AN INTRODUCTORY SURVEY
OF METHODOLOGIES
FOR ESTIMATING PROBABILITY OF DETECTION

Derek Sturges
GE Aircraft Engines
Quality Technology Center
10270 St. Rita Lane
Cincinnati, Ohio 45215

Telephone: 513-552-4612

1996 ASNT Spring Conference
Norfolk, Virginia

GOALS

To provide an introduction to concepts
involved in quantifying the capability
of detection methods,
typically expressed in terms of the
Probability of Detection (POD)
and (sometimes) the
Probability of False Alarm (PFA)

To review the principal features of several alternative methods

To illustrate use of those methods

OUTLINE

WHY?

SOME APPLICATIONS OF POD AND PFA

Introductory definitions

Examples of applications

PROBABILITIES OF DETECTION (POD) AND OF FALSE ALARM (PFA)

- **WHAT IS DETECTION?**
  - **Identifying conditions suggestive of the presence of a defect**

- **WHAT IS POD?**
  - **Quantitative expression of flaw-detection capability of NDE processes**
  - **The probability that a defect will be detected, given that it is present**
  - **The proportion of flaws (if present) that a process is likely to detect**
  - **The probability that a defect will be correctly identified as a defect**
    - **For some purposes a defect may be any detectable discontinuity**
    - **Otherwise, only discontinuities exceeding a critical size are defects**

- **WHAT IS PFA?**
  - **Probability that a non-defect will be incorrectly identified as a defect**
    - **The indication may be from "noise" or from a small discontinuity**

SOME APPLICATIONS OF POD AND PFA

- o **POD & PFA PROVIDE A MEANINGFUL BASIS FOR:**
  - **Measuring the capability of NDE methods**
    - **What size defects can a specific method detect?**
    - **What proportion of defects will be detected ... or missed?**
  - **Comparing inspection methods or inspection systems**
    - **Which penetrant is best?**
    - **Does eddy current detect smaller defects than FPI?**
    - **Does changing instrumentation affect the inspection?**
  - **Measuring the reliability of inspection methods**
    - **How consistently are defects likely to be detected?**
  - **Evaluating inspector training and performance**
    - **Is training improving performance?**
    - **Is performance adequate to meet business needs?**

- o **POD PLAYS AN ESSENTIAL ROLE IN LIFE MANAGEMENT**
  - **Relates inspection capability to life of cyclically-stressed products**
    - **Measures effect of better inspection in improving product life**
  - **Provides estimate of "initial flaw size" in fracture mechanics programs**
    - **Basis for scheduling in-service inspections**

WHY?

WHY DEAL WITH STATISTICAL PROBABILITIES?

Capability from Calibration
Capability from Precedent
Process Control Applications of NDE
Capability for Real Defects

MEASURING "CAPABILITY"

- o NDE CAPABILITY HAS RARELY BEEN MEASURED
  - Our perception of "capability" usually comes from calibration or precedent
- o MEANINGFUL QUANTIFICATION REQUIRES CAREFUL PLANNING
  - Usually start with a group of known flaws
  - Measure the proportion detected
  - Express the results in terms of POD
  - False alarm rates may also be needed

CAPABILITY FROM CALIBRATION

- o **TRUE STATEMENT**
  - **Ultrasonic inspection can reject all indications larger than that from a 0.0008 in^2 planar void**
- o **THE SUPPORTING EVIDENCE**
  - **Inspection is calibrated from a 3/64" diameter flat-bottomed hole (FBH)**
  - **Indications above 50% of that reference level are rejected**
- o **THE PROBLEM**
  - **Does not address what size <u>defect</u> would be detected or rejected**
  - **Might be misinterpreted as stating that all 0.0008 in^2 <u>defects</u> rejected**

CAPABILITY FROM PRECEDENT

- o **TRUE STATEMENTS**
  - **1 mm long cracks can be detected with Fluorescent Penetrant Inspection**
  - **FPI is <u>capable</u> of detecting 1 mm long cracks**
- o **THE SUPPORTING EVIDENCE?**
  - **An indication detected by FPI was caused by a 1 mm crack**
  - **This crack was preserved in a metallographic mount**
- o **THE PROBLEM**
  - **No measure of how many 1 mm long cracks were missed!**
  - **Might be misinterpreted as stating <u>all</u> 1mm cracks are detectable**

TYPICAL INDUSTRIAL UNDERSTANDING OF CAPABILITY

- LIMITED TO SIMPLE STATEMENTS LIKE THE PRECEDING
  - Correct and apparently innocuous as stated
  - Open to potentially dangerous misinterpretation
- WHAT SIZE DEFECT CAN NDE PROCESSES DETECT?
  - Honest answer: it has usually not been established!
- HOW LONG HAVE WE HAD TO GET THE ANSWER?
  - Major NDE processes have been used for 50 to 100 years!
- HOW DO WE EXPLAIN THIS APPARENT GAP?
  - The question has rarely been asked!

PROCESS CONTROL APPLICATIONS OF NDE

- MOST NDE APPLICATIONS IN MOST INDUSTRIES
  - Emphasis on consistency
    - Inspection should be independent of when or where conducted
  - Calibrate from artificial defects
    - Need reproducible reference targets for each site
- NDE USED IN "PACKAGE" OF PROCESS CONTROLS
  - Typically monitor from raw material through finished product
    - Chemical, mechanical, dimensional, etc.
  - Success is judged by a favorable field-service record
- TRUE CAPABILITY NOT ESSENTIAL FOR PROCESS CONTROL
  - Not essential to know capability of individual elements in the package

CAPABILITY FOR REAL DEFECTS?

- FLAW DETECTION DEPENDS ON NUMEROUS PARAMETERS
  - Material and manufacturing process properties
  - Flaw properties
  - Inspection equipment and inspection process parameters

- EXAMPLE: ULTRASONIC INSPECTION
  - Material properties:
    - Alloy composition and phase, grain size, forging flow pattern, etc.
  - Manufacturing processes:
    - Material configuration, surface texture, etc.
  - Flaw properties:
    - Size, shape, orientation, location, acoustic impedance, etc.
  - Inspection equipment parameters:
    - Beam properties, frequency response, electronic noise, linearity, etc.
  - Inspection process parameters:
    - Wave mode, calibration, scan speed & index, operator skills, etc.

CAPABILITY FOR REAL DEFECTS?

- CAPABILITY IS EXPRESSED IN PROBABILISTIC TERMS
  - Inspection processes are too complex to be described deterministically
  - NDE typically provides too little information
    - Example: ultrasonics provides only 2 outputs: amplitude and delay but ultrasonic response depends on far more than 2 factors

- FOUR OUTCOMES TO ANY INSPECTION
  - Detection of a defect that is present ("true positive")
    - Probability of Detection (POD)
  - Non-detection of a defect that is present ("false negative")
    - Probability = (1 - POD)
  - Apparent detection of a defect that is not present ("false positive")
    - Probability of False Alarm (PFA)
  - Non-detection of a defect that is not present ("true negative")
    - Probability = (1 - PFA)

WHEN?

MAJOR DEVELOPMENTS AND THEIR MOTIVATIONS

Overview
Signal Detection Theory
Power Generation Industry
Aerospace Industry
Modeling POD

DEVELOPMENT OF POD CONCEPTS

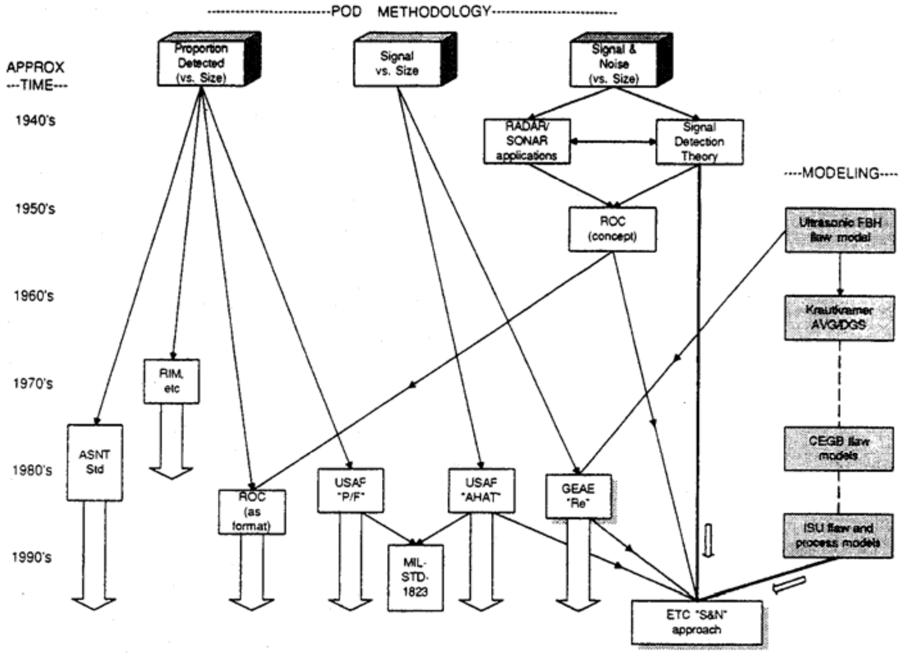
- **EARLIEST RELATED WORK: 1940's**
  - **Development of Signal Detection Theory**
  - **Applications to RADAR and SONAR**
  - **Not applied to NDE**

- **FIRST ATTEMPTS TO QUANTIFY NDE CAPABILITY?: 1950's and 1960's**
  - **Krautkramer ultrasonic flaw models**
  - **Power generation rotor life predicted from estimated flaw size**

- **MAJOR AEROSPACE EFFORTS: 1970's and 1980's**
  - **NASA and US Air Force programs**
  - **First standardized procedures for estimating POD**

- **APPLYING SIGNAL DETECTION THEORY TO NDE: 1990's**
  - **Federal Aviation Administration programs**

OVERVIEW OF THE DEVELOPMENT OF POD CONCEPTS



D.J STURGES, GE AIRCRAFT ENGINES

SIGNAL DETECTION THEORY

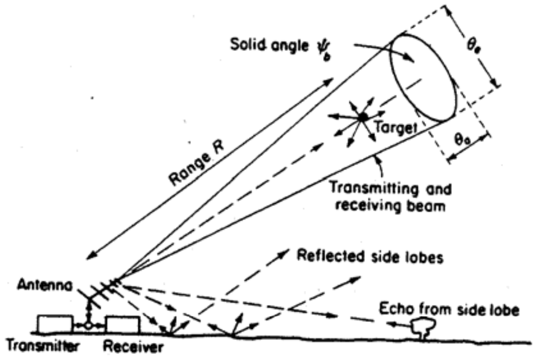
- **RADIO-WAVE DETECTION THEORY**
  - **Application of statistical decision/detection theory**
    - **Probabilistic descriptions of uncertainty about signals and noise**
  - **Gives relationship between Probability of Detection (POD), Probability of False Alarm (PFA), and Signal-to-Noise ratio (S/N)**
  - **Founded on work by North (RCA, 1943), Uhlenbeck (MIT, 1943), Rice (Bell Telephone, 1944), Marcum & Swerling (Rand, 1947), and others**

- **RADAR/SONAR APPLICATIONS (1943 et seq.)**
  - **Measurement of performance in detection of aircraft or ships**
  - **Improvement in receiver circuitry to optimize signal detection**
  - **Development of autodetection hardware to eliminate the human observer**

SIGNAL DETECTION THEORY

- **BASIS**
  - **Measurement of signal & noise distributions, or of POD and PFA**
- **ADVANTAGES**
  - **A rationale for evaluating and optimizing detection system performance**
    - **A framework for examining "co-variation" of detection and false alarm**
- **LIMITATIONS**
  - **Results depend on the specific signal and noise distributions**
    - **May be qualitatively - but not quantitatively - transferable**
  - **Threshold optimization depends on the specific detector (circuitry)**
    - **Caution: most radar analyses assume time-varying Gaussian noise**
    - **Caution: radar analyses are based on detector types uncommon in NDE**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS



Typical radar-target geometrical relationships

- **DETECTION GOAL**
  - **Sensing the presence of the sought-after target in the presence of competing indications which arise from background radiation, undesired echoes, or noise generated in the receiver**

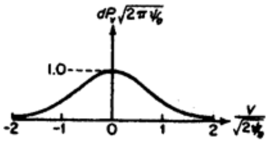
SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

- **RADAR "SIGNALS"**
  - **A delayed series of echo-pulses of radio-frequency energy**
    - **Affected by beam pattern, target reflectivity & transmission medium**
    - **Doppler-shifted by radial velocity of target**
- **RADAR "NOISE"**
  - **Thermal noise**
    - **Human-origin, atmospheric, solar or galactic noise**
    - **Waveguide, duplexer, or receiver noise**
  - **Non-thermal noise**
    - **Target signals from side-lobe echo-paths**
    - **Ground "clutter" - steady (e.g. buildings) or time-varying (e.g. waves)**
    - **"Jamming" (i.e. deliberate interference)**

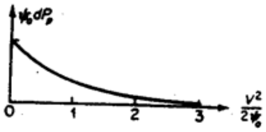
SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

- **INITIAL GOAL**
  - **Optimizing receiver design for detecting target-reflected signals**
- **INITIAL APPROACH (D.O.North)**
  - **Analyze probability that a threshold would be exceeded by:**
    - **A single sample of random noise**
    - **A single sample of sinusoidal signal added to random noise**
  - **Assumptions:**
    - **Random noise confined to narrow frequency band by (i-f) filtering**
    - **Envelope detection used (i.e. "video" signals)**

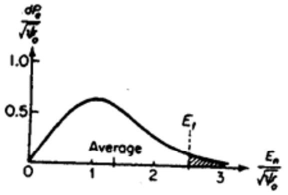
SIGNAL DETECTION THEORY FOR RADAR SYSTEMS



Gaussian distribution of i-f noise



Exponential distribution of i-f noise power

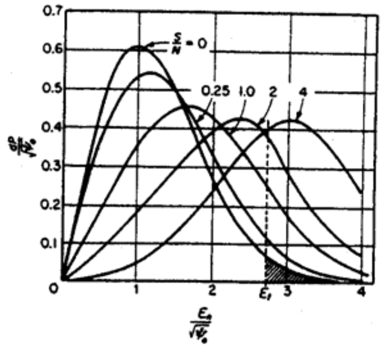


Rayleigh distribution of detected envelope

- PROBABILITY DISTRIBUTIONS OF NOISE
  - Distributions corresponding to North's assumptions
  - Probability (PFA) that a single sample of detected noise exceeds threshold level E_t is given by the area under the curve to the right of E_t

---

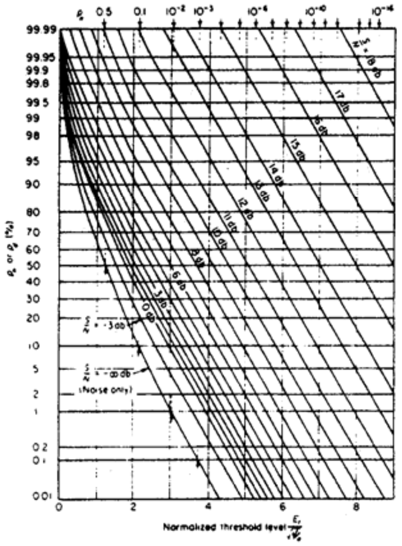
SIGNAL DETECTION THEORY FOR RADAR SYSTEMS



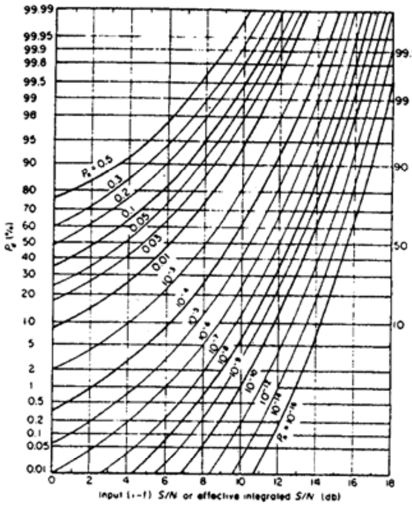
Probability distributions of the envelope of signal-plus-noise

- DETECTED OUTPUT WITH SIGNAL PRESENT
  - Envelope detection: narrow-band Gaussian noise, with power N
    - Steady sinusoidal signal, with received signal power S
  - Single-pulse detection probability (POD) is the area to the right of E_t under the curve for each value of S/N

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

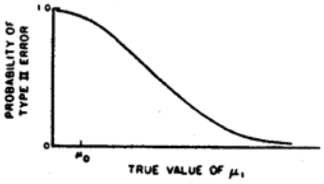


Probability of crossing threshold (numerical analysis by Rice)

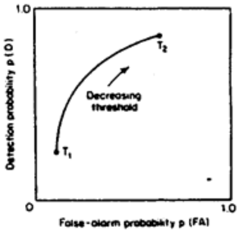


Probability of detection vs. S/N (single-pulse POD for radar systems)

SIGNAL DETECTION THEORY



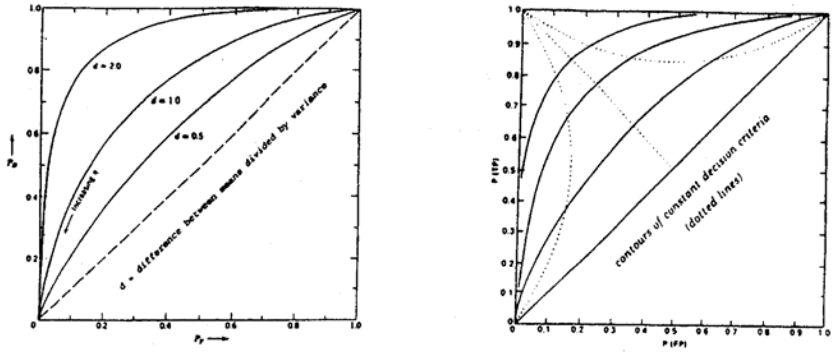
Example of an Operating Characteristic



Example of a Relative Operating Characteristic

- **OPERATING CHARACTERISTICS**
  - **Statisticians' designation for specific types of probability plots**
    - **The evaluation component of detection theory**
  - **ROC: a graphic way of comparing two operating characteristics**
    - **For Radar applications, known as <u>Receiver</u> Operating Characteristics**
    - **Later renamed <u>Relative</u> Operating Characteristics**
  - **Merged with aspects of information theory and systems analysis**
    - **Founded on work of Peterson & Birdsall (University of Michigan, 1953), J.A.Swets (Bolt Beranek & Newman, 1954), and others**

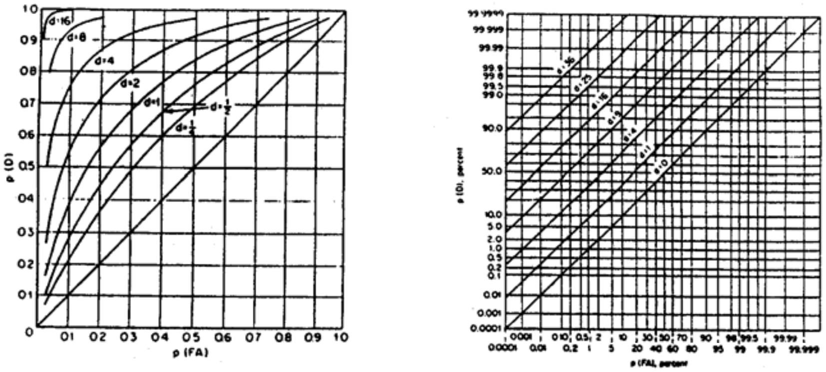
SIGNAL DETECTION THEORY FOR RADAR SYSTEMS



Single-pulse POD curves replotted in ROC format

- **RECEIVER OPERATING CHARACTERISTIC**
  - Each curve plots POD against PFA for a single S/N as E_t is varied
  - High POD with low PFA is usually the most desirable goal
    - Implies high S/N ratio
    - Provides a rationale for selecting optimum threshold

SIGNAL DETECTION THEORY FOR SONAR SYSTEMS

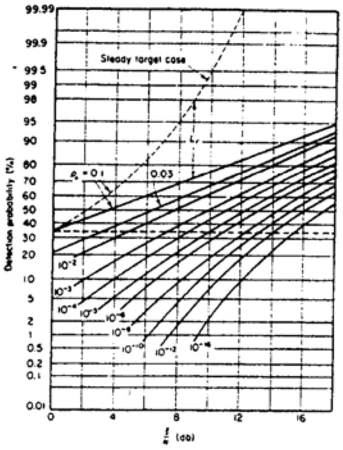


ROC curves reformatted on probability coordinates

- **ROC CURVES IN BINORMAL COORDINATES**
  - Noise and signal-plus-noise taken as Gaussian, with equal variance
    - ROC curves become straight lines of unit slope
  - Use is made in Psychoacoustics of a Detection Index, d
    - Defined as squared difference between means divided by variance

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

Probability of detection for a fluctuating target



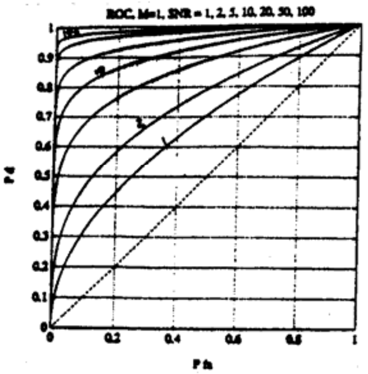
- **EFFECT OF VARYING TARGET SIGNAL**
  - **Single-pulse detection probabilities in Gaussian noise**
  - **Effect of a Rayleigh-distributed echo voltage**
  - **Fluctuation raises low POD's and reduces high POD's**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

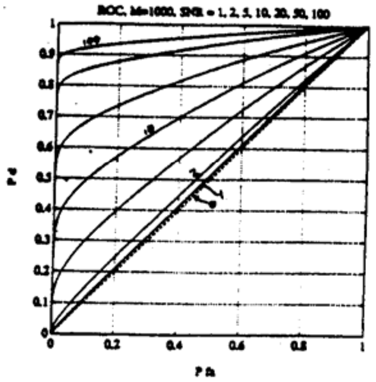
- **INTEGRATION OF PULSE TRAINS**
  - **Radar system detection is rarely based on a single pulse**
  - **Enhanced POD results from efficient use of a train of pulses**
    - **An ideal detector would add energy from n pulses in a matched filter**
      - **Would give integrated S/N ratio n times the single-pulse value**
      - **Practical radars using video integration achieve about $n^{0.6}(S/N)_1$**

- **SEARCH (SCANNING) RADAR**
  - **Fast scanning increases chance of intercepting target on more than one scan**
    - **Cumulative POD is increased (if scans are statistically independent)**
    - **$P_n = 1 - (1 - P_i)^n$**
  - **But faster scanning reduces number of pulses integrated per scan**
  - **System parameters may be chosen for optimum net POD**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS



Inspection of a single cell



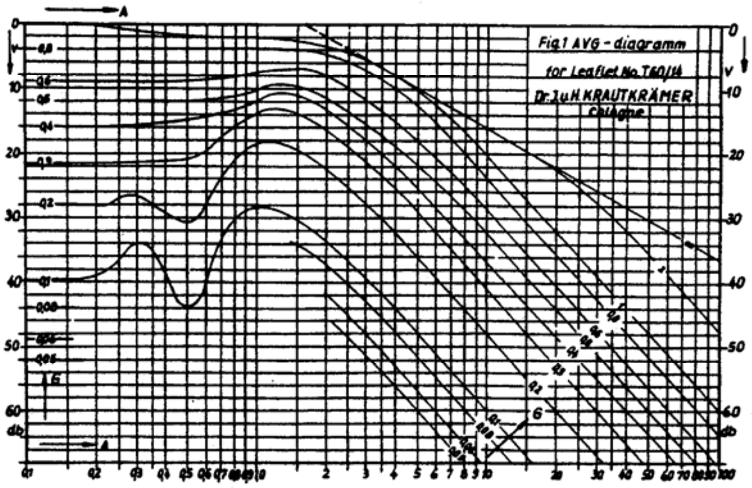
Inspection of 1000 cells

- o **BROAD AREA SEARCH EFFECTS**
  - **Increase the number of area or volume elements searched from 1 to M**
    - **Chance that noise in one cell will exceed threshold increases**
    - **E_t must be raised to keep PFA constant**
    - **Signal-to-noise ratio must be raised to keep POD constant**
  - **ROC curves tend to assume "peripheral threshold" shape**

EARLY APPROACHES TO QUANTIFYING DETECTION CAPABILITY

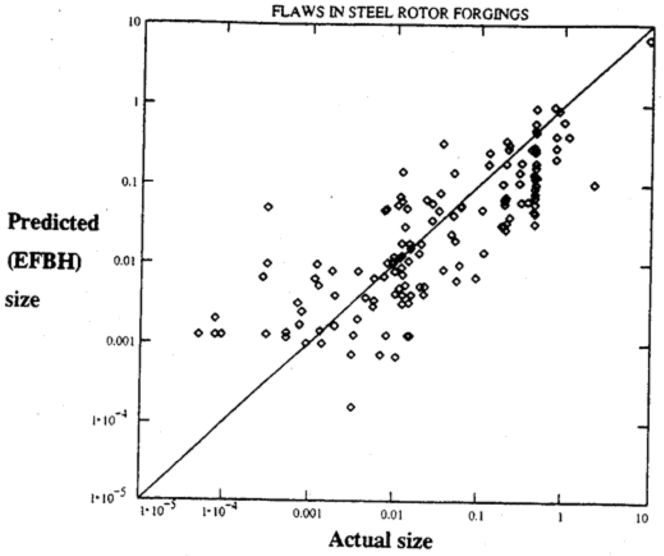
- MATHEMATICAL MODELLING OF ULTRASONICS (1950's)
  - Theoretical description of transducer/reflector interaction
  - The flat-bottomed hole (FBH) as a defect model
- INDUSTRIAL APPLICATION BY THE KRAUTKRAMERS (and others)
  - Krautkramer AVG/DGS diagram
    - Quantitative relation between backwall and planar disk responses
    - Defect size in terms of "Equivalent FBH" (FBH giving same indication)
  - Basis for life prediction for steam turbine & generator rotors
    - GE (1960's); ASTM Standard A418 (1974)

THE KRAUTKRAMER DISTANCE-GAIN-SIZE DIAGRAM



A = Distance (in near-field lengths)
V = Gain (in dB)
G = Size (of a FBH relative to the transducer diameter)

POWER GENERATION INDUSTRY

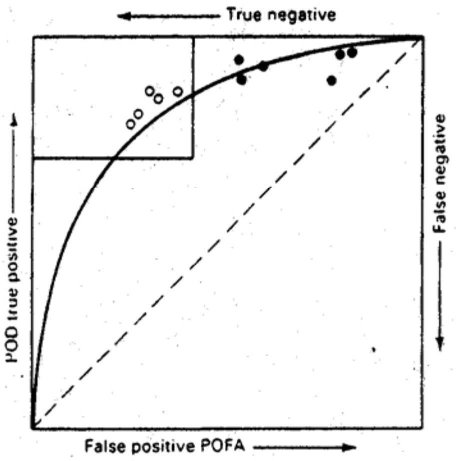


- COMPARISON OF EFBH AND ACTUAL DEFECT SIZES (Schwant & Timo, 1984)
  - Data used to adjust EFBH estimate for individual indications
  - Modified EFBH size used in predicting rotor life

POWER GENERATION INDUSTRY

- POD/RELIABILITY MODELLING APPROACH
  - Central Electricity Generating Board (Coffey, 1977 et seq.)
  - Theoretical estimates for upper bounds on crack detection and sizing
- RELATIVE OPERATING CHARACTERISTIC (ROC) APPROACH
  - ROC format adopted for presentation of reactor inspection data
    - Used mostly to compare inspector performance
  - More on this later ....

POWER GENERATION INDUSTRY



Example of ROC presentation of data

AEROSPACE INDUSTRY

- o **USAF, NASA, and US AEROSPACE INDUSTRY**
  - **Work began in the late 1960's**
  - **Motivated by need for data for fracture-mechanics analysis**
    - **Needed to counter high cost of maintenance & replacement**
- o **AIRFRAME INDUSTRY PROGRAMS**
  - **First detailed study published in 1968**
  - **Fracture-control procedures initiated after loss of F-111 (1969)**
  - **Several major sponsored programs run in the mid-1970's**
    - **Lockheed-Georgia evaluated over 500,000 inspection opportunities**
    - **Reported POD as a function of process, material, inspector, etc.**

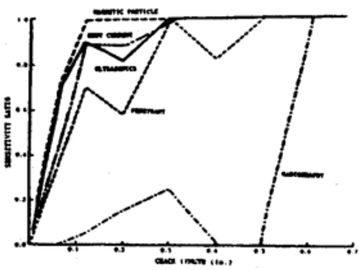
AEROSPACE INDUSTRY

- o **PAST NASA PROGRAMS**
  - **Awarded contracts for acquisition of quantitative NDE data (1973)**
  - **Produced Shuttle Fracture Control Plan (1974)**
    - **Defect detectability based on penetrant POD data**
    - **Distinguished "standard" and "special" NDE capability standards**
  - **Fracture Control applied to Shuttle main engine (1980)**
  - **NASA FLAGRO (fracture analysis program) released (1986)**
- o **FUTURE NASA PROGRAMS**
  - **Distribution of NASA NDE Reliability Computer Program**
  - **Large-scale effort for fabrication of defect samples**

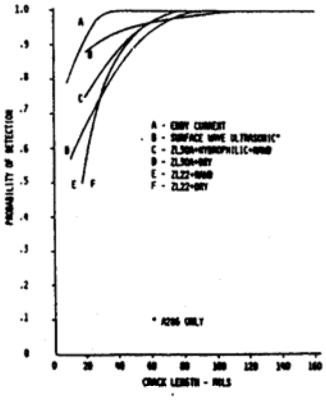
AEROSPACE INDUSTRY

- o US AIR FORCE PROGRAMS
  - AIRCRAFT STRUCTURAL INTEGRITY PROGRAM (ASIP)
    - Requirements defined in MIL-STD-1530A (USAF), issued in 1975
      - First applied rigorously to B-1 bomber
      - ASIP criteria have been applied to over 20 US military planes
  - DURABILITY AND DAMAGE TOLERANCE ASSESSMENT (DADTA)
    - Retrospective review of existing engines (initiated 1978)
  - ENGINE STRUCTURAL INTEGRITY PROGRAM (ENSIP)
    - Requirements defined in MIL-STD-1783 (USAF), issued in 1984

AEROSPACE INDUSTRY



Airframe data (early 1970's)



Typical 1970-period POD data

Engine data (late 1970's)

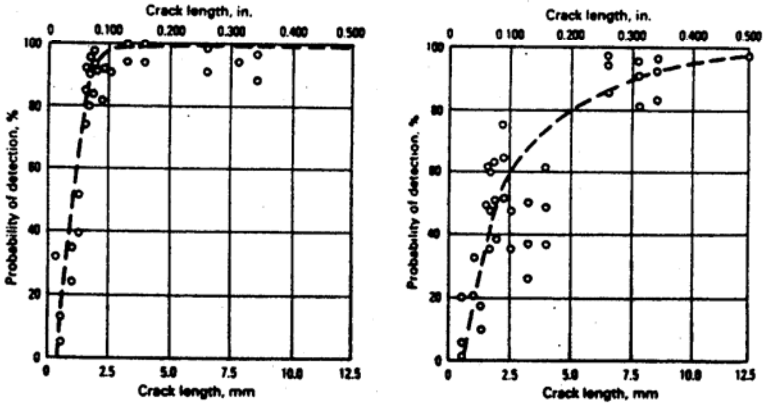
- o INDIVIDUAL MANUFACTURER PROGRAMS
  - Interest grew in use of POD concepts for other purposes (1970's)
    - Comparison of effectiveness of two penetrants
    - Comparison of effectiveness of different NDE techniques

AEROSPACE INDUSTRY

- EARLY ANALYTICAL METHODS
  - Early analyses based on binomial statistics
    - Each defect reported as "hit" or "miss"
    - Arbitrary averaging used to obtain smoothly increasing curves
  - ASNT Standard published in 1982
    - Focuses on POD for single flaw size and confidence level
- USAF LED DEVELOPMENT OF NEW STATISTICAL METHODS
  - Parametric regression techniques make better use of data
    - Incorporated in MIL-STD-1823 for NDE Reliability Assessment (pending)
- GE INCLUDED PHYSICAL MODELING IN AN ALTERNATIVE METHOD
  - Developed to deal with special needs for ultrasonic inspections
    - Combines linear regression with simple physical model for flaw response

---

AEROSPACE INDUSTRY



Eddy current inspection

X-ray inspection

328 cracks, lengths 0.3 to 18 mm, machined and etched surfaces

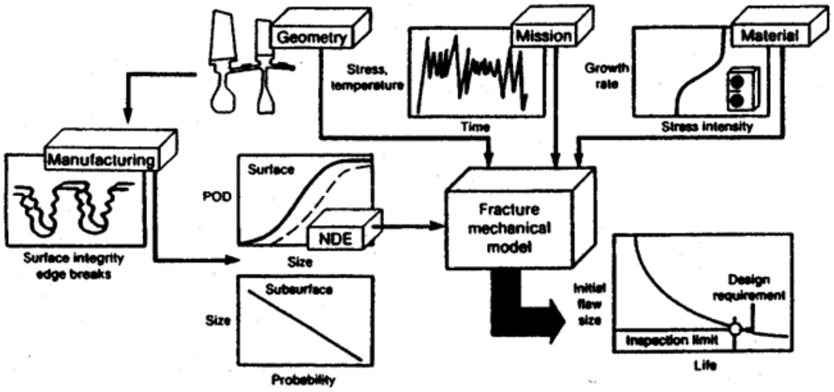
Typical format for current aerospace POD data

AEROSPACE INDUSTRY

- **DESIGN AND LIFING PHILOSOPHIES ARE CHANGING**
  - **"Durability" being replaced by "Damage Tolerance"**

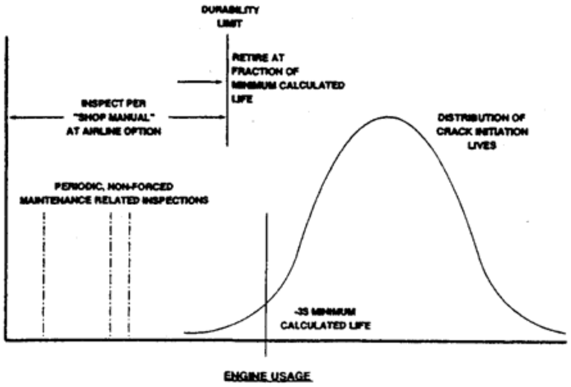
- **DAMAGE TOLERANCE**
  - **Ability of a structure to resist failure due to flaws or other damage for a specified period of unrepaired usage**
  - **Invokes fracture-mechanics and POD**
  - **USAF is encouraging this transition**
  - **FAA has recommended manufacturers to incorporate damage tolerance concepts (1990 Titanium Review Team recommendations)**

AEROSPACE INDUSTRY



Outline of a fracture-mechanics model for life-management calculations

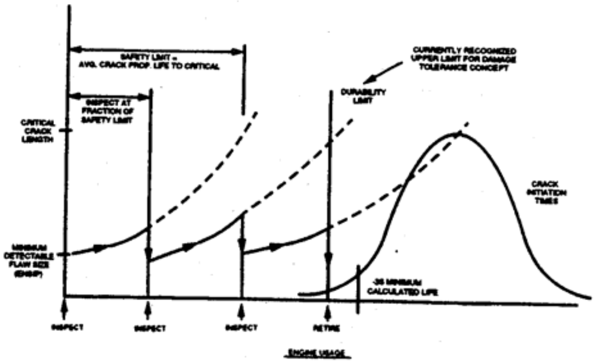
AEROSPACE INDUSTRY



- **ROLE OF NDE IN DURABILITY DESIGNS**
  - Design process does not invoke specific capability of NDE processes
    - Production parts are assumed to contain no (detectable) defects
    - Assumes no in-service damage (resulting in detectable defects) occurs
  - NDE is used during manufacture, as a check on process control
    - Capability is less important than repeatability and reproducibility
  - In-service "inspections of opportunity" are used as added safeguards

---

AEROSPACE INDUSTRY



- **ROLE OF NDE IN DAMAGE TOLERANT DESIGNS**
  - Design process requires knowledge of capability of NDE processes
    - Inspected parts are assumed to contain defects (below detection limit)
  - NDE is used during manufacture as a vital part of life-management process
    - Capability is expressed in terms of Probability of Detection
  - In-service inspections are mandated at intervals set (in part) by POD

AEROSPACE INDUSTRY: ENGINE TITANIUM CONSORTIUM

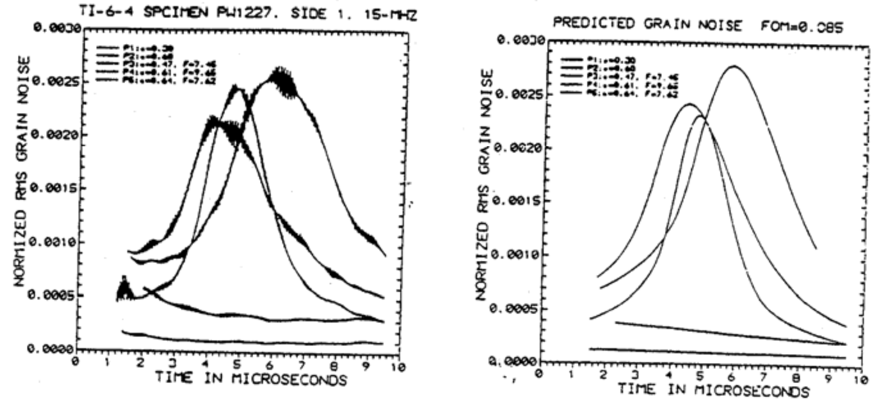
- **FEDERAL AVIATION ADMINISTRATION**
  - **Titanium Rotating Component Review Team report (1990)**
  - **Formed Engine Titanium Consortium (ETC)**
    - **Allied Signal, GE Aircraft Engines, Iowa State University, Pratt & Whitney**
    - **To encourage implementation of TRCRT recommendations**

- **OBJECTIVES, TASK 3 - PROBABILITY OF DETECTION**
  - **To develop methodologies to estimate and/or determine the probability of detection of flaws, especially subsurface, in (titanium) materials**
  - **To verify the methodologies**
  - **To provide appropriate quantitative information to allow risk and life management studies to be carried out**

AEROSPACE INDUSTRY: ENGINE TITANIUM CONSORTIUM

- **FOCUS FOR ETC POD TASK**
  - **Develop consensus on POD needs for aircraft engine applications**
    - **Highlighted by consideration of ultrasonic inspection of titanium**
    - **Methodology should apply to other techniques and materials**
  - **Identify practical, flexible, cost-effective approach:**
    - **Decouple effects of flaw, material, instrumentation and technique parameters**
    - **Good prospect for improving transferability of POD data**
  - **Starting point:**
    - **Base data acquisition on measured distributions of "signal" and "noise"**
    - **Use physical modelling to predict influence of individual parameters**
    - **Use Detection Theory concepts in statistical analysis of data**
  - **Target implementation through a formal (national) standard practice**
    - **Define terminology and concepts, and detail a procedure**

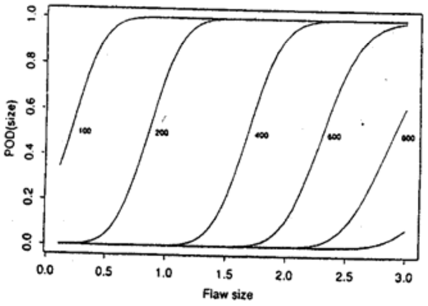
AEROSPACE INDUSTRY: ENGINE TITANIUM CONSORTIUM



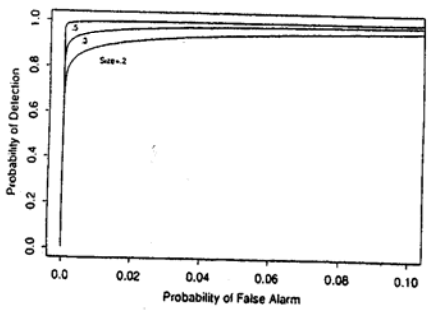
Measurements and modeled noise for five transducers

- o **STATUS OF CONSORTIUM POD TASK**
  - **Improved capability to model material noise and flaw response**

AEROSPACE INDUSTRY: ENGINE TITANIUM CONSORTIUM



POD for FBH's at various thresholds



POD and PFA for various size FBH's

- o **STATUS OF CONSORTIUM POD TASK**
  - **Improved capability for synthesizing flaws demonstrated**
    - **Disks, spheres and cylinders with realistic acoustic properties**
  - **Detection Theory approach demonstrated on voids and synthetic inclusions**
    - **POD and PFA derived from measurements of S and N distributions**

INTER-INDUSTRY PROGRAMS

- o REVIEW OF PROGRESS IN QUANTITATIVE NDE
  - Started in 1974 under DARPA/AFWAL sponsorship
  - Currently organized by the Iowa State University Center for NDE
    - Cosponsored by ASNT, DOE, FAA, JHU, NIST, NSF, ONR, USAF ....
  - Extensive conference Proceedings published annually
    - Numerous papers relating to Reliability, Capability, POD, etc.

- o POD MODELING (ISU Center for NDE)
  - Several studies of POD modeling published (since 1985)
    - Mostly for ultrasonic and eddy-current inspection
    - Work in progress on adding POD capability to radiography models
  - Recent studies have included ROC-type analyses

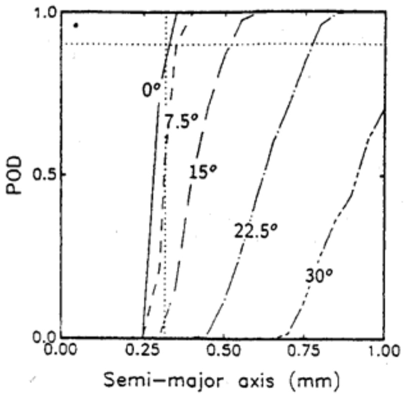
MODELING POD

- o MUST BE BASED ON PHYSICS OF THE INSPECTION METHOD
  - Model as many of the parameters contributing to detectability as possible
  - Establish the range of validity
  - Validate the model with experimental data
  - Rely on empirical data to cover parameters that cannot be modeled

- o PHYSICAL MODELING INVOLVES TWO TYPES OF ASSUMPTIONS:
  - The assumed flaw has the same properties as the natural flaw
  - Governing equations and boundary conditions are adequately satisfied

- o ADVANTAGES OF MODELING
  - Allows studying effects of varying each parameter independently
  - Allows evaluating cost-effectiveness trade-offs
  - Provides basis for adapting measured POD data to different parameters

MODELING POD

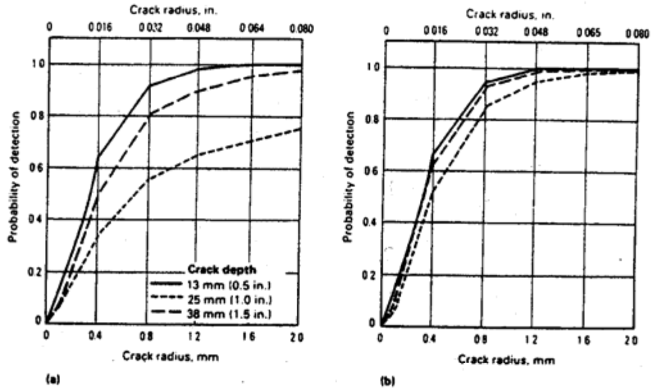


POD as a function of beam angle relative to mean normal to flaws

(T.A.Gray, POD Tutorial Workshop, 1993 ASNT Spring Conference)

- o **EXAMPLE: Predicting POD for silicate inclusions aligned by forging flow**
  - **Inclusions were assumed to be uniformly distributed within 5° of flow lines**
  - **10 MHz unfocused ultrasonic transducer, 30° refracted angle, 0.1" scan index**
  - **Concluded that single probe angle would not provide adequate POD**

MODELING POD



- o **EXAMPLE: Predicting the influence of scan index on POD**
  - **Ultrasonic detection of cracks at various depths below a cylindrical surface**
    a) axial index = 2.5 mm; circumferential index = 2.5 mm
    b) axial index = 5.0 mm; circumferential index = 1.3 mm
  - **Cylindrical surface focuses beam at approximately 25 mm depth**

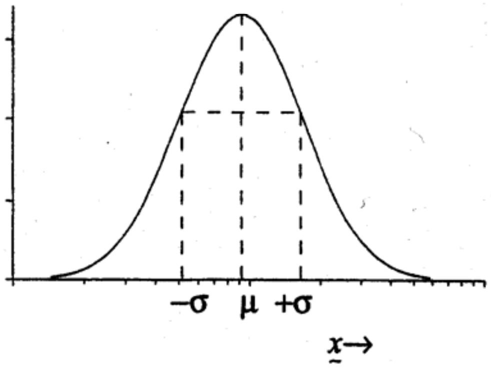
(J.N Gray et al., Metals Handbook, 1989)

WHAT?

STATISTICAL CONCEPTS

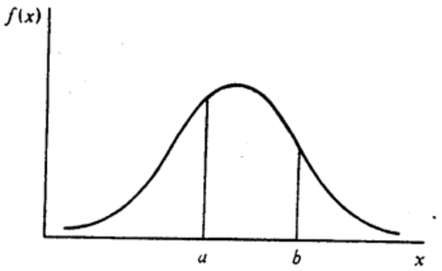
Probability Distributions
Populations and Sampling
Likelihood
Statistical Intervals
Statistical Independence
Applications to POD

PROBABILITY DISTRIBUTIONS



- o **Engineers are familiar with probability distributions of physical characteristics**
- o **Statisticians work with probability distributions of (estimates of) population parameters - i.e. with statistics**

PROBABILITY DENSITY FUNCTIONS



- A Probability Distribution of a continuous variable is usually called a Probability Density Function (PDF)
- The PDF describes the relationship between X and the relative frequency of X
- The area under the curve, bounded by the X-axis:
  - gives the probability of X assuming a value in the range between 'a' and 'b', when computed over the range between 'a' and 'b';
  - is equal to one, when computed over the range for which f(X) is defined

---

POPULATION PARAMETERS

- DESCRIBE THE UNDERLYING BEHAVIOR OF THE DATA
  - For example, <u>mean</u> and <u>standard deviation</u> (or <u>location</u> and <u>scale</u>)
- ARE FIXED
  - Behavior of the population is described by a distribution
  - The population parameters have no distributions
- ARE NOT FUNCTIONS OF THE OBSERVATIONS
  - Population parameters can be estimated from the observations
  - Such parameter estimates <u>are</u> functions of the data
- TELL HOW OFTEN TO EXPECT A GIVEN VALUE OF X

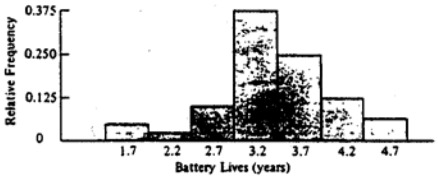
SAMPLING

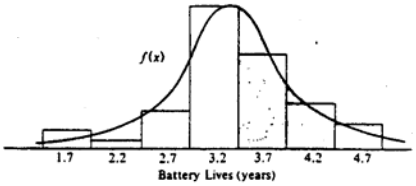
- **THE "POPULATION" CAN RARELY BE MEASURED**
  - **We estimate the nature of the distribution from the sample data**
    - **e.g. we test for conformance to selected standard distributions**
  - **We estimate population parameters based on the properties of a sample**
    - **e.g. we base our estimate of the population mean on the sample mean**
  - **Compare forecasting a presidential election from questioning 600 people**

- **SAMPLES MUST BE SENSIBLY STRUCTURED**
  - **In general there may be random and non-random elements in the sample**
  - **Compare public opinion polls: the sample must be "representative"**
    - **e.g. randomly sample within selected communities?**

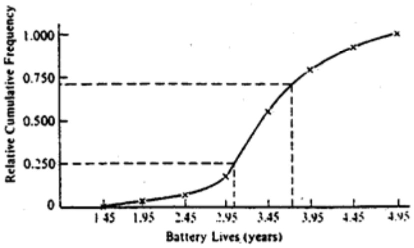
- **WE TRY TO EXPLOIT KNOWN PROPERTIES OF DISTRIBUTIONS**
  - **Confidence limits can easily be added to POD if data fit a normal distribution**
  - **Compare this with the "margin of error" in public opinion polls**
    - **Candidate A: 48.5% .. Candidate B: 51.5% .. Margin of Error: 4% ..?**

EMPIRICAL DISTRIBUTIONS



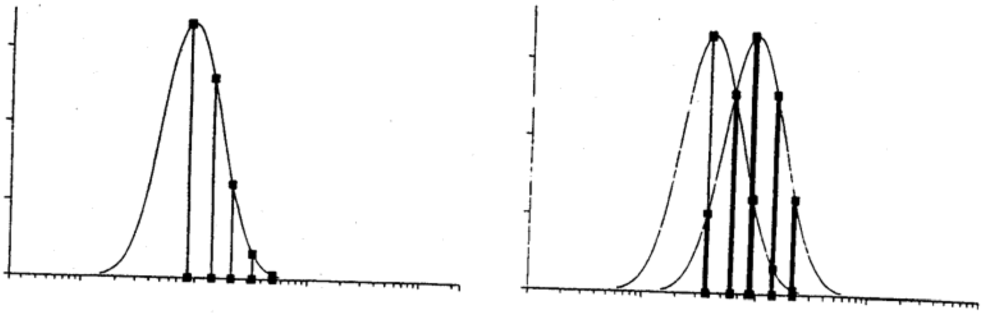


- **Example: fitting a normal distribution to empirical data on battery life**



- **Testing normality by plotting cumulative probability**
  - **Data should fit a straight line if truly "normal"**

LIKELIHOOD



- o Likelihood describes the behavior of the population parameter estimates, given the data
- o Likelihood has a maximum (for well-behaved situations)
- o Maximum Likelihood Estimators are normally distributed
  - The sample mean is a MLE

STATISTICAL INTERVALS

TABLE 5. Factors for Calculating One-Sided 95% Probability Limits for a Normal Distribution

Number of Given Observations	Factors for Calculating Prediction Limit to Contain All of k Future Observations with 95% Probability						Factors for Calculating Prediction Limit to Contain Mean of k = ∞ Future Observations with 95% Probability	Factors for Calculating Tolerance Limit to Contain at Least the Proportion p of the Population with 95% Probability			Factors for Calculating 95% Confidence Limit to Contain the Population Mean μ
	k							p			
n	1	2	5	10	20	∞		0.90	0.95	0.99	
4	2.63	3.40	4.47	5.28	6.06	4.21	1.66	4.16	5.15	7.04	1.18
5	2.34	2.95	3.79	4.42	5.03	3.79	1.35	3.41	4.20	5.74	0.95
6	2.18	2.72	3.43	3.97	4.49	3.58	1.16	3.01	3.71	5.06	0.82
7	2.08	2.57	3.22	3.70	4.17	3.45	1.04	2.76	3.40	4.64	0.73
8	2.01	2.47	3.07	3.52	3.95	3.37	0.95	2.58	3.19	4.35	0.67
9	1.96	2.40	2.97	3.38	3.79	3.32	0.88	2.45	3.03	4.14	0.62
10	1.92	2.35	2.89	3.28	3.67	3.28	0.82	2.36	2.91	3.98	0.58
11	1.89	2.30	2.82	3.21	3.58	3.26	0.77	2.28	2.82	3.85	0.55
12	1.87	2.27	2.78	3.14	3.50	3.24	0.73	2.21	2.74	3.75	0.52
15	1.82	2.20	2.67	3.01	3.34	3.21	0.64	2.07	2.57	3.52	0.45
20	1.77	2.13	2.57	2.89	3.19	3.19	0.55	1.93	2.40	3.30	0.39
25	1.74	2.09	2.52	2.82	3.11	3.20	0.48	1.84	2.29	3.16	0.34
30	1.73	2.07	2.48	2.78	3.05	3.21	0.43	1.78	2.22	3.06	0.31
40	1.71	2.04	2.44	2.72	2.99	3.24	0.38	1.70	2.13	2.94	0.27
60	1.69	2.01	2.40	2.67	2.92	3.30	0.31	1.61	2.02	2.81	0.22
∞	1.64	1.95	2.32	2.57	2.80	∞	0	1.28	1.64	2.33	0

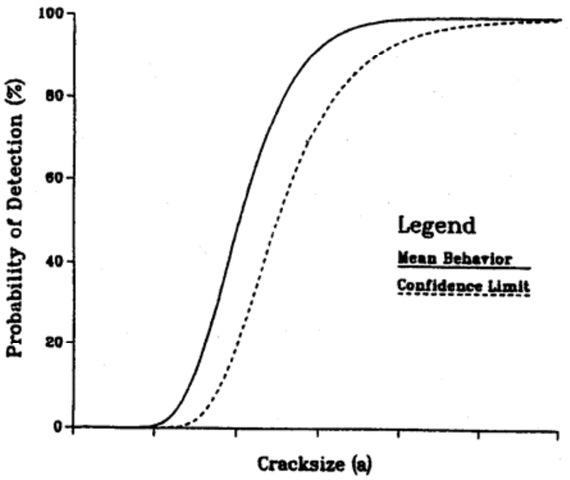
Lower 95% Limit is $\bar{y} - c'(n, 0.95)s$ where $c'(n, 0.95)$ is the appropriate tabulated value and $\bar{y}$ and s are the mean and the standard deviation of the given sample of size n.
Upper 95% Limit is $\bar{y} + c'(n, 0.95)s$.

- o Various statistical intervals have been evaluated for Normal Distributions
  - Each is based on the sample mean, sample variance, and sample size
- o Confidence Limits for bounding the population mean are used with POD
  - Use of a Lower One-Sided 95% Confidence Limit was selected by the USAF

STATISTICAL INDEPENDENCE

- o TWO EVENTS ARE STATISTICALLY INDEPENDENT IF P(A|B) = P(A)
  - i.e. the probability of A, given B, is equal to the probability of A
  - i.e. the outcome of event B has no influence on the outcome of event A
- o SIMPLE EXAMPLES
  - Independent events: outcome of tossing a coin twice in succession
  - Dependent events: drawing two cards <u>without replacement</u>
    - If first card was an ace, probability of another ace is 3/51, not 4/52
- o COMBINING PROBABILITIES:
  - Independent events: P(A and B) = P(A) * P(B)
  - Dependent events: P(A and B) = P(A) * P(B|A)
    or P(A and B) = P(B) * P(A|B)

APPLICATION TO POD CALCULATIONS



- o SO WHAT DOES ALL THIS MEAN FOR POD?
  - We can estimate POD(size) behavior from experimental data and provide a mean POD(size) relationship and associated confidence limits

APPLICATION TO POD CALCULATIONS

- o SAMPLING
  - We generate "representative samples" of cracks for use with surface NDE
  - We use detected flaws as "representative samples" for subsurface UT
  - Ideally, tests should be based on random selections from these samples

- o DISTRIBUTIONS
  - Log-normal or log-logistics distributions are implicit in USAF/UDRI software
  - Log-normal & Largest Extreme Value distributions are used for ultrasonics
  - Ideally, we should test whether a specific distribution is appropriate
    - We can test whether the sample conforms to some standard distributions
    - This does not guarantee that the population will also fit

---

APPLICATION TO POD CALCULATIONS

- o SAMPLE MEANS
  - Sample means are used as the "best estimate" of the population mean
  - If we had the luxury of rerunning the test with a new sample set, we should not be surprised if we got a different estimate of the mean
    - The new estimate may be above or below the previous one

- o CONFIDENCE LIMITS
  - A 95% (lower one-sided) confidence limit implies that we expect 19 out of 20 successive sample sets to give estimates of the population mean that are above the original confidence limit
    - Note: this means we should not be surprised if 1 out of 20 such estimates of the population mean fall below the original 95% confidence limit
    - i.e. our POD calculations yield estimates, not immutable absolute values

MULTIPLE INSPECTIONS

- **DOES USING TWO INSPECTIONS IMPROVE NET POD?**
  - **ONLY IF the inspections are statistically independent**

Probability of $\text{Miss}_{1\text{ or }2} = (1 - \text{POD}_1) * (1 - \text{POD}_2)$

e.g. if $\text{POD}_1 = 0.9$ and $\text{POD}_2 = 0.8$

$$\text{POD}_{1\text{ or }2} = (1 - \text{POM}_{1\text{ and }2}) = (1 - 0.1 * 0.2) = 0.98$$

- **THIS IMPROVEMENT IS RARELY JUSTIFIED!**
  - **Two inspections with the same method are likely to be highly dependent**
  - **Two inspections with totally different methods might be independent**
    - **Need to establish independence by measurement**
  - **If two inspections are fully dependent:**

$$\text{POD}_{1\text{ or }2} = (1 - \text{POM}_{1\text{ and }2}) = (1 - \text{POM}_1) = (1 - 0.1) = 0.9$$

- **GENERALLY BEST TO TRY IMPROVING THE BETTER INSPECTION!**

WHAT?

DETECTION THEORY APPLIED TO NDE

Defining "Detection"

NDE Methods in a Signal and Noise Context

Selected Elements of Signal Detection Theory

DEFINING "DETECTION"

- **NDE "DETECTION" IS RARELY DISCUSSED EXPLICITLY**
  - We probably understand it - but we don't define it
- **DICTIONARY DEFINITIONS** (Webster's New World Dictionary)
  - Detect: - to catch or discover, as in a misdeed
    - to discover or manage to perceive (something hidden or not easily noticed)
  - Detection: - a finding out (said especially of what tends to elude notice)
- **A RADAR DEFINITION** (D.K.Barton, Radar System Analysis)
  - Target Detection: the process by which the presence of a sought-after object, or target, is sensed in the presence of competing indications which arise from background radiation, undesired echoes, or noise generated in the receiver

DEFINING "DETECTION"

- o **WE NEED TO DEFINE "DETECTION" BEFORE WE MEASURE "POD"**
  - **Standard NDE handbooks offer little help**
  - **The Radar definition is close to what we need**

- o **LET'S REVIEW SOME NDE TECHNIQUES**
  - **Clarify our implicit definition of detection**
  - **Assess whether the Radar definition is suitable for NDE use**

INSPECT, TEST, OR EVALUATE?

- o **DISTINGUISH BETWEEN NDT, NDI, AND NDE?**
  - **NDT, NDI, and NDE are often used interchangeably**
  - **Nondestructive Evaluation can be used as a more comprehensive term**
    - **NDE comprises all NDT and NDI activities**
    - **NDE comprises detection, location, and characterization**

- o **DETECTION AND CHARACTERIZATION ARE DISTINCT PROCESSES**
  - **In what follows, POD will refer to detection only**
  - **Caution: this distinction is not always made**
    - **Example: "POD" is sometimes used to express a measurement of performance in detecting and classifying (e.g. sizing) defects**
    - **Better termed "Joint Probability of True Positive" (Joint P{TP}) or "Probability Of Detection and Correct Interpretation" (PODCI)**

DEFECT DETECTION OR DEFECT-FREE MATERIAL?

- SEPARATE BUT RELATED ISSUES
  - High POD alone does not guarantee defect-free material
- DEFECTS SUCCESSFULLY DETECTED
  - High POD suffices for selecting NDE techniques
  - High POD is a necessary but insufficient condition for high product life
- DEFECTS ESCAPING DETECTION
  - High product life requires a low probability that inspected material still contains any defects (compare Avioli's definition of Reliability)
    - Requires high POD <u>and</u> low probability of occurrence of defects
    - High POD may be less important if there is a very low probability of there being a defect in the uninspected material

---

MAGNETIC PARTICLE AND PENETRANT INSPECTIONS

- SENSOR (the eye)
  - Responds to indication color contrast (dye) or brightness (fluorescent)
  - Responds to indication shape and size (principally length)
- DISCRIMINATOR (the brain)
  - Records high-contrast and/or large indications - the "signal"
    - Classifies these as "flaws" or "defects" (i.e. above threshold)
  - Ignores low-contrast and/or small indications - the "noise"
    - Classifies these as "irrelevant" or "background" (i.e. below threshold)
- THRESHOLD (set by Specifications, Drawing Notes, etc.)
  - Typically defines maximum acceptable indication length

RADIOGRAPHIC INSPECTION

- o SENSOR (the eye)
  - Responds to film contrast and density
  - Responds to indication shape and size
- o DISCRIMINATOR (the brain)
  - Records high-contrast and/or large indications - the "signal"
    - Classifies these as "flaws" or "defects" (i.e. above threshold)
  - Ignores low-contrast and/or small indications - the "noise"
    - Classifies these as "irrelevant" or "background" (i.e. below threshold)
- o THRESHOLD (set by Specifications, Drawing Notes, etc.)
  - Typically defines maximum acceptable indication diameter (or length)

EDDY-CURRENT OR ULTRASONIC A-SCAN INSPECTION

- o SENSOR (electronic instrumentation)
  - Responds to indication amplitude
  - Responds to indication depth
- o DISCRIMINATOR (an electronic threshold)
  - Records large indications - the "signal"
    - Classifies these as "flaws" or "defects" (i.e. above threshold)
  - Ignores small indications - the "noise"
    - Classifies these as "irrelevant" or "background" (i.e. below threshold)
- o THRESHOLD (set by Specifications, Drawing Notes, etc.)
  - Typically defines a maximum acceptable indication amplitude
    - Threshold may be related to indication phase (ET) or depth (UT)

EFFECTS OF AUTOMATION

- o **SENSORS**
  - The eye is replaced by electronic or electro-optical devices
    - Example: machine vision for automated penetrant systems
- o **DISCRIMINATORS**
  - The brain is replaced by an automatic electronic device
    - Example: ultrasonic indications signalled by electronic alarm
  - The discrimination may be more quantitative, and computer-based
    - Example: penetrant indications may involve measured brightness and size
- o **THRESHOLDING**
  - Thresholds may be defined more precisely, and may be more complex
    - Example: "detection" of an indication may be determined by the value of a function of several measured variables

---

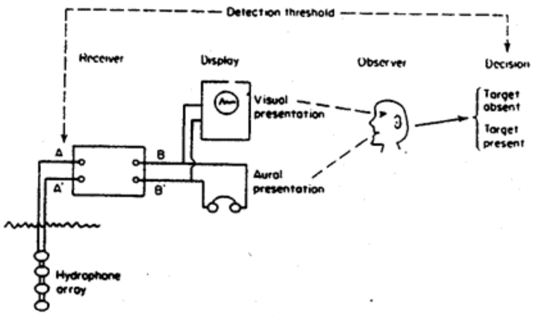
EFFECTS OF IMAGING

- o **SPATIAL CORRELATION**
  - Imaging presents data from successive interrogations more efficiently
    - Takes advantage of spatial correlation of information
  - Facilitates use of automatic pattern-recognition and decision-making
    - Example: a computer can decide if X-ray CT wall-thickness is acceptable
- o **IMAGING DOES NOT CHANGE THE BASIC DETECTION PROCESS**
  - POD is not necessarily improved by imaging
    - Example: an A-scan strip-chart contains the same data as a C-scan image
  - POD <u>may</u> be improved as a result of use a lower discrimination threshold
    - In effect, the threshold may be lowered on a local basis

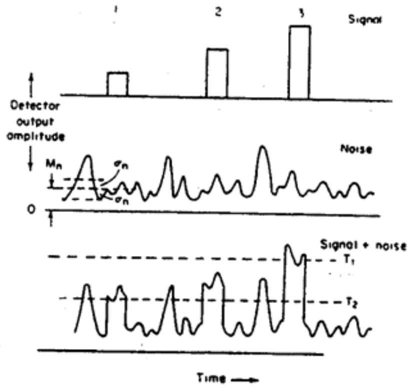
DETECTION AS A SIGNAL-TO-NOISE PROCESS

- **NDE TECHNIQUES ARE THRESHOLDED S:N PROCESSES**
  - **Easiest to see this for "electronic" techniques (e.g. ET, UT) and for automated inspection techniques (e.g. computed tomography)**
  - **Basically true for other techniques, too (e.g. MT, PT, RT)**
- **"THRESHOLD"**
  - **Represents a level of discrimination between "signal" and "noise"**
- **"SIGNAL"**
  - **Represents the response from the defects - the sought-after target**
- **"NOISE"**
  - **Competing signals from sources other than defects**

DETECTION AS A SIGNAL-TO-NOISE PROCESS



Elements of a receiving system



Signals with additive noise

- **SONAR DETECTION SYSTEM**
  - **Many similarities to ultrasonic NDE**
  - **Fundamental concepts are common to most NDE processes**

SOURCES OF NOISE

- o NOISE FROM THE MATERIAL UNDERGOING INSPECTION
  - Generic examples:
    - Surface roughness
    - Edge effects
  - Process-specific examples:
    - Local changes in conductivity (MT, ET)
    - Indications from surface pores in castings (PT)
    - Grain-boundary reflections (UT)
    - Scattering (RT)
- o NOISE FROM THE INSPECTION PROCESS ITSELF
  - Examples:
    - Electronic noise (ET, UT)
    - Grain structure in film (RT)

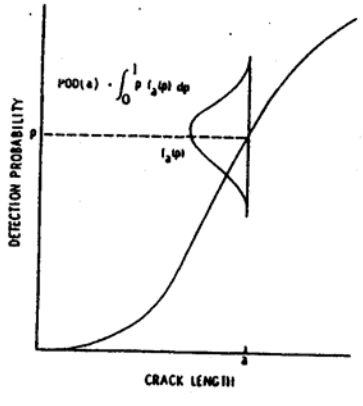
VARIATION IN THE DETECTION PROCESS

- o DEFECTS
  - Each defect has slightly different characteristics
    - Differences in size, shape, orientation, location, nature
- o MATERIAL
  - Each material sample has slightly different characteristics
    - Differences in conductivity, permeability, absorbtivity, etc.
    - Differences in shape, surface texture, etc.
- o INSPECTION SYSTEM
  - Each inspection process has slightly different characteristics
    - Differences in penetrant concentration, etc.
    - Differences in sensitivity, linearity, frequency response, etc.
  - Each inspector has slightly different capabilities

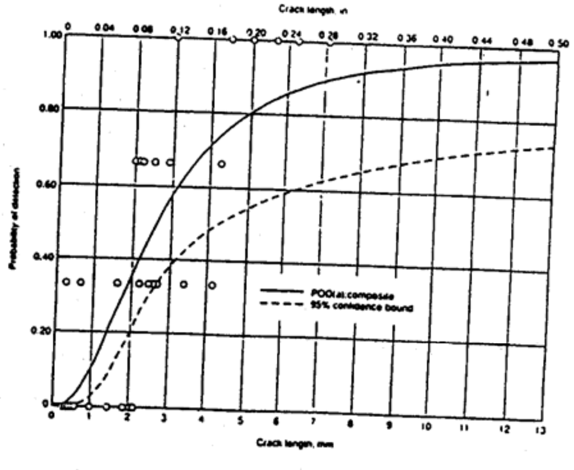
EFFECTS OF VARIATION

- o **NUMEROUS PARAMETERS CAN INFLUENCE DETECTION**
  - **Some have predictable (deterministic) effects**
  - **Many effects are not quantitatively predictable**
    - **Consequently, detection is a <u>probabilistic</u> phenomenon**
- o **IMPERFECTLY CONTROLLED INSPECTION PARAMETERS**
  - **Perfection is a worthy goal - but always out of reach!**
- o **SCATTER IN MEASURED DATA**
  - **Both signal and noise data incorporate the effects of variation**
    - **Measurements will yield <u>distributions</u> of values**

EFFECTS OF SCATTER ON POD

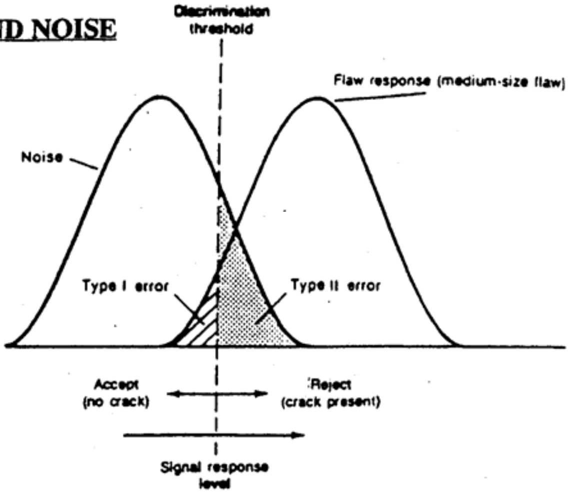


Probability density function of crack detection probabilities for cracks of a specific length



Probability of Detection data with lower one sided confidence bound representing the effects of sample size and scatter in the data

DISTRIBUTIONS OF SIGNAL AND NOISE



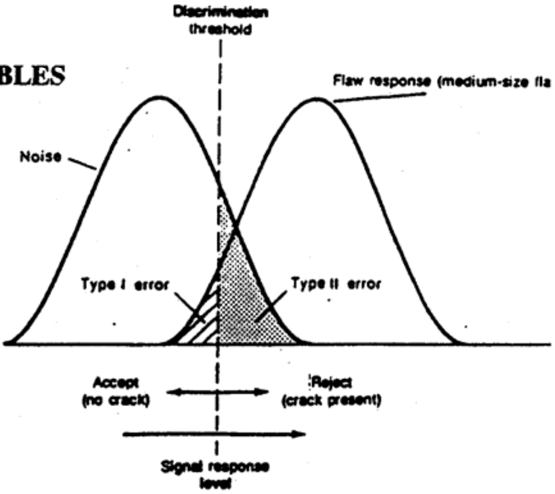
A signal-to-noise
view of detection

- o AREAS ENCLOSED RELATE TO THE ALTERNATIVE OUTCOMES

- Detection of a defect that is present	(POD)
- Non-detection of a defect that is present	(1 - POD)
- Apparent detection of a defect that is not present	(PFA)
- Non-detection of a defect that is not present	(1 - PFA)

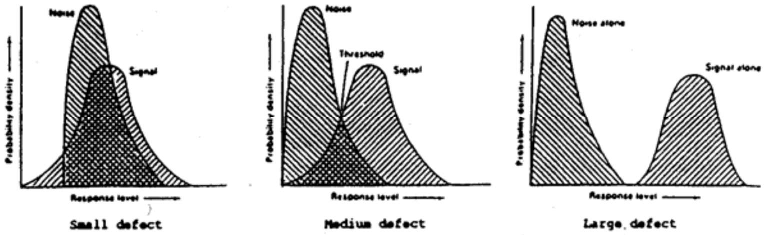
DISTRIBUTIONS OF SIGNAL AND NOISE

- o POD AND PFA ARE CO-VARIABLES
  - POD is not a function of PFA
  - **Both** vary with the threshold
    - A lower threshold increases POD and PFA



- o CHOICE OF THRESHOLD
  - Difficult to keep design and manufacturing engineers happy!
    - Reducing Type I errors represents improved product life
    - Increasing Type II errors represents increased manufacturing cost

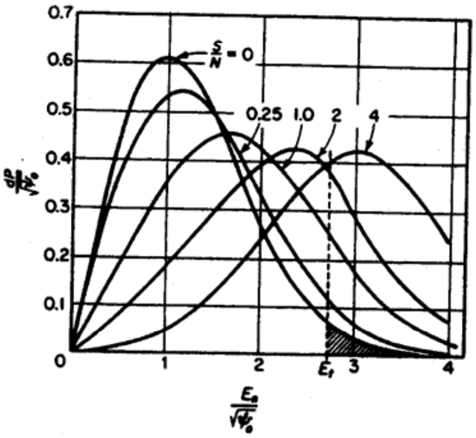
DISTRIBUTIONS OF SIGNAL AND NOISE



Small defect Medium defect Large defect

- **DEPENDENCE ON DEFECT SIZE**
  - **Larger defects generally give larger signals**
  - **For a given POD, the PFA will tend to decrease with increasing defect size**
  - **For a given PFA, the POD will tend to increase with increasing defect size**
  - **For a given defect "size", POD also depends on other parameters**
    - **Shape, orientation, location, nature**

DISTRIBUTIONS OF SIGNAL AND NOISE



- **AN ALTERNATIVE WAY OF LOOKING AT SIGNAL AND NOISE**
  - **Probability distributions of envelope of signal-plus-noise**
  - **Plotted as a function of signal-to-noise ratio (S/N)**

DISTRIBUTIONS OF SIGNAL AND NOISE

- A GATEWAY TO EXISTING SIGNAL DETECTION THEORY
  - Fifty years of theoretical development and practical application
  - NDE applications have been largely ignored
- STATISTICAL THEORIES OF SIGNAL DETECTION
  - Original applications were to radio communications and radar
    - Predominantly devoted to analyzing reception of pulsed signals
  - Based on statistical criteria for testing hypotheses and making decisions
    - Allow derivation of POD and PFA from noise and signal-plus-noise
    - Allow optimization of receiver/detector design
    - Provide rationale for choosing a specific detection threshold

---

DETECTION THEORY

- RESPONSE TO A BINARY STIMULUS
  - Two possible states of the stimulus
    - Noise alone (N)
    - Signal plus noise (SN)
  - Two possible detection decisions (responses)
    - Noise alone present
    - Signal plus noise present
  - Four possible outcomes
    - N true; choose N .... true negative ..... probability (1 - PFA)
    - N true; choose SN .... false positive ...... probability PFA
    - SN true; choose SN .... true positive ...... probability POD
    - SN true; choose N .... false negative .... probability (1 - POD)
- THRESHOLDED DETECTION
  - All observed values greater than the threshold are classed as SN

DETECTION THEORY

- o **OBSERVED PROBABILITIES**
  - Probabilities of decision outcomes are estimated from observed frequencies
  - These are conditional probabilities (conditional upon the stimuli)
    - PFA is the probability of an N response given an SN stimulus
    - POD is the probability of an SN response given an SN stimulus
- o **DECISION CRITERIA**
  - Rules to help make a choice between the two responses
    - Based on attaching relative importance to the four outcomes
  - Of many such criteria, two are of most interest
    - Bayes
    - Neyman-Pearson

DETECTION THEORY

- o **BAYES CRITERION**
  - Assumptions: Observer has information about the source prior to the test
    - Prior (or "a priori") probabilities (P_1, P_0) are known
    - A "cost" (C_{10}, C_{00}, C_{01}, C_{11}) can be associated with each outcome
    - These may be truly known or may be educated guesses
  - Goal: design the decision criterion to minimize average cost
  - Problem: it is often difficult to assign prior probabilities and costs
- o **NEYMAN-PEARSON CRITERION**
  - Assumptions: only the measured conditional probabilities are available
  - Goal: design a test to minimize PFA and maximize POD
  - Problem: these are usually conflicting objectives
  - Solution: set an upper limit to PFA, and maximize POD

DETECTION THEORY

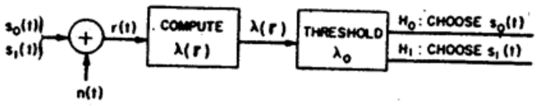
- **LIKELIHOOD RATIO TEST**
  - **Bayes and Neyman-Pearson criteria both lead to likelihood ratio tests**
    - **Likelihood Ratio =** $\dfrac{\textbf{Probability of choosing SN if SN is true}}{\textbf{Probability of choosing N if N is true}}$
  - **The decision depends on whether the likelihood ratio exceeds a threshold**
  - **The threshold used depends on the test**
    - **Bayes test: threshold involves the prior probabilities and costs**

      $$E_t = [P_0(C_{10} - C_{00})]/[P_1(C_{01} - C_{11}]$$

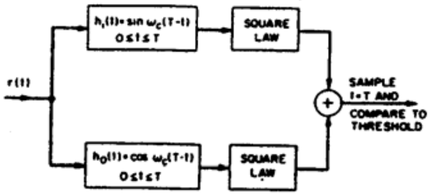
    - **Neyman-Pearson test: threshold chosen so that PFA = constrained value**

- **AN OPTIMUM RECEIVER**
  - **One that presents at its output the likelihood ratio for each receiver input**
    - **LR =** $\dfrac{\textbf{Probability that given amplitude represents SN if SN is true}}{\textbf{Probability that given amplitude represents N if N is true}}$

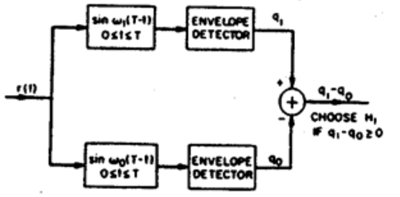
DETECTION THEORY



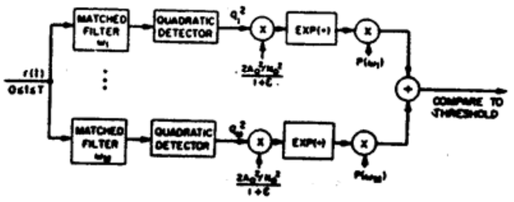
Binary communication signals



Signals of random phase



Signals of random amplitude and phase



Rayleigh fading signals of unknown frequency

- **EXAMPLES OF OPTIMUM RECEIVERS**
  - **Signals with known or random parameters**
  - **Signals corrupted by additive white Gaussian noise**

DETECTION THEORY

- ALTERNATIVE DECISION CRITERIA
  - "Optimal" performance can be defined in many ways
    - Maximize True Positive while restricting False Positive
    - Maximize expected value (minimize "costs") of decision
    - Maximize percentage of correct decisions
  - Individual inspectors may apply their own criteria
    - Give priority to avoiding false alarms, or to avoiding misses, etc.
- HUMAN FACTORS IN THE DECISION PROCESS
  - Inadequately trained inspectors use widely varying decision criteria
    - Human criteria tend to change with experience and/or training
      - e.g. both POD and PFA tend to decrease with time if SN is small
- AUTOMATED INSPECTION
  - Decision criteria may vary, but are applied consistently

DETECTION THEORY

- MEASUREMENT OF CONDITIONAL PROBABILITIES
  - Produce the stimulus and measure the response
  - Base estimates of probabilities of outcomes on observed frequencies
- COMMUNICATIONS, RADAR, SONAR: relatively easy to accomplish
  - Switch on the signal source(s), or tell the pilot where to fly
- NONDESTRUCTIVE EVALUATION: much more of a challenge
  - Stimuli are much more varied
  - True stimuli are natural, not manufactured
  - A "referee" technique is needed to identify the SN stimulus
    - For surface defects this might be optical microscopy
    - For subsurface defects no such technique exists

WHAT?

RELATIVE OPERATING CHARACTERISTIC

Psychophysical Applications

Empirical NDE Applications

Applications based on Signal Detection Theory

PSYCHOPHYSICAL APPLICATIONS 0F ROC's

- **TYPICAL HUMAN DISCRIMINATION CHALLENGES**
  - **How well do individuals respond to sensory stimuli?**
    - **Brightness, hue, loudness, pitch, taste, smell, etc.**
    - **Size, distance, direction, time, etc.**
  - **Is an unbiased measure of cognitive discrimination possible?**
    - **Psychological measurements are plagued by covert discrimination criteria**
      - **Expectations and motivations**
      - **Probabilities and utilities**
- **TWO METHODS HAVE BEEN FOUND TO MINIMIZE BIAS**
  - **Present two alternative stimuli, and ask which is A and which is B**
    OR
  - **Use single stimuli, and plot data in the ROC format**

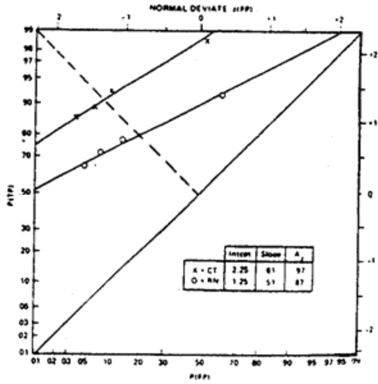
PSYCHOPHYSICAL APPLICATIONS 0F ROC's

- ANALYSIS OF HUMAN HEARING (1954-57)
  - An unexpected application of signal detection theory to human perception
  - Traditional view: a sound will always be perceived if it is big enough
  - New view: perception is based in part on the risk of false alarm
    - More sounds are perceived by people willing to accept more false alarms
- APPLICATIONS IN OTHER FIELDS (1960's and later)
  - Clinical Medicine (initial diagnosis)
  - Criminal Justice System (effect of court trials)
  - Library Science (information retrieval)
  - Power Generation industry (evaluation of inspectors)
  - Psychology (memory testing)
  - Radiography (evaluation of film readers)
  - Weather Forecasting

PSYCHOPHYSICAL APPLICATIONS 0F ROC's

- SOME CONCLUSIONS ABOUT DETECTION AND RECOGNITION
  - Measures of sensitivity that do not isolate effects of changes in the decision criterion ignore a substantial source of variation
  - Measurement of data in terms compatible with signal detection theory reduces test-to-test differences
  - ROC analyses help distinguish between causes of change
    - Threshold increase is the commonest cause of POD decreasing with time
      - May be due to differences between training and actual testing
      - The stimulus probability can affect the response probabilities
- RELEVANCE TO NDE
  - Likely to be greatest for processes with human sensor/receiver

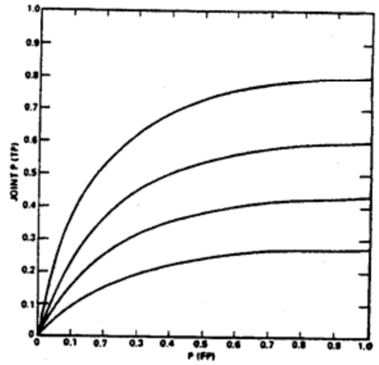
ADDITIONAL ROC EXAMPLES



Empirical ROC's from studies of imaging of brain lesions
CT: computed tomography
RN: radionuclide

- **BINORMAL ROC PLOTS WORK WELL FOR MANY DISTRIBUTIONS**
  - **Some non-normal distributions approximate to straight lines**
    - **e.g. exponential, rectangular, etc.**
  - **Many empirical ROC plots can be fitted by straight lines**
    - **These may be modeled by distributions with different variances**
    - **Slope equals ratio of standard deviations of noise and signal-plus-noise**

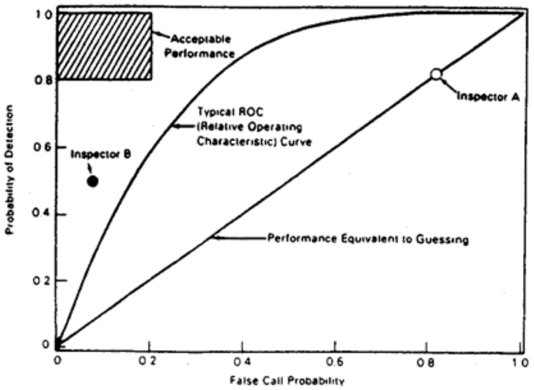
ADDITIONAL ROC EXAMPLES



Illustrative joint ROC's:
Probability of detection and correct classification

- **SIGNAL DETECTION AND CLASSIFICATION**
  - **Joint ROC approach**
    - **Extension of binary ROC to include classification (type, location, etc.)**
    - **Index of classification accuracy independent of detection criterion**
  - **Sequential binary response approach**
    - **Separate binary decisions about detection, type, location, size, etc.)**

NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS



An Operating Characteristic presentation of POD and PFA

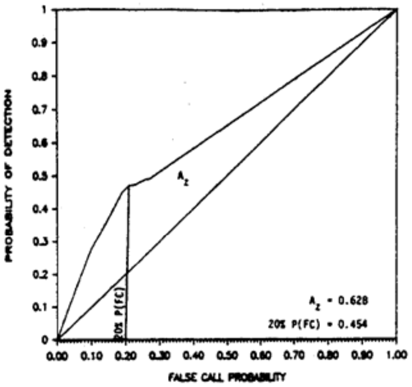
- **POD AND PFA FROM KNOWN-DEFECT SAMPLES**
  - **Diagonal line is equivalent to guessing (target strength = 0)**
  - **The ROC curve indicates the typical effect of changing reporting criteria**
  - **Desirable performance is within the marked box**
    - **Is Inspector A's performance better than Inspector B's?**

NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS

- **EVALUATION OF A GROUP OF INSPECTORS**
  - **Important to recognize that performance will reflect subtle changes**
    - **Inspection environment, fatigue, perceived management expectations ...**
  - **Performance of a group of inspectors should fit an ROC curve**
    - **Approach to the ideal "box" is limited by the sample(s) provided**
  - **Well-trained inspectors will try to use consistent reporting criteria**
    - **Performance will not wander along the ROC curve**
    - **Single-point evaluation is appropriate**
    - **Must use composite criteria (such as high POD <u>plus</u> low PFA)**

- **EVALUATION OF AN INSPECTION PROCEDURE**
  - **An ROC curve presents a more complete description than a single point**
    - **Can be used to determine proper reporting criteria**
  - **Allows comparison of inspection procedures**
    - **The best has highest ROC curve (or bounds the largest area)**

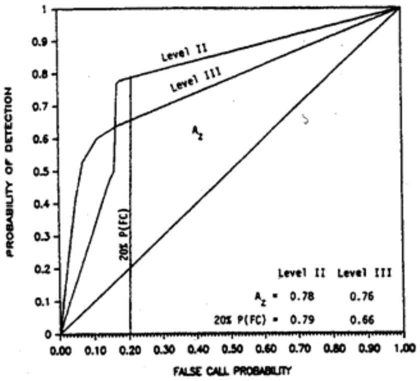
NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS



Average inspector performance during an IGSCC round-robin

- o **DETECTION AND IDENTIFICATION OF DEFECTS**
  - **12 technicians inspected samples of intergranular stress-corrosion cracks**
  - **Recorded detection and rated certainty of identification of "crack"**
    - **These are "Probability of Detection and Correct Identification" data**
  - **Area A_Z under ROC curve chosen as an evaluation parameter**

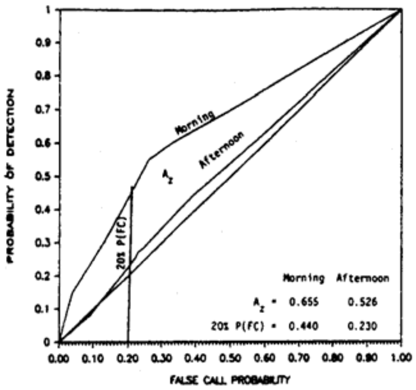
NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS



Comparison of Level II and Level III inspector performance

- o **CERTIFICATION NOT A GUARANTEE OF PERFORMANCE**
  - **PODCI & PFA data from the same IGSCC round-robin**
  - **Differences between Level II and Level III not statistically significant**
    - **Level III inspectors reported lack of recent experience**

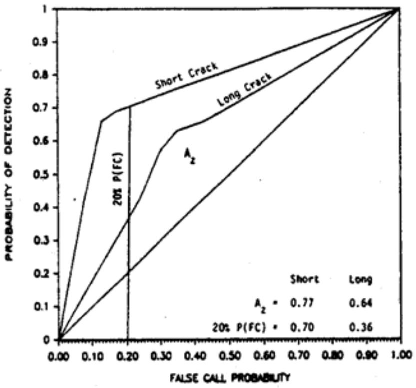
NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS



Comparison of performance on far-side short cracks

- **DETRIMENTAL INFLUENCE OF FATIGUE**
  - **Fatigue contributes to incorrect calls**
    - **Long hours (wearing protective clothing) and limited time off**

NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS



Comparison of performance on short and long near-side cracks

- **PODCI NOT ALWAYS CORRELATED WITH DEFECT SIZE**
  - **Detectability (POD) of long and short cracks was approximately equal**
  - **Technicians size predictions were better for short than for long cracks**
    - **PODCI better for short than long cracks**

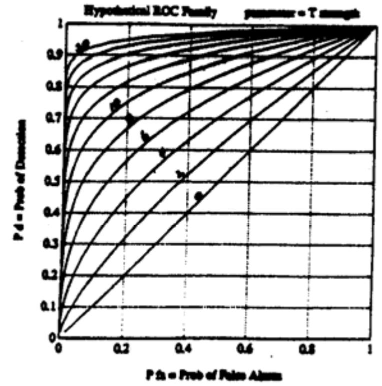
APPLYING SIGNAL DETECTION THEORY TO NDE

- o PARALLELS BETWEEN RADAR, SONAR, AND ULTRASONIC NDE
  - Clearly many similarities
    - Pulsed radio-frequency echo detection
    - Affected by beam pattern, target reflectivity & transmission medium
    - Thermal and non-thermal noise sources
  - Some important differences about ultrasonic NDE
    - Stationary targets
    - Noise usually dominated by time-invariant echoes from microstructure
- o TRANSFERABILITY OF RESULTS
  - Radar-derived results are of relevance, but must be used with caution
    - Offer qualitative insight
    - Not quantitatively applicable
  - Theories cover only simplified cases (i.e. they only approximate to reality)

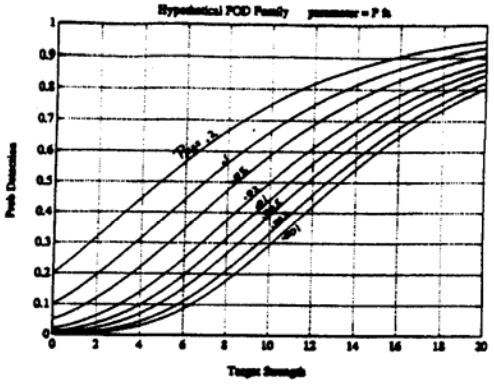
APPLYING SIGNAL DETECTION THEORY TO NDE

- o GENERAL APPLICATIONS TO NDE
  - Signal detection theory applicable to the extent that S & N are measurable
    - Appears possible for some NDE processes (e.g. eddy-current & ultrasonics)
    - Likely to be difficult for other processes (e.g. ultrasonics, penetrants)
  - Radar targets can be steered into the beam to measure "signal"
    - Equivalent may be possible for surface or synthetic NDE targets
    - Generally not possible with naturally-occurring subsurface targets
  - Operating Characteristics may be used even without signal distributions
    - Good format for combining "empirical" POD and PFA data
- o EXAMPLES (to follow)
  - Deriving dependence of POD on defect size from signal theory
  - Experimental data displayed in ROC and POD-vs-size formats
  - Comparing experimental data with optimized ROC predictions

AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS



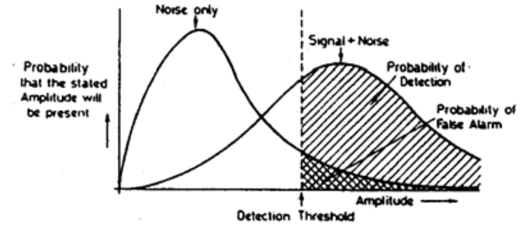
Hypothetical ROC curves



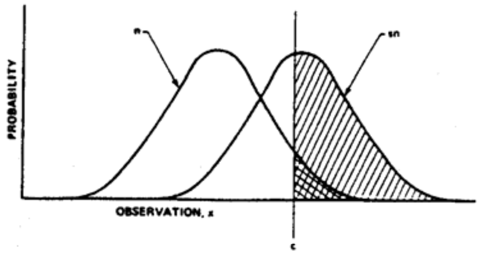
Equivalent plot of POD versus "size"

- **TAKE S/N AS A MEASURE OF "TARGET STRENGTH"**
  - **Assume only that "target strength" increases as defect size increases**
  - **Select a single Probability of False Alarm**
    - **For low PFA, the result is a typical POD-versus-size plot**

AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS



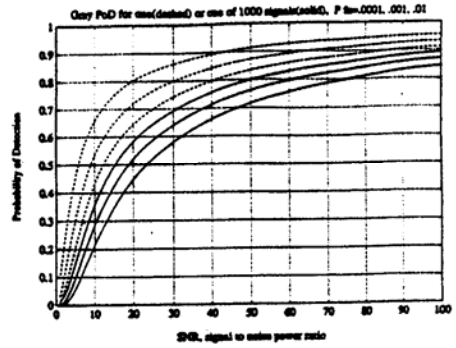
PDF broadened and shifted by signal



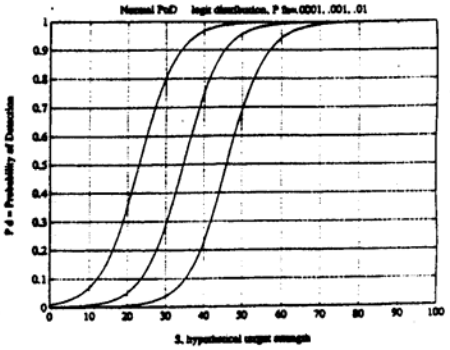
PDF shifted without broadening

- **DIFFERENCES IN NOISE AND SIGNAL-PLUS-NOISE DISTRIBUTIONS**
  - **Distributions depend on phenomena studied and on detection circuitry**
    - **Characterized by shape, mean (location), and variance (scale)**
  - **Central limit theorem: large samples often approach Gaussian distributions**

AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS



Typical "gray" POD curves

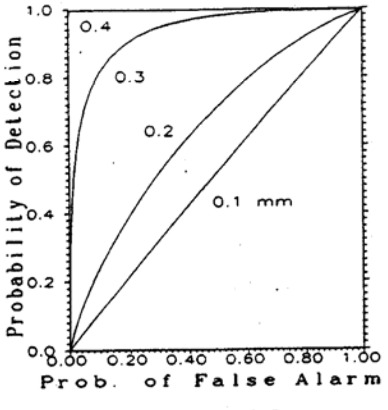


Typical "B&W" POD curves

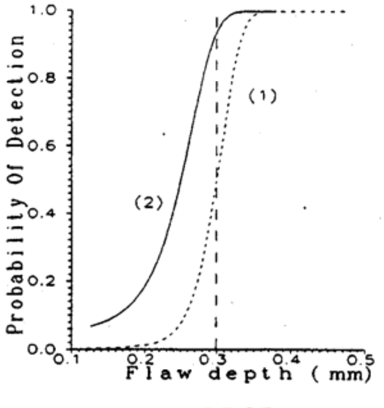
- **DEPENDENCE ON SHAPE OF S-plus-N DISTRIBUTION**
  - Presence of defect can affect probability density function (pdf) differently
  - "Gray" curves result when the pdf is broadened by presence of defect
  - "Black & white" curves result when pdf is shifted without broadening

---

AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS



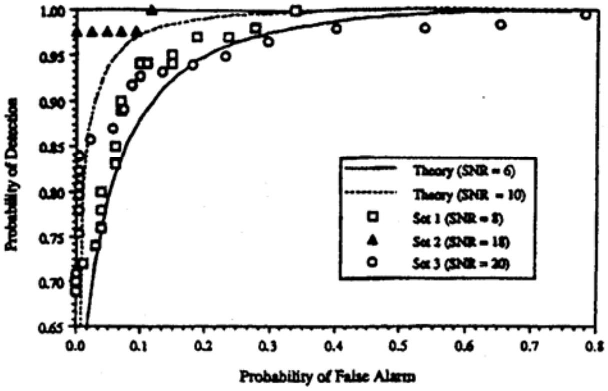
Predicted ROC



Predicted POD

- **PREDICTIONS OF POD AND PFA**
  - Eddy-current inspection of tight cracks
  - Theoretical predictions of defect signals and experimental noise data
  - ROC for various thresholds and four semi-elliptical crack sizes
  - POD for two specific threshold levels

AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS



Comparison of optimized ROC curves with three sets of simulated data

- o OPTIMIZED PREDICTIONS OF POD AND PFA
  - Neyman-Pearson likelihood-ratio criterion maximizes POD for given PFA
- o SIMULATED DEFECT AND NOISE DATA
  - Two large sets: simulated spherical defects plus real grain noise data
  - One small set (#2): Monte-Carlo simulation of signal and noise

---

SUMMARY

- o SIGNAL DETECTION THEORY
  - Useful tool for modelling detectability in communications and in NDE
    - Facilitates studying role of individual inspection parameters
    - Valid only to the extent that the model is valid
  - Extensive work on radar detection readily available
    - Can be applied to NDE with caution (usually not quantitatively useful)
    - Similar studies directly applicable to NDE remain to be done
- o RELATIVE OPERATING CHARACTERISTICS
  - Provide useful framework for displaying parametric POD/PFA relationship
  - Can be used with or without knowledge of signal-plus-noise distributions
  - A different way of looking at NDE capability, not a panacea
    - Problems exist in measuring S/N distributions as well as POD or PFA

WHAT?

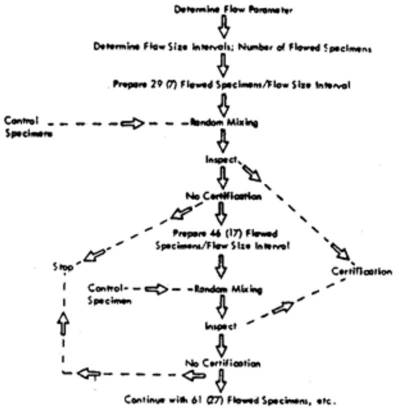
ASNT RECOMMENDED PRACTICE (1982)

Demonstration of NDE Reliability
(on Aircraft Production Parts)

ASNT RECOMMENDED PRACTICE

- **EFFORT INITIATED BY NEED TO COMPLY WITH MIL-STD-1530**
  - **Aircraft Structural Integrity Program, Airframe Requirements (1972)**
- **DEVELOPED BY ASNT AIRFRAME SUBCOMMITTEE**
  - **Submitted for publication in February 1976**
  - **Published in "Materials Evaluation", Vol. 40, pp. 922-932 (August 1982)**
- **PURPOSE**
  - **Define the limiting flaw size detectable with given probability & confidence**
  - **Guidelines for developing repeatable data for fracture mechanics applications**
- **SCOPE - Describes:**
  - **Specimen design and selection**
  - **Choice of operating parameters**
  - **Personnel and environmental variables**
  - **Data analysis**

ASNT RECOMMENDED PRACTICE



Procedure Flow Chart

TABLE B-1 Sample Sizes and Permitted Failures to Demonstrate a Minimum Success Rate at 90% with Indicated Confidence

No. Failures	Confidence Level 50	60	70	80	90	95	99	99.9
0	7	9	12	16	22	29	51	66
1	17	20	24	29	38	46	72	89
2	27	31	36	42	52	61	90	108
3	37	42	47	54	65	76	106	126
4	47	52	58	66	78	89	122	143
5	57	63	70	78	91	103	137	159
6	67	73	81	90	104	116	152	175
7	77	84	91	101	116	129	167	190
8	87	94	102	113	128	142	181	205
9	97	105	113	124	140	154	195	220
10	107	115	124	135	152	167	209	235
11	117	125	135	146	164	179	222	249
12	127	136	145	157	175	191	236	263
13	137	146	156	169	187	203	249	277
14	147	156	167	180	199	215	262	291
15	157	167	177	191	210	227	275	305
16	167	177	188	202	222	239	288	318
17	177	187	199	213	233	251	301	332
18	187	197	209	224	245	263	314	345
19	197	208	220	234	256	275	327	358
20	207	218	230	245	267	286	340	372

Sample Size Tables for 90% POD

- **For use with Hit/Miss data**
- **Orthodox binomial statistical analysis**
- **Requires relatively large sample size**
  - **At least 29 defects to establish 90% POD with 95% confidence at a single flaw size**

ASNT RECOMMENDED PRACTICE: OUTLINE OF METHOD

- **IDENTIFY FLAW SIZE, PROBABILITY & CONFIDENCE REQUIREMENTS**
- **DETERMINE APPROPRIATE NUMBER OF FLAW SPECIMENS**
  - **e.g. minimum of 7 for 90% POD @ 50% CL (or 29 for 90%/95%)**
- **MAKE APPROPRIATE SPECIMENS**
  - **Method focuses on cracks in plates and fastener holes**
  - **Randomly mix flawed and unflawed specimens**
- **INSPECT**
  - **Record number of flaws detected**
- **ANALYZE**
  - **If all 7 (90/50) or all 29 (90/95) flawed specimens have been detected in a flaw size interval and in all larger flaw size intervals, the procedure is considered certified at the (mean of the) smallest such flaw size interval**
  - **If no flaw size interval is certified, decide whether to continue certification by making additional specimens, and reinspecting**

WHAT?

POD AS A FUNCTION OF FLAW SIZE

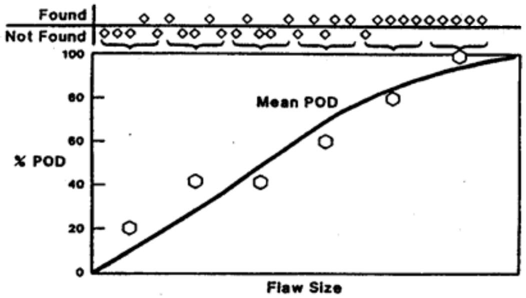
Data Cumulation Techniques

Regression Techniques

Model-based Regression

POD from distributions of Signal and Noise

DATA CUMULATION: RANGE-INTERVAL METHOD



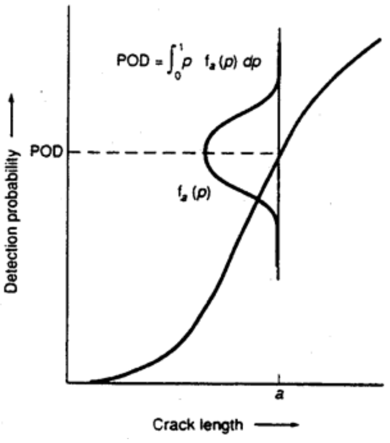
- **Intuitive (or physics-based) expectation that POD will depend on flaw size**
  - **Data "cumulation" methods tried to achieve this by grouping and averaging dat**
- **The Range-Interval method (RIM) provided a conceptually simple approach**
  - **Proportion of flaws detected in each interval was equated to POD**
  - **Data typically showed erratic fluctuations superimposed on expected behavior**
- **Now replaced by regression techniques**

DATA CUMULATION: OTHER TECHNIQUES

- Various other techniques were applied to smooth the POD-versus-size plots
- "Overlapping 60-point method"
  - POD was calculated for largest 60 flaws, and plotted at largest flaw size in interval
  - Next interval included the first 60 flaws below the median of first interval
  - POD was calculated for that interval, and plotted at largest flaw size in interval
  - Averaging was continued until smallest flaws were included
- "Optimized probability method"
  - Ordered POD data were arranged in intervals of successively increasing size range
  - POD was calculated for the largest size range
  - These data were combined with next smallest size range, and POD was recalculated
  - This process was repeated, adding more and more size ranges
  - The largest POD was plotted at the largest flaw size in the composite range
  - The process was repeated, but starting from the second largest size range
  - Etc!
- These techniques have been replaced by regression techniques

---

REGRESSION TECHNIQUES



Distribution of detection probabilities for cracks of fixed length

- Regression techniques provide a means to fit a smooth curve to POD data
  - Assume each defect of size a has detection probability p
  - Let probability density function of detection probabilities = $f_a(p)$
  - Conditional probability of a random defect being detected = $p*f_a(p)*dp$
- POD for crack of size a = integral of this expression over p
  - This POD function is the curve through the averages of individual pdf's

REGRESSION TECHNIQUES

- **Many alternative expressions for this POD function were evaluated**
  - **Evaluation was conducted under US Air Force auspices (around 1980)**
  - **The log-logistics ("log-odds") model was the best fit to the available data**

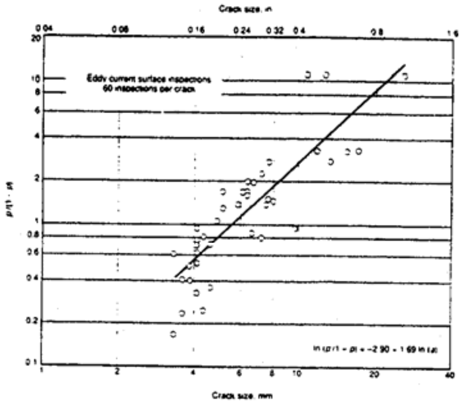
$$\mathbf{POD(a)} = \frac{[A + B^{*}\ln(a)]}{1 + \exp[A + B^{*}\ln(a)]}$$

  - **The log-logistics model has been used principally with hit/miss inspections**

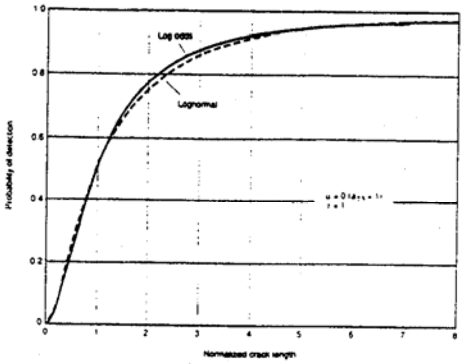
- **Linear regression**
  - **The curve-fitting described above is equivalent to:**
    - **Transforming the data into new coordinates showing a linear relation**
    - **Using regression techniques to fit a straight line to the transformed data**
    - **Transforming the line back into a smooth curve in the original coordinates**

- **Advantages of regression techniques**
  - **Produce smooth monotonically increasing curves**
  - **Make good use of relatively small data sets**

REGRESSION TECHNIQUES



Example of linear relation between log-odds and log crack size



Comparison of log-odds and cumulative log-normal models

- o **Cumulative log-normal model gives similar results to log-logistics**
  - **Has been used principally with response-versus-size ("A-HAT vs A") data**
- o **US Air Force sponsored development of POD software based on regression**
  - **Software written at University of Dayton Research Institute (UDRI)**

UDRI's "PF" and "AHAT" PROGRAMS

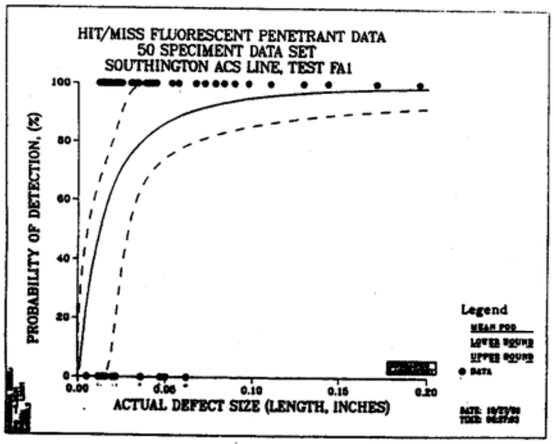
- o **"PF"**
  - **For use with pass/fail (hit/miss) data recording**
    - **Qualitative information about the presence or absence of a flaw**
  - **Major application: "manual"/"eyeball" inspection techniques**
    - **Principal usage: POD for fluorescent penetrant inspections**
- o **"AHAT"**
  - **For use with "A-HAT versus A" (response versus size) data recording**
    - **Quantitative information about the apparent size of detected flaws**
  - **Major application: "electronic" inspection systems**
    - **Principal usage: POD for eddy-current inspections**

UDRI's "PF" and "AHAT" PROGRAMS

- o DESCRIPTIONS OF USAF/UDRI METHODOLOGY
  - Metals Handbook, 9th edition, Volume 17 (ASM International, 1989)
    - "NDE Reliability Analysis", by Alan P. Berens, pp. 689-701
  - MIL-STD-1823 (pending)
    - "Non-destructive Evaluation System Reliability Assessment"

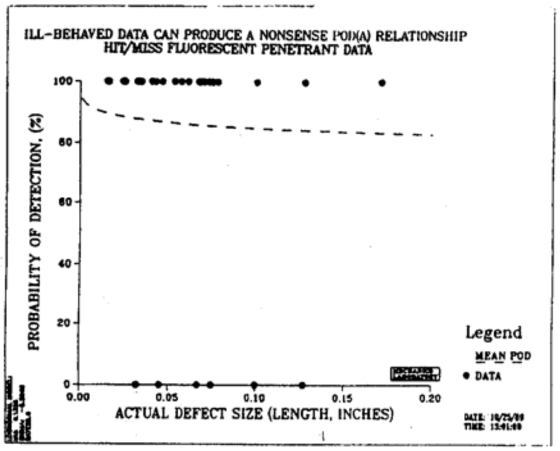
- o ENQUIRIES ...... (POD software and MIL-STD-1823)
  - SHARON VUKELICH (Telephone: 513-255-9594, Ext. 4179)
    ASC/ENFP, Flight Systems Engineering Division
    Wright-Patterson AFB, OH 45433-7809

ANALYSIS OF HIT/MISS DATA



- o The method works well for "a well-designed experiment of sufficient sample size for which the log-odds model is a reasonable representation of the POD(a) function" (A.P.Berens, Metals Handbook, 9th Ed., Vol.17, p.696)

ANALYSIS OF HIT/MISS DATA



- o The method will fail, or can produce apparently nonsensical results, if:
  - there is no overlap in the flaw size ranges of the hits and misses
  - there are too few flaws in the region of rapid increase of the POD function
  - the test plan is poorly designed
  - the inspection process is not in control

SIGNAL RESPONSE ("A-HAT vs.A") ANALYSIS: GENERAL APPLICATION

- o GOAL
  - Find a coordinate system giving linear dependence of response on flaw size
    - In general, this could take many forms
    - In practice, a log-log transform has been found to work well for most data
  - Fit a straight line to the transformed data
  - Determine the distribution of residuals about this line
  - Use the parameters of the fitted line and of the distribution to estimate POD
  - Revert to the original coordinate system, and plot POD versus size
- o RESPONSE
  - May be a simple measured parameter, such as peak response (e.g. mV)
  - May be a more complicated function of several parameters
    - For example, indication amplitude <u>and</u> apparent flaw area
    - Allows effective use of imaging (C-scan) techniques

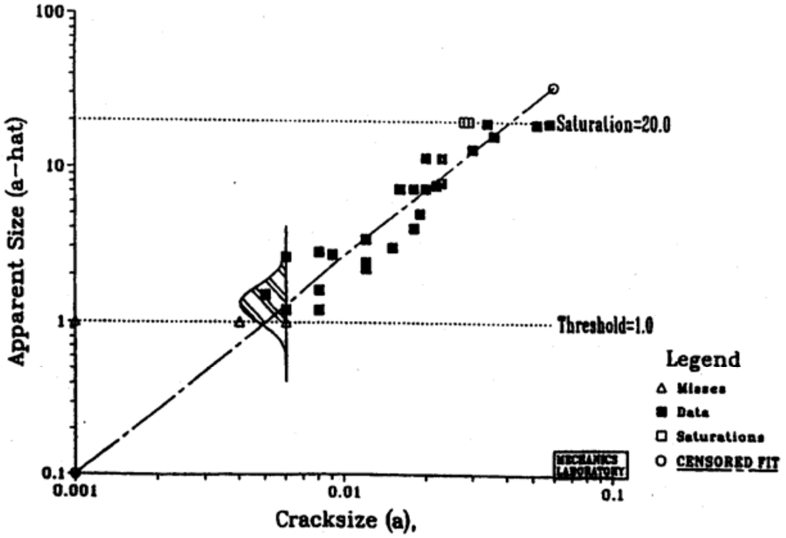
SIGNAL RESPONSE ("A-HAT vs.A") ANALYSIS: TYPICAL APPLICATION

- **STANDARD TRANSFORMED COORDINATES**
  - **Data are plotted in form ln(response) vs. ln(flaw size)**
    i.e. ln(A-HAT) vs. ln(A)
  - **Simple linear regression allows estimation of POD(A) and confidence limits**
    - **Note: the least squares regression is a maximum likelihood estimation**

- **RESPONSE ("A-HAT") VALUES**
  - **Assumed to have a normal distribution for a flaw of size A**

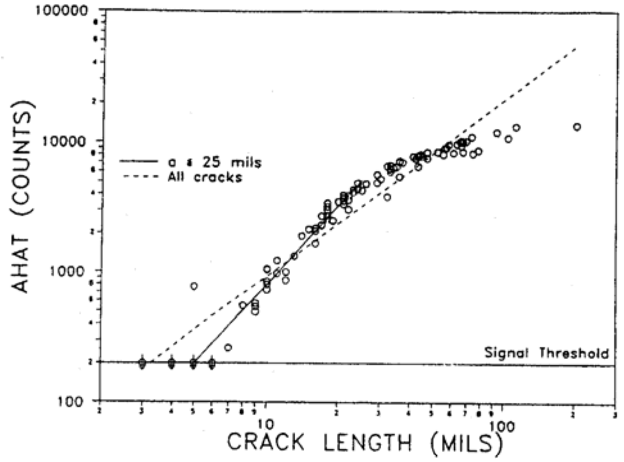
- **THRESHOLDING & SATURATION**
  - **In the UDRI program, the likelihood function is partitioned into 3 regions:**
    - **1) A-HAT values falling below the recording signal threshold**
    - **2) A-HAT values recorded and assumed to be proportional to A**
    - **3) A-HAT values falling above a saturation level**
  - **The outputs depend on the threshold and saturation levels used**

SIGNAL RESPONSE ("A-HAT vs.A") ANALYSIS



- **Example of A-HAT vs A data** **(from the MIL-STD for USAF NDE System Reliability Assessment)**
  - **Shaded region represents the POD for flaws of size 0.006**

SIGNAL RESPONSE ("A-HAT vs. A") ANALYSIS



- o Successful use of A-HAT vs A depends on similar criteria to those for Hit/Miss
- o Data must also conform to requirements implicit in linear regression
  - Relationship must be reasonably linear, with uniform normal residuals
  - Caution: these requirements are not flagged in the A-HAT software

---

MODEL-BASED REGRESSION

- o REGRESSION WORKS WELL WITH "WELL-BEHAVED" DATA
  - Requires a large data set with good linearity
  - Widely scattered data may produce implausible regression lines
    - This typifies many sets of real-defect ultrasonic data!
- o A PHYSICAL MODEL CAN HELP GUIDE THE REGRESSION
  - Can produce more plausible regression lines for widely scattered data
  - GE applied model-based regression to ultrasonic inspection
    - Developed contemporaneously with UDRI programs (early 1980's)
    - Invoked simple linear area-amplitude model (i.e. zero intercept)

MODEL-BASED REGRESSION

- o INITIAL APPROACH WAS NOT VIEWED AS REGRESSION
  - Based on use of flat-bottomed hole as flaw model
  - Inspection was calibrated using FBH's
  - Naturally-occurring defects were modeled as FBH's
  - Defect size predicted from model was compared with metallographic size
  - Ratio was called "EFFECTIVE REFLECTIVITY" (or R_e)

- o R_e IS A STATISTIC DESCRIBING DEVIATIONS FROM THE MODEL
  - Includes effects of defect size, shape, orientation, & character
  - Usually found to be approximately log-normally distributed
  - Properties of log-normal distribution were used to predict POD
  - Method became known as the Effective Reflectivity approach

---

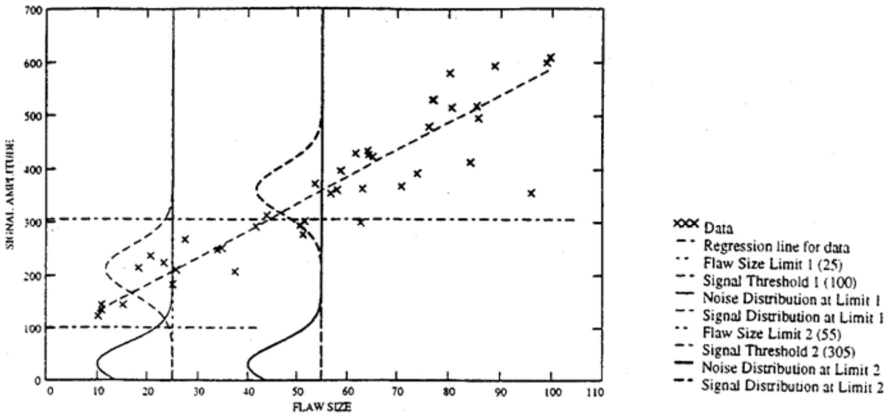
MODEL-BASED REGRESSION

- o R_e APPROACH IS EQUALLY WELL VIEWED AS LINEAR REGRESSION
  - The chosen flaw model preselects the intercept for the regression line
  - The simple FBH model gives zero response for zero size
  - Alternative models would have different effects

- o ADVANTAGES
  - Closely related to the standard "A-HAT versus A" approach
  - POD estimates are based on natural flaw properties
  - Provides plausible first-order estimates based on sparse data

- o MODIFIED APPROACH
  - Original version was relatively easy to apply
    - Results were potentially biased by limiting attention to detected defects
  - Modified version compensates for bias through maximum-likelihood process
    - Significantly more difficult to apply!

DISTRIBUTIONS OF SIGNAL AND NOISE



- o **TAKING NOISE INTO ACCOUNT**
  - **Flaw signals: the distribution mean increases with increasing flaw size**
  - **Noise signals: the mean is independent of flaw size**
  - **PFA decreases as signal detection threshold is raised**
  - **For a given flaw size, POD decreases as signal threshold is raised**

DISTRIBUTIONS OF SIGNAL AND NOISE

- o **ENGINE TITANIUM CONSORTIUM**
  - **Developing a new POD methodology for NDE applications**
    - **Based on Detection Theory**
    - **Based on physical model of noise and flaw response signals**
  - **Initially targeting ultrasonic inspection**
    - **Methodology should work for other quantitative-response data**
- o **METHODOLOGY WILL BE PUBLISHED BY FAA ON COMPLETION**
  - **Has been demonstrated for simple flaw types**
  - **Further development in progress**
- o **FOR FURTHER INFORMATION ....**
  - **Contact LISA BRASCHE, Associate Director (Telephone: 515-294-4999)**
    FAA Center for Aviation Systems Reliability, 185 ASCII, Ames, IA 50011

HOW?

DATA COLLECTION CONDITIONS

General Planning Guidelines

Capability or Reliability?

The "Known Flaw" Sample

GENERAL PLANNING GUIDELINES

- **POD IS NOT A PANACEA**
  - **Quantifying capability in terms of POD is an attractive concept**
  - **But ..... it is not necessarily easy to accomplish**
    ..... it is not always necessary

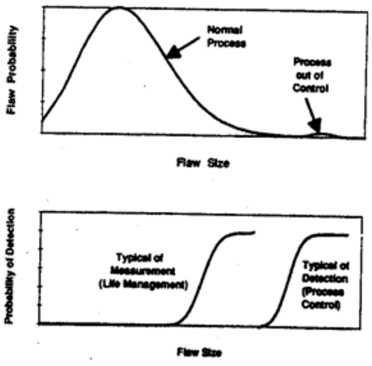
- **APPLICATIONS MUST BE JUDICIOUSLY SELECTED**
  - **Establish need**
    - **Is conventional Process Control inspection adequate?**
    - **Is accurate quantification truly needed?**
  - **Establish practicality**
    - **What parameters contribute significantly to POD?**
    - **Can all of these parameters be controlled or measured?**
  - **Establish timing and cost**
    - **Are available funds and time compatible with data acquisition?**

PROCESS CONTROL OR LIFE MANAGEMENT?

- o PSEUDO-QUANTITATIVE DATA FOR "PROCESS CONTROL"
  - Emphasis on consistency: independent of when or where conducted
    - Calibrate from artificial defects: reproducible targets for each site
  - Sensitivity usually set to detect <u>unexpected</u> (abnormal) defects
    - Primary goal is to sense loss of control in prior processes
- o QUANTITATIVE DATA FOR PREDICTING PRODUCT LIFE
  - Continue emphasis on consistency
    - Standardized inspection conditions
    - Calibrate from artificial defects ... but probably at increased sensitivity
  - Must have high proven POD for natural flaws that might limit product life
    - Inspection goals may include detection of some "normal" defects
      e.g. possible defects arising from intrinsic material non-uniformities

---

PROCESS CONTROL OR LIFE MANAGEMENT?



- o TYPICAL EFFECT OF CHANGING INSPECTION GOALS
  - Process Control inspections detect signs of loss of control
    - Sensitivity is often set empirically to detect the "maverick" defects
  - Life Management inspections may also need to detect smaller flaws
    - Sensitivity is set to achieve the required POD for specific size flaws

FACTORS INFLUENCING DETECTABILITY

- DEFECT PROPERTIES
  - Size, shape, orientation, location, character, etc.
- MATERIAL PROPERTIES
  - Composition, configuration, surface texture, grain size, cleanliness, etc
  - Absorptivity, attenuation, conductivity, permeability, porosity, etc.
- INSPECTION MATERIALS AND CONDITIONS
  - Composition, concentration, time, temperature, etc.
- INSTRUMENTATION
  - Frequency response, linearity, stability, sensitivity, etc
- SENSORS/TRANSDUCERS
  - Frequency response, size, shape, field pattern, film type, etc.
- INSPECTION PROCEDURE
  - Scan speed and index, wave mode, frequency, coupling, calibration, etc.
- HUMAN FACTORS
  - Skill, attentiveness, visual acuity, training, motivation, etc

PRACTICALITY

- VARIABLES
  - Significant sources of variability must be controlled or measured
    - Experience or preliminary measurement help to identify major factors
    - e.g. effective automation greatly reduces influence of human factors
  - Try to control as many sources of variability as possible
    - Write specifications for instruments, sensors, inspection parameters, etc.
  - Some parameters may be difficult to control or to measure
    - Are conditions identical at all sites where inspection is conducted?
    - If not, can the effect of each set of conditions be evaluated?
- IDENTIFY THE GOAL OF THE POD PROGRAM
  - Best attainable POD? -- or POD with real-world limitations?
  - Capability or Reliability?

CONCEPTS OF CAPABILITY AND RELIABILITY

- DICTIONARY DEFINITIONS
  - CAPABLE: "having capacity or ability; able to do things well; skilled; competent; efficient; able"
  - RELIABLE: "can be relied on; dependable; trustworthy"

- INDUSTRIAL QUALITY DEFINITION (Juran & Gryna, Quality Planning & Analysis)
  - RELIABILITY: "the probability of a product performing without failure a specified function under given conditions for a specified period of time"

CONCEPTS OF CAPABILITY AND RELIABILITY

- NDE DEFINITIONS OF RELIABILITY
  - "A quantitative measure of the efficiency of (an NDE) procedure in finding flaws of a specific type and size"

    (Packman et al., Metals Handbook, 8th edition)
  - "The probability that no defect which threatens loss of structural integrity exists in a component given that NDE says no such defect exists"

    (Avioli, 1993 ASNT Spring Conference)

- AND A COMMENT:
  - "NDE Reliability depends on POD,
    False Alarm Rate (FAR),
    the Probability of a critical flaw P(F),
    knowing where to look and knowing where you can't look"

    (Avioli, 1993 ASNT Spring Conference)

CONCEPTS OF CAPABILITY AND RELIABILITY

- **FURTHER SUGGESTED DISTINCTIONS:** (Haines, Sturges, Abernethy, Thompson)
  - **RELIABILITY:**
    - Composite effect of the following sources of variation
  - **CAPABILITY:** Defect detectability determined by physics
    - Detection of defects of the same nominal size
  - **REPEATABILITY:** Changes in performance with time
    - Reinspection by the same inspector or inspection system
  - **REPRODUCIBILITY:** Differences in nominally identical systems
    - Reinspection after process changes
  - **VARIABILITY:** Effects of human factors
    - Reinspection by another inspector or inspection system

CONCEPTS OF CAPABILITY AND RELIABILITY

- **CAPABILITY OR RELIABILITY?**
  - Different concepts - but often used interchangeably!
    - Easy to confuse - both are usually expressed in terms of POD
    - Packman's definition works for both
    - Avioli's definition introduces quite different concepts
  - Inspection processes involve numerous variables
    - Many of these affect Capability and Reliability differently

- **THERE IS NEED FOR AGREEMENT ON TERMINOLOGY**
  - Requests for POD analyses must be carefully defined
    - What is the real need?
    - Can the real need be met?

CAPABILITY OR RELIABILITY?

- o WHAT GETS MEASURED?
  - Ideally, "Reliability" (as defined above by Haines et al.)
    - Often, only a subset of conditions is considered
    - For example, experience may show that some factors are negligible ….. …… or costs for a full study may be prohibitive
  - For some purposes, "Capability" is adequate
    - For example, comparison of two penetrants …..?

- o HOW ARE THEY MEASURED?
  - The proportion of "known" flaws detected is measured
    - False alarm rates may also be needed
  - What is the source of those "known flaws"?

THE "KNOWN FLAW" SAMPLE

- o REPRESENTATIVENESS
  - Must realistically represent the population of natural defects

- o USE NATURAL DEFECTS?
  - Good manufacturing and design limit the number occurring
  - Destructive characterization may be needed .. ending their usefulness

- o USE SYNTHETIC DEFECTS?
  - Simple defects are relatively easy to make (notches, FBH's, etc.)
    - Large numbers and controllable properties
    - Detectability is likely to differ from that of natural defects
  - Realistic simulations of natural defects are difficult to make
    - Detectability may still differ from natural defects?
  - Use of synthetic defects results in POD for synthetic defects!
    - Need to compare detectability of natural and synthetic defects

THE "KNOWN FLAW" SAMPLE

- o **HOW LARGE SHOULD THE FLAW SAMPLE BE?**
  - **Depends on the method of analysis**
- o **ASNT Method: POD for a single flaw size**
  - **A sample as small as 7 may establish 90% POD with 50% confidence**
    - **Sample size increases with increasing confidence and if misses occur**
    - **e.g. 46 flaws for 90% POD at 95% confidence, with 1 nondetection**
  - **POD is established for the mean flaw size in the sample**
  - **Sample should contain at least as many specimens unflawed as flawed**
- o **USAF/UDRI Methods: POD for a range of flaw sizes**
  - **MIL-STD-1823 recommends:**
    - **a minimum of 60 flaws for hit/miss ("PF") data**
    - **a minimum of 40 for quantitative response ("A-HAT") data**
    - **at least 3 times as many inspection sites unflawed as flawed (for PFA)**

DATA ACQUISITION

- o **CONDITIONS MUST MATCH THE FINAL APPLICATION**
  - **Detectability depends on numerous factors**
  - **Must use appropriate conditions for measurement of POD database**
  - **The effect of changes can be predicted qualitatively at best**
    - **Example: Increasing an ultrasonic scan index will decrease POD**
      ..... but we generally don't know by how much!
- o **HUMAN FACTORS ARE A PARTICULAR PROBLEM**
  - **It is difficult to avoid changes due to awareness of the measurement**
- o **RECORD ALL DATA PERTINENT TO DETECTABILITY**
  - **The conditions of POD measurement establish range of validity of results**

HOW?

DATA ANALYSIS

Options

Examples

DATA ANALYSIS AND PRESENTATION

- **OPTIONS FOR "HIT/MISS" DATA: Methods and typical outputs**
  - **ASNT method:**
    - **A statement of POD, confidence and flaw size**
  - **USAF "PF" method:**
    - **A graph of POD versus flaw size at 50% and/or 95% confidence**
  - **Empirical ROC method:**
    - **A graph of POD versus PFA (sometimes distinguishing flaw sizes)**
- **OPTIONS FOR "QUANTITATIVE RESPONSE" DATA:**
  - **USAF "A-HAT" method**
    - **A graph of POD versus flaw size at 50% and/or 95% confidence**
  - **GE "Effective Reflectivity" method**
    - **A graph of POD versus flaw size at 50% and/or 95% confidence**

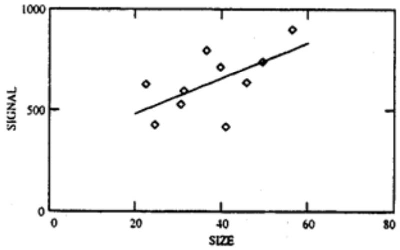
WHO CAN REALLY RUN THESE ANALYSES?

- o **SIMPLE METHODS OF ANALYSIS**
  - **Range-Interval Method? ... ASNT Method? ... Empirical ROC?**
  - **Record what is detected and what is missed**
  - **POD is simply the proportion of defects detected**
  - **PFA is measured as a rate "per opportunity" .. /pulse .. /part .. etc.**
- o **MORE COMPLEX METHODS**
  - **Linear regression on response-size data? ... original R_e method?**
  - **Can be done with a pocket calculator and introductory statistics text**
  - **Easier with general-purpose statistics software (such as Minitab)**
- o **SPECIALIZED METHODS**
  - **USAF "PF"? ... USAF "A-HAT"? ... Need the USAF software**
  - **Modified R_e method? ... Need a good general understanding of statistics and some general-purpose software (e.g. MathCad)**

EXAMPLES OF ANALYSES: LEAST-SQUARES LINEAR REGRESSION

Samples n = 10	Size, X	Signal, Y	X*Y	X^2	Y^2
1	31.4	593	18620	986	351649
2	22.6	627	14170	511	393129
3	49.6	740	36704	2460	547600
4	39.9	712	28409	1592	506944
5	36.6	795	29097	1340	632025
6	30.7	526	16148	942	276676
7	41.0	419	17179	1681	175561
8	45.9	636	29192	2107	404496
9	24.8	426	10565	615	181476
10	56.3	899	50614	3170	808201
Totals:	ΣX=378.8	ΣY=6373	ΣX*Y= 250698	ΣX^2= 15404	ΣY^2= 4277768
Means:	$\bar{X}$=37.88	$\bar{Y}$=637.3			

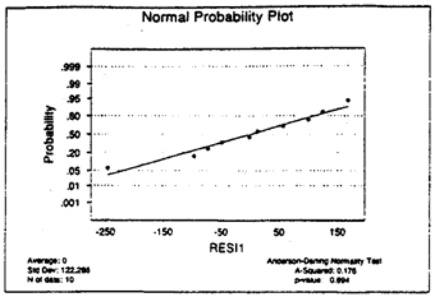
Numerical Example



Linear?

- o **Check that dependence appears to be linear**
  - **Independent variable (e.g. flaw size):** X
  - **Dependent variable (e.g. flaw response):** Y
- o **Fit regression line:** $Y_x = a + b \cdot X$
  - **Coefficients:** $a = \bar{Y} - b \cdot \bar{X}$ $\quad b = \dfrac{\Sigma X \cdot Y - n \cdot \bar{X} \cdot \bar{Y}}{\Sigma X^2 - n \cdot \bar{X}^2}$
  - **Standard Error:** $S_{Y.X} = \sqrt{\dfrac{\Sigma\left(Y - \bar{Y}_x\right)^2}{n - 2}} = \sqrt{\dfrac{\Sigma Y^2 - a \cdot \Sigma Y - b \cdot \Sigma X \cdot Y}{n - 2}}$

EXAMPLES OF ANALYSES: LEAST-SQUARES LINEAR REGRESSION



Cumulative normal plot

DISTANCE FROM AVERAGE DIVIDED BY STANDARD DEVIATION x/σ	AREA	DISTANCE FROM AVERAGE DIVIDED BY STANDARD DEVIATION x/σ	AREA
0.0	0.00000	2.0	0.47725
0.1	0.03983	2.1	0.48214
0.2	0.07926	2.2	0.48610
0.3	0.11791	2.3	0.48928
0.4	0.15542	2.4	0.49180
0.5	0.19146	2.5	0.49379
0.6	0.22575	2.5758	0.49500
0.7	0.25804	2.6	0.49534
0.8	0.28814	2.7	0.49653
0.9	0.31594	2.8	0.49744
1.0	0.34134	2.9	0.49813
1.1	0.36433	3.0	0.49865
1.2	0.38493	3.1	0.49903
1.3	0.40320	3.2	0.49931
1.4	0.41924	3.3	0.49952
1.5	0.43319	3.4	0.49966
1.6	0.44520	3.5	0.49977
1.7	0.45543	3.6	0.49984
1.8	0.46407	3.7	0.49989
1.9	0.47128	3.8	0.49993
1.96	0.47500	3.9	0.49995
		4.0	0.49997

Normal Distribution

- **ESTIMATION OF POD AND FLAW SIZE**
  - **Check that residuals are normally distributed**
  - **Select value $Z = (x/\sigma)$ from normal distribution to match desired mean POD**
    - **e.g. 1.282 corresponds to 90%** (i.e. area = 0.4 below mean + 0.5 above mean)
  - **Calculate flaw size X_t for this POD and selected threshold Y_t:**

$$X_T = \frac{Y_T + (x/\sigma)^* S_{Y,X} - a}{b}$$

EXAMPLES OF ANALYSES: LEAST-SQUARES LINEAR REGRESSION

Regression Analysis

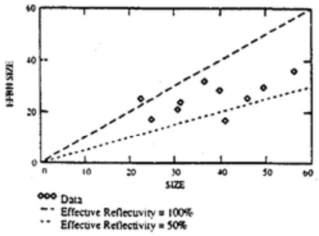
```
The regression equation is
Signal = 304 + 8.80 Size

Predictor       Coef        Stdev     t-ratio        p
Constant       304.1        156.9        1.94    0.089
Size           8.801        3.999        2.20    0.059

s = 129.7       R-sq = 37.7%     R-sq(adj) = 29.9%

Analysis of Variance

SOURCE       DF          SS          MS         F        p
Regression    1       81459       81459      4.84    0.059
Error         8      134585       16823
Total    ·    9      216044

Unusual Observations
Obs.      Size      Signal        Fit  Stdev.Fit   Residual    St.Resid
  7       41.0       418.9      664.8       42.9     -245.9      -2.01R

R denotes an obs. with a large st. resid.
```

Example of "Minitab" output for the previous example

- **Commercially available software can assist with this calculation**
  - **General-purpose mathematics or statistics programs (e.g. Minitab, etc.)**
  - **Special-purpose programs (e.g. USAF "A-HAT")**
    - **Caution: the "A-HAT" program <u>automatically</u> uses log-log coordinates**

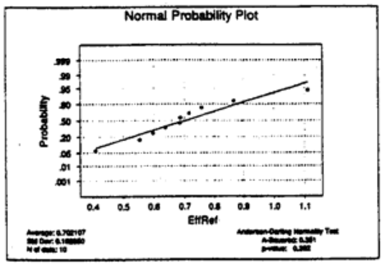
EXAMPLES OF ANALYSES: ORIGINAL EFFECTIVE REFLECTIVITY METHOD

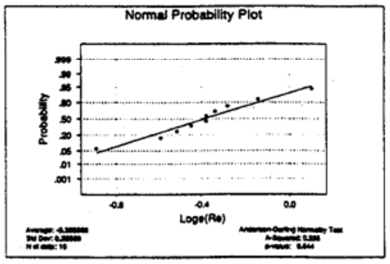
Samples n = 10	Signal	Predicted size	Measured Size	Re	loge(Re)
1	593	23.7	31.4	0.757	-0.278
2	627	25.1	22.6	1.111	+0.105
3	740	29.6	49.6	0.597	-0.516
4	712	28.5	39.9	0.715	-0.336
5	795	21.1	36.6	0.685	-0.378
6	526	25.4	30.7	0.554	-0.591
7	419	16.8	41.0	0.409	-0.894
8	636	17.1	45.9	0.686	-0.377
9	426	31.8	24.8	0.868	-0.142
10	899	36.0	56.3	0.639	-0.448
		Mean:		0.702	-0.385
		Standard Deviation:		0.189	0.266



- **SELECT APPROPRIATE FLAW MODEL: e.g. Flat-Bottomed Hole**
  - **Establish minimum FBH size detectable during inspection**
  - **This is equivalent to a detection threshold A_t (e.g. 20% of #2 FBH = 0.000153 in^2)**
- **INSPECT MATERIAL TO ESTABLISH R_e DATABASE**
  - **Select a group of detected defects**
  - **Calculate "Equivalent FBH size"**
  - **Examine metallographically to determine actual size**
  - **Ratio of EFBH to actual size is "Effective Reflectivity" for each defect**

EXAMPLES OF ANALYSES: ORIGINAL EFFECTIVE REFLECTIVITY METHOD





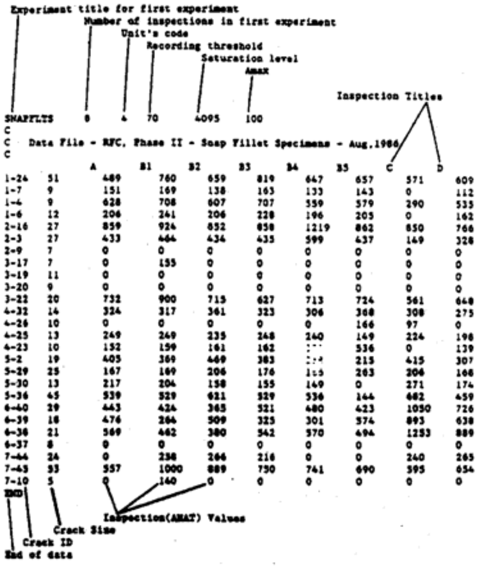
- **SELECT APPROPRIATE STANDARD DISTRIBUTION**
  - **Test for normality or log-normality**
    - **Either distribution could be used to describe the above data**
    - **The log-normal distribution is a somewhat better fit (larger "p")**
  - **Calculate mean and standard deviation for R_e observations**
    - **Calculate in units of the best-fitting distribution**
    - **Normal distribution:** μ_{Re} σ_{Re}
    - **Log-normal distribution:** $\mu_{\log_e(Re)}$ $\sigma_{\log_e(Re)}$

EXAMPLES OF ANALYSES: ORIGINAL EFFECTIVE REFLECTIVITY METH

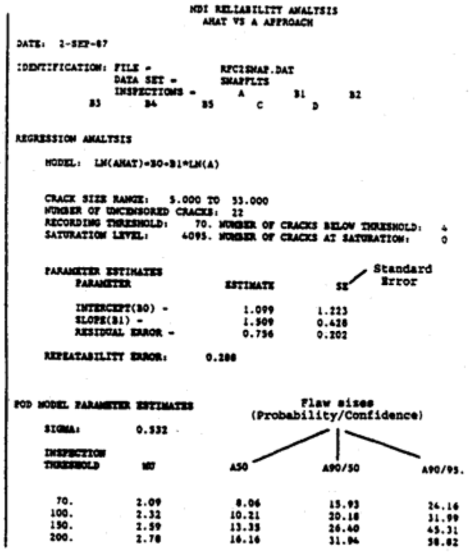
- o **ESTIMATION OF POD AND FLAW SIZE**
  - **Select value $Z = \left(\frac{x}{\sigma}\right)$ from normal distribution to match desired mean POD**
    - **e.g. $Z = 1.282$ corresponds to 90% POD**
  - **Calculate flaw size for this mean POD and threshold A_t**
    - **Normal distribution:** $X_t = \dfrac{A_t}{\mu_{Re} - \left(\frac{x}{\sigma}\right)^{*}\sigma_{Re}}$
    - **Log-normal distribution:** $X_t = \dfrac{A_t}{e^{\left(\mu_{loge(Re)} - \left(\frac{x}{\sigma}\right)^{*}\sigma_{loge(Re)}\right)}}$
- o **EXAMPLE:**
  - **Inspection threshold = 20% of #2 FBH ... i.e. $A_t = 0.000153$ in^2**
  - **Required mean POD = 90% ... i.e. $Z = 1.282$**
  - **Corresponding flaw size = 0.000153/[0.702 - 1.282*0.189] = 0.000333 in^2 (normal solution)**

    or = 0.000153/[e$^{(-0.385 - 1.282*0.266)}$] = 0.000316 in^2 (log-normal solution)

EXAMPLES OF ANALYSES: USAF/UDRI "A-HAT"



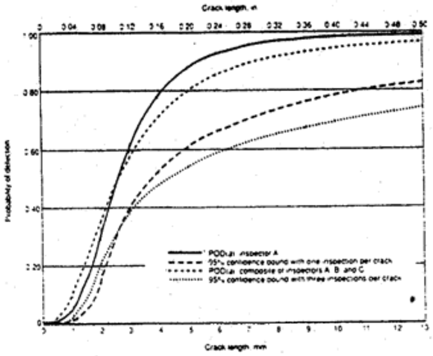
```
SHAPFLTS    8    4   70     4095    100
C
C   Data File - RFC, Phase II - Soap Fillet Specimens - Aug.1986
C
              A      B1     B2     B3     B4     B5     C      D
1-24   51    489    760    659    819    647    657    571    609
1-7    9     151    169    138    163    133    143    0      112
1-4    9     628    708    607    707    559    579    290    535
1-6    12    206    241    206    228    196    205    0      162
2-16   27    859    924    852    858    1219   862    850    766
2-3    27    433    464    434    435    599    437    149    328
2-9    7     0      0      0      0      0      0      0      0
3-17   7     0      155    0      0      0      0      0      0
3-19   11    0      0      0      0      0      0      0      0
3-20   9     0      0      0      0      0      0      0      0
3-22   20    732    900    715    627    713    724    561    648
4-32   14    324    317    361    323    306    368    308    275
4-26   10    0      0      0      0      0      166    97     0
4-25   13    249    249    235    248    240    149    224    198
4-23   10    152    159    161    162    [illegible]    536    0      139
5-2    19    405    369    469    383    [illegible]    215    415    307
5-29   25    167    169    206    176    [illegible]    263    206    168
5-30   13    217    204    158    155    149    0      271    174
5-36   45    539    529    621    529    536    144    482    459
6-40   29    443    424    365    521    480    423    1050   726
6-39   18    476    284    509    325    301    574    893    638
6-38   21    569    462    380    542    570    494    1253   889
6-37   8     0      0      0      0      0      0      0      0
7-44   24    0      238    266    216    0      0      240    265
7-45   33    557    1000   889    730    741    690    595    654
7-10   5     0      140    0      0      0      0      0      0
END
```



```
                              NDI RELIABILITY ANALYSIS
                               AHAT VS A APPROACH

DATE:  2-SEP-87

IDENTIFICATION: FILE -          RFC2SNAP.DAT
                DATA SET -      SNAPFLTS
                INSPECTIONS -      A          B1         B2
           B3         B4         B5         C          D

REGRESSION ANALYSIS

     MODEL:  LN(AHAT)=B0+B1*LN(A)

     CRACK SIZE RANGE:    5.000 TO  33.000
     NUMBER OF UNCENSORED CRACKS:  22
     RECORDING THRESHOLD:     70. NUMBER OF CRACKS BELOW THRESHOLD:    4
     SATURATION LEVEL:      4095. NUMBER OF CRACKS AT SATURATION:     0

     PARAMETER ESTIMATES
          PARAMETER                   ESTIMATE      SE

          INTERCEPT(B0) -               1.099      1.223
          SLOPE(B1) -                   1.509      0.428
          RESIDUAL ERROR -              0.756      0.202

     REPEATABILITY ERROR:        0.288

POD MODEL PARAMETER ESTIMATES

     SIGMA:          0.532

     INSPECTION
     THRESHOLD        MU           A50          A90/50        A90/95.

        70.          2.09          8.06          15.93          24.16
       100.          2.32         10.21          20.18          31.99
       150.          2.59         13.35          26.40          45.31
       200.          2.78         16.16          31.94          58.82
```

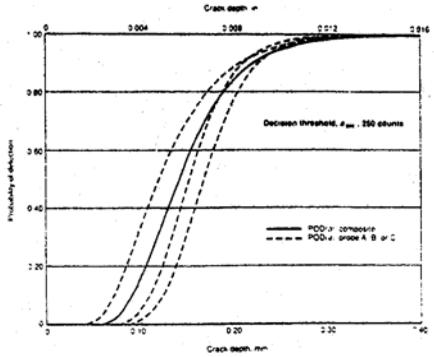
- **Examples of input and output data sheets** (Berens et al., 1988)

EXAMPLES OF ANALYSES: USAF/UDRI "PF" AND "A-HAT"



"P/F": Penetrant inspection

(three inspectors)



"A-HAT": Automated eddy-current inspection

(three probes)

- **Examples of POD data for flat-plate cracks** (Berens, Metals Handbook, 1989)

CONCLUDING COMMENTS

- THE MAJOR GOAL OF NDE IS DETECTION OF DEFECTS
  - Well-chosen NDE techniques do successfully detect defects
  - Use of well-chosen techniques does not <u>guarantee</u> detection

- FLAW DETECTION DEPENDS ON NUMEROUS PARAMETERS
  - Material and manufacturing process properties
  - Flaw properties
  - Inspection equipment and inspection process parameters

- PROBABILITY OF DETECTION
  - Detection capability is measured from a representative sample of flaws
  - Standardized statistical procedures used to estimate "population" properties
  - Several alternative methodologies exist
    - Each has advantages and limitations

CONCLUDING COMMENTS

POD DATA ACQUISITION REQUIRES:

- WELL-DEFINED AND CONTROLLED INSPECTION CONDITIONS
  - Essential since detectability depends on inspection conditions

- WELL-DEFINED DEFECT TYPE AND MATERIAL PROPERTIES
  - Essential since detectability depends on properties of defects and material

- AVAILABILITY OF APPROPRIATE SAMPLE DEFECTS
  - Naturally-occurring or realistically simulated
  - Suitably distributed properties to cover all relevant variables
  - Adequate number to give sufficient confidence in the results

- AVAILABILITY OF A "REFEREE" TECHNIQUE
  - Need to establish "true" flaw number and size
  - Nondestructive technique desirable to avoid loss of flaw samples

CONCLUDING COMMENTS

- o APPLICABILITY OF POD DATA
  - Strictly applicable only to the circumstances used in their acquisition

- o ANY CHANGE POTENTIALLY INVALIDATES THE DATA
  - Changes in how the inspection is performed
  - Changes affecting the transmitted signals
  - Changes affecting the signals reaching the flaw
  - Changes affecting the flaw responses
  - Changes affecting the received flaw signals and how they are processed
  - Changes affecting the received noise signals and how they are processed
  - Changes in how the outputs are perceived and judged

- o POD SHOULD BE REMEASURED AFTER ANY CHANGE
  - Application under changed circumstances involves assumptions
    - Any such assumptions MUST be tested as soon as possible

---

CONCLUDING COMMENTS

- o MEASUREMENT OF POD ALONE
  - Acceptable: POD is not dependent on PFA
    - Knowledge of POD is adequate for determining capability and reliability
    - Knowledge of POD is adequate for predicting product life
  - Caution: thresholds must be set realistically far above "noise"
    - Inconsistent thresholding may occur in non-automated inspections

- o MEASUREMENT OF POD AND PFA
  - Recommended: a more complete description of the detection process
    - Knowledge of PFA helps in selecting an optimum threshold
    - Knowledge of PFA helps identify manufacturing costs
    - Knowledge of PFA provides a check on consistency of thresholding
  - Caution: empirical ROC studies often ignore the size-dependence of POD

CONCLUDING COMMENTS

- **POD CONCEPTS PROVIDE:**
  - **Improved understanding of capabilities and limitations of NDE techniques**
  - **A common basis for review between design, manufacturing and NDE functions**
  - **A rational basis for matching NDE techniques to engineering needs**
  - **A basis for estimation of residual life of cyclically-loaded components**
  - **An option for measuring the success of training programs**
  - **An option for qualifying and monitoring personnel performance**

- **POD DATA:**
  - **Are strictly valid only under very specific conditions**
  - **Should be used only with good understanding of NDE technology**

- **POD IS NOT A PANACEA**
  - **Quantification of inspection capability is generally desirable**
  - **Quantification of inspection capability is not always practicable**

BIBLIOGRAPHY

C.Annis
"Quantifying Non-Destructive Evaluation Capability and Reliability", Probability of Detection Tutorial Workshop, 1993 ASNT Spring Conference

C.Annis, A.Berens, F.Bray, K.Erland, G.Hardy, W.Herron, W.Hoppe, & S.Vukelich
"Proposed MIL-STD for USAF NDE System Reliability Assessment", AF Contract No. F33615-81-C-5002 (1989)

M.J.Avioli, Jr.
"Defining and Measuring NDE Reliability", presentation at the 1992 Fall ASNT Conference

M.J.Avioli, Jr.
Untitled presentation, Probability of Detection Tutorial Workshop, 1993 ASNT Spring Conference

M.J.Avioli, Jr. & S.-N.Liu
"Nuclear Power Plant Costs as a Function of NDE Reliability", Review of Progress in QNDE, Vol. 10B, pp. 2225-2233 (1991)

D.K.Barton
"Radar Systems Analysis" (Prentice-Hall, 1964)

R.E.Beissner, K.A. Bartels & J.L.Fisher
"Prediction of the Probability of Eddy Current Flaw Detection", Review of Progress in QNDE (Plenum Press, 1988), Vol. 7B, pp. 1753-1760

N.C.Bellinger, A.Fahr, A.K.Koul, G.Gould & P.Stoute
"The Reliability and Sensitivity of NDI Techniques in Detecting LCF Cracks in Fastener Bolt Holes of Compressor Discs", National Aeronautical Establishment Laboratory Report LTR-ST-1651 (National Research Council Canada, 1988)

A.P.Berens
"NDE Reliability Data Analysis", in Metals Handbook, 9th edition, Vol. 17, pp. 689-701 (ASM International, 1989)

A.P.Berens, P.W.Hovey, R.M.Donahue & W.N.Craport
"User's Manual for Probability of Detection Software System (POD/SS)", University of Dayton Research Institute (1988)

T.G.Birdsall
"POD, ROC, Noise and Reality", presentation at the Fall 1992 ASNT meeting

R.H.Burkel, R.S.Gilmore, D.J.Sturges & W.T.Tucker
"Probability of Detection for Applied Ultrasonic Inspections", Review of Progress in QNDE (Plenum Press, 1996), Vol. 15 (to be published)

C-P.Chiou, R.B.Thompson & L.W.Schmerr
"Model-based Signal Processing Techniques for Ultrasonic Flaw Detection: Simulation Studies", Review of Progress in QNDE, Vol.12A, pp. 703-710 (1993)

J.M.Coffey & R.K.Chapman
"Application of elastic scattering theory for smooth flat cracks to the quantitative prediction of ultrasonic defect detection and sizing", Nuclear Energy 22, 319 (1983)

H.W.Cole
"Understanding Radar" (Collins, 1985)

J.G.Costa, R.E.Gonzalez, R.E.Guyotte, R.J.Koenig, D.P.Salvano & T.Swift
"Titanium Rotating Components Review Team Report", Volume 1 (US Federal Aviation Administration, 1990)

W.D.Cowie
"Fracture Control Philosophy", in Metals Handbook, 9th edition, Vol. 17, pp. 666-673 (ASM International, 1989)

W.B.Davenport & W.L.Root
"An Introduction to the Theory of Random Signals and Noise" (McGraw-Hill, 1958)

M.K.Davis
"Proficiency Evaluation of NDE Personnel Utilizing the Ultrasonic Methodology", Review of Progress in QNDE (Plenum Press, 1988), Vol. 7B, pp. 1777-1789

K.W.Fertig
"Nondestructive Measurement System Performance: Statistical Characterization", Review of Progress in QNDE (Plenum Press, 1985), Vol. 4, pp. 1297-1304

J.N.Gray, T.A.Gray, N.Nakagawa & R.B.Thompson
"Models for Predicting NDE Reliability", in Metals Handbook, 9th edition, Vol. 17, pp.702-715 (ASM International, 1989)

D.J.Hagemaier
"Reliable Supplemental Inspections of Aging Commercial Aircraft", presented at the 1987 Fall ASNT Conference

N.F.Haines
Report RD/B/5209N81, PWR/RCC/MWG/P{81}56, (Central Electricity Generating Board (U.K.), 1981)

W.M.Hall
"Prediction of Pulse Radar Performance" (Proc. IRE 44, 224 (1956)

C.A.Hawkins & J.E.Weber
"Statistical Analysis" (Harper & Row, 1980)

P.G.Heasler, D.J.Bates, T.T.Taylor & S.R.Doctor
NUREG/CR-4464 (U.S.Nuclear Regulatory Commission, 1986)

J.M.Juran & F.M.Gryna, Jr
"Quality Planning and Analysis" (McGraw-Hill, 1980)

A.K.Koul, A.Fahr, G..M.Gould & N.C.Bellinger
"Influence of Sensitivity and Reliability of NDI Techniques on Damage Tolerance Based Life Prediction of Turbine Discs", National Aeronautical Establishment Laboratory Report LTR-ST-1665 (National Research Council Canada, 1988)

J.Krautkramer & H.Krautkramer
"Ultrasonic Testing of Materials" (Springer-Verlag, 1969)

F.J.Margetan, I.Yalda & R.B.Thompson,
"Predicting Gated-Peak Grain Noise Distributions for Ultrasonic Inspections of Metals", Review of Progress in QNDE (Plenum Press, 1996), Vol. 15 (to be published)

W.Q.Meeker
"Methodology for Estimating NDE Capability", presented at Engine Titanium Consortium Open Forum, 199

W.Q.Meeker, R.B.Thompson, C.-P.Chiou, S.L.Jeng & W.T.Tucker
"Methodology for Estimating Nondestructive Evaluation Capability", Review of Progress in QNDE (Plenum Press, 1996), Vol. 15 (to be published)

N.Nakagawa, M.W.Kubovich & J.C.Moulder
"Modelling Inspectability for an Automated Eddy Current Measurement System", Review of Progress in QNDE, Vol. 9A, pp. 1065-1072 (Plenum Press, 1990)

C.Nockemann, H.Heidt & N.Thomsen
"Reliability in NDT: ROC study of radiographic weld inspections", NDT&E International, Vol. 24, pp. 235-245 (1991)

C.Nockemann, G.-R.Tillack & H.Wessel
"ROC within a Global Conception of Reliability in NDE" presented at the 1994 ASNT Spring Conference

D.O.North
"Analysis of Factors Which Determine Signal-to-Noise Discrimination in Radar", Report PTR-6c (RCA Laboratories, 1943)

P.F.Packman, S.J.Klima, R.L.Davies, J.Malpani, J.Moyzis, W.Walker, B.G.W.Yee &
D.P.Johnson, "Reliability of Flaw Detection by Nondestructive Inspection", in Metals Handbook, 8th edition, Vol.11, p.414 (American Society for Metals, 1976)

W.W.Peterson & T.G.Birdsall
"The Theory of Signal Detectability", Technical Report No. 13, Electronic Defense Group (University of Michigan, 1953); and Trans. IRE PGIT-4, 171 (1954) {with W.C.Fox}

S.B.Richmond
"Statistical Analysis" (Ronald, 1964)

W.D.Rummel
"Recommended Practice for a Demonstration of Nondestructive Evaluation Reliability on Aircraft Production Parts", Materials Evaluation 40, 922 (1982)

W.D.Rummel
"Assessment and Demonstrationof the Capabiliyties of NDI Processes, Equipment and Personnel", presented at the 1989 NATO AGARD Specialist Meeting on NDE/NDI

W.D.Rummel, B.K.Christner & D.L.Long
"Methodology for Analysis and Characterization of Nondestructive Inspection Capability Data", Review of Progress in QNDE (Plenum Press, 1988), Vol. 7B, pp. 1769-1776

W.D.Rummel, G.L.Hardy, & T.D.Cooper
"Applications of NDE Reliability to Systems", in Metals Handbook, 9th edition, Vol. 17, pp. 674-688 (ASM International, 1989)

C.Salkowski
"NASA Use and Development of Quantitative NDE for Fracture Control and Damage Tolerance", presentation at the Fall 1992 ASNT Conference

R.C.Schwant & D.P.Timo
"Inspection and Life Evaluation of General Electric Turbine Rotors", ASME Paper 84-JPGC-Pwr-4 (1984)

J.C.Spanner, Jr.
"Human Factors Impact on NDE Reliability", Review of Progress in QNDE (Plenum Press, 1988), Vol. 7B, pp. 1791-1798

F.Spencer
"Procedure Assessment for EC Inspection Reliability", Proceedings of the Second FAA Inspection Reliability Workshop, 1995

D.J.Sturges
"Quantification of Inspection Process Capability", Proceedings of the National Conference on Nondestructive Testing Monitoring Diagnostics, (Italian Society for Nondestructive Testing Monitoring Diagnostics, Milano, 1990)

D.J.Sturges

"The Role of NDE in Predicting Residual Life for Aircraft Engine Components", Proceedings of the International Conference on Monitoring and Predictive Maintenance of Plants and Structures (Italian Society for Nondestructive Testing Monitoring Diagnostics, Firenze, 1992)

D.J.Sturges

"A Subsurface Statistical Safari and The Pursuit of a Portable POD", Proceedings of the First FAA Inspection Reliability Workshop, 1993

D.J.Sturges

"Use of Simulated Defects for Disk Inspection", Proceedings of the Second FAA Inspection Reliability Workshop, 1995

D.J.Sturges & R.B.Abernethy

"Desirable Distinctions in NDE Reliability Terminology and Practice", in "Reliability Analysis Workshop Summary Report", ed. V.E.Panhuise & T.D.Cooper (WPAFB, 1982, unpublished); and Review of Progress i QNDE, Vol. 12B, pp. 2381-2388 (1993)

D.J.Sturges, R.S.Gilmore & P.W.Hovey

"Estimating Probability of Detection for Subsurface Ultrasonic Inspection", Review of Progress in QNDE, Vol. 5A, 929 (1986)

J.A Swets

"The Relative Operating Characteristic in Psychology", Science 182, 990 (1973)

J.A.Swets

"Assessment of NDT Systems - Part I: The Relationship of True and False Detections", Materials Evaluation 41, 1294-1298 (1983)

J.A.Swets

"Assessment of NDT Systems - Part II: Indices of Performance", Materials Evaluation 41, 1299-1303 (1983)

J.A.Swets & R.M.Pickett

"Evaluation of Diagnostic Systems: Methods from Signal Detection Theory" (Academic, 1982)

R.B.Thompson et al.

Review of Progress in QNDE 3, 13 (1984)

R.B.Thompson

"The Role of Physical Modeling of Inspection Processes", Probability of Detection Tutorial Workshop, 1993 ASNT Spring Conference

R.B.Thompson

"Overview of Engine Titanium Consortium Task 3: Probability of Detection, or, What? Another POD Study? Proceedings of the First FAA Inspection Reliability Workshop, 1993

W.Tucker

"Likelihood Ratio Methods for Ti Inspection", Proceedings of the First FAA Inspection Reliability Workshop, 1993

R.J.Urick

"Principles of Underwater Sound" (McGraw-Hill, 1983)

H.L.Van Trees

"Detection, Estimation, and Modulation Theory, Part I" (Wiley, 1968)

A.D.Whalen

"Detection of Signals in Noise" (Academic, 1971)

W.A.Wheeler, W.L.Rankin, J.C.Spanner, R.V.Badalemente & T.T.Taylor,

"Human Factors Study Conducted in Conjunction with a Mini-Round Robin Assessment of Ultrasonic Technician Performance", NUREG/CR-4600 (U.S.Nuclear Regulatory Commission, 1986)