Quantifying NDE Reliability from ENIQ Qualification Information

Iikka VIRKKUNEN *, Mika KEMPPAINEN *, Kaisa MIETTINEN *
* Trueflaw Ltd., Espoo, Finland

Abstract. The current trend towards risk-informed inspection planning, increasing requirements on plant safety and aging of power plants increase the importance of quantifying nondestructive evaluation (NDE) reliability in the Nuclear industry. At the same time, there's large body of work already done to ensure NDE reliability in the form of inspection qualification. In Europe, this mostly takes the form of European network for inspection qualification (ENIQ) -style qualification. However, attempts to infer quantitative NDE reliability information from existing qualification data have met with limited success. In particular, numerous approaches have been tried to estimate probability of detection (POD)curves based on qualification data. These include the MIL-HDBK-1823A statistical approach, Bayesian approach and others. Unfortunately, this work has, to date, been largely unsuccessful due to lack of data or test pieces for statistical analysis, or in some cases due to improper distribution of available data.

The present paper introduces an alternate approach for estimating NDE reliability from existing qualifications. Instead of focusing on the actual inspection results gathered from the qualification trials (which are few in number), the approach focuses on the qualification requirements, i.e. test piece trials and related pass-fail criteria. A set of relevant performance criteria (i.e. POD curve or sizing error) is then tested against the qualification requirements to determine the pass probability for an inspector having such a performance. The pass probabilities can be calculated (e.g.) using Monte Carlo method. The highest performance that will likely fail qualification can be used as a lower limit estimate of performance for inspectors who will pass the qualification.

1. Introduction

Nondestructive evaluation (NDE) is one of the key tools for ensuring continued safe and reliable operation of nuclear power plants. Thus, the required reliability for NDE is high. The flaws that NDE needs to find are quite challenging. Service induced cracks may be tight, small and otherwise difficult to detect. Yet, the expected reliability of inspections is very high.

After some early round robin exercised showed [1,2] clearly insufficient performance, the nuclear industry has taken steps to assert sufficient NDE performance. In essence, it is now required that sufficient performance is demonstrated (qualified) before using any NDE method or procedure for in service inspection of nuclear power plant. Also, it was noted, that significant variability between inspectors exists and thus also used personnel must be qualified for nuclear inspections.

Two distinct qualification schemes or frameworks are currently widely used for nuclear NDE qualification. In the U.S., the american society of mechanical engineers (ASME) code adopted a statistical screening approach [3]. The code defines a required sample set and pass/fail criteria for inspection qualification. The inspector is, in principle, allowed to use any method necessary to make the inspection, as long as he or she is able to complete the inspection of defined sample and pass the set criteria.

The ASME criteria were developed so that an inspector with unacceptable performance would fail the test with high probability whereas inspector with acceptable performance would pass the test with high probability. Such a test is always a compromise between number of samples or inspections and accuracy of the test. A set of "power curves" was developed to quantify this idea and to define sample set size and pass fail criteria that would provide good compromise between number of samples required and demonstrated reliability. Sometimes quoted example of the criteria is that if an examiner identifies 90% of the flaws in the specimen test set and does not exceed 10% false call rate, the examiners probability of passing the qualification is 90%. [3, 4]

At the same time, the European network for inspection qualification (ENIQ) took somewhat different approach. In ENIQ, it was postulated, that statistical evidence alone would not be sufficient to guarantee the required level of performance and that the number of test samples necessary for such demonstration would be prohibitively large. Furthermore, the European context necessitated greater flexibility for requirements due to varying technical situation and authority requirements in various countries. Consequently, the ENIQ developed a more flexible framework for inspection qualification. In this framework, a separate input information document is created for each qualification case. The input information defines the inspection requirements, expected flaw types and the structural integrity context of the inspection. Next, a technical justification (TJ) is written for the NDE procedure to be qualified. This justification includes the physical reasoning and possibly references to other evidence that demonstrates high-expected reliability for the inspection. The TJ is also used to select test blocks and defects to test the reliability. This allows use of "worst case" defects to reduce needed sample set. Finally, the qualification is completed with open practical trials for procedure qualification and blind practical trials for personnel qualification. The TJ and practical trials together demonstrate that the NDE system has required performance.

Neither the ASME qualification nor ENIQ qualification provide quantitative data on the attained performance levels. The results are given as pass/fail or acceptable/unacceptable.

In recent years, there's been increasing need for quantitative performance data for NDE. In part, this is due to advances in risk-informed in service inspection (RI-ISI). The rationale behind RI-ISI is that inspections should be focused to components and locations where they are most beneficial. Likewise, the inspection intervals should be chosen with maximum expected benefit from the inspection. Unsurprisingly, the expected NDE reliability has significant influence in these calculations and the optimal inspection strategy varies depending on inspection reliability. In particular, it's often necessary to demonstrate very high inspection reliability to get significant advantage for performing NDE in the RI-ISI calculations.

Quantifying inspection reliability is also becoming increasingly important for the qualification itself. It's now over 30 years since the first inspection qualification programs started. Consequently, there's also an ever-increasing body of completed qualifications. These include qualification done tens of years between them and with wildly different NDE equipment. It's difficult, in particular fort the ENIQ type qualifications, to enforce and demonstrate consistent requirements across qualifications. Furthermore, there's increasing interest to take advantage of already completed work on previous qualifications or qualifications done elsewhere. This is also complicated by the lack of quantitative measures that would make qualifications comparable across time and between countries.

Finally, in the ENIQ methodology, the input information is used to define scope and performance target for the inspection procedure. Structural integrity significance and previous degradation data, when available, are used to determine the inspection scope. This allows, at least in principle, for the NDE to focus on significant damage modes and work with reasonable detection targets. However, it's currently difficult to assess how well the qualification really addresses these requirements and what is the confidence level that qualified inspectors meet the targets set in the input information.

There's been significant effort to address this increasing need for quantitative performance demonstration data. Given the great body of available qualification data, the focus has mostly been to extract quantitative data from existing qualification data. However, this has proven quite challenging. Currently, the work has culminated in defining POD-curve, as required by the RI-ISI calculations, from completed or planned qualification exercise. There's a series of ENIQ reports [5-7] detailing various approaches that show the extent of this challenge.

At first, it was recognized, that since in an ENIQ qualification the performance is demonstrated with combination of technical justification and practical trials, any quantitative measure should take credit for both parts. It was a challenge to quantify the value of the TJ and to combine this with practical trials. Gandossi et al. [5,6] noted that as it is, the adequacy of the TJ is judged by an expert judgement in the qualification body. Consequently, the problem was, in essence, quantifying this expert judgement. However, it proved difficult for the experts to provide quantitative assessment of the TJs'. The quantitative evaluation was finally improved by introducing the concept of "equivalent test blocks". The experts were asked to compare the TJ to a practical trial and asses the number of test blocks required to provide same confidence of NDE performance that the TJ alone provides. The quantified expert judgement and information from practical trials were then combined to provide final estimate for the quantified NDE reliability using Bayesian inference.

However, even the improved expert judgements contained significant variability and thus the reliability of the judgement remained somewhat questionable. Furthermore, when the expert judgement and practical trial information was combined, often the TJ did not show significant contribution to the overall demonstrated performance (and sometimes even provided negative contribution).

Over the same period of time, the aerospace industry developed their own methodology for demonstrating NDE performance. This methodology relied more heavily on statistical evidence and developed advanced methodology to extract POD information from limited set of cracked samples [8-9]. The most recent version is documented in the MIL-HDBK-1823A and recently standardised as the ASTM E2862. Sadly, attempts to apply this approach to extract POD curves from nuclear industry qualification data have met with limited success. The most obvious problem has been lack of samples: the MIL-HDBK-1823A hit/miss analysis requires 60 samples. For nuclear qualification, it's uncommon to have more than 20 flaws in qualification. There are also more subtle problems. The ASTM E2862 directly states, that POD cannot be modelled (as continuous function of discontinuity size) if all discontinuities are found (or if none are found). In nuclear qualification it is expected that all the cracks will be found. Furthermore, the use of various worst-case defect locations makes the assumption of monotonously increasing POD for the test set questionable. Consequently, it would require significant changes to current qualification practices for the ASTM E2862 to be applicable; most notably increased number of cracks and cracks with low to medium probability of detection.

There's also some other approaches developed in the aerospace industry. These have not, to our knowledge, been applied in the nuclear industry nor do they seem to present a viable solution for the problem at hand. Some are summarized here because they provide alternate solution to similar problems or expose criticism within the aerospace industry.

A set of independent hit/miss inspections can be modelled with binomial distribution (with certain limitations). It follows that finding 29 cracks out of 29 cracked samples is consistent with 90% lower limit POD estimate at 95% confidence level (i.e., the true POD is >= 90% with 95% confidence, or conversely there is a 5% chance that the true POD is < 90%). Thus, the 29/29 requirement has been used in some cases in the aerospace industry [13]. This is, perhaps, the closest equivalent the aerospace industry uses to the nuclear industry qualification (ASME-type qualification, in particular). The requirement does not provide POD curve, but it provides statistical assurance (at 95% confidence level) of sufficient performance (90% POD).

The DOEPOD model [10-12] is also based on the binomial view of hit/miss data. The main motivation for the DOEPOD model is, that using model-based POD estimation (e.g. ASTM E2862) assumes POD as a function of flaw size follows certain model. In particular, the POD is continuous, monotonically increasing function of flaw size a. This assumption may not always be justifiable, e.g. when the method sensitivity varies for different flaw sizes due to different probes, beam focusing or for some other reason. The DOEPOD model does not assume functional relationship between POD and flaw size. Instead, the inspection results are grouped and analysed, simply stated, as groups to make sure that the 29/29 condition is fulfilled for certain flaw size and flaw sizes above it.

In summary, it can be said that the previous attempts to quantify NDE performance based on ENIQ qualifications have not proven successful. Furthermore, the approaches tried so far have clear problems or incompatibilities and thus success with these seems improbable. At the same time, it's generally agreed that qualification has significantly improved NDE reliability. Thus, it should also be possible to quantify the improvement. There's no doubt that the vast amount of qualification data contains valuable information about the NDE performance.

The topic of this paper is to present an alternate approach to quantifying NDE reliability based on ENIQ qualification data.

2. Qualification as a screening test

The root cause for many of the problems in quantifying NDE performance from ENIQ qualification is, that the qualification was designed to be a screening test and not a POD experiment or performance evaluation. Consequently, it's closer to the 29/29 requirement than it is to ASTM E2862 exercise.

In order to get quantitative data from the current nuclear inspection, the problem set must be changed to reflect the character of the qualification. The chosen problem set is then: given that the inspector has passed qualification, what's the lower limit performance we can expect from this inspector? To make the problem easier to solve, we first solve the inverse problem: given, that the inspector has certain performance, what's the probability that he will pass the qualification. With this information, the inverse problem can be solved iteratively, given certain limitations for inspector performance. The approach can be seen as an extension to the ASME power curve approach or to the 29/29 demonstration criteria.

For further simplification, we first solve the problem concerning the blind test results only. The role and significance of the TJ and open trials is discussed in section 5.

Solving probability of passing a qualification, given the inspector performance is rather straightforward task. Sometimes, the inspection task and the pass/fail criteria are simple enough to be solved analytically. Often, the pass/fail criteria contain multiple overlapping criteria, and forming analytical solutions is cumbersome. Consequently, numerical solution via Monte Carlo simulation was chosen for this work. The Monte Carlo simulation is easy to formulate, and solves in couple of seconds. The numerical Monte Carlo solution easily accommodates various extended pass/fail criteria and false calls. (See paragraph 4. for details of the simulation.)

This intermediary result has certain practical significance. In the ENIQ process, input information is used to review possible degradation mechanisms and their possible structural significance. With this information, the input information sets the goal or target for qualification, i.e. the performance required for the qualified personnel. Probability of passing the qualification with this given performance is a measure of the confidence level provided by the qualification (assuming different performance targets are comparable; see paragraph 3. for further discussion). Any inspector, who has lower performance than specified, has lower probability of passing the qualification. Thus, the probability of the target performance passing can be seen as the risk that we are willing to accept for any inspector up to this performance level qualifying.

In many of the previous approaches, false calls do not significantly alter the outcome of the analysis. They can often be handled by simply removing false call data from the analysis. In contrast, the false call rate of the inspector and related pass-fail criteria significantly affect the probability of passing the qualification. The qualification body cannot, in principle, separate the inspectors that passed due to high false call rate from those that passed due to the required skill. The amount of blank samples (opportunities to make false calls) and the number of allowed false calls affect the probability of passing qualification as well as the "optimal" level of false calls that maximize candidates chance to pass.

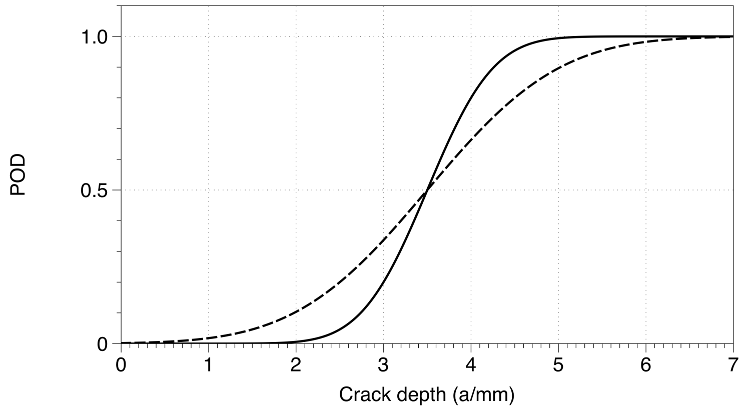
The discussion here focuses on probability of detection. However, the approach is easily extendible to sizing or other performance criteria.

3. POD vs. a dependence

Different approaches for solving probability of detection as a function of crack size (POD (a)) require different assumptions concerning the POD(a) behaviour. The present approach is no different. The MIL-HDBK-1823A approach requires that POD be continuous and increasing function of a. In addition, it is sometimes assumed that the limiting POD for large cracks sizes is 100%. These assumptions are integral in the success of the approach and significantly improve the amount of information that can be extracted from limited sample data. At the same time, both of these assumptions have been questioned and may be difficult to justify in some cases. The DOEPOD approach specifically addresses the assumption that POD is an increasing function of a and provides assurance of POD without this assumption.

For present purpose, different assumptions regarding POD(a) dependence can be made. If information about this dependence is available, it can be used to provide better estimate for the expected POD. In cases where such information is not available, simplified models can be used, at the expense of justified performance.

The present approach enforces one additional requirement for whatever POD(a) dependence is assumed. Since the required/postulated performance is used as the lower limit performance estimate, any chosen POD(a) dependence must provide a uniquely comparable set of performance criteria. For example, Picture 1 shows a set of two POD curve estimations. Both of these curves show an area where POD surpasses the other curve and, consequently, neither is ambiguously better. In other words, is the inspector allowed to compensate lower-than-required performance for certain flaw sizes with better-than-require performance for other flaw sizes?



Pic. 1. Two possible POD curves neither of which is unambiguously better than the other.

For this reason, four models of POD(a) dependence are included, each of which is useful under certain conditions. Firstly, the usual normal-cfd functional dependence is used. This is a commonly used approximation for POD(a) dependence. As shown above, the curves are not, in general, unambiguously comparable and thus additional discretion is needed from the operator to choose a comparable subset of curves. Secondly, the normal-cfd function with additional maximum POD is provided. This provides additional flexibility, when the assumption of POD limiting to 1 cannot be justified.

Both of these models require significant assumption regarding the POD(a) dependence, which may not be available. The third model is a stepwise curve with adjustable POD at top level. It is conventional to use detection target and minimum POD to determine inspection requirements in the input information. This functional form corresponds to such target definition. Stepwise POD curve is also sometimes used in the RI-ISI analysis [14].

The stepwise POD curve assumes equal POD over significant flaw size range. If POD is an increasing function of flaw size, then the candidate may compensate less-than-required performance at the detection target with higher-than-required performance on the bigger flaws. To calculate POD at detection target, the fourth model "counts" only cracks at the smallest size. Other flaw sizes are not used for calculating the probability of passing qualification. If, on the other hand, some of the bigger flaw sizes are missed, then the test is considered a failure. This is similar to including only cracks with size matching the detection target in the flaw set.

4. Software implementation

To provide a practical implementation for this method, a web-based software code was written and made publicly available at http://www.trueflaw.com/qualificationhelper. The implementation has a user interface usable with modern web browser and a computation engine running on the server that computes the computationally intensive Monte Carlo results. The back-end is written in Objective-C and is heavily optimized to provide responsive user interface.

The front-end provides fields to input given inspector performance (as outlined above). Two sets of performance criteria can be input: one for lower limit (unacceptable) performance and one for upper limit (satisfactory) performance. The lower limit performance can be used to iteratively solve the performance that can be reliably expected from an inspector that has passed the qualification. The upper limit performance can be used to iteratively solve the performance, which can be expected to pass the qualification (i.e. the requirements seen by the inspector).

Fields are provided to input qualification test setup information: number of cracks and corresponding crack sizes, number of blanks (for false call analysis) and pass/fail criteria: misses allowed and false calls allowed.

The front end then gathers this data and sends it to the server for analysis. The server calculates the corresponding pass probabilities and reports them back to the front end, which shows them to the user. The tool also has similar analysis implemented for crack sizing performance. This is not considered in present paper.

The back-end receives the candidate performance data (POD as function of crack size and false call probability), trial information and pass/fail criteria. It then calculates the corresponding pass probability using ordinary Monte Carlo analysis. The algorithm used for the Monte Carlo simulation is outlined below in pseudo code:

```
repeat for N Monte Carlo trials:
    for each test block (crack or blank):
        draw a random number R between 0.0 ... 1.0
        if R < given false call rate and test block contains a crack, report detection
        if R < given false call rate and test block is a blank, report false call
        if R > given false call rate and test block is a blank, report correct clean block
        if R > given false call rate and test block contains a crack of size a,
            draw another random number S between 0.0 ... 1.0
            if S < POD(a), report detection
            else, report miss
    compare reported misses and false calls with given criteria and
    determine the trial to be pass or fail
report number of passes / number of trials as the pass probability
```

5. The role of the technical justification

As noted in section 1, in the ENIQ methodology the evidence for sufficient performance is formed by the technical justification and the practical trials. So far, the discussion has only considered practical trials. Consequently, it may be argued that the present discussion gives overly pessimistic view of the true performance of the inspection system. Thus, the TJ should be considered as well.

The role of the TJ in ENIQ qualification is somewhat controversial. It's generally agreed that the TJ is important and valuable part of the ENIQ qualification. However, there are marked differences in the way the TJ is valued both between qualification bodies and between qualification cases. Thus, the contribution of TJ is presented in four different ways:

a) TJ as necessary but insufficient condition

One approach is to say, that the role of the TJ is to set the limits of applicability for the inspection in question. It is thus necessary and important part of the qualification. However, in this approach, it's not considered sufficient to show performance. Thus, the reliability of the inspection should be valued on the practical trials alone, and the TJ seen as a precondition.

b) TJ as necessary and sufficient condition

Second view states, that the TJ in itself provides assurance of NDE reliability. In this case, the contribution should be quantifiable. To use TJ's contribution in connection with present framework, its contribution should be quantified in terms of equivalent hits. That is, the TJ adds test blocks, which are automatically found, and this improves the statistical confidence limits. Such quantification is similar to what was done by Gandossi et al. [5,6], and it has proven challenging to do in practice.

c) TJ as justification for lower statistical confidence level

The basic premise of the ENIQ qualification is, that practical trials alone will not provide sufficient statistical confidence and thus the TJ is needed to provide sufficient overall confidence to the NDE performance. This notion offers another way to quantify the effect of the TJ: the TJ can be used to justify that lower statistical confidence can be considered sufficient. This would be another way to quantify the expert judgement applied in current ENIQ qualifications. It can also be calculated after the fact, by analysing a qualification case in connection with the stated detection target and POD requirement and calculating the confidence level at which this is attained.

d) TJ as justification for POD(a) dependence

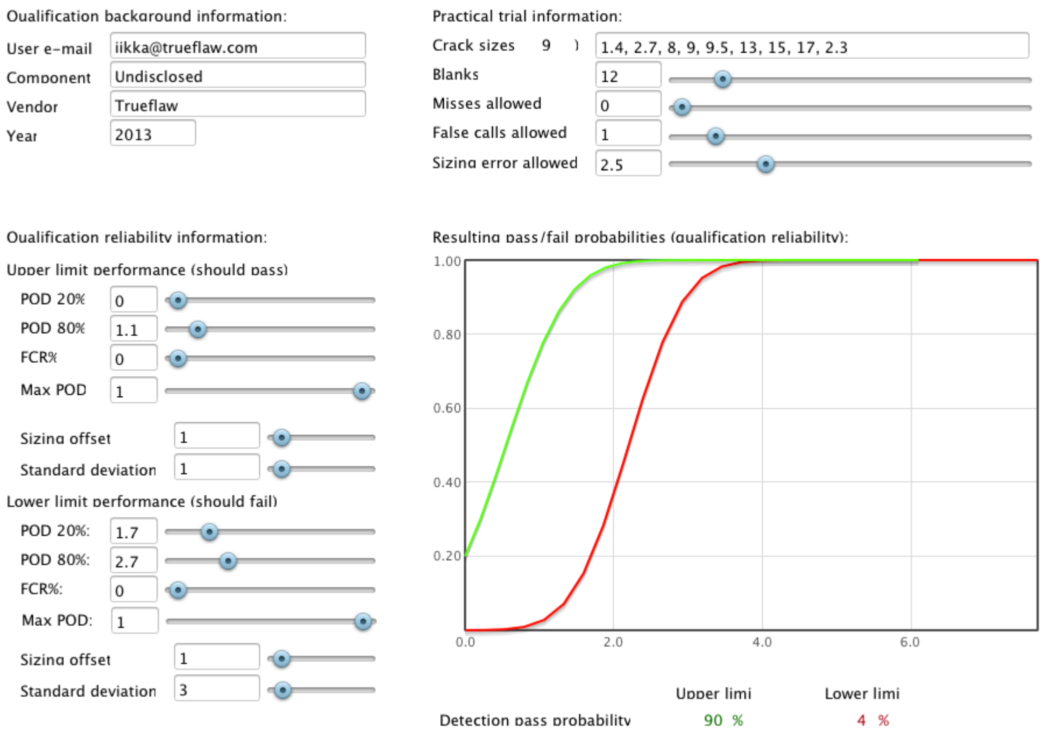
Finally, the TJ can be used to justify certain form of POD(a) dependence. Better knowledge about this dependence can increase the amount of information that can be extracted from inspection data.

Depending on the qualification case, each one of these may be justified. The list is not exhaustive and some approaches can be used in combination (especially c and d). In any case, the decision is a form of expert judgement applied to the qualification case.

6. Example analysis of qualification data

To show current approach in practical context, an example analysis was completed with real data from a Finnish qualification. Due to confidentiality of the qualification data, the analysis is completed on open sample data. The blind sample data can be assumed to be similar enough for present purposes. Also, the exact details of the case in question are left undisclosed.

In first analysis, the shape of the POD curve is assumed known and POD is assumed to reach 100%. Picture 2 shows screen capture with POD curves giving 90% pass probability (performance which vendors need to have to pass qualification with high probability), and 4% pass probability. The inspectors are assumed to have 0% false call rate, to simplify the analysis. The curves were found by manually iterating to find acceptable risk that a passed inspector does not, in fact, have better than desired capability.

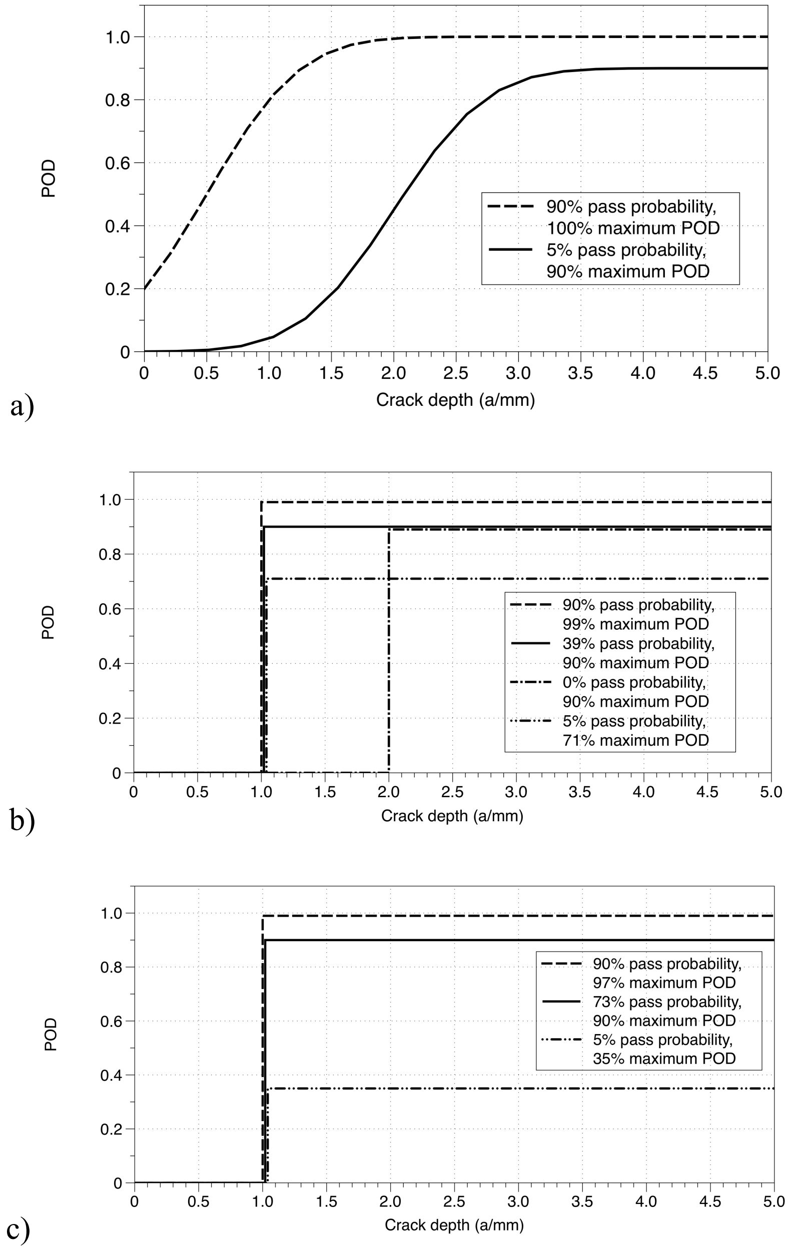


Pic. 2. Example analysis with assumed known POD curve shape reaching 100%. POD curves giving 90% pass probability (green curve) and 4% pass probability (red curve). The shape and slope of the curve is assumed to be known.

Second analysis assumes the shape of the POD curve to be known, but the lower-limit POD is expected to level off at 90% (Picture 3a; for the following analysis only the POD curves and resulting pass probabilities are shown; qualification information is as shown in Picture 2). Now the decrease in maximum POD must be compensated with higher POD for the lower flaw sizes, which is seen as a slight shift of curves to the left.

Thirdly, it is assumed, that the Input information states detection target of 1 mm and assumes step-wise POD-curve with maximum POD levelling off at 90% (Picture 3b). Now it can be seen, that the qualification has 39% probability of passing inspectors at the 90% limit. At the same time, 99% POD is needed for the vendor to be confident of his possibility to pass the qualification. The risk of passing at the lower limit curve can be decreased by lowering the required POD. For 71% maximum POD, the risk of passing is 5%. Alternately, the detection threshold can be increased above the smallest crack size in the test, which decreases the the pass probability to zero (the stepwise function assumes 0% POD below the threshold, so crack sizes are automatically missed. Since the pass/fail criteria do not allow any misses, this fails the qualification.)

Finally, it is assumed, that the POD may, or may not increase after 1 mm detection target (Picture 3c.). Thus the confidence of POD at this level, is only affected by cracks near this size. Thus the amount of flaws is reduced significantly, and there's 73% risk of passing POD of 90%. Conversely, there's 5% risk of passing POD of 35%. (It should be noted, that this is quite conservative assumption.)



Pic. 3. Example analysis with different assumed known POD curves. See text for details. (Different stepwise POD-curves are drawn with slight offset to aid readability.)

7. Conclusions

The approach presented here provides an alternate way to quantify the performance guaranteed by existing ENIQ-qualification (with given the risk that a passed inspector does not, in fact, have better than desired capability). It can be applied on existing information and it allows for different ways to take advantage of the technical justification included in the ENIQ qualification.

The example analysis here shows, that with limited sample set the level of risk for passing lower than desired capability associated with the practical trials alone is somewhat high (even assuming zero false call rate), and the technical justification is important to justify that this level of practical demonstration is sufficient to show that the required performance is met in practice.

Acknowledgements

The authors wish to thank Jennifer R. Brown for her most valuable comments on the manuscript.

References

[1] Lemaintre, P., Koblé, T. D. & Doctor, S. R. 1996. "Summary of the PISC round robin test results on wrought and cast austenitic steel weldments, part III: cast-to-cast capability study". International Journal of Pressure Vessels and Piping (69) pp. 33-44.

[2] Lemaitre, P., Iacono, I. & Vergucht, P. (eds.) 1999. Results of the Destructive Examination of the ENIQ Pilot Study: Defect Catalogue. ENIQ Report 19, EUR 19024 EN, JRC Petten, The Netherlands

[3] Cowfer, C.D. 1991. Basis / Background for ASME Code Section XI proposed Appendix VIII: Ultrasonic examination performance demonstration. Nuclear Engineering and Design, 131, pp. 313 - 317.

[4] Becker, F. L. 2001. Examination Effectiveness Based on Performance Demonstration Results for Flaws at the Clad-to-Base Metal Interface, Proc. 3rd Int. Conf. NDE in Relation to Structural Integrity for Nuclear and Pressurized Components, Seville, Spain, November 14–16.

[5] Gandossi, L., Simola, K. A. 2007. Bayesian Framework For Quantitative Modelling of the ENIQ Methodology for Qualification of Non-Destructive Testing. European Communities, DG JRC, Institute for Energy, EUR22675EN, ISSN 1018-5593.

[6] Gandossi, L., Simola, K., Shepherd, B. 2010. Application of a Bayesian Model for the Quantification of the European Methodology for Qualification of Non-Destructive Testing. International Journal of Pressure Vessels and Piping, 87, pp. 111-116.

[7] Gandossi, L., Annis, C. 2010. Probability of Detection Curves: Statistical Best-Practices. European Communities, DG JRC, Institute for Energy, ENIQ Report nr. 41, EUR 24429 EN, ISBN 978-92-79-16105-6, ISSN 1018-5593.

[8] Anon. 2009. Nondestructive Evaluation System Reliability Assessment. Department of Defence Handbook. MIL-HDBK-1823A. 171 p.

[9] Anon. 2012. Standard Practice for Probability of Detection analysis for Hit/Miss Data. American Society for Testing and Materials, ASTM E2862-12.

[10] Generazio, E.R. 2009. Design of Experiemnts for Validating Probability of Detection Capability of NDT Systems and for Qualification of Inspectors. Materials Evaluation, June, pp. 730 – 738.

[11] Generazio, E.R. 2011. Validating Design of Experiments for Determining Probability of Detection Capability for Fracture Critical Applications. Materials Evaluation, december, pp. 1399 – 1407.

[12] Generazio, E.R. 2011. Binomial Test Method for Determining Probability of Detection Capability for Fracture Critical Applications. NASA/TP–2011–217176. National Aeronautics and Space Adminstration, Langley Research Center. 33 p.

[13] Rummel, W.D., Matzkanin, G.A. 1997. Nondestructive Evaluation (NDE) Capabilities Data Book. 3rd. ed. Nondestructive Testing Information Analysis Center. NTIAC-DB-97-92. NTIAC-1997-A286978. 625 p.

[14] Shepherd, B., Gandossi, L., Simola, K. 2009. Link Between Risk-Informed In-Service Inspection And Inspection Qualification. ENIQ Report No 36, European Communities, DG JRC, Institute for Energy, EUR 23928 EN, ISSN 1018-5593.