Theoretical and Applied Fracture Mechanics 120 (2022) 103385

Contents lists available at ScienceDirect

Theoretical and Applied Fracture Mechanics

journal homepage: www.elsevier.com/locate/tafmec

Interlaboratory comparison study of crack size measurements performed with two different methods

Enrico Lucon

Applied Chemicals and Materials Division, National Institute of Standards and Technology (NIST), Boulder, CO, United States

ARTICLE INFO

Keywords:
9-Point Average method
Area Average method
ASTM E691-18
Crack size measurements
Fatigue and fracture test methods
Interlaboratory study

ABSTRACT

An interlaboratory study (ILS) involving 15 laboratories from North America, Europe, and Japan, was conducted to validate crack size measurements on various types of fracture toughness specimens by means of two commonly used methodologies: the 9-Point Average (9PA) method and the Area Average (AA) method. The results of this ILS confirm the findings of previous investigation published by the author in 2021, which showed that crack sizes measured by the two methods are in close agreement, and therefore we recommend the AA method be included as an alternative to the other approach in all fatigue and fracture test methods that require crack sizes to be measured.

1. Introduction

In 2021, the author published a comparative study [1,2] on two crack size measuring methodologies in fatigue and fracture mechanics specimens: the 9-Point Average (9PA) method and the Area Average (AA) method. The study, conducted on 140 fracture toughness specimens of various materials and different geometries, concluded that the two methods provide measurements that are in close agreement, particularly for initial cracks obtained by fatigue precracking. Therefore, based on the author's opinion, a recommendation was formulated to include the AA method in ASTM fatigue and fracture test standards, as an alternative (not a replacement) to the more popular *X*PA method, where *X* indicates the number of measurements prescribed (between 2 and 9).

For the benefit of the reader, the main features of the two methodologies (9PA and AA) are summarized hereafter, with reference also to Fig. 1 [1].

1.1. 9-Point Average (9PA) method

The 9PA method uses 9 equally spaced measurement locations across the specimen thickness or net thickness, extending to 0.5 % of the specimen width from the specimen sides (plane-sided specimens) or side-groove roots (side-grooved specimens). The crack size is calculated by a weighted average, where the two near-surface measurements are averaged, and then the result is averaged with the remaining seven measurements. See left side of Fig. 1.

1.2. Area Average (AA) method

The AA method measures the surface area of a region of the fracture surface delimited by a reference line (zero line for crack size measurements) and the crack front. This area is then divided by the specimen thickness or net thickness, as applicable, to obtain an estimate of the crack size. See right size of Fig. 1. To the author's knowledge, the only normative document that mentions the Area Average (or "Area-Averaged") method is the 1992 ESIS P2-92 Procedure [3], which stated "*Although the area-averaged method should provide the most reliable results, the 9-point average method produces acceptable results.*" No fracture or fatigue ISO or ASTM test method presently allows the use of the Area Average method.

When the investigation results were presented by the author during the ASTM E08.07.05 subcommittee in November 2021, it was suggested to initiate an interlaboratory study (ILS), with the following objectives:

(a) Confirming the substantial equivalence between 9PA and AA measurements on the basis of a larger group of investigators, using different measurement tools.
(b) Establishing the between-laboratory reproducibility of crack size measurements obtained from the two methods.

This paper reports the results of this ILS, which was conducted in the period December 2021-March 2022 using digital pictures of fracture surfaces from five specimens of different materials and test geometries, both plane-sided and side-grooved, including two miniaturized samples, MC(T) and MSE(B).

https://doi.org/10.1016/j.tafmec.2022.103385
Received 24 March 2022; Received in revised form 4 May 2022; Accepted 5 May 2022
Available online 10 May 2022
0167-8442/Published by Elsevier Ltd.

Nomenclature

0.5TC(T)	Compact Tension specimen with thickness $B = 0.5$ in $= 12.7$ mm
1TC(T)	Compact Tension specimen with thickness $B = 1$ in $= 25.4$ mm
20 % SG PCCv	precracked Charpy-type specimen with 20 % side-grooves
9PA	9-Point Average (crack size measurement method)
α	confidence level for a *t*-test
AA	Area Average (crack size measurement method)
a_f	final crack size (mm)
$a_{f,9PA}$	final crack size measured with the 9-Point Average method (mm)
$a_{f,AA}$	final crack size measured with the Area Average method (mm)
a_o	initial crack size (mm)
$a_{o,9PA}$	initial (original) crack size measured with the 9-Point Average method (mm)
$a_{o,AA}$	initial (original) crack size measured with the Area Average method (mm)
ASTM	American Society for Testing and Materials
B	specimen thickness (mm)
B_N	net thickness for a side-grooved specimen (mm)
d	deviation of an individual crack size measurement from the average (mm)
h	repeatability consistency statistic
$h_{critical}$	critical value of h, which depends on the number of laboratories
ILS	Interlaboratory Study
MC(T)	miniature Compact Tension specimen (thickness = 4 mm)
MSE(B)	miniature Single-Edge Bend specimen (thickness 1.65 mm)
p	probability value for a *t*-test
R	95 % reproducibility limit (mm)
s_R	reproducibility standard deviation (mm)
$s_{\bar{x}}$	standard deviation of crack size measurements (mm)
x	individual crack size measurement (mm)
$\bar{\bar{x}}$	average of crack size measurements (mm)
XPA	*X*-Point Average (crack size measurement method), with $X = 2$ to 9

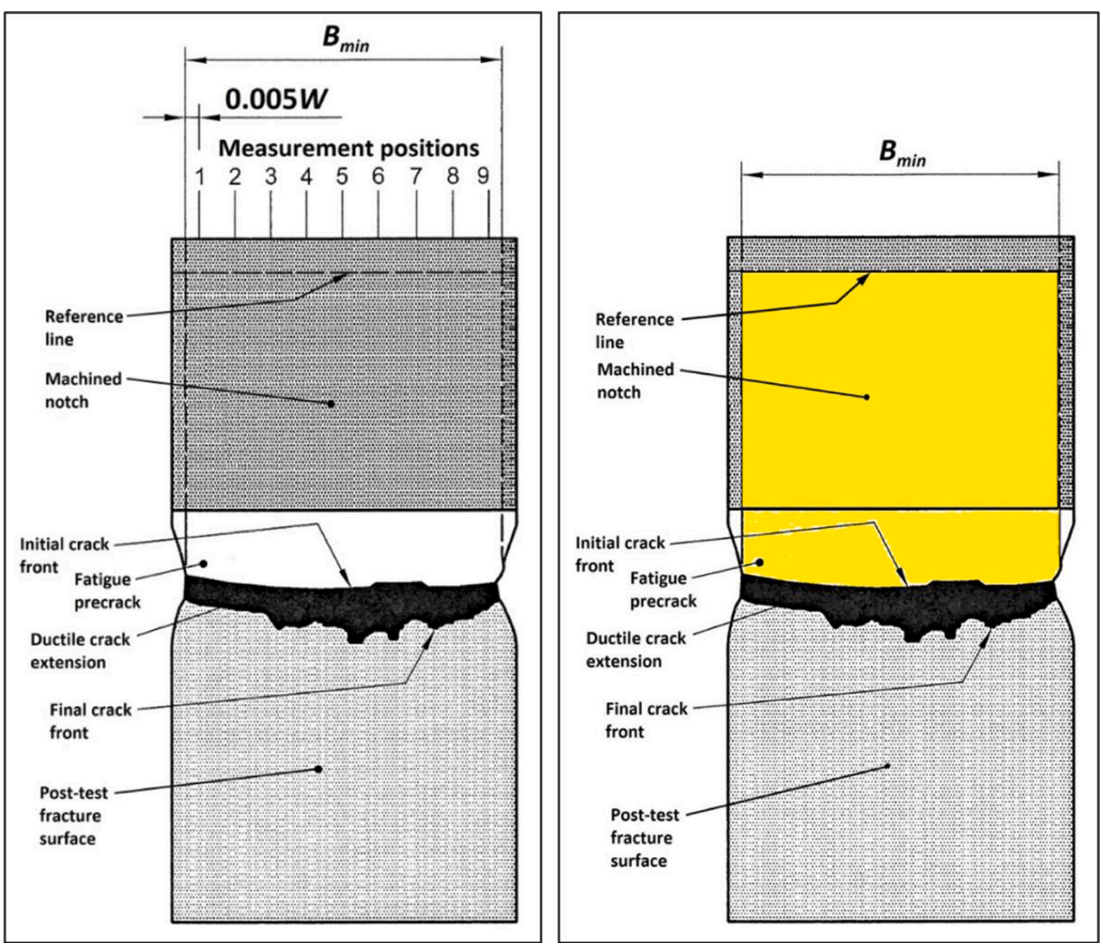

Fig. 1. Illustration of the 9PA (left) and AA (right) methods [1].

2. Description of the interlaboratory study

At the end of 2021, a call for participation was sent out to all the members of ASTM E08.07.05 (*Ductile Crack Initiation and Growth*) and E08.07.06 (*Ductile-Brittle Transition*) subcommittees. A sizeable number of responses was received, and the 15 participating laboratories who

Table 1
Details of participating institutes (listed in random order).

Participating Institute	City, Country	Contact Person
Blade Energy Partners	Houston, TX, US	Ken George
CanmetMATERIALS	Hamilton, Ontario, Canada	Jidong Kang
CanmetMATERIALS	Calgary, Alberta, Canada	Lin Yang
Comtes FHT	Dobrany, Czech Republic	Jan Dzugan
CRIEPI	Tokyo, Japan	Masato Jamamoto
Framatome GmbH	Erlangen, Germany	Johannes May
Hy-Performance Materials Testing	Bend, OR, US	Kevin Nibur
IWM Fraunhofer	Freiburg, Germany	Johannes Tlatlik
NIST	Boulder, CO, US	Enrico Lucon
NIST	Boulder, CO, US	Dash Weeks
SCK•CEN	Mol, Belgium	Johan Schuurmans
TWI	Cambridge, UK	Alex Pargeter
University of California	Davis, CA, US	Christine Smudde
University of California	Davis, CA, US	Tony Huang/Mike Hill
US Naval Academy	Annapolis, MD, US	Rick Link

succeeded in providing measurement results by the deadline provided are listed (in random order) in Table 1.

Participants were provided 5 digital pictures of varying quality, each corresponding to a different specimen size and configuration. Some relatively low-quality pictures were also included, as representative of “difficult” conditions for crack size measurements (low resolution, imperfect focus, etc.). The specimens represented (Fig. 2) were: 0.5TC(T), 1TC(T), MC(T), MSE(B), and 20 % SG PCCv. Only the 1TC(T) specimen featured measurable ductile crack extension, while for the remaining specimens only the initial crack size could be measured.

Each picture featured a reference line from which measurements had to be made. Based on the definition of crack size in the standards, the reference line for Compact Tension specimens corresponds either to the load-line, or the centerline of the pin holes; for bend-type specimens, *i.e.*, MSE(B) and PCCv, the reference line is located on the front face of the sample. Finally, to establish the scale of the pictures, either the specimen thickness (B) or net thickness (B_N) was provided to participants.

3. Interlaboratory study results

Individual measurements provided by participating laboratories, randomly numbered from 1 to 15, are shown in Table 2, while Table 3 presents basic statistics calculated from the measurements received (mean value, standard deviation, standard error, minimum and maximum values, and corresponding range).

The mean differences between 9PA and AA average crack sizes in Table 3 range between –0.02 mm (0.5TC(T) specimen) and 0.24 mm (1TC(T), a_f). Differences between individual measurements (9PA–AA) range from a minimum of –0.28 mm (0.5TC(T), Lab 8) up to a maximum of 1.39 mm (1TC(T), a_f, Lab 8). AA crack size measurements appear slightly more scattered than 9PA measurements.

Crack size measurements obtained from 9PA and AA methods are graphically compared in Fig. 3 (crack sizes between 1.5 mm and 6 mm) and Fig. 4 (crack sizes between 17.5 mm and 30 mm). Only three data points, all corresponding to a_f values for the 1TC(T) specimen, fall outside the $\pm$ 0.5 mm scatter band, corresponding to a reasonable level of precision, shown in Fig. 4.

3.1. *t-tests For the statistical comparisons between 9PA and AA crack size measurements*

Sets of 9PA and AA crack size measurements provided by the participants were statistically compared by performing *t*-tests at a confidence level $\alpha = 0.05$. The *t*-test [4] is a statistical hypothesis test that is used to compare the means of two groups, and indicates that the two groups are not statistically different when the calculated probability value (p) is greater than the confidence level used, in this case $p \geq 0.05$.

The *t*-test results for the 5 specimens and 6 crack sizes considered are summarized in Table 4. In all cases, the 9PA and AA measurement groups are not statistically different ($p > 0.05$).

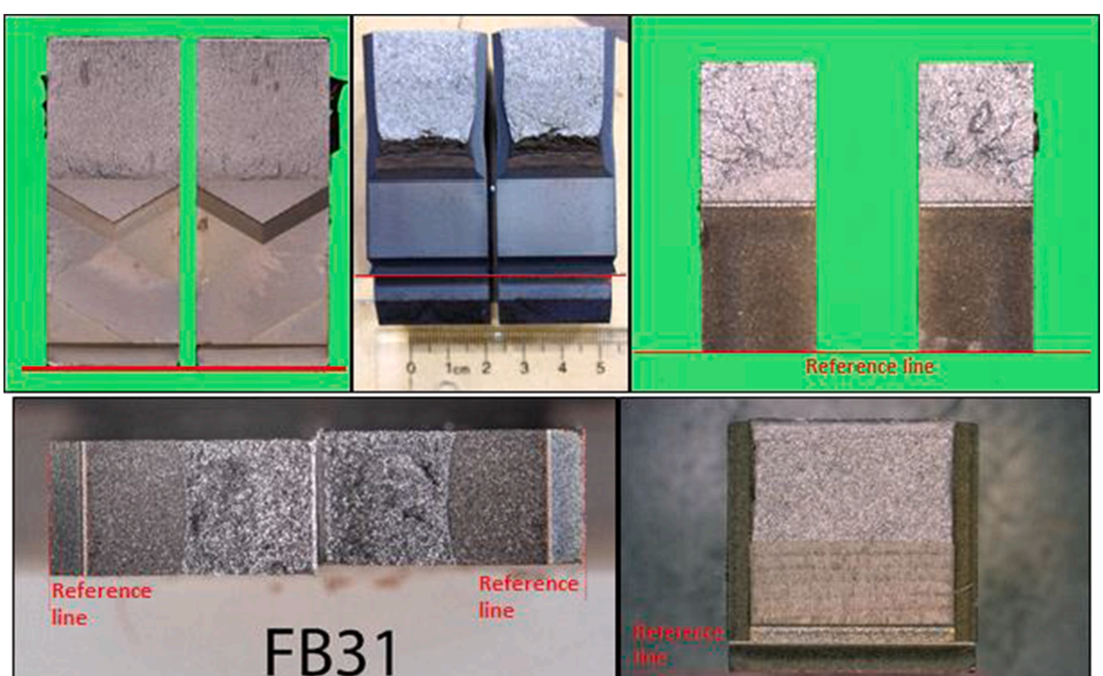

Fig. 2. Digital pictures provided to participants for crack size measurements (starting top left and going clockwise: 0.5TC(T), 1TC(T), MC(T), 20 % SG PCCv, MSE (B)). Reference lines are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Individual crack size measurements provided by the 15 participating laboratories.

Lab #	0.5TC(T) $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)	1TC(T) $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)	$a_{f,9pA}$ (mm)	$a_{f,AA}$ (mm)	MC(T) $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)	MSE(B) $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)	20 % SG PCCv $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)
1	18.14	18.13	25.19	25.09	29.14	28.40	5.94	5.93	1.65	1.63	5.24	5.23
2	17.76	17.76	25.26	25.47	29.15	29.22	5.95	5.97	1.63	1.64	5.33	5.32
3	18.03	18.06	24.76	24.84	28.18	28.26	5.96	5.97	1.63	1.61	5.10	5.10
4	17.83	17.94	24.78	25.04	28.46	28.47	5.94	5.99	1.63	1.63	5.18	5.00
5	17.96	18.02	24.51	24.62	28.38	28.50	5.98	5.97	1.63	1.62	5.14	5.16
6	18.08	18.09	24.47	24.66	28.34	28.44	5.95	5.95	1.64	1.64	5.17	5.18
7	17.98	17.94	24.72	24.57	28.11	27.94	5.95	5.92	1.63	1.64	5.14	5.13
8	17.87	18.15	24.57	24.41	28.44	27.04	6.00	5.94	1.62	1.64	5.15	5.16
9	17.72	17.79	24.64	24.86	28.38	27.98	5.99	5.96	1.65	1.63	5.17	5.17
10	18.06	18.01	25.05	25.01	28.16	28.17	5.91	5.89	1.64	1.66	5.14	5.20
11	17.76	17.73	25.03	25.03	28.89	28.81	5.98	5.97	1.63	1.63	5.21	5.21
12	17.78	17.82	24.80	24.91	28.67	28.41	5.93	5.96	1.63	1.64	5.16	5.18
13	17.69	17.64	24.77	24.57	28.59	28.02	5.91	5.91	1.63	1.62	5.13	5.11
14	18.03	18.02	24.72	24.67	28.33	27.85	5.92	5.90	1.64	1.62	5.19	5.20
15	18.01	17.89	24.75	24.72	27.95	28.11	5.95	5.96	1.64	1.64	5.21	5.15

Table 3
Basic statistics calculated from participants' results.

Statistic (mm)	0.5TC(T) $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)	1TC(T) $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)	$a_{f,9pA}$ (mm)	$a_{f,AA}$ (mm)	MC(T) $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)	MSE(B) $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)	20 % SG PCCv $a_{o,9pA}$ (mm)	$a_{o,AA}$ (mm)
Mean value	17.91	17.93	24.80	24.83	28.48	28.24	5.95	5.94	1.63	1.63	5.18	5.17
Standard deviation	0.15	0.16	0.23	0.27	0.36	0.49	0.03	0.03	0.01	0.01	0.06	0.07
Standard error	0.04	0.04	0.06	0.07	0.09	0.13	0.01	0.01	0.002	0.003	0.01	0.02
Minimum value	17.69	17.64	24.47	24.41	27.95	27.04	5.91	5.89	1.62	1.61	5.10	5.00
Maximum value	18.14	18.15	25.26	25.47	29.15	29.22	6.00	5.99	1.65	1.66	5.33	5.32
Range	0.45	0.52	0.79	1.06	1.21	2.18	0.09	0.10	0.03	0.05	0.23	0.32

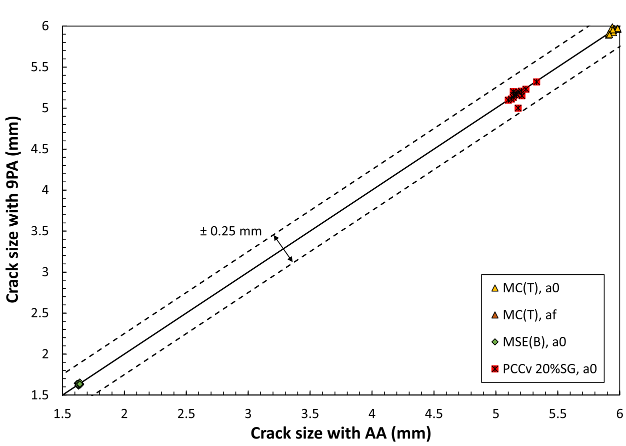

Fig. 3. Comparison between 9PA and AA measurements for crack sizes up to 6 mm.

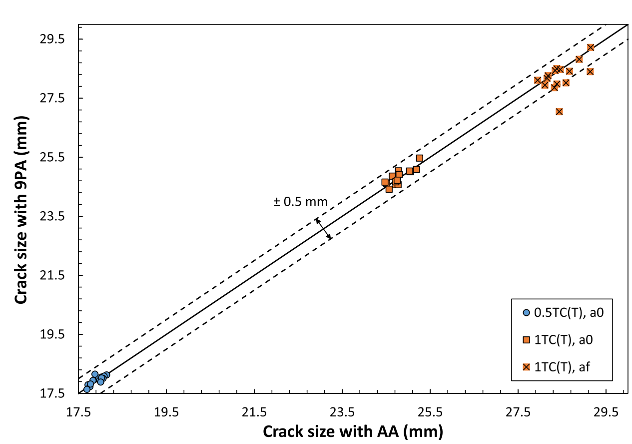

Fig. 4. Comparison between 9PA and AA measurements for crack sizes between 17.5 mm and 30 mm.

3.2. Determination of the precision of the measurement methods in accordance with ASTM E691-18

The ASTM E691-18 standard [5] is used for planning and conducting an interlaboratory study (ILS), and for analyzing its results so that the precision statement of a test method (repeatability and reproducibility) can be formulated.

In this case, E691 was used to establish and compare the reproducibility (between-laboratory) statistics of the two crack size measurement methods, 9PA and AA. No information on the repeatability (within-laboratory) of the methods could be generated, as participants reported single values of crack size for each specimen, and no details about the uncertainty of individual measurements was available.

For each of the available data sets (specimen/initial or final crack size/9PA or AA), the following parameters were calculated:

- average of crack size measurements, $\overline{\overline{x}}$;
- standard deviation of crack size measurements, $s_{\overline{x}}$.

For each individual crack size measurement ×, the following

Table 4
Results of t-tests on the groups of 9PA and AA measurements.

Statistic	0.5TC(T)		1TC(T), a_o		1TC(T), a_f		MC(T)		MSE(B)		20 % SG PCCv	
	9PA	AA	9PA	AA	9PA	AA	9PA	AA	9PA	AA	9PA	AA
Mean value	17.91	17.93	24.80	24.83	28.48	28.24	5.95	5.94	1.63	1.63	5.18	5.17
Variance	0.02	0.02	0.05	0.07	0.13	0.24	0.001	0.001	6.6E-5	1.5E-4	0.003	0.005
Degrees of freedom	28		28		28		28		28		28	
t statistic	0.34832		0.33336		1.510516		0.567753		0.308034		0.483679	
t critical (two-tail)	2.048407		2.048407		2.048407		2.048407		2.048407		2.048407	
***p* (two-tail)**	0.730207		0.741349		0.142116		0.574727		0.760336		0.632375	

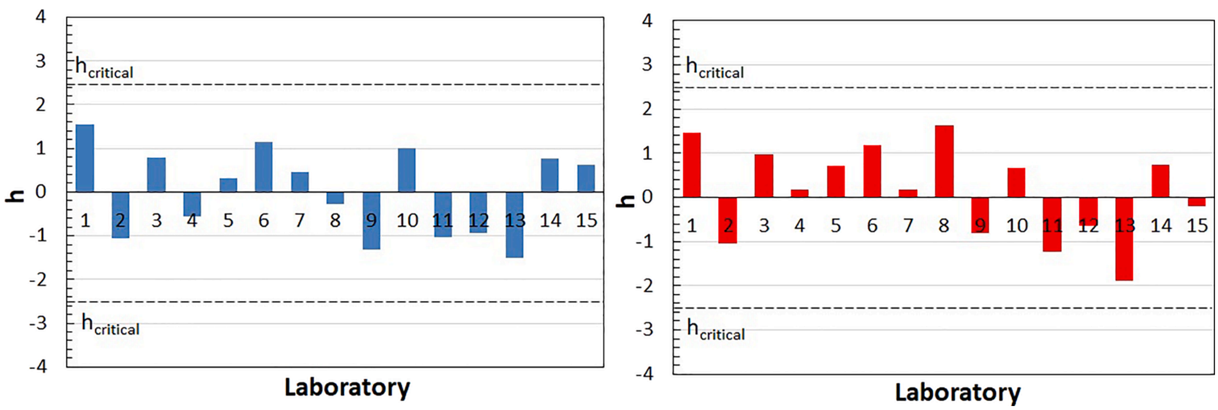

Fig. 5. Consistency statistic h for a_o measurements on the 0.5TC(T) specimen, using 9PA (left) and AA (right).

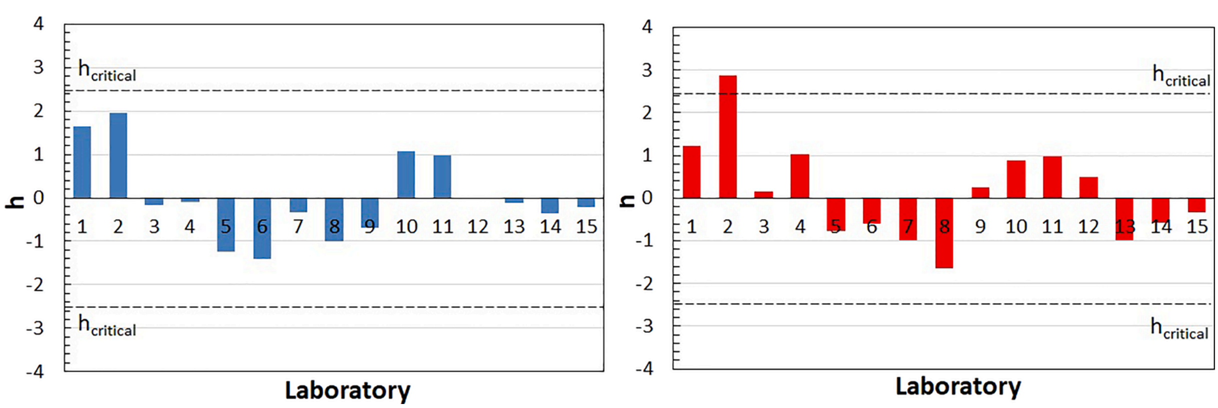

Fig. 6. Consistency statistic h for a_o measurements on the 1TC(T) specimen, using 9PA (left) and AA (right).

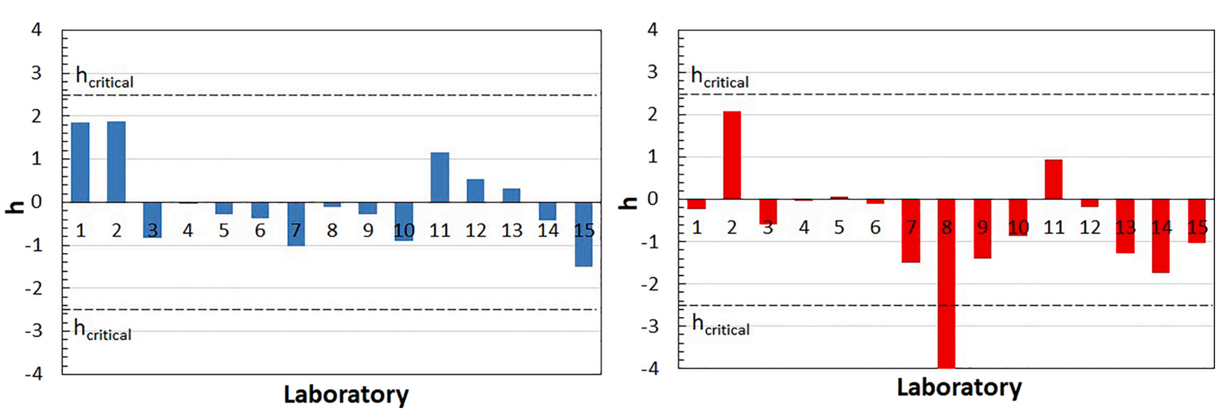

Fig. 7. Consistency statistic h for a_f measurements on the 1TC(T) specimen, using 9PA (left) and AA (right).

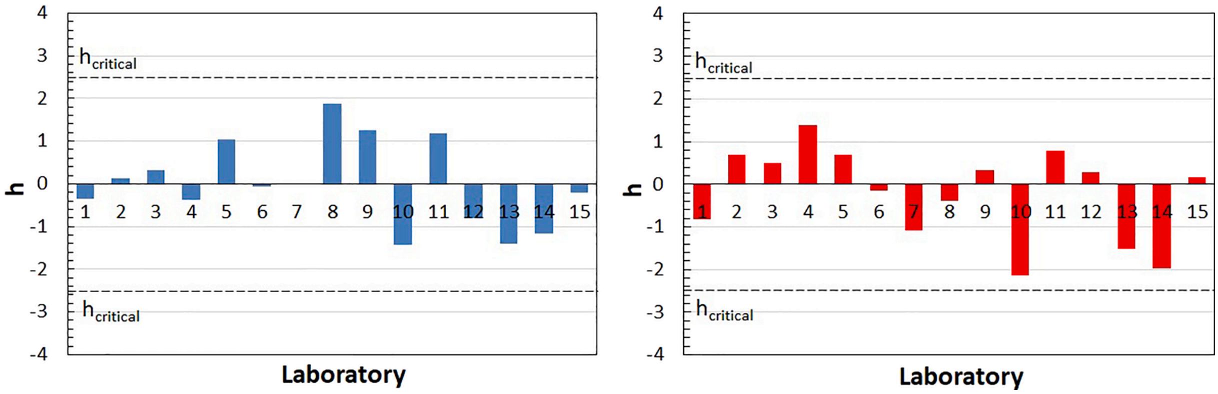

Fig. 8. Consistency statistic h for a_o measurements on the MC(T) specimen, using 9PA (left) and AA (right).

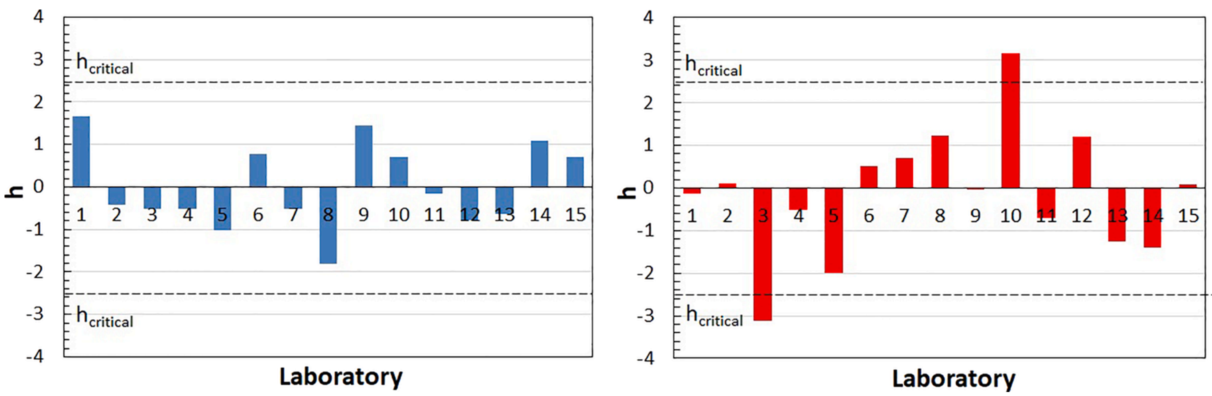

Fig. 9. Consistency statistic h for a_o measurements on the MSE(B) specimen, using 9PA (left) and AA (right).

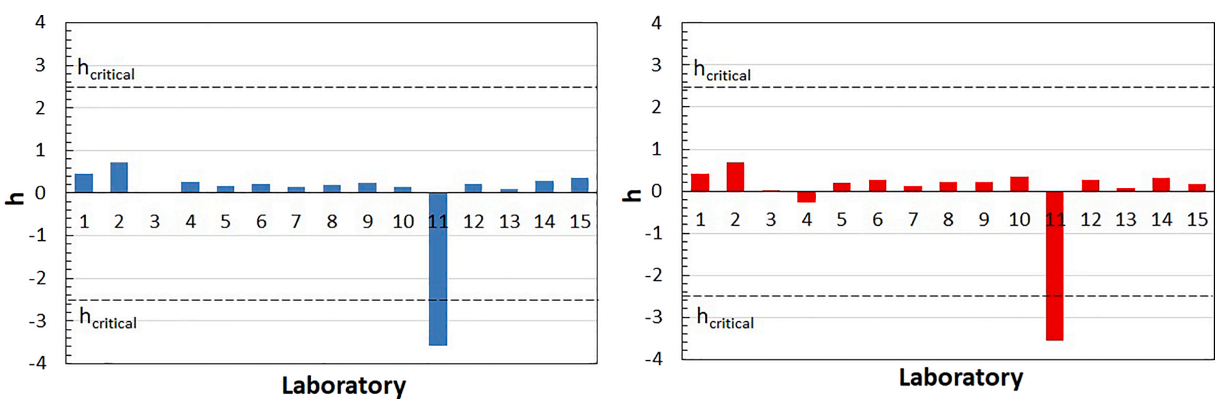

Fig. 10. Consistency statistic h for a_o measurements on the PCCv 20 % SG specimen, using 9PA (left) and AA (right).

statistics were calculated:

- deviation of the individual measurement from the measurement average, or cell deviation, $d = x - \overline{\overline{x}}$;
- repeatability consistency statistic, $h = \frac{d}{s_{\overline{x}}}$.

Cell deviations are plotted for each specimen/crack size/method in Figs. 5-10 and compared to a critical value, $h_{critical}$, corresponding to a 0.5 % significance level. The value of $h_{critical}$ depends only on the number of laboratories, p; in our case ($p = 15$), $h_{critical} = 2.47$.

Any measurements for which $h > h_{critical}$ (“flagged” cell) should be scrutinized for clerical, sampling, and/or procedural errors. If no such errors are found, the data should be retained in the calculation of the precision statistics.

Based on the examination of Figs. 5-10, the following six measurements were found to correspond to flagged cells:

(a) Lab 2, 1TC(T), a_0, AA method ($d = 2.86$).
(b) Lab 8, 1TC(T), a_f, AA method ($d =$ -4.02).
(c) Lab 3, MSE(B), AA method ($d =$ -3.11).
(d) Lab 10, MSE(B), AA method ($d = 3.17$).
(e) Lab 11, PCCv 20 % SG, 9PA method ($d =$ -3.57).
(f) Lab 11, PCCv 20 % SG, AA method ($d =$ -3.56).

No obvious mistakes were found for measurements (a), (b), (c), and (d) above. As for items (e) and (f), from the examination of the measurement report submitted by Lab 11 it was discovered that crack sizes had been measured from a different reference line than the one specified

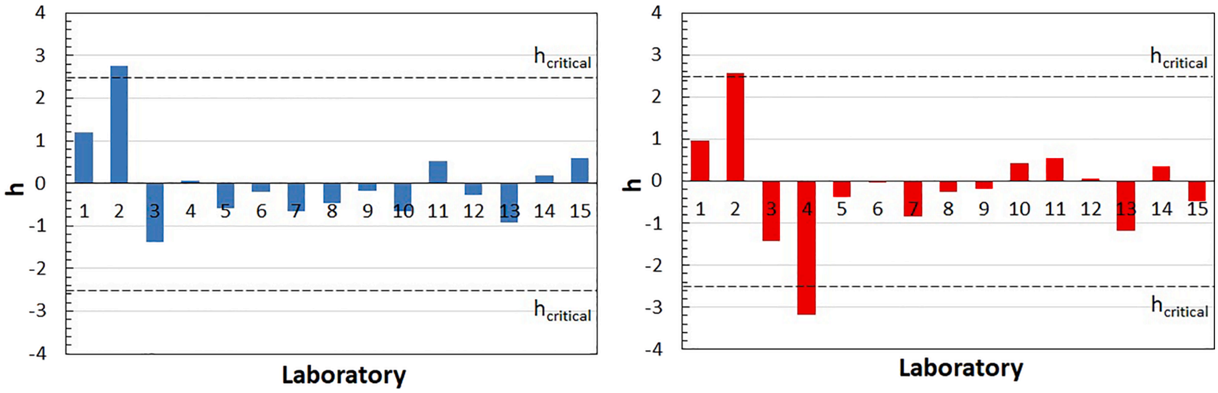

Fig. 11. Revised consistency statistic h for a_o measurements on the PCCv 20 % SG specimen, using 9PA (left) and AA (right), after correcting Lab 11 measurements.

Table 5
Repeatability precision statistics calculated for 9PA and AA crack size measurements from ILS results.

Specimen	Measurement	Method	$\bar{\bar{x}}$ (mm)	s_R (mm)	R (mm)
0.5TC(T)	a_o	9PA	17.91	0.15	0.41
		AA	17.93	0.16	0.44
1TC(T)	a_o	9PA	24.80	0.23	0.66
		AA	24.83	0.27	0.76
	a_f	9PA	28.48	0.36	1.00
		AA	28.24	0.49	1.36
MC(T)	a_o	9PA	5.95	0.03	0.08
		AA	5.94	0.03	0.08
MSE(B)	a_o	9PA	1.63	0.01	0.02
		AA	1.63	0.01	0.03
PCCv 20 % SG	a_o	9PA	5.18	0.06	0.16
		AA	5.17	0.07	0.20

in the original digital picture for the 20 % SG PCCv specimen. Measurements (e) and (f) could therefore be corrected and the revised consistency plots shown in Fig. 11 were obtained. After correction, all the *h* values changed, and two more flagged cells (for which no obvious errors were found) were observed:

(e) Lab 2, PCCv 20 % SG, 9PA method ($d = 2.75$).
(f) Lab 4, PCCv 20 % SG, AA method ($d = -3.18$).

We note that 5 out of 6 flagged cells correspond to AA measurements, which confirms that the AA method is subject to somewhat more variability than the 9PA method. This is to be expected, as the 9PA method only considers 9 points along the crack front, whereas AA accounts in principle for the whole crack front. We also note that the 6 flagged cells correspond to 5 different labs, indicating that none of the participants should be singled out for the questionable quality of their measurements.

Once all ILS results have been checked for possible errors, precision statistics can be calculated. In accordance with ASTM E691-18, when no repeatability information is available, the reproducibility standard deviation, s_R, is equal to $s_{\bar{x}}$. Furthermore, the 95 % reproducibility limit R, defined in accordance with ASTM E177-14 [6], is given by $R = 2.8 \cdot s_R$.

The final precision statistics calculated from the ILS results are summarized in Table 5.

4. Conclusions

The outcomes of an interlaboratory study (ILS) on crack size measurements using the 9-Point Average and Area Average methods, which involved 15 laboratories in North America, Europe, and Japan, fully corroborated the conclusions of a comparative study published by the author at the end of 2021.

Specifically:

- Measurements of initial and final crack sizes on fracture mechanics specimens of different size and configuration are in close agreement between the two methods investigated. This was verified by performing statistical *t*-tests on the different sets of ILS results, as well as by establishing the precision of the two measurement methods in accordance with ASTM E691-18 and ASTM E177-14.
- The variability of AA measurements is slightly larger than that of 9PA measurements.
- All 15 ILS participants demonstrated good proficiency in measuring crack sizes using both methods.

For initial crack fronts that are relatively uniform and straight, the excellent agreement between the two method is predictable. For more irregular and jagged final crack front, differences can be expected to be larger, but it is the author's opinion that the Area Average methods provides a more accurate representation of the actual crack front.

All things considered, the results of this ILS support the previously formulated recommendation to include the Area Average method in all fracture and fatigue test methods that require measuring crack sizes, as an alternative (not as a replacement!) to the widely popular X Point Average method (with X = between 2 and 9 points, depending on the specific test method).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author gratefully acknowledges the collaboration and timeliness of all ILS participants, many of whom provided detailed reports, and not merely measurement results.

References

[1] E. Lucon, An assessment of different approaches for measuring crack sizes in fatigue and fracture mechanics specimens, Theor. Appl. Fract. Mech. 116 (2021) 103119, https://doi.org/10.1016/j.tafmec.2021.103119.
[2] E. Lucon, Corrigendum to "An assessment of different approaches for measuring crack sizes in fatigue and fracture mechanics specimens", Theor. Appl. Fract. Mech. 117 (February 2022), 103230 https://doi.org/10.1016/j.tafmec.2021.103230.
[3] European Structural Integrity Society, *ESIS Procedure for Determining the Fracture Behaviour of Materials*, ESIS TC1 Subcommittee on Fracture Mechanics Testing Standards, January 1992.

[4] NIST Engineering Statistics Handbook, *1.3.5.3. Two Sample t-test for Equal Means*, https://www.itl.nist.gov/div898/handbook/eda/section3/eda353.htm (accessed 23 March 2022).

[5] ASTM E691-18, "Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method," ASTM International, West Conshohocken, PA, approved April 1, 2018.

[6] ASTM E177-14, "Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods," ASTM International, West Conshohocken, PA, approved May 1, 2014.