Evaluation of test results and ranking criteria for Varestraint testing of an austenitic high-temperature alloy

Mikael M. Johansson [1,2] · Peter Stenvall [1] · Leif Karlsson [2] · Joel Andersson [2]

Abstract

Varestraint testing is commonly used to evaluate hot cracking susceptibility of materials. In this paper, the dependence of operators and evaluation technique on test results is studied for a high-temperature austenitic stainless steel (UNS S31035). Samples were tested at six different strain levels ranging from 0.7 to 3.8%. Four different operators evaluated the same samples following the same instructions on how to measure the cracks manually in an optical microscope at × 25 magnification. The largest variation among operators evaluation was found for low strain levels where small and few cracks were found. In addition, one of the four operators used image analysis to evaluate the samples at × 50 magnification. The average total crack length and total number of cracks in fusion zone and heat-affected zone were approximately 1.5 times higher when using image analysis compared with manual evaluation. Image analysis at × 50 made it possible to detect smaller cracks compared with manual evaluation at × 25 magnification, contributing to an increased number of cracks detected.

The maximum crack length using image analysis at × 50 was similar to manual evaluation made at × 25 magnification and was the criterion that showed the least variation in this study. However, further comparisons using other magnifications are needed to verify the agreement between manual evaluation and image analysis found in this study. An advantage with evaluation using image analysis is that it provides traceable results. A harmonized standard for Varestraint testing, and especially for evaluation, would decrease the variation among operators and laboratories.

Keywords Varestraint testing · High-temperature austenitic stainless steel · Hot crack evaluation · Ranking criteria

1 Introduction

More than 100 different techniques have been developed to quantify hot cracking susceptibility over the years and some

✉ Mikael M. Johansson
mikael.m.johansson@sandvik.com

Peter Stenvall
peter.stenvall@sandvik.com

Leif Karlsson
leif.karlsson@hv.se

Joel Andersson
joel.andersson@hv.se

[1] Sandvik Materials Technology, SE-811 81 Sandviken, Sweden

[2] Department of Engineering Science, University West, SE-461 86 Trollhättan, Sweden

have been standardized [1, 2]. Recently, a round robin evaluation of different hot cracking tests used worldwide was initiated by Commission IX of the International Institute of Welding. The purpose of the study was to evaluate the test equipment, testing procedures, specimen preparation, and the test results provided by the participating organizations. The results from the round robin study gave conflicting predictions of cracking susceptibility both for different testing techniques and between different laboratories using the same testing technique [3]. It was stated that to achieve better correlation between test result development of standardized equipment and procedures was necessary.

The longitudinal Varestraint test quantifies weld solidification cracking in the fusion zone (FZ), liquation cracking, and ductility dip cracking (DDC) in the heat-affected zone (HAZ), whereas the Transvarestraint test is used primarily to quantify the weld solidification cracking [4]. Both methods are weldability tests with focus on one type of welding defect, namely hot cracks, and make it possible to rank alloys according to a hot cracking index (criteria) such as cracking

threshold, total crack length (TCL), or maximum crack length (MCL) [5]. A challenge when using Varestraint testing is to properly interpret the data within the limits of the test's usefulness in relation to real applications [6].

The major studies on evaluation of the Varestraint test include how different testing parameters correlate to the cracking response for a selected alloy [5–8]. Comparison of results of crack evaluation between operators is less common than testing of effects of different parameters. However, work has been performed which shows that a greater variation can be expected when two or more operators evaluate the same specimen, as compared with if all evaluation are done by the same operator, especially at the lower strain levels where cracks are small and few [6, 7].

As optical microscopes equipped with high-resolution digital cameras have become available, evaluation can also be performed using an image analysis software. This software can be used to measure cracks and the results can be recorded and saved as data, which makes results of evaluation traceable and controllable. This part of the evaluation could earlier only be carried out manually by utilizing a calibrated reticle in the eyepiece of the stereo microscope and writing down the length of each crack.

In this study, the longitudinal Varestraint test was used, from now on abbreviated Varestraint test, for hot cracking evaluation. All Varestraint test parameters were kept fixed while operators and techniques used for evaluation and measurements of the hot cracks were varied. The focus in this study is on how different evaluation techniques, commonly used for crack measurement, might affect the reproducibility and comparability of the results for Varestraint testing of the high-temperature austenitic stainless steel UNS S31035 base material (Sandvik Sanicro 25) [9]. This steel can potentially be used in super-heaters and reheaters in the next generation of advanced ultra-super critical power plants [10–13]. The material possesses both high creep strength and good corrosion resistance at temperatures up to 700 °C. Its high temperature properties are positioned between other commercially available high-temperature austenitic stainless steels and high-temperature nickel-based alloys. It is well known that an austenitic solidification mode combined with a fully austenitic microstructure makes alloys susceptible to hot cracking [14]. This will become an increased problem for thick-walled material in multipass welding and could compromise the integrity of the welded component. The steel was therefore used in the present evaluation of reproducibility in results of hot cracking testing.

In the present study, the evaluation was first carried out by using manual crack measurement in a stereo microscope by a number of operators, and secondly by utilizing digital imaging software to measure the cracks in the captured images. In many publications, different Varestraint test parameters are statistically evaluated, but crack evaluation is rarely discussed [7, 8, 15]. This work is therefore focusing on the difference between operators in manual evaluation and how the transition from manual to a digital measurement of hot cracks could affect the choice of a suitable Varestraint hot cracking criteria.

2 Experimental

2.1 Material

Sanicro 25 (UNS S31035) base material from bar was used for Varestraint testing. The chemical composition can be found in Table 1 and were tested on a Varestraint coupon from position 1 in Fig. 1.

Varestraint coupons were machined from bar material with dimensions: ⌀246 × 250 mm in length in the longitudinal direction. A figure that shows how the coupons were extracted can be found in Fig. 1. The final Varestraint coupon dimension after machining was width × length × thickness = 40 mm × 250 mm × 8 mm. To minimize the influence of possible segregations on Varestraint test results, since no homogenization was made on the material prior to testing, the center part and the outer 13 mm of the of the bar were excluded.

2.2 Varestraint testing

Varestraint testing was performed at the Sandvik R&D welding lab in Sandviken, Sweden. A total of 24 coupons were tested at 6 different radii. Autogenous TIG welding was made using parameters in Table 2. A pneumatically-actuated loading yoke was used for bending.

The removable die block of the desired radius was prepared and fastened prior testing. The arc was initiated on the centerline of the specimen, approximately 90 mm from the specimen's unclamped end. A bending force was applied as the center of the arc passed a point, close to the point of tangency between the curved surface of the die block and the fixed end of the specimen. The specimen was then bent downward until it conformed to the radius of curvature of the top surface of the die block. The rate of arc travel was constant from its point of initiation to its point of termination. The augmented strain ε

Table 1 Chemical composition (wt%) of UNS S31035 base material from bar on coupon from position 1 in Fig. 1

C	Si	Mn	P	S	Cr	Ni	W	Co	Cu	Nb	B	N	Fe
0.068	0.27	0.43	0.019	0.0005	22.2	25.23	3.41	1.47	2.92	0.47	0.0024	0.216	Bal.

was calculated using the following equation [16]:

$$\varepsilon = \frac{t}{2R + t} \qquad (1)$$

Here, t is the thickness of the coupon and R is the bend radius. The radii and corresponding strain levels used in this investigation can be found in Table 3.

2.3 Crack evaluation

In this study, Varestraint testing was following parts of both ISO/TR 17641-3 and ANSI/AWS B4.0-2007 [4, 16]. For manual crack evaluation, × 25 magnification was used as suggested in ISO/TR 17641-3. Varestraint samples were evaluated in the as-tested condition. Image analysis software was also used for the crack evaluation but at × 50 magnification instead of × 25, as suggested in ANSI/AWS B4.0-2007 (× 40–80).

Optical microscopy was performed using a Zeiss Discovery V8 stereo microscope with a LED-type of light source, capable of up to × 50 magnification and equipped with at 3 Megapixel color camera, for the crack measurement. The image analysis software and camera were calibrated and checked on a regular basis. Crack evaluation was performed by counting the total number of cracks and measuring the length of each crack that was found in the as-welded

specimen, in both FZ and HAZ. The following hot cracking criterions were used: total crack length (TCL), total number of cracks, and the maximum crack length (MCL).

A schematic illustration of different regions and boundaries used in the crack evaluation is seen in Fig. 2. The following regions can be found: fusion zone (FZ) and heat-affected zone (HAZ), and finally the fusion line (FL) boundaries.

The MCL was measured for each bending radius, which was the longest crack found in all four samples for each radius independent of location. The reported values of TCL are average values of total crack length for each tested radius.

Crack evaluation was done (1) by four different operators at × 25 magnification manually in the stereo microscope and (2) by one of the operators using image analysis of photos taken at × 50 magnification. The purpose was to see if the results would vary between operators when using the same instructions and to see how manual evaluation vary compared with image analysis. In the manual evaluation a calibrated, checked with a stage micrometer, micrometer reticle in the eyepiece was used and the cracks were measured from start point to the end point of the crack in a straight line, whereas in image analysis, measurements were performed following the actual shape of the cracks. In Fig. 3, a schematic figure (left photo) shows how the photos were taken on the Varestraint coupons; normally 4–8 photos were needed to cover all cracks at × 50 magnification. This was also the pattern followed and

Fig. 1 Schematic illustration of how Varestraint coupons were extracted from the cross section of UNS S31035 bar material, the figure is not according to scale. Chemical analysis methods used: HFIR, XRF, EXTR, OESG, and ICP
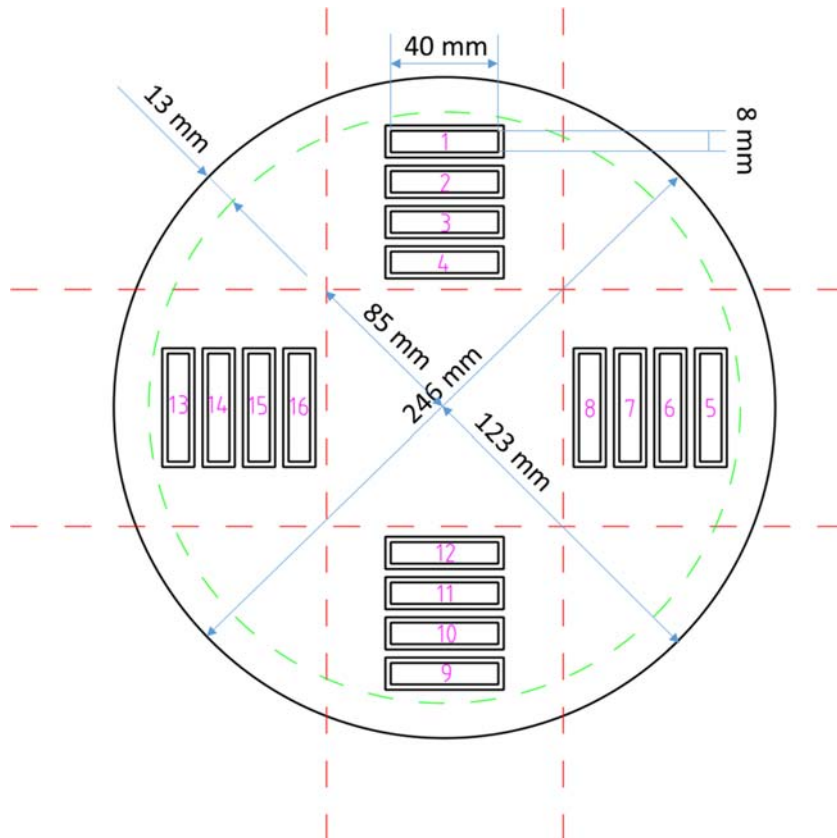
Table 2 Parameters for autogenous TIG welding in the Varestraint testing

Variable	Setting
Current	200 A
Voltage	14 V
Travel speed	150 mm/min
Arc energy	1.1 kJ/mm
Electrode	W-1.5La$_2$O$_3$, 2.4 mm diameter
Gas cup size (height)	10 mm
Arc length	2 mm
Shielding gas	100% argon
Gas flow	10 l/min
Ram speed	≥ 150 mm/s
Weld width	8–9 mm
Total weld length	150 mm
Distance between position of arc when bending and end of weld	60 mm

order of how the manual evaluation was performed. The red lines in the right photo in Fig. 3 illustrate crack measurements using the image analysis software. The extended depth of focus (EDF) tool box was used to combine multiple images into one single photo producing a sharp photo despite the limited depth of focus which is a consequence of using higher magnifications.

3 Results

Evaluation of Varestraint test specimens showed the presence of solidification cracks and liquation cracks. However, no indication of ductility dip cracking was found.

3.1 Crack evaluation

The average total crack lengths for manual evaluation by all operators in both FZ and HAZ are shown in Fig. 4. Operators 3 and 4 measured lower values at most strain levels and operator 1 measured the highest values at all strain levels; however, all were within one standard deviation from the average value of all operators. Furthermore, the average TCL manually evaluated at × 25 and with image analysis at × 50 by the same operator is found in Fig. 5. The highest values were

Table 3 Die block radii and strain levels used in the Varestraint testing

Radius (mm)	Strain (%)
100	3.8
135	2.9
200	2.0
270	1.5
400	1.0
540	0.7

measured with the image analysis software for all strain levels. However, error bars from both data series were overlapping at all strain levels except at 2% and 3.8%.

The average total crack lengths for manual evaluation by all operators in FZ and HAZ can be found in Fig. 6. It clearly shows the small contribution from HAZ cracks to the total amount.

The total number of cracks for manual evaluation by all operators in both fusion zone and heat-affected zone are shown in Fig. 7. Operator 1 measured the highest total number of cracks at all strain levels. The average total number of cracks for operator 1 and operator 3, at 2.9% and 3.8% strain levels respectively, were outside the error bars for the average value for all operators. Average numbers of cracks evaluated by the same operator manually at × 25 and with image analysis at × 50 magnification are shown in Fig. 8. Significantly higher values for average total number of cracks were measured when using the image analysis software at all strain levels except for the 1% strain level, where the one standard deviation error bars were overlapping.

The maximum crack lengths for manual evaluation by all operators in both FZ and HAZ are shown in Fig. 9. Small variations can be observed between operators with operator 3 having the largest deviation from the average of all operators. The MCL of UNS S31035 base material was evaluated by the same operator both manually at × 25 and with image analysis at × 50 magnification, and the results are presented in Fig. 10. No significant difference between the two data series can be found.

4 Discussion

The purpose of the study was twofold: (1) to analyze how results varied between different operators evaluating the same

Fig. 2 Illustration of different regions and boundaries that were used in the crack evaluation of UNS S31035 base material

samples and (2) to compare image analysis at ×50 magnification to manual evaluation in a stereo microscope at ×25 magnification. The magnifications used represented suggested magnifications in ISO/TR 17641-3 and AWS B4.0M for Varestraint testing [4, 16]. Three different criteria were used for each tested strain level: (1) the total crack length, (2) the total number of cracks, and (3) the maximum crack length. These will first be discussed individually; comparisons with pros and cons for the different criteria used are then given and what needs to be further investigated is commented on.

4.1 Total crack length

For manual evaluation of TCL at ×25 magnification, it could be seen that all operators reported values within the error bars for the average of all operators (Fig. 4). When comparing evaluations made by the same operator (operator 1) but with different methods, manual evaluation at ×25 and image analysis at ×50 magnification, it was clearly seen that image analysis results in higher TCL. The difference is not that large at lower strain levels, but above 1% a clear difference can be observed (Fig. 5). The
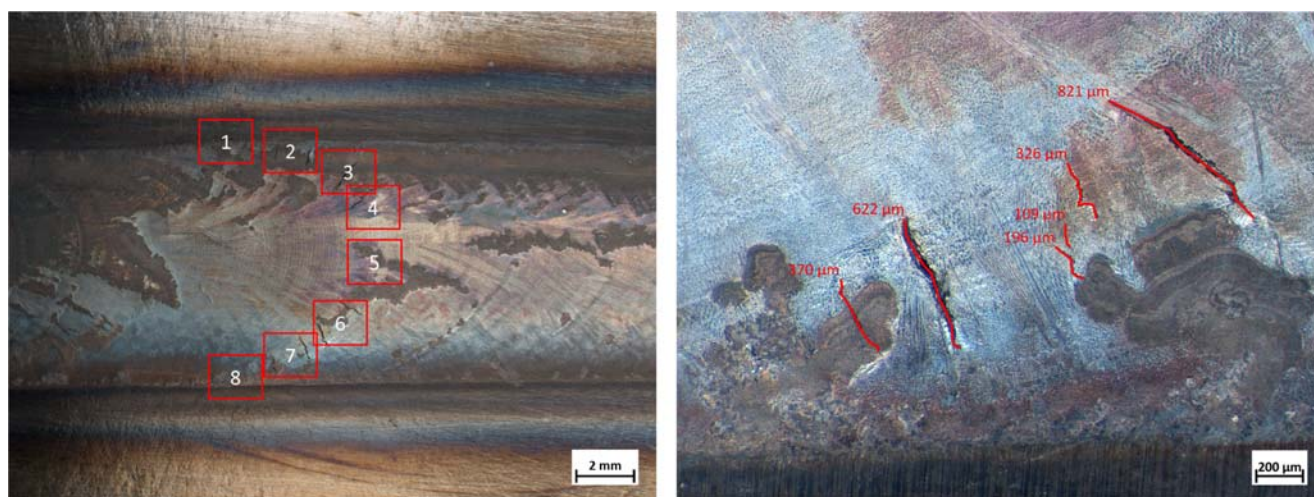
Fig. 3 Schematic representation of how photos were taken, red boxes, when using image analysis to measure cracks on the Varestraint coupons (left photo). Illustration of how cracks were measured inside the red boxes, red lines, at ×50 original magnification (right photo)

Fig. 4 Average total crack length (TCL) in FZ + HAZ versus augmented strain for UNS S31035 base material manually evaluated by four different operators at × 25 magnification. Operator 1 measured the highest TCL at all strain levels, but all operators were within one standard deviation from the average TCL of all operators. Error bars represent one standard deviation from average value of all operators
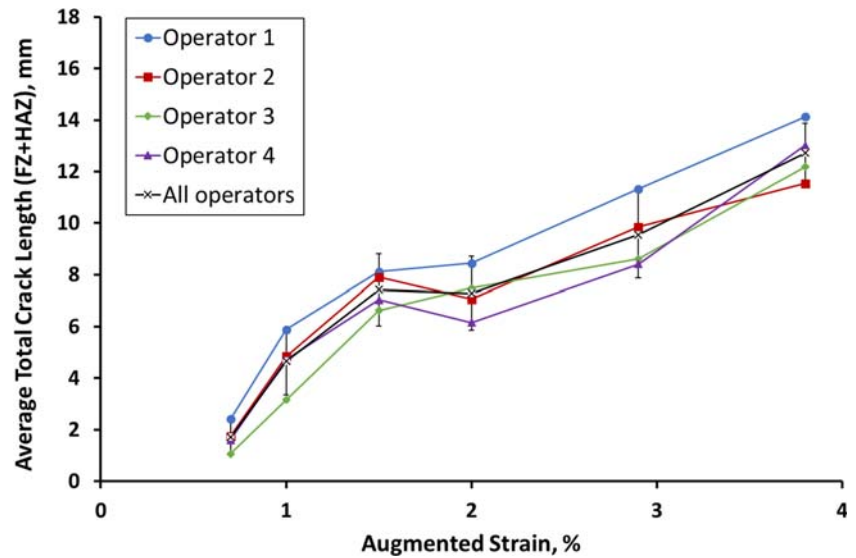

higher TCL measured by image analysis might be explained by the following factors. (1) The higher magnification used in image analysis × 50 compared with × 25 for manual evaluation could give higher accuracy and (2) the difference in crack length when following the crack shape in image analysis compared with using straight lines to measure from the start point to the end point of the crack. The TCL criterion will not be affected by whether the operator considers a crack to consist of many small or fewer larger cracks, since it is the total length of all measured cracks found in the specimen that is evaluated [7].

4.2 Total number of cracks

Manual evaluation of total number of cracks at × 25 magnification gave some results outside the error bars for operator 1 and operator 3 (Fig. 7). The difference between manual evaluation at × 25 and image analysis at × 50 was much larger than for the

evaluation of TCL and shows that only the results at 1% strain level had overlapping error bars (Fig. 8). Also, the higher measured total number of cracks for image analysis might be explained by the following factors. 1) The higher magnification of × 50 used in image analysis as compared with × 25 for manual evaluation will find smaller cracks. (2) Whether the operator considers a crack to consist of many small or fewer larger cracks. The first factor might be more influential for the total number of cracks since a higher magnification, especially at the lower strain levels, makes it easier to find the smallest cracks. However, the second factor will probably have a higher impact when different operators evaluate the same specimen compared with when the same operator uses manual evaluation at × 25 compared with image analysis at × 50. This factor is also dependent on the magnification to some extent, since if a low magnification is used, then it might be impossible to judge if it has one crack or many small cracks.

Fig. 5 Average total crack length (TCL) in FZ + HAZ versus augmented strain for UNS S31035 base material evaluated by the same operator, manually at × 25 and with image analysis at × 50 magnification. The highest TCL was measured at all strain levels when using image analysis. At 2% and 3.8% strain levels, the two data series were fully separated at the one standard deviation level. Error bars represent one standard deviation from average value
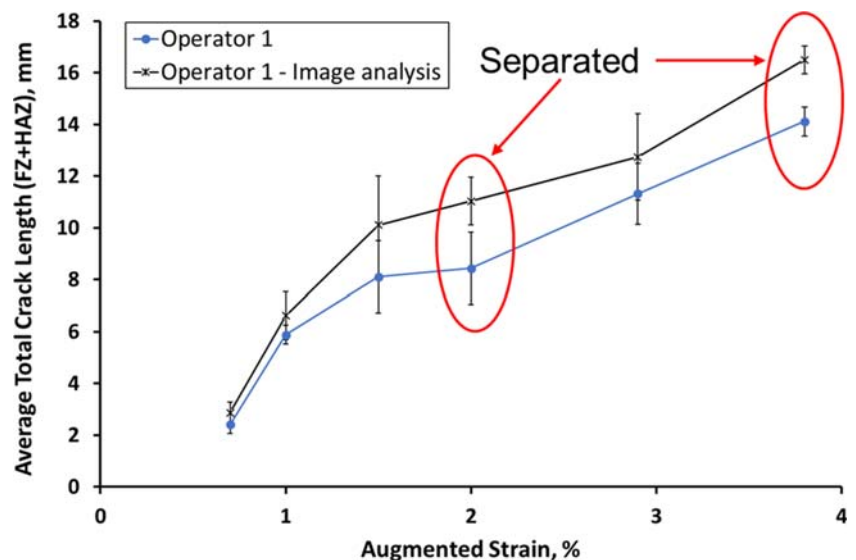
Fig. 6 Average total crack length (TCL) in FZ and HAZ versus augmented strain for UNS S31035 base material manually evaluated by four different operators at × 25 magnification. Operator 1 measured the highest TCL at all strain levels and in both regions. Error bars represent one standard deviation from average value of all operators
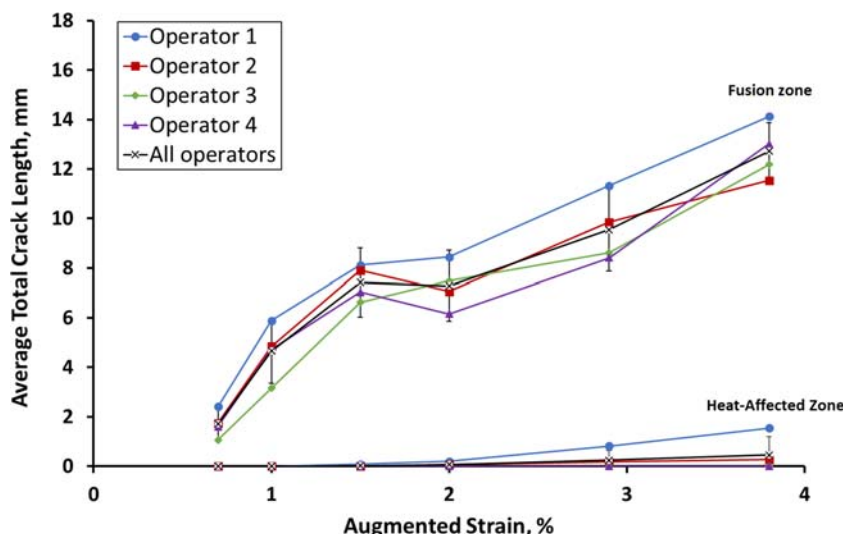
4.3 Maximum crack length

When the MCL criterion was used and compared for operator 1 with manual evaluation at × 25 with image analysis at × 50 magnification, a different behavior could be seen compared with the other evaluated criteria (Fig. 10). No significant difference between the two data series can be found, which indicates that the MCL criterion is not very sensitive to the choice of evaluation method compared with the TCL and total number of cracks criteria in this study. Since the MCL is the longest crack measured in each sample and strain level [18], a lower magnification would still be enough to measure a relative accurate length compared with when using a higher magnification. Furthermore, if measurement is carried out using straight lines to measure from the start to the end point of the crack instead of following the actual crack path, the difference in crack length will not be that significant since it is the longest

crack that is evaluated. A small difference in length will have a higher percentage difference for a short crack compared with a minor difference in length for a longer crack.

4.4 Operators influence on results

The coefficient of variation (CV) is a measure of relative variability. It is the ratio of the standard deviation to the mean (average). As it is a dimensionless number, it is particularly useful when comparing data sets with different units [17]. The coefficient of variation was therefore calculated for the manual crack evaluation by the four operators for average TCL, average total number of cracks, and average MCL and is presented in Fig. 11.
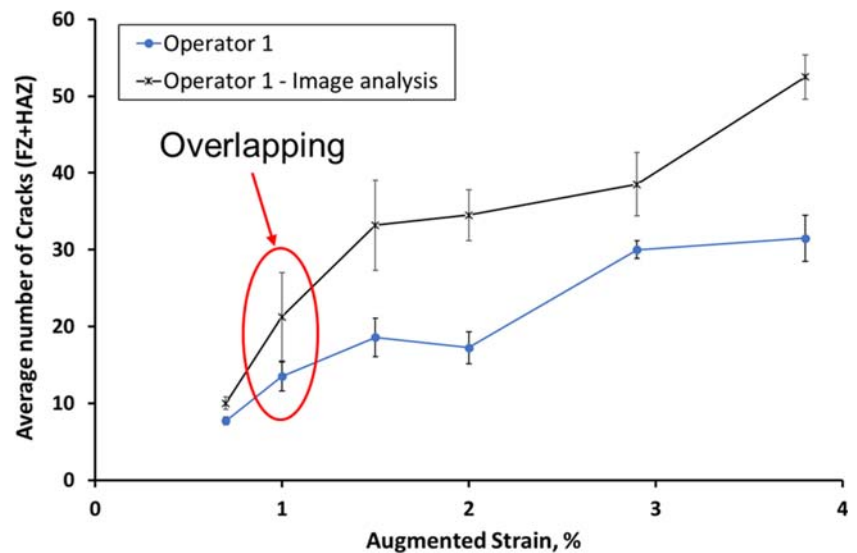
Coefficient of variation $\qquad$ (2)

$$= (\text{standard deviation}/\text{mean}) \times 100, c_v = \frac{\sigma}{\mu} . \times 100$$

Fig. 7 Average total number of cracks in FZ + HAZ versus augmented strain for UNS S31035 base material manually evaluated by four different operators at × 25 magnification. The highest number of cracks was measured by operator 1. Operator 1 and operator 3 measurements were outside the error bars at 2.9% and 3.8% strain levels. Error bars represent one standard deviation from average value of all operators
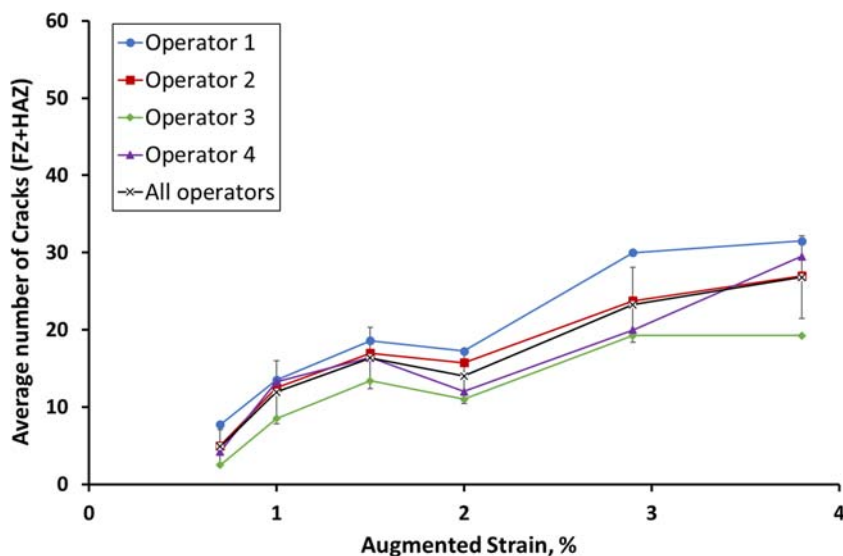
Fig. 8 Average total number of cracks in FZ + HAZ versus augmented strain for UNS S31035 base material evaluated by the same operator, manually at × 25 and with image analysis at × 50 magnification. The highest number of cracks was measured for image analysis at all strain levels except for 1% where error bars were overlapping. Error bars represent one standard deviation from average value
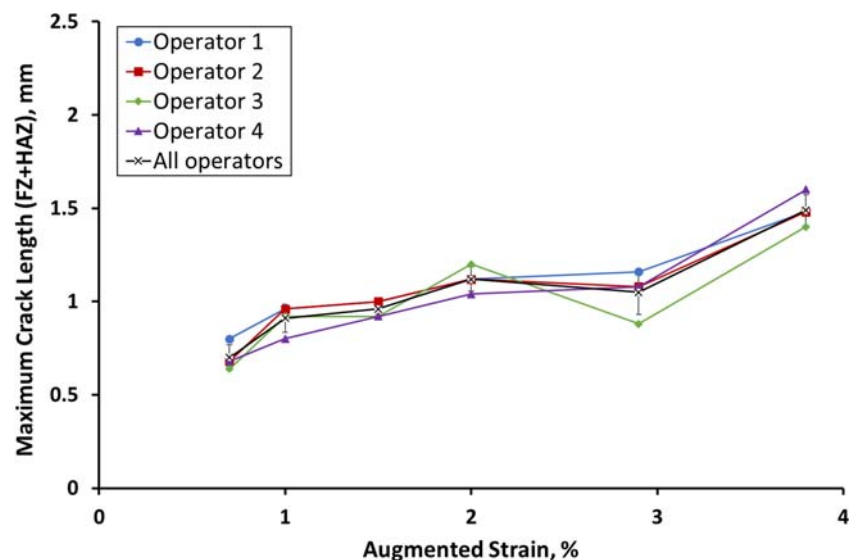
It was found that at low strain levels, resulting in low TCL, the coefficient of variation (variation between operators) was larger, which was also the case for the criterion "total number of cracks". This relation was also found for average TCL criterion by other authors [7]. It has earlier been reported in studies of the correlation for TCL that a 10% variation can be expected [6]. Furthermore, other authors found that the operator's subjectivity, in particular when judging what is a crack and what is not a crack, as well as lack of training, could give a greater variation [7]. The smallest variation was observed for the MCL criterion followed by TCL and number. of cracks. Furthermore, the lower sensitivity of the MCL criterion to operator variation compared with the other two criteria can clearly be observed in Fig. 11.

4.5 Comments on ranking criteria

Different recommendations are given in ISO/TR 17641-3 and ANSI/AWS B4.0 regarding Varestraint testing and parameters

for evaluation, which makes it difficult to compare results between different operators and experiments. Post weld cleaning (if performed) of the welds prior to hot crack evaluation can vary between organizations and laboratories giving different prerequisites. A harmonized standard for Varestraint testing, especially for the evaluation part of the test, would improve reproducibility and decrease the scatter of the test results between different operators and laboratories. This has been proposed for the Varestraint test by different authors over time [3, 18, 19].

The results from the Varestraint hot cracking ranking criteria used in this study, using fixed welding and Varestraint testing parameters, indicate that the MCL criterion probably is the best candidate to use if comparison should be made between different operators within the same study. Since MCL correlates to the solidification temperature range for a given welding procedure and particular level of augmented strain, MCL will probably be a better predictor of the crack susceptibility than TCL from a metallurgical point of view. This was also the criteria that had the

Fig. 9 Maximum crack length (MCL) in FZ + HAZ versus augmented strain for UNS S31035 base material manually evaluated by four different operators at × 25 magnification. Error bars represent one standard deviation from average value of all operators

Fig. 10 Maximum crack length (MCL) in FZ + HAZ versus augmented strain for UNS S31035 base material evaluated by the same operator manually at × 25 and with image analysis at × 50 magnification. Both data series are very similar in shape and there is no significant difference between them. Error bars represent one standard deviation from average value
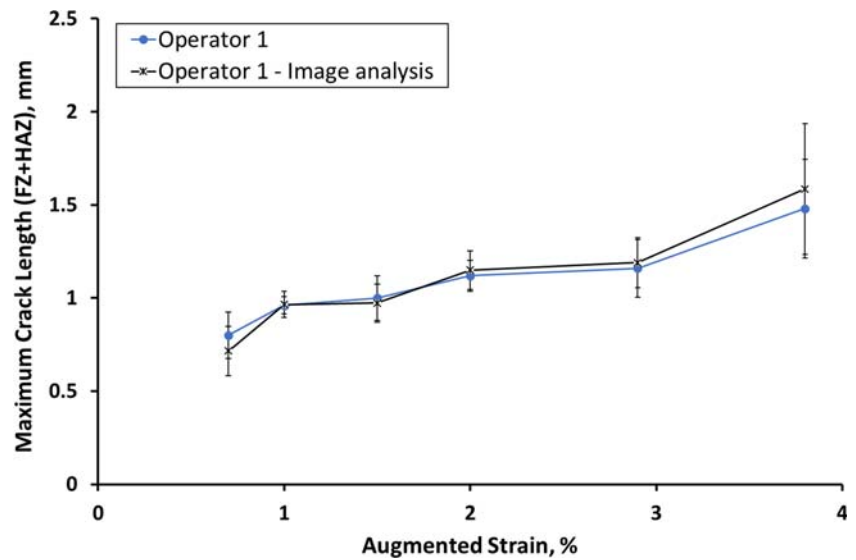
smallest difference between manual evaluation and image analysis. However, factors such as welding parameters, stroke rate, and sample thickness will probably overshadow the influence of the evaluation part of the test if they vary when comparing results between organizations and laboratories using different equipment and procedures. Other authors have reported that the TCL criterion is best suited as a quantitative index of the cracking sensitivity if the Varestraint test conditions are varied [5].

Work has been performed to statistically find the best Varestraint testing parameters for materials for a specific equipment [7, 8, 15]. It is demonstrated that moving from analog to digital measurements is likely to reveal more and longer cracks in
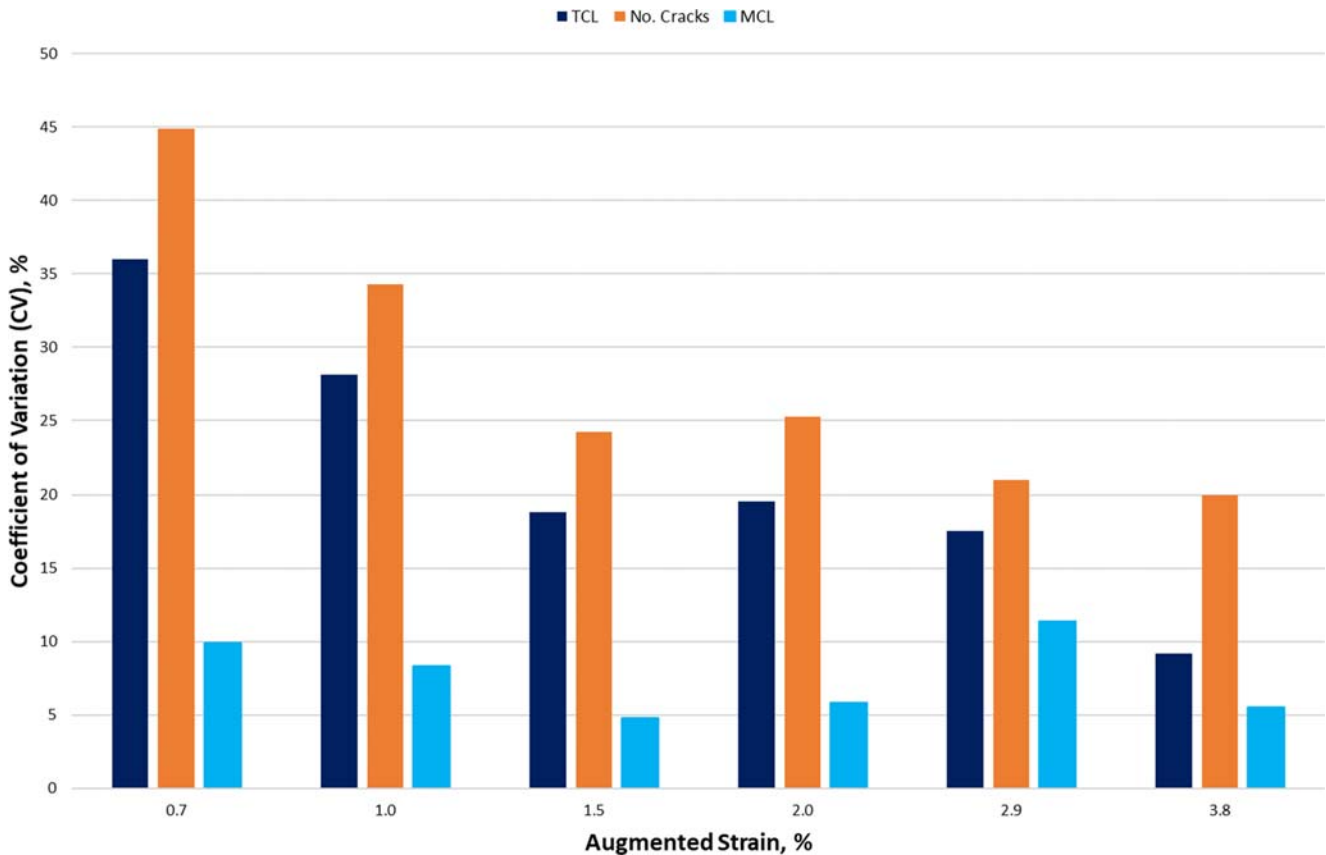
Fig. 11 Coefficient of variation for three Varestraint hot crack evaluation criteria expressed in percent: average TCL for all operators, average total number of cracks for all operators, and average MCL for all operators when manually evaluating UNS S31035 base material at × 25 magnification. Average values are based on four measurements at each strain level

weldability test samples than analog measurements. In addition to magnification, camera resolution and quality of the optical system and lighting are going to play a major role in the transition to digital. However, the importance from our study that we would like to stress is the fact that digital image analysis provides traceable results which is very important, not at least from an industrial perspective. Further investigations would therefore be valuable to increase understanding of differences between manual evaluation and image analysis.

5 Conclusions

The influence of operators and evaluation technique on results of Varestraint testing at six strain levels was evaluated for a high-temperature austenitic stainless steel (UNS S31035). Four operators evaluated TCL, number of cracks, and MCL using manual crack measurement at × 25 magnification and one of these also used image analysis at × 50 magnification. The following was concluded:

- The individual operator average TCL was within one standard deviation of the average value of all operators. A similar standard deviation was found by manual and image analysis.
- The average TCL and total number of cracks were approximately 1.5 times higher when using image analysis compared with manual evaluation.
- The difference in TCL can be explained by: (1) the crack path was followed in image analysis while straight lines were used in manual evaluation and (2) the higher magnification in image analysis made it possible to detect smaller cracks.
- The MCL results achieved with image analysis were similar to those of manual evaluation and were the criterion that showed the best agreement between the two methods.
- The largest variation between operators for manual evaluation was found for low strain levels where small and few cracks were found.
- Varestraint hot crack evaluation using image analysis provides traceable results. However, further studies of how magnification affects the results would be valuable.

References

1. Kannengiesser T, Boellinghaus T (2014) Hot cracking tests—an overview of present technologies and applications. Weld World 58:397–421
2. Lippold JC (2005) Recent developments in weldability testing for advanced materials. In: Joining of Advanced and Specialty Materials VII (#05116G). ASM International
3. Finton TL, Lippold JC (2002) Comparison of weld hot cracking tests–summary of an iiw round robin study. EWI Report No MR0205
4. ISO/TR 17641–3 (2005) Destructive test on welds in metallic materials–hot cracking test for weldments–arc welding processes–Part 3: External loaded tests
5. Savage WF, Lundin CD (1965) The Varestraint test. Weld J 44(10): 433–442
6. Lingenfelter AC (1972) Varestraint testing of nickel alloys. Weld J 51(9):430–436
7. Andersson J, Jacobsson J, Lundin CD (2016) A historical perspective on Varestraint testing and the importance of testing parameters. In: Crack Phenomena in Welds IV, 3–23
8. Finton TL, Lippold JC (2004) Standardization of the transvarestraint test: a statistical study using austenitic stainless steels and nickel-base alloys. EWI Report No MR0402
9. Sandvik Sanicro 25 (Tube and Pipe, seamless): https://www.materials.sandvik/en/materials-center/material-datasheets/tube-and-pipe-seamless/sanicro-25/, Accessed 2019-10-10
10. Przemyslaw J, Maria S (2013) Characteristics of the structure of welded joints in Sanicro 25 steel. Solid State Phenom 212:71–74
11. Chai G, Bodström M, Olaison M, Forsberg U (2013) Creep and LCF behaviors of newly developed advanced heat resistant austenitic stainless steel for A-USC. Procedia Eng 55:232–239
12. Rautio R, Bruce S (2005) Sandvik Sanicro 25, a new material for ultra supercritical coal fired boilers. Proceedings of the 4th International Conference on Advances in Materials Technology for Fossil Power Plants, 274–290
13. Chai G, Nilsson J-O, Boström M, Högberg J, Forsberg U (2010) Advanced heat resistant austenitic stainless steels. Proceedings of ICAS 2011 56
14. Lippold JC (2015) Welding metallurgy and weldability, p100. Wiley, Hoboken
15. Lee JW, Nichols EE, Goodman S (1968) Varestraint testing of cast 70 Cu–30 Ni Alloy. Weld J 47(8):371–377
16. Standard methods for mechanical testing of welds (2007). ANSI/AWS B4.0–2007, American Welding Society
17. Everitt B (1998) The Cambridge dictionary of statistics. Cambridge University Press, Cambridge, UK New York ISBN 978-0521593465
18. Lundin CD, Lingenfelter AC, Grotke GE, Lessman GG, Matthews SJ (1982) The Varestraint test. WRC Bull 280
19. Wilken K (1999) Investigation to compare hot cracking tests–externally loaded specimen. doc. IX-1945-99, IIS/IIW-Subcommission IX-H/II-C