental variability (scatter) have to be fulfilled [4]. The operational procedure in use by the Charpy Verification Program at NIST evaluates the suitability of a Charpy Verification Lot by examining two groups of 100 specimens supplied by one of the qualified vendors, called *pilot lot* and *production lot* respectively. Evaluation consists of dimensional controls, hardness measurements, and 75 impact tests performed on the reference machines (25 tests per machine).

If the pilot lot evaluation is successful^b, the vendor is authorized to heat-treat and machine the remaining specimens of the Verification Lot. An additional group of 100 specimens (production lot), randomly sampled, is then evaluated in the same way. If the results are also

^b The requirements for the acceptability of a pilot or production lot are addressed in the following section.

acceptable, the entire Verification Lot is qualified, and the test results from both pilot and production lots are jointly analyzed to establish the certified value of absorbed energy (KV_{ref}) and the associated uncertainties (standard and expanded).

If the evaluation of either the pilot or the production lot is unsuccessful, the Verification Lot is rejected. When this happens at the pilot lot stage, the vendor is notified and can try to modify the heat treatment of the remaining blanks to improve the quality of the Verification Lot.

The following two sections of this Internal Report describe the analytical procedures used by the Charpy Verification Program for the qualification of pilot and production lots, as well as the calculation of reference energies and associated uncertainties for an acceptable Charpy Verification Lot.

2. Qualification of a Pilot or Production Lot

As mentioned above, 75 specimens from a pilot or production lot are normally tested on the three NIST reference machines (25 on each machine). The remaining 25 specimens are held in reserve for possible additional testing. The test temperature is $-40\text{ }^{\circ}\text{C}$ for low-energy (13 J to 20 J) and high-energy (88 J to 136 J) specimens, and room temperature ($21\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$) for super-high-energy (176 J to 244 J) specimens.

The grand average absorbed energy is defined as:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (1)$$

where x is absorbed energy (KV) and n is the number of tests performed. For tests performed on an individual machine, normally $n = 25$; considering all tests performed on a pilot or production lot, normally $n = 75$.

The standard deviation (for an individual machine or for the whole lot) is given by:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}}, \quad (2)$$

where $x_1 \dots x_n$ are the individual test results, while the pooled standard deviation is defined as:

$$s_p = \sqrt{\frac{\sum_{j=1}^P (n_j - 1) s_j^2}{\sum_{j=1}^P n_j - 1}}, \quad (3)$$

^c Occasionally, less than 25 tests are available for a specific machine ($n < 25$), due to various causes (energy values not recorded, specimen not positioned correctly, etc.). In a few cases, more than 25 tests have been performed.

where n_j and s_j are the number of tests and the standard deviation for the specific reference machine, and P is the number of reference machines (three)^d.

When the number of tests is 25 for each reference machine, eq. (3) simplifies to:

$$s_p = \sqrt{\frac{s_1^2 + s_2^2 + s_3^2}{P}}, \quad (4)$$

where s_1, s_2, s_3 are the standard deviations of the three reference machines.

The determining parameter for the qualification/acceptance of a pilot/production lot is the **sample size**, which corresponds to the minimum number of specimens that should be tested by a customer for a statistically reliable machine verification. In the case of 3 machines, it is defined as:

$$n_{SS} = \left(\frac{3s_p}{E} \right)^2, \quad (5)$$

where E is the larger of 1.4 J or 5 % of the mean energy. **If the sample size of a pilot or production lot is smaller than or equal to than 5.0 ($n_{SS} \leq 5.0$), the lot is considered acceptable^e.**

For low-energy specimens, $E = 1.4$ J, and therefore the maximum acceptable pooled standard deviation is $s_p = 1.04$ J, based on eq. (5). For high-energy and super-high-energy specimens, $E = 0.05 \overline{KV}$, and the maximum acceptable pooled standard deviation is:

$$s_p = 0.037 \overline{KV}, \quad (6)$$

which corresponds to a coefficient of variation (given by the standard deviation divided by the mean absorbed energy) $CV = 0.037$.

The determination of the sample size is accompanied by an outlier analysis, which is performed using box-and-whiskers plots to provide a graphical representation of the data. An outlier is defined as a test result that is lower than the first quartile, or higher than the third quartile, by more than 1.5 times the absolute difference between the first and third quartiles. If a lot has more than 5 % but less than 10 % outliers, it may be rejected (depending on the outcome of further analyses, such as additional Charpy tests); if a lot has more than 10 % outliers, it shall be rejected. Note that an outlier is not automatically removed from the analyses, unless the specimen shows evidence of jamming, significant material flaws, errors in tests execution, or other reasons that can justify its atypical behavior.

Finally, the energy variation for each machine is assessed by means of the parameter k , which is the ratio between the machine standard deviation and the pooled standard deviation:

^d The choice of the standard deviation to be used for subsequent evaluations depends on whether the machines meet the variability requirement k , see eq. (7). If the requirement is met by all machines, s_p is used. If one or more machines do not meet the requirement, the standard deviation corresponds to the largest value of s among the individual machines.

^e As long as the results of the dimensional and hardness measurements are also acceptable.

$$k_j = \frac{s_j}{s_p} \quad , \quad (7)$$

where j is the machine index ($j = 1, 2$, or 3). If k is larger than the critical value 1.28 (assuming 25 specimens tested on each machine), the variability of that machine can be questioned and additional measures are taken (repairs, testing of additional samples, etc.).

3. Establishment of Certified Absorbed Energy (KV_{ref}) and Associated Uncertainties

The certified value of absorbed energy for the Charpy Verification Lot (KV_{ref}) is calculated as the mean of the means of the pilot lot and the production lot^f:

$$KV_{ref} = \frac{\overline{KV}_{pilot} + \overline{KV}_{production}}{2} \quad , \quad (8)$$

where:

$$\overline{KV}_{pilot} = \frac{\sum_{j=1}^3 \overline{KV}_{j,pilot}}{3} \quad , \quad (9)$$

$$\overline{KV}_{production} = \frac{\sum_{j=1}^3 \overline{KV}_{j,production}}{3} \quad , \quad (10)$$

and $\overline{KV}_{j,z}$ is the grand average of the measurements from machine j for $z = \text{pilot or production lot}$, **provided there is no statistical difference between the pilot lot and the production lot**. The latter statement is assessed on the basis of a F -test for the equality of variances [5,6] and a t -test for the equality of means [5,7]. If one or both tests lead to rejecting the null hypothesis that the variances and/or the means are equal, the certified value of absorbed energy is the grand average for the production lot only, given by Eq. (10), as this is considered more representative of the entire Charpy Verification Lot.

Similarly, the associated uncertainties are those of the two combined lots (pilot and production) if there are no statistical differences in terms of means and/or variances, or just those of the production lot otherwise.

For the calculation of standard uncertainties, the standard error (SE) for each machine has to be calculated:

$$SE_j = \frac{s_j}{\sqrt{n_j}} \quad , \quad (11)$$

where s_j and n_j are the standard deviation and the number of tests for machine j , respectively. The standard uncertainty for the pilot or production lot is given by:

$$u = \sqrt{\sum_{j=1}^3 \left(\frac{SE_j}{3} \right)^2} \quad . \quad (12)$$

^f If 50 test results are available for each reference machine (25 from the pilot lot and 25 from the production lot), this corresponds to the grand average of all 150 measurements.

If pilot and production lots are statistically equivalent, the standard uncertainty for the combined lot is:

$$u_{combined} = \sqrt{\frac{u_{pilot}^2 + u_{production}^2}{2}} \quad (13)$$

If they are not statistically equivalent, $u_{production}$ is used for the entire lot.

Finally, the expanded uncertainty is obtained by multiplying the standard uncertainty by a coverage factor k (typically, in the range 2 to 3), chosen on the basis of the desired level of confidence:

$$U = k \cdot u \quad (14)$$

In practice, k is calculated as the two-tailed inverse of the Student's t -distribution for a specific number of degrees of freedom. For a pilot or production lot, the number of degrees of freedom is calculated as:

$$DF = \frac{u^4}{\sum_{j=1}^3 \left[\frac{\left(\frac{SE_j}{3} \right)^4}{n_j - 1} \right]} \quad (15)$$

For a combined pilot/production lot, the corresponding number of degrees of freedom is given by:

$$DF_{combined} = \frac{u_{combined}^4}{\frac{\left(\frac{u_{pilot}}{2} \right)^4}{DF_{pilot}} + \frac{\left(\frac{u_{production}}{2} \right)^4}{DF_{production}}} \quad (16)$$

4. Software Used at NIST for the Analysis and Qualification of Charpy Verification Lots

For the first 15 years of the Charpy Verification Program at NIST, the statistical analyses described in the previous two sections were performed with the aid of a commercially available statistical package, called SYSTAT^g.

In the mid-2005s, two codes in R language^h were written by Jolene Splett of the Statistical Engineering Division of NIST, one for the determination of the sample size of pilot and production lots and one for the establishment of reference values and uncertainties for the Verification Lot. The first R code (sample size calculation) also offered a graphical representation of the test results in the form of both box-and-whiskers plots and histograms. These two codes replaced the SYSTAT package and have been used, until recently, in its lieu.

However, the analyses performed by means of SYSTAT or R did not include the outlier analysis mentioned in section 2.

This Internal Report describes the operation and use of two spreadsheet templates in Microsoft Excel 2016^e that have recently been implemented and successfully tested, which closely match the statistical and graphical outcomes of the above-mentioned software, but also include an explicit outlier analysis. The spreadsheets can be used without the need to install a platform for running R codes. Only the standard Microsoft Office package, which includes Excel, is required.

Note that, in order to use the spreadsheets, the user has to enable macros in Excel.

The two templates have been checked and validated against calculations performed on several pilot, production, and Verification Lots by means of the previously used R-language codes.

4.1 Template for the Analysis of Pilot and Production Lots

The name of the Excel template is *Charpy lot statistical analyses.xltn*. A partial screenshot of the first sheet (“Input data”) is shown in Figure 1.

Columns B and C contain the results of the qualification tests (absorbed energy in J in column B and hardness measurements HRC in column C). Column A indicates the machine on which each test was performed. NOTE – Results for the individual machines must be reported in this order: first SI, then TK, and finally TO.

If the sheet already contains data in columns B and C, the user can click the button “CLEAR DATA” (❶ in Figure 1) to erase the columns (except the headings). Test results (absorbed energy and hardness) should now be entered in the columns.

^g Commercial names are identified in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the NIST, nor does it imply that they are necessarily the best available for the purpose.

^h R is a programming language and free software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing, and is widely used among statisticians and data miners for developing statistical software and data analysis.

	A	B	C	D	E	F	G	H	I	J	K
1	Machine	Energy	Hardness	1	CLEAR DATA	2	Lot Id: LL-171 Type of lot: Pilot Test date: February 26, 2018				
2	SI	13.745	45.45								
3	SI	14.237	44.85								
4	SI	14.307	45.65								
5	SI	14.166	45.4		# of tests : 75						
6	SI	12.624	45.45		# of machines : 3						
7	SI	14.589	45.75								
8	SI	14.096	45.7								
9	SI	13.745	45.75	3	IMPORT MACHINE RESULTS						
10	SI	12.974	45.5								
11	SI	12.904	45.55								
12	SI	14.801		4	START ANALYSES						
13	SI	12.136									
14	SI	14.589									
15	SI	15.225									
16	SI	14.801									
17	SI	14.519									
18	SI	13.254									
19	SI	13.885									
20	SI	14.801									
21	SI	14.589									
22	SI	14.589									
23	SI	14.307									
24	SI	14.66									
25	SI	12.205									
26	SI	15.013									
27	TK	12.174	45.45								
28	TK	12.773	45.65								
29	TK	12.873	45.4								
30	TK	13.572	45.2								
31	TK	12.573	45.4								
32	TK	12.474	45.45								
33	TK	12.973	45.55								
34	TK	11.576	45.85								
35	TK	13.073	45.5								
36	TK	12.274	45.6								
37	TK	13.073									

Information about the lot id, the lot type (pilot or production), and the test date are entered in the green box on the right (2 in Figure 1). The test date can be entered in any format, since it's formatted as generic text.

To import the values of absorbed energy obtained from tests performed with an individual machine, the user can click the button "IMPORT MACHINE RESULTS" (3 in Figure 1). These values are typically available in an ASCII file (text, tab delimited) that has the format shown in Appendix 1.

After clicking the button, the user selects and opens the text file containing the KV values. Automatically, unnecessary rows and columns are removed, and only the absorbed energy values are selected and copied into the clipboard. The text file is then closed without saving. The user can now position the cursor on the cell of column B corresponding to the first result of the corresponding machine, and press $Ctrl+V$ (or click the "Paste" icon). The KV values are then pasted in the spreadsheetⁱ.

Once all test results are entered, cells F5 and F6 will contain the total number of tests (typically 75) and the total number of machines (typically 3), respectively.

ⁱ The user should verify that the number of test results pasted (typically 25) effectively corresponds to the number of cells for which column A contains the corresponding machine id (SI, TK, or TO).

Analyses are initiated by clicking on the button “START ANALYSES” (4 in Figure 1). The sheet “Analyses” (screenshot in Figure 2) contains the following results:

- Statistics for all machines (rows 4-13).
- Statistics for each individual machine (rows 18-20).
- Test for the equality of variances (Levene's test [8,9], rows 25-26).
- Test on the variation of energy values for each machine (ASTM Pass/Fail criterion, rows 32-34 and columns A-E).
- Pooled standard deviation (cell I30) and maximum pooled standard deviation (cell I33).
- Sample size (cell F39^j), with the value of E used in eq. (5) in cell F38.
- Outlier analysis for the individual machines and the entire lot (rows 44-54^k).

NOTE: columns N-AI contain individual machine data and detailed calculations for the outlier analysis.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

	A	B	C	D	E	F	G	H	I	J	K	L	M	
All Machines					<div>Lot Id : LL-171 Type of Lot : Pilot Test Date : February 26, 2018</div>									
			Energy	Hardness										
Number of tests N =	75	30												
Mean =	13.552	45.517												
Standard Deviation =	1.115	0.200												
Variance =	1.244	0.040												
Degrees of freedom =	74	29												
Standard Error =	0.129	0.037												
Minimum =	10.679	44.85												
Maximum =	15.381	45.85												
Range =	4.702	1												
CV =	0.082	0.004												
Machine Statistics														
Machine	Mean	n	St.dev.	Variance	DOF	St. Err.	Min	Max	Range	CV				
SI	14.030	25	0.874	0.764	24	0.175	12.136	15.225	3.089	0.062				
TK	12.534	25	0.700	0.490	24	0.140	10.679	13.572	2.893	0.056				
TO	14.090	25	0.980	0.960	24	0.196	10.76	15.381	4.621	0.070				
Equality of Variances: Levene's Test														
N = 75		k = 3		Significance level $\alpha = 0.05$				p-value = 0.763						
Levene's test statistic W = 0.272				$F_{critical} = 3.124$				Variances are equal (W < F_critical)						
ASTM Pass/Fail criterion					Pooled Standard Deviation									
					$s_p = 0.859$ J									
Machine	Deviation	Pass/Fail	k	k critical										
SI	0.479	Pass	1.018	1.28										
TK	-1.018	Pass	1.015	1.28										
TO	0.539	Pass	1.140	1.28										
Sample Size														
E = 1.40 J														
$n_{LS} = 3.390$														
Outlier Analysis														
Machine: SI				Machine: TK				Machine: TO						
First quartile SI = 13.4995 J				First quartile TK = 12.174 J				First quartile TO = 13.707 J						
Third quartile SI = 14.6245 J				Third quartile TK = 13.023 J				Third quartile TO = 14.694 J						
Lower limit SI = 11.812 J				Lower limit TK = 10.901 J				Lower limit TO = 12.227 J						
Upper limit SI = 16.312 J				Upper limit TK = 14.297 J				Upper limit TO = 16.175 J						
Low outliers SI = 0				Low outliers TK = 1				Low outliers TO = 2						
High outliers SI = 0				High outliers TK = 0				High outliers TO = 0						
Total outliers SI = 0				Total outliers TK = 1				Total outliers TO = 2						
				Total low outliers = 3										
				Total high outliers = 0										
				Total outliers = 3 (4%)										
THE NUMBER OF OUTLIERS IS NOT A CONCERN FOR THIS LOT														

<

>

Input data

Analyses

Box-and-Whiskers Plot

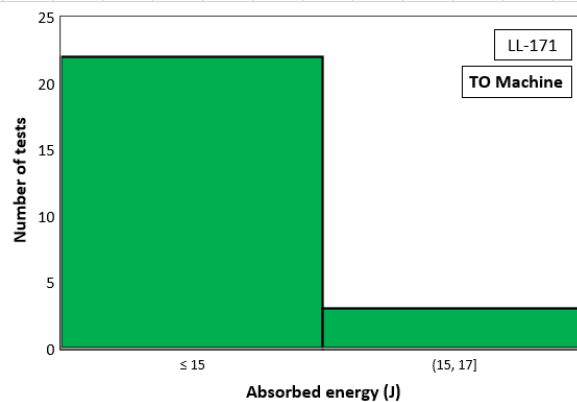
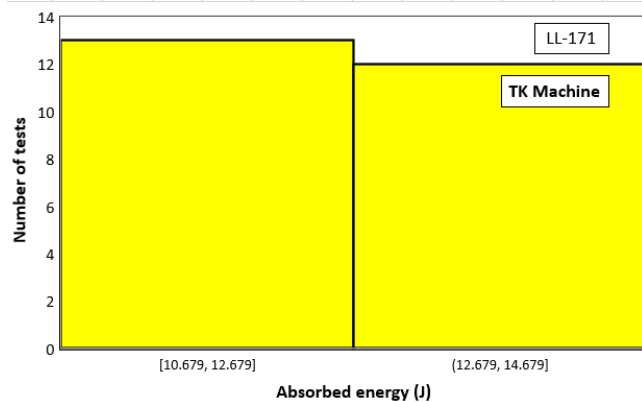
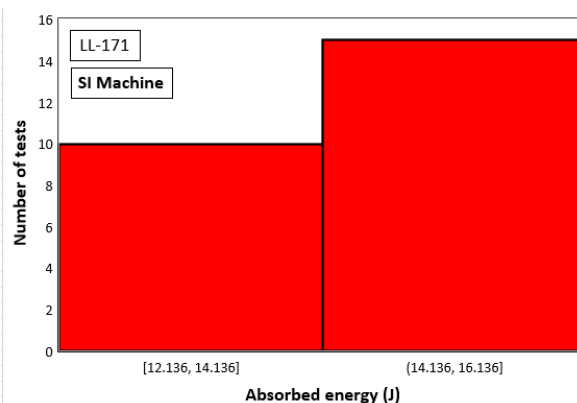
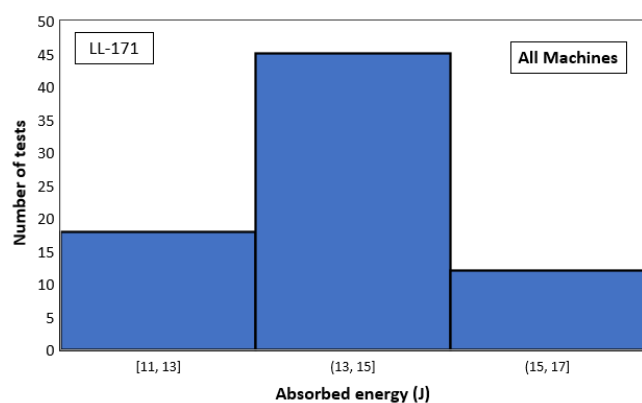
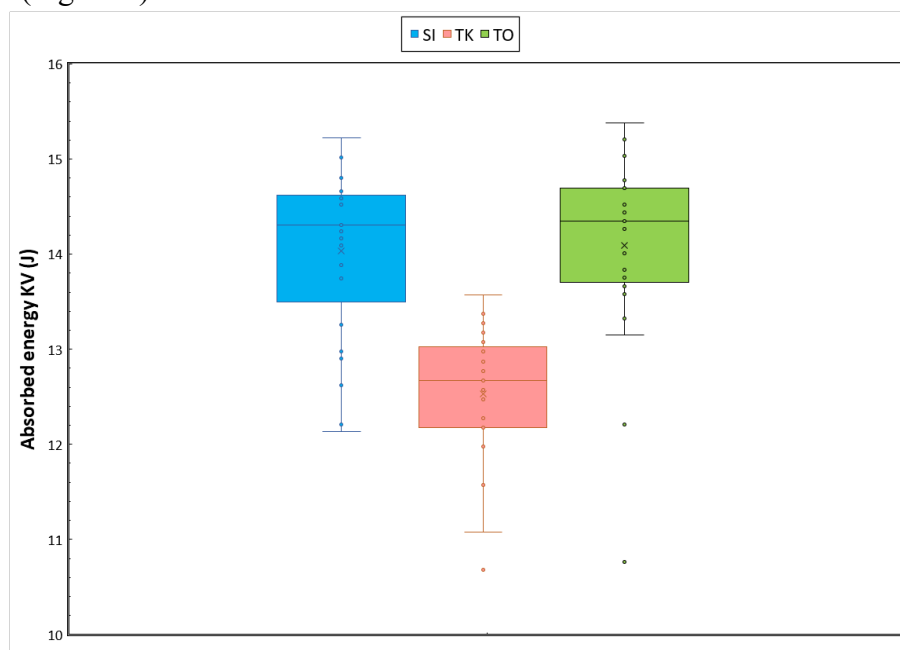
Histograms

+

^j Cell F39 is formatted according to its value and acceptability (green if $n_{SS} \leq 5.0$, see Figure 2, red if $n_{SS} > 5.0$).

^k Row 54 is formatted according to the outcome of the outlier analysis (green if total outliers $\leq 5\%$, see Figure 2, yellow if total outliers between 5% and 10% , and red if total outliers $> 10\%$).

Graphical output is contained in the sheets “Box-and-Whiskers Plot” (Figure 3) and “Histograms” (Figure 4).



The analytical and graphical output of the spreadsheet can be printed (using the system default printer) by clicking the button “PRINT RESULTS” on the “Analyses” sheet (Figure 2). An example of printed output is provided in Appendix 2.

4.2 Template for the Calculation of Reference Values and Uncertainties

The name of the Excel template is *Reference & Uncertainties Charpy lots.xltm*. A partial screenshot of the sheet “Pilot lot data” (the sheet “Production lot data” is basically identical) is shown in Figure 5.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Machine	Energy	Hardness												
2	SI	16.054	46.2												
3	SI	16.622	46.95												
4	SI	16.267	47												
5	SI	15.912	47.05												
6	SI	15.558	46.95												
7	SI	14.922	47.15												
8	SI	15.487	47.4												
9	SI	14.71	46.95												
10	SI	15.77	46.95												
11	SI	15.346	47.2												
12	SI	14.429													
13	SI	16.693													
14	SI	16.409													
15	SI	15.078													
16	SI	16.338													
17	SI	14.71													
18	SI	13.586													
19	SI	16.196													
20	SI	15.275													
21	SI	15.204													
22	SI	16.125													
23	SI	15.841													
24	SI	16.409													
25	SI	16.693													
26	SI	16.054													
27	TK	14.572	46.65												
28	TK	12.773	46.95												
29	TK	13.172	46.75												
30	TK	14.072	47.05												
31	TK	13.372	47.2												
32	TK	13.972	47.3												
33	TK	14.572	46.7												
34	TK	12.274	46.8												
35	TK	14.572	46.75												
36	TK	14.372	47												
37	TK	13.472													

Lot id and date of analysis are entered by the user in cell I1 and I3, respectively. Any pre-existing data in columns B-C and N-P can be erased by clicking on the button “CLEAR DATA”. New data can be copied and pasted, for example, from the corresponding analysis spreadsheet into columns B and C (remember to check that the machine information in column A is correct). To initiate the analyses, click the button “START ANALYSES”.

The calculated statistics are provided in rows 9-11 for the individual machines and in rows 13-17 for the whole lot, including the reference value for absorbed energy, the standard uncertainty, and the degrees of freedom. Pooled standard deviation (cell J16) and sample size (cell J17) must coincide with those calculated previously by means of the other template.

After filling out the sheet “Production lot data” in a similar way, the statistical comparisons between pilot and production lot, namely the F -Test for the Equality of Variances and the t -Test for the Equality of Means, are presented in the sheet “Analyses of lots” (Figure 6).

In the lower part of the sheet, statistics are provided for the production lot (row 18) and for the combined lots (row 19). If either or both the comparative tests indicate a significant difference between pilot and production lots, the statistics for the production lots are highlighted in bold blue font, and an arrow appears in column A (Figure 6). If no statistical difference is detected between the lots, the same graphical highlighting is applied to the combined lots in row 19.

	A	B	C	D	E	F	G	H	I	J	K	L
1	F-Test for Equality of Variances						t-Test for Equality of Means					
2												
3		Uncertainty of pilot lot :	0.094	J				Average of pilot lot :	15.107	J		
4		Uncertainty of production lot :	0.089	J				Average of production lot :	15.609	J		
5												
6		Test statistic F =	1.0980					Test statistic t =	3.8846			
7		$F_{critical}$ =	1.557					$t_{critical}$ =	0.0001509			
8		Significance level α =	0.05					Significance level α =	0.05			
9		p-value =	0.3668103					p-value =	0.0003018			
10												
11		Variances are not different						Means are different				
12												
13												
14	Calculation of reference values and associated uncertainties											
15												
16										95 % conf. interval	Sample	
17			Ref. value	u(ref)	DOF	k factor	U(exp)	Lower	Upper	size		
18	====>	Production lot	15.609	0.089	50	2.0086	0.179	15.430	15.789	2.750		
19		Combined lots	15.358	0.065	119	1.9801	0.128	15.230	15.486	2.885		
20												
21												
22												
23												

The complete results of the analyses are collected in the sheet “Results”, and can be printed on the default system printer by clicking the button “PRINT RESULTS”. The final parameters for the Charpy Verification Lot, to be reported on its SRM Certificate, are at the bottom of the page, namely:

- Reference value for absorbed energy (KV_{ref}).
- Expanded uncertainty (U).
- Effective degrees of freedom.
- Coverage factor (k), corresponding to an approximate 95 % confidence interval on the reference value.
- Sample size (n_{SS}) for the production lot or the combined lots (depending on the statistical comparison).

An example of printed output is reproduced in Appendix 3.

References

- [1] ASTM E23-16b, “Standard Test Methods for Notched Bar Impact Testing of Metallic Materials,” 03.01 Book of Standards, ASTM International, West Conshohocken, PA.
- [2] ISO 148-2:2016, “Metallic materials — Charpy pendulum impact test — Part 2: Verification of testing machines,” International Standards Organization.
- [3] “Mechanical Testing – The Charpy V-Notch Impact Test,” Quality Assurance Pamphlet AMXMR-P-702-104, Army Materials and Mechanics Research Center, 1 December 1969.
- [4] C. N. McCowan, T. A. Siewert, and D. P. Vigliotti, “The NIST Charpy V-Notch Verification Program: Overview and Operating Procedures,” in *Charpy Verification Program: Reports Covering 1989-2002*, NIST Technical Note 1500-9, Materials Reliability Series, September 2003 (http://ws680.nist.gov/publication/get_pdf.cfm?pub_id=851238).
- [5] G. W. Snedecor and W. G. Cochran, *Statistical Methods*, Eighth Edition, Iowa State University Press, 1989.
- [6] NIST/SEMATECH *e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/handbook/eda/section3/eda359.htm>, retrieved February 28th, 2018.
- [7] NIST/SEMATECH *e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/handbook/eda/section3/eda353.htm>, retrieved February 28th, 2018.
- [8] H. Levene, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et al., Eds., Stanford University Press, 1960, pp. 278-292.
- [9] NIST/SEMATECH *e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm>, retrieved February 28th, 2018.

Appendix A – Charpy ASCII Data File

3Ni2_TK_012017.txt							
ID#	DIM. in	P.E. J	B.E. J	STR. ft-lbf/in	BREAK TYPE	E CORR. J	LIMITS ENERGY STRENGTH
0001	1.0000	355.74	193.64	142.82		0.0000	
0002	1.0000	355.74	192.48	141.97		0.0000	
0003	1.0000	355.74	191.23	141.04		0.0000	
0004	1.0000	355.74	196.22	144.73		0.0000	
0005	1.0000	355.74	196.60	145.01		0.0000	
0006	1.0000	355.74	221.38	163.28		0.0000	
0007	1.0000	355.74	204.94	151.16		0.0000	
0008	1.0000	355.74	195.17	143.95		0.0000	
0009	1.0000	355.74	190.65	140.62		0.0000	
0010	1.0000	355.74	196.32	144.80		0.0000	
0011	1.0000	355.74	192.39	141.90		0.0000	
0012	1.0000	355.74	195.07	143.88		0.0000	
0013	1.0000	355.74	187.85	138.55		0.0000	
0014	1.0000	355.74	184.93	136.40		0.0000	
0015	1.0000	355.74	213.79	157.69		0.0000	

Appendix B – Printed Output of Template *Charpy lot statistical analyses*

All Machines

	Energy	Hardness
Number of tests N =	75	30
Mean =	13.552	45.517
Standard Deviation =	1.115	0.200
Variance =	1.244	0.040
Degrees of freedom =	74	29
Standard Error =	0.129	0.037
Minimum =	10.679	44.85
Maximum =	15.381	45.85
Range =	4.702	1
CV =	0.082	0.004

Lot Id : LL-171
Type of Lot : Pilot
Test Date : February 26, 2018

Machine Statistics

Machine	Mean	n	St.dev.	Variance	DOF	St. Err.	Min	Max	Range	CV
SI	14.030	25	0.874	0.764	24	0.175	12.136	15.225	3.089	0.062
TK	12.534	25	0.700	0.490	24	0.140	10.679	13.572	2.893	0.056
TO	14.090	25	0.980	0.960	24	0.196	10.76	15.381	4.621	0.070

Equality of Variances: Levene's Test

N = 75	k = 3	Significance level $\alpha = 0.05$	p-value = 0.763
Levene's test statistic W = 0.272	$F_{critical} = 3.124$	Variances are equal ($W < F_{critical}$)	

ASTM Pass/Fail criterion

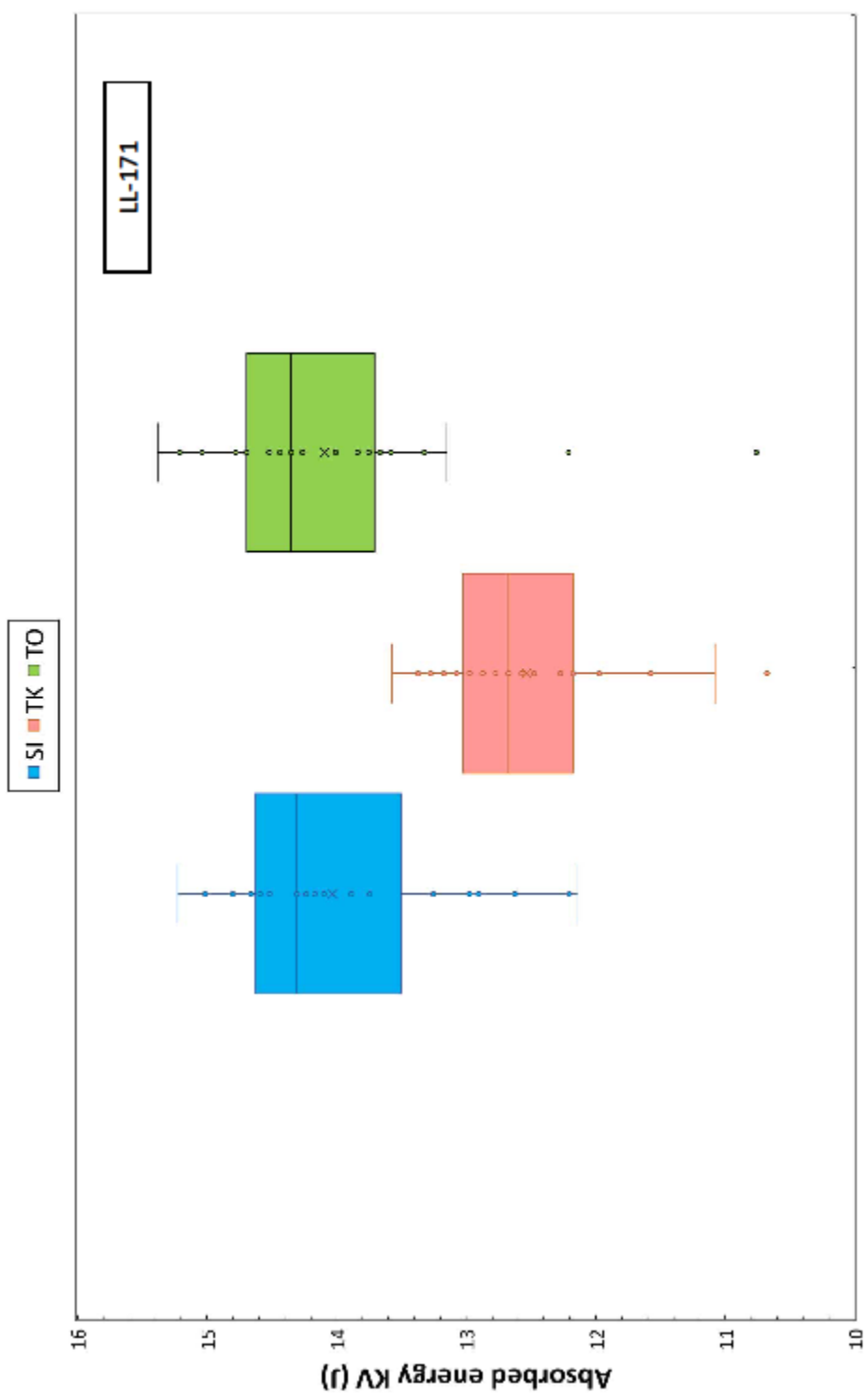
Machine	Deviation	Pass/Fail	k	k critical
SI	0.479	Pass	1.018	1.28
TK	-1.018	Pass	0.815	1.28
TO	0.539	Pass	1.140	1.28

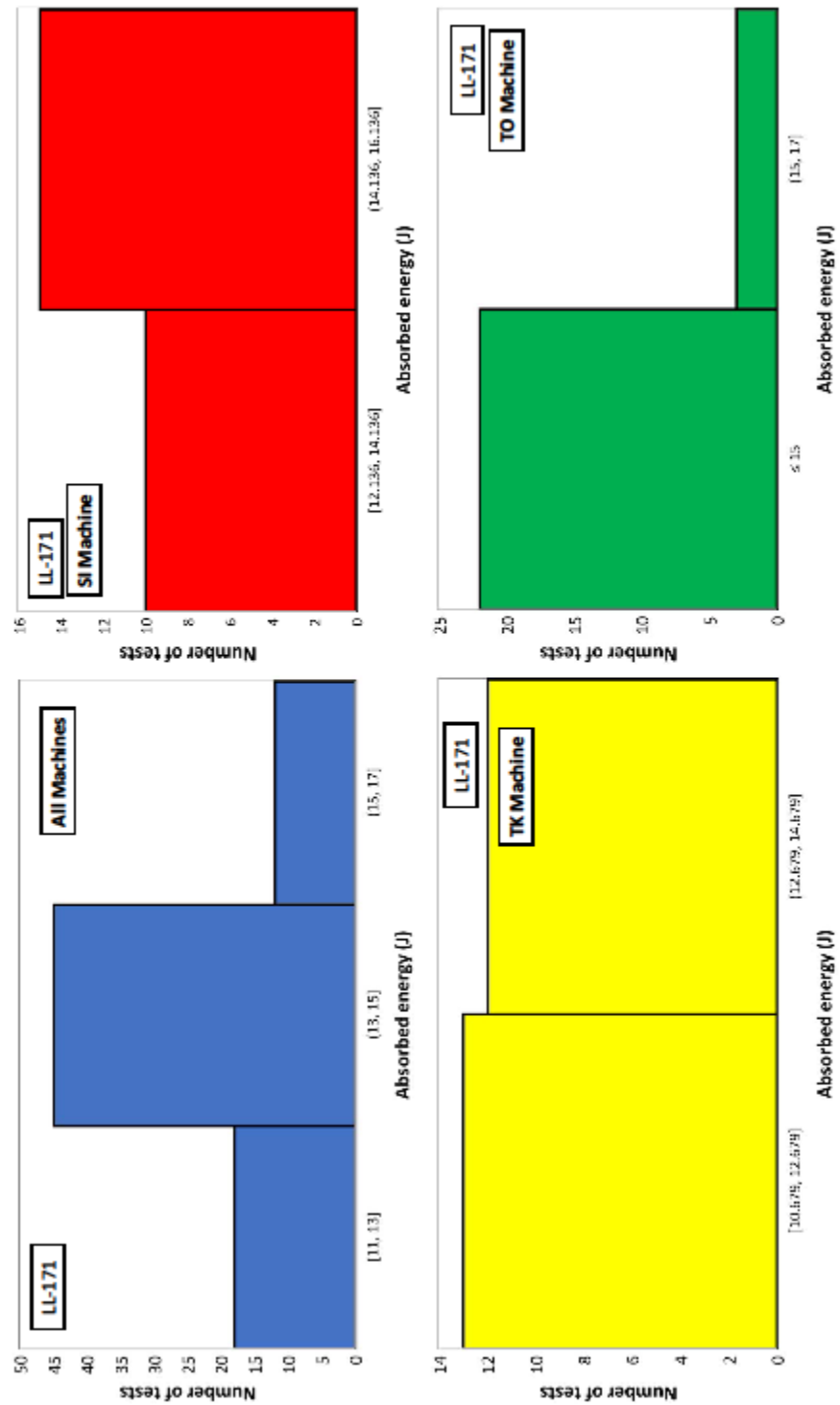
Pooled Standard Deviation

$s_p = 0.859$ J
Maximum Pooled Standard Deviation
Max $s_p = 1.043$ J

Sample Size
E = 1.40 J
$n_{ss} = 3.390$

Outlier Analysis		
Machine: SI	Machine: TK	Machine: TO
First quartile SI = 13.4995 J	First quartile TK = 12.174 J	First quartile TO = 13.707 J
Third quartile SI = 14.6245 J	Third quartile TK = 13.023 J	Third quartile TO = 14.694 J
Lower limit SI = 11.812 J	Lower limit TK = 10.901 J	Lower limit TO = 12.227 J
Upper limit SI = 16.312 J	Upper limit TK = 14.297 J	Upper limit TO = 16.175 J
Low outliers SI = 0	Low outliers TK = 1	Low outliers TO = 2
High outliers SI = 0	High outliers TK = 0	High outliers TO = 0
Total outliers SI = 0	Total outliers TK = 1	Total outliers TO = 2
Total low outliers = 3		
Total high outliers = 0		
Total outliers = 3 (4 %)		
THE NUMBER OF OUTLIERS IS NOT A CONCERN FOR THIS LOT		





Appendix C – Printed Output of Template *Reference & Uncertainties* *Charpy Lots.xltn*

Lot id: LL-159

Date: November 2017

Reference values, standard uncertainties and degrees of freedom for the individual lots

	$KV_{ref} (J)$	$u_{KVref} (J)$	DOF
Pilot Lot	15.107	0.094	71
Production Lot	15.609	0.089	50

F-test for the equality of variances

F-value	1.0980	\Rightarrow	Variances are not different	$(p > 0.05)$
p-value	0.3668			

t-test for the equality of means

t-value	3.8846	\Rightarrow	Means are different	$(p < 0.05)$
p-value	0.0003			

Results for Production Lot and Pilot + Production Lots Combined

	$KV_{ref} (J)$	$u_c (J)$	DOF	k	$U_{KVref} (J)$	95 % conf. interval		Sample size
						Lower (J)	Upper (J)	
Production	15.609	0.089	50	2.0086	0.179	15.430	15.789	2.750
Combined	15.358	0.065	119	1.9801	0.128	15.230	15.486	2.885

Reference Value and Expanded Uncertainty for lot LL-159

Absorbed energy, $KV_{ref} = 15.609 \text{ J}$
 Expanded uncertainty, $U = 0.089 \text{ J}$
 Effective degrees of freedom = 50
 Coverage factor, $k = 2.0086$
 Sample size, $n_{ss} = 2.750$

NOTES The coverage factor k is the t- table value corresponding to an approximate 95 % confidence interval on the reference values based on the effective degrees of freedom. U is the combined standard uncertainty, u_c multiplied by k .

Appendix D – Change Log

The following changes were made in November 2021.

On page 2:

- Eq. (3) for the calculation of the pooled standard deviation has been updated, to account for the fact that sometimes more or less than 25 test results may be available for a single reference machine (footnote c has also been added to address such occurrence).
- The simplified equation for the case that 25 test results are available for each machine is now eq. (4).

On page 4:

- Eqs. (8-10) for the calculation of the certified value and the pilot and production lot means have been added, to account for when less than 25 test results are available for one or more reference machine. Footnote f has been added to address the case that every machine has 50 test results (25 from the pilot lot and 25 from the production lot).
- A typo has been corrected in the formula for the standard error, currently eq. (11).
- Another typo has been corrected in the formula for the combined standard uncertainty, currently eq. (13).
- Below eq. (13), a statement has been added to clarify that the standard uncertainty of the production lot is used if pilot and production lots are not statistically equivalent.

On page 5:

- A typo has been corrected in the formula for the number of degrees of freedom for a pilot or production lot, currently eq. (15).