

# Konzept für den Einsatz der Bayes'schen Statistik: In der Zuverlässigkeitsbestimmung von zerstörungsfreien Diagnosesystemen – Das Problem zu kleiner experimenteller Datenmengen

Daniel KANZLER\*, Christina MÜLLER\*,  
Jorma PITKÄNEN\*\*

\*BAM Bundesanstalt für Materialforschung und -prüfung, Berlin<sup>1</sup>

\*\*Posiva Oy, Finnland<sup>2</sup>

**Kurzfassung.** Die statistische Qualitätskennzahl „Probability of Detection (POD)“ hängt von der Anzahl der zugrunde liegenden, brauchbaren Ergebnisse ab. Ist die Datenmenge niedrig, so ist auch die Aussage der POD eher beschränkt. Dennoch heißt eine große Zahl von Daten, nicht gleichzeitig, dass die POD für den realen Einsatz in der Produktion genügt. Es muss unterschieden werden, auf welcher Basis – unter welchen konkreten Bedingungen - die Daten erhoben wurden.

Als wichtigste Ergebnisse zählen die Messergebnisse aus der Produktion der später verwendeten Prüfobjekte für den konkreten Einsatz. Noch bevor die zerstörungsfreien Prüfmethoden zuverlässig in der Produktion eingesetzt wurde, muss jedoch nachgewiesen werden, dass die real vorkommenden Fehler gefunden werden können. Das bedeutet, zu einem Zeitpunkt, an dem noch nicht ausreichend Daten vorhanden sind.

Während für die realistische Prüfung die notwendige Anzahl der verschiedenen Fehler meist nicht erreicht wird, so fehlen bei der Bewertung der künstlichen Fehler wichtige Parameter und Einflussgrößen. Es muss somit eine Zwischenlösung gesucht werden, in der die Informationen von beiden Prüfungen kombiniert werden. Mit der Bayes'schen Statistik ermöglicht die Mathematik diese Informationen miteinander zu kombinieren und somit beide Informationen mit in die Bewertung einfließen zu lassen.

Praktisch angewandt und erprobt wird die Vorgehensweise gemeinsam mit der Firma POSIVA, die sich mit dem Bau eines Endlagers für hoch-radioaktiven Abfall in Finnland beschäftigt. Hier wird als eine der technischen Barrieren ein Kupferbehälter für die Aufbewahrung der verbrauchten Brennstäbe genutzt. Dieser Kupferbehälter wird mit mehreren zerstörungsfreien Prüfmethoden geprüft. Anhand von Radiographiedaten wird gezeigt, dass die Aussagekraft der realistischen Fehlerdaten mit Hilfe der künstlichen Referenzfehlerdaten untermauert werden kann.

<sup>1</sup> BAM - Unter den Eichen 87, 12205 Berlin

<sup>2</sup> POSIVA OY – Oikiluoto – FI-27160 Eurajoki - Finnland

## 1. Einführung

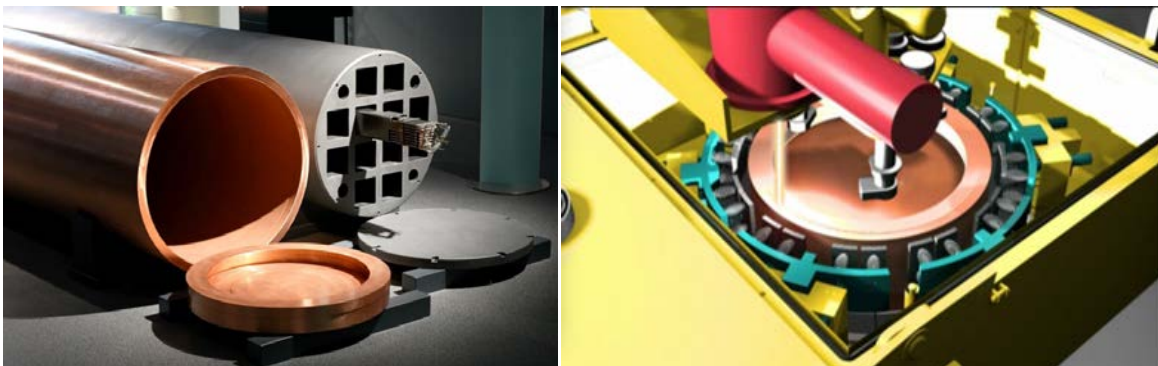
Die Zuverlässigkeit eines zerstörungsfreien Prüfsystems (ZfP) wurde im zweiten Europäischen-Amerikanischen Workshop zur Zuverlässigkeit in der zerstörungsfreien Prüfung wie folgt definiert: „Die ZfP-Zuverlässigkeit ist das Maß mit dem das ZfP-System in der Lage ist den vorgesehenen Zweck zu erfüllen bzgl. der Detektion, der Charakterisierung und der Falschanzeigen“ [1].

Das weitverbreitetste Verfahren zur Bestimmung einer Zuverlässigkeitskenngröße ist die Fehlerdetektionswahrscheinlichkeit (POD) von Berens anhand von Referenzfehlern. Die Referenzfehler sollen hierbei reale Fehler im zu prüfenden Bauteil repräsentieren [2]. Doch obwohl die POD oft als Qualitätskennzahl eingeführt wird, sind den Anwendern die Grenzen des Verfahrens nicht immer bewusst. Eine der größten Hürden ist die Anzahl der Experimente, die benötigt wird, um die statistische Korrektheit des Verfahrens zu gewährleisten. Besonders für reale Fehlerprüfungen reicht diese Anzahl in vielen Fällen nicht aus.

Dieser Artikel beschreibt eine, auf das Theorem von Bayes basierenden, Ansatz, um trotz geringer Datenmengen möglichst korrekte Aussagen in der Zuverlässigkeitsbestimmung zu erlangen. Der Artikel beschreibt anfangs das Einsatzgebiet der POD und die Entstehung der zerstörungsfreien Prüfergebnisse. Darauf folgen eine Einführung in die POD-Berechnung und eine Einführung in den Bayes-Ansatz. Anhand beider Werkzeuge wird im Folgenden eine Qualitätskennzahl erstellt, die Informationen aus realistischen und künstlichen Fehlern erfasst. Die Evaluierung wird anhand von Radiographiedaten durchgeführt. Der Abschluss bildet eine Zusammenfassung und einen Ausblick auf die weiteren Schritte.

## 2. Projektbeschreibung

Das Problem der Endlagerung von hoch-radioaktivem Abfall muss in der nahen Zukunft gelöst werden, Ein Ansatz ist die langfristige unterirdische Endlagerung. Dieses Konzept wird von Finnland unterstützt, das momentan ein Endlager für ihren hoch-radioaktiven Abfall baut. Um sicher zu gehen, dass dieser Abfall für den geplanten Zeitraum untertage bleibt, wurde ein drei Barrieren-Sicherheitssystem entworfen. Drei unabhängige technische und natürliche Barrieren sollen die sichere Verwahrung der radioaktiven Abfälle sicherstellen. Eine technische Barriere stellt der Kupferkanister dar, der vor Korrosion schützen soll. Der Kupferbehälter ist in Abbildung 1 zu sehen.



**Abbildung 1:** Die Komponenten des Endlagerkanisters für hoch-radioaktiven Abfall und der schematische Prozess des Elektronenstrahlschweißens (POSIVA)

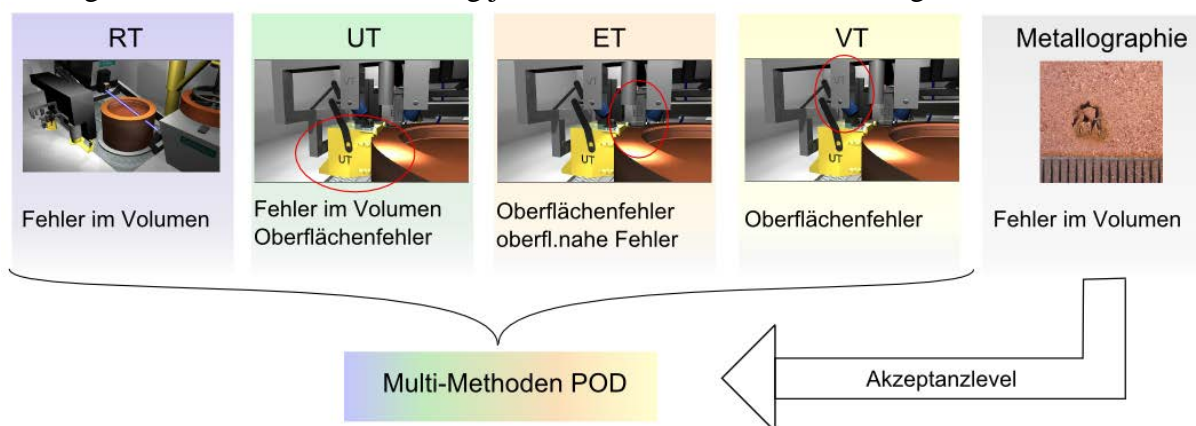
Der Kupfermantel wird durch den Kupferdeckel mit Hilfe der Elektronenstrahlschweißung verschlossen (Abbildung 1 rechts). Das Elektronenstrahlschweißen ist ein Prozess, bei dem durch den Aufprall des Elektronenstrahls so viel Wärme entsteht, dass

das Kupfer schmilzt und Deckel und Mantel miteinander verschweißt werden. In dieser Schweißnaht und in der Wärmeeinflusszone können jedoch verschiedene Materialfehler entstehen, die die Funktionsfähigkeit des gesamten Behälters gefährden können. Zu nennen sind hier beispielsweise Porosität, Hohlräume im Material, Lunker und zu tiefes Eindringen des Elektronenstrahls ins Grundmaterial. Um die Funktionsfähigkeit der Komponenten zu gewährleisten, werden zerstörungsfreie Prüfsysteme eingesetzt, um Materialfehler ab der kritischen Größe zu finden.

Um sicherzugehen, dass jeder kritische Fehler gefunden wird, werden vier verschiedene zerstörungsfreie Prüfverfahren eingesetzt (Abbildung 2):

- Radiographie (RT)
- Ultraschallprüfung (UT)
- Wirbelstromprüfung (ET)
- Visuelle Prüfung (VT)

Die Verfahren basieren auf verschiedenen physikalischen Prinzipien, wodurch unterschiedliche Arten und Fehler gefunden werden können. Die Aussage, ob das Objekt fehlerfrei ist, wird aus der kombinierten Antwort aller Verfahren gebildet und für einzelne Fehler mit der Metallographie überprüft (Abbildung 2). Die Überprüfung mit Hilfe der Metallographie baut darauf auf, dass die ZfP-Methoden nur Signalbilder des Fehlers wiedergeben, die zerstörende Prüfung jedoch die absolute Wahrheit zeigt.



**Abbildung 2:** ZfP – Verfahren zur Prüfung der Schweißnaht des Kupferbehälters (POSIVA)

Die Praxis hat gezeigt, dass jedes Prüfsystem, das an den Grenzen seiner Fähigkeiten angewendet wird, nicht mehr alle Fehler findet. Von mehreren Fehlern der gleichen Größe, werden nicht alle gefunden [3]. Dies kann auch nicht durch Wiederholungsprüfung ausgeglichen werden. Aus diesem Grund hat sich für jede Anwendung der ZfP deren Ausfall große Folgen mit sich ziehen würde, eine Zuverlässigkeitsbewertung etabliert. Für die Gesamtbewertung des oben beschriebenen Systems wird zunächst jedes einzelne ZfP-System hinsichtlich seiner Fehlerdetektionswahrscheinlichkeit bewertet. In der zukünftigen Projektphase ist eine Fusion der Einzelergebnisse geplant [5].

Zur Bewertung betrachtet man drei voneinander getrennte Teile der Prüfdurchführung. Im ersten Teil beschränkt sich die Prüfung nur auf die bekannten Referenzfehler. Im zweiten Schritt werden realistischen Fehler, die absichtlich durch Abweichung von den idealen Schweißbedingungen hergestellt wurden, zur Bewertung herangezogen. Die eigentliche Aussage ist jedoch während des Produktionsprozesses ausschlaggebend. Die Prüfungen der künstlichen und realistischen Fehler geben somit eine Aussage, ob die Prüfmethode und das Prüfequipment generell für die spätere Prüfung der produzierten Kanister verwendet werden kann.

### 3. Probability of Detection

Das meist verbreitetste Verfahren zur Quantifizierung der Zuverlässigkeit ist die Probability of Detection (POD) [2]. Diese wurde erstmals in den 70er Jahren von Berens eingeführt, um die Detektion von Rissen in Turbinenschaufeln von Flugzeugen mit Hilfe des Wirbelstroms zu bewerten [4]. Seither gilt die POD in vielen industriellen Bereichen als die wichtigste Qualitätskennzahl für die zerstörungsfreie Prüfung.

Das Vorgehen zur Erstellung der POD wertet die maximale Amplitude des ZfP-Verfahrens (als „ $\hat{a}$ “ bezeichnet) im Gegensatz zum signifikanten Fehlerparameter („ $a$ “) aus [4]. Aus Sicht des Einsatzes der Kupferkanister ist die verbleibende Wanddicke das Hauptmerkmal. Je größer die von Fehlern unbeschädigte Wanddicke ist, umso wahrscheinlicher ist es, dass der Kanister für die gesamte Lebenszeit standhält. Ein Materialfehler verringert die verbleibende Wanddicke und somit auch die Zuverlässigkeit des Behälters. Dadurch dass der Materialfehler die Strahlung unterschiedlich zum fehlerfreien Material absorbiert, kann die Radiographieprüfung diesen Fehler finden. Im Fall der Radiographie entspricht somit die durchstrahlte Länge des Fehlers dem Parameter „ $a$ “ der POD. Aus den Grauwerten des Röntgenbilds können Kontrast und Rauschen abgelesen werden, die das „ $\hat{a}$ “ bilden und die Detektionsgrenze „ $\hat{a}_{dec}$ “ bestimmen. In diesem einfachen Fall wird angenommen, dass nur ein Parameter Einfluss auf die POD hat. Komplexere Zusammenhänge wie geometrische Einflüsse spielen in diesem Artikel keine Rolle.

Für eine ganzheitliche Aussage ist ein kontinuierlicher Zusammenhang von Interesse. Bei den Daten der Radiographieprüfung wurde gezeigt, dass ein linearer Zusammenhang dadurch erzielt werden kann, dass beide Achsen logarithmiert werden [5]. Der daraus entstehende Graph wird nach seiner Achsenbezeichnung „ $\hat{a}$  vs.  $a$ “-Graph genannt und ist für die künstlichen Referenzfehler in Abbildung 3 zu sehen.

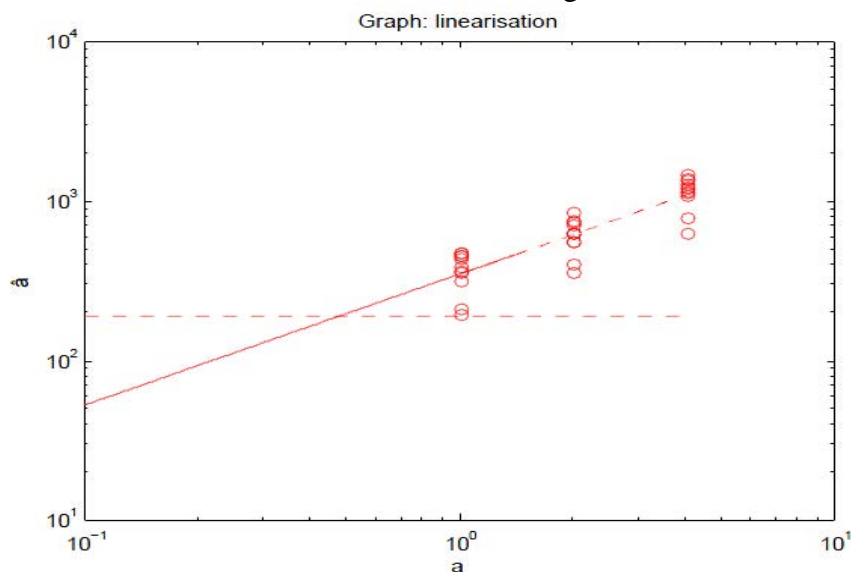


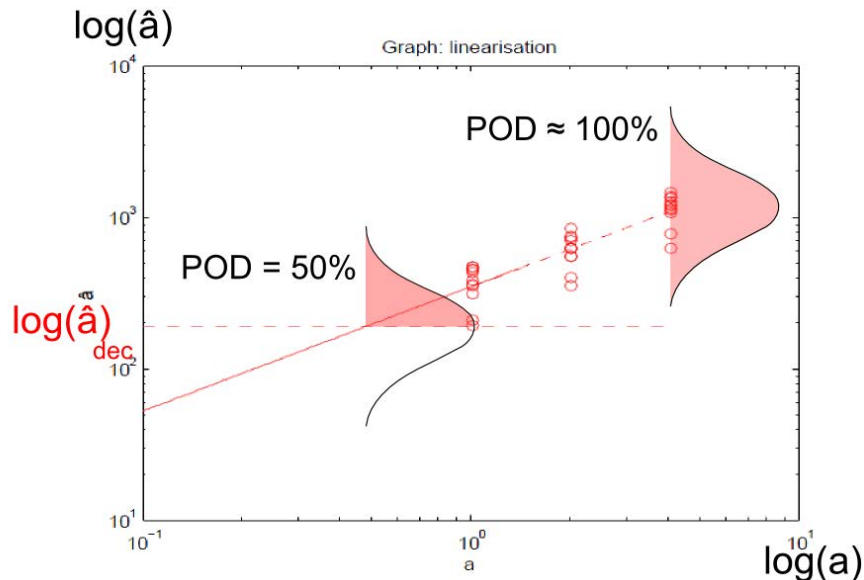
Abbildung 3: Linearer  $\hat{a}$  vs.  $a$  Graph für Radiographie für künstlichen Referenzfehler

Wie an der Abbildung zu sehen ist, wird das Absorptionsgesetz anhand des Anstiegs der interpolierten Geraden ersichtlich. Die Streuung der Messwerte um die Gerade kann zurückgeführt werden auf die Meßunsicherheit des Meßverfahrens. Mit Hilfe der interpolierten Geraden und der Streuung kann nun für jeden Status des Parameters „ $a$ “ ein dazugehöriges „ $\hat{a}$ “ abgeschätzt werden.

Desweiteren wird die Detektionsschwelle „ $\hat{a}_{dec}$ “ eingeführt. Diese gibt eine Aussage, welche Amplituden klar vom Rauschen unterschieden werden können. Der „ $\hat{a}$  vs.  $a$ “-Graph und Detektionsschwelle „ $\hat{a}_{dec}$ “, geben Auskunft über die Detektionswahrscheinlichkeit des Verfahrens für die Fehler. Bei dem „ $a$ “-Wert, bei dem deren Messwerte überwiegend

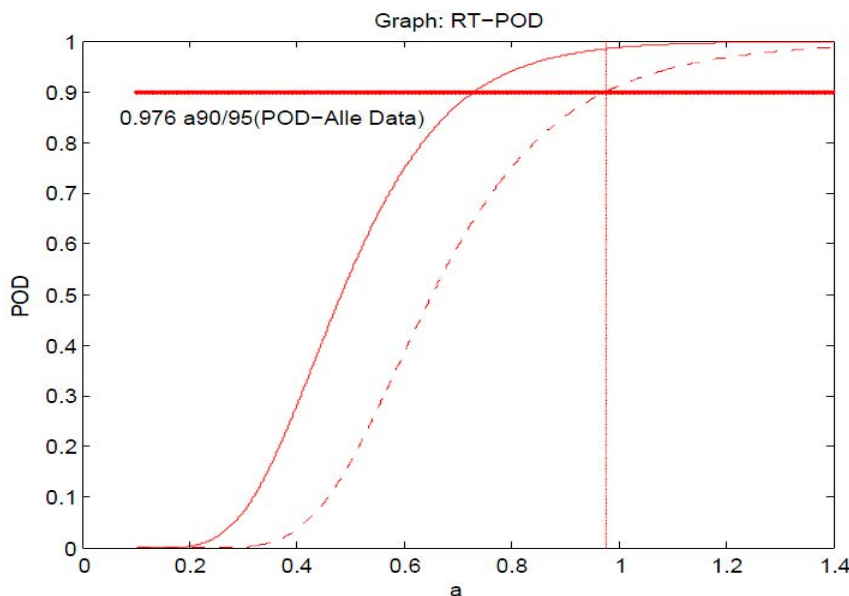
oberhalb der Detektionsschwelle liegen, kann gesagt werden, dass diese Fehlerdimension eher gefunden werden. Während Messwerte die überwiegend unterhalb der Detektionsschwelle liegen, eher nicht gefunden werden.

Mit der Annahme, dass die Messwerte um den Mittelwert jeweils normalverteilt streuen, kann eine quantitative Fehlerdetektionswahrscheinlichkeit berechnet werden. Es kann nun für jede Größe „a“ die Detektionswahrscheinlichkeit als Fläche unter der Normalverteilung oberhalb der Detektionsschwelle berechnet werden (Siehe hierzu die roteingefärbte Fläche in Abbildung 4).



**Abbildung 4:** Normalverteilte Streuung der Messwerte um die Mittelwerte (Regressionsgerade)

Graphisch wird die Gesamtaussage der einzelnen kumulierten Wahrscheinlichkeitswerte in der POD-Kurve mit Konfidenzintervall zusammengefasst [4]. Wie in Abbildung 5 zu sehen ist, wird dabei die X-Achse weiterhin die Fehlergröße bezeichnen, während auf die Y-Achse die Fehlerdetektionswahrscheinlichkeit aufgetragen werden.



**Abbildung 5:** POD Graph für künstliche Daten

An der Stelle an der das 95%-Konfidenzintervall die 90% Detektionswahrscheinlichkeit (POD) schneidet, wird der „a“-Wert  $a_{90/95}$  bestimmt. Dieser Wert gibt eine Aussage

über den Fehler, der höchstwahrscheinlich gefunden wird. Der  $a_{90,95}$  Wert wird mit dem kritischen Fehler, der gefunden werden muss, verglichen. Nur falls der  $a_{90,95}$  Wert kleiner als der kritische Fehler ist, kann das Prüfsystem für die Prüfung akzeptiert werden.

#### 4. Bayes'sche Statistik

Ein wichtiger Punkt bei der Probability of Detection ist die Anzahl der zur Verfügung stehenden Messwerte. Eine POD, die auf zu wenigen Datenpunkten basiert, erfüllt nicht die statistischen Anforderungen und muss sehr kritisch betrachtet werden und im schlimmsten Fall als unbrauchbar und falsch verworfen werden.

Aus der Sichtweise von Normen, behilft man sich mit den künstlichen Referenzfehlern. Auf der Basis der künstlichen Fehler, wie beispielsweise Flachbodenbohrungen für die Ultraschallprüfung, kann geprüft werden, ob generell Fehler in kritischer Größe gefunden werden. Die Kosten für adäquate Referenzfehler in vergleichbarem Material sind im Allgemeinen verkraftbar. Dennoch ist es nicht ohne Weiteres möglich, von dem Ergebnis der künstlichen Referenzfehlerprüfung auf das Ergebnis der Prüfung von realen Fehlern zu schließen.

Deshalb können realistische Fehler zur Bewertung herangezogen werden. Bei realistischen Fehlern handelt es sich, um reale Materialfehler, die durch absichtliche Abweichungen von optimalen Produktionsparametern, z.B. Temperaturabweichung beim Schweißen, hergestellt werden. Die Kosten für realistische Fehler sind jedoch höher als für künstlichen Fehler und können nur zu einem beschränkten Maß hergestellt werden. Dennoch ist die Aussagekraft für die später in der Produktion auftretenden Fehler deutlich höher.

Es muss somit eine Zwischenlösung gesucht werden, in der die Informationen von beiden Prüfungen kombiniert werden. Leider handelt es sich um verschiedenartige Prüfungen und somit ist eine reine Mischung der Daten nicht zielführend.

Ein Ansatz beide Daten miteinander zu kombinieren, ist der Weg über die Bayes'sche Statistik. Hierbei werden die künstlichen Fehlerprüfungen als a priori Wissen, so zu sagen als Zusatzinformation, für die Experimente (Likelihood Funktion) bereitgestellt. In diesem Artikel soll dies an einem einfachen Beispiel gezeigt werden. Um das Problem möglichst einfach zu halten, wurde mit gleich vielen Daten gearbeitet, wobei sich Daten aus der Fehlerprüfung mit künstlichen Daten und der der realen Daten sehr ähnlich verhalten. Hierdurch wurde eine Gewichtung der Daten vermieden.

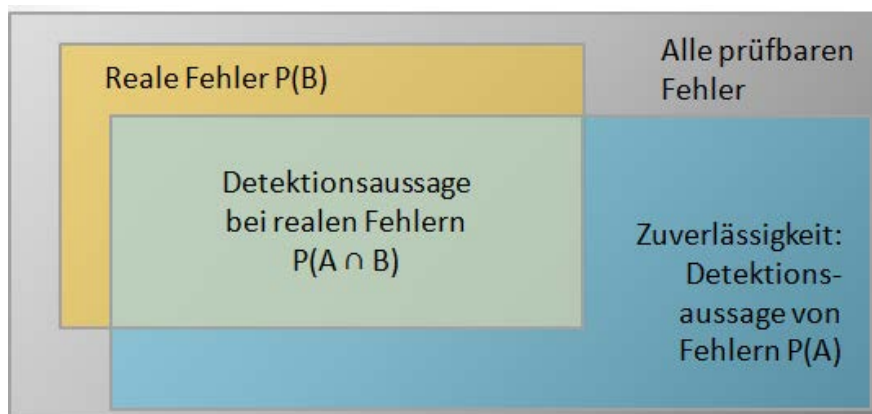
Bei der Darstellung der die Grundlagen der mathematischen Theorie der Bayes'schen Statistik soll hier auf das Nötigste beschränkt werden. Dennoch sind im Folgenden die Grundlagen der Bayes'schen Theorems aufgezeigt: Das Bayes'sche Theorem beschreibt den Zusammenhang zweier bedingter Wahrscheinlichkeiten [7]. Ist die Wahrscheinlichkeit von B unter der Bedingung A („ $P(B/A)$ “) bekannt, so kann mit Hilfe der Wahrscheinlichkeit A („ $P(A)$ “) eine Aussage über die Wahrscheinlichkeit von A unter der Bedingung B („ $P(A/B)$ “) getroffen werden.

$$P(A|B) \sim P(B|A) \cdot P(A)$$

Abbildung 6: Theorem von Bayes

Speziell auf den Fall der Prüfung von realen Fehlern übertragen, ist die Likelihoodfunktion, die Wahrscheinlichkeit, dass ein realer Fehler vorhanden ist, unter der Bedingung, dass er detektiert wurde ( $P(B/A)$ ). Die A priori Information ist die Detektionswahrscheinlichkeit von Fehlern des Prüfsystems  $P(A)$ . Über diese Wahrscheinlichkeit wurden schon mit Hilfe von Referenzprüfungen Erkenntnisse erlangt. Die Qualität des Systems, das bewertet werden soll, kann übersetzt werden als die

Wahrscheinlichkeit, dass es sich um eine Detektion handelt, unter der Bedingung, dass ein realer Fehler vorhanden ist  $P(A/B)$ .



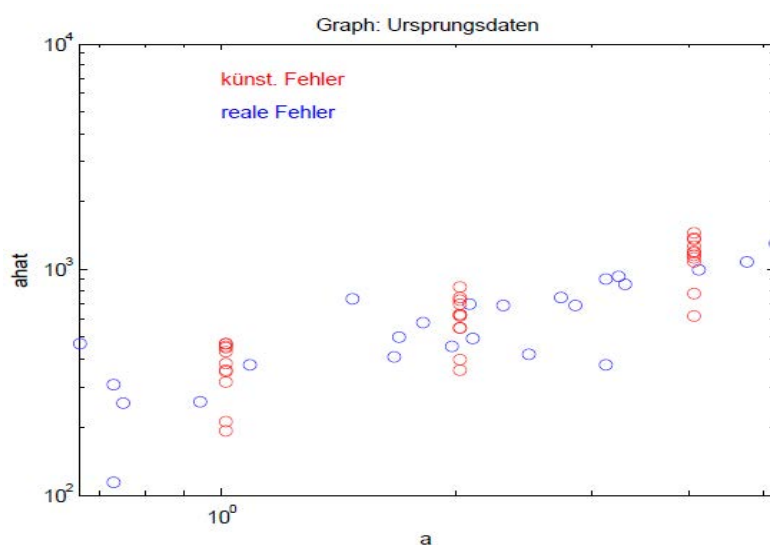
**Abbildung 7:** Bedingte Wahrscheinlichkeiten und Theorem von Bayes in der zerstörungsfreien Prüfung

Diese Theorie soll auf einen konkreten Anwendungsfall übertragen werden.

## 5. Konkreter Anwendungsfall

Im konkreten Beispiel werden Radiographiedaten von künstlichen und realen Fehler untersucht. Bei der generellen Überprüfung des Vorgehens werden folgende Anforderungen gestellt:

- Für die Prüfergebnisse der künstlichen Fehler kann eine Probability of Detection aufgestellt werden. Diese POD besteht aus einer Log - Normalverteilung.
- Es existiert eine a priori Log-Normal-Verteilung mit Mittelwert und Standardabweichung, die eine Aussage über die Zuverlässigkeit der Prüfung zulässt.
- Es gibt eine begrenzte Anzahl von Prüfergebnissen realer Fehler (Likelihoodfunktion), aus denen auch eine POD auf der Grundlage der Log- Normalverteilung erstellt werden kann.



**Abbildung 8:** Datenmenge der künstlichen und realen Fehler

Aus den künstlichen Fehlerdaten, die das apriori Wissen bilden, ergibt sich, dass eine Aussage über Anstieg (*Absorptionsgesetz*) und Streuung (*apparativ bedingt*) des

Defektkontrastes als Funktion der durchstrahlten Länge getroffen wird. Diese Informationen widerspiegeln die grundlegende Natur des Verfahrens und rechtfertigen die Mischung der Daten.

Aus mathematischer Sicht werden die Wahrscheinlichkeitsdichteverteilungen (PDF) der a priori Funktion (rot) und der Likelihood Funktion (blau) mit Hilfe der numerischen Multiplikation durchgeführt (Abbildung 9). Die resultierende a posteriori Funktion (grün) muss im darauffolgenden Schritt normiert werden, um eine kumulierte Gesamtwahrscheinlichkeit von 100% zu erreichen.

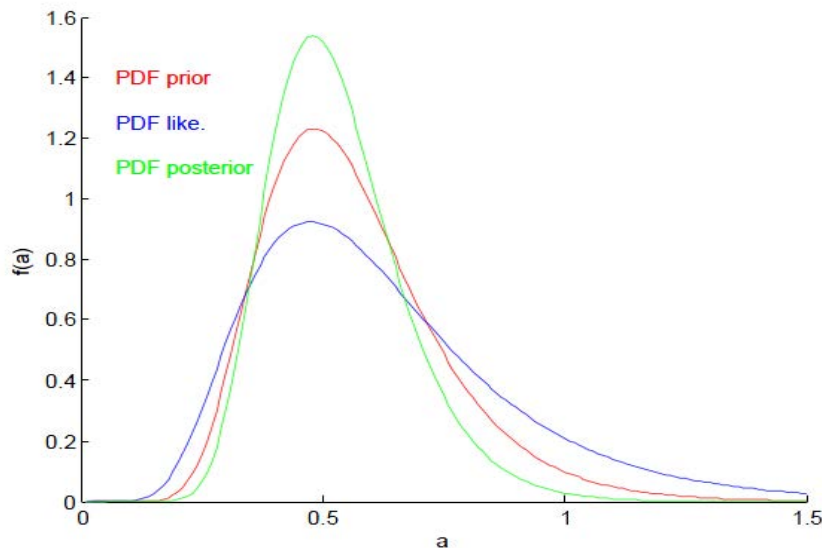


Abbildung 9: Dichtefunktionen der Bayes-Bestandteile

Die a posteriori Wahrscheinlichkeitsdichtefunktion wird als Grundlage der POD genutzt. Im direkten Vergleich zwischen den Daten der POD aus realen Fehlerdaten und der POD auf Basis des Bayes-Ansatzes, kann erkannt werden, dass durch die größere Anzahl der Werte bzw. durch die physikalische Information aus den künstlichen Fehlern die POD-Kurve steiler wird. Die Ungenauigkeit der kleineren Datenanzahl konnte somit verringert werden.

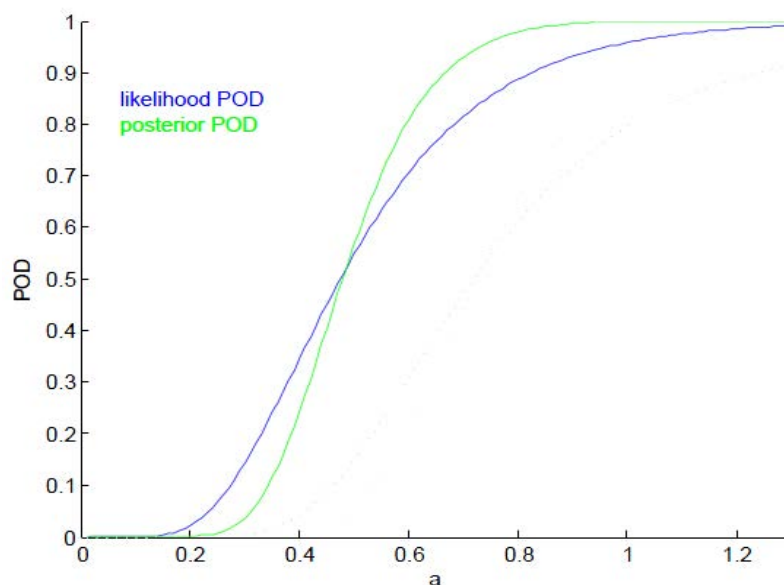


Abbildung 10: Vergleich der PODs von reinen realen Daten und kombinierten Daten



## 6. Zusammenfassung

Die POD ist ein wichtiges und mächtiges Werkzeug der Zuverlässigkeitsbestimmung von ZfP-Systemen. Für die Erstellung von PODs mit realen und realistischen Fehlern ist jedoch meist die statistische Grundlage nicht gegeben. Es ist somit erstrebenswert Daten aus Prüfungen mit realen Fehlern durch Referenzprüfungen zu untermauern. Eine Möglichkeit der Kombination von künstlichen Daten und von realen / realistischen Daten liefert der Einsatz der Bayes'schen Statistik. Anhand eines einfachen Beispiels konnte gezeigt werden, dass die Ungenauigkeit, die von der Anzahl der Werte stammt verringert wurde. Auf dieser Basis werden in der Zukunft komplexere Beispiele untersucht. Generell konnte jedoch schon hier gezeigt werden, dass die Bayes-Statistik auch in der Zuverlässigkeitsbestimmung Einzug erhalten kann. Jedoch hängt das Gesamtergebnis von der Likelihoodfunktion und der A priori Funktion ab. Es ist somit unerlässlich, dass das benutzte Bayes-Apriori-Modell kritisch hinterfragt wird, um kein falsches Ergebnis zubekommen.

In der Zukunft sind einerseits komplexere Beispiel geplant. Desweiteren ein mathematisch komplexer Ansatz, um die unterschiedliche Gewichtung von Daten in das System einzuführen. Außerdem eine Erweiterung auf andere ZfP-Verfahren die möglicherweise ohne einen Multiparameter Ansatz nicht zu lösen sind. Als weiterer unerlässlicher Punkt ist die Einführung verschiedener Test geplant, um eine korrekte Aussage der a priori Funktion zu gewährleisten.

## References

- [1] Müller C, Scharmach M, Konchina Vera, Piscenec Z, Markucic D, "General Principles Of Reliability Assessment Of Nondestructive Diagnostic Systems And Its Applicability To The Demining Problem" *8th ECNDT*, Barcelona, June 2002
- [2] Singh, R. inc., K. T. (Ed.) "Three Decades of Reliability Assessment" *Report No. Karta-3510-99-01*, 2000
- [3] Müller, C., Pavlovic, K. Takahashi, U. Ewert, M. Rosenthal, G. Brekow, R. Böhm, U. Ronneteg, J. Pitkänen, 2007, POD Evaluation of NDE Techniques for Canister-Components for Risk Assessment of Nuclear Waste Encapsulation, *6th International conference on NDE in Relation to Structural integrity for Nuclear and Pressurized Components*, Budapest, 2007
- [4] Berens A P, NDE Reliability Data Analysis – Metals Handbook, Volume 17, 9th Edition: *Nondestructive Evaluation and Quality Control*, ASM International, OH, 1989
- [5] D. Kanzler, S. Milsch, C. Müller, M. Pavlovic, J. Pitkänen „Concept of Total Reliability of NDT Methods for Inspection of the EB Weld of Copper Canister Used for a Long-Term Storage of Spent Nuclear Fuel“, *8th International Conference on NDE in Relation to Structural Integrity for Nuclear and Pressurised Components*, Berlin, 2010
- [6] Müller, C., Elagin, M., Scharmach, M., Bellon, C., Jaenisch, G.-R., Bär, S., Redmer, B., Goebbels, J., Ewert, U., Zscherpel, U., Böhm, R., Brekow, G., Erhard, A., Heckel, T., Tessaro, U., Tscharnatke D. & Ronneteg, U., 2006, Reliability of nondestructive testing (NDT) of the copper canister seal weld, *Svensk kärnbränslehantering AB report R-06-08*, 158p.
- [7] Yudkowsky E., 2011 An intuitive Explanation of Bayes' Theorem