Reliability research on automatic ultrasonic testing (AUT) technology of land long distance pipeline with automatic welding joint

Yanhua Hu1,2, Huli Niu1,2, Deyu Tang1,2, Jian Liu1,2, Huaqing Yang1,2, Yanwei Shen1,2

1 CNPC Engineering Technology Research Co. Ltd, No. 40 Jintang Road Tanggu Tianjin, 300451, P. R., China
2 Key Laboratory of Offshore Engineering of CNPC, No. 40 Jintang Road Tanggu Tianjin 300451, P.R., China

E-mail address: huyh@cnpc.com.cn (Yanhua Hu)

Abstract. The complete automatic ultrasonic testing (AUT) reliability theory and evaluation system were established for automatic welding joint of long distance pipeline. The reliability of AUT system was verified by several indexes, including Probability of Detection (POD), Probability of Sizing (POS), Probability of False Alarm (PFA), Probability of Rejection (POR) and Receiver Operating Characteristic (ROC). In order to evaluate AUT process and analysis reliability of AUT system, the sample size was optimized to 46 and the quantitative data of defects in AUT, RT and macroscopic metallographic section were analyzed statistically, which could directly provide reliable quantitative bases for the automatic welding quality acceptance of AUT Technology. The results show that POD and reliability of AUT are not less than RT, which can provide support for AUT application to automatic welding of the domestic long distance pipeline.

1. Induction

In the past, both X-radiographic testing (RT) and automatic ultrasonic testing (AUT) were used for quality inspection of domestic pipeline welding joints. When the results of AUT were consistent with those of RT, the AUT process was considered to be credible and could be used for pipeline detection. Actually, the detection principle as well as criteria of AUT is different from those of RT. Thus the reliability of AUT for pipeline quality acceptance should not be verified by RT.

In this paper, a complete automatic ultrasonic testing (AUT) reliability theory and evaluation system, including the generalized reliability index and its reliability analysis method, were established for automatic welding joint of long distance pipeline. In order to evaluate the reliability of AUT process, the sample size was optimized and the quantitative data of defects in AUT, RT and macroscopic metallographic section were analyzed statistically to provide reliable quantitative bases for the automatic welding quality acceptance of AUT Technology [1-5].

2. Inspection requirements of automatic welding construction

Nowadays, automatic welding technology is widely used in long distance pipeline construction. The construction speed of the pipeline operation is so fast that it is necessary for the corresponding nondestructive testing technology to be developed in high speed. AUT can produce detection results
on site with high detection efficiency. Additionally, it can detect the on-site automatic welding quality in real time, which is favorable for the welding unit to adjust the welding parameters in time, so as to ensure the welding quality of the welding joint [6].

The welding quality of automatic welding joint mainly depends on the control precision of the preset welding parameters. Incomplete fusion defects are common in automatic welding process. Interlayer incomplete fusion defects on the X-ray film are not generally shown or displayed, while they can be clearly seen in the AUT scanning chart. Besides, AUT can precisely locate the depth of defects by A scanning and TOFD scanning, which helps to determine whether the damage of the defects need repair. In comparison, RT can only rely on the projection principle to locate the position of defects and this method is easy to produce positioning error [6]. Therefore, AUT detection is recommended for the detection of automatic welding construction, and RT detection can also be used as needed.

3. Establishment and analysis of generalized reliability evaluation index [7]

In this paper, the reliability of AUT system was verified by five indexes, which include Probability of Detection (POD), Probability of Sizing (POS), Probability of False Alarm (PFA), Probability of Rejection (POR) and Receiver Operating Characteristic (ROC). These indexes can solve the technical problem of quantitative evaluation of AUT generalized reliability from the aspects of human, machine, material, method and environment and are helpful for the comprehensive evaluation of the application ability of AUT detection technology.

3.1. Confidence level (CL)

CL refers to the confidence level of detection probability, that is, the degree of probability assurance, also known as reliability or confidence coefficient. There is a correspondence between the confidence level and the sample size (n), the confidence interval (in a certain confidence level, the error range between the sample statistics and the general parameter values). Generally speaking, the higher the confidence interval is, the higher the confidence level is; the higher the confidence level is, the larger the sample capacity is.

3.2. POD and PFA

In the process of nondestructive testing, the normal distribution curve of probability density can be obtained if the test is repeated. As the noise in the nondestructive testing system also obeys the normal distribution, the output signal beyond noise amplitude is generally considered as the signal of defects. In the curve, when the defect detection signal is greater than the set threshold, it means that defects can be detected, which can be expressed as POD. Meanwhile, if false alarm is detected, PFA can be statistically counted.

The discrimination of flaw responses from application noise responses is a simple process; POD will be high, and the PFA will be low. In practical engineering applications, the flaw size is not fixed (and is rarely large), and the discrimination process is more complex. Indeed, the discrimination process is applied to a continuous range of flaw sizes, where the capability for discrimination is dependent on the inherent performance characteristic of the nondestructive testing procedure and on the separation of the signal (plus noise) from the noise response of the process.

3.3. POR

The method used for the evaluation of POR is the same as POD. The difference between POD and POR is defined by the criteria used on the observations for “hit-miss”. A POD is based on hit-miss by the initial reporting measure at the AUT system, usually echo amplitude or indication in TOFD. The POR is based on hit-miss evaluated from the same measure as used in the acceptance criteria, often defect height in mm.
3.4. POS
POS should be evaluated by an accommodation of the sizing inaccuracy observations to the normal distribution and determines the quantitative accuracy of the defect prefabrication and interpretation, which directly affects the accuracy of other reliability indexes such as POD, POR.

3.5. ROC
ROC is an important index to evaluate the reliability of AUT system, such as operator, equipment and process, and the controlled state of the system. Generally speaking, the system has higher reliability under the condition of larger POD and lower PFA.

4. Reliability evaluation of AUT system

4.1. Establishment of defect samples
The sample capacity is proportional to POD and CL. The structure of the pipeline is a typical structure with slow crack growth, and the reliability level should be set as 90% POD under 95% confidence level.

In this study, calculation model of the defect sample is binomial distribution model based on Hit-Miss. For the samples with same artificial defects, there are only two detecting results, detected or missed, so its capacity calculation can be described based on Hit-Miss binomial model.

\[
P(s = n) = \binom{N}{n} p^n q^{N-n}
\]

\[
\binom{N}{n} = \frac{N!}{n!(N-n)!}
\]

(1)

In Equation (1), \( N \) stands for the total sample size; \( S \) stands for the detectable sample; \( P \) stands for probability of detection: \( q \) stands for probability of missed; \( q=1-p \).

When the reliability level is determined (POD/CL=90/95), the solution of the equation (1) can be obtained, as shown in Table 1. Considering the complexity and uncertainty from personnel, equipment and construction environment on site, and factors combined with the construction experience of automatic welding and historical data of automatic ultrasonic detection, the sample capacity for reliability evaluation is set to be 46.

| Reliability level | Sample | Detectable sample |
|-------------------|--------|-------------------|
| 90/95             | 29     | 29                |
|                   | 46     | 45                |
|                   | 61     | 59                |
|                   | 75     | 72                |
|                   | 89     | 85                |
|                   | 103    | 98                |

4.2. Quantitative distribution of defect size
The quantitative distribution of artificial defects in critical range should be pre-optimized in the process of prefabrication. It is assumed that the POD curve can be described by two-parameter Weibull distribution [8]:

\[
P_D(a) = 1 - \exp \left[ - \left( \frac{a}{\beta} \right)^\alpha \right]
\]

(2)

Two critical sizes of the defects can be obtained by the following formula:

\[
a_{10} = \beta [\ln(0.9)]^{\frac{1}{\alpha}} a_{90} = \beta [\ln(0.1)]^{\frac{1}{\alpha}}
\]

(3)
In Equation (3), \( a_{10} \) stands for the defect height corresponding to 10% detection rate and \( a_{90} \) stands for the defect height corresponding to 90% detection rate.

4.3. Procedure qualification and reliability evaluation of AUT process

In this study, the total sample size was expanded on the basis of the sample size, and the total amount of the sample was 263. It should be noted that these 263 defects were detected by AUT, RT and macroscopic metallographic section, including the artificial defects and the natural defects produced in the welding process. All defects were verified by the macroscopic metallographic sections.

4.3.1. Overall detection rate and rejection rate.

Of all the 263 defects, only 255 defects were detected in the A scan channel of AUT and the overall detection rate was about 96.96%. Among the defects found in AUT, the minimum height was 0.5 mm and the minimum length was 8 mm, both of which were verified by macroscopic metallographic sections. There were 8 defects including 5 volumetric defects in the AUT band diagram that could not be determined and their heights were less than 0.8 mm. And these defect types were incomplete fusion in filling area (3), linear hole (2) and weld porosity (3) respectively. By improving existing evaluation methods and adopting the data of TOFD channel and B scan channel to evaluate defects, these 263 defects could be detected completely and the overall detection rate would be 100%.

Of all the 263 defects, only 187 defects were detected by RT and the overall detection rate of RT was about 71.1%. Among the defects found by RT, the minimum height was 0.35 mm and the minimum length was 9 mm, both of which were also verified by macroscopic metallographic sections. Those defect types that could not be determined were incomplete fusion in weld root (9), incomplete fusion in filling area (12), incomplete fusion in weld zone (21), incomplete fusion in filling area (3) and lack of penetration in root face (6) respectively.

The overall detection rates of AUT and RT were shown in Figure 1, which indicated that the detection rate of AUT is 28% higher than that of RT.

![Figure 1. The overall detection rates of AUT and RT.](image)

4.3.2. Rejection rate and overlap rate
Rejection and overlap analyses of AUT and RT were illustrated in Figure 2. From Figure 2 and Table 2, there were 41 defects that AUT rejected but RT accepted, and 27 of them RT missed, which indicated that defects in the 45° heat welding zone were not easy to be detected by RT rays. However, these undetected defects may have large influence on the safety of welding joints, resulting in a dangerous area.

Table 2. Analysis of AUT rejection.

|  | Root face | Hot weld | Filling pass | Cover pass |
|---|-----------|----------|--------------|------------|
| AUT rejected, RT detected but accepted | 2 | 5 | 6 | 1 |
| AUT rejected, RT missed | 3 | 20 | 4 |

As shown in Table 3, when specimens were merely judged by the A scan channel results, the 24 defects rejected by RT were accepted by AUT, 2 of which were not even detected by AUT. It indicated that the AUT band figure had certain difficulty in detecting small volumetric defects. Be that as it may, such defects were rather obvious in the volume channel and TOFD channel of AUT system. Thus the problems can be solved by improving the evaluation criterion (acceptance criteria part) of AUT system. Together with Figure 2, it was also found that the positioning errors of RT technology for defect depth were prone to be misjudged.

Table 3. Analysis of RT rejection.

|  | Volumetric defects | Criterion difference | Misjudgements due to ray positioning |
|---|-------------------|----------------------|--------------------------------------|
| RT rejected, AUT detected but accepted | 4 | 11 | 4 | 3 |
| RT rejected, AUT missed | 2 |

To summarize, 182 defects were detected by both AUT and RT, accounting for 69.2% of the total 263 defects. The different results were mainly caused by the inconsistent acceptance criteria of the two detection methods. Considering that AUT and RT may have inconsistent rejecting results to the same defect, it is meaningless to evaluate the reliability of the two methods by the index of rejection rate.

4.3.3. Evaluation of reliability.

Testing results of AUT and RT were shown in Figure 3. It is found that the reliability level of AUT was 90/95: 1.105 mm, which was higher than that of RT (90/95:1.875mm). Together with the accuracy of defect sizing by means of AUT, which was shown is Figure 4, it was concluded that AUT was more stable and reliable. However, the reliability index indicates that both AUT and RT are reliable, and we can make a wise choice in the engineering project according to the operation environment, welding procedures and so on.
As was shown in Table 4, AUT has a lower detection rate in the root and fusion zone; thus testing procedures of AUT should be adjusted accordingly in order to improve the overall reliability and feasibility.

![Figure 3](image1.png)

**Figure 3.** (a) POD diagram of RT (90/95=1.875mm); (b) POD diagram of AUT (90/95=1.105mm).

![Figure 4](image2.png)

**Figure 4.** POD diagram of AUT.

| Statistics of defect heights under 95% confidence level with 90% detection rate |
| --- |
| **Statistics type** | POD with 20% FSH threshold (The whole weld) | POD with 20% FSH threshold (Root) | POD with 20% FSH threshold (Hot weld) | POD with 20% FSH threshold (Filling) | POD with 20% FSH threshold (Covering) |
| The total number of samples | 205 | 1.105 | 2.515 | 2.045 | 1.135 | 1.865 |

**5. Conclusions**

In this paper, the reliability of AUT system was verified by RT and macroscopic metallographic section, and evaluated by indexes of POD, POS, PFA, POR and ROC. The main findings were as follows:

1. The reliability index (POD/CL = 90/95) of AUT system was determined. Based on Hit - Miss binomial model and Weibull distribution model, the sample size for AUT procedure qualification was
determined as 46, which could meet the operational requirements according to the construction experience.

(2) The reliability of AUT system was verified by five indexes, which included Probability of Detection (POD), Probability of Sizing (POS), Probability of False Alarm (PFA), Probability of Rejection (POR) and Receiver Operating Characteristic (ROC). These indexes can solve the technical problem of quantitative evaluation of AUT generalized reliability from aspects of human, machine, material, method and environment and are helpful for the comprehensive evaluation of the application ability of AUT detection technology.

(3) By analyzing testing results of AUT, RT and macroscopic metallographic section, it was found that the reliability level of AUT (90/95:1.105 mm) was higher than that of RT (90/95:1.875 mm). The existing evaluation method of AUT can be improved by integrating the TOFD channel and B scan channel data into a comprehensive evaluation. Then AUT can detect the defects detected by RT, and had a higher detection rate than RT.

(4) For testing of land long distance pipeline with automatic welding joint, AUT had a much higher detection rate and reliability than RT, and could detect all of the harmful defects, minimum size of which was smaller than that of RT. Moreover, the height of defects was limited in the acceptance standard of AUT while it was not in that of RT. Therefore, AUT is reliable enough to replace RT. In addition, with the improvement of AUT technology evaluation technology and the quality management system, AUT will be widely applied to provide technical support for the promotion of automatic welding technology, and play a more and more important role in the construction of long-distance pipeline.

References

[1] Georgiou, G.A. Probability of detection (POD) curves derivation, application and limitations [R]. Health and Safety Executive, 2006. (London, UK)

[2] Visser Consultancy Limited. POD/POS curves for non-destructive examination [R]. Health and Safety Executive. 2000. (London, UK)

[3] FUCSKF. MuLLERC. SCHARMACH M. Measuring of the reliability of NDE [C] //8th International Conference of the Slovenian Society for Non-Destructive Testing, 2005:173-180. (Portorož, Slovenia)

[4] FOR SYTHDS, FAHRA, LEEMANS D V, et al. “Development of POD from in-service NDI data” review of progress in quantitative non-destructive evaluation [M]. V01.19B. New York, 2000:2167-2174.

[5] NATH S.K., BALASUBRAMANIAM Krishnan, KRISHNAMURTHY C.V., NARAYANA B. H.. Reliability analysis of time off light diffraction characterization of inclined cracks [J]. Material Evaluation, 2009,67(3):342-353. (Columbus, USA)

[6] Li Guangbo, Application of Automatic Ultrasonic Inspection in the Construction of West Pipeline [J]. OIL-GASFIELD SURFACE ENGINEERING, 2010, 29(8): 92-93.

[7] Li Yingying, Initial Exploration on the Reliability of Phased Array Ultrasonic Testing and Numerical Simulation of POD [D]. Dalian: Dalian University of Technology, 2014.

[8] FENG Zhen Yu, LI Zhen Xing. The Reliability-Based Method to Estimate POD of Crack. Nondestructive Testing, 2010, 32(4): 249-252.