Probability of Detection in Nondestructive Testing

InnoTesting 2016
February 25-26th, Wildau

Daniel KANZLER (KNC-Reliability)
Christina MÜLLER (Bundesanstalt für Materialprüfung und -forschung)
<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Motivation:</strong></td>
</tr>
<tr>
<td><strong>State of the art:</strong></td>
</tr>
<tr>
<td><strong>Evaluation methods:</strong></td>
</tr>
<tr>
<td><strong>Advanced approaches:</strong></td>
</tr>
<tr>
<td><strong>Use of the POD:</strong></td>
</tr>
</tbody>
</table>
Manufacturing process of spacecrafts

“Today`s technology and manufacturing processes don`t permit construction of a defect-free spacecraft. Thus defects are present and test discrepancies need to be remedied.”

- Lot = 1
- Handwork
- Prices between 100M$ - 1B$

“For e.g. the spacecraft industry, where the price of failure is too high, the technology is so complex testing is currently required to ensure a product that performs reliably to expectations.”

Risk-Averse: Failure was not an option in spacecraft – is not an option in a lot other industries: This leads to an culture, where tests and NDT plays an important role.
We cannot find every defect

Undetected cracks lead to the changing of thinking:
damage tolerance concepts and probabilistic evaluations of ndt

Undetected Forging Crack in F-111

This 1969 failure of a new F-111 aircraft was caused by an undetected forging defect that quickly grew to failure by fatigue in the high-strength steel wing pivot fitting structure.

Figure source: Thompson, D. O.: “EVOLUTION OF QNDE'S CORE INTERDISCIPLINARY SCIENCE AND ENGINEERING BASE” Review of Quantitative Nondestructive Evaluation, 2010, 3-25
Add here the title of the accident, e.g. Irsching power plant accident, 31. December 1987

“[An example of a failure of NDT to detect a structure breaking defect] ... is the accident in the power plant Irsching, Bavaria, in 1987. Back then, the middle course of a turbine exploded into 30 pieces. One piece with a weight of more than one ton skidded up in the air and landed in a distance of about 1 kilometer to the power plant. Nobody was injured, but the property damage was enormous. After the accident, the [...] experts perfected the testing technology for turbines.”

“... the spectacular crash had more than one causes: The fracture toughness of the material was on the lower limits of the tolerance; at the same time a production defect was missed, due to a wrong interpretation of an UT-indication; the low temperature at the time of the accident did the rest”

Sources:
http://www.heise.de/tr/artikel/Gegrillte-Maeuse-fliegende-Turbinen-280433.html
Probability of Failure

When does it happen?

Probability of Occurrence
- There is a defect

Probability of Detection
- One defect is critical
  - The defect was not found while testing

Probability of critical Defect after ndt
- One component has a critical defect after testing
  - Load of the defective position / component

What to do?
- Testing with ndt
- Evaluation of ndt Systems
- Damage Tolerant Design
- Damage Limitation
Further definitions of reliability and probability

The detection ratio is not directed into the future but in the past.

The most commonly used procedure to evaluate the reliability:
Mostly the Probability of Detection (POD) is made out of data from artificial defects, which represent real defects in the tested object (Singh, R. Three decades of NDI Reliability Assessment 2000).

Probability:
The probability can only be estimated by the relative frequency and requires a very big amount of experimental data (Sachs: Angewandte Statistik 1999 - Applied Statistics).

Detection Ratio:
If the number of tested defects is not big enough, the result is called ratio of detection instead of probability of detection (Shepherd, B. Research SKI Report 2007).
Standards: A common way to define the capability of NDT systems

Example: RT

Standards are important tools for defining the basic capability of NDT systems:

- EN ISO 19232-1 (etc.)
- EN ISO 19232-2
- (EN 462-2)
- ASTM E1025
The answers the “wires” cannot give us:

- Why do we miss defects while testing?
- What are the reasons for misinterpretation?
- How can we handle these facts?
- ...

Definition of NDE reliability:
NDE reliability is the degree that a NDE system is capable of achieving its purpose regarding detection, characterization and false calls.

(Second European American Workshop on NDE Reliability, 1999).

The answer of these question can be found in evaluating the NDT situation. For the correct evaluation probabilistic evaluations are used.
**Consequences of decisions based NDT**

<table>
<thead>
<tr>
<th>Truth</th>
<th>Decision</th>
<th>Costs</th>
<th>Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Defect</td>
<td></td>
<td>Min. costs: NDT costs</td>
<td>0</td>
</tr>
<tr>
<td>Defect</td>
<td></td>
<td>NDT costs + scrapping / repairing costs</td>
<td>$ ( + threat banished)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Defect</td>
<td></td>
<td>NDT costs + scrapping (unnecessary)</td>
<td>$$$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+ loss of trust in the NDT department)</td>
<td></td>
</tr>
<tr>
<td>Defect</td>
<td></td>
<td>NDT costs + failure of a defective component</td>
<td>$$$$$$$ ( + high level of risk)</td>
</tr>
</tbody>
</table>
Modular Reliability Model (Müller, 2009)

Organisational Context

Application Parameters
- e.g. Environment

Human Factors
- e.g. Experience

Intrinsic Capability

NDT System

Physics

Consequences
An investigation in the area of spacecrafts showed, that for a general testing situation:

- “Nearly half of the discrepancies (a functional or structural anomaly, which may reveal itself as a deviation from requirements or specifications) were caused by the testing equipment and employee / operator error”
- “A research finds that 29% of discrepancies reported are related to test equipment.”

Discrepancies

- Equipment
- Human related
- Other reasons

A. Weigel and J. Warmkessel 2001: Understanding the Enterprise Value of Test: Characterizing System Test Discrepancies in the Spacecraft Industry
Relationship between the signal and the defect parameters in RT

The contrast of “small” defects is lower than for “bigger” defects
POD – measurements with fluctuations

Theory vs. Reality!

We don’t measure the same contrast every time.

RT:
\( a: \) penetrated length of the material defect
\( \widehat{a}: \) contrast of the indication

Statistics provides models to describe the fluctuations!
The decision threshold is a part of the testing procedure: It often depends on the detectable contrast, the noise, etc..

The probability of detection correlates with the area **under the distribution above the decision threshold** for each parameter size $a$. 
POD from the linear $\hat{a}$ vs. $a$ graph to the POD curve

**Linear $\hat{a}$ vs. $a$ graph**

- $\hat{a}$ vs. $a$
- POD = 50%
- POD ≈ 100%

**POD curve**
The confidence band gets smaller as the amount of data increases.
Additionally, as the amount of data increases, the parameters are closer to the theory.
In the original POD it was suggested to use at least 40 data points for \( \alpha \) vs \( a \).
What are we doing with the result of the POD?

POD

90%

not - acceptable critical defect size

acceptable critical defect size

D. Kanzler: Innotesting 2016
Summary: POD in radiographic testing

- **specimen under test** → intrinsic factor → NDT method → environmental factors

- POD diagram
  - Linear
  - POD = 50%
  - POD = 100%
  - POD cumulative distribution

- POD values:
  - \(a_{90/95}\)

- System acceptability:
  - \(a_{90/95} > a_{crit}\) → Not acceptable
  - \(a_{90/95} < a_{crit}\) → acceptable

- \(a_{dec}\) → detection level
Final deposit of spent nuclear fuel (KBS-3)

Concept for a final deposit of POSIVA

1. Final disposal canister
2. Bentonite buffer
3. Tunnel backfill
4. Bedrock

Figure source: POSIVA Oy
Radiography and the connection to the POD

Figure source: SKB Oskarshamn Sweden

9 MeV linac (Varian 3000)
angle 10°
Digital linear detector

| Reliability Characteristics | remaining wall thickness $RW = W - a$
| NDE Systems                | $W =$ wall thickness
|                            | $a =$ penetrated length of the defect
|                            | $\Rightarrow \hat{a}$ and $\hat{a}_{\text{dec}}$

Reliability Characteristics
- remaining wall thickness $RW = W - a$
- $W =$ wall thickness
- $a =$ penetrated length of the defect
- $\Rightarrow \hat{a}$ and $\hat{a}_{\text{dec}}$

NDE Systems
- contrast
- noise
Examples: POD evaluation of an RT-system

\( \hat{a} \) vs. a graph: linear regression

Artificial defects

POD graph

Artificial and real defects

POD in use
Thank you for your attention!
Any questions?

Daniel Kanzler  M. Eng.
Email: KanzlerD@knc-reliability.com