

# Human Factors in der POD - ist das möglich?

Marija BERTOVIĆ<sup>1</sup>, Joseph GIVEN<sup>1</sup>, Vamsi Krishna RENTALA<sup>2</sup>, Johannes LEHLEITNER<sup>1</sup>, Daniel KANZLER<sup>2</sup>, Thomas HECKEL<sup>1</sup>, Viktoriya Tkachenko<sup>1</sup>

<sup>1</sup> Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin

<sup>2</sup> Applied Validation of NDT, Berlin

Kontakt E-Mail: marija.bertovic@bam.de

**Kurzfassung.** Menschliche Faktoren sind ein häufig genanntes Thema, wenn wir über die Zuverlässigkeit der zerstörungsfreien Prüfung (ZfP) sprechen. Die Probability of Detection (POD), das üblicherweise verwendete Maß für die Zuverlässigkeit der ZfP, betrachtet jedoch nur die technische Fähigkeit eines ZfP-Systems, einen Fehler zu entdecken. Nach mehreren Jahrzehnten der Erforschung des Einflusses der menschlichen Faktoren auf die Zuverlässigkeit der ZfP gibt es immer noch keinen allgemein akzeptierten Ansatz, um menschliche Faktoren in der Zuverlässigkeitsbewertung sichtbar zu machen. Dieser Beitrag gibt einen Überblick über verschiedene verfügbare Methoden zur Einbeziehung menschlicher Faktoren in die Zuverlässigkeitsbewertung. Das Thema ist ein essenzieller Bestandteil des laufenden WIPANO-Projekts "normPOD", das sich als Ziel gesetzt hat, die Standardisierung der Zuverlässigkeitsbewertung in Deutschland voranzutreiben und im Vergleich zu den schon bekannten internationalen Normen einen Fokus auf den Umgang mit den menschlichen Faktoren zu legen.

## Einführung

Menschliche Faktoren (Human Factors, HF) sind zu einem unvermeidlichen Thema geworden, wenn wir über die Zuverlässigkeit der zerstörungsfreien Prüfung sprechen. Das überrascht nicht, denn die Zuverlässigkeit von ZfP-Systemen wird bekanntermaßen durch die intrinsische Fähigkeit, Anwendungsfaktoren, menschliche Faktoren und den organisatorischen Kontext beeinflusst [1](**Abb. 1**). Dabei beziehen sich die menschlichen Faktoren auf Faktoren aus der Umwelt, der Organisation und der Arbeit wie auch menschliche oder individuelle Charakteristika, die einen sicherheits- und gesundheitsrelevanten Einfluss auf das Arbeitsverhalten haben.“ [2]. Der organisatorische Kontext bezieht sich auf die Betriebsanweisungen, das Management, die Normen und die industrielle Beziehungen der Organisation, für die die ZfP Dienstleistungen erbringt [3].

HF haben immer noch den Ruf, ein schlecht erforschtes Thema zu sein. In der Tat gab es in den letzten 50 Jahren nur ein Dutzend Institute, die auf diesem Gebiet geforscht haben. Wie dem auch sei, durch diese Studien wurden viele Erkenntnisse gewonnen, aber ihre Ergebnisse oder Überlegungen sind der breiten ZfP-Öffentlichkeit nach wie vor relativ unbekannt. Die Arbeit des letzten Jahrzehnts an Instituten wie der BAM, dem Electric Power Research Institute (EPRI) und der U.S. Nuclear Regulatory Commission (NRC) zeigt unbestreitbar, dass menschliche Faktoren aus einer Vielzahl von Faktoren bestehen (**Abb. 2**), die einen Einfluss auf die ZfP-Leistung haben können [4]–[6].



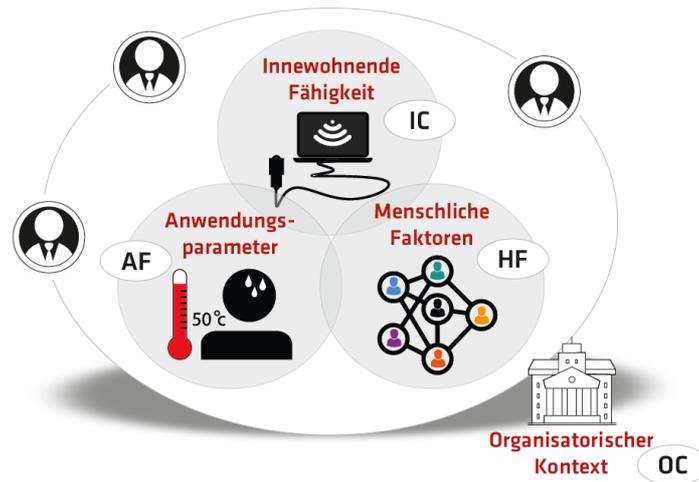


Abb. 1. Modulares Zuverlässigkeitsmodell (nach Müller et al., 2013 [1])

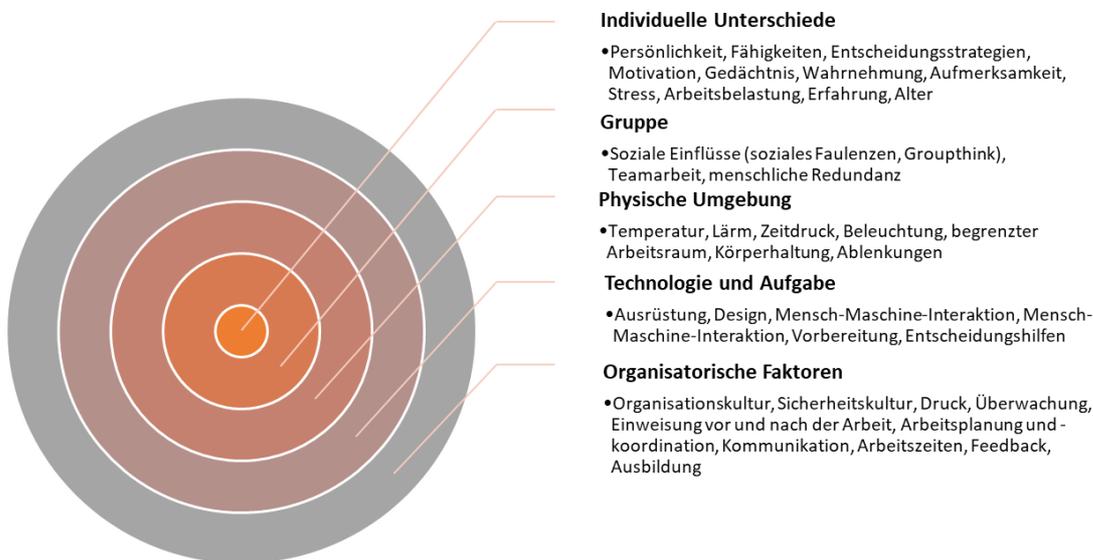


Abb. 2. Sozio-technisches Systemmodell der HF-Einflüsse auf die ZfP-Leistung. Die aufgeführten Beispiele umfassen Ergebnisse empirischer Studien oder theoretische und praktische Überlegungen zu den HF-Einflüssen.

Obwohl schon immer bekannt war, dass HF eine entscheidende Rolle für die Zuverlässigkeit der zerstörungsfreien Prüfung spielen, wurden sie unabhängig von der technischen Zuverlässigkeit bewertet, untersucht und beurteilt. Gleichzeitig wurden HF in den Berichten über die technische Zuverlässigkeit nicht berücksichtigt. Der Grund dafür war, dass es fast unmöglich ist, HF zu quantifizieren und daher in die Bewertung der Fehlerauffindwahrscheinlichkeit (Probability of Detection, POD) einzubeziehen. Neuere Versuche von Spies & Rieder [7] und Bato et al. [8] sind vielversprechend, dass solche Versuche unternommen werden könnten.

Eine mögliche Quantifizierung von HF und deren Integration in die Zuverlässigkeitsbewertung wird derzeit im Rahmen des vom Bundesministerium für Wirtschaft und Klimaschutz (BMWK) geförderten WIPANO-Projektes (Förderkennzeichen 03TN0006A) unter dem Namen "Normung zur probabilistischen Bewertung der Zuverlässigkeit für zerstörungsfreie Prüfverfahren" (normPOD) diskutiert [9]. Ziel dieses laufenden Verbundprojektes ist es, einen Vorschlag für eine Normung der

Zuverlässigkeitsbewertung zu erarbeiten. Erstmals werden dabei auch HF und deren Einfluss auf die Zuverlässigkeit berücksichtigt.

Nach mehreren Jahrzehnten der Forschung über den HF-Einfluss auf die ZfP-Zuverlässigkeit gibt es immer noch keinen allgemein akzeptierten Ansatz, um HF in der Zuverlässigkeitsbewertung sichtbar zu machen. Ziel dieses Beitrags ist es, verschiedene mögliche Methoden zur Quantifizierung von HF und deren Integration in die Zuverlässigkeitsbewertung hinsichtlich ihrer Durchführbarkeit, Vor- und Nachteile zu bewerten und die Frage zu beantworten, ob es möglich ist, HF in die POD einzubeziehen.

## 1. Quantifizierung menschlicher Faktoren und menschlicher Fehler

Der erste Schritt zur Beantwortung der Frage, ob HF in den POD integriert werden kann, besteht darin, herauszufinden, ob und wie HF quantifiziert werden können. Dazu ist es wichtig, den Unterschied zwischen menschlichen Faktoren (human factors) und menschlicher Fehler (human error) zu kennen.

### 1.1. Definition der Begriffe

Wie bereits erwähnt, gehören zu den *menschlichen Faktoren (HF)* eine Vielzahl von Faktoren (individuelle und gruppenspezifische Merkmale, physische Umgebung, Technologie, Aufgabeneigenschaften und organisatorische Faktoren), die sich auf den Menschen und damit auf seine Leistung auswirken. Auch wenn HF oft mit Fehlern gleichgesetzt werden, wirken sie sich nicht unbedingt negativ auf die Leistung aus. Landung des US-Airways-Flugs 1549 im Hudson River [10] und die Rettung von 150 Passagieren durch Missachtung von Regeln und Befehlen ist ein bekanntes Beispiel dafür, wie sich HF positiv auf die Leistung und damit auf Zuverlässigkeit und Sicherheit auswirken können. Schließlich wird die zerstörungsfreie Prüfung seit Jahrzehnten eingesetzt, um die Sicherheit der Strukturen zu gewährleisten, für die sie konzipiert wurde, und sie leistet dabei großartige Arbeit, größtenteils dank der positiven Auswirkungen von HF.

*Menschliche Fehler* werden definiert als "alle Fälle, in denen eine geplante Abfolge geistiger oder körperlicher Aktivitäten ihr beabsichtigtes Ergebnis nicht erreicht, und wenn dieses Versagen nicht auf das Eingreifen einer zufälligen Instanz zurückgeführt werden kann" [11] (p. 9). Vereinfacht ausgedrückt: Wenn eine Handlung ihr beabsichtigtes Ergebnis nicht erreicht, sprechen wir von menschlichen Fehlern. Menschliche Fehler beziehen sich in der Regel auf Fehler, Ausrutscher und Entgleisungen (Engl. *mistakes slips & lapses*). Wir verwenden den Begriff menschlicher Fehler, um menschliches Handeln oder ein Ereignis zu erklären, das sich ereignet hat (beobachtbares Versagen), oder als Ursache für ein Missgeschick oder einen Unfall. Darüber hinaus ist menschlicher Fehler, auch wenn es in den Händen des Bedieners liegt, bekanntlich nur ein Symptom für Probleme, die tiefer im System liegen (fehlerhafte Verfahren, nicht verwendbare oder defekte Ausrüstung, Fehlkommunikation usw.) [12]–[14], d. h. die HF.

### 1.2. Methoden zur Quantifizierung menschlicher Faktoren und menschlicher Fehler

HF ist ein Studienbereich, der sich hauptsächlich mit dem Verständnis der verschiedenen Einflüsse befasst, die negative Auswirkungen auf die Zuverlässigkeit und Sicherheit von Organisationen haben können, um diese Auswirkungen zu verhindern. Andererseits sind Organisationen mit hohen Zuverlässigkeits- und Sicherheitsanforderungen (Kernkraftwerke, Eisenbahn, Luft- und Raumfahrt, chemische Industrie usw.) nicht nur an der Art der Einflüsse, sondern auch an probabilistischen Schätzungen der Zuverlässigkeit und Sicherheit interessiert. Die Nuklearindustrie verwendet beispielsweise die probabilistische

Sicherheitsbewertung (*Probabilistic Safety Assessment*, PSA). Ein Teil der PSA besteht darin, menschliche Handlungen zu untersuchen, die sich auf die Risikowahrscheinlichkeit, die Wahrscheinlichkeit signifikanter Ereignisse und deren Folgen auswirken [15]. Dies geschieht mit Hilfe einer Human Reliability Analysis (HRA) [16]. Im Folgenden wird zwischen quantitativen und qualitativen Methoden zur Untersuchung von HF und der Quantifizierung menschlicher Fehlerwahrscheinlichkeiten mit Hilfe der HRA unterschieden. Darüber hinaus bietet die Praxis des Risikomanagements [17] eine Vielzahl von Methoden zur Risikobewertung an, von denen einige semi-quantifizierbare Methoden verwenden, z. B. die menschenorientierte Fehlzustandsart- und -auswirkungsanalyse (FMEA)[18].

### 1.2.1. *Quantitative vs. qualitative Studien*

Nach dem Wörterbuch der American Psychological Association (APA) beruht die *quantitative* Forschung auf der Messung von Variablen mit Hilfe eines numerischen Systems, der Analyse dieser Messungen mit Hilfe einer Vielzahl von statistischen Modellen und der Darstellung von Beziehungen und Assoziationen zwischen den untersuchten Variablen. Bei diesen Variablen kann es sich z. B. um Testergebnisse oder Messungen der Reaktionszeit handeln. Das Ziel ist es, die Natur eines Phänomens zu verstehen, zu beschreiben und vorherzusagen, insbesondere durch die Entwicklung von Modellen und Theorien. Beispiele hierfür sind Experimente und Umfragen. [19].

*Qualitative* Forschung liefert deskriptive (nicht numerische) Daten, z. B. Verhaltensbeobachtungen oder persönliche Erfahrungsberichte. Ziel ist es, zu untersuchen, wie Einzelpersonen die Welt aus verschiedenen Blickwinkeln wahrnehmen können. Beispiele hierfür sind Tiefeninterviews, Fokusgruppen, teilnehmende Beobachtung und Fallstudien [20].

Die Auswirkungen von HF auf die Leistung im Allgemeinen lassen sich am besten durch die Überprüfung von Hypothesen anhand von Experimenten untersuchen. Das bedeutet, dass der Einfluss bestimmter Faktoren, z. B. Erfahrung, Zeitdruck, Temperatur, Lärm, Körperhaltung, in einem kontrollierten Experiment variiert wird, während alle anderen Faktoren konstant gehalten werden. Das Ergebnis einer solchen Studie ist die Wahrscheinlichkeit, dass dieser Effekt mit 95%iger Sicherheit ( $p < 0,05$ ) in der Population vorhanden ist. Der Vorteil dieser Methode ist der eindeutige Zusammenhang zwischen Ursache und Wirkung.

Alternativ kann man eine Korrelationsstudie durchführen, in der man die Beziehung zwischen zwei oder mehreren Variablen untersucht. Korrelationsstudien werden häufig mit Hilfe psychometrischer Instrumente (z. B. Persönlichkeits- oder Fähigkeitsskalen) durchgeführt und die Ergebnisse z. B. mit dem Ergebnis einer Prüfung in Beziehung gesetzt. Es wird erwartet, dass ein Anstieg einer Variablen (z. B. Erfahrung) zu einem Anstieg oder Rückgang der anderen Variablen (z. B. Leistung) führt. Der Nachteil dieser Methode ist, dass wir die Ursache-Wirkungs-Beziehung nicht beurteilen können; wir wissen nicht, ob eine Variable die andere verursacht hat, sondern nur, dass sie miteinander in Beziehung stehen.

Studien, die nur eine kleine Anzahl von Teilnehmern enthalten (wie häufig bei HF in der ZfP-Forschung), können ebenfalls als qualitativ angesehen werden, da sie eher der Beschreibung der Auswirkungen als deren Quantifizierung dienen.

### 1.2.2. *Analyse der menschlichen Zuverlässigkeit*

Die Human Reliability Analysis (HRA) ist die am häufigsten angewandte Technik zur Abschätzung menschlicher Fehler [16]. Sie wurde für die Zwecke der PSA in Kernkraftwerken entwickelt, hat aber seither eine breite Anwendung gefunden. Sie quantifiziert die Wahrscheinlichkeit des Auftretens eines menschlichen Fehlers bei einem

Vorgang innerhalb eines komplexen Systems zu einem bestimmten Zeitpunkt unter gegebenen Umständen [21]. Viele HRA-Techniken, wie die Technique for Human Error Rate Prediction (THERP) [16] oder die Cognitive Reliability and Error Analysis Method (CREAM) [22] schätzen menschliche Fehlerwahrscheinlichkeiten (Human Error Probability, HEP) für Teilaufgaben des Systems. Die HEPs werden in der Regel von Experten (z. B. Betriebsleitern) geschätzt, die tatsächliche Betriebsdaten und zuvor aufgetretene menschliche Fehlerereignisse analysieren. HEPs können durch leistungsbeeinflussende Faktoren (Performance Shaping Factors, PSF) positiv oder negativ beeinflusst werden, z. B. durch Beleuchtung, Temperatur, Körperhaltung, Ermüdung und Zeitdruck. Obwohl HRA-Techniken noch keine validen und zuverlässigen Quantifizierungsmethoden für kognitive oder organisatorische Faktoren bieten und stark auf Expertenurteile angewiesen sind, bieten sie die Möglichkeit, die menschliche Zuverlässigkeit in komplexen soziotechnischen Systemen systematisch zu bewerten und abzuschätzen.

### 1.2.3. Menschenorientierte FMEA

Die Fehlzustandsart- und -auswirkungsanalyse (*Failure Modes and Effects Analysis*, FMEA) ist ein Standardinstrument zur Risikobewertung. Es wurde ursprünglich von den US-Armed Forces im Jahr 1949 entwickelt und 1980 überarbeitet (MIL-STD 1629A). Ursprünglich zur Ermittlung von Fehlern und Fehlermodi technischer Systeme konzipiert, hat sie im Laufe der Zeit verschiedene Verwendungszwecke gefunden, darunter auch die Ermittlung von Fehlermodi und -effekten beim Menschen. Es wird eingesetzt, um potenzielle Fehlerarten (die Art und Weise, wie ein Fehler auftritt) zu ermitteln, ihre Ursachen und Auswirkungen (Folgen) auf den Betrieb des Systems zu bestimmen, die Fehlermechanismen zu identifizieren und Maßnahmen zur Vermeidung und/oder Abschwächung der Auswirkungen des Fehlers auf das System zu ermitteln. Im Mittelpunkt dieses Prozesses steht die Vorhersage dessen, was versagen oder schief gehen könnte. Menschenorientierte FMEA [24] kann zu Verbesserungen auf persönlicher, ergonomischer (optimierte Arbeitsbedingungen) und organisatorischer Ebene (optimiertes Zusammenspiel von Mensch, Technik, Material und Methode) führen, was eine erhebliche Verringerung von Systemausfällen zur Folge haben kann. Auf die FMEA kann eine Kritikalitätsanalyse folgen, die jedem Fehlermodus eine Bedeutung zuweist (IEC/ISO 31010:2009). Die Risikoprioritätszahl (*Risk Priority Number*, RPN) ist eine halbquantitative Methode zur Bestimmung der Kritikalität, bei der jedem Fehlermodus ein numerischer Wert in Bezug auf seine Schwere (oder Relevanz), sein Auftreten und seine Entdeckungswahrscheinlichkeit zugewiesen wird. Da die RPN durch Multiplikation dieser drei Werte ermittelt wird, ist das Ergebnis eine Bewertung der Kritikalität des Systems, wobei eine höhere RPN mit den risikoreichsten Elementen des Prozesses verbunden ist. Die menschenzentrierte FMEA wurde bereits mehrfach eingesetzt, um Risiken in der ZfP zu identifizieren und zu priorisieren [25], [26].

## 2. Menschliche Faktoren und die Bewertung der Zuverlässigkeit

### 2.1. Frühere Versuche, menschliche Faktoren in MAPOD zu berücksichtigen

Die Vernachlässigung von HF bei der POD-Modellierung (*model-assisted POD*, MAPOD) wurde als eine der größten Einschränkungen von Modellierungsansätzen identifiziert [27]–[29]. Infolgedessen gab es mehrere Versuche, HF in die MAPOD einzubeziehen. Im Folgenden werden einige Ansätze zur Einbeziehung von HF in die MAPOD aufgeführt:

- Wall et al. [30], [31] haben vorgeschlagen, menschliche Faktoren über eine Transferfunktion in die POD einzubeziehen oder Expertenurteile zu verwenden, um geeignete POD-Daten für die Integritätsbewertung zu ermitteln.
- Harding et al. [32] beziehen menschliche Einflüsse in ihre POD-Modelle ein, indem sie davon ausgehen, dass der Schwellenwert für die Beobachtung eines Fehlersignals durch eine Zufallsvariable mit einer zu charakterisierenden Verteilung dargestellt werden kann. Wenn die Anzahl der vom Prüfer nicht beobachteten Fehler bekannt ist, kann eine so genannte Beobachtungswahrscheinlichkeit unter Verwendung der Hit-Miss-Information berechnet werden. Dieser Ansatz gilt für Verfahren, bei denen der HF-Einfluss auf die Interpretation der erfassten Daten im Vordergrund steht und eignet sich daher eher für die automatisierte UT-Prüfung.
- Meyer et al. [33] werteten die MAPOD- und HF-Literatur aus und schlugen die folgenden Möglichkeiten vor, wie HF in POD-Modelle für MAPOD-Anwendungen integriert werden kann: a) Verwendung von HF als zusätzlicher Parameter im Multiparameter-POD, b) Verwendung von HF als Zufallsvariable im POD-Modell und/oder c) Verwendung von HF bei der Definition des Schwellenwerts für die Fehlererkennung.
- Dominguez et al. [27] schlugen die Verwendung eines operationellen ZfP-Simulators vor, der es ermöglicht, Prüfscenarien in realen ZfP-Konfigurationen, mit realen Teilen, in der realen Umgebung und von realen Prüfern durchzuführen, jedoch mit synthetisierten Daten (die simuliert, experimentell gemessen oder aus beiden Arten von Daten zusammengesetzt sein können). Auch wenn dieser Ansatz nicht die gesamte mögliche Variation zwischen einzelnen Personen berücksichtigt, ist er doch wertvoll, weil er die tatsächliche Variation zwischen Personen beim Bewegen des Messwertaufnehmers, beim Lesen und Interpretieren der Signale auf dem Bildschirm und beim Treffen von Entscheidungen einbezieht. Diese Art von Betriebssimulator kann für POD-Studien, aber auch für die Ausbildung von Personal unter realen Bedingungen verwendet werden.
- Spies und Rieder [7] verwendeten empirische Daten aus der Studie "Programme for the Inspection of Steel Components (PISC 3) [34] um die Variabilität der Inspektoren in einem praktischen Versuch abzuschätzen. In dieser Studie erreichten die Teilnehmer eine durchschnittliche Wirksamkeit von 80 % bei der Erkennung von Ermüdungsrissen. Die Autoren nutzten diese Information, um 18,2 % der zufällig ausgewählten Amplituden unter den Geräuschpegel zu senken. Der Vergleich der  $a_{90/95}$ -Werte zwischen MAPOD unter Berücksichtigung menschlichen Versagens und einem Standard-MAPOD deutet auf eine Erhöhung von  $a_{90/95}$  um etwa 0,5 mm (ca. 13 %) zugunsten der Berücksichtigung von HF hin, was dies zu einem konservativeren Ansatz macht.
- Bato et al. [8] verwendeten Daten aus anderen Studien zu ausgewählten Fähigkeiten (Erfahrung, Zertifizierungsniveau, mechanische Kenntnisse und Rechts-/Linkshänder) sowie Daten aus ihrer eigenen internen Erhebung, um die menschliche Variabilität in ihre Modellierung einzubeziehen. Darüber hinaus simulierten sie drei verschiedene Arbeitsumgebungen (ein "Labor" und zwei "Flugzeugszenarien") und zeichneten die Prüfkopfbewegungen der Prüfer auf. Sie beobachteten einen Zusammenhang zwischen den Sondenbewegungen und den Ergebnissen der Inspektoren.

## 2.2. Vorgeschlagene quantifizierbare Methoden zur Berücksichtigung von HF bei der Zuverlässigkeitsbewertung

Tabelle 1 listet verschiedene mögliche Methoden zur Einbindung von HF in den POD auf und beschreibt sie im Hinblick auf ihren Output für den POD, ihre Vor- und Nachteile. Beachten Sie, dass die meisten Überlegungen in der folgenden Tabelle zusammen mit Martin Wall für die Zwecke des „Guidance on Human Factors for Model-Assisted Qualification (MAQ) Protocol“ (Anhang C) angestellt wurden [35].

**Tabelle 1.** Vor- und Nachteile verschiedener quantifizierbarer Methoden zur Berücksichtigung von HF bei der Zuverlässigkeitsbewertung

Methode	Beschreibung	Vorteile	Nachteile
1. Berücksichtigung der Variabilität der individuellen Leistung, die in der vorherigen HF in ZfP-Studien ermittelt wurde	a) Aus den Rohdaten etwas über die durchschnittliche Erkennbarkeit lernen und diese Information als Parameter in der Simulation verwenden oder eine ausgewählte Anzahl von Amplituden unter den Rauschlevel senken.	<ul style="list-style-type: none"> <li>▪ Einbeziehung einer gewissen Variation der menschlichen Leistung in die POD</li> <li>▪ Die Informationen sind in dem Wert<sub>90/95</sub> enthalten.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Zu wenige Studien über menschliche Faktoren in der ZfP verfügbar</li> <li>▪ Die Unterschiede in der ZfP-Leistung aus diesen Studien könnten auf eine Vielzahl von Faktoren zurückzuführen sein, die möglicherweise nicht den gleichen Einfluss auf die Variation in der Leistung in dem spezifischen untersuchten Fall haben</li> <li>▪ Erfasst nicht alle relevanten leistungsbeeinflussenden Faktoren (z. B. Organisation, Umfeld)</li> </ul>
	b) Erfassung verfügbarer Daten aus verschiedenen Studien oder (wenn keine vollständigen Daten verfügbar sind) Verwendung von Mittelwerten, Standardabweichungen usw., um Datenverteilungen zu randomisieren. Die Simulation/Interpolation kann auch verwendet werden, um fehlende Daten zu ergänzen. Auf diese Weise können entweder Verteilungen für eine Reihe von Variablen oder eine Zusammenfassung (nach Vorverarbeitung) erstellt und eine oder eine begrenzte Anzahl von Variablen als Verteilung(en) zur Verwendung in der POD bereitgestellt werden.	<ul style="list-style-type: none"> <li>▪ Verwendung echter Ergebnisse, die für die - begrenzte ZfP-Population relevant sind.</li> <li>▪ Simulationen sind kosteneffizient und berücksichtigen kleine Populationen, die sonst nicht getestet werden könnten.</li> <li>▪ Die Informationen sind in dem Wert<sub>90/95</sub> enthalten.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Künstliche Daten, selbst wenn sie von tatsächlichen Daten abgeleitet sind, unterliegen Beschränkungen und können niemals schlüssige Ergebnisse liefern.</li> <li>▪ Es gibt nur wenige Studien über menschliche Faktoren in der zerstörungsfreien Prüfung mit brauchbaren oder aussagekräftigen Ergebnissen.</li> <li>▪ Die Verwendung von Variablen kann auf die Bereiche und Bedingungen der jeweiligen Studien beschränkt sein.</li> <li>▪ Einzelne Maßnahmen können niemals ein vollständiges Bild ergeben.</li> </ul>
2. Berücksichtigung der Variabilität der individuellen Leistung, die in den vorangegangenen HF in ZfP-Studien ermittelt wurde + Expertenbewertungen	Erstellung einer Liste bekannter Einflüsse und Aufforderung an Experten, diese nach ihrem Schweregrad einzustufen. Die daraus resultierenden Rangfolgen können dann in statistischen Modellen (deterministische und probabilistische - modellbasierte Verteilungen) verwendet werden.	<ul style="list-style-type: none"> <li>▪ Methode zur Datenerhebung mit geringem Aufwand.</li> <li>▪ Leicht zu bewerten und umzusetzen.</li> <li>▪ Die sich daraus ergebenden Verhältnisse könnten in POD-Modellen verwendet werden.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Abhängig von Selbstauskünften (Gefahr der menschlichen Voreingenommenheit, i.e. <i>Bias</i>).</li> <li>▪ Liefert keine direkten Informationen über Abhängigkeiten zwischen den Daten (obwohl Korrelationen hilfreich sein können).</li> <li>▪ Es handelt sich um Intervalldaten und nicht um kontinuierliche Daten; daher wären große Teilnehmer Stichproblem ratsam.</li> </ul>

Methode	Beschreibung	Vorteile	Nachteile
<b>3. Berücksichtigung der Variabilität der individuellen Leistung, die in anderen (nicht-ZfP) Psychologie-Studien ermittelt wurde</b>	a) Wie 1b, aber Verwendung von Studien aus Bereichen, die der ZfP nahestehen, wie z. B. CT, MRT oder Ultraschall im Gesundheitswesen.	<ul style="list-style-type: none"> <li>▪ Ergänzt seltene Daten aus dem eher begrenzten Bereich der ZfP.</li> <li>▪ Die Informationen können in den Wert<sub>90/95</sub> aufgenommen werden.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Künstliche Daten können keine schlüssigen Ergebnisse liefern.</li> <li>▪ Die Daten stammen von außerhalb des Untersuchungsgebiets, so dass sich die Schwerpunkte auf allgemeinere HF-Effekte beschränken können.</li> <li>▪ Einzelne Maßnahmen können niemals ein vollständiges Bild ergeben.</li> <li>▪ Vor allem in der Medizin sind erhöhte Motivationen / kognitive Stressoren denkbar, so dass diese Bereiche nicht mit der ZfP vergleichbar sind.</li> </ul>
	b) Verwendung von Rohdaten aus Psychologiestudien, in denen die für die ZfP relevanten Eigenschaften (visuelle Suche usw.) untersucht wurden, und Verwendung dieser Daten in der Simulation auf die gleiche Weise wie in 1a beschrieben.	<ul style="list-style-type: none"> <li>▪ Einbeziehung einer gewissen Variation der menschlichen Leistung in der Simulation</li> <li>▪ Die Informationen sind in dem Wert<sub>90/95</sub> enthalten.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Zu viele, manchmal widersprüchliche, psychologische Studien, die die Unterschiede in der menschlichen Leistung bei einer Reihe von verschiedenen Aktivitäten gemessen haben.</li> <li>▪ Es ist schwierig, die für einen bestimmten ZfP-Fall geeigneten Verfahren auszuwählen.</li> <li>▪ Die Unterschiede in der Leistung dieser Studien könnten auf eine Vielzahl von Faktoren zurückzuführen sein, die einen Einfluss auf die Variation in der ZfP-Leistung haben können oder auch nicht.</li> <li>▪ Mangelndes psychologisches Fachwissen derjenigen, die die MAPOD-Berechnungen durchführen</li> <li>▪ Erfasst nicht alle relevanten leistungsbeeinflussenden Faktoren (z. B. Organisation, Umfeld)</li> </ul>
<b>4. Durchführung weiterer HF-Experimente, um mehr Informationen über die HF-Variabilität zu erhalten</b>	Experimentelle Überprüfung der Hypothesen und Verwendung der erhaltenen Daten (Verteilungen) für die POD-Berechnung.	<ul style="list-style-type: none"> <li>▪ Enthält relevante Informationen über HF auf der Grundlage von Hypothesen, die für den genauen Kontext, für den die Zuverlässigkeitsbeurteilung durchgeführt wird, aufgestellt wurden</li> </ul>	<ul style="list-style-type: none"> <li>▪ Kostenaufwendig</li> <li>▪ Unmöglich, die gesamte Variabilität, die in einem bestimmten Kontext auftreten kann, zu erfassen</li> <li>▪ Begrenzt auf eine kleine Anzahl von Einflussfaktoren</li> <li>▪ Kostenaufwendig</li> <li>▪ Unwirksam, wenn nur ein paar Prüfer an den Studien teilnehmen</li> </ul>
<b>5. Durchführung neuer Experimente + Simulationen</b>	Verwendung von Experimenten für einige wenige Parameter, die experimentell überprüft werden müssen, und Simulation für Parameter, die keine weitere experimentelle Validierung erfordern. Verwendung dieser Variation in der Simulation. Sicherstellen, dass die Teilnehmerstichprobe groß genug und repräsentativ für die Population der Prüfer*innen ist (z. B. Unterschiede in Bezug auf Alter, Erfahrung, Qualifikationsniveau usw.).	<ul style="list-style-type: none"> <li>▪ Einbeziehung von anwendungsrelevanten Variationen der menschlichen Leistung in die Modellierung</li> <li>▪ Durchführbar</li> <li>▪ Benötigt nicht viele Prüfgegenstände und Fehler, sondern viele Menschen</li> <li>▪ Die Informationen sind in dem Wert<sub>90/95</sub> enthalten.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Um sicherzustellen, dass die Teilnehmer repräsentativ sind, sind Kenntnisse erforderlich</li> <li>▪ Erfasst nicht alle relevanten leistungsbeeinflussenden Faktoren (z. B. Organisation, Umfeld)</li> </ul>

Methode	Beschreibung	Vorteile	Nachteile
	Durchführung weiterer HF Experimente mittels Simulationen mit Virtual-Reality-Brillen (VR, Deutsch: Virtuelle Realität) einschl. integrierter Blickverfolgung und Pupillometrie. Quantifizierung von Erregung und/oder Konzentration durch Verfolgung des Blicks und Messung der Pupillengröße während der Testaufgaben — auch als Augmented Reality (AR, Deutsch: Augmentierte Realität) möglich.	<ul style="list-style-type: none"> <li>▪ Erstellen von virtuellen Testobjekten mit Materialeigenschaften, um Szenarien ohne aufwändige Herstellung auszuprobieren</li> <li>▪ Direkte quantitative Messung von Erregung / Konzentration im Kontext von Aufgaben.</li> <li>▪ Kann weltweit als eigenständige Software-Anwendung eingesetzt werden, sofern die Teilnehmer Zugang zu einem ähnlichen Gerät haben.</li> <li>▪ Könnte die Teilnehmerzahl stark erhöhen unter ähnlichen Testbedingungen.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Erfordert eine Anfangsinvestition in Hardware und Programmierzeit.</li> <li>▪ Erfordert eine Validierung mit realen Experimenten.</li> </ul>
<b>6. Verwendung von Simulationen</b>	Durchführung kleinerer Messexperimente, Verwendung der Ergebnisse zur weiteren Simulation der menschlichen Variabilität bei Messaufgaben, Überprüfung der experimentellen Ergebnisse mit einer ausreichend großen Stichprobe von Prüfer*innen und Einbeziehung in die POD durch Absenken einer ausgewählten Anzahl von Amplituden unter den Rauschlevel.	<ul style="list-style-type: none"> <li>▪ Einbeziehung von anwendungsrelevanten Variationen der menschlichen Leistung in die Modellierung</li> <li>▪ Durchführbar</li> <li>▪ Benötigt nicht viele Prüfgegenstände und Fehler, sondern viele Menschen</li> <li>▪ Die Informationen sind in dem Wert<sub>90/95</sub> enthalten.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Geringe externe Validität (je nachdem, wie viele Prüfende zur Überprüfung der Ergebnisse herangezogen werden)</li> <li>▪ Erfordert eine größere Anzahl von Prüfpersonal (als normalerweise für die Überprüfung der Ergebnisse eingesetzt wird)</li> </ul>
<b>7. Durchführung von POD-Experimenten mit einer großen Stichprobe von Prüfer*innen und unter realistischen Bedingungen</b>	Berücksichtigung menschlicher Faktoren bei POD-Versuchen. Das heißt, die POD-Versuche werden von einer repräsentativen Stichprobe von Prüfer*innen (>30) und unter realistischen Bedingungen durchgeführt, die dem Kontext entsprechen, für den die Zuverlässigkeitsbewertung vorgenommen wird.	<ul style="list-style-type: none"> <li>▪ Große externe Validität</li> <li>▪ HF wird direkt in die POD-Kurve einbezogen</li> </ul>	<ul style="list-style-type: none"> <li>▪ Kostenaufwendig</li> <li>▪ Erfordert eine größere Anzahl von Prüfer*innen (&gt;30)</li> <li>▪ Erfordert eine sorgfältige Gestaltung und Planung</li> <li>▪ Erfordert Fachkenntnisse in der Versuchsplanung und HF</li> <li>▪ Simulation realistischer Bedingungen manchmal nicht möglich (z. B. Strahlung)</li> <li>▪ Vollständige Repräsentativität der Bedingungen (z. B. organisatorisch) nicht machbar</li> </ul>
<b>8. Verwendung der Modellierung der menschlichen Leistungsfähigkeit</b>	Modelle der menschlichen Leistung sagen das menschliche Verhalten in einer Aufgabe, einem Bereich oder einem System voraus. Diese prädiktiven Auswertungen können in der rechnergestützten Modellierung weiterverwendet werden und werden typischerweise bei der Entwicklung von Mensch-Maschine-Systemen eingesetzt [36].	<ul style="list-style-type: none"> <li>▪ Umfassende Analyse der menschlichen Faktoren</li> <li>▪ Die Informationen sind in dem Wert<sub>90/95</sub> enthalten.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Schwierigkeiten bei der Auswahl der zu modellierenden Faktoren oder bei der Berücksichtigung einer ausreichenden Zahl relevanter Faktoren</li> <li>▪ Erfordert Erfahrung und Fachkenntnisse</li> </ul>
<b>9. Anwendung von Methoden zur Quantifizierung menschlicher Fehler</b>	Schätzung menschlicher Fehlerwahrscheinlichkeiten ( <i>Human Error Probabilities</i> , HEPs) für spezifische Prüfaufgaben unter Verwendung von Expertenurteilen durch die Anwendung von Human Reliability Analysis (HRA)-Techniken (z.B. THERP, CREAM).	<ul style="list-style-type: none"> <li>▪ Enthält anwendungsrelevante Fehlerwahrscheinlichkeiten</li> </ul>	<ul style="list-style-type: none"> <li>▪ Könnte in Umfang und Zeit zu aufwendig sein.</li> <li>▪ Erfordert Kenntnisse und Erfahrung in der HRA, die für jeden spezifischen Anwendungsbereich durchgeführt werden muss</li> <li>▪ Erfordert erfahrene Personen zur Schätzung der Fehlerwahrscheinlichkeiten</li> <li>▪ Die Maßnahme ist subjektiv (auf der Grundlage von Expertenschätzungen)</li> </ul>

Methode	Beschreibung	Vorteile	Nachteile
<b>10. Anwendung von Methoden zur Risikobewertung</b>	Schätzung des menschlichen Risikos als Funktion der Fehlerwahrscheinlichkeit, der Relevanz und der Entdeckungswahrscheinlichkeit unter Verwendung der Human-FMEA. Das Ergebnis ist eine Risikoprioritätszahl oder eine Fehlerwahrscheinlichkeit für bestimmte Aufgaben.	<ul style="list-style-type: none"> <li>Enthält anwendungsrelevante Fehlerwahrscheinlichkeiten</li> </ul>	<ul style="list-style-type: none"> <li>Erfordert Kenntnisse und Erfahrungen mit Risiken im Zusammenhang mit Menschen und Verständnis für menschliches Versagen</li> <li>Erfordert erfahrene Personen zur Schätzung der Fehlerwahrscheinlichkeiten</li> <li>Die Maßnahme ist subjektiv (auf der Grundlage von Expertenschätzungen)</li> </ul>
<b>11. Verwendung von Data science<sup>1</sup></b>	Zusammenbringen von menschlichen und technischen Faktoren in einem gemeinsamen fundierten datenbasierten Modell; Identifikation von Zusammenhängen zwischen menschlichen und technischen Faktoren, mit dem Ziel qualitative Aussagen in quantitative Schlussfolgerungen zu überführen.	<ul style="list-style-type: none"> <li>Zusammenbringen von menschlichen und technischen Faktoren in einer POD Analyse</li> </ul>	<ul style="list-style-type: none"> <li>Erforderliche Kenntnisse und Erfahrung mit KI-Algorithmen</li> <li>Benötigt viele Daten über HF-Einflüsse, die noch nicht vorhanden sind</li> </ul>

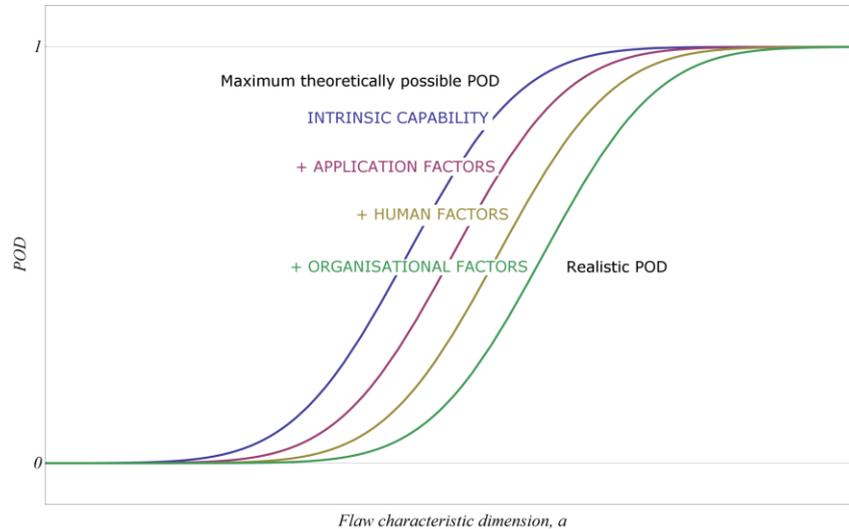
### 2.3. Statistische Methoden zur Einbeziehung der HF in den POD

Im Allgemeinen unterscheiden sich die statistischen POD-Verfahren je nach Art des ZfP-Ergebnisses, z. B. qualitative oder quantitative Daten. Qualitative ZfP-Techniken wie die Eindringprüfung liefern Informationen über die Inspektion in Form von binären Ergebnissen wie ja oder nein, entdeckt oder nicht entdeckt, 1 oder 0. Bei quantitativen ZfP-Techniken wie der Ultraschallprüfung oder der Wirbelstromprüfung wird das Ergebnis in Form eines messbaren Parameters wie Amplitude oder Volt angegeben. Diese qualitativen und quantitativen ZfP-Daten werden mit Hilfe der statistischen Verfahren Hit-Miss und  $\hat{a}$  (Signalantwort) vs.  $a$  (Fehlergröße) verarbeitet [37]. Diese POD-Verfahren vom Typ Hit-Miss und  $\hat{a}$  vs.  $a$  basieren auf der log-odds oder log-logistischen Verteilungsfunktion bzw. der kumulativen log-normalen Verteilungsfunktion. Bei beiden statistischen Verfahren ist die POD-Funktion hauptsächlich von den Modellparametern  $\mu$  und  $\sigma$  abhängig [37]. Der Parameter  $\mu$  spiegelt die Fehlergröße wider, die in 50 % der Fälle erkannt wird, oder den  $a_{50}$ -Wert. Darüber hinaus spiegelt dieser Parameter  $\mu$  auch die Entscheidungsschwelle wider, die bei der Verarbeitung der Daten im Falle des signalantwortbasierten POD-Verfahrens verwendet wird. Der Parameter  $\sigma$  hingegen zeigt die Streuung oder Variabilität in den Daten an und beeinflusst somit die Steilheit der POD-Kurve. Mit anderen Worten, wenn der Parameter  $\sigma$  kleiner ist, d. h. wenn die Daten nicht stark gestreut sind, ergibt die POD-Funktion eine viel steilere POD-Kurve. Es ist also klar, dass diese beiden Parameter, d. h.  $\mu$  und  $\sigma$ , die Form der POD-Kurve und schließlich den POD-Metrikwert  $a_{90/95}$  beeinflussen. Wenn die Einflüsse (auf eine der in Tabelle 1 beschriebenen Weisen) auf diese POD-Parameter dargestellt werden können, erhält man die POD-Kurven unter diesen Einflüssen.

Aus dem modularen Modell geht hervor, dass die Zuverlässigkeit der zerstörungsfreien Prüfung von der intrinsischen Fähigkeit, den Anwendungsparametern und den menschlichen Faktoren im organisatorischen Kontext beeinflusst wird [1]. Wenn die POD-Bewertungen in der Laborumgebung mit realistischen oder repräsentativen Stichproben durchgeführt werden, ist die in den Daten beobachtete Variabilität auf die inhärenten Fähigkeiten der in der Studie verwendeten Testsysteme zurückzuführen. Werden diese POD-Bewertungen in der Feldumgebung durchgeführt, so ist die in den Daten beobachtete Variabilität auf die Variabilität aufgrund der Anwendungs-/Umgebungsparameter zurückzuführen. Darüber hinaus führt die Berücksichtigung der auf

<sup>1</sup> Diese Idee entstand im Rahmen eines gemeinsamen Projektantrags und ist Eigentum von SVTI (Dr. D. Algernon, P. Kicherer, Dr. A. Muller und Dr. C. Thurnherr)

HF zurückzuführenden Schwankungen in den POD-Daten zu POD-Kurven in realistischen Szenarien. Mit anderen Worten: Alle diese Einflüsse erhöhen die Streuung der POD-Daten. Wenn die Gesamtstreuung aufgrund all dieser verschiedenen Einflüsse in den POD-Berechnungen berücksichtigt werden kann, erhält man die realistischen POD-Kurven, wie in **Abb. 3** dargestellt.



**Abb. 3.** Einfluss der verschiedenen Faktoren auf die POD-Kurve (Abbildung M. Pavlovic<sup>2</sup>)

Es ist jedoch eine Herausforderung, den HF-Einfluss in die POD-Studien einzubeziehen. In der aktuellen Arbeit zur Berücksichtigung des Einflusses von HF auf die POD-Kurven basiert die Hauptidee daher auf der Identifizierung der Mehrzahl der mit dem Menschen verbundenen Aufgaben bei der Durchführung der Prüfverfahren in Bezug auf verschiedene statistische Metriken, wie in **Tabelle 1** erläutert. Die Mehrzahl der mit dem Menschen zusammenhängenden Faktoren kann in Form von verschiedenen statistischen Verteilungen, Expertenrankings, Fehlerwahrscheinlichkeiten für jede Aufgabe usw. quantifiziert werden. Diese Art der Darstellung in Form von statistischen Metriken würde es ermöglichen, all diese menschenbezogenen Einflüsse in den Streuparameter  $\sigma_{HF}$  einzubeziehen. Mit anderen Worten, jeder menschliche Einfluss ist für eine gewisse Streuung in den Daten verantwortlich und die Gesamtstreuung aufgrund menschlicher Faktoren ( $\sigma_{HF}$ ) in den Daten ist eine Summe all dieser individuellen Streuungen, wie in Gleichung 1 dargestellt.

$$\sigma_{HF} = \sigma_{HF1} + \sigma_{HF2} + \sigma_{HF3} \dots \dots + \sigma_{HF_n} \quad (1)$$

wobei  $\sigma_{HF1}$ ,  $\sigma_{HF2}$ ,  $\sigma_{HF3}$  usw. die Streuungen aufgrund von n verschiedenen menschlichen Faktoren wie Erfahrung, Wissen, Fähigkeiten, Zeitdruck usw. sind. Wenn die POD-Kurven unter Verwendung dieser  $\sigma_{HF}$  zusätzlich zu den intrinsischen Streuungen und dem Parameter  $\mu$  auf der Grundlage einer geeigneten Entscheidungsschwelle aufgezeichnet werden, können die POD-Kurven sowie die  $a_{90/95}$  -Werte unter dem HF-Einfluss ermittelt werden. Wenn die Gesamtvariabilität aufgrund menschlicher Einflüsse anhand der oben genannten statistischen Metriken geschätzt wird, können auch validierte Modelle zur Simulation der POD-Daten unter Berücksichtigung dieser menschlichen Einflüsse verwendet werden. Darüber hinaus können die POD-Kurven unter den menschlichen Einflüssen durch experimentell unterstützte Modellierung und Simulationen ermittelt werden. Da die Streuungen aufgrund der verschiedenen HF nicht identifiziert werden können, besteht eine der Alternativen zur Berücksichtigung der Streuungen darin, Literaturinformationen (wie in **Tabelle 1** erwähnt)

<sup>2</sup> Übersetzung aus der englischen Sprache: M. Bertovic

über die durch HF bedingten Variabilitäten einzuholen und die POD-Kurven zu schätzen [38]. Zusätzlich zur Aktualisierung des Streuungsparameters haben Vamsi et. al. [38] in ihrer Arbeit auch einen Faktor der Verzerrung eingeführt. Diese Verzerrung kann auf mangelnde Motivation, Konzentrationsmangel, Stimmungsschwankungen oder andere zufällige Schwankungen zurückgeführt werden. Die menschlichen Fehler, die auf diese Verzerrungsfaktoren zurückzuführen sind, können zu zwei Arten von Fehlern führen, nämlich zu einer Fehlanzeige oder zur Zuweisung falscher Signalamplituden aufgrund falscher Kalibrierungsverfahren. Im Falle von Fehlanzeigen können die POD-Daten mit einer bestimmten prozentualen Zuordnung für diese Verzerrungsfaktoren auf der Grundlage einiger realistischer experimenteller Beobachtungen geändert werden. Im Falle der Zuweisung falscher Signalamplituden kann eine ausgewählte Anzahl von Amplituden unter den Rauschpegel in den Daten oder mittels Simulationen abgesenkt werden. Die beste Methode zur Einbeziehung von HF in die POD wäre natürlich eine umfassende Variation in den POD-Experimenten (vorgeschlagene Methode 7), indem POD-Experimente nicht nur an verschiedenen Testfehlern, sondern an realistischen Testfehlern und unter kontextspezifischen und realistischen Bedingungen mit einer ausreichend großen Stichprobe von Prüfer\*innen durchgeführt werden. Diese Methode ist jedoch kaum durchführbar, weshalb die weitere Arbeit im normPOD-Projekt die Untersuchung anderer erwähnter Methoden und die Überprüfung ihrer Durchführbarkeit umfassen wird. Das Ziel des Projekts besteht nicht darin, die perfekte Methode auszuwählen, sondern verschiedene Optionen zu prüfen und abzuwägen und so zu einer realistischeren Zuverlässigkeitsbewertung zu gelangen.

#### *2.4. Schwierigkeiten bei der Quantifizierung menschlicher Faktoren in der zerstörungsfreien Prüfung*

Die Schwierigkeit bei der Suche nach einer geeigneten Methode zur Quantifizierung menschlicher Faktoren und ihrer Einbeziehung in die Zuverlässigkeitsbewertung besteht darin, a) die beste Methode zur Quantifizierung von HF zu finden und b) die geeigneten Informationen (z. B. Häufigkeit, Prozentsatz, Verteilung) bereitzustellen, die in den POD aufgenommen werden können.

Eine der größten Schwierigkeiten ist der gravierende Mangel an empirischen Studien über HF in der ZfP mit quantifizierbaren Daten darüber, wie groß der Einfluss menschlicher Faktoren auf die Zuverlässigkeit ist. Nur eine Handvoll Studien in diesem Bereich waren experimentell (z. B. [3], [34], [39]), so dass uns nicht viel zur Verfügung steht. Wenn die Studie nicht über eine Kontrollgruppe verfügt (eine Gruppe von Teilnehmern, auf die die unabhängige Variable keinen Einfluss hatte, z. B. ein HF-Einfluss), ist es unmöglich zu sagen, wie groß der Effekt ist. In der PISC-3-Studie wird behauptet, dass ein Leistungsabfall von 18,2 % zwischen den leichten, laborähnlichen Bedingungen und den harten, praxisnahen Bedingungen, die sie simulierten, beobachtet wurde [34] was es Spies & Rieder ermöglicht [7] diese Informationen bei der Erstellung ihres HF-POD zu nutzen. Die PISC-Studie hatte jedoch ihre methodischen Schwierigkeiten, so dass wir uns fragen, ob 18,2 % wirklich das Ausmaß des HF-Einflusses sind.

Eine weitere Schwierigkeit besteht darin, dass HF nicht unbedingt immer negativ sind. Methoden der menschlichen Fehlerwahrscheinlichkeit (HEP), wie im Rahmen der HRA (siehe Abschnitt 1.2.2), weisen Wahrscheinlichkeiten des menschlichen Versagens zu. Was ist jedoch mit den positiven Ergebnissen, die sich aus der großen Erfahrung oder Einsicht der Personen ergeben, die die Prüfungen durchführen? Darüber hinaus neigen Methoden zur Quantifizierung menschlicher Fehler dazu, HEPs bestimmten Aufgaben und nicht der gesamten Prüfaufgabe zuzuordnen. Dies wirft die Frage auf, wie diese Einzelwahrscheinlichkeiten in eine allgemeine Zuverlässigkeitsbewertung einbezogen

werden können. HEPs und die in der FMEA zugewiesene Risikopriorität beruhen auf subjektiven Expertenschätzungen und werden daher in hohem Maße von der Anzahl, dem Fachwissen und der Persönlichkeit (z. B. wie kritisch man ist) der Experten beeinflusst.

Die Erforschung menschlicher Faktoren hat sich bisher auf die Vielfalt und Unterschiedlichkeit der verschiedenen Einflüsse konzentriert. Quantifizierungsmethoden erfordern oft nur eine Zahl, eine Verteilung oder einen Prozentsatz. Dadurch wird die HF auf eine einzige Zahl reduziert, was auf den Black-Box-Ansatz zurückgeht, den die ZfP seit ihren Anfängen verfolgt hat.

Und schließlich sind die HF in hohem Maße von der Situation abhängig, in der die Inspektionen durchgeführt werden. Per Definition beziehen sich HF auf Umwelt-, Organisations- und Arbeitsplatzfaktoren sowie auf menschliche und individuelle Eigenschaften, die das Verhalten bei der Arbeit beeinflussen. All diese Faktoren sind bei verschiedenen Anwendungen, Umgebungen, Organisationen und Menschen sehr unterschiedlich. Dies führt zu der Schlussfolgerung, dass quantifizierbare Methoden niemals in der Lage sein werden, den menschlichen Faktoren gerecht zu werden.

### **3. Schlussfolgerungen: Können menschliche Faktoren in die POD-Kurve integriert werden?**

Die Zuverlässigkeitsbewertung von ZfP-Systemen hat sich bisher immer auf die technischen Fähigkeiten der ZfP konzentriert und die Einflüsse menschlicher und organisatorischer Faktoren weitgehend vernachlässigt. Dies war darauf zurückzuführen, dass es nicht möglich war, ein geeignetes Maß zur Quantifizierung der HF zu finden.

Menschliche Zuverlässigkeitsanalysen bieten verschiedene Methoden zur Quantifizierung der Wahrscheinlichkeit menschlicher Fehler bei verschiedenen Aufgaben. Andere mögliche Ansätze sind die Ableitung von Erkenntnissen aus früheren Studien, die Modellierung menschlicher Faktoren, Methoden zur Risikobewertung und Risikopriorisierung, die Verwendung von Simulationen oder die Durchführung neuer Experimente, um die erforderlichen Erkenntnisse zu gewinnen. Obwohl sie mit vielen Herausforderungen konfrontiert sind, sind die Methoden zur Quantifizierung menschlicher Faktoren und menschlicher Fehler vielversprechend, da sie quantifizierbare Informationen über die menschliche Zuverlässigkeit liefern können. Der Vorteil dieses Unterfangens ist, dass HF endlich ihren Platz in der Zuverlässigkeitsbewertung finden und dass der Endanwender praxisnahe Informationen über die Zuverlässigkeit seines Systems erhält, wenn es im Feld eingesetzt wird. Der Nachteil von Quantifizierungsmethoden ist, dass sie die Vielfalt menschlicher Faktoren auf eine Zahl, d. h. eine Blackbox, reduzieren und nicht anwendungsspezifisch sind.

Die Forscher sind sich im Allgemeinen einig, dass keine der Lösungen die gesamte durch menschliche Faktoren bedingte Variabilität abdecken wird. Einige der größten Herausforderungen, die es zu bewältigen gilt, sind vielleicht die Komplexität der HF-Einflüsse und ihre Interaktion sowie der Mangel an (angemessen berichteten) Studien über menschliche Faktoren, die die wichtigsten leistungsbestimmenden Faktoren identifizieren.

Daher wird die Zuverlässigkeitsbewertung am besten durchgeführt, indem sowohl quantifizierbare als auch nicht quantifizierbare Ansätze für menschliche Faktoren verwendet werden. Die Quantifizierung wird niemals die qualitativen Bemühungen ersetzen können, die notwendig sind, um die Einflüsse menschlicher Faktoren zu verstehen, die negativen Auswirkungen zu minimieren und die Zuverlässigkeit zu verbessern.

## 4. Danksagung

Die in diesem Artikel geäußerten Überlegungen betreffen das Projekt "normPOD", das vom Bundesministerium für Wirtschaft und Klimaschutz unter dem Namen "Standardisierung zur probabilistischen Bewertung der Zuverlässigkeit für zerstörungsfreie Prüfverfahren" unter dem Förderkennzeichen 03TN0006D gefördert wird.

## 5. Referenzen

- [1] C. Mueller *et al.*, "Paradigm Shift in the Holistic Evaluation of the Reliability of NDE Systems," *Mater. Test.*, vol. 55, no. 4, pp. 261–269, Apr. 2013.
- [2] HSE, *Reducing error and influencing behaviour (HSG48)*, 2. Edition. Health and Safety Executive, HSE Books, 1999.
- [3] M. Bertovic, M. Gaal, C. Müller, and B. Fahlbruch, "Investigating human factors in manual ultrasonic testing: testing the human factors model," *Insight*, vol. 53, no. 12, pp. 673–676, 2011.
- [4] M. Bertovic, *Human Factors in Non-Destructive Testing (NDT): Risks and Challenges of Mechanised NDT*. Berlin: Doctoral dissertation. Technische Universität Berlin. BAM-Dissertationsreihe Band 145. Bundesanstalt für Materialforschung und -prüfung (BAM), 2016.
- [5] EPRI, "Human Factors in Nondestructive Evaluation (NDE): A literature Review and Field Observations. Report No. 3002010462," Electric Power Reserch Institute (EPRI), Palo Alto, CA, 2017.
- [6] A. D'Agostino, S. Morrow, C. Franklin, and N. Hughes, "Review of Human Factors Research in Nondestructive Examination," Office of Nuclear Reactor Regulation, U.S. Nuclear Regulatory Commission, Washington, DC, 2017.
- [7] M. Spies and H. Rieder, "An approach to the question 'How to account for human error in MAPOD?,'" in *12th European Conference on Non-Destructive Testing (ECNDT 2018), Gothenburg 2018, June 11-15, 2018*, pp. 1–5.
- [8] M. R. Bato, A. Hor, A. Rautureau, and C. Bes, "Experimental and numerical methodology to obtain the probability of detection in eddy current NDT method," *NDT E Int.*, vol. 114, p. 102300, Sep. 2020.
- [9] D. Kanzler *et al.*, "Lasst uns mal POD standardisieren!," in *DGZfP Jahrestagung 2022, Kassel, 2022*, pp. 1–10.
- [10] National Transportation Safety Board, "Accident Report NTSB/AAR-10/03: Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River/US Airways Flight 1549/Airbus A320-214," 2009.
- [11] J. Reason, *Human error*. New York: Cambridge University Press, 1990.
- [12] S. W. Dekker, *The field guide to human error investigations*. Aldershot, England: Ashgate, 2002.
- [13] E. Hollnagel, "The phenotype of erroneous actions," *Int. J. Man. Mach. Stud.*, vol. 39, no. 1, pp. 1–32, 1993.
- [14] J. Reason, *Managing the Risks of Organizational Accidents*. Farnham, Surrey: Ashgate, 1997.
- [15] P. Pyy, *Human reliability analysis methods for probabilistic safety assessment (VTT Publications 422)*, vol. 4, no. 2. Espoo: VTT Technical Research Centre of Finland., 2000.
- [16] A. Swain and H. Guttman, "Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report [No. NUREG/CR-1278; SAND-80-0200]," U.S. Nuclear Regulatory Commission, Washington, D.C., 1983.
- [17] ISO 31000, "Risk management — Principles and guidelines.," International Organization for Standardization (ISO), Geneva, Switzerland, 2009.
- [18] IEC/ISO 31010, "Risk management - Risk assessment techniques," ISO/IEC, Geneva, Switzerland, 2009.
- [19] American Psychological Association, "Quantitative Research," *APA Dictionary of Psychology*, 2022. [Online]. Available: <https://dictionary.apa.org/quantitative-research>. [Accessed: 08-Apr-2022].
- [20] American Psychological Association, "Qualitative Research," *APA Dictionary of Psychology*, 2022. [Online]. Available: <https://dictionary.apa.org/qualitative-research>. [Accessed: 08-Apr-2022].
- [21] V. Di Pasquale, R. Iannone, S. Miranda, and S. Riemm, "An Overview of Human Reliability Analysis Techniques in Manufacturing Operations," in *Operations Management*, InTech, 2013.
- [22] Hollnagel, *Cognitive reliability and error analysis method: CREAM*. Oxford, UK: Elsevier, 1998.
- [23] MIL-STD 1629A, "Procedures for performing a Failure Mode, Effects and Criticality Analysis. Military standard." U.S. Department of Defense, Washington, DC, 1980.
- [24] J. Algedri and E. Frieling, *Human-FMEA*. Munich: Carl Hanser Verlag, 2001.
- [25] M. Bertovic, "Assessing and Treating Risks in Mechanised NDT: A Human Factors Study.," *ZfP Zeitung*, vol. 161, pp. 52–62, 2018.

- [26] M. Bertovic, “Human-Related Risks in the Non-Destructive Testing (NDT) of Hollow Railway Axles: Implications for the Education and Training of the NDT Personnel,” in *Proceedings of the 29th European Safety and Reliability Conference*, 2019, pp. 294–300.
- [27] N. Dominguez, D. Rodat, F. Guibert, A. Rautureau, and P. Calmon, “POD Evaluation Using Simulation: Progress and Perspectives regarding Human Factor,” in *Proceedings of the 19th World Conference on Non-Destructive Testing, Munich, 13-17.06.2016*, 2016, pp. 1–7.
- [28] B. Chapuis, F. Jenson, P. Calmon, G. DiCrisci, J. Hamilton, and L. Pomić, “Simulation supported POD curves for automated ultrasonic testing of pipeline girth welds,” *Weld. World*, vol. 58, no. 4, pp. 433–441, 2014.
- [29] P. Calmon, B. Chapuis, F. Jenson, and E. Sjerne, “The use of simulation in POD curves estimation: An overview of the IIW best practices proposal,” in *Proceedings of the 19th World Conference on Non-Destructive Testing, Munich, 13-17.06.2016*, 2016, pp. 1–7.
- [30] M. Wall, F. A. Wedgwood, and S. Burch, “Modelling of NDT Reliability (POD) and applying corrections for human factors,” in *Proceedings of the ECNDT*, 1998, vol. 3, no. 7, pp. 1–8.
- [31] M. Wall, S. Burch, and J. Lilley, “Human factors in POD modelling and use of trial data,” *Insight - Non-Destructive Test. Cond. Monit.*, vol. 51, no. 10, pp. 553–561, 2009.
- [32] C. A. Harding, G. R. Hugo, S. J. Bowles, D. O. Thompson, and D. E. Chimenti, “Application of model-assisted POD using a transfer function approach,” in *AIP Conference Proceedings*, 2009, vol. 1096, no. 1, pp. 1792–1799.
- [33] R. M. Meyer, S. L. Crawford, J. P. Lareau, and M. T. Anderson, “Review of Literature for Model Assisted Probability of Detection - PNNL-23714,” Richland, WA, 2014.
- [34] R. Murgatroyd, R. Chapman, S. Crutzen, H. Seed, A. Willets, and G. Worrall, “Human Reliability in Inspection, Final Report on Action 7 in the PISC III Programme [PISC III Report n° 31],” JRC & NEA, 1994.
- [35] M. Wall, “Model Assisted Qualification of Inspection - Protocol. [Report No. 30304/1v4/20 MAQ Protocol],” TWI Ltd, Camebridge, UK, 2020.
- [36] A. Sebok, C. D. Wickens, and R. Sargent, “Using Meta-Analyses Results and Data Gathering to Support Human Performance Model Development,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 57, no. 1, pp. 783–787, Sep. 2013.
- [37] A. P. Berens, “NDE Reliability Data Analysis,” in *ASM Metals Handbook, Volume 17, Nondestructive Evaluation and Quality Control*, 9th Ed., Materials Park, OH: American Society of Metals International, 1989, pp. 689–701.
- [38] V. K. Rentala, D. Kanzler, and P. Mylavarapu, “Impact of NDE 4.0 on the life revisions of aero-engines| Rentala | Publications | Spie,” in *Proc. SPIE PC12049, NDE 4.0, Predictive Maintenance, and Communication and Energy Systems in a Globally Networked World*, 2022.
- [39] J. Enkvist, A. Edland, and O. Svenson, “Effects of operator time pressure and noise on manual ultrasonic testing,” *Insight*, vol. 43, no. 11, pp. 725–730, 2001.