

# Partnerspezifischer Schlussbericht

## JHP GmbH

### Verbundprojekt

RDV – Real Driving Validation: Erweiterung der Überprüfbarkeit von Continuous SW Integration in Kommunikation mit Fahrzeugen im Feld



**Laufzeit des Vorhabens** 01.01.2022 bis 30.06.2025

---

**Projektkoordinator** dSPACE GmbH

---

**Erstellungsdatum** 12.10.2025

## Inhaltsverzeichnis

1	Aufgabenstellung.....	4
1.1	Gesamtziel des Vorhabens .....	4
1.2	Partnerspezifische Projektziele .....	4
2	Voraussetzungen, unter denen das Vorhaben durchgeführt wurde .....	4
3	Planung und Ablauf des Vorhabens .....	5
4	Wissenschaftlicher und technischer Stand, an den angeknüpft wurde.....	6
5	Zusammenarbeit mit anderen Stellen.....	7
6	Verwendung der Zuwendung und Ergebnisse, mit Gegenüberstellung der vorgegebenen Ziele .....	7
6.1	Statistische Auswertung von Validationsdaten von Perzeptionsalgorithmen .....	7
6.2	Abschätzung des unknown risk entsprechend sotif.....	16
6.3	Statistischer Vergleich von autonomen Fahrzeugen mit menschlichen Fahrern.....	17
7	Wichtigste Positionen des zahlenmäßigen Nachweises.....	19
8	Notwendigkeit und Angemessenheit der geleisteten Arbeit.....	19
9	Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans.....	20
10	Während der Durchführung des Vorhabens bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen .....	20
11	Erfolgte oder geplante Veröffentlichungen des Ergebnisses.....	20



## 1 AUFGABENSTELLUNG

Eine kontinuierliche Qualitätsaussage über automatisierte Fahrfunktionen von Level 3 bis 5 auf Fahrzeugebene („Grey box“) unter umfassender Berücksichtigung aller möglichen Szenarien, Umweltbedingungen und Verkehrsteilnehmer (Gesamtsystem) ist eine sehr große Herausforderung und bedarf einer zunehmenden Vernetzung sowie einem effizienten Wissenstransfer von Wirtschaft, Wissenschaft und Industrie.

### 1.1 GESAMTZIEL DES VORHABENS

Auf Grundlage realer Fahrdaten aus dem Feld sowie aus Hauptuntersuchungen wurde durch das Projekt Real Driving Validation (RDV) eine effiziente und sichere Methodik für OEM's und Tier 1/Tier 2 zur Validierung von KI-basierten Fahrfunktionen zur Verfügung gestellt. RDV nutzte stochastische Modelle und Ausfallwahrscheinlichkeiten, um anhand der realen Betriebszustände von Fahrzeugen im Feld kritische Szenarien gezielter und schneller identifizieren und testen zu können. Diese Szenarien können in digitale Ökosysteme wie GAIA-X oder zur Weiterentwicklung der industriellen Normung und Standardisierung genutzt werden.

Mittels Markov-Modell und numerischen Simulationsmethoden wurde zudem eine methodisch und analytisch-quantitativ abgesicherte Validierung der hochautomatisierten Fahrfunktionen (z.B. bei Softwareupdates) gewährleistet, wodurch optimierte Entwicklungs- und Zulassungsprozesse die Innovationszyklen durch Real Driving Validation beschleunigen.

Über ein AD-Referenzsystem für zukünftige „fail-operational“ agierender Fahrzeuggenerationen wurden anhand eines automatischen AD MW & Computing Test-Systems Erkenntnisse in Bezug auf Testbarkeit, speziell im Middleware-Segment des AD-Stacks und der on-board Datenkommunikation, gewonnen.

### 1.2 PARTNERSPEZIFISCHE PROJEKTZIELE

Aus der Vorhabensbeschreibung und der übergeordneten Projektstruktur, welche sich an verschiedenen Use-Cases orientierte, leiteten sich drei Projektziele für dieses Teilvorhaben ab. Diese sind:

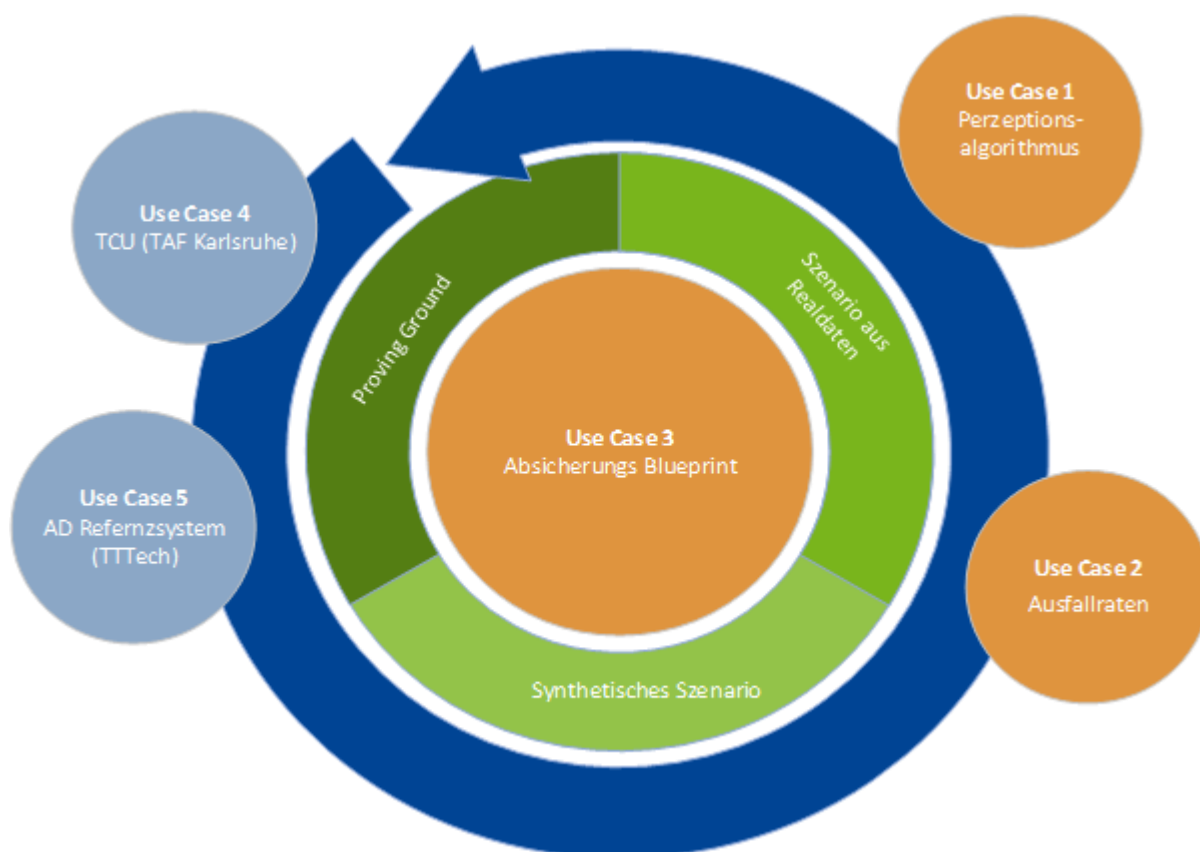
- Statistische Auswertung von Perzeptionsalgorithmus-Validierungsdaten,
- Statistische Auswertung von autonomen Fahrzeugen im realen Feldbetrieb und
- Ableitung einer statistischen Methode zur Risikobewertung bzgl. unentdeckter Fahrscenarien, welche potentiell zu Unfällen führen können.

## 2 VORAUSSETZUNGEN, UNTER DENEN DAS VORHABEN DURCHFÜHRT WURDE

Der Projektpartner Jäger & Hitziger Beratungsgesellschaft für Produktzuverlässigkeit mbH (JHP) beschäftigt sich seit mehr als 20 Jahren mit Fragen der Produktzuverlässigkeit und -sicherheit. JHP berät Kunden verschiedener Industriebereiche und führt Projekte durch. Insbesondere im Fahrzeugbereich hat sich JHP von Beginn an eine breite Wissensbasis erarbeitet. Im Vordergrund der Arbeit steht die Kombination aus Technologie und Statistik. Alle bisherigen Standards der Funktionalen Sicherheit basieren auf dieser Kombination. Beim autonomen Fahren wird über den Bereich der funktionalen Sicherheit hinausgegangen, da das Gesamtsystem nun unter allen Bedingungen permanent funktionieren muss. Hierbei steht dann insbesondere die Funktionssoftware autonomer Systeme im Vordergrund. JHP beschäftigt sich seit langer Zeit mit Zuverlässigkeitsmodellen verschiedener Technologien und führt über die Reliability Engineering Academy internationale Ausbildungsgänge zu diesen Themen durch. JHP baut technologiespezifische Zuverlässigkeitsvalidierungsprogramme auf, wertet Felddaten aus und baut Lebensdauermodelle für neue Technologien auf. Basis dafür sind die statistischen Ansätze der Zuverlässigkeitstechnik und die Kenntnis von technologiespezifischen Schadensmechanismen.

### 3 PLANUNG UND ABLAUF DES VORHABENS

Das Gesamtvorhaben Real Drive Validation lief von Januar 2022 nach einer kostenneutralen Verlängerung bis Ende Juni 2025. Nach einem Wechsel des Konsortialführers 2023 gab sich das Verbundprojekt eine Use-Case-Struktur in welcher die Projektpartner an insgesamt 5 Use-Cases zusammenarbeiteten.



**Abbildung 1:** Use Case basierte Projektstruktur nach Neufassung der Vorhabensbeschreibung

Use Case 1 diente der Qualitätsbewertung von Perzeptionsalgorithmen. In Use Case 2 wurde die Modellierung von Ausfällen autonomer Fahrzeuge betrachtet. In Use Case 3 stand die Absicherung (Validierung) der Funktionsfähigkeit und Zuverlässigkeit autonomer Fahrzeuge im Fokus. Use Case 4 arbeitete am Datentransfer für Teleoperation und Use Case 5 erarbeitete ein automatisches AD Middleware & Computing Test-System. JHP war in den Use Cases 1 und 3 mit seinen Themen eingebunden. In der ursprünglichen Vorhabensbeschreibung waren die Projektbestandteile von JHP dem Arbeitspunkt 3 zugeordnet. Durch den Wechsel des Konsortialführers und die Neufassung der Vorhabensbeschreibung kamen die Projektpartner Liangdao und die Technische Hochschule Augsburg hinzu. Hierdurch wurde es JHP möglich, verstärkt an der statistischen Bewertung von Perzeptionsalgorithmen zu arbeiten. Dies kam im Projekt zum richtigen Zeitpunkt, denn es stellte sich zu Beginn des Projektes in der ursprünglichen Konstellation heraus, dass Hardwareaspekte weniger im Fokus stehen, wenn es um autonome Fahrzeuge geht, sondern die Safety of the Intended Function (SOTIF).

In Use Case 1 arbeitete JHP zusammen mit den Projektpartnern Liangdao, IQZ und dSPACE am Aufbau einer prototypischen Entwicklungs- und Validierungskette für Perzeptionsalgorithmen. JHP prüfte die Anwendbarkeit von bekannten statistischen Konzepten auf die Daten, welche im Rahmen der Validierung von KI-basierten Perzeptionsalgorithmen anfallen. Es gelang, diese Daten statistisch auszuwerten und anhand verschiedener Fallbeispiele den Nutzen solcher Analysen für den Entwicklungs- und Optimierungsprozess darzustellen. JHP gelang es mit den statistischen Analysen Aussagen zu machen über das Verbesserungspotenzial durch weiteres Training des Perzeptionsalgorithmus, die Robustheit gegenüber verzerrten Originaldaten und die Identifikation von Optimierungspotenzialen.

In Use Case 3 arbeitete JHP an Themen, welche für die Absicherung (Validierung) von autonomen Fahrzeugen oder Teilsystemen relevant sind. Anhand einer großen und gut dokumentierten Datenmenge eines Feldversuchsprogramms mit autonomen Fahrzeugen in San Francisco konnte JHP statistische Auswertungen zum Unfallgeschehen bei diesen Fahrzeugen machen. Zudem war es möglich einen statistischen Vergleich mit menschlichen Fahrern durchzuführen. Es konnte gezeigt werden, dass auch durch Feldversuch zumindest im innerstädtischen Bereich der Nachweis der Eignung autonomer Fahrzeuge möglich ist.

Das zweite Thema, welches JHP im Rahmen von Use Case 3 bearbeitete war die Erarbeitung einer statistischen Methode zur Bewertung des Restrisikos noch vorhandener aber bisher unidentifizierten Fahrzenarien innerhalb eines ODDs. Hierbei wurde intensiv mit dem Projektpartner dSPACE zusammengearbeitet. Es gelang JHP einen statistischen Ansatz aus der Zuverlässigkeitstechnik auf diese Fragestellung anzuwenden.

JHP arbeitete unterstützend auch im Use Case 2 mit. Insbesondere zu der Definition von Ausfallszenarien und Ursachen von Fehlverhalten autonomer Fahrzeugfunktionen konnte JHP mit seinem Praxiswissen beitragen.

#### **4 WISSENSCHAFTLICHER UND TECHNISCHER STAND, AN DEN ANGEKNÜPFT WURDE**

Die Aufgabe von JHP im RDV-Projekt war die Prüfung und Anwendung verschiedenster bekannter statistischer Herangehensweisen auf die Fragestellungen, welche sich bei der Entwicklung und Absicherung autonomer Fahrzeuge und Teilsysteme ergeben. Der Neuheitsgrad liegt hierbei weniger in der Erarbeitung neuer statistischer Ansätze – diese sind bezogen auf ingenieurwissenschaftliche Anwendungen seit 50 Jahren bekannt – sondern in der konkreten Prüfung der Anwendbarkeit auf eine bestimmte Fragestellung und ggfs. die Herstellung der Anwendbarkeitsvoraussetzungen durch mathematische Vorarbeiten.

Insbesondere die Identifikation der nötigen Methodenvoraussetzungen in der Fragestellung der Projektpartner ist hierbei eine wesentliche Aufgabe. Dies war auch bei den Teilvorhaben von JHP im RDV-Projekt so. Der wesentliche Aspekt bestand darin, den Projektpartnern aufzuzeigen, welche Aussagen durch die Anwendung statistischer Methoden möglich sind.

Dementsprechend gibt es keine spezifischen wissenschaftlichen Vorarbeiten, an welche angeknüpft werden kann. Im Hinblick auf die Anwendung der Statistik in konkreten Entwicklungsprojekten handelt es sich fast immer um eine Neuheit für die Beteiligten, da deren Anwendung in der Hochschulausbildung (leider) oft immer noch nicht ausreichend aufgezeigt wird und es vielen Menschen schwerfällt, die Anwendbarkeit bestimmter statistischer Methoden auf bestimmte Fragestellungen zu erkennen.

Einmal angewendet, ist das Ergebnis aber in vielen Fällen von hohem praktischem Nutzen. So war es auch im RD- Projekt. JHP konnte hier bei allen Themen aufzeigen, dass durch die Anwendung von bekannten statistischen Methoden ein praktischer bisher unbekannter Nutzen möglich wird. Dementsprechend resultierten aus den Arbeiten von JHP im RDV-Projekt zwei Veröffentlichungen, davon eine in der praxisrelevanten Zeitschrift „HANSER automotive“.

Bei den Arbeiten von JHP im RDV- Projekt wurden folgende statistischen Konzepte verwendet, welche im Sinne dieses Kapitels als wissenschaftliche Vorarbeiten betrachtet werden können:

- Eindimensionale Regressionsanalyse
- Zweiseitige Vertrauensbereiche
- Hypothesentest: Test auf Normalverteilung
- Hypothesentest: t-Test
- Chi<sup>2</sup>-Verteilung
- Exponentialverteilung
- Nicht homogener Poisson Prozess
- Crow-AMSAA Modell

## 5 ZUSAMMENARBEIT MIT ANDEREN STELLEN

Durch die Use Case Struktur des RDV-Projektes war die Zusammenarbeit mit den anderen Projektpartnern für JHP sehr intensiv. Wie oben dargestellt lassen sich statistische Methoden praktisch nur dann anwenden, wenn die Voraussetzungen für deren Anwendbarkeit vorliegen. Die Prüfung und eventuelle Herstellung dieser Voraussetzungen war im RDV Projekt Ergebnis intensiver Diskussion und Abstimmung mit den relevanten Projektpartnern.

Der Konsortialführer dSPACE unterstützte die Kommunikation hierbei durch 2-wöchige Termine, an denen alle Projektpartner teilnahmen und eine definierte Agenda.

Es gab in ca. halbjährigem Abstand physische Konsortialtreffen, bei denen die aktuellen Teilstände der Projekte vorgestellt, intensiv diskutiert und mit dem Zielbild abgeglichen wurden.

Konkret arbeitete JHP während des gesamten Projektes sehr intensiv zusammen mit den Projektpartnern Liangdao und dSPACE.

## 6 VERWENDUNG DER ZUWENDUNG UND ERGEBNISSE, MIT GEGENÜBERSTELLUNG DER VORGEgebenEN ZIELE

Im Folgenden werden die Arbeitsergebnisse von JHP beschrieben. Zu Beginn jedes Teilkapitels erfolgt ein Abgleich mit den definierten Projektzielen.

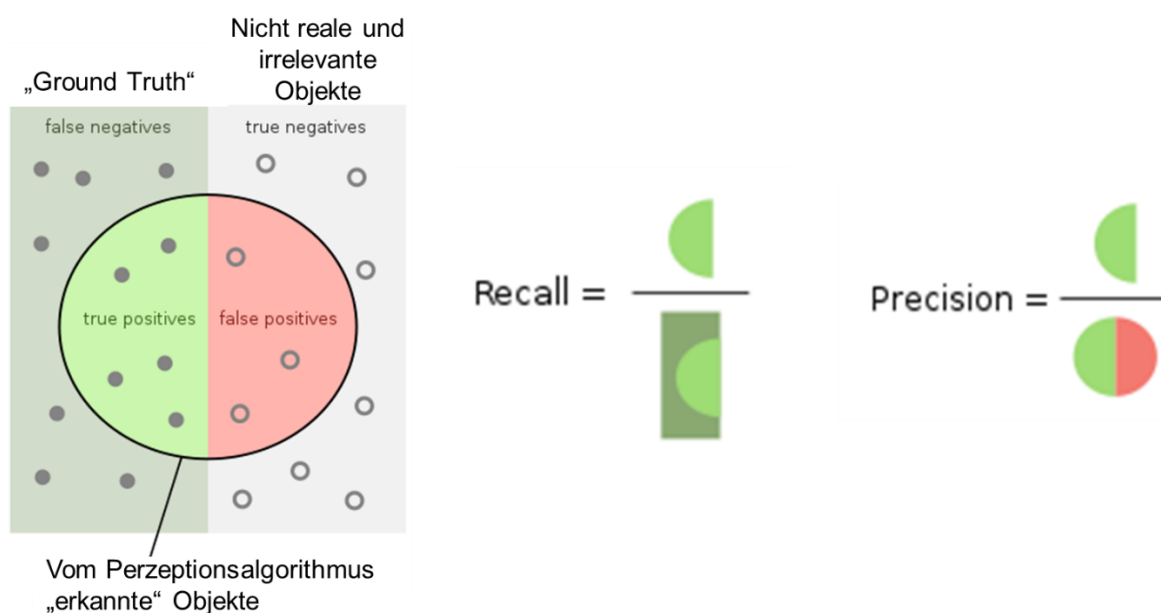
### 6.1 STATISTISCHE AUSWERTUNG VON VALIDATIONS DATEN VON PERZEPTIONSALGORITHMEN

**Ziel des Arbeitspaketes:** Prüfung der Anwendbarkeit statistischer Methoden auf Validationsdaten von KI-Perzeptionsalgorithmen und Auswertung dieser Daten

**Zielerreichung:** Das vorgegebene Ziel wurde vollumfänglich erreicht. Die Daten konnten ausgewertet werden und der Nutzen der Ergebnisse in der Entwicklungs- und Validierungsphase konnte nachgewiesen werden.

Es wurde prototypisch eine Datenstrecke für die Validierung von Perzeptionsalgorithmen aufgebaut. Im Fokus standen hierbei die Gütemaße Precision und Recall. Abbildung x beschreibt die Bedeutung der Maße Precision und Recall grafisch. Der Perzeptionsalgorithmus erkennt während einer Fahrt Objekte. In einem idealen Fall erkennt er nur Objekte, welche wirklich existieren (Ground Truth) und davon alle. In Realität wird der Perzeptionsalgorithmus allerdings nur einen Teil der realen Objekte erkennen (true positives) aber zusätzlich leider auch nicht reale Objekte (false positives), z.B. aufgrund von Sensorschatten.

Hieraus lassen sich die beiden Gütemaße Precision und Recall ableiten. Recall ist der Quotient aus True Positives und Ground Truth, d.h. es geht nur darum, wie gut der Perzeptionsalgorithmus alle realen Objekte erkennt. Ein hoher Recallwert ist wünschenswert, da das Nichterkennen von Objekten potentiell das Unfallrisiko erhöhen kann. Der Precisionwert beschreibt den Quotienten aus True Positives und der Summe aller erkannten Objekte. Dieses Gütemaß beschreibt somit die Effizienz des Perzeptionsalgorithmus. Wenn ein hoher Recallwert durch einen niedrigen Precisionwert „erkauft“ wird resultieren dadurch ebenfalls praktische Nachteile. So wird z.B. durch das Erkennen nicht realer Objekte eine Notbremsung ausgelöst, welche ebenfalls negative Auswirkungen haben kann, wie z.B. eine Notbremsung und in der Folge einen Auffahrunfall.



**Abbildung 2:** Erkennungsverhalten von Perzeptionsalgorithmen und Definitionsbasis für die Gütemaße Precision und Recall

Die originalen Perzeptionsdaten (Lidar-Datensätze) stemmen vom Projektpartner Liangdao und wurden von dem Projektpartner dSPACE. JHP wertete diese Daten statistisch aus. Hierbei wurde zunächst geprüft, ob mit einfacher Normalverteilungsstatistik zufriedenstellende Ergebnisse erzielt werden können. Es stellte sich heraus, dass es möglich ist, die Streuung der Gütemaße einer Gesamtvalidationsdatenmenge gut mittels des Normalverteilungsmodells zu beschreiben. Um dies zu erreichen, wurden die Gütemaße für Einzelframes (Lidardatensatz für ein Bild) in Pakete mit jeweils neun Einzelframes gebündelt. Dies führte zu einer besseren Auflösung der Gütemaße und erlaubte die Anpassung von Normalverteilungen. Würden die Datensätze einzeln ausgewertet ohne die Bündelung in 9er-Pakete wäre die Auflösung der Precision und Recall Werte zu gering und eine Anpassung an die Normalverteilung wäre unmöglich.

Die Normalverteilung wird oft deshalb angestrebt, weil mit diesem Verteilungsmodell viele klassische und vergleichsweise einfache Tests, wie z.B. der t-Test möglich sind. Solche Hypothesentests werden verwendet, um statistisch basierte Aussagen zu machen, wie z.B. Algorithmus A ist besser als Algorithmus B. Die notwendige Basis, um solche Aussagen zu machen, ist die Bestimmung der sog. Irrtumswahrscheinlichkeit. Wenn die Daten normalverteilt sind, können hier vergleichsweise einfache und bekannte Ansätze verwendet werden. Die auszuwertenden Datensätze hatten folgende Struktur.

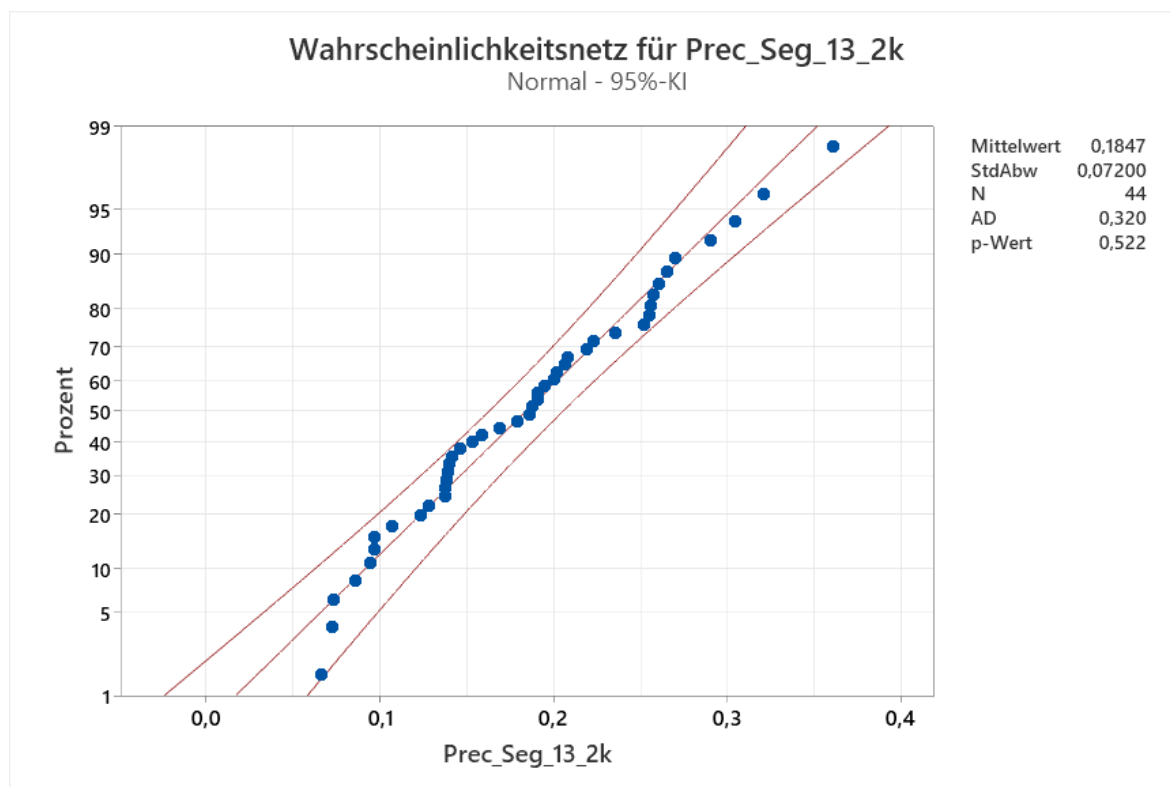
Frames	True Positive	False Positive	Ground Truth	Precision	Recall
1	24	78	126	0,23529	0,19048
2	9	116	126	0,07200	0,07143
3	21	141	146	0,12963	0,14384
.	.	.	.	.	.
.	.	.	.	.	.
44	24	37	108	0,39344	0,22222

**Tabelle 1:** Darstellung der Datenstruktur als Basis für die statistische Auswertung

Jede Zeile stellt einen Gesamtdatensatz („Frame“) aus 9 „Einzelframes“ dar. Beim ersten Frame ergeben sich 24 True Positive Objekte, 78 False Positive Objekte und es lagen insgesamt 126 reale Objekte vor (Ground Truth). Entsprechend der in Abbildung 2 gezeigten Definition der Werte für Precision und Recall ergeben sich dafür sehr niedrige Werte für Precision und Recall.

Auf Basis eines solchen Datensatzes lassen sich die Werte für Precision und Recall nun statistisch auswerten, da in dem in Tabelle y gezeigten Datensatz 44 Frames mit den jeweiligen Werten für Precision und Recall vorhanden sind. Die statistische Auswertung resultiert in Punktesfolgen und Regressionsanpassung in Normalverteilungs-Wahrscheinlichkeitsnetzen, abgebildet in Abbildung 3.

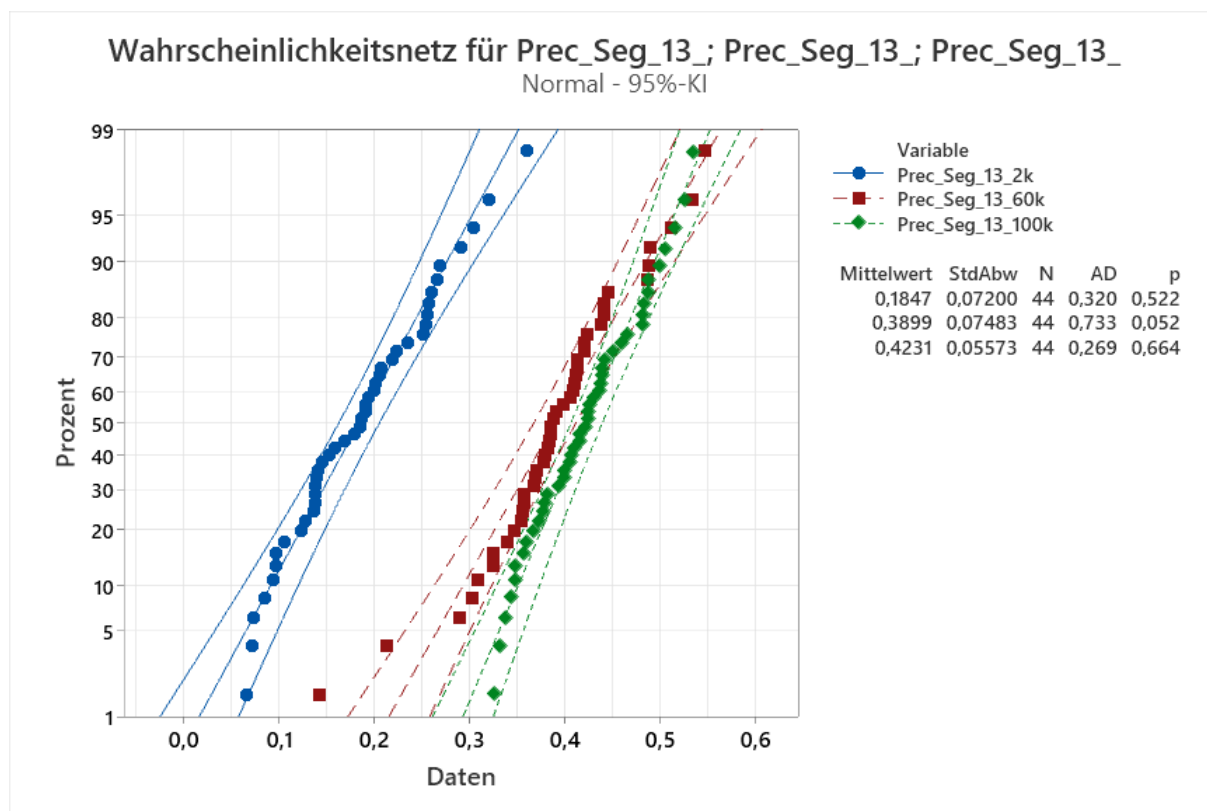
Auf der x-Achse dieser Diagramme sind die jeweiligen Precision- oder Recallwerte der einzelnen Datensätze abgetragen. In Abbildung z sind entsprechend Tabelle y 44 blaue Punkte eingetragen; jeder steht für einen „Frame“ aus dem zugehörigen Datensatz. Die Werte auf der y-Achse sind Wahrscheinlichkeitswerte und werden aufgrund statistischer Regeln zugeordnet. Es ist erkennbar, dass die y-Achse nicht linear ist. Dies entsprechende Transformation der y-Achse führt dazu, dass wenn es sich bei den Werten auf der x-Achse um normalverteilte Werte handelt, die 44 Punkte gut durch eine Gerade im Diagramm beschrieben werden können. Dies ist bei den Daten in Abbildung 3 der Fall. In dem Diagramm werden auch der Mittelwerte und die Standardabweichung – die Bestimmungsparameter der Normalverteilung – ausgegeben. Diese beiden Werte korrespondieren mit der Ausgleichsgerade durch die Datenpunkte. Die einhüllenden Kurven definieren den sog. Vertrauensbereich und berücksichtigen den Umstand, dass die Schätzung der Verteilungsparameter (und somit der Lage und Steigung der Gerade) auf einer endlich großen Stichprobe basiert. In den folgenden Auswertungen wird standardmäßig ein 95%-Vertrauensbereich verwendet (95%-KI). Die Interpretation diese Vertrauensbereichs ist, dass die „wahre“ Verteilung (Gerade) mit einer Wahrscheinlichkeit von 95% innerhalb des Vertrauensbereichs liegt.



**Abbildung 3:** Exemplarische Auswertung eines Datensatzes in einem Normalverteilungs-Wahrscheinlichkeitsnetz.

Ausgewertet wurden 3 verschiedene Trainingsstände eines Perzeptionsalgorithmus (2k, 60k und 100k) welche vom Projektpartner Liangdao gestellt wurden und vom Projektpartner dSPACE vorverarbeitet wurden. Die Bezeichner 2k, 60k und 100k beziehen sich dabei auf die Größe des Trainingsdatensatzes, mit welchem die drei Algorithmen trainiert wurden. Die verschiedenen Trainingsstände des Perzeptionsalgorithmus (im Folgenden PATS genannt) werden auf verschiedene Verkehrssituationen aus dem innerstädtischen und Autobahnverkehr angewendet. Diese verschiedenen Verkehrssituationen werden im Folgenden als Segmente bezeichnet (abgekürzt in den Diagrammen als „Seg“). Hierbei liegen je Verkehrssituation mehrere Segmente vor. Dies erlaubt -wie später noch beschrieben wird - tiefere Analysen. Zusätzlich wurden die Lidardatensätze mit einem Perlin Noise beaufschlagt, um eine statistische Aussage über die Robustheit der PATS zu machen. Im Folgenden werden anhand der statistischen Auswertung verschiedene Aussagen ermöglicht, indem verschiedene Datensätze miteinander verglichen werden. In Abbildung 4 werden die Precision-Werte der drei verschiedenen Trainingsstände 3k, 60k und 100k miteinander verglichen:

- Es gibt eine große Verbesserung von der Version 2k zur Version 60k.
- Die Verbesserung zwischen 60k und 100k fällt gering aus. Hier ist der Unterschied nur noch so gering, dass die Vertrauensbereichsgrenzen bereits miteinander überlappen. In solchen Fällen müssen für statistisch relevante Aussagen sog. Hypothesentests verwendet werden. Diese liefern für eine bestimmte statistische Aussage eine Irrtumswahrscheinlichkeit, welche als Entscheidungsbasis verwendet wird.
- Es ist (unter bestimmten Annahmen) wahrscheinlich, dass ein weiteres Training nur noch bedingt Verbesserungen erbringt, da der zusätzliche Trainingsaufwand von 60k auf 100k nur noch eine sehr kleine Verbesserung erbracht hat.

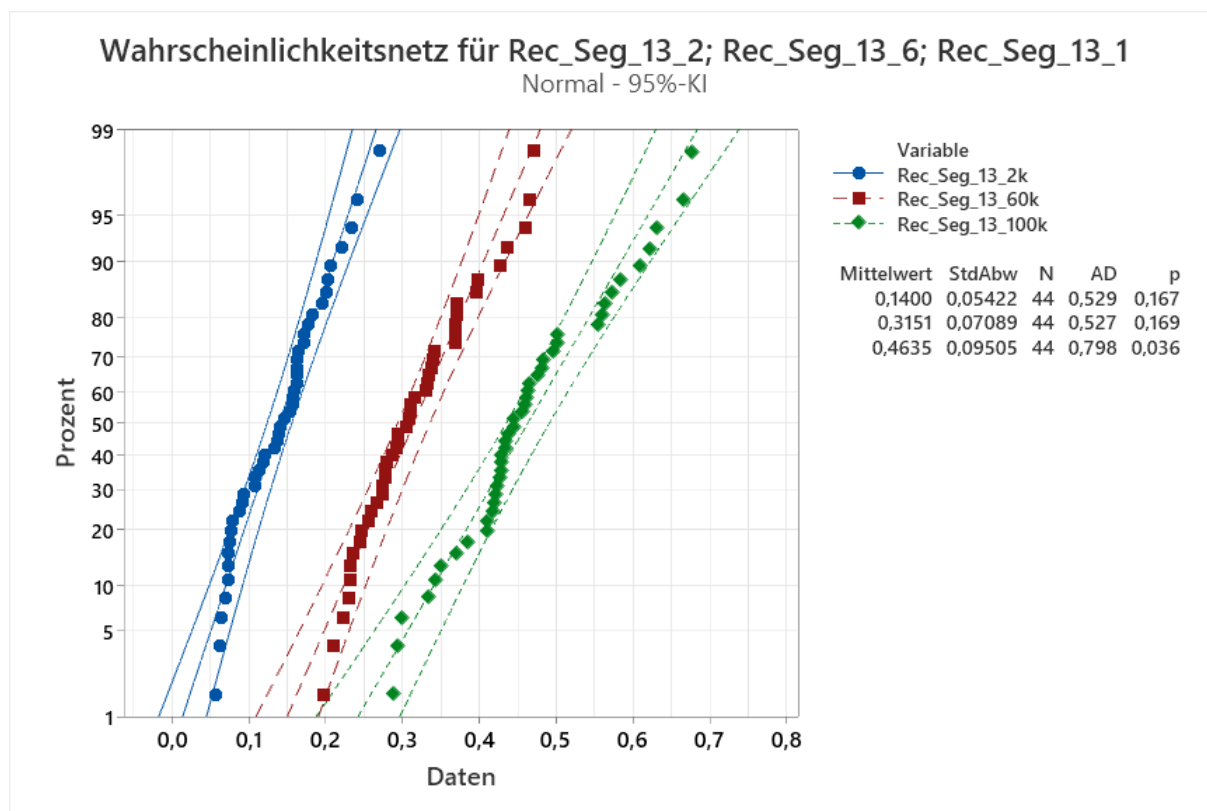


**Abbildung 4:** Vergleich der Trainingsstände 2k, 60k und 100k bzgl. des Precision-Wertes

In Abbildung 5 wird für die 3 Algorithmen die Veränderung des Recall-Wertes untersucht:

- Es ist zu erkennen, dass es noch einen großen „Sprung“ der Recall-Werte vom 60k Trainingsstand auf den 100k Trainingsstand gibt.
- Es gibt hier noch keine „Sättigung“, wie beim Precision-Wert.
- Ein weiteres Training hat noch die Chance weitere Verbesserungen zu erzielen.

Es ist also festzustellen, dass die Werte Precision und Recall ein „Eigenleben“ führen und nicht vollständig miteinander korrelieren.



**Abbildung 5:** Vergleich der Trainingsstände 2k, 60k und 100k bzgl. des Recall-Wertes

Die Robustheit von Perzeptionsalgorithmen stellt ein wichtiges Kriterium für deren Qualität dar. Dementsprechend wurde im Rahmen der durchgeführten Studien auch die Auswirkung von Perlin-Noise Verzerrungen auf die Originaldaten untersucht. Die Einstellungsparameter des Perlin-Noise blieben hierbei immer konstant. Im Folgenden wird die Auswirkung des Perlin-Noise auf die Precision- und Recall-Werte untersucht. Abbildung 6 zeigt zusätzlich die Precision-Werte des 100k-Trainingsstandes bei verzerrten Daten des Datensatzes.

Die beiden Punkteserien sind nahezu deckungsgleich und somit liegt auch nur ein marginaler Unterschied zwischen den beiden hinterliegenden Normalverteilungen vor. Dies ist auch an den in Abbildung 6 angegebenen Werten für den Mittelwert und die Standardabweichung erkennbar, die nur marginal verschieden sind für den Datensatz 100k und 100k\_m (manipuliert). Hieraus lässt sich ableiten, dass die Precision des Perzeptionsalgorithmus 100k robust ist gegenüber dem eingestellten Perlin Noise. Hier könnten weitergehende Studien mit verschiedenen Perlin Noise Leveln die Grenzen der Robustheit aufzeigen.

Analog zur Robustheitsbetrachtung für den Precisionwert kann dies auch für den Recall-Wert durchgeführt werden. Die entsprechenden Ergebnisse sind in Abbildung 7 illustriert. Es ist auch hier erkennbar, dass der Recall-Wert für den 100k Perzeptionsalgorithmus nur minimal vom manipulierten Datensatz abweicht. Somit ist die Robustheit auch hier gegeben.

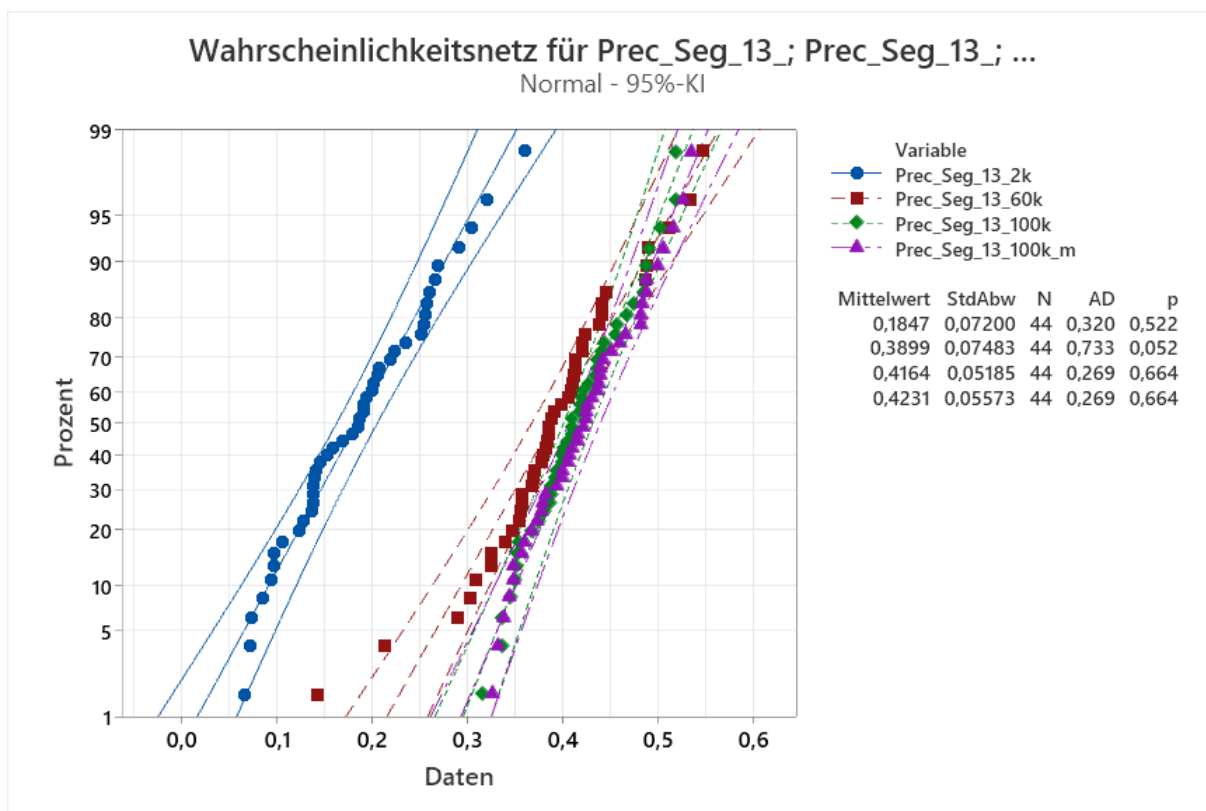


Abbildung 6: Trainingsstand 100k vs. 100k\_m bzgl. Precision.

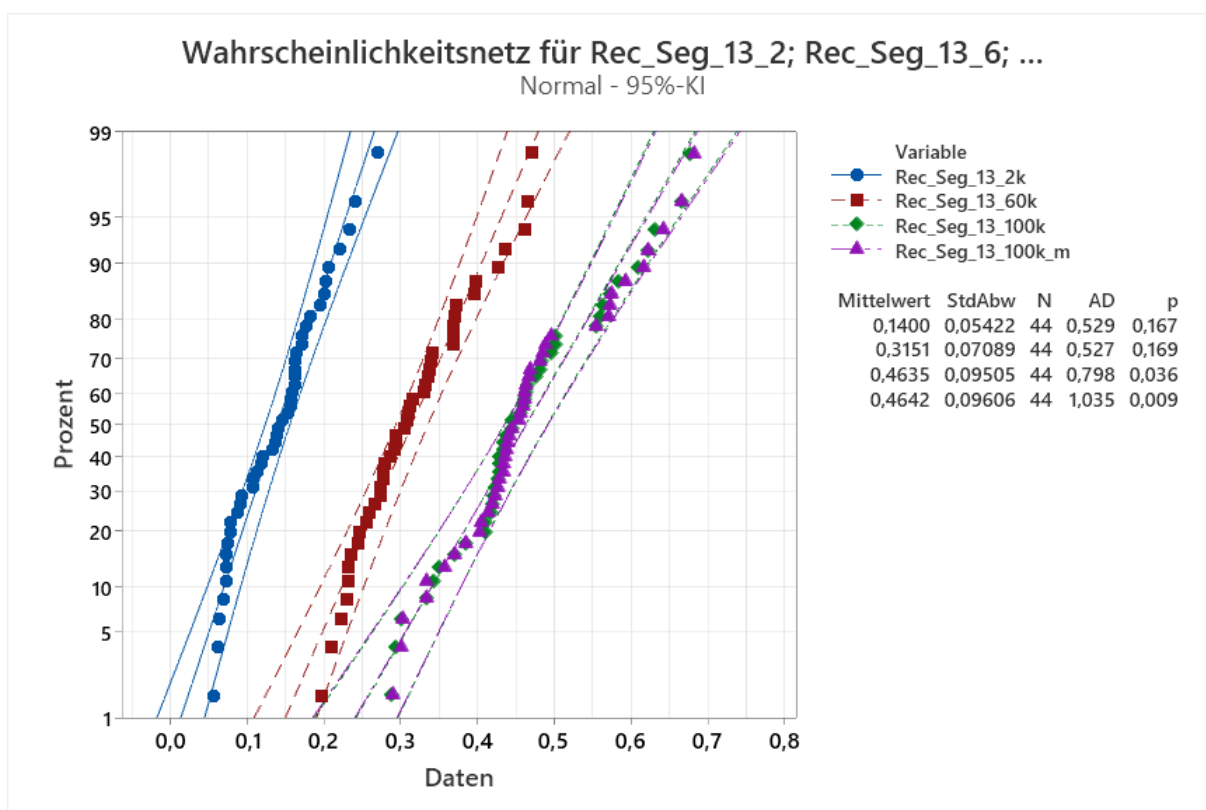
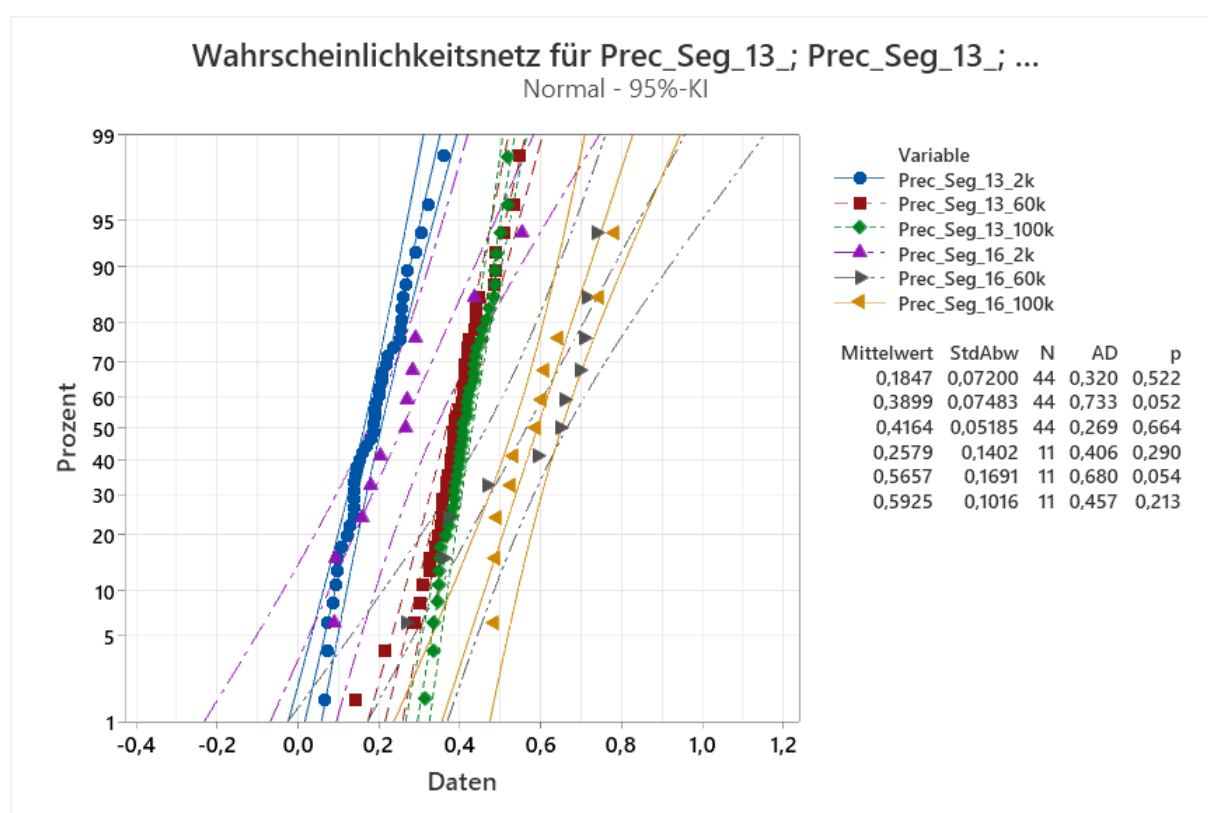


Abbildung 7: Trainingsstand 100k vs. 100k\_m bzgl. Recall.

Die bisherigen Analysen können im realen Entwicklungsprozess verwendet werden, um die Reife, das weitere Verbesserungspotential und die Robustheit von Perzeptionsalgorithmen zu bewerten. Im Rahmen der Entwicklung (von Perzeptionsalgorithmen) können mit statistischer Analyse auch tieferegehende Analysen durchgeführt werden, um mögliche Ursachen für bestimmte Sachverhalte zu identifizieren.

Solche Analysen sind notwendig, um Verbesserungspotentiale zu identifizieren, aber auch um Einsatzgrenzen eines Perzeptionsalgorithmus zu verstehen. In Abbildung 8 sind die beiden Segmente 13 und 16 miteinander verglichen worden bzgl. der Precision. Bei beiden Segmenten handelt es sich um die Verkehrssituation des innerstädtischen Verkehrs. Dennoch ist in Abbildung 8 zu sehen, dass es signifikante Unterschiede des Precisionwertes gibt. Dies deutet darauf hin, dass es zusätzlicher Variablen bedarf, um die unterschiedlichen Verteilungen des Precisionwertes zu erklären.

Auch hierbei kann die Statistik oder auch der Einsatz von KI-Algorithmen helfen, um anhand der vorhandenen Variablen ein Modell bzw. eine Erklärung für die Unterschiede zu finden. Wie in Tabelle 1 zu sehen, lag für jeden Datensatz die Ground Truth Anzahl an echten Objekten vor. Zusätzlich war für jeden Datensatz (Zeile in Tabelle 1) die Aufteilung dieser Objekte in Objekttypen (Cars und Pedestrians) vorhanden. Dementsprechend besteht die Möglichkeit, eine Korrelation zwischen den Variablen Objekttyp und Echtobjektanzahl und den Precisionwerten durchzuführen. Entsprechend dem statistischen Grundsatz, dass Korrelation die notwendige Voraussetzung für Kausation ist, ermöglicht die Korrelationsanalyse die Identifikation von möglichen (aber noch zu verifizierenden) Ursachen für den Unterschied der Precisionwerte bei „gleicher“ Verkehrssituation.



**Abbildung 8:** Es ist bei den 60k und 100k Trainingsständen des Segments 16 eine signifikant bessere Precision vorhanden wie bei den entsprechenden Trainingsständen des Segmentes 13.

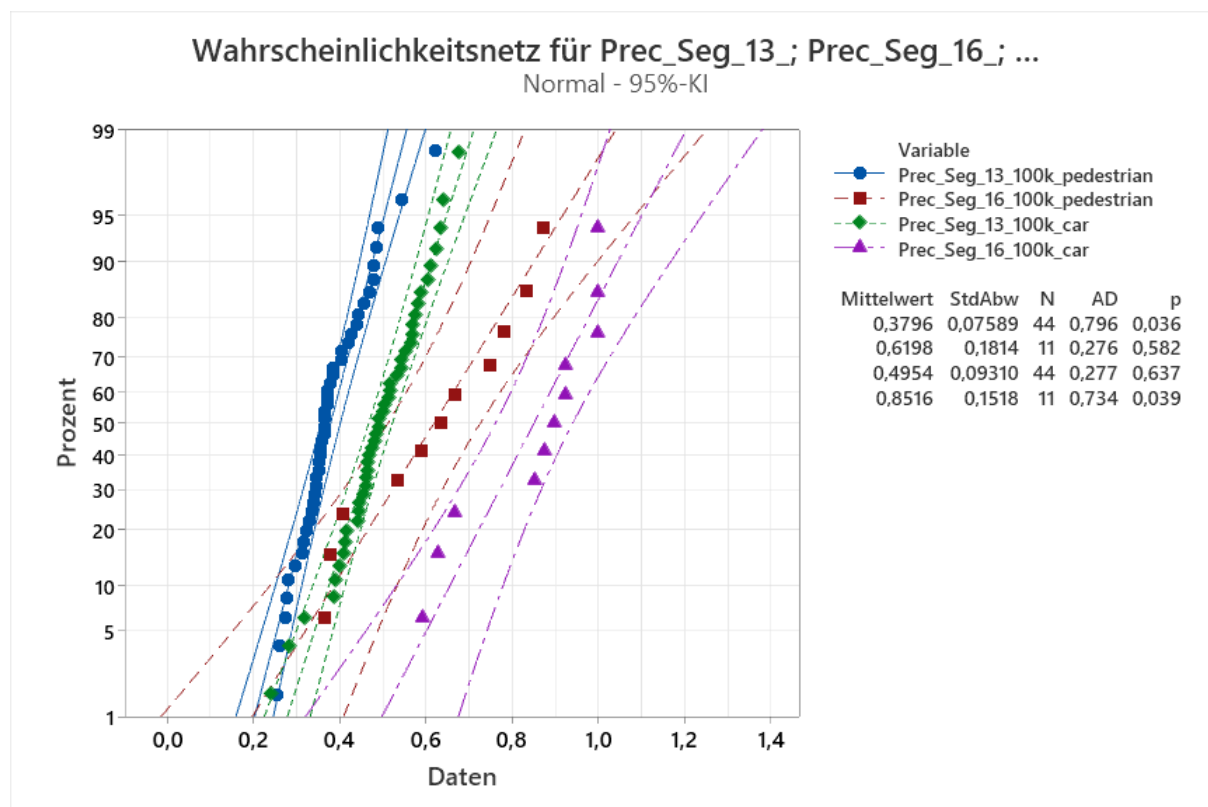
Abbildung 9 zeigt die Auswirkung der Auftrennung der Objekte in die Typen „Car“ und „Pedestrian“. Es ist zu erkennen, dass der Unterschied zwischen Segment 13 und Segment 16 sich bei beiden Objekttypen zeigt. Hieraus lässt sich vermuten, dass der Objekttypus nicht notwendigerweise die erklärende Variable für Unterschied ist.

### Exkurs

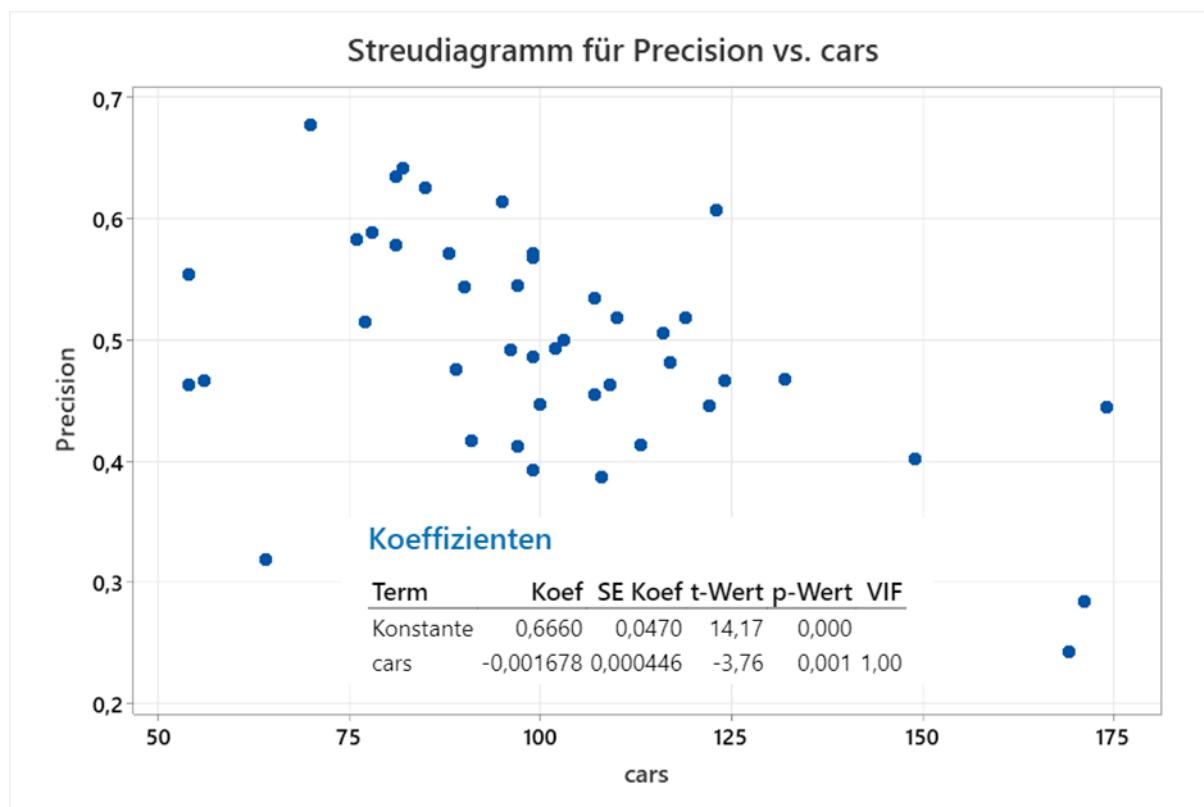
Es lässt sich aber erkennen, dass die Steigung der Normalverteilungen für den Typ „Cars“ deutlich niedriger ist wie beim Typ „Pedestrians“, Hieraus lässt sich schließen, dass die Varianz der Precision abhängig ist vom Objekttypus

### Ende Exkurs

Eine Korrelationsanalyse zwischen Echtojektanzahl und Precisionwerten für den Objekttyp „Cars“ resultiert in Abbildung 10. Die Korrelation ist schon gut grafisch zu erkennen. Je höher die Objektanzahl, desto niedriger werden die Precisionwerte. Es liegt hier eine hohe Streuung vor, was es nötig macht, die Korrelation statistisch auf Signifikanz zu untersuchen. Die entsprechende Analyse ist ebenfalls in Abbildung 10 zu erkennen. Es liegt eine Irrtumswahrscheinlichkeit (p-Wert) von 0,001 vor, d.h. die Korrelation ist hochsignifikant. Eine erhöhte Objektanzahl führt also zu einer niedrigeren Präzision bei der Objekterkennung. Dieser Zusammenhang erscheint auch plausibel und ist somit mit einer hohen Wahrscheinlichkeit auch ein relevanter Einfluss auf die Höhe des Precisionwertes.



**Abbildung 9:** Es ist bei den 60k und 100k Trainingsständen des Segments 16 eine signifikant bessere Precision vorhanden wie bei den entsprechenden Trainingsständen des Segmentes 13.



**Abbildung 10:** Korrelationsanalyse zwischen Echtobjektanzahl und Precisionwerten für den Typus „Cars“

## 6.2 ABSCHÄTZUNG DES UNKNOWN RISK ENTSPRECHEND SOTIF

**Ziel des Arbeitspaketes:** Prüfung der Möglichkeit einer Vorhersagemethodik für das „unknown risk“ entsprechend SOTIF und Ausarbeitung dieser Methodik

**Zielerreichung:** Das vorgegebene Ziel wurde vollumfänglich erreicht. Es wurde eine praktisch verwendbare Methodik gefunden, um einen Schätzwert für das unknown risk basierend auf bisherigen Messdaten zu bestimmen.

Perzeptionsalgorithmen sind letztendlich Software. Input wird verarbeitet und es gibt einen Output. Im Sinne autonom Fahren der Fahrzeuge ist dieser Output hochrelevant für die Zuverlässigkeit und Sicherheit des Fahrzeugs.

Im Rahmen der Entwicklung von Perzeptionsalgorithmen und autonomer Fahrzeuge werden in einer ODD (operational design Domain) viele instrumentierte Messfahrten durchgeführt, um die verschiedenen Szenarien innerhalb einer ODD zu erfassen. Sehr hohe Anzahlen an Messfahrten bei verschiedenen Bedingungen (z.B. Wetter, Tageszeit, Jahreszeit) sollen sicherstellen, dass ein sehr hoher Anteil der möglichen Szenarien identifiziert wurde. Wenn im Idealfall alle ODD-relevanten Szenarien gefunden wurden und darauf basierend ein Perzeptionsalgorithmus trainiert wurde dürften (so die Annahme) keine Perzeptionsfehler basierend auf der unvollständigen Erfassung aller ODD-relevanten Szenarien auftreten. Diese Logik liegt dem Software-Engineering inne. Es gibt viele Beispiele von Softwareausfällen basierend auf einer unvollständigen Erfassung aller Anwendungsfälle (z.B. den Absturz der Ariane 5 Rakete bei ihrem Erstflug).

Die praktische Herausforderung in diesem Themenfeld ist es, mit der Unsicherheit umzugehen, dass eben nicht alle ODD-relevanten Szenarien identifiziert wurden.

Im Rahmen der Zusammenarbeit mit dSPACE konnte JHP ein statistisches Modell entwickeln, welches basierend auf den Ergebnissen einer durchgeführten Messfahrtkampagne eine Auftretensrate für bisher unbekannte Szenarien ermittelt. Hiermit ist es möglich, einen quantitativen Beitrag im Rahmen der Sicherheitsanalyse von autonomen Fahrzeugen zu liefern.

Die gesamtheitliche Darstellung der Zusammenhänge und Ergebnisse erfolgte in der in Kapitel 11 genannten Veröffentlichung [1]. Es soll hier deshalb nur der Kern des Ansatzes beschrieben werden.

Während der Durchführung einer Messkampagne mit instrumentierten Fahrzeugen werden permanent Verkehrs- und Fahrszenarien erfasst. Diese Szenarien werden in einer Datenbank erfasst. Über einen Abgleich der bisher schon identifizierten mit den bei der Messkampagne fortlaufend aufgenommenen Szenarien werden „Neuszenarien“ identifiziert. Über die Dauer der Messkampagne hinweg fällt somit eine Sequenz an Neuszenarien an, versehen mit einem Zeitstempel ihrer Identifikation im Verlauf der Messkampagne.

Erwartungsgemäß müsste die Häufigkeit von Neuszenarien mit zunehmender Dauer der Messkampagne sinken, weil immer weniger bisher unbekannte Szenarien „übrig“ sind. Abbildung 11 zeigt exemplarisch eine solche Ereignissequenz auf einem Zeitstrahl.



**Abbildung 11:** Ereignisse auf Zeitstrahl mit sinkender Ereignisintensität

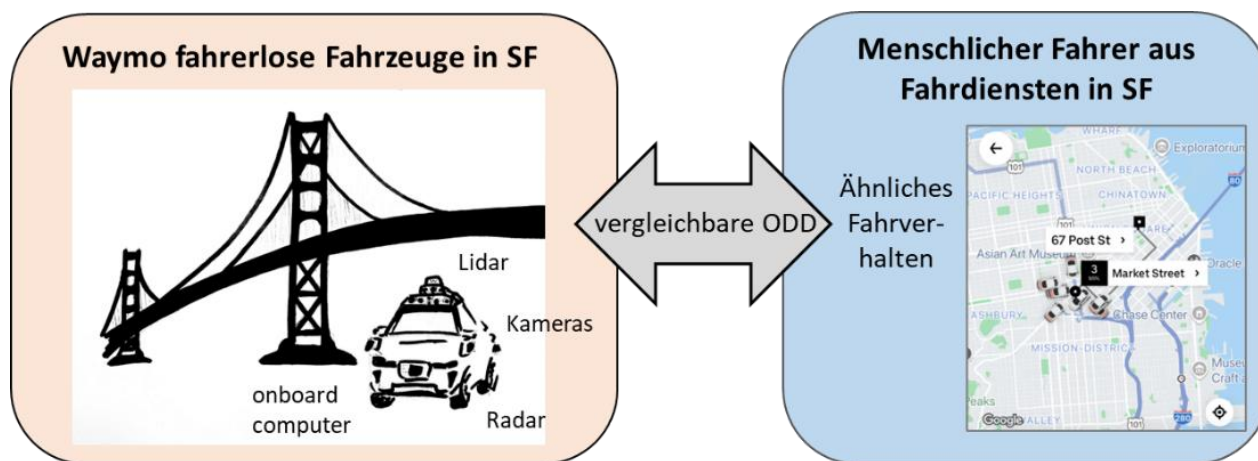
Die Basis des statistischen Ansatzes zur Ableitung eines Wahrscheinlichkeitsmaßes für das Auftreten bisher unbekannter Szenarien ist das Verständnis des Auftretens bisher unbekannter Fahrszenarien während einer Messkampagne als sog. nichthomogenen Poissonprozess. Basierend auf der bisherigen Sequenz an Neuszenarien, welche während einer Messkampagne angefallen ist, lässt sich ein stochastisches Modell parametrieren. Das hier verwendete Modell ist das sog. Crow-AMSAA Modell, welches in der Zuverlässigkeitstechnik von großer Bedeutung ist bei der Erprobung und Reifegradbewertung von Systemen. Das Modell liefert auf Basis der Daten eine Schätzung für die Auftretensrate bisher unbekannter Fahrszenarien und kann somit einen wichtigen Beitrag liefern für die Entwicklung und Absicherung autonomer Fahrzeuge.

### 6.3 STATISTISCHER VERGLEICH VON AUTONOMEN FAHRZEUGEN MIT MENSCHLICHEN FAHRERN

**Ziel des Arbeitspaketes:** Erarbeitung eines praktisch anwendbaren Vorgehens für den statistischen Abgleich von Feldunfallintensitäten als Basis für die Erbringung eines Feldnachweises autonomer Fahrzeuge.

**Zielerreichung:** Das vorgegebene Ziel wurde vollumfänglich erreicht. Es konnte gezeigt werden, dass mit bekannten statistischen Methoden ein solcher Nachweis darstellbar ist. Zusätzlich dazu wurde gezeigt, dass ein praktischer feldversuchsbasierter Nachweis nicht notwendigerweise mit extremen Erprobungsfahrleistungen verbunden sein muss, da die menschliche Unfallhäufigkeit vergleichsweise hoch ist.

Im Sinne einer Absicherung autonomer Fahrfunktionen und Gesamtfahrzeuge stellt auch der Abgleich von autonomen Fahrzeugen mit menschlichen Fahrern ein interessantes Arbeitsgebiet dar, da gefordert wird, dass autonome Fahrzeuge hinsichtlich ihres Risikos nicht schlechter sein dürfen als menschliche Fahrer. Dieser praktisch relevante Abgleich wurde von JHP im RDV-Projekt prototypisch durchgeführt. Eine große Schwierigkeit dabei ist es eine vergleichbare Datenbasis von menschlichen Fahrern und autonomen Fahrzeugen mit ähnlichem Fahrverhalten und vergleichbaren ODD (Operational Design Domain) zu schaffen. Während den letzten Jahren wurden Tests mit autonomen Fahrzeugen auf öffentlichen Straßen durchgeführt, wobei eine hohe Anzahl von gefahrenen Meilen zurückgelegt wurde. Die Datenbank des California Department of Motor Vehicles (DMV) enthält Informationen über autonome Fahrzeuge und ihre Unfälle sowie über Disengagements, wenn der/die Testfahrer/-in übernehmen musste. Eine solide und umfassende Datenbasis liefert die fahrerlose Waymo Flotte, die als Robotaxis in San Francisco betrieben werden. Eine vergleichbare Datenbasis menschlicher Fahrer aus Fahrdiensten liefert eine im September 2023 veröffentlichte Fallstudie. Es wurden über einen Zeitraum von zwei Jahren Unfalldaten sowie Laufleistungen von menschlichen Fahrern aus "ridehail" Fahrdiensten im selben Gebiet von San Francisco gesammelt und Unfallraten zur Verfügung gestellt. Eine vergleichbare Datenbasis war somit gegeben.



**Abbildung 12:** Ein realistischer Abgleich erfordert Daten aus demselben ODD

Die Ursachen für Verkehrsunfälle durch menschliche Fahrer im Vergleich zu autonomen Fahrzeugen werden sich stark unterscheiden. Unfälle durch menschliche Fahrer entstehen u.a. durch abgelenktes Fahren, überhöhte Geschwindigkeit, Trunkenheit am Steuer oder Müdigkeit. Autonome Fahrzeuge hingegen können Unfälle insbesondere aufgrund von Softwarefehlern (SOTIF) verursachen. Der Vergleich erfolgt daher im Allgemeinen.

Die Schlüsselfrage für eine feldbasierte Absicherung autonomer Fahrzeuge lautet: Wie viele Meilen müssen autonome Fahrzeuge unfallfrei fahren, um nachzuweisen, dass sie hinsichtlich des Unfallgeschehens mindestens so zuverlässig sind wie der menschliche Fahrer?

JHP erarbeitete für diese Fragestellung anhand der genannten Daten ein praktisch einsetzbares statistisches Nachweisvorgehen, welches unterschiedliche Unfallschweren berücksichtigen kann und den Umstand miteinbezieht, dass ein Unfall meistens von einem der beiden Unfallbeteiligten verursacht wird. Die hinterliegende Statistik basiert auf der sog. Chi-Square-Verteilung, welche verwendet werden kann, um den Vertrauensbereich exponentialverteilter Ereignisintensitäten und mittlerer Zeitdauern zwischen Ereignissen (MTBF) zu bestimmen.

Der einseitige Vertrauensbereich der MTBF für exponentiell verteiltes Ausfallverhalten beträgt:

$$MTBF_{\min} = \frac{2 \sum_{i=1}^n t_i}{\chi_{2(x+1);1-\alpha}^2} = \frac{2 \sum_{i=1}^n t_i}{\chi_{2(x+1);C}^2}$$

$t_i$  = Laufleistung pro Fahrzeug  $i$

$x$  = Anzahl Ausfälle

$n$  = Stichprobenumfang

$\chi^2$  = Wert der Chi<sup>2</sup> Verteilung

$\alpha$  = Irrtumswahrscheinlichkeit

$C$  = Confidence Level

Der statistische Nachweis basiert dabei auf drei Schritten. Zunächst wird die MTBF für den menschlichen Fahrer anhand von Unfalldaten bestimmt. Dieser Wert stellt die Referenz dar. Dann wird bestimmt, wie hoch die MTBF der fahrerlosen Fahrzeuge sein muss, um mit einer Aussagesicherheit von  $C$  aussagen zu können, dass letztere eine niedrigere Unfallintensität (höherer MTBF) haben als die menschlichen Fahrer. Dieser Wert kann dann verglichen werden mit den realen Testergebnissen der autonomen Fahrzeuge.

Das vollständige Vorgehen wurde in der in Kapitel 11 genannten Veröffentlichung [2] dargelegt, weshalb hier auf weitere Einzelheiten verzichtet wird.

## 7 WICHTIGSTE POSITIONEN DES ZAHLENMÄßIGEN NACHWEISES

Die wichtigste Position des zahlenmäßigen Nachweises sind die Personalkosten. Genaue Angaben zur Höhe und die weiteren Positionen können dem zahlenmäßigen Nachweis entnommen werden.

## 8 NOTWENDIGKEIT UND ANGEMESSENHEIT DER GELEISTETEN ARBEIT

Um die in diesem Teilvorhaben umgesetzten Arbeiten durchzuführen, war eine externe Finanzierung notwendig, da es sich bei den von JHP durchgeführten Arbeiten und Ansätzen um bisher in der Industrie in dem Arbeitsfeld autonomes Fahren und KI-Perzeptionsalgorithmen um neue Betrachtungsweisen handelt. Dies war bei der Diskussion unserer Ansätze mit den Projektpartnern deutlich zu erkennen. Unsere Betrachtungsweisen und Vorschläge waren für die RDV-Projektpartner neuartig. Dementsprechend fiel zu Beginn des Projektes auch viel Aufwand für die Herstellung des gegenseitigen Verständnisses an. Nur durch die Struktur eines Verbundprojektes war der intensive Austausch und die Entwicklung eines gegenseitigen Verständnisses und Arbeitsklimas möglich.

JHP konnte im Rahmen des RDV-Projektes aufzeigen, dass bisherige Betrachtungsweisen teilweise falsch sind und dass es Lösungen für bisher unbeantwortete Probleme gibt. So wurde zu Beginn des Projektes von vielen Projektpartnern die Sichtweise vertreten, dass ein Feldnachweis autonomer Fahrzeuge praktisch unmöglich sei, weil die Erprobungsleistungen viel zu hoch wären. Dies konnte teilweise – zumindest für den Stadtbetrieb – widerlegt werden. Die erarbeitete Vorgehensweise zur Quantifizierung des Risikos bisher unentdeckter Fahrzenarien basierend auf bisherigen Messfahrtdaten war gänzlich unbekannt und somit neuartig. Auch die Nutzung statistischer Ansätze in der Entwicklung von KI-Perzeptionsalgorithmen und deren Nutzen war den Projektpartnern bisher unbekannt.

Mit diesem Teilvorhaben wurden somit Beiträge erarbeitet, die hohen praktischen Nutzen haben.

## **9 VORAUSSICHTLICHER NUTZEN, INSBESONDERE DER VERWERTBARKEIT DES ERGEBNISSES IM SINNE DES FORTGESCHRIEBENEN VERWERTUNGSPLANS**

Die JHP GmbH ist für unsere Kunden aus der Industrie ein Beratungs- und Projektpartner. Unsere Aufgabe in den Projekten ist es, mit unseren Betrachtungsweisen konkrete Fragestellungen schnellst möglich zu lösen. Dazu ist es notwendig, den Kunden und seine Fragestellung zu verstehen aber auch den fachlichen Hintergrund nachvollziehen zu können, da für die Erarbeitung eines gemeinsamen Verständnisses typischerweise keine Zeit vorhanden ist. Dementsprechend ist der praktische Nutzen der Arbeiten im Rahmen des RDV-Projektes für die JHP GmbH sehr hoch, da wir uns mit dem Thema autonomes Fahren / KI-Perzeptionsalgorithmen in ein neues Themenfeld einarbeiten konnten und ein Verständnis dafür entwickeln konnten, welche Fragestellungen hier für unsere Kunden praktisch relevant sind und wie mögliche Lösungen aussehen können.

Insbesondere die Quantifizierung des Risikos unbekannter Fahrscenarien ist aus unserer Sicht eine Lösung an der viel Interesse seitens der Industrie besteht. Um unsere Sichtbarkeit hierzu zu steigern, wurde mit dem Projektpartner dSPACE eine themenbezogene Veröffentlichung in einer Fachzeitschrift erarbeitet.

Über unsere Ausbildungsfirma Reliability Engineering Academy GmbH & Co. KG schulen wir Mitarbeiter aus Industriefirmen zu den Themen der Zuverlässigkeitstechnik. Die im Rahmen des RDV-Projektes erarbeitete Ansätze und Vorgehensweisen können dort in das Schulungsmaterial eingehen, wenn es um Fragen der Softwarezuverlässigkeit geht.

## **10 WÄHREND DER DURCHFÜHRUNG DES VORHABENS BEKANTT GEWORDENER FORTSCHRITT AUF DEM GEBIET DES VORHABENS BEI ANDEREN STELLEN**

Während des laufenden Projektes haben Recherchen keine Ergebnisse erbracht, welche für die Arbeiten oder Ergebnisse unseres Teilvorhabens im RDV-Projekt relevant waren.

## **11 ERFOLGTE ODER GEPLANTE VERÖFFENTLICHUNGEN DES ERGEBNISSES**

Die Arbeiten dieses Teilvorhabens mündeten in zwei Veröffentlichungen:

[1] Jäger, P., Harlammert, N., De Candido, O., Yildiz, G. (2025). Vom Unsichtbaren zum Messbaren. In HANSER automotive Ausgabe 4/2025 Seite 22-25.

[2] Jäger, P., Krolo, M., Jäger, U. (2025). Approach for a reliability proof of autonomous vehicles compared to human drivers through real-world road tests. In Proceedings of the 35th European Safety and Reliability Conference (ESREL 2025), 15.-19.06.2025, Stavanger, Norway.