Probability of Detection in Nondestructive Testing

InnoTesting 2016
February 25-26th, Wildau

Daniel KANZLER (KNC-Reliability)
Christina MÜLLER (Bundesanstalt für Materialprüfung und -forschung)

Contents

Motivation:	What can we learn from the history of aerospace?
State of the art:	Detectability as the basis of probability of failure
Evaluation methods:	Consequences of decisions based on ndt
Advanced approaches:	Probability of Detection: Evaluation of ndt systems are important!
Use of the POD:	Out of the box: Finnish and Swedish project for final disposal of spent nuclear fuel

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."

Figure source: http://www.cbc.ca/news

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

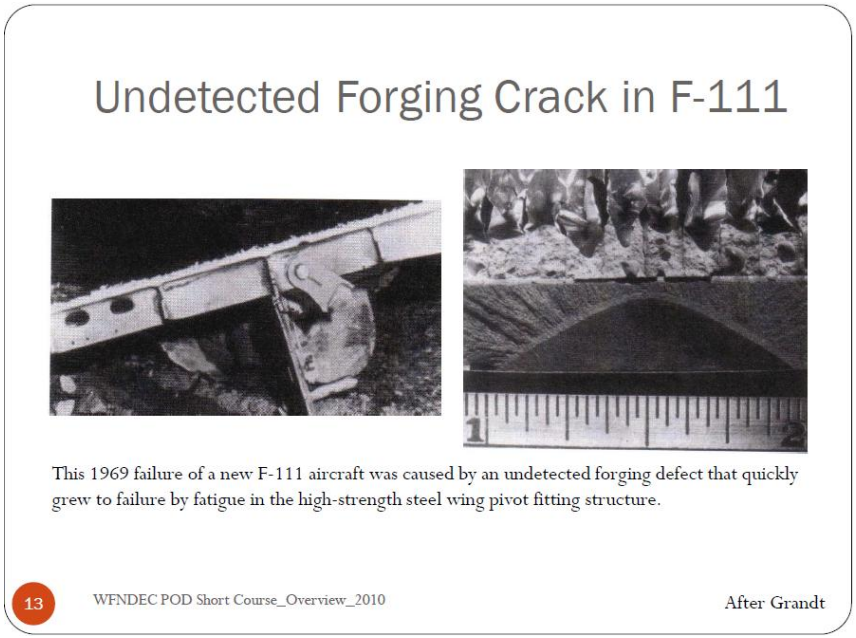
"For e.g. the spacecraft industry, where the price of failure is too high, the technology is so complex **testing is** currently **required** to ensure a product that performs reliably to expectations."

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



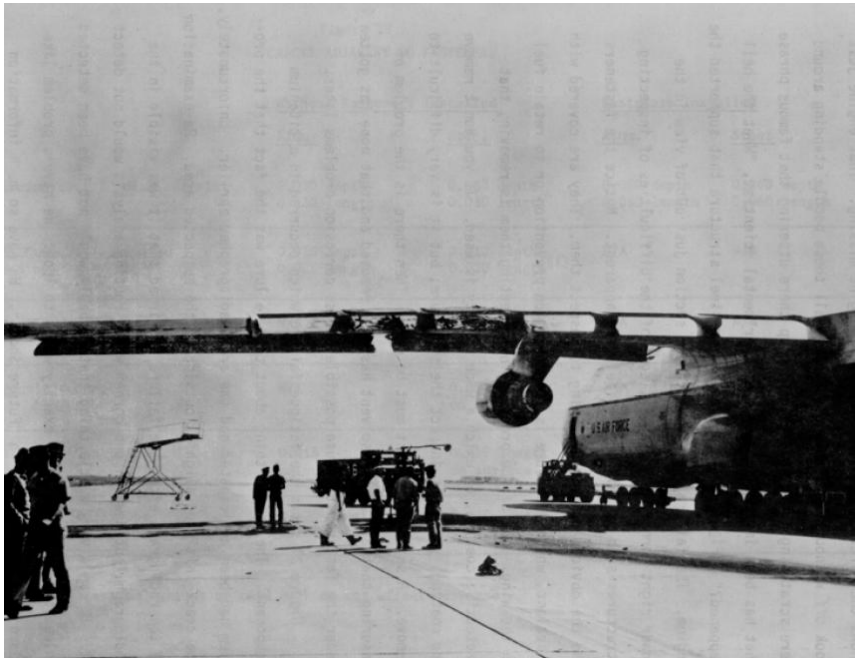


Figure source: Thompson, D. O.: "EVOLUTION OF QNDE'S CORE INTERDISCIPLINARY SCIENCE AND ENGINEERING BASE" *Review of Quantitative Nondestructive Evaluation,* **2010**, 3-25

Not an aerospace problem

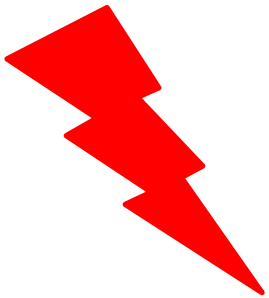
Add here the title of the accident, e.g. Irsching power plant accident, 31.December 1987

"[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."

Sources:
http://www.heise.de/tr/artikel/Gegrillte-Maeuse-fliegende-Turbinen-280433.html
https://www.allianz.com/de/presse/news/geschaeftsfelder/versicherung/news_2012-11-21.html/

„… 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"

Probability of Failure

	Probability of Occurrence	Probability of Detection	Probability of critical Defect after ndt	Probability of Failure
When does it happen?	There is a defect	One defect is <u>critical</u> ⊕ The defect was not found while testing	One component has a <u>critical defect</u> 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:***

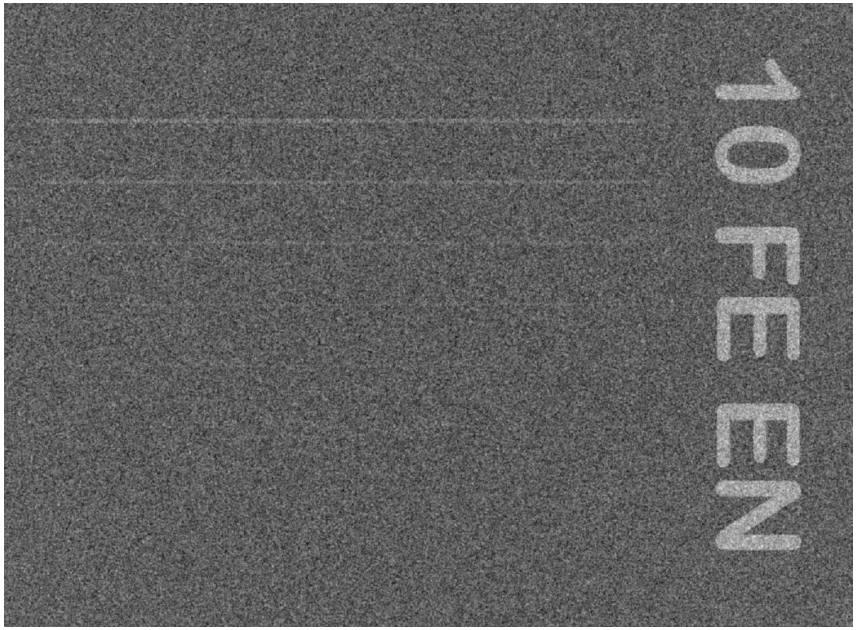
  The probability can only be estimated by the relative frequency and requires a very big amount of experimental data (Sachs: Angewandte Statistik 1999 - Applied Statistics).

- ***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).

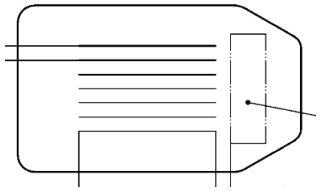
Standards: A common way to define the capability of NDT systems

Example: RT

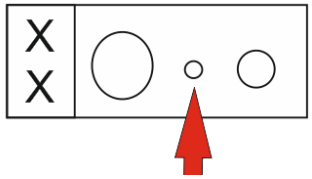


Standards are important tools for defining the basic capability of NDT systems:

- 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?
- …

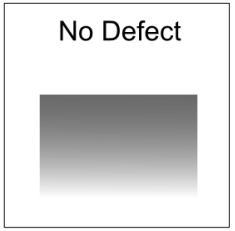
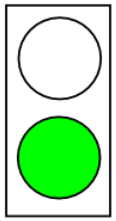
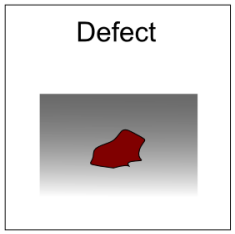
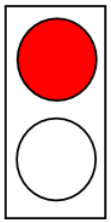
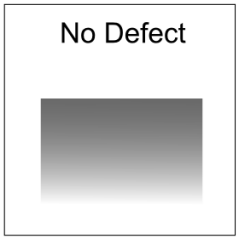
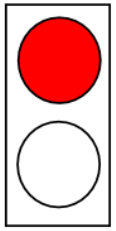
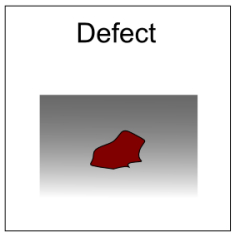
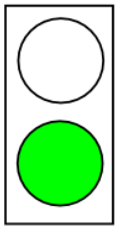
Definition of NDE reliability:

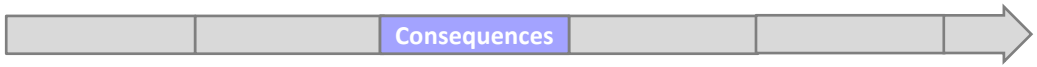
NDE reliability is the degree that a NDE system is capable of achieving its purpose regarding detection, characterization and false calls.

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

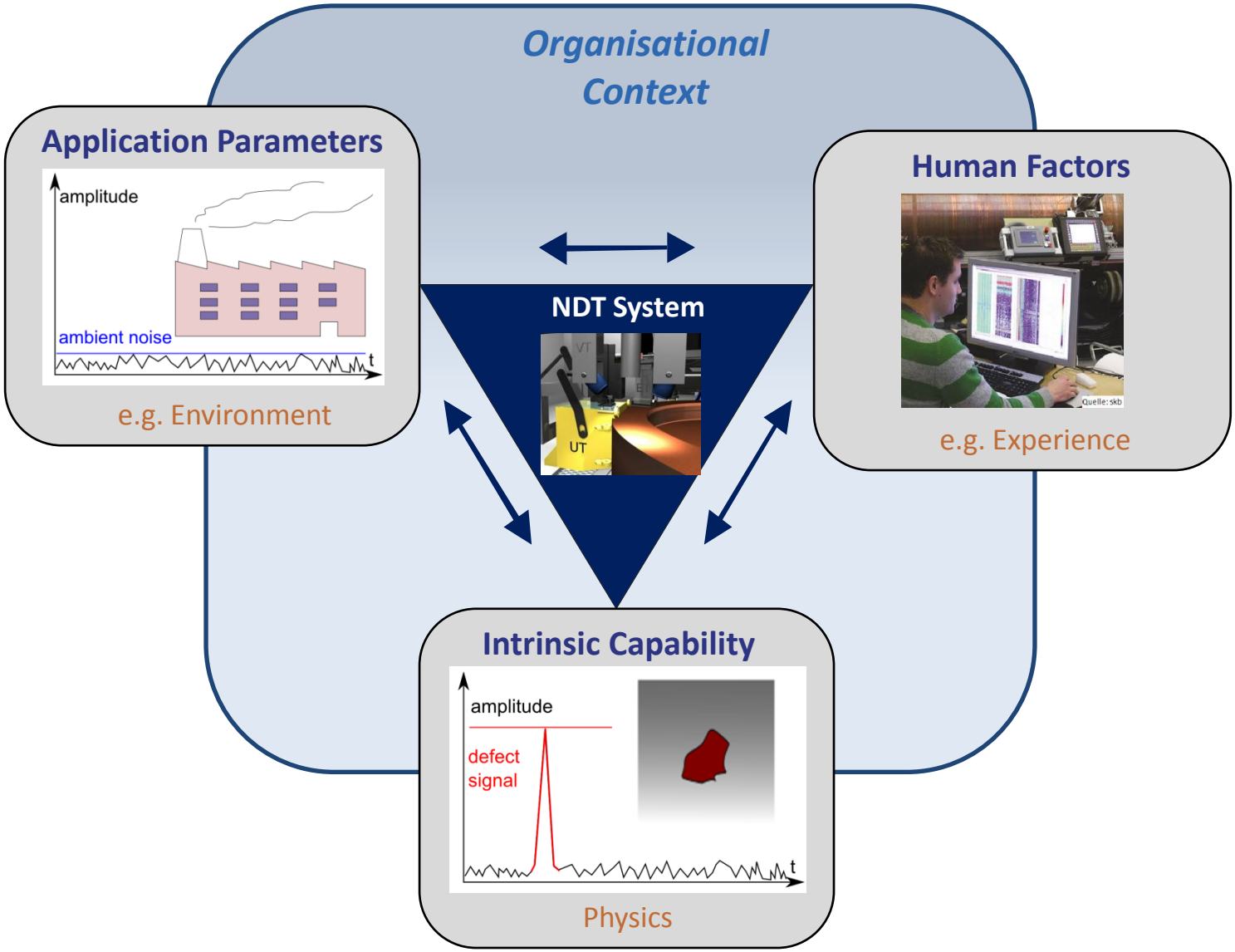
The answer of these question can be found in evaluating the NDT situation. For the correct evaluation probabilistic evaluations are used.

Consequences of decisions based NDT

Truth	Decision	Costs	Consequences
		Min. costs: NDT costs	0
		NDT costs + scrapping / repairing costs	$ (+ threat banished)
		NDT costs + scrapping (unnecessary) (+ loss of trust in the NDT department)	$$$
		NDT costs + failure of a defective component	$$$$$ 



Modular Reliability Model (Müller, 2009)

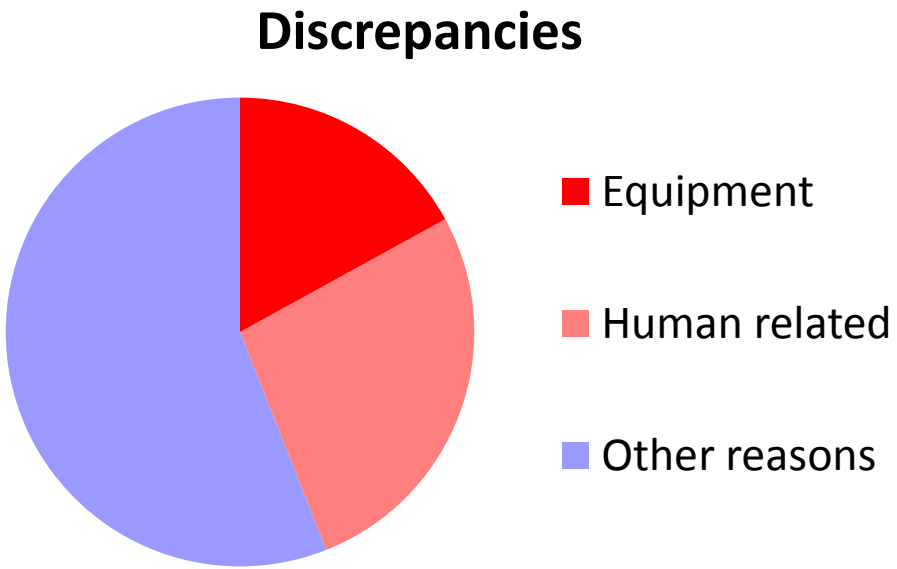


Consequences

Wrong decision happens

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

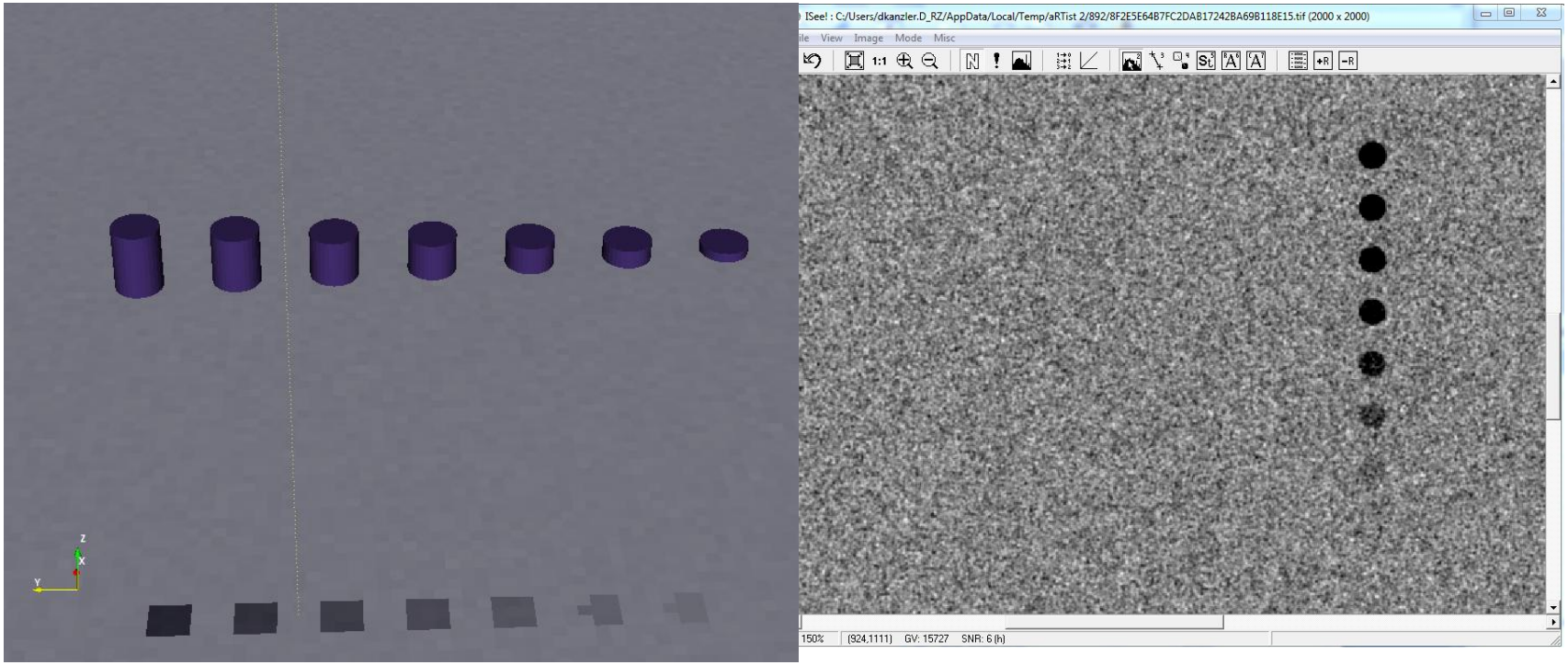
- "Nearly half of the discrepancies (a functional or structural anomaly, which may reveal itself as a deviation from requirements or specifications) were caused by the testing equipment and employee / operator error"
- "A research finds that 29% of discrepancies reported are related to test equipment."



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

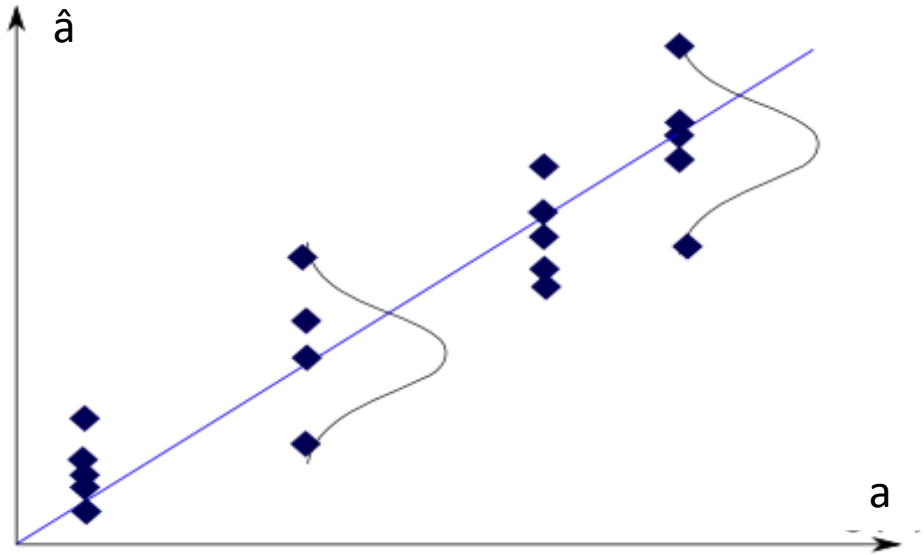
The contrast of "small" defects is lower than for "bigger" defects



POD – measurements with fluctuations

Theory vs. Reality!

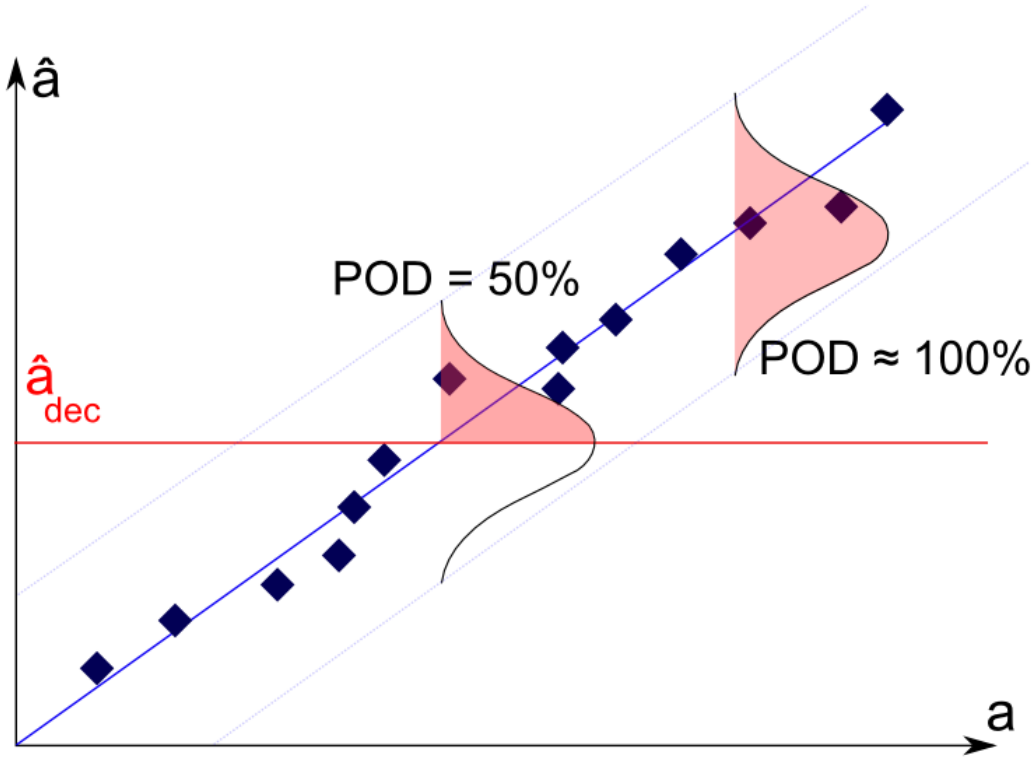
We don‘t measure the same contrast every time.



RT:
a: penetrated length of the material defect
â: contrast of the indication

Statistics provides models to describe the fluctuations!

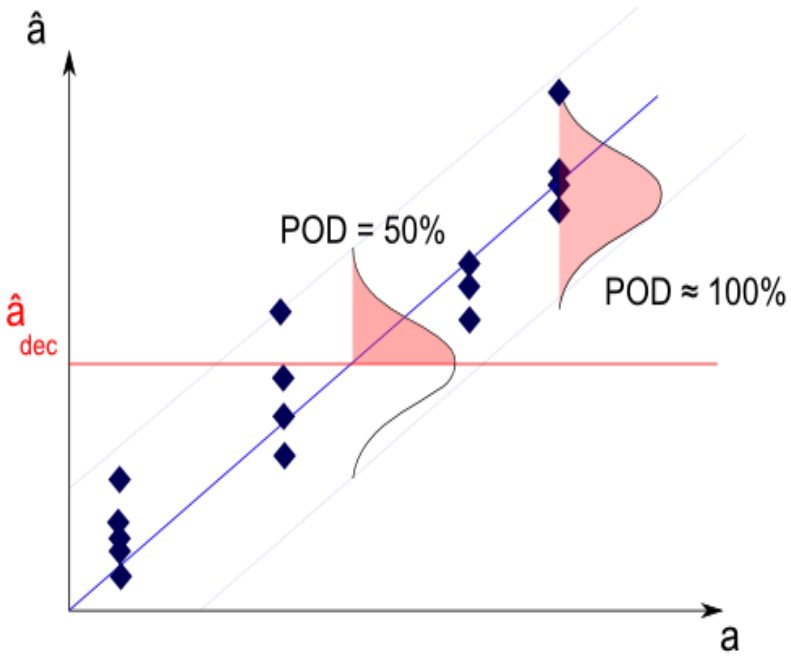
Linear â vs. a graph: decision threshold



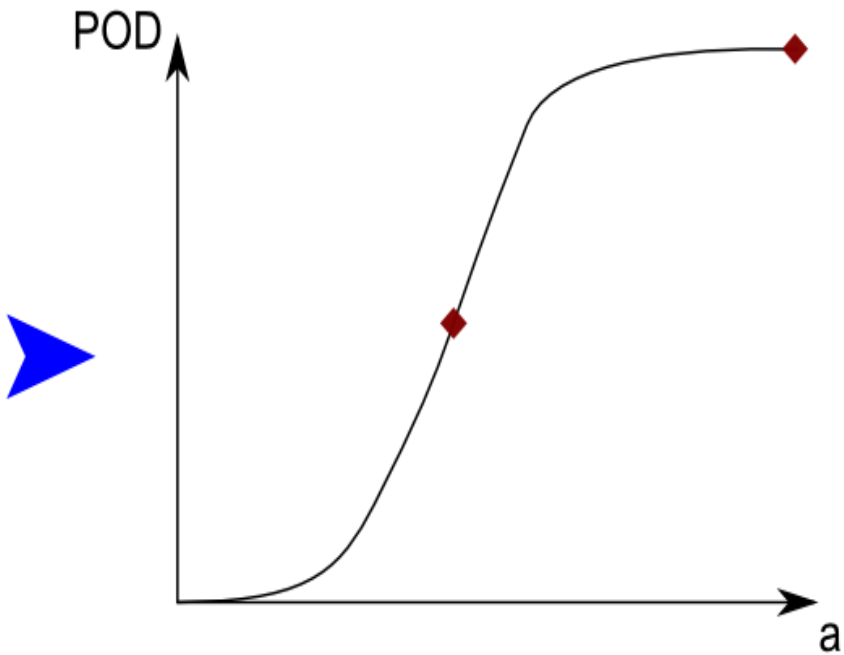
- The decision threshold is a part of the testing procedure: It often depends on the detectable contrast, the noise, etc..
- The probability of detection correlates with the area **under the distribution above the decision threshold** for each parameter size a.

POD from the linear â vs. a graph to the POD curve

Linear â vs. a graph

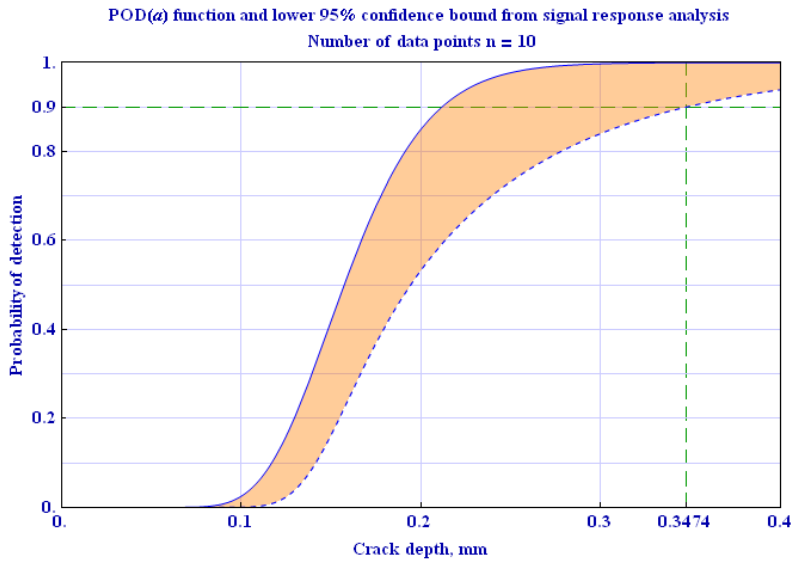


POD curve

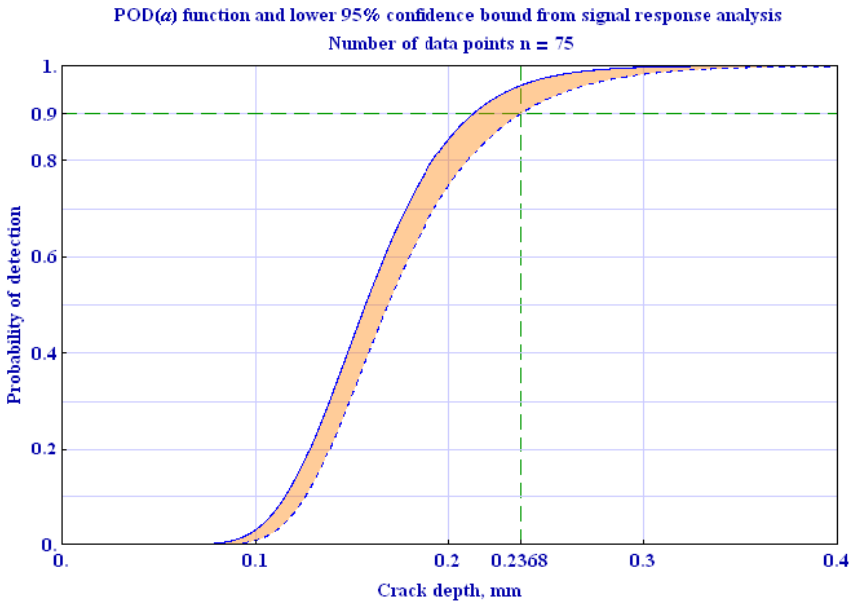


Amount of data and the confidence band

10 data points



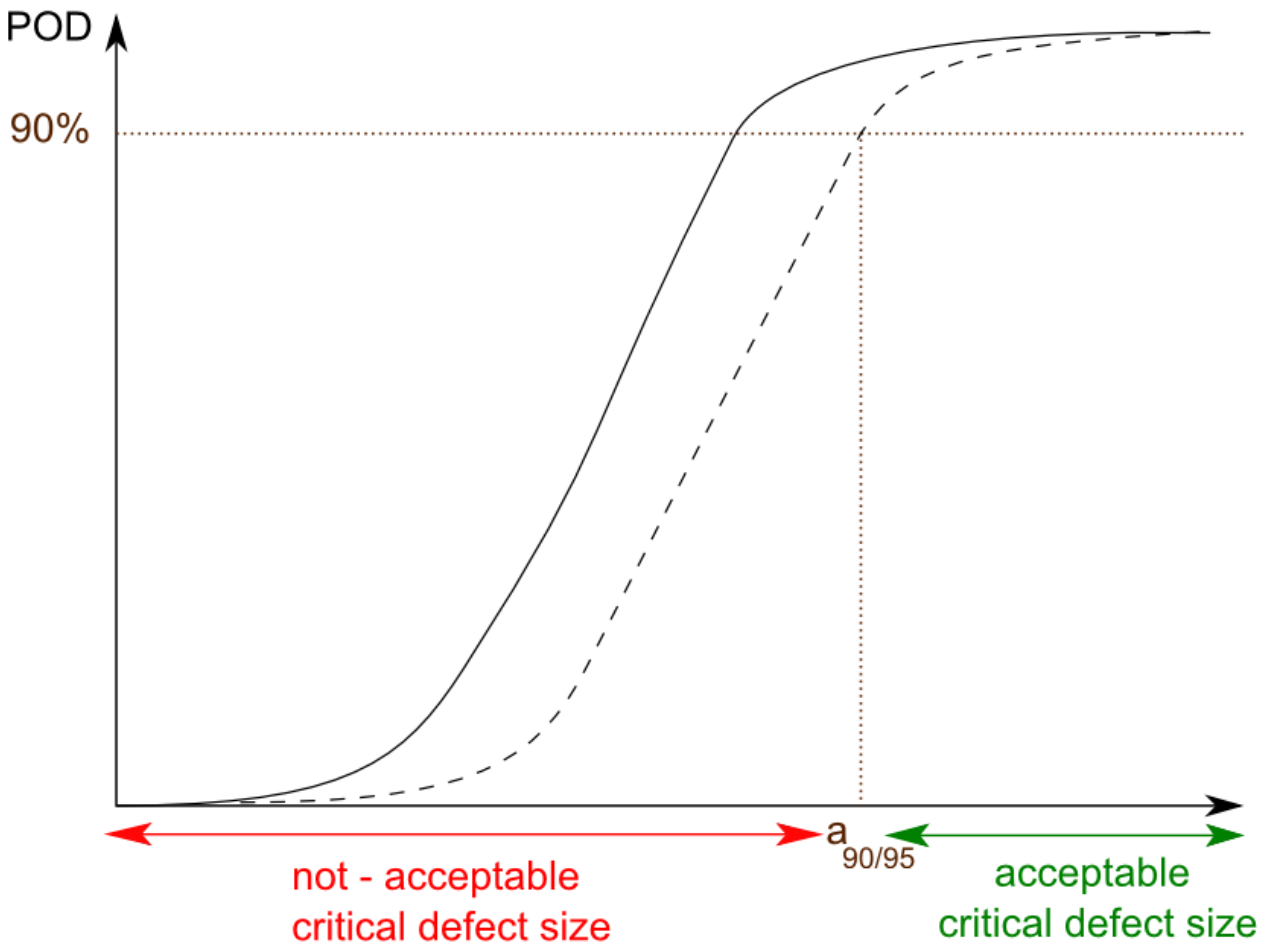
75 data points



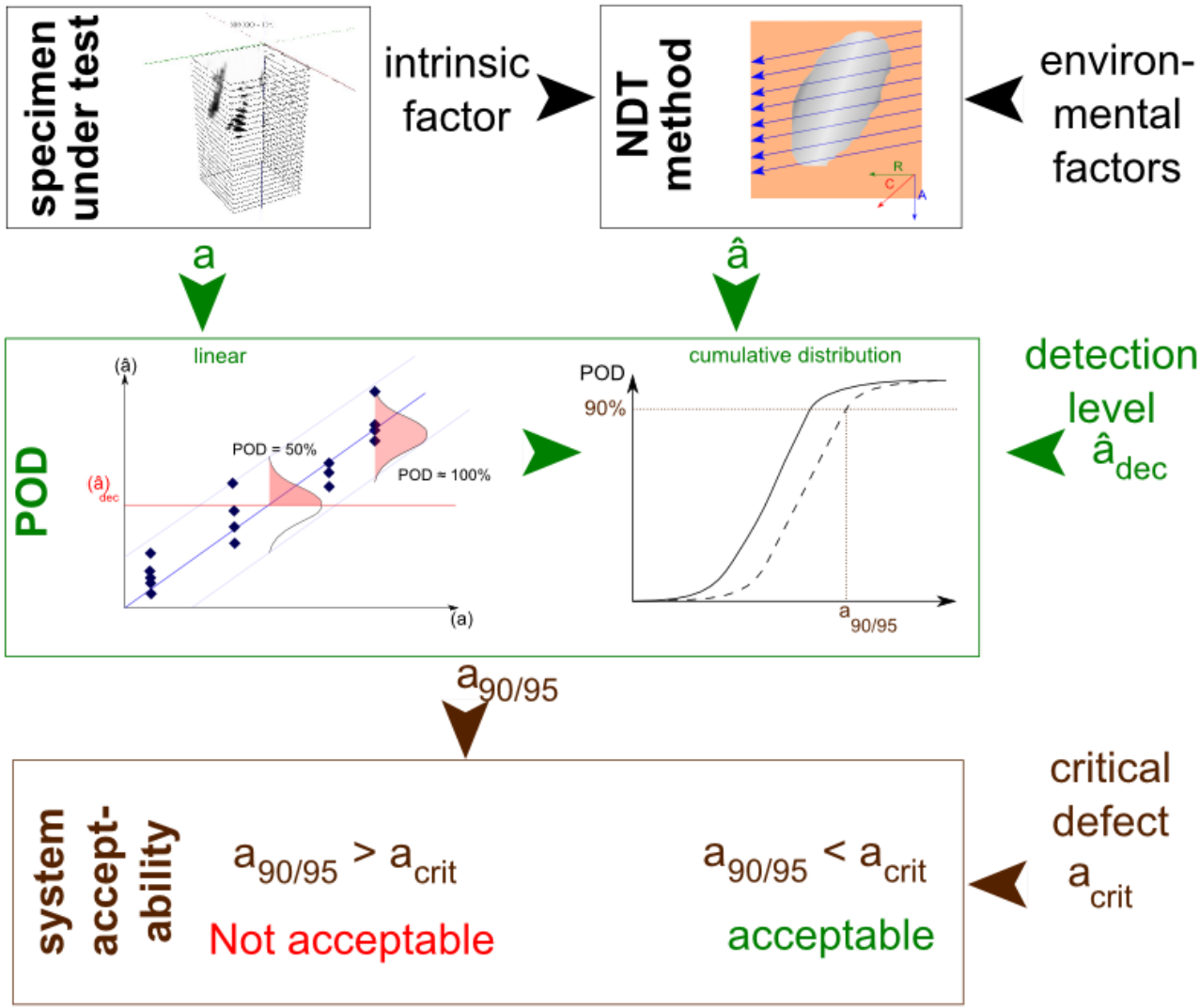
Pavlovic 2011

- 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 â vs a!**

What are we doing with the result of the POD?

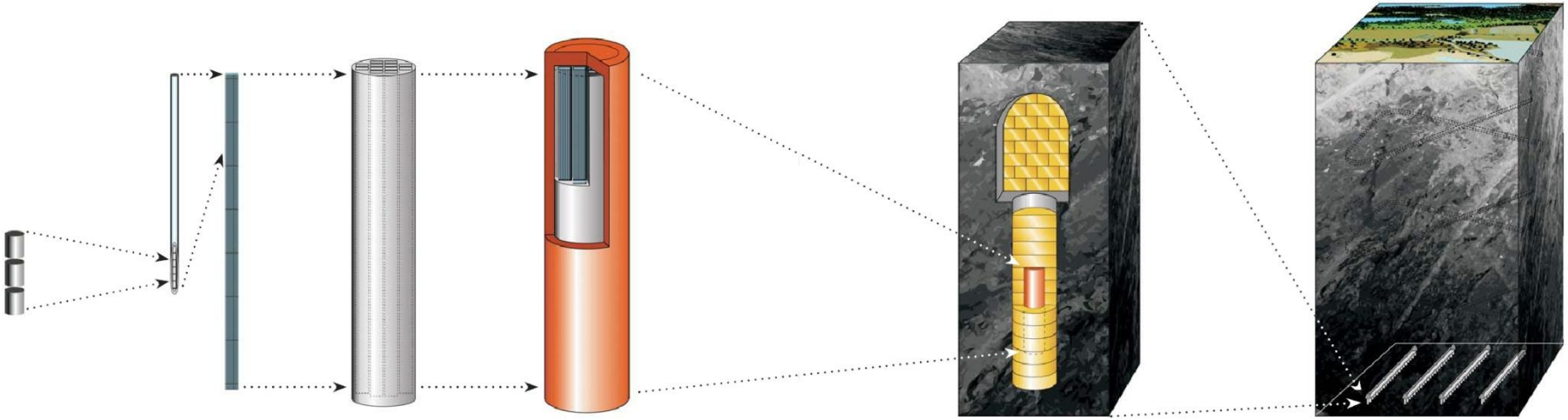


Summary: POD in radiographic testing



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

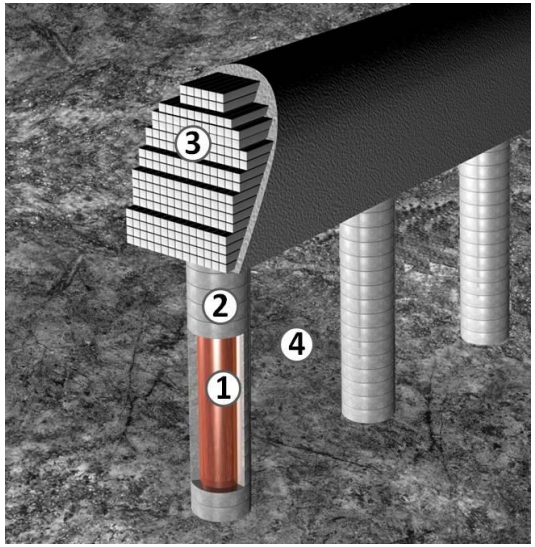


Figure source:
POSIVA Oy

Radiography and the connection to the POD

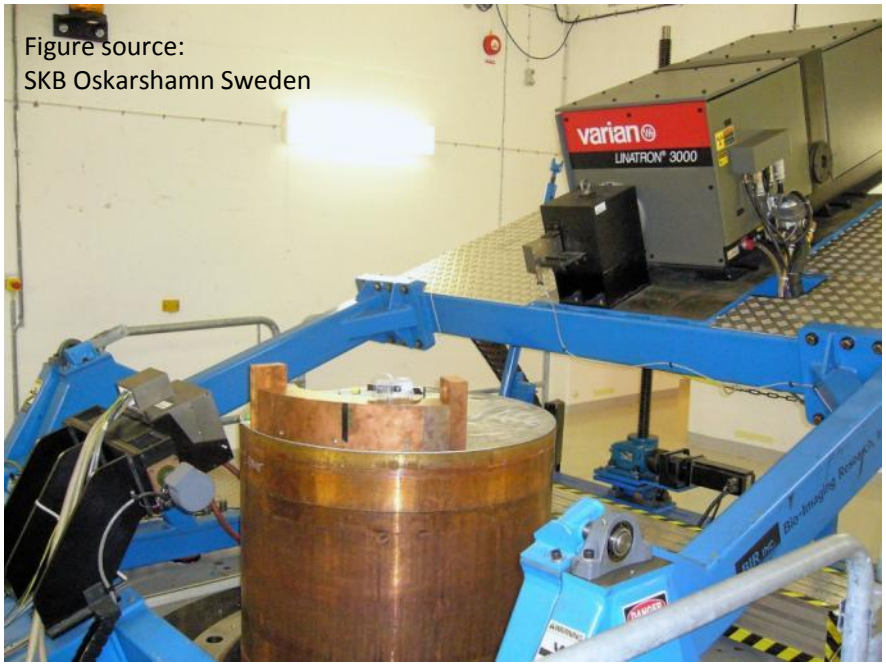
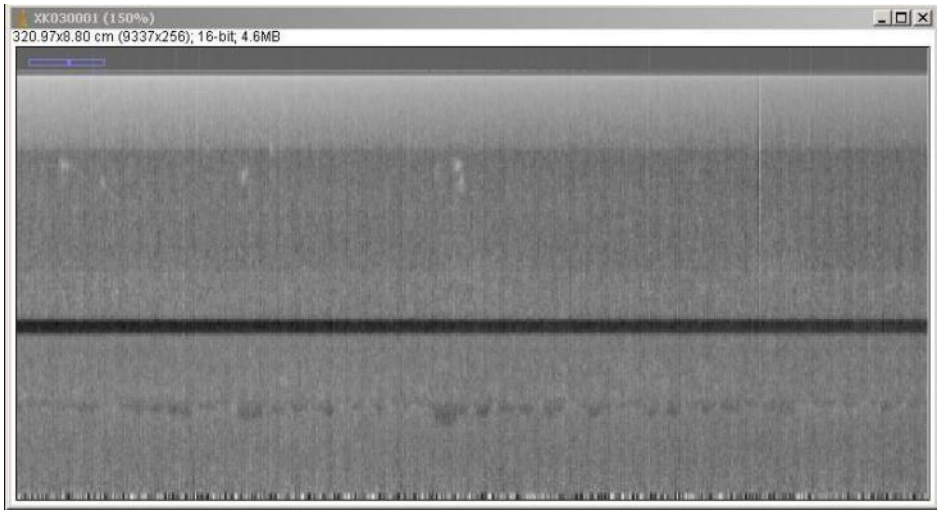


Figure source:
SKB Oskarshamn Sweden



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

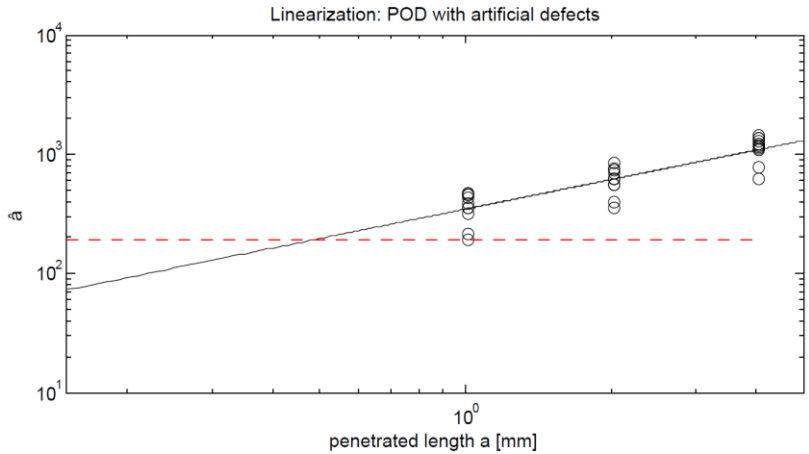
Reliability Characteristics	remaining wall thickness RW = W – a	
	W = wall thickness	
	a = penetrated length of the defect	=> $\hat{a}$ and $\hat{a}_{dec}$
NDE Systems	contrast	
	noise	

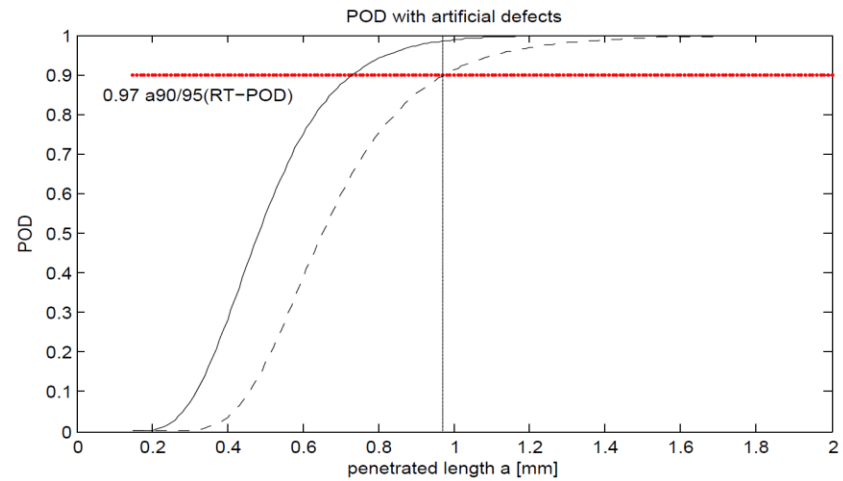
Examples: POD evaluation of an RT-system

â vs. a graph: linear regression

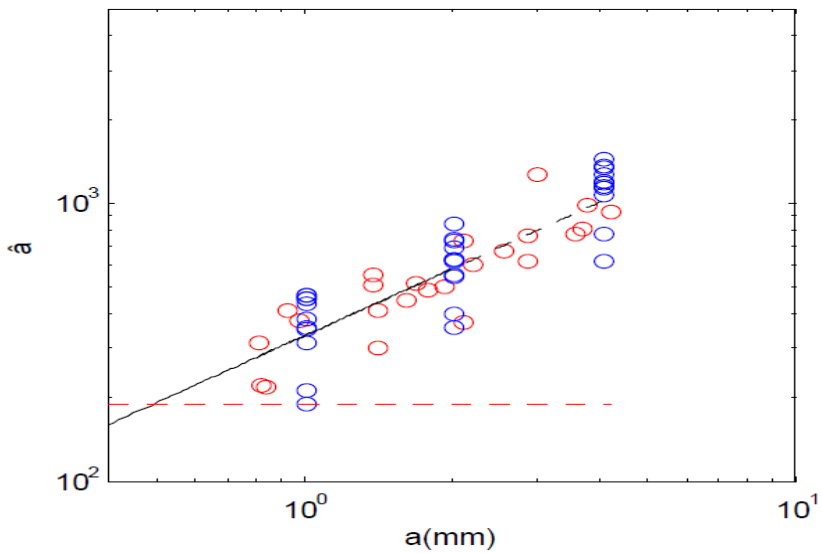
POD graph

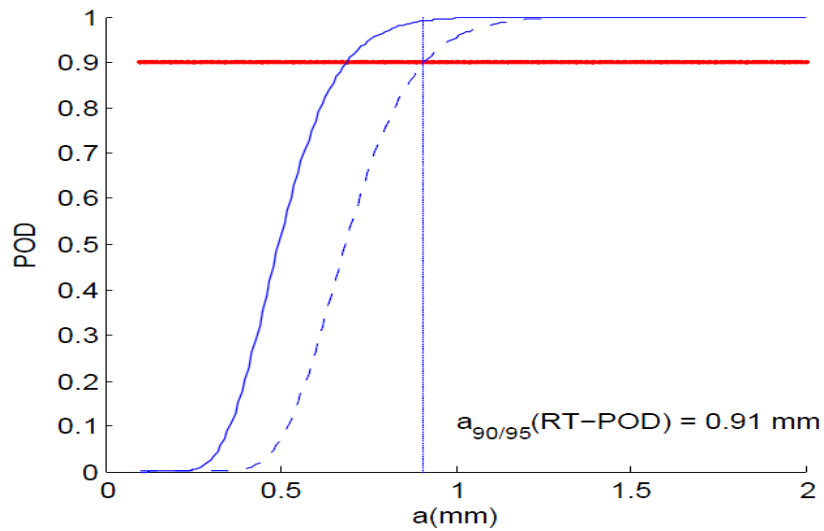
Artificial defects





Artificial and real defects





Thank you for your attention!

Any questions?

Daniel Kanzler M. Eng.

Email: KanzlerD@knc-reliability.com