

# Technische Hochschule Ingolstadt

## Masterarbeit

### Clustern von realen Messdaten zur Identifikation von Fahrscenarien

vorgelegt von

**Eduard Springer**

**Matrikelnummer: 58652**

**Studiengang:** Informatik

**Fakultät:** Informatik

**Ausgabetermin:** 12.07.2021

**Abgabetermin:** 31.08.2021

**Erstprüfer:** Prof. Dr. Melanie Kaiser

**Zweitprüfer:** Prof. Dr. Michael Botsch

**Betreuer:** M. Sc. Francesco Montanari, AUDI AG

# Danksagung

Diese Masterarbeit wurde im Rahmen des Informatikstudiums an der Technischen Hochschule Ingolstadt in Kooperation mit AUDI AG angefertigt.

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Masterarbeit unterstützt und motiviert haben.

Zuerst gebührt mein Dank **Frau Prof. Melanie Kaiser**, die meine Masterarbeit betreut und begutachtet hat. Für die hilfreichen Anregungen und die konstruktive Kritik bei der Erstellung dieser Arbeit möchte ich mich herzlich bedanken.

Ein besonderer Dank gilt meinem Betreuer **Francesco Montanari** für die ausgezeichnete Betreuung, ohne ihn diese Arbeit nicht hätte entstehen können. Mein Dank gilt seiner Hilfsbereitschaft und seinen interessanten Ideen, die maßgeblich dazu beigetragen haben, dass diese Masterarbeit in dieser Form vorliegt.

Ich danke auch allen **Mitarbeitern des SAVE-Projekthauses und der Abteilung EZ-I** für die angenehme Arbeitsatmosphäre, Zusammenarbeit und für die informativen Beiträge in den Videokonferenzen.

Abschließend möchte ich mich bei meinen Eltern **Valentina und Alexander** und meiner Freundin **Anastasia** bedanken, die mir mein Studium durch ihre Unterstützung ermöglicht haben.

## Kurzfassung

Mit der Einführung automatisierter und komplexer Fahrfunktionen ist es auf Dauer nicht mehr wirtschaftlich, diese im Rahmen von klassischen Testfahrten zu validieren. Aus diesem Grund müssen zunehmend Verkehrssituationen, sogenannte Fahrszenarien, für den Test- und Freigabeprozess des automatisierten Fahrens herangezogen werden. In dieser Arbeit wird eine Methode entwickelt, die Fahrszenarien in realen Fahrzeugmessdaten erkennt und anhand eines maschinellen Lernalgorithmus ähnliche Muster gruppiert. Diese Gruppierungen von gefundenen und ähnlichen Mustern werden anschließend als potenzielle Szenarien betrachtet. Die entwickelte Methode basiert auf einer bereits veröffentlichten Methode zur Erkennung von Fahrszenarien in realen Messdaten, welche jedoch noch viel Potenzial bietet. Dieses Potenzial wird ausgenutzt, sodass Erweiterungen eingeführt werden mit dem Ziel, die Möglichkeiten des Auffindens weiterer Fahrszenarien zu demonstrieren.

## Abstract

With the introduction of automated and complex driving functions, validating them within the framework of classic test drives is no longer economical in the long term. For this reason, traffic situations, so-called driving scenarios, increasingly have to be used for the test and approval process of automated driving. In this work, a method is developed that recognises driving scenarios in real vehicle measurement data and groups similar patterns using a machine learning algorithm. These groupings of found and similar patterns are then considered as potential scenarios. The developed method is based on an already published method for recognising driving scenarios in real measurement data, which however still offers a lot of potential. This potential is utilised so that extensions are introduced with the aim of demonstrating the possibilities of finding further driving scenarios.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>I</b>
<b>Tabellenverzeichnis</b>	<b>I</b>
<b>Diagrammverzeichnis</b>	<b>II</b>
<b>Abkürzungsverzeichnis</b>	<b>IV</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Herausforderung bei der Absicherung und Freigabe des automatisierten Fahrens . . . .	1
1.2 Zielsetzung . . . . .	2
1.3 Struktur der Arbeit . . . . .	3
<b>2 Grundlagen des automatisierten Fahrens</b>	<b>4</b>
2.1 Begriffsdefinitionen . . . . .	4
2.2 Szenarien zur Entwicklung, Absicherung und Erprobung . . . . .	7
2.3 Definition verschiedener Automatisierungsgrade . . . . .	9
<b>3 Grundlagen des maschinellen Lernens</b>	<b>11</b>
3.1 Maschinelles Lernen zur Mustererkennung . . . . .	12
3.2 Clusteranalyse . . . . .	13
3.3 Agglomeratives Cluster-Verfahren . . . . .	16
<b>4 Forschungsprojekte und Stand der Technik</b>	<b>18</b>
4.1 Forschungsprojekte zum automatisierten Fahren . . . . .	18
4.2 Szenariobasierte Ansätze . . . . .	20
4.2.1 Wissensbasierter Ansatz . . . . .	20
4.2.2 Datenbasierter Ansatz . . . . .	21
4.2.2.1 Überwachtes Lernen . . . . .	21
4.2.2.2 Unüberwachtes Lernen . . . . .	22
<b>5 Explorativer Ansatz zur Identifikation von Fahrscenarien</b>	<b>25</b>
5.1 Idee und Aufbau . . . . .	25
5.2 Einführung von Erweiterungen . . . . .	28
5.2.1 Einsatz von großen Fahrzeugmessdatensmengen . . . . .	28
5.2.2 Kategorisierung von ausgewählten Signalen . . . . .	28
5.2.3 Gleichzeitige Untersuchung von mehreren Signalen . . . . .	29
5.2.4 Anwendung von weiteren Features . . . . .	29



---

<b>6</b>	<b>Ergebnisse und Evaluierung</b>	<b>31</b>
6.1	Kategorie 1: Verkehrsteilnehmer . . . . .	31
6.2	Kategorie 2: Ego . . . . .	40
6.3	Kategorie 3: Umwelt . . . . .	46
6.4	Grenzen der Methode . . . . .	55
6.5	Diskussion . . . . .	55
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>57</b>
7.1	Zusammenfassung . . . . .	57
7.2	Ausblick . . . . .	58
	<b>Anhang</b>	<b>IV</b>
	<b>Literaturverzeichnis</b>	<b>XVIII</b>
	<b>Eidesstattliche Erklärung</b>	

# Abbildungsverzeichnis

1.1	Struktur der Arbeit . . . . .	3
2.1	Ein Beispiel für eine Trajektorie im Koordinatensystem . . . . .	4
2.2	UML-Diagramm zu den Begriffen eines Szenarios . . . . .	6
2.3	Verschiedene Abstraktionsebenen von Szenarien im V-Modell-basierten Entwicklungsprozess . . . . .	7
2.4	UML-Diagramm zu den verschiedenen Abstraktionsebenen von Szenarien . . . . .	8
2.5	5-Ebenen-Modell zur Beschreibung von Szenarien . . . . .	8
2.6	Definition der Automatisierungsebenen nach SAE J3016 . . . . .	9
3.1	Übersicht über die verschiedenen Zweige der künstlichen Intelligenz . . . . .	11
3.2	Darstellung eines Dendrogramms basierend auf vorgegebener Datenpunkte . . . . .	17
4.1	Prozess des wissensbasierten Ansatzes zur Erstellung von Szenarien . . . . .	20
4.2	Prozess des datenbasierten Ansatzes zur Erstellung von Szenarien . . . . .	21
4.3	Prozess des Clustering-Ansatzes nach Montanari, German und Djanatliev . . . . .	23
4.4	Ein Beispiel für die Zerlegung einer Zeitreihe anhand des Beschleunigungssignals . . . . .	24
4.5	Ein Beispiel für die Zerlegung einer Zeitreihe mit Verarbeitungsschritt . . . . .	24
5.1	Erweiterungsschritte in Hinblick auf die Prozesskette des Ansatzes nach Montanari, German und Djanatliev 2020 . . . . .	25
5.2	Aufbau der Methode dieser Arbeit . . . . .	26
6.1	Erkannte potenzielle Szenarien für die Verkehrsteilnehmer-Kategorie . . . . .	39
6.2	Erkannte potenzielle Szenarien für die Ego-Kategorie . . . . .	45
6.3	Erkannte potenzielle Szenarien für die Umwelt-Kategorie . . . . .	54
7.1	Vorschläge für die möglichen Anwendungen der ausgearbeiteten Methodik . . . . .	58

# Tabellenverzeichnis

5.1	Übersicht über die kategorisierten Fahrzeugsignale . . . . .	29
-----	--	----

5.2 Übersicht über die in dieser Arbeit verwendeten Features pro Kategorie . . . . . 30

6.1 Überblick über die verwendeten Signale und Features für die Kategorie Verkehrsteilnehmer 31

6.2 Überblick über die verwendeten Signale und Features für die Ego-Kategorie . . . . . 40

6.3 Ego-Kategorie: Überblick über die vorgeschlagenen Werte für die optimale Clusteranzahl . 42

6.4 Überblick über die verwendeten Signale und Features für die Umwelt-Kategorie . . . . . 46

6.5 Umwelt-Kategorie: Überblick über die vorgeschlagenen Werte für die optimale Clusteranzahl 48

6.6 Vergleich Francescos Methode mit meiner . . . . . 56

## Diagrammverzeichnis

6.1 Ein Beispiel für das Schneiden des Signals „Position y“ der Verkehrsteilnehmer-Kategorie mit Prominenzwert 1 . . . . . 32

6.2 Ein Beispiel für das Schneiden des Signals „Position y“ der Verkehrsteilnehmer-Kategorie mit Prominenzwert 5 . . . . . 33

6.3 Berechnung einer optimalen Clusteranzahl mittels Distortion . . . . . 34

6.4 Berechnung einer optimalen Clusteranzahl mittels Calinski-Herabsz . . . . . 34

6.5 Berechnung einer optimalen Clusteranzahl mittels Silhouette . . . . . 35

6.6 Ein Dendrogramm mit 20 Clustern als Ergebnis des Clusters von Signalen der Verkehrsteilnehmer-Kategorie . . . . . 36

6.7 Trajektorien der Verkehrsteilnehmer . . . . . 37

6.8 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 19 . . . . . 38

6.9 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 16 . . . . . 38

6.10 Ein Beispiel für das Schneiden des Signals „Geschwindigkeit“ der Ego-Kategorie mit Prominenzwert 1 . . . . . 41

6.11 Ein Beispiel für das Schneiden des Signals „Geschwindigkeit“ der Ego-Kategorie mit Prominenzwert 8 . . . . . 41

6.12 Ein Dendrogramm mit 7 Clustern als Ergebnis des Clusters von Signalen der Ego-Kategorie . . . . . 42

6.13 Alle geclusterten Signale der Ego-Kategorie . . . . . 43

6.14 Alle geclusterten Signale der Ego-Kategorie mit Referenz zur Null-Linie . . . . . 44

6.15 Ein Boxplot mit Feature-Werten des Signals „Geschwindigkeit“ für alle Cluster der Ego-Kategorie . . . . . 45

6.16 Ein Beispiel für das Schneiden des Signals „Linie rechts“ der Umwelt-Kategorie mit Prominenzwert 0.1 . . . . . 47

6.17 Ein Beispiel für das Schneiden des Signals „Linie rechts“ der Umwelt-Kategorie mit Prominenzwert 1 . . . . . 47

6.18 Ein Dendrogramm mit zehn Clustern als Ergebnis des Clusters von Signalen der Umwelt-Kategorie . . . . . 48

6.19 Signale der Umwelt-Kategorie aus Cluster 1 . . . . . 49

6.20 Signale der Umwelt-Kategorie aus Cluster 6 . . . . . 50

6.21 Signale der Umwelt-Kategorie aus Cluster 7 . . . . . 50

6.22 Signale der Umwelt-Kategorie aus Cluster 8 . . . . . 51

6.23 Signale der Umwelt-Kategorie aus Cluster 9 . . . . . 51

6.24 Signale der Umwelt-Kategorie aus Cluster 10 . . . . . 52

6.25 Ein Boxplot mit Feature-Werten des Signals „Linie links“ für alle Cluster der Umwelt-Kategorie . . . . . 53

6.26 Ein Boxplot mit Feature-Werten des Signals „Linie rechts“ für alle Cluster der Umwelt-Kategorie . . . . . 53

A.1 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 1 . . . . . IV

A.2 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 2 . . . . . V

A.3 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 3 . . . . . V

A.4 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 4 . . . . . VI

A.5 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 5 . . . . . VI

A.6 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 6 . . . . . VII

A.7 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 7 . . . . . VII

A.8 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 8 . . . . . VIII

A.9 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 9 . . . . . VIII

A.10 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 10 . . . . . IX

A.11 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 11 . . . . . IX

A.12 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 12 . . . . . X

A.13 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 13 . . . . . X

A.14 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 14 . . . . . XI

A.15 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 15 . . . . . XI

A.16 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 17 . . . . . XII

A.17 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 18 . . . . . XII

A.18 Signale der Verkehrsteilnehmer-Kategorie aus Cluster 20 . . . . . XIII

A.19 Ein Boxplot mit Feature-Werten des Signals „Position x“ für alle Cluster der Verkehrsteilnehmer-Kategorie . . . . . XIII

A.20 Ein Boxplot mit Feature-Werten des Signals „Position y“ für alle Cluster der Verkehrsteilnehmer-Kategorie . . . . . XIV

A.21 Signale der Umwelt-Kategorie aus Cluster 2 . . . . . XV

A.22 Signale der Umwelt-Kategorie aus Cluster 3 . . . . . XV

A.23 Signale der Umwelt-Kategorie aus Cluster 4 . . . . . XVI

A.24 Signale der Umwelt-Kategorie aus Cluster 5 . . . . . XVI

# Abkürzungsverzeichnis

<b>AININ</b>	Artificial Intelligence Network Ingolstadt gGmbH
<b>BMVI</b>	Bundesministerium für Verkehr und digitale Infrastruktur
<b>CARISSMA</b>	Center of Automotive Research on Integrated Safety Systems and Measurement Area
<b>FAS</b>	Fahrerassistenzsystem
<b>GPS</b>	Global Navigation Satellite System
<b>IEC</b>	International Electrotechnical Commission
<b>ISO</b>	International Organization for Standardization
<b>ML</b>	maschinelle Lernen
<b>PEGASUS</b>	Projekt zur Etablierung von generell akzeptierten Gütekriterien, Werkzeugen und Methoden sowie Szenarien und Situationen zur Freigabe hochautomatisierter Fahrfunktionen
<b>SAE</b>	Society of Automotive Engineers
<b>SAVe</b>	Funktions- und Verkehrs-Sicherheit im Automatisierten und Vernetzten Fahren
<b>THI</b>	Technische Hochschule Ingolstadt

# 1 Einleitung

»Wir können aufhören weitere Fahrfunktionen zu entwickeln, wenn wir keine Methoden finden, um die Systeme zu testen und abzusichern.« [Schuldt 2017, S. 3]

(Prof. Dr.-Ing. Markus Maurer, TU Braunschweig)

## 1.1 Herausforderung bei der Absicherung und Freigabe des automatisierten Fahrens

Die Fahrzeuge verfügen bereits heute über eine breite Auswahl an Fahrerassistenzsystem (FAS), die dem Fahrer Sicherheit, Komfort und Effizienz ermöglichen. Auf dem Weg vom assistierten Fahren zum vollautomatisierten Fahren spielen Testmethoden für die Freigabe von Fahrerassistenzsystemen eine immer wichtigere Rolle, da die Systeme und ihre gegenseitige Interaktion beim Übergang zum vollautomatisierten Fahren ein hohes Maß an Komplexität erreichen. Eine der größten Herausforderungen ist die Absicherung und Freigabe von Fahrfunktionen nach ISO26262 im Hinblick auf die funktionale Sicherheit. Es bestehen hierfür eine Reihe von Ansätzen. Laut Prof. Maurer benötigt ein gängiger streckenbasierter Fahrtest 6,62 Milliarden Kilometer, um eine einzige Funktion des automatisierten Fahrzeugs zu qualifizieren [vgl. Maurer u. a. 2015, S. 458].

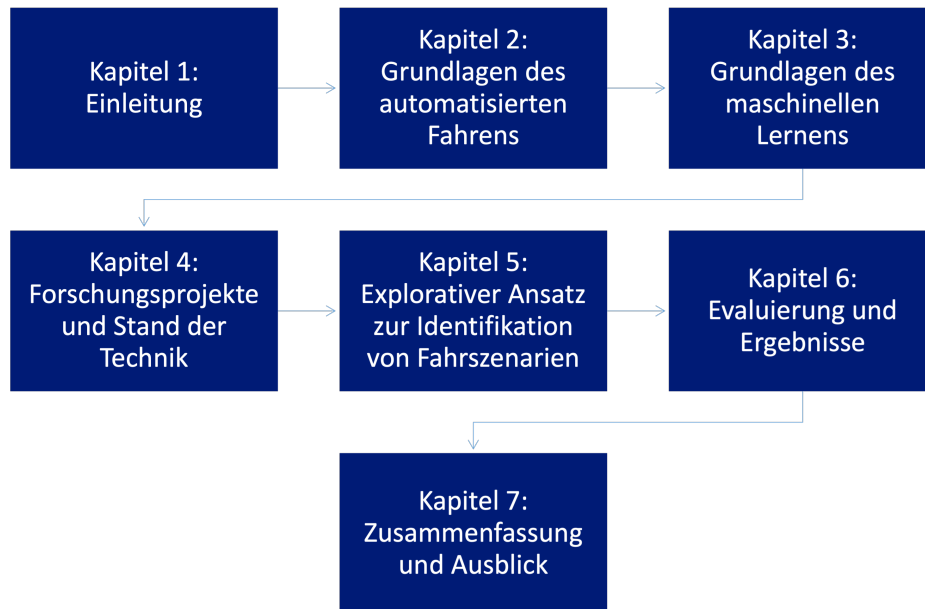
Dieser streckenbasierte Ansatz ist für die Automobilindustrie aus zeitlicher und wirtschaftlicher Sicht nicht praktikabel. Eine andere Möglichkeit ist das szenariobasierte Testen von automatisierten Fahrfunktionen. Dabei werden Testkataloge aus einzelnen Fahrscenarien, z.B. von Domänenexperten, erstellt und anschließend validiert. Die Anzahl der gefundenen Szenarien hängt dabei vom Wissen der Experten ab. Das bedeutet im Umkehrschluss, dass solche Testkataloge möglicherweise lückenhaft beschriebene oder unrealistisch dargestellte Szenarien enthalten. Wesentlich erfolgversprechender ist die Identifizierung von Verkehrssituationen, den sogenannten Fahrscenarien, in realen Messdaten. Exemplarisch wird ein solcher Ansatz in der Veröffentlichung von Montanari, German und Djanatliev beschrieben. Bei diesem Ansatz geht es darum, aus realen Fahrzeugmessdaten automatisiert Fahrscenarien zu extrahieren und darüber hinaus auf die komplexe und zeitaufwändige Erstellung von Vordefinitionen und Regeln zu verzichten. Das bedeutet, dass in den Messdaten nach Mustern gesucht wird und diese als mögliche Fahrscenarien gedeutet werden. Zu diesem Zweck werden die eingelesenen Fahrzeugmessdaten mittels eines intelligenten Schnitalgorithmus in Teile zerlegt. Im nächsten Schritt werden diese Teile mittels maschinellen Lernens den entstehenden Gruppen (Clustern) zugeordnet. Jedes Cluster entspricht einem potenziellen Szenario, das zum Testen von automatisierten Fahrfunktionen genutzt werden kann [vgl. Montanari, German und Djanatliev 2020, S. 1].

## 1.2 Zielsetzung

Die szenariobasierte Entwicklung und Erprobung des automatisierten und vernetzten Fahrens erfordert eine bisher unerreichte Anzahl von Fahrscenarien. Im Rahmen dieser Arbeit soll ein datenbasiertes Verfahren entwickelt werden, das aus realen Fahrzeugmessdaten automatisch Fahrscenarien ableitet. Damit dieses Verfahren selbstständig für Fahrscenarien typische Muster findet und diese Muster den möglichen Fahrscenarien zuführt, bietet sich der Einsatz von Algorithmen z.B. aus dem Bereich des maschinellen Lernens an. Die resultierenden Gruppen aus ähnlichen Mustern werden dabei als potenzielle Fahrscenarien für den Validierungsprozess einer automatisierten Fahrfunktion interpretiert. Dazu wird der zuvor beschriebene Ansatz nach Montanari, German und Djanatliev als geeignete Grundlage und somit auch als Teilziel dieser Arbeit für die zu implementierende Methodik herangezogen und um weitere Aspekte erweitert. Der Lösungsansatz von Montanari, German und Djanatliev weist ein großes Potenzial für Erweiterungen auf. So kann beispielsweise die Messdatenmenge einer 30-minütigen Datenaufzeichnung auf einer Autobahn auf eine längere Fahrzeit von zwei Stunden erhöht werden, da für diese Arbeit große Mengen an Buskommunikationsdaten von der Firma AUDI AG zur Verfügung gestellt werden. Ein weiterer ausbaufähiger Punkt ist die begrenzte Verwendung von sogenannten Features, die als Eingangsdaten für die Anwendung eines Clustering-Verfahrens notwendig sind. Es werden also mehr Features für das zu implementierende Verfahren verwendet. Auch werden mehr Signale aus den Buskommunikationsdaten und gleichzeitig untersucht mit dem Ziel, mehr mögliche Szenarien in den Messdaten zu finden. Die Menge der zu analysierenden Signale kann zu einer Unübersichtlichkeit führen, weshalb diese in Kategorien eingeteilt werden, denn laut dem Autor Ebner bestehen die Signale aus drei Komponenten Ego-Fahrzeug, Verkehrsteilnehmer und Umwelt. Damit wird ein weiterer positiver Effekt erzielt: Der Umfang von Test- und Freigabeprozessen für automatisierte Fahrfunktionen kann reduziert werden, wenn dafür nur Signale der entsprechenden Kategorie verwendet werden können. [vgl. Ebner 2014, S. 17].

## 1.3 Struktur der Arbeit

Insgesamt besteht die Arbeit mit dem Titel „Clustern von realen Messdaten zur Identifikation von Fahrscenarien“ aus sieben aufeinander aufbauenden Kapiteln (siehe [Abbildung 1.1](#)):



**Abbildung 1.1:** Struktur der Arbeit

Im ersten Teil der Arbeit werden die Motivation für die Thematik, die Zielsetzung, sowie die Struktur der Arbeit erläutert.

Der Grundlagenteil ist in zwei separate Kapitel untergliedert. Zunächst wird in [Kapitel 2](#) das automatisierte Fahren thematisiert. Hierbei werden relevante Definitionen und Grundlagen zum automatisierten Fahren vermittelt. In [Kapitel 3](#) werden die Grundlagen des „maschinellen Lernens“ und seiner Teilbereiche behandelt. Insbesondere wird hierzu auf das hierarchische agglomerative Cluster-Verfahren eingegangen. Die Forschungsprojekte und der aktuelle Stand der Technik werden in [Kapitel 4](#) vorgestellt. Der Schwerpunkt liegt dabei vor allem auf den verschiedenen szenariobasierten Ansätzen zur Identifizierung von Fahrscenarien in Messdaten. Anschließend wird in [Kapitel 5](#) ein Ansatz zur Erkennung von Fahrscenarien in realen Messdaten mithilfe des maschinellen Lernens vorgestellt. Zum einen wird die Idee und der Aufbau der umgesetzten Methodik basierend auf den Referenzansatz nach [Montanari, German und Djanatliev 2020] präsentiert. Zum anderen werden die einzelnen umgesetzten Erweiterungsschritte aufgezählt. In [Kapitel 6](#) werden die Ergebnisse der entwickelten Methode dargestellt und anschließend bewertet. Zudem werden in diesem Abschnitt die Grenzen der entwickelten Methodik beschrieben. Im Anschluss folgt eine Gegenüberstellung der entwickelten Methode und der Referenzmethode nach [Montanari, German und Djanatliev 2020].

Die vorliegende Arbeit wird in [Kapitel 7](#) mit einer Zusammenfassung sowie einem Ausblick auf das Anwendungspotenzial der entwickelten Methodik zur Identifikation von Fahrscenarien mithilfe des Clusters von realen Messdaten abgeschlossen.



## 2 Grundlagen des automatisierten Fahrens

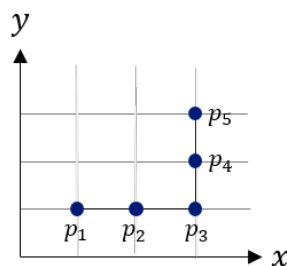
### 2.1 Begriffsdefinitionen

Dieser Abschnitt definiert eine einheitliche Terminologie für die Entwicklung und das Testen von automatisiert fahrenden Fahrzeugen.

**Definition 2.1 (Trajektorie)** „Eine zeitliche Abfolge von Wegpunkten  $T = (p_1, \dots, p_n)$  einer Entität, bei der jeder Punkt  $p = ((x, y, t))$  eine räumliche Position  $(x, y)$  zu einem Zeitpunkt  $\tau$  in einem Koordinatensystem darstellt, wird als Trajektorie bezeichnet“ [Furtado, Pilla und Bogorny 2018, S. 19].

Die physischen oder virtuellen Objekte werden in der International Electrotechnical Commission (IEC)-Norm 15459-3:2014, 3.1 als Entität bezeichnet. Im Zusammenhang des automatisierten Fahrens wird der Begriff Entität für das eigene Fahrzeug oder für andere Fahrzeuge, die am Verkehr teilnehmen, verwendet. Weiterhin wird in dieser Arbeit zwischen dem eigenen Versuchsfahrzeug und weiteren Verkehrsfahrzeugen unterschieden. So werden das eigene Fahrzeug als Ego-Fahrzeug und weitere Verkehrsfahrzeuge als Objekte betitelt [ISO/IEC 2018].

Mithilfe von Global Navigation Satellite System (GPS) Informationen kann die Fortbewegung in Form einer Trajektorie dargestellt werden. Eine Trajektorie einer Entität bildet räumliche Orte auf Zeitpunkte ab (siehe Abbildung 2.1) [vgl. Furtado, Pilla und Bogorny 2018, S. 19].



**Abbildung 2.1:** Ein Beispiel für eine Trajektorie im Koordinatensystem [Orientiert an Furtado, Pilla und Bogorny 2018, S. 19]

**Definition 2.2 (Autonomes Fahren)** Der Verfasser Winner beschreibt das autonome Fahren folgendermaßen: „Unter dem Begriff des autonomen Fahrens wird hier die Fortbewegung mithilfe eines nicht an eine dezidierte Infrastruktur gebundenen Straßenfahrzeugs verstanden (also eines Personen- oder Lastkraftwagens ohne Schienenführungssysteme), das ausschließlich durch die Eingabe oder Adaption einer Mission vom Menschen bedient wird oder sich sogar eigenständig eine Mission zuweist (z. B. Fahrt zu einer Ladestation nach erfolgreicher Transportmission): Die Mission besteht dabei immer aus einer Transportaufgabe von einem Standort A zu einem Standort B mit Transport von Gütern, Personen oder nur dem Fahrzeug selbst“ [Winner 2015, S. 1142].

Die Definition einheitlicher Schnittstellen ist für die Entwicklung und das Testen von Funktionsmodulen, wie z. B. Kontextmodellierung und Verhaltensplanung, eines automatisierten Fahrzeugs notwendig. Nach Ulbrich et al. fungiert eine Szene als zentrale Schnittstelle zwischen Wahrnehmung und Verhaltensplanung [vgl. Ulbrich u. a. 2015, S. 982]:

**Definition 2.3 (Szene)** *„Eine Szene beschreibt eine Momentaufnahme des Umfelds, welche die Szenerie, dynamische Elemente, die Selbstrepräsentation aller Akteure und Beobachter wie auch die Verknüpfung dieser Entitäten umfasst. Einzig eine Szenenrepräsentation in einer simulierten Welt kann allumfassend sein (objektive Szene, Ground Truth). In der realen Welt ist sie immer unvollständig, fehlerbehaftet, unsicherheitsbehaftet und aus der Perspektive eines oder mehrerer Beobachter (subjektive Szene)“* [Ulbrich u. a. 2015, S. 983].

Die drei wesentlichen Komponenten einer Szene sind:

1. Szenerie,
2. dynamischen Elemente und
3. Selbstrepräsentation aller Akteure und Beobachter.

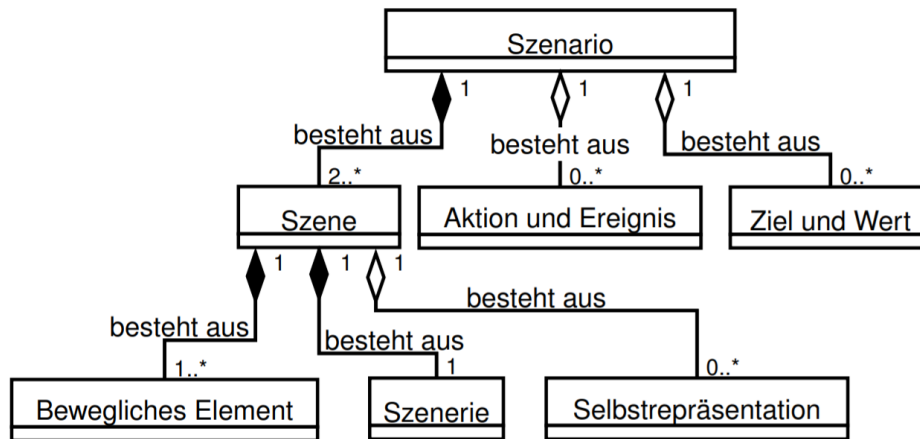
Die Szenerie enthält alle räumlich stationären Aspekte, wie z. B. das Fahrbahnnetz und die Umgebungsbedingungen. Die dynamischen Elemente bewegen sich mithilfe von Energie. Ein entsprechendes Beispiel hierfür ist das Anhalten an Ampeln. Die Eigenschaften und Fähigkeiten, beispielsweise Sichtweite, der Akteure und Beobachter werden in der Komponente Selbstdarstellung aller Akteure und Beobachter beschrieben [vgl. Ulbrich u. a. 2015, S. 983].

**Definition 2.4 (Szenario)** Laut Ulbrich et al. wird ein Szenario wie folgt definiert: *„Ein Szenario beschreibt die zeitliche Entwicklung von Szenenelementen innerhalb einer Folge von Szenen, welche mit einer Startszene beginnt. Aktionen und Ereignisse ebenso wie Ziele und Werte können spezifiziert werden, um diese zeitliche Entwicklung in einem Szenario festzulegen. Im Gegensatz zu Szenen decken Szenarien eine gewisse Zeitspanne ab“* [Ulbrich u. a. 2015, S. 986].

Die Bestandteile eines Szenarios sind [Steimle u. a. 2018, S. 5]:

- mindestens eine Szene,
- Aktionen,
- Ereignisse,
- Ziele und
- Werte.

Um den zeitlichen Verlauf eines Szenarios zu beschreiben, besteht jedes Szenario aus mindestens einer Anfangsszene bzw. abhängig vom Anwendungsfall auch aus einem Satz von Szenen, optionalen Aktionen und Ereignissen, sowie möglichen Zielen und Werten. Beispiele für die Ziele und Werte sind in diesem Zusammenhang die Domäne oder die lokalen regulatorischen Aspekte [vgl. Ulbrich u. a. 2015, S. 986]. Die Beziehung der beiden Begriffe Szene und Szenario einschließlich ihrer Komponenten kann in einem UML-Diagramm wie folgt dargestellt werden:



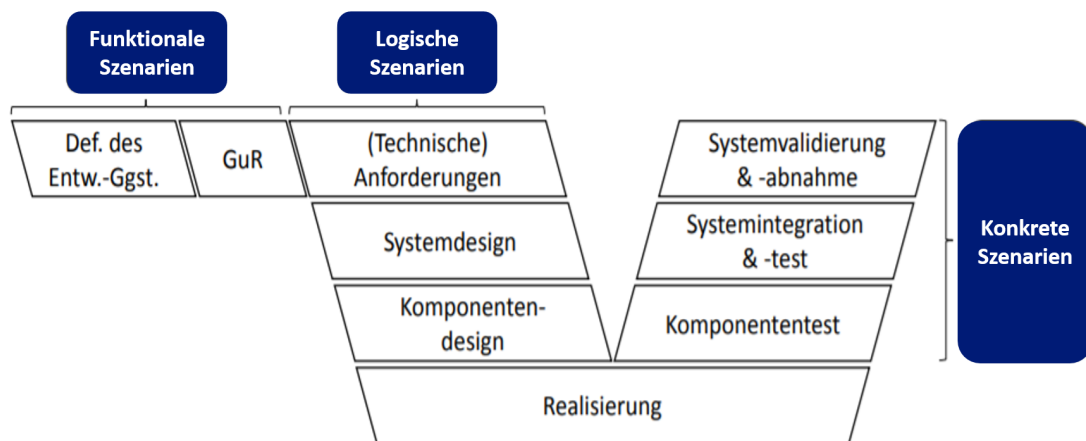
**Abbildung 2.2:** UML-Diagramm zu den Begriffen eines Szenarios [Steimle u. a. 2018, S. 5]

**Definition 2.5 (Situation)** Eine Situation ist im Sinne von Ulbrich et al. auf folgende Weise zu verstehen: „Eine Situation beschreibt die Gesamtheit der Umstände, die für die Auswahl geeigneter Verhaltensmuster zu einem bestimmten Zeitpunkt zu berücksichtigen sind. Sie umfasst alle relevanten Bedingungen, Möglichkeiten und Determinanten von Handlungen. Eine Situation wird aus der Szene durch einen Prozess der Informationsauswahl und -augmentierung abgeleitet, basierend auf transienten (z. B. missionspezifischen) wie auch permanenten Zielen und Werten. Folglich ist eine Situation immer subjektiv, indem die Sicht eines Elements repräsentiert“ [Ulbrich u. a. 2015, S. 985].

Ausgehend von einer Szene und den Zielen und Werten eines Systems kann eine Situation vollständig abgeleitet werden, sodass ein Überschneidungsbereich zwischen einer Szene und einer Situation hinsichtlich der darin enthaltenen Informationen entsteht. Dieser Überlappungsbereich enthält alle relevanten Aspekte der Szene, wie z. B. die dynamischen Elemente, die Szenerie oder die Selbstdarstellung. Darüber hinaus wird eine Situation, die nur funktional relevante Informationen enthält, um funktional relevante Ziele und Werte sowie um relevante funktionspezifische Situationsaspekte erweitert. Ein Beispiel für die Ziele und Werte sind Fahrhinweise zur Bewältigung einer Aufgabe eines automatisierten Fahrzeugs. Ein Situationsaspekt ist unter anderem das Ergebnis einer Situationsbewertung [vgl. Ulbrich u. a. 2015, S. 985 f.].

## 2.2 Szenarien zur Entwicklung, Absicherung und Erprobung

Für die sicherheitskritischen elektrischen oder elektronischen Systeme, zu denen auch Fahrerassistenz- und Fahrzeugführungssysteme zählen, stellt die Norm International Organization for Standardization (ISO) 26262 einen Entwicklungsprozess dar. Der zuvor definierte Begriff Szenario hat in den verschiedenen Phasen des Entwicklungsprozesses unterschiedliche Inhalte und Darstellungsweisen, sodass eine Aufteilung des Begriffs Szenario in die drei Abstraktionsebenen funktional, logisch und konkret vorgenommen wird (siehe Abbildung 2.3) [vgl. Menzel, Bagschik und Maurer 2018, S. 1824].



**Abbildung 2.3:** Verschiedene Abstraktionsebenen von Szenarien im V-Modell-basierten Entwicklungsprozess [Bengler 2017, S. 134]

### Funktionale Szenarien

Die funktionalen Szenarien definieren den Entwicklungsgegenstand in der Konzeptphase im V-Modell und dienen der Erstellung der Gefährdungsanalyse und Risikobeurteilung [vgl. Bengler 2017, S. 133]. Die abstrakte und allgemeine Beschreibung der am Szenario beteiligten Entitäten und ihrer Beziehungen liegt auf der sprachlichen Ebene. Dabei wird das Szenario unter der Verwendung eines Vokabulars oder einer domänenspezifischen Sprache beschrieben [vgl. Menzel, Bagschik und Maurer 2018, S. 1824 - 1825].

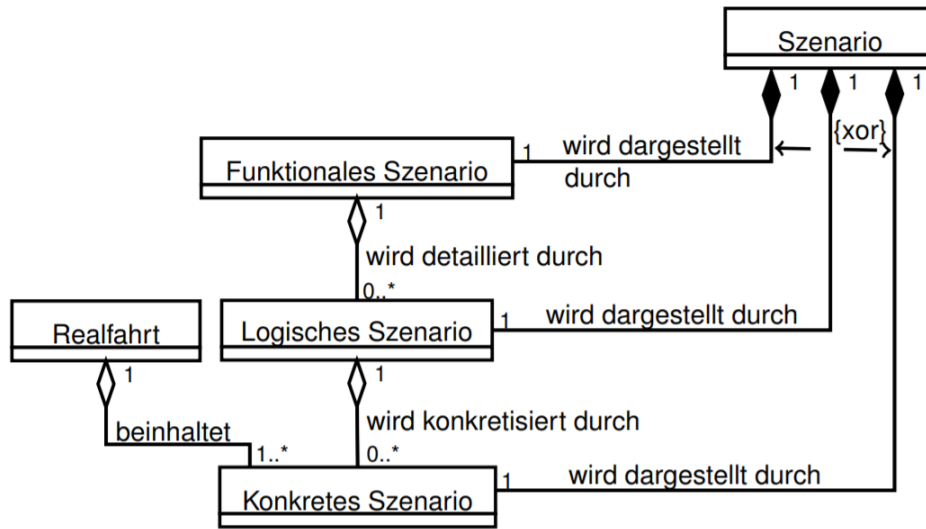
### Logische Szenarien

Die logischen Szenarien prüfen die technischen Voraussetzungen für die Entwicklung einer Funktion. Funktionale Szenarien, die detailliert und im physikalischen Zustandsraum beschrieben sind, werden logische Szenarien genannt [vgl. Bengler 2017, S. 133].

### Konkrete Szenarien

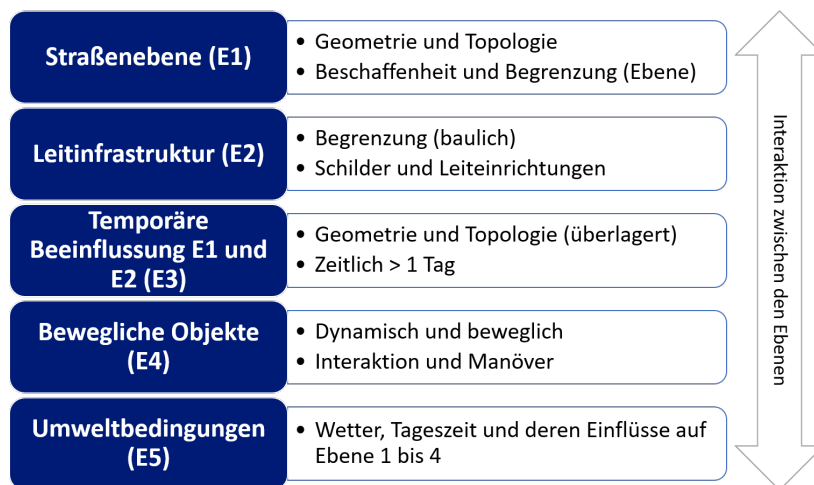
Die Testfälle in der Testphase im V-Modell basieren auf den konkreten Szenarien, die zuvor um die Testinfrastruktur und das Sollverhalten des Testobjekts erweitert werden. Jedes logische Szenario kann durch Konkretisierung der Parameterbereiche für die Entitäten in ein konkretes Szenario transferiert werden [vgl. Bengler 2017, S. 133 f.].

Die Relationen zwischen den oben genannten Typen von Szenarien können in einem UML-Diagramm folgendermaßen dargestellt werden:



**Abbildung 2.4:** UML-Diagramm zu den verschiedenen Abstraktionsebenen von Szenarien [Steimle u. a. 2018, S. 6]

Darüber hinaus können die Inhalte von Szenarien über fünf Informationsebenen von Szenarien definiert werden (siehe Abbildung 2.5). In der **ersten Ebene** werden die Straßengeometrie und -topologie sowie der Zustand der Straße, z. B. die Abmessungen und Anzahl der Fahrspuren, beschrieben. Die Infrastruktur wird auf der **zweiten Ebene** dargestellt. Dazu gehören bauliche Begrenzungen, Verkehrsschilder und -ampeln. **Ebene 3** beschreibt den vorübergehenden Einfluss der ersten beiden Ebenen, z. B. Straßenarbeiten oder Verkehrsunfälle. Die dynamischen Objekte samt deren Interaktionen und Manövern befinden sich auf der **Ebene 4**. In der **fünften Ebene** werden Umweltbedingungen festgehalten. Dazu gehören beispielsweise das Wetter oder die Tageszeit und deren Auswirkungen auf die Ebenen 1 bis 4 [vgl. Bagschik, Menzel und Maurer 2018, S. 1816 - 1817].



**Abbildung 2.5:** 5-Ebenen-Modell zur Beschreibung von Szenarien [Bagschik, Menzel und Maurer 2018, S. 1817]

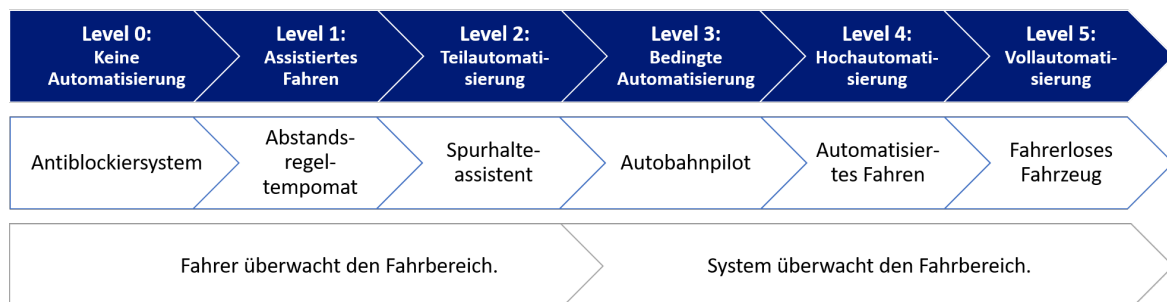
Des Weiteren bestehen, laut dem Autor Ebner, die Szenarien aus drei wesentlichen Komponenten:

- Ego-Fahrzeug,
- Verkehrsteilnehmer,
- Umwelt.

Es ist nicht notwendig, alle Informationen von den Szenarien für die Entwicklung des FAS zu berücksichtigen, es genügt, eine passende Selektion vorzunehmen [vgl. Ebner 2014, S. 17].

## 2.3 Definition verschiedener Automatisierungsgrade

In diesem Abschnitt werden die verschiedenen Stufen des automatisierten Fahrens vorgestellt. In der Automobilindustrie hat die Organisation Society of Automotive Engineers (SAE) die Level für automatisierte Fahrzeuge definiert (siehe Abbildung 2.6).



**Abbildung 2.6:** Definition der Automatisierungsebenen nach SAE J3016 [Orientiert an SAE International 2021, S. 25 - 26]

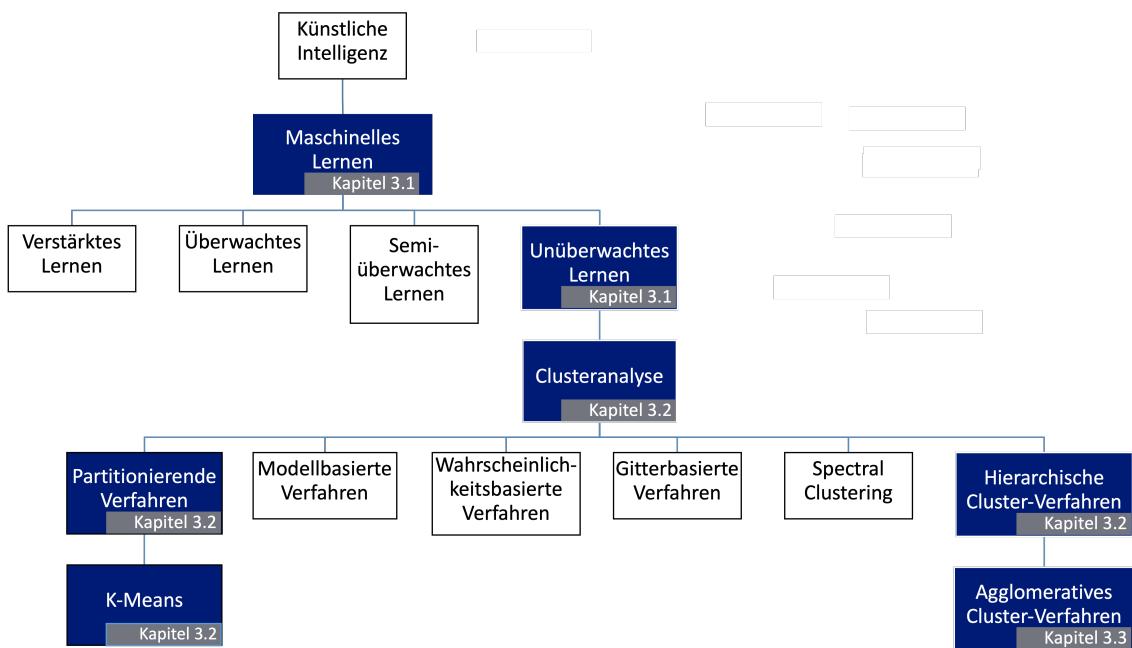
Bei **Level 0** erfolgt keine Automatisierung des Fahrens. Der Fahrer wird von Systemen, wie z. B. vom Antiblockiersystem, unterstützt. Darüber hinaus übernimmt der Fahrer die gesamte Kontrolle über das Fahrzeug. Assistiertes Fahren gehört zu **Level 1** und setzt voraus, dass der Fahrer den Verkehr vollständig kontrolliert und jederzeit eingreifen kann. Ein Beispiel hierfür ist der Abstandsregeltempomat, der den Fahrer bei den Einzelaufgaben der Längs- oder Querverführung entlastet. Solche Systeme sind oft von klimatischen Bedingungen oder der Geschwindigkeit des Fahrzeugs abhängig. **Level 2**, teilautomatisiertes Fahren, umfasst Systeme, wie den Spurhalteassistenten und erfordert ebenfalls die ständige Überwachung vom Fahrer. In diesem Level kann das Fahrzeug jedoch in manchen Situationen die Längs- und Querverführung zur gleichen Zeit übernehmen und so zum Beispiel der Autobahnspur folgen und parallel dazu den Abstand zu anderen Fahrzeugen halten. Das Fahrzeug Audi Q7, ab dem Baujahr 2015, kann dieser Stufe zugeordnet werden. Für **Level 3** ist eine ständige Überwachung durch den Fahrer nicht mehr erforderlich. Allerdings muss der Fahrer nach einer vordefinierten Zeit die Kontrolle über das Fahrzeug übernehmen, sobald das System ihn dazu auffordert. Zugelassene Fahrfunktionen zu dieser Stufe existieren noch nicht, auch wenn Systeme,

wie beispielsweise der Audi A8 - AI Staupilot, technisch nahe an der Serienreife sind. In **Level 4** fährt das Fahrzeug hochautomatisiert. Dabei kann der Fahrer in bestimmten Situationen die Fahraufgaben dem System vollständig übergeben. Zu dem begrenzten Anwendungsbereich zählen zum Beispiel die Autobahn- und Landfahrten, sowie das Einparken. Das vollautomatisierte Fahren wird in **Level 5** beschrieben. Der Anwendungsbereich ist unbegrenzt und der Fahrer wird als Fahrgast gezählt. Dabei übernimmt das System selbstständig alle Fahraufgaben, ohne dass der Fahrer eingreifen muss [vgl. Botsch und Utschick 2020, S. 11].

Weitere Definitionen verschiedener Grade zur Klassifizierung der Automatisierung von Fahrfunktionen schlagen die Bundesanstalt für Straßenwesen [Gasser 2012, S. 9] und die National Highway Traffic Safety Administration [National Highway Traffic Safety Administration (NHTSA) 2021] vor.

### 3 Grundlagen des maschinellen Lernens

Zum Verständnis des methodischen Ansatzes dieser Arbeit, der das hierarchische agglomerative Clustering zur automatisierten Mustererkennung in realen Fahrzeugdaten verwendet, werden zunächst in Abschnitt 3.1 die Grundlagen des maschinellen Lernens sowie dessen Subgebiete näher erläutert. Einer der Teilbereiche ist das unüberwachte Lernen, das ebenfalls in Abschnitt 3.1 vorgestellt wird. Die Clusteranalyse ist eine Methode des unüberwachten Lernens und wird zusammen mit ihren Ansätzen in Abschnitt 3.2 beschrieben. Darüber hinaus wird insbesondere auf den Algorithmus der Partitionierungsmethode K-Means und die hierarchische Clustermethode „agglomerative Cluster-methode“ eingegangen, da die hierarchische agglomerative Clustermethode die Grundlage für die Methode dieser Arbeit darstellt und demzufolge in Abschnitt 3.3 separat behandelt wird. Zum anderen dienen der Algorithmus „K-Means“ und die dafür verwendeten Werkzeuge als Unterstützung für die Methode dieser Arbeit. Im Folgenden werden die in der Abbildung 3.1 farblich markierten Teilbereiche, die im unmittelbaren Zusammenhang mit dieser Arbeit stehen, näher erläutert. Im anderen Fall werden sie zur Darstellung von Alternativen und zur besseren Verständlichkeit des maschinellen Lernens kurz abgehandelt.



**Abbildung 3.1:** Übersicht über die verschiedenen Zweige der künstlichen Intelligenz [Eigene Darstellung]



## 3.1 Maschinelles Lernen zur Mustererkennung

Das maschinelle Lernen (ML), ein Teilgebiet der künstlichen Intelligenz, ermöglicht die Prädiktion zukünftiger Daten, indem es durch Signalverarbeitungsmethoden den statistischen Korrelationen in den Daten nachgeht. Dieser Vorgang wird auch als künstliche Generierung von Wissen aus Erfahrung bezeichnet. Aufgrund der ständig steigenden Rechenleistung von Computern und der anfallenden Daten ist das ML, das auf mathematischer Statistik basiert, bei der Bearbeitung von Aufgaben häufig gefragt. Insbesondere der Bereich der Ingenieurwissenschaften zur Lösung anspruchsvollen Praxisanwendungen profitiert davon. Beim ML werden grundsätzlich vier Haupttypen des Lernens unterschieden:

1. Verstärktes Lernen,
2. Überwachtes Lernen,
3. Semi-überwachtes Lernen,
4. Unüberwachtes Lernen.

Bei der Methode des **verstärkten Lernens** halten sich Algorithmen an Regeln und erhalten dafür Belohnungen, während sie eine Sequenz von Aktionen lernen. Beim **überwachten Lernen** besteht das Ziel darin, Vorhersagen für weitere Eingabedaten zu treffen, indem Algorithmen auf der Grundlage bereits bekannter gelabelter Eingabedaten angelernt werden. Die gelabelten Eingabedaten werden im Voraus bestimmten Ausgabedaten zugewiesen. Das **semi-überwachte** Lernen realisiert die Prädiktion für künftige Eingabedaten, indem ein Anteil an gelabelten und ein anderer Anteil an ungelabelten Eingabedaten verwendet wird. Das **unüberwachte Lernen** bietet Algorithmen, die ohne vorheriges Anlernen selbstständig die Datenstruktur in den vorliegenden Eingabedaten erkennen und daraus ein geeignetes Modell generieren [vgl. Botsch und Utschick 2020, S. 253 f.]. Der hierfür entsprechende Trainingsdatensatz kann folgendermaßen formuliert werden:  $D = \{v_1, \dots, v_m\}$ . Beim unüberwachten Lernen ist die Clusteranalyse, auch bekannt als Clustering, eine der zentralen Aufgaben. Hier geht es darum, Ähnlichkeiten in nicht gelabelten Daten zu entdecken und die Datenpunkte einer bestimmten Anzahl von Gruppen zuzuordnen. Je nach der Menge der zu untersuchenden Daten und dem Anwendungsbereich sind einige Algorithmen dieser Lerntypen des ML besser für den Einsatz geeignet im Vergleich zu anderen [vgl. Botsch und Utschick 2020, S. 342].

Handelt es sich um umfangreiche oder komplexe Daten, kann die Mustererkennung mithilfe von ML die Anzahl der Möglichkeiten stark reduzieren, die dann kombinatorisch weiter untersucht werden können. Bei der Mustererkennung wird eine Verknüpfung von menschlicher Assoziation und Kreativität mit maschineller Kombinationsfähigkeit angestrebt. Dabei ist die menschliche Assoziation und damit die Kreativität stark ausgeprägt, während die maschinelle Assoziation und Kreativität begrenzt ist. Der Mensch greift in für ihn ungewohnten Situationen auf seine Erfahrungswerte zurück und passt sie auf kreative Weise entsprechend an. Die kombinatorische Fähigkeit der Maschine ist jedoch höher als die des Menschen, da der Mensch neue Informationen mit bekannten Informationen vergleicht, während die Maschine alle mathematisch möglichen Kombinationen von Eingabewerten untersucht und das optimale Ergebnis errechnet [vgl. Mockenhaupt und Rüttgers 2021, S. 142 f.]

## 3.2 Clusteranalyse

Wie im vorigen Abschnitt angekündigt, fällt die Clusteranalyse unter das unüberwachte Lernen. Die Clusteranalyse dient dazu, die Daten verständlicher zu strukturieren, verwaltbarer zu gestalten und die Daten für weitere Prozessstufen der Verarbeitung von Signalen zu vereinfachen. Zu diesem Zweck unterteilt die Clusteranalyse die ungelabelten Eingangsdaten auf automatische Weise in Gruppen, die auch als Cluster bezeichnet werden und aus ähnlichen Datenpunkten bestehen. Je nach Art der Eingabedaten kann die Clusteranalyse in eine ähnlichkeitsbasierte und eine merkmalsbasierte Clusteranalyse unterteilt werden. Bei ähnlichkeitsbasierten Methoden ist eine  $M \times M$ -Matrix der Ausgangspunkt, die die Ähnlichkeiten bzw. Unähnlichkeiten zwischen allen Elementen in  $D$  beinhaltet. Bei den merkmalsbasierten Methoden hingegen wird eine  $M \times N$ -Matrix als Ausgangspunkt betrachtet, die die  $M$  Elemente der Dimension  $N$  von  $D$  enthält. Weiterhin lässt sich die Clusteranalyse anhand ihrer Durchführungsart in folgende Cluster-Verfahren einteilen:

1. Partitionierende Verfahren,
2. Modellbasierte Verfahren,
3. Wahrscheinlichkeitsbasierte Verfahren,
4. Gitterbasierte Verfahren,
5. Spectral Clustering,
6. Hierarchische Verfahren.

Bei den ersten vier genannten Verfahren handelt es sich um merkmalsbasierte Methoden. Spectral Clustering und hierarchische Verfahren gelten als ähnlichkeitsbasierte Verfahren [vgl. Botsch und Utschick 2020, S. 342 ff.].

Mit **partitionierenden Clustering-Verfahren** können die Datenpunkte aus  $D$  in eine bestimmte Anzahl von  $k$  Clustern  $C_1, \dots, C_k$  eingeteilt werden, wobei eine Kostenfunktion auf ein Minimum reduziert wird und jeder Punkt  $v_m$  aus  $D$  genau einem Cluster zugeordnet wird. Das Ziel ist es, die  $k$  Cluster und die Zuordnung jedes Datenpunktes aus  $D$  zu einem Cluster zu bestimmen. Das Ziel ist erreicht, wenn die Optimierungsaufgabe gelöst ist, d.h. ein Repräsentant  $c_k$  für jedes Cluster  $C_k$  gefunden ist, sodass alle Datenpunkte aus  $D$  in der Nähe ihres Repräsentanten liegen. Der gebräuchlichste Algorithmus für die Partitionierungscluster-Methode ist der K-Means-Algorithmus. Die Optimierungsaufgabe wird durch den K-Means-Algorithmus mit folgenden Schritten gelöst:

1. Bestimmung der Anzahl  $k$  von Clustern  $C_1, \dots, C_k$ .
2. Initialisierung von  $k$  Clusterzentren nach dem Zufallsprinzip.
3. Zuordnung aller Eingangsvektoren  $v_m$  aus  $D$  zu den Clusterzentren mit dem kleinsten Abstand zu ihnen [Botsch und Utschick 2020, S. 343]:

$$\{z_1, \dots, z_M\} = \underset{\{z'_1, \dots, z'_M\}}{\operatorname{argmin}} \left\{ \sum_{m=1}^M d(v_m, c_{z'_m}) \right\}. \quad (3.1)$$

4. Berechnung neuer Clusterzentren unter Berücksichtigung der konstanten Zuordnung der Datenpunkte  $v_m$  zu den Clustern [Botsch und Utschick 2020, S. 343]:

$$\{c_1, \dots, c_k\} = \underset{\{c'_1, \dots, c'_k\}}{\operatorname{argmin}} \left\{ \sum_{m=1}^M d(v_m, c'_{z'_m}) \right\}. \quad (3.2)$$

Dies ist ein iterativer Prozess, bei dem die Schritte 3 und 4 wiederholt werden. Der Prozess wird gestoppt, und zwar sobald sich die Zentren der Cluster  $c_1, \dots, c_k$  nicht mehr maßgeblich verändern. Der Nachteil von K-Means ist, dass die Anzahl  $k$  der Cluster im Vorhinein bekannt sein muss, d.h. die Wahl der Anzahl  $k$  bestimmt das Ergebnis des Clusters. Um diesem Nachteil entgegenzuwirken, werden z.B. sogenannte Qualitätsmaße verwendet, die eine optimale Anzahl  $k$  von Clustern berechnen. Zu diesem Zweck kann zum Beispiel das Qualitätsmaß Silhouettenkoeffizient verwendet werden. Die Silhouette  $s(v_m)$  wird für einen Datenpunkt  $v_m$  berechnet. Dazu wird der Abstand des Datenpunktes zum jeweiligen Clusterzentrum  $d(v_m, c_{z_m})$  und die Distanz  $d(v_m, c_{z_m})$  von  $v_m$  zu einem weiteren nächstgelegenen Clusterzentrum  $c_{q_m}$  berechnet [vgl. Botsch und Utschick 2020, S. 342 ff.]. Die vollständige Formel hierfür lautet [Botsch und Utschick 2020, S.344]:

$$s(v_m) = \frac{d(v_m, c_{q_m}) - d(v_m, c_{z_m})}{\max \{d(v_m, c_{q_m}), d(v_m, c_{z_m})\}} \in [-1, 1]. \quad (3.3)$$

Für den berechneten Silhouetten-Koeffizienten können die Ergebnisse von -1 bis 1 variieren.  $s(v_m)$  von -1 entspricht dabei einer schwachen Zuordnung, während +1 eine starke Zuordnung des Datenpunktes  $v_m$  zu seinem Clusterzentrum bedeutet.

Die **modellbasierten Cluster-Verfahren** streben eine Anpassung der Datenpunkte aus  $D$  an ein vorgegebenes statistisches Modell an, wobei die Datenpunkte ursprünglich mittels einer Mischverteilung erstellt wurden. Auf der Grundlage der gegebenen Modellannahmen wird die am besten zu dem Modell passende Clusterzerlegung gesucht. Bei wahrscheinlichkeitsbasierten Clustering-Verfahren sind die Datenpunkte innerhalb eines Clusters, definiert als Gebiete im  $N$ -dimensionalen Raum, einander räumlich verbunden. In diesem Fall erfolgt die Abtrennung von Clustern in den Gebieten, in denen die Datenpunkte weniger stark benachbart sind. Für die **gitterbasierten Clustering-Verfahren** werden alle Datenpunkte aus  $D$  einer Gitterzelle eines Gitters im  $N$ -dimensionalen Raum zugewiesen. Daneben wird die Punktdichte jeder Gitterzelle berechnet, um anschließend Gitterzellen zu entfernen, wenn sie unter einen bestimmten Schwellenwert fallen beziehungsweise, um Cluster mit Gitterzellen mit hoher Punktdichte zu bilden. Das **Spectral Clustering** baut auf der Graphentheorie auf, indem es die Datenpunkte aus  $D$  als Knoten eines Graphen betrachtet. Zusätzlich werden die Kanten anhand der Distanzen bzw. Dissimilaritäten zwischen den jeweiligen Datenpunkten gewichtet. Über die Projektion aller Datenpunkte aus ihrer  $M$ -dimensionalen Darstellung in den niederdimensionalen spektralen Bereich mit  $K$  Dimensionen kann dann ein partitionierendes Verfahren eingesetzt werden, um Cluster zu finden, deren Grenzen im ursprünglichen Raum nicht konvex sind. Die **hierarchischen Cluster-Verfahren** werden durch eine Clusterstruktur in Form eines Baumes, eines sogenannten Dendrogramms, dargestellt. Dabei werden die Cluster mit minimalem Abstand zusammengeführt

und enthalten Datenpunkte, die enger zusammenliegen als zu Datenpunkten der restlichen Cluster. Das Dendrogramm bildet die hierarchische Aufteilung der Daten in  $D$  in zunehmend kleinere Teilmengen ab. Dabei steht die Wurzel des Dendrogramms für alle Datenpunkte in  $D$ , während die Blätter für die einzelnen Datenpunkte  $v_m$  stehen. Jede der inneren Knoten des Dendrogramms stellt ein Cluster dar, das aus der Gesamtheit der Datenpunkte des darunter befindlichen Teilbaums besteht. Nachdem das Dendrogramm erstellt ist, kann die aufgabenorientierte Clusteranzahl durch einen Schnitt des Dendrogramms in einer sinnvollen Höhe bestimmt werden. Dabei wird die Position mit einer großen Distanzdifferenz in fusionierten Clustern als geeignet angesehen. Der Vorteil eines Dendrogramms ist, dass die Anzahl der Cluster nicht im Vorfeld festgelegt werden muss. Es hat jedoch den Nachteil, dass bei der Erstellung des Dendrogramms Zuordnungsfehler auftreten können, die nicht mehr im Nachhinein behoben werden können [vgl. Botsch und Utschick 2020, S. 346 ff.].

Es existieren zwei hierarchische Cluster-Verfahren:

1. Divisive Clustering-Verfahren,
2. Agglomerative Clustering-Verfahren.

Der Unterschied zwischen den beiden Verfahren liegt in der Vorgehensweise beim Durchlaufen der Baumstruktur. **Divisives Clustering-Verfahren** beruht auf der Top-Down-Methode, während **agglomeratives Clustering** die Bottom-Up-Methode realisiert [vgl. Botsch und Utschick 2020, S. 351].

### 3.3 Agglomeratives Cluster-Verfahren

Das agglomerative Cluster-Verfahren folgt dem Bottom-up-Prinzip, das bedeutet, dass man ausgehend von den Blättern des Baumes Schritt für Schritt bis zur Wurzel des Baumes vorgeht. Im Wesentlichen handelt es sich um einen iterativen Prozess mit den unten drei aufgeführten Schritten:

1. Die Distanz bzw. die Ähnlichkeit zwischen allen Datenpunkten  $v_m$  aus  $D$ , die ursprünglich als einzelne Cluster betrachtet werden, wird mittels eines Distanzmaßes bzw. Ähnlichkeitsmaßes ermittelt.
2. Die Cluster, die den geringsten Abstand oder die größte Ähnlichkeit aufweisen, werden zu einem neuen Cluster fusioniert.
3. Die Distanz oder Ähnlichkeit des neuen Clusters zu den anderen wird mittels einer Distanzfunktion  $d()$  ermittelt.

Bei jeder Iteration verringert sich die Anzahl der Cluster. Dieser Prozess ist beendet, sobald alle Datenpunkte zu einem zentralen Cluster, d.h. zur Baumwurzel, vereinigt sind. Die Ergebnisse der Clusterbildung werden von der Distanzfunktion bestimmt, die zur Berechnung des Abstands zwischen den Clustern gewählt wird. Zu den gängigsten Distanzfunktionen gehören Single-Linkage und Complete-Linkage, die sich in ihrer Clusterbildung unterscheiden. Mit Single-Linkage werden tendenziell ausgedehnte Cluster mit einer hohen Streuung gebildet, während bei Complete-Linkage der Trend zu kleinen und kompakten Clustern geht [vgl. Botsch und Utschick 2020, S. 351 f.]. Folglich ergeben sich die folgenden Formeln für die Single-Linkage- und Complete-Linkage-Distanz [Botsch und Utschick 2020, S. 352]:

$$d_{\text{Single-Linkage}}(C_i, C_j) = \min_{v_m \in C_i, v_l \in C_j} \{d(v_m, v_l)\}, \quad (3.4)$$

$$d_{\text{Complete-Linkage}}(C_i, C_j) = \max_{v_m \in C_i, v_l \in C_j} \{d(v_m, v_l)\}, \quad (3.5)$$

wobei die Funktion  $d(v_m, v_l)$  den Abstand beziehungsweise die Unähnlichkeit zwischen den Datenpunkten  $v_m$  aus  $C_i$  und  $v_l$  aus  $C_j$  angibt.

Eine weitere bekannte Funktion zur Berechnung des Abstands zwischen Clustern ist die Distanzfunktion „Ward“ [Ward 1963]. Es wird wie folgt berechnet [Amorim 2015, S. 47]:

$$\text{Ward}(C_i, C_j) = \frac{N_{C_i} N_{C_j}}{N_{C_i} + N_{C_j}} d(c_{C_i}, c_{C_j}). \quad (3.6)$$

Die Funktion  $d()$  liefert die Distanz zwischen den Clusterzentren  $c_{C_i}$  und  $c_{C_j}$  der beiden Cluster  $C_i$  und  $C_j$ . Dabei repräsentieren  $N_{C_i}$  und  $N_{C_j}$  die Kardinalitäten, also die Mengen der Elemente der Cluster  $c_{C_i}$  und  $c_{C_j}$  [vgl. Amorim 2015, S. 47]. Das herkömmliche Clustern mit der agglomerativen Clustermethode kann eine Laufzeit von  $O(M^3)$  zur Folge haben, da es  $O(M^2)$  Möglichkeiten für jede Iteration gibt. Bei der Anwendung von Single-Linkage beziehungsweise Complete-Linkage können die Algorithmen SLINK [Sibson 1973] beziehungsweise CLINK [Defays 1977] mit einer geringeren

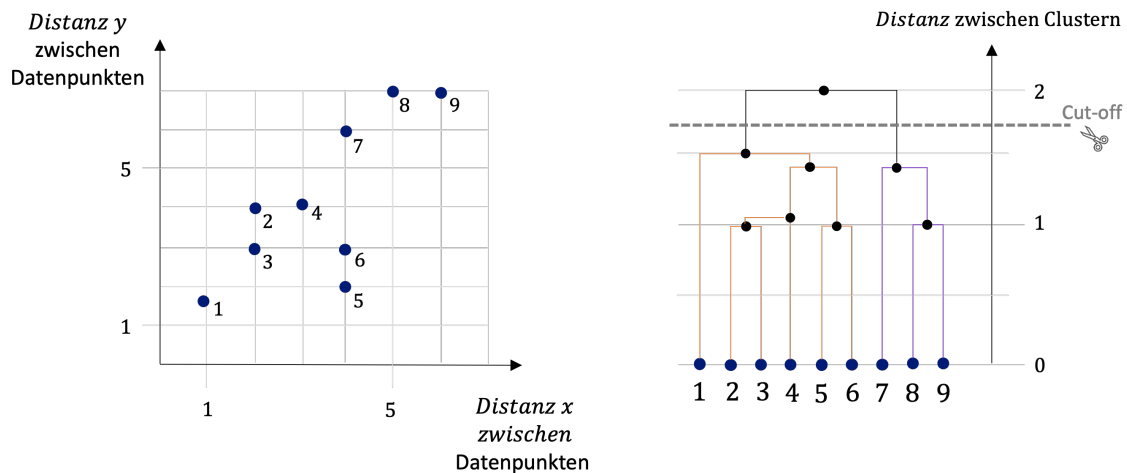
Komplexität von  $O(M^2)$  für die Single-Linkage- und Complete-Linkage-Distanzfunktion eine effiziente Lösung darstellen [vgl. Botsch und Utschick 2020, S. 352].

Zur Berechnung des Abstands zwischen allen numerischen Datenpunkten wird häufig als Distanzmaß die „Euklidische Distanz“ zu Hilfe genommen [Ester und Sander 2000, S. 47]:

$$dist(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}, \tag{3.7}$$

wobei  $x = (x_1, \dots, x_d)$  den Datensätzen entspricht und die Funktion  $dist()$  einen Wert für die Ähnlichkeit der Datenpunkte liefert. Je kleiner der Wert ist, desto ähnlicher sind die Datenpunkte zueinander [vgl. Ester und Sander 2000, S. 46 f.].

Eine beispielhafte Darstellung eines Dendrogramms bei Anwendung der Distanzfunktion „Single-Linkage“ ist auf der rechten Seite der Abbildung 3.2 zu finden. Auf der linken Seite dieser Abbildung sind die entsprechenden Datenpunkte 1 bis 9 mit ihren Abständen zueinander dargestellt. Mithilfe einer manuellen Beobachtung und Untersuchung des Dendrogramms ist es möglich, durch einen horizontalen Schnitt des Dendrogramms eine konkrete Zerlegung der Datenmenge zu definieren. In diesem Beispiel wird ein Schnitt (Cut-off) bei einer Distanz zwischen Clustern von 1.75 gewählt, was zu einer Anzahl von 2 Clustern führt [vgl. Ester und Sander 2000, S. 77 f.].



**Abbildung 3.2:** Darstellung eines Dendrogramms basierend auf vorgegebener Datenpunkte [Orientiert an Ester und Sander 2000, S. 78]

## 4 Forschungsprojekte und Stand der Technik

In Kapitel 4.1 werden Forschungsprojekte, die sich mit der Generierung von Fahrscenarien zum Testen von automatisierten Fahrfunktionen des automatisierten Fahrens und mit Methoden zur Identifikation von Fahrscenarien in realen Messdaten beschäftigen, vorgestellt. Im Rahmen dieser Arbeiten sollen sogenannte szenariobasierte Ansätze verfolgt werden, um die Fahrfunktionen abzusichern und freizugeben. Daraufhin wird ein umfassender Überblick über eine Auswahl von szenariobasierten Ansätzen gegeben (Kapitel 4.2).

### 4.1 Forschungsprojekte zum automatisierten Fahren

In diesem Teil werden Projekte bzw. Forschungsinstitute samt ihrer Projekte vorgestellt, die grundlegende Ansätze und Lösungen sowie geeignete Werkzeuge zur Ermöglichung hochautomatisierter Fahrfunktionen untersuchen und ihre Forschungsergebnisse der Öffentlichkeit präsentieren.

#### **Projekt PEGASUS**

Projekt zur Etablierung von generell akzeptierten Gütekriterien, Werkzeugen und Methoden sowie Szenarien und Situationen zur Freigabe hochautomatisierter Fahrfunktionen (PEGASUS) ist ein mit öffentlichen Mitteln gefördertes Projekt zur Erarbeitung von Methoden zur Sicherstellung der Sicherheit beim automatisierten Fahren am Beispiel der „Autobahn-Chauffeur-Funktion“, die das Fahren auf der Autobahn hochautomatisiert übernimmt und somit nach SAE dem Level 3 zugeordnet werden kann. Dabei wurden folgende Hauptziele gesetzt:

1. Entwicklung eines Verfahrens zur Festlegung von Kriterien für die Auslegung,
2. Entwurf des Prozesses zur Freigabe von hochautomatisierten Systemen für Fahrzeuge,
3. Entwurf, Aufbau und Vorführung von Komponenten eines effizienten Instrumentariums für Simulation, Prüfgelände und Feldversuch,
4. Einbeziehung der Erfahrungen in die Branche, sowie Verbreitung der Resultate und Erfahrungen zur Standardisierung.

Mit der im Projektzeitraum vom 01.01.2016 bis 30.06.2019 entwickelten Methode konnte ein Beitrag zur Gesamtsicherheitsüberprüfung von hochautomatisierten Fahrfunktionen geleistet und eine Definition für den neuen Status zur Freigabe von hochautomatisierten Fahrfunktionen gesetzt werden. Szenarien bildeten den Kern des PEGASUS-Ansatzes. Das Ergebnis war ein szenariobasierter

Ansatz zur Verifizierung und Validierung von hochautomatisierten Fahrfunktionen, der sich deutlich von dem distanzbasierten Ansatz unterscheidet [vgl. Luft- und Raumfahrt e. V. (DLR) 2020, S. 1 - 5].

### **CARISSMA und AININ**

Das Forschungs- und Testzentrum Center of Automotive Research on Integrated Safety Systems and Measurement Area (CARISSMA) der Technische Hochschule Ingolstadt (THI) trägt zur Erhöhung der Verkehrssicherheit im In- und Ausland bei. Im Fokus steht die Vision von der völligen Vermeidung tödlicher Verkehrsunfälle. Es vereint drei Institute mit ihren jeweiligen Projektschwerpunkten [vgl. Technische Hochschule Ingolstadt 2021]:

1. CARISSMA Institute of Safety in Future Mobility (C-ISAFE): u. a. Absicherung bei ungünstigen klimatischen Bedingungen,
2. CARISSMA Institute of Automated Driving (C-IAD): u. a. Entwicklung, Erprobung und Validierung von Fahrfunktionen für das automatisierte Fahren,
3. CARISSMA Institute of Electric, Connected and Secure Mobility (C-ECOS): u. a. IT-Sicherheit in der Automobilindustrie.

Die Institute sind eng miteinander vernetzt und arbeiten intensiv zusammen, insbesondere zum Thema Testen, Absicherung und sicheres automatisiertes Fahren. Ein weiteres Forschungszentrum Artificial Intelligence Network Ingolstadt gGmbH (AININ), das ebenfalls an der THI angesiedelt ist, kooperiert vor allem mit dem C-IAD Institut [vgl. Technische Hochschule Ingolstadt 2021]. Projekte, wie zum Beispiel die Fahrzeugsicherheit, die durch künstliche Intelligenz unterstützt wird, stellen einen der zahlreichen Forschungsschwerpunkte des gesamten Portfolios dar [vgl. Artificial Intelligence Network Ingolstadt gGmbH 2021].

### **Projekt SAVE und SAVeNoW**

Gesamtziel des von Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) geförderten Projekts Funktions- und Verkehrs-Sicherheit im Automatisierten und Vernetzten Fahren (SAVE) war es, durch die Verknüpfung von virtueller und physischer Validierung von Verkehrsszenarien eine Grundlage für zukünftige Automobil- und auch Verkehrsinfrastrukturentwicklungen zu schaffen. Zu diesem Zweck wurden Streckenabschnitte in Ingolstadt digitalisiert, um virtuelle Entwicklungs- und Umsetzungsmethoden für künftige automatisierte Fahrfunktionen zu etablieren. Dabei wurde ein digitaler Zwilling der Stadt Ingolstadt erstellt. Zudem wurden die entstehenden Szenarien auf ihre Eignung geprüft. Das Projekt startete im Mai 2018 und wurde am 01.03.2021 durch das aktuelle Folgeprojekt SAVeNoW abgelöst. Hier erfolgt derzeit die Skalierung und Operationalisierung der von SAVE entwickelten referenzierten Architektur zur virtuellen Entwicklung, Verifikation und Validierung von hochautomatisierter und zugleich vernetzter Mobilität. Dies wird zur Lösung komplexer Probleme der urbanen Mobilität beitragen. Diese Arbeit wurde im Rahmen des SAVeNoW-Projekts verfasst. Ein untergeordnetes Projekt von SAVeNoW befasst sich mit Methoden zur szenariobasierten Entwicklung und Erprobung des automatisierten und vernetzten Fahrens, die im nächsten Unterkapitel behandelt werden [vgl. BMVI 2021].



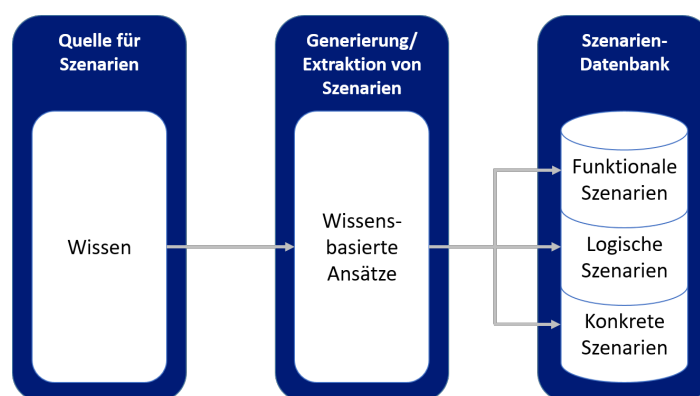
## 4.2 Szenariobasierte Ansätze

Die szenariobasierte Ansätze, dessen Teilgebiete, sprich wissenschaftlicher Ansatz und datenbasierter Ansatz, werden in den Abschnitten 4.2.1 und 4.2.2 beschrieben. Ein großer Vorteil solcher szenariobasierten Ansätzen ist, dass die Szenen irrelevanter Handlungen ausgelassen werden, wodurch sich der Umfang des Tests verringert [vgl. Riedmaier u. a. 2020, S. 3]. Diese Ansätze haben jedoch ebenfalls ihre Grenzen, z. B. das Risiko, dass eine unrealistische Auswahl der zu testenden Szenarien getroffen wird oder dass aufgrund lückenhafter Messdaten aus Testfahrten nicht die tatsächlichen Szenarien erstellt werden [vgl. Eckstein 2020, S. 69 - 70].

### 4.2.1 Wissensbasierter Ansatz

Der wissenschaftliche Ansatz verwendet Normen, Richtlinien, wie beispielsweise Unfallstatistiken und Richtlinien für die Erstellung von Verkehrsinfrastruktur, sowie auch Expertenwissen als Grundlage für die Generierung funktionaler, logischer oder konkreter Szenarien (siehe Abbildung 4.1).

Bei der wissenschaftlichen Erstellung von Szenarien werden üblicherweise Ontologien zur Ablage und Strukturierung des zu verwendenden Expertenwissens eingesetzt [vgl. Riedmaier u. a. 2020, S. 5 - 6]. Ein solcher Ansatz wird von den Autoren Bagschick et al. mit Unterstützung von Mitgliedern des PEGASUS-Projekts vorgeschlagen, wobei verschiedene Wortkombinationen zur Erstellung der Szenarien verwendet werden. Insbesondere beziehen sie sich auf die fünf Schichten des 5-Ebenen-Modells und generieren Szenen der deutschen Autobahn. Ein Vorteil dieser Vorgehensweise ist, dass durch die Verwendung der Ontologie als Wissensbasis der resultierende Szenarienkatalog nicht überprüft werden muss, sondern ausschließlich die modellierte Wissensbasis. Für den nächsten Schritt empfehlen die Verfasser die Durchführung von Analyseschritten im Szenarienkatalog, z. B. bei der Gefahrenanalyse und Risikobewertung, um eine bessere Benutzbarkeit des Katalogs für die Spezialisten zu ermöglichen [vgl. Bagschick, Menzel und Maurer 2018, S. 1 ff.].



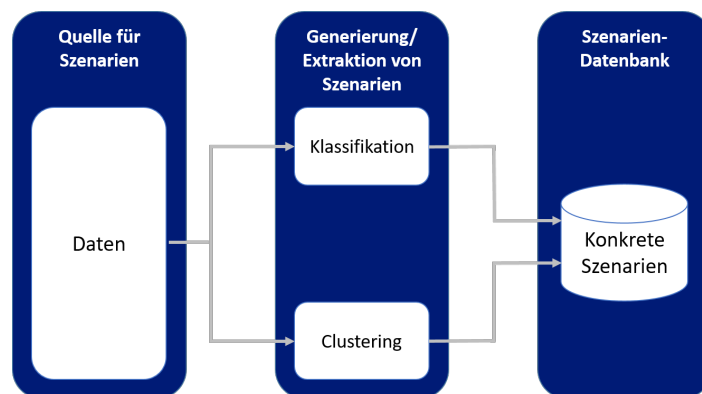
**Abbildung 4.1:** Prozess des wissenschaftlichen Ansatzes zur Erstellung von Szenarien [Riedmaier u. a. 2020, S. 6]

Der derzeitige Stand der Technik im Bereich der Entwicklung und Erprobung von FAS wird durch Ansätze zur manuellen Erstellung von Szenarienkatalogen auf Basis von Expertenwissen repräsentiert [vgl. Montanari, German und Djanatliev 2020, S. 591]. Für komplexere Fahrfunktionen bzw. ab

der zweiten Automatisierungsstufe nach SAE (vgl. Abbildung 2.6) ist die Ermittlung einer großen Anzahl von Szenarien notwendig, um aussagekräftige Testergebnisse zu erhalten. Die Domänenexperten sind zwar gut darin, schwierig zu handhabende Szenarien zu identifizieren, es ist jedoch unwahrscheinlich, dass es ihnen gelingt, auf der Grundlage des ihnen zugänglichen Wissens alle möglichen Szenarien zu finden. Weitere entscheidende Nachteile der manuellen Fertigung von Szenarienkatalogen durch Fachexperten bestehen in dem hohen zeitlichen und finanziellen Aufwand [vgl. Bagschik, Menzel und Maurer 2018, S. 1].

## 4.2.2 Datenbasierter Ansatz

Erfolgsversprechende Lösungen für die Absicherung und das Testen von FAS sind datenbasierte Ansätze. Der datenbasierte Ansatz verwendet Daten, z. B. aus Fahrversuchen, als Quelle für die Erstellung von Szenarien. Hier wird die Vollständigkeit dieser Fahrdaten zur entscheidenden Voraussetzung für die Qualität der erzielten Szenarien. Häufig sind solche Datensätze nicht öffentlich zugänglich, wobei jedoch in letzter Zeit zunehmend mehr Institutionen sie zur Verfügung stellen. Gängige Ansätze für den datenbasierten Ansatz sind Methoden des maschinellen Lernens. Bei den Klassifizierungs- bzw. Clustering-Methoden werden die Daten nach ihrer Ähnlichkeit Szenario-Gruppen zugeordnet, wodurch sich konkrete Szenarien ergeben. Die Clustering-Methode verwendet dabei unüberwachtes Lernen, um die Szenarien Gruppen zuzuordnen, die aus ähnlichen Szenarien zusammengesetzt sind. Dagegen verwendet die Klassifikationsmethode überwachtes Lernen, um Szenarien vordefinierten Szenario-Gruppen zuzuordnen (vgl. Abbildung 4.2) [vgl. Riedmaier u. a. 2020, S. 5 - 7].



**Abbildung 4.2:** Prozess des datenbasierten Ansatzes zur Erstellung von Szenarien [Orientiert an Riedmaier u. a. 2020, S. 7]

### 4.2.2.1 Überwachtes Lernen

Um das überwachte Lernen anwenden zu können, müssen im Vorfeld Szenarien erstellt werden, um die Szenarien klassifizieren zu können. Eine solche Vorstufe beschreiben Weber u. a. in ihrer Arbeit, indem sie logische Szenarien definieren, die auf möglichen Kollisionen zwischen dem Ego-Fahrzeug und einem anderen Fahrzeug basieren. Die Methodik ist auf die Szenarien auf Autobahnen

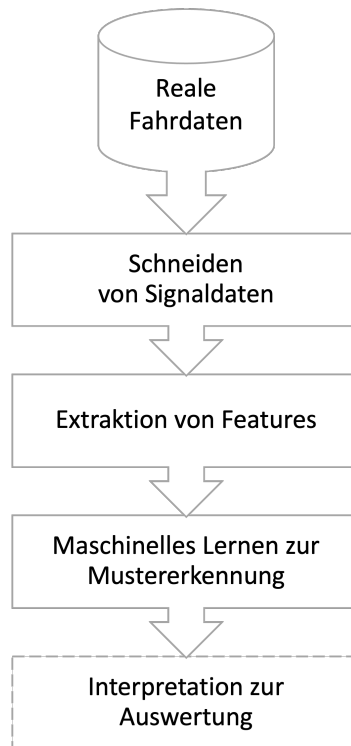
spezialisiert und kann Randbedingungen einbeziehen, wie z. B. Einschränkungen der Handlungsmöglichkeiten des Ego-Fahrzeugs durch weitere Verkehrsteilnehmer. Das Ergebnis ist ein Katalog von sicherheitsrelevanten Szenarien, der im Rahmen des PEGASUS-Projekts als Grundlage für die Klassifizierung von Szenarien entstand [vgl. Weber u. a. 2019, S. 65].

Beglorovic et al. setzen reale Messdaten von Autobahn- und Landstraßenfahrten ein, um Fahrszenarien für ein Spurhalteassistenzsystem zu klassifizieren. Unter Verwendung eines Deep-Learning-Ansatzes als maschinelle Lernmethode werden relevante Fahrszenarien aus verschiedenen Kanälen klassifiziert. Die entwickelte Methodik funktioniert sowohl offline, also nach der Fahrt, als auch online, also während der Fahrt. Im ersten Fall wird eine bessere Genauigkeit bei der Einordnung erreicht [vgl. Beglerovic u. a. 2018, S. 1]

#### 4.2.2.2 Unüberwachtes Lernen

Zur Ermittlung ähnlicher Szenarien aus generierten Fahrdaten und deren automatischer Einteilung in Szenariengruppen verwenden die Autoren Kruber et al. im letzten Teil ihrer Arbeit eine unüberwachte Cluster-Methode. Das Verfahren wird auf die verschiedenen Autobahnszenarien angewandt. Bei diesem Verfahren können weitere neu hinzukommende Messdaten den bestehenden Gruppen zugeordnet werden, wenn sie einen definierten Ähnlichkeitswert überschreiten, andernfalls wird eine neue Gruppierung gebildet [vgl. Kruber u. a. 2019, S. 590].

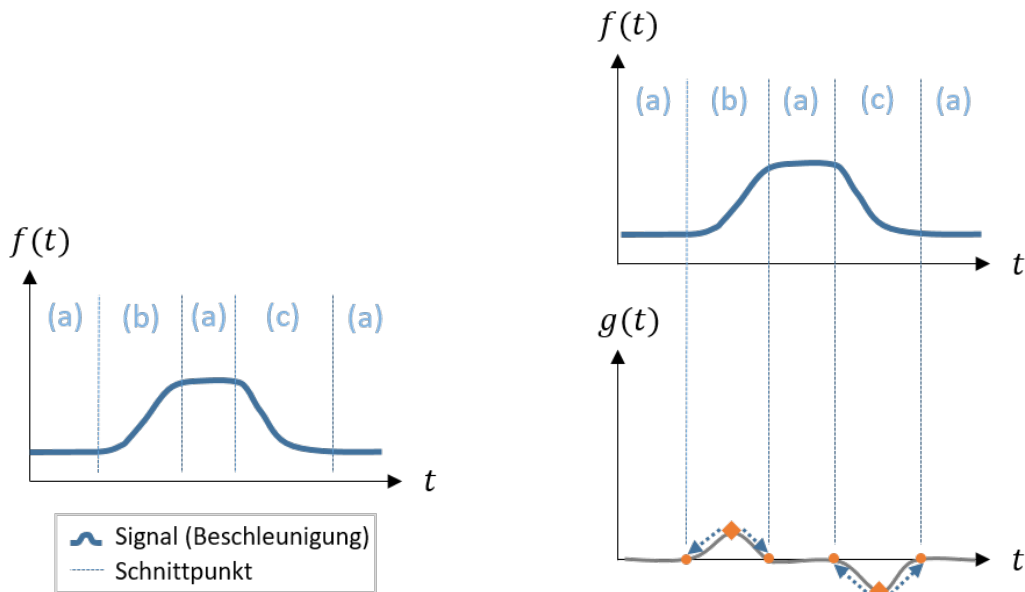
Die im Laufe des SAVe-Projekts entwickelte Methode des unüberwachten Lernens zur Ermittlung von Szenarien aus realen Fahrzeugmessdaten nach [Montanari, German und Djanatliev 2020] wird im Folgendem vorgestellt. Da auf dieser Methode die Methode der vorliegenden Arbeit aufbaut, werden aus diesem Grund zunächst die einzelnen Schritte (vgl. Abbildung 4.3) des letztgenannten Ansatzes im Detail beschrieben [vgl. Montanari, German und Djanatliev 2020, S. 592 ff.].



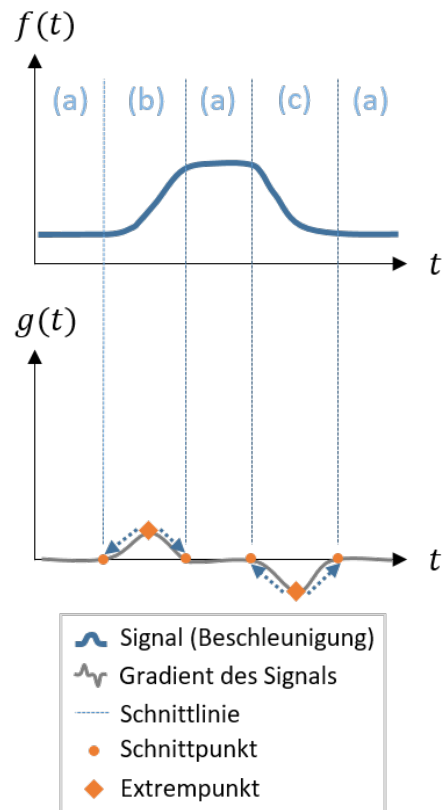
**Abbildung 4.3:** Prozess des Clustering-Ansatzes nach Montanari u. a. [Eigene Darstellung]

**Schritt 1 (Reale Fahrdaten)** Zur Entwicklung des datenbasierten Ansatzes als Basis für die Identifikation von Fahrscenarien in den Fahrzeugmessdaten, ohne Vorgabe oder Ableitung von Regeln, werden reale Buskommunikationsdaten von seriennahen und Serienfahrzeugen des Herstellers Audi verwendet. Der Sensordatensatz von Autobahnfahrten besteht aus zahlreichen Signalen, wie zum Beispiel Fahrzeuggeschwindigkeit oder GPS-Informationen des Fahrzeugs.

**Schritt 2 (Schneiden von Signalen)** Für die Anwendung des unüberwachten Lernens werden die ausgewählten Signale zunächst in kleinere Teile zerlegt. Zu diesem Zweck verwenden die Autoren einen Algorithmus, bei dem Zeitreihen an bestimmten Punkten getrennt werden. Solche Punkte sind Knickpunkte, die ein bestimmtes Ereignis im Signal signalisieren. Als Beispiel wird in Abbildung 4.4 die Steigung des Signals Fahrzeuggeschwindigkeit verwendet. Durch die Ermittlung der Knickpunkte gelingt es die konstanten Teile von den sinkenden beziehungsweise steigenden Teile des Signals zu trennen. Daraus ergeben sich drei verschiedene Muster, und zwar die konstante Geschwindigkeit (a), Beschleunigung (b) und Verlangsamung (c). Um die Steigung eines Signals zu bestimmen, wird zunächst eine Filterung mittels Gradientenfilter durchgeführt. Danach werden die Extrempunkte, sprich die Wendepunkte im Signal, mit einer Prominenz von mindestens 0.3 des Maximums gewählt. In diesem Fall bestimmt die Prominenz die Anzahl der Werte für lokales Maximum. Die Wendepunkte sind zum Beispiel die Teile mit der größten Änderung. Ausgehend von den Extrempunkten entfernt man sich in beide Richtungen, d. h. nach rechts und nach links entlang der Steigung des Signals, bis man auf den Wert der Steigung von null trifft, um so den Schnittpunkt eines Signals zu bestimmen (siehe Abbildung 4.5). Bei aufeinanderfolgenden Schnittbereichen, das bedeutet, wenn der Anfangsschnittpunkt eines Bereichs der Endschnittpunkt seines Vorgängers ist, werden beide Schnittbereiche als ein Schnittbereich gezählt.



**Abbildung 4.4:** Ein Beispiel für die Zerlegung einer Zeitreihe anhand des Beschleunigungssignals [Angelehnt an Montanari, German und Djanatliev 2020, S. 593]



**Abbildung 4.5:** Ein Beispiel für die Zerlegung einer Zeitreihe mit Verarbeitungsschritt [Angelehnt an Montanari, German und Djanatliev 2020, S. 593]

**Schritt 3 (Maschinelles Lernen zur Mustererkennung)** Um eine Menge von Signalabschnitten unterschiedlicher Größe in Gruppen aufzuteilen, damit gleichartige Abschnitte der derselben Gruppe zugeordnet werden, wird das agglomerative hierarchische Clustering praktiziert. Als Eingangsdaten werden sogenannte Features wie Minimum, Maximum und Mittelwert des Signals oder jeweils der Steigung des Signals verwendet.

**Schritt 4 (Interpretation zur Auswertung)** Zur Auswertung wird die Interpretation eines Dendrogramms herangezogen, indem die resultierenden Gruppen zu möglichen Funktionsszenarien gehören und die einzelnen Signalabschnitte dieser Gruppen als eventuelle konkrete Szenarien angesehen werden können.

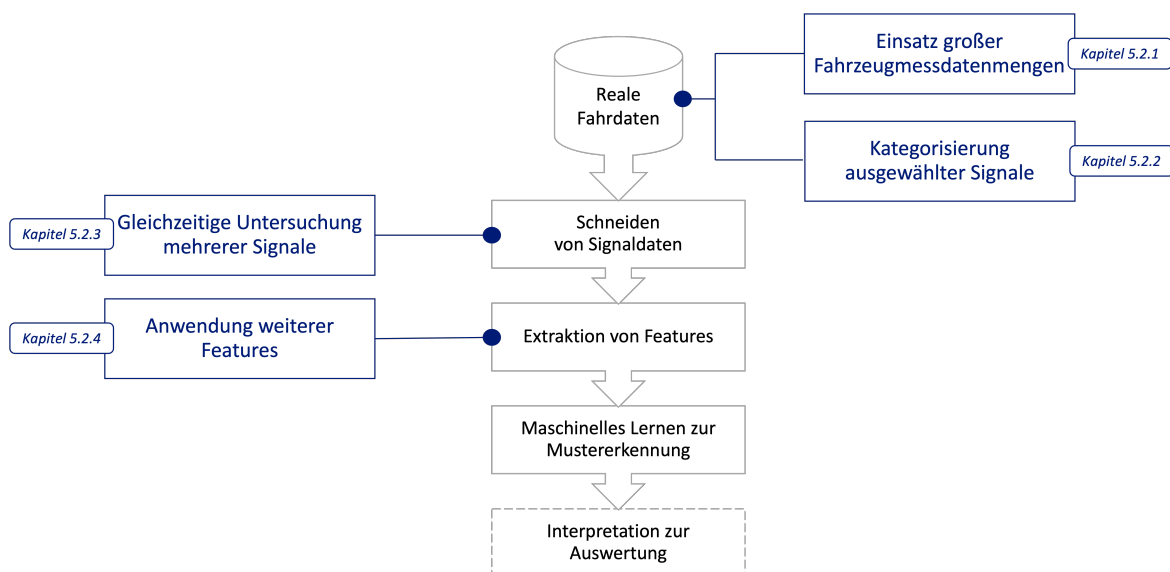
Die hier beschriebene Methode liefert zwar aussichtsreiche und überzeugende Erkenntnisse, weist aber in einigen Aspekten ungenutztes Potenzial auf, wie z. B. die geringe Auswahl der für das maschinelle Lernen verwendeten Merkmale oder die nicht abgestimmten Parameterwerte für die Prominenz und die Größe des Gradientenfilters hinsichtlich der Verarbeitung unterschiedlicher Signale. Weiterhin werden Signale nur einzeln untersucht und somit die Betrachtung der Kombinationsmöglichkeit von Signalen außen vor gelassen [vgl. Montanari, German und Djanatliev 2020, S. 592 ff.].

## 5 Explorativer Ansatz zur Identifikation von Fahrscenarien

In diesem Kapitel wird die Idee für das Konzept und den Aufbau von der Verarbeitung der realen Fahrzeugmessdaten bis zum Auswertungsschritt er zu entwickelnden Methode erläutert. Darüber hinaus werden die einzelnen Ausbauschnitte vorgestellt, die sich aus den ungenutzten Möglichkeiten des zuletzt in Kapitel 4.2.2.2 beschriebenen Ansatzes ergeben haben.

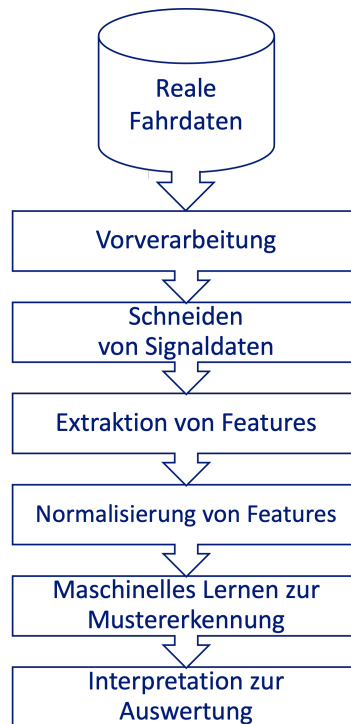
### 5.1 Idee und Aufbau

Die Idee hinter der hier vorgestellten Methode ist es, das Clustern von realen Messdaten für die Erkennung von Fahrscenarien zu nutzen. Um dieses Ziel zu erreichen, wird ein geeignetes Konzept aus der aktuellen Literatur als Ausgangspunkt gewählt. Weiterhin wird das ungenutzte Potenzial dieses Konzepts nach [Montanari, German und Djanatliev 2020] ausgenutzt, um darauf aufbauend Erweiterungen einzuführen und programmatisch umzusetzen. Die einzelnen Erweiterungsschnitte werden in der Abbildung 5.1 dargestellt. Eine Erweiterung davon ist die Anwendung großer Datenmengen der Fahrzeugfahrten (Kapitel 5.2.1). Anschließend werden die Kategorien für ausgewählte Fahrzeugmessdaten eingeführt (Kapitel 5.2.2). Die Untersuchung mehrerer Signale zur gleichen Zeit findet im Prozessschritt „Schneiden von Signaldaten“ statt (Kapitel 5.2.3). Im Anschluss daran folgt die Einführung weiterer angewandter Features, die für den Einsatz des Clustering-Verfahrens notwendig sind (Kapitel 5.2.4).



**Abbildung 5.1:** Erweiterungsschritte in Hinblick auf die Prozesskette des Ansatzes nach Montanari, German und Djanatliev 2020 [Eigene Darstellung]

Die endgültige Architektur der Methode dieser Arbeit ist in der Abbildung 5.2 abgebildet. Im Gegensatz zu dem bereits in Kapitel 4.3 vorgestellten Referenzprozess sind die Prozessschritte „Vorverarbeitung“ und „Normalisierung von Features“ enthalten, da diese in der Publikation des Ansatzes nach Montanari u. a. nicht erwähnt oder verwendet werden.



**Abbildung 5.2:** Aufbau der Methode dieser Arbeit [Eigene Darstellung]

**Prozessschritt 1 (Vorverarbeitung)** Zunächst werden die realen Messdaten aus dem Bussystem der Fahrzeuge aufbereitet, indem sie bereinigt und in ein geeignetes Format für die programmatische Weiterverarbeitung konvertiert werden. Während des Bereinigungsprozesses werden vordefinierte Standardwerte oder unrealistische Werte mit Nullwerten versehen, um fehlende Werte für weitere Schritte entsprechend zu behandeln und zu berücksichtigen. Ein solcher Messdatensatz enthält zahlreiche Signale.

**Prozessschritt 2 (Schneiden von Signaldaten)** Ausgewählte Signale werden im nächsten Schritt geschnitten, d.h. es werden zusammenhängende Segmente der Signale von anderen getrennt und anschließend für weitere Vorgänge gesammelt.

**Prozessschritt 3 (Extraktion von Features)** Da die gesammelten Segmente unterschiedlich lang sind, werden aus den Segmenten sinnvolle Features abgeleitet, die dann für das Clustering verwendet werden.

**Prozessschritt 4 (Normalisierung von Features)** Zusätzlich werden die Features normalisiert, indem ihre Werte auf den Wertebereich von 0 bis 1 skaliert werden. Dieser Vorgang ist notwendig, da viele Signale gleichzeitig untersucht werden und ihre Werte stark voneinander abweichen können,

was zu einer unerwünschten Dominanz von Features eines Signals mit höheren Werten führen kann. Beispielsweise gibt es ein Signal mit Geschwindigkeitswerten von 0 bis etwa 160 und das Signal mit Abstandswerten zu den Spurlinien, die im Bereich von 0 bis etwa 3 liegen. Dementsprechend würden die Geschwindigkeitswerte bei der Clusterbildung eine entscheidendere Rolle einnehmen.

**Prozessschritt 5 (Maschinelles Lernen zur Mustererkennung)** Als Clustering-Methode für eine geeignete Gruppierung der gefundenen Muster in den Daten wird hier das agglomerative hierarchische Clustering gewählt. Zudem wird die „Euklidische Distanz“ als Distanzmaß für die Berechnung der Abstände zwischen den Datenpunkten und die „Ward-Methode“ als Distanzfunktion zur Berechnung der Abstände zwischen Clustern gewählt. Wie in Kapitel 3.3 bereits beschrieben wird bei diesem Cluster-Verfahren ein Dendrogramm erstellt. Um für die Durchführung eines Schnittes in dem Dendrogramm die optimale Clusteranzahl zu ermitteln, wird nicht nur das manuelle Betrachten und das Analysieren der Baumstruktur durch menschliche Fähigkeiten realisiert, sondern auch das aus dem Kapitel 3.2 bekannte Qualitätsmaß „Silhouette“ herangezogen. Hinzu kommen noch weitere Qualitätsmaße: „Calinski-Herabsz“ und „Distortion“. Distortion berechnet die Summe der quadratischen Abstände aller Punkte zu den ihnen zugeordneten Clusterzentren, während Calinski-Herabsz die optimale Anzahl der Cluster durch Berechnung des Verhältnisses der Streuung zwischen und innerhalb der Cluster berechnet. Die drei Qualitätsmaße und ihre Bewertungsgrafiken werden von der Yellowbrick-Bibliothek für die Entwicklung mit der Programmiersprache „Python“ bereitgestellt. [vgl. The scikit-yb developers 2021].

**Prozessschritt 6 (Interpretation zur Auswertung)** Nach dem Schnitt des Dendrogramms nach der vorgegebenen Anzahl von Clustern werden Cluster gebildet. Diese Cluster, die aus ähnlichen Signalmustern bestehen, werden mithilfe von weiteren Grafiken detailliert analysiert und anschließend in Form von Szenarien interpretiert. Dieser Schritt ist notwendig, da es sich um unüberwachtes Lernen handelt und somit keine direkten Szenarien als Ergebnis dieses Prozesses entstehen, sondern nur Cluster von ähnlichen Signalmustern. In diesem Zusammenhang nutzt die Methode dieser Arbeit hauptsächlich die maschinellen Fähigkeiten zur Mustererkennung in den realen Fahrzeugmessdaten, und die identifizierten geclusterten Muster werden dann von den menschlichen Fähigkeiten als mögliche Fahrscenarien interpretiert. Darüber hinaus werden die menschlichen Fähigkeiten u. a. auch bei der Auswahl der realen Messdaten und des Lernalgorithmus sowie der dafür benötigten Parameter eingesetzt.



## 5.2 Einführung von Erweiterungen

Die einzelnen Erweiterungen, die für die Methode dieser Arbeit implementiert werden, werden im Einzelnen besprochen. Folgende Erweiterungen wurden umgesetzt:

1. Einsatz von großen Fahrzeugmessdatenmengen,
2. Kategorisierung von ausgewählten Fahrzeugmessdaten,
3. Gleichzeitige Untersuchung von mehreren Signalen,
4. Anwendung von weiteren Features.

### 5.2.1 Einsatz von großen Fahrzeugmessdatenmengen

Als Eingabedaten für den Prozess, gemäß der Abbildung 5.2, werden reale Fahrdaten verwendet. Hierfür stehen Fahrzeugmessdaten, die aus Fahrten von seriennahen oder Prototyp-Fahrzeugen des Herstellers Audi stammen, zur Verfügung. In der Vorverarbeitungsstufe werden die Messdaten nicht nur im Hinblick auf ungültige oder fehlende Werte bereinigt, sondern auch nach Autobahndaten gefiltert. Aufgrund der unterschiedlichen Infrastruktur und Topologie sowie des Fahrstils auf städtischen, ländlichen oder Autobahnstraßen ist es notwendig, die Autobahndaten von anderen Daten zu trennen. So können insgesamt 20 Autobahnfahrstunden von 45 realen Fahrstunden für die Untersuchung extrahiert werden.

### 5.2.2 Kategorisierung von ausgewählten Signalen

Wie in Kapitel 2.2 erwähnt, enthalten die Szenarien die folgenden Bestandteile: Verkehrsteilnehmer, Ego-Fahrzeug und Umwelt. Um eine bessere Übersicht über die vielen Signale zu erhalten, können diese nach den Bestandteilen der Szenarien eingeteilt werden. So kann eine Kategorisierung der gewählten Signale für die Auswertung in Verkehrsteilnehmer, Ego und Umwelt vorgenommen werden. Ein weiterer Vorteil einer solchen Kategorisierung ist die Minimierung des Aufwands für den Test- und Freigabeprozess von FAS. So können beispielsweise für den Test einer Fahrfunktion für den einfachen Spurhalteassistenten nur die Szenarien der Kategorie Umwelt herangezogen werden, da es die Signalwerte zu den beiden Fahrspurlinien enthält. Die für diese Arbeit verwendeten Signale, einschließlich ihrer Beschreibung sind unter Angabe ihrer Kategorie in Tabelle 5.1 aufgeführt. Die Auswahl der Signale basiert auf Vorüberlegungen hinsichtlich der Relevanz der Identifikation von Szenarien und der Vollständigkeit sowie der Qualität der Signalwerte.

Kategorie	Signalbezeichnung	Beschreibung
Verkehrsteilnehmer	Position x	Relativer Abstand vom Ego-Fahrzeug zum anderen Verkehrsteilnehmer in Längsrichtung
	Position y	Relativer Abstand vom Ego-Fahrzeug zum anderen Verkehrsteilnehmer in Querrichtung
Ego	Geschwindigkeit	Geschwindigkeit des Ego-Fahrzeugs
Umwelt	Linie rechts	Abstand zur rechten Spurlinie
	Linie links	Abstand zur linken Spurlinie

**Tabelle 5.1:** Übersicht über die kategorisierten Fahrzeugsignale [Eigene Darstellung]

### 5.2.3 Gleichzeitige Untersuchung von mehreren Signalen

In dem Ansatz nach [Montanari, German und Djanatliev 2020] werden Signale separat ausgewertet, sodass die Kombinationsmöglichkeit mehrerer Signale nicht verfolgt wird. Um möglichst gezielt viele sinnvolle Muster in den Fahrdaten zu finden, kann die Untersuchung von mehreren Signalen zur gleichen Zeit in Erwägung gezogen werden. Ausgehend von Vorüberlegungen ermöglicht die parallele Analyse der Signale, beispielsweise „Position x“ und „Position y“ die Erkennung von Mustern wie zum Beispiel das Szenario für einen Spurwechsel der Verkehrsteilnehmer oder des Ego-Fahrzeugs. Dementsprechend kann die gleichzeitige Untersuchung aller Signale der jeweiligen Kategorie aus Tabelle 5.1, eine zusätzliche Anzahl von Mustern für bestimmte Szenarien zu erfassen.

### 5.2.4 Anwendung von weiteren Features

Eine weitere Möglichkeit, die Erfolgswahrscheinlichkeit der Erkennung aussagekräftiger Muster in den Fahrdaten zu erhöhen, besteht darin, etliche Features aus den Signalen zu extrahieren und sie im nächsten Prozessschritt für die Clusterbildung zu verwenden. Ein Beispiel hierfür ist die Verwendung des Anfangs- und Endwerts der „relativen Position y“ des Signals, um ein weiteres Muster in den Fahrdaten zu identifizieren, z. B. für das Szenario eines mehrmaligen Fahrspurwechsels oder eines doppelten Fahrspurwechsels des Verkehrsteilnehmers.

Die Auswahl der Features basiert auf der Veröffentlichung von Montanari et al. und wird um die Features mit Start- und Endwerten sowie Absolutbetrag ergänzt. In der Tabelle 5.2 sind die für die Methode dieser Arbeit eingesetzten Features pro Kategorie gekennzeichnet. Die betroffenen Signale, von denen die Features abgeleitet werden, werden in den einzelnen Unterkapiteln des Kapitels 6 zu Beginn erläutert. Aufgrund der parallelen Auswertung von mehreren Features werden diese zusätzlich auf den Wertebereich  $[0, 1]$  normalisiert.

Feature	Beschreibung	Verkehrsteilnehmer	Ego	Umwelt
s_med	Medianwert des Signals			
s_max	Maximalwert des Signals			
s_min	Minimalwert des Signals			
s_start	Anfangswert des Signals	x		
s_end	Endwert des Signals	x		x
g_med	Medianwert des abgeleiteten Signals	x	x	
g_max	Maximalwert des abgeleiteten Signals		x	x
g_min	Minimalwert des abgeleiteten Signals		x	x
g_start	Startwert des abgeleiteten Signals			
g_end	Endwert des abgeleiteten Signals			
g_big	Absolutbetrag des maximalen Wertes von g_max und g_min			

**Tabelle 5.2:** Übersicht über die in dieser Arbeit verwendeten Features pro Kategorie [Eigene Darstellung]

## 6 Ergebnisse und Evaluierung

Der folgende Abschnitt befasst sich mit den Ergebnissen, die bei der Anwendung der Methode dieser Arbeit erzielt werden, und dessen anschließenden Bewertung. Die umgesetzte Methode ist an den Ansatz der Publikation von den Autoren [Montanari, German und Djanatliev 2020] orientiert und durch Einsatz von großen Fahrzeugmessdaten, Kategorisierung von ausgewählten Signalen, gleichzeitige Untersuchung von mehreren Signalen und Anwendung weiterer Features erweitert. In Anbetracht der Einteilung der verwendeten Signale aus den Fahrdaten in die Kategorien Verkehrsteilnehmer, Ego und Umwelt werden im weiteren Verlauf die Ergebnisse zu den genannten Kategorien in der Folge dargestellt, interpretiert und evaluiert. In Kapitel 6.4 werden die Grenzen dieser Methode erläutert. In Kapitel 6.5 erfolgt anschließend die Diskussion, in der insbesondere die Resultate, aber auch die Erreichung der festgelegten Ziele der entwickelten Methodik hinterfragt werden. Ebenso wird in diesem Kapitel die Methodik der Veröffentlichung von den Autoren [Montanari, German und Djanatliev 2020] und dieser Arbeit gegenübergestellt.

### 6.1 Kategorie 1: Verkehrsteilnehmer

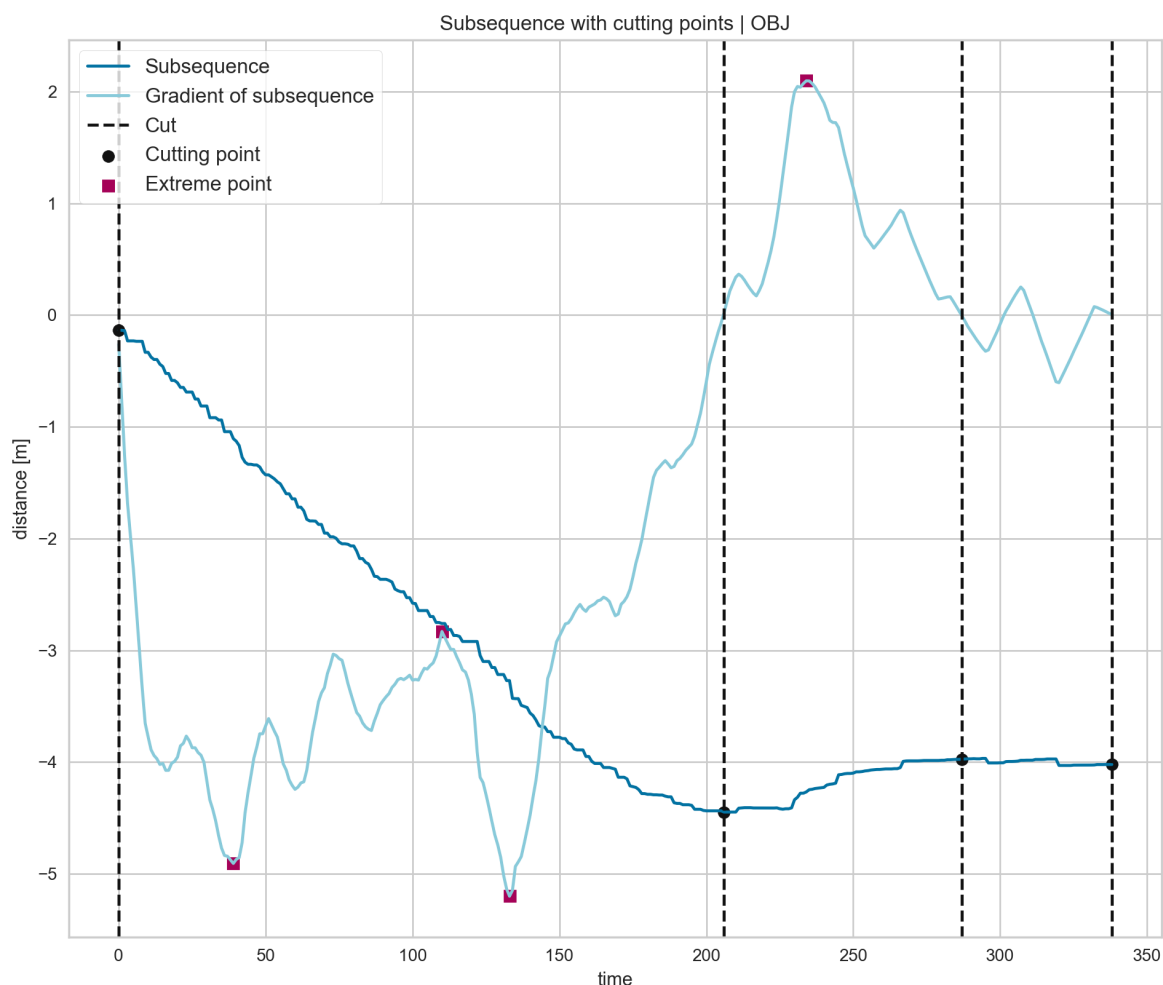
In diesem Abschnitt werden die Ergebnisse für die Kategorie „Verkehrsteilnehmer“ vorgestellt. Als Erstes wird in tabellarischer Form ein kurzer Überblick über die für den Analysezweck verwendeten Daten, nämlich die Signale und die zugehörigen Features, gegeben. Danach folgen die Ergebnisse der Anwendung des Schnittalgorithmus unter der Verwendung unterschiedlicher Prominenzwerte. Als Nächstes werden Grafiken zur Bestimmung der optimalen Cluster Anzahl mittels Qualitätsmaße für den Schnitt des Dendrogramms vorgestellt. Nachfolgend werden weitere Darstellungen zur Erkennung möglicher Szenarien diskutiert. Zusammenfassend werden die in den durchgeführten Auswertungen identifizierten Szenarien dargestellt.

Die für die Verkehrsteilnehmer-Kategorie verwendeten Signale und die daraus abgeleiteten Features, die in Kapitel 5.2 bereits vorgestellt wurden, können der Tabelle 5.1 entnommen werden.

Signal	Feature
Position x	s_start
	s_end
	g_med
Position y	s_start
	s_end
	g_med

**Tabelle 6.1:** Überblick über die verwendeten Signale und Features für die Kategorie Verkehrsteilnehmer [Eigene Darstellung]

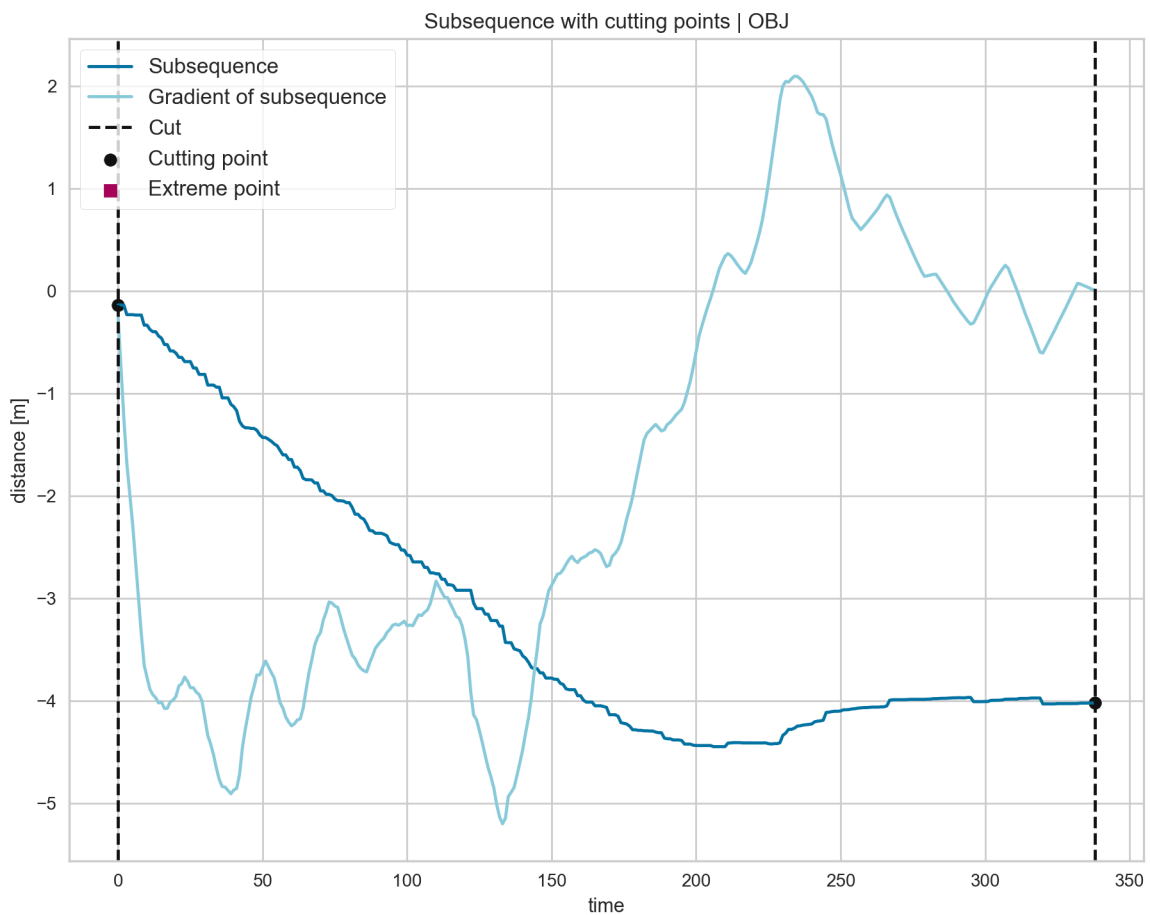
Für das Schneiden von Signalen werden kategoriespezifisch Prominenzwerte gesetzt. In dem Diagramm 6.1 wird ein ausgewähltes Signal „Position y“ (Subsequence) mit dem festgelegten Wert für die Prominenz von 1 dargestellt. Das Diagramm zeigt die Distanz in y-Richtung des Verkehrsteilnehmers zum Ego-Fahrzeug im Zeitverlauf an. Des Weiteren ist der Gradient des Signals sowie die Extrempunkte abgebildet. Auf Basis der Extremwerte und des Prominenzwertes von 1 wird das gesamte Signal in drei Abschnitte geteilt. Hierbei ist anzumerken, dass das Ego-Fahrzeug im Zeitverlauf stets die y-Koordinate von 0 aufweist. Somit ist im ersten Abschnitt ein Muster für die Entfernung des Verkehrsteilnehmers zum Ego-Fahrzeug zu beobachten, da sich das Signal von der Null-Linie über den Zeitverlauf bis zum Distanzwert von ca. -0.45 entfernt. Im nächsten Abschnitt nähert sich der Verkehrsteilnehmer dem Ego-Fahrzeug wieder, indem es der Null-Linie näher kommt. Im dritten Abschnitt ist ein Muster für eine konstante Entfernung des Verkehrsteilnehmers zum Ego-Fahrzeug zu erkennen, da der Distanzwert sich in diesem Abschnitt fast nicht ändert. Somit ist der erste Abschnitt ein Muster, in dem sich der Verkehrsteilnehmer von einem Ego-Fahrzeug entfernt. Der zweite Abschnitt zeigt die wiederum eine Annäherung des Verkehrsteilnehmers zum Ego-Fahrzeug. Im dritten Abschnitt ändert sich die Distanz vom Verkehrsteilnehmer zum Ego-Fahrzeug kaum und kann als konstante Distanz gedeutet werden.



**Diagramm 6.1:** Ein Beispiel für das Schneiden des Signals „Position y“ der Verkehrsteilnehmer-Kategorie mit Prominenzwert 1 [Eigene Darstellung]

Bei einem höheren Prominenzwert von 5 ist es anhand des Diagramms 6.2 ersichtlich, dass dasselbe Signal nicht mehr in Teile geschnitten wird, sondern als ein vollständiges Signal erfasst wird. Über den zeitlichen Verlauf ist anhand des Signals (Subsequence) im Gesamten eine Entfernung des Verkehrsteilnehmers vom Ego-Fahrzeug festzustellen. Durch die Wahl einer höheren Prominenz werden in diesem Beispiel Muster zur Annäherung und Muster für eine konstante Distanzierung des Verkehrsteilnehmers zum Ego-Fahrzeug ausgelassen.

Grundsätzlich kann man davon ausgehen, dass die Wahl der Prominenz die Anzahl der erhaltenen Signalabschnitte beeinflusst und somit auch die Ergebnisse des Cluster-Verfahrens.



**Diagramm 6.2:** Ein Beispiel für das Schneiden des Signals „Position y“ der Verkehrsteilnehmer-Kategorie mit Prominenzwert 5 [Eigene Darstellung]

In dem nächsten Diagramm 6.3 wird eine Ellbogen-Methode verwendet, um bei der Auswahl der optimierten Anzahl von Clustern zu unterstützen. Wenn die Linie einem Arm ähnelt, ist der Wendepunkt dieser Kurve ein gutes Zeichen dafür, dass das Modell an dieser Stelle geeignet ist. Die Distortion-Methode schlägt einen Wert von 6 als optimale Clusteranzahl vor, während die Calinski-Herabsz (siehe Diagramm 6.4) und Silhouette (siehe Diagramm 6.5) eine Anzahl von 3 bestimmen.

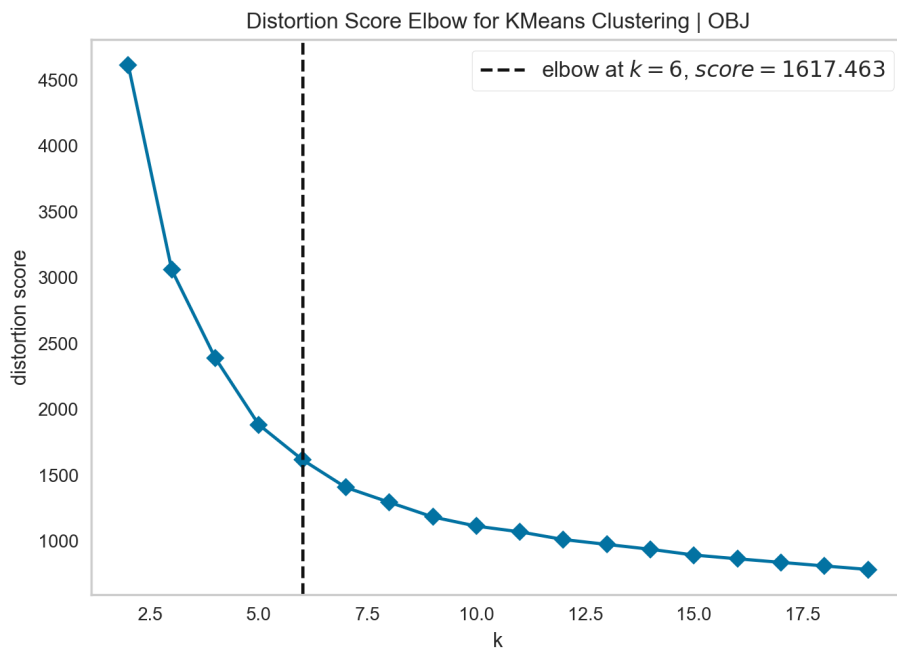


Diagramm 6.3: Berechnung einer optimalen Clusteranzahl mittels Distortion [Eigene Darstellung]

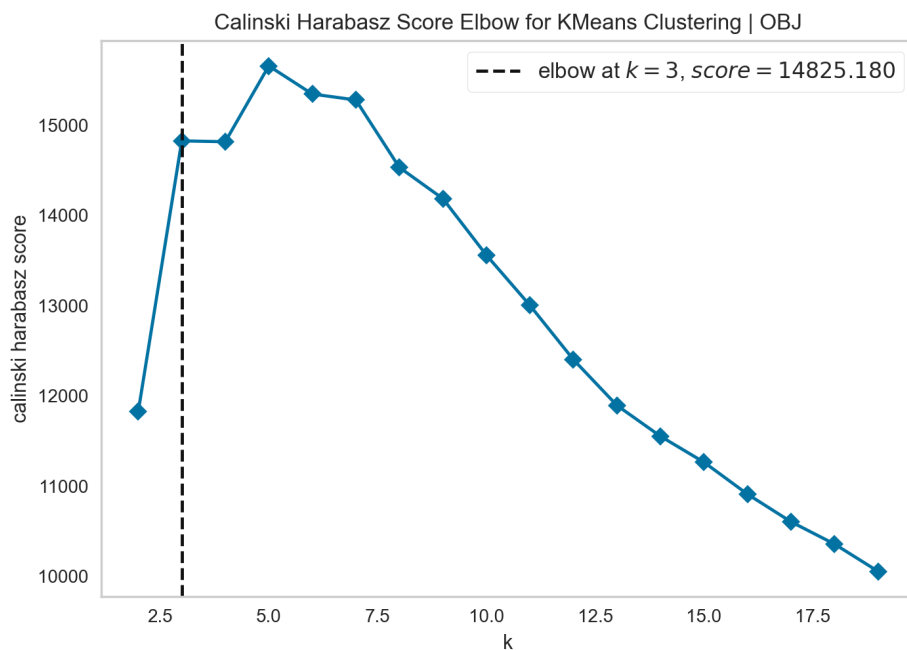
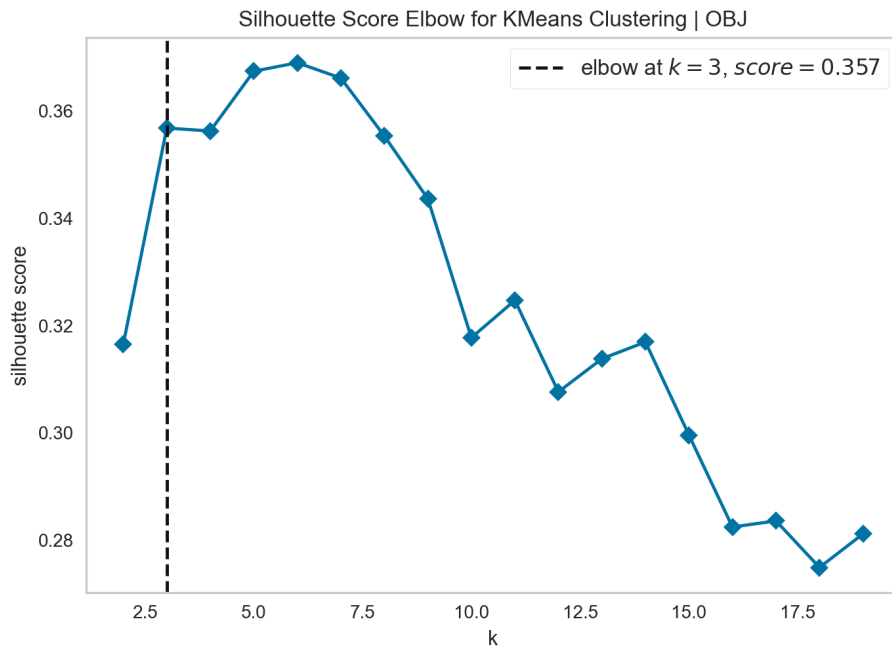


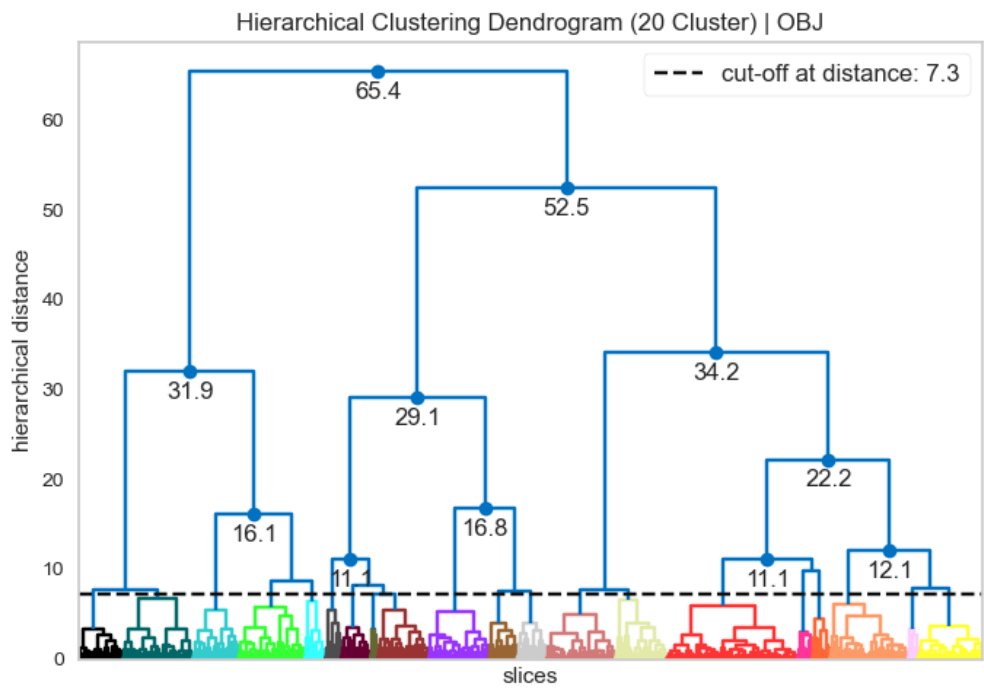
Diagramm 6.4: Berechnung einer optimalen Clusteranzahl mittels Calinski-Herabsz [Eigene Darstellung]



**Diagramm 6.5:** Berechnung einer optimalen Clusteranzahl mittels Silhouette [Eigene Darstellung]

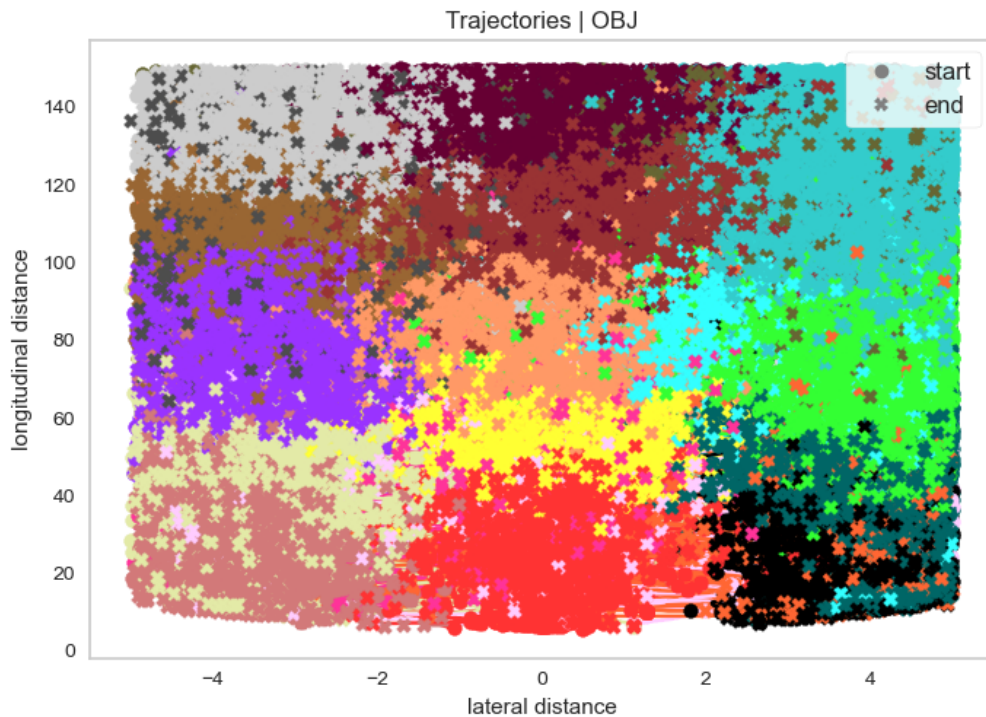


Aufgrund der vorgeschlagenen, im niedrigen Bereich gehaltenen Werte für die optimale Anzahl von Clustern durch die Qualitätsmaße, ist es schwierig, weitere Darstellungen, insbesondere der einzelnen Cluster, zu interpretieren. Diese Darstellungen können dem Anhang dieser Arbeit entnommen werden. In diesem Zusammenhang wird manuell eine höhere Zahl von 20 bestimmt, welches einer Distanz zwischen Clustern von 7.3 im Dendrogramm des Diagramms 6.6 entspricht.



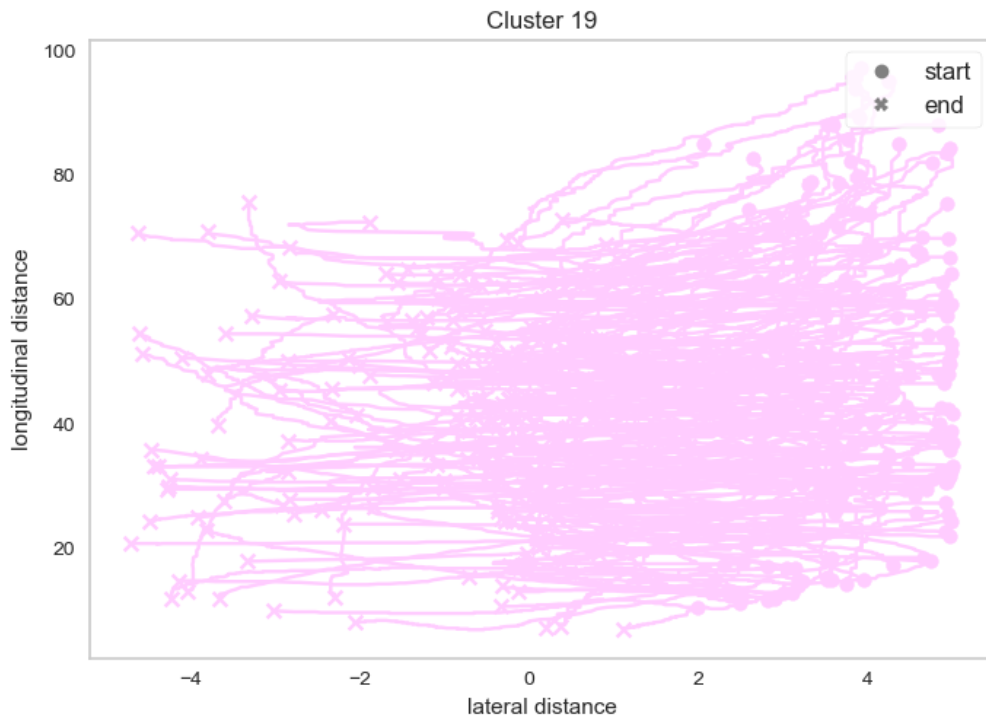
**Diagramm 6.6:** Ein Dendrogramm mit 20 Clustern als Ergebnis des Clusters von Signalen der Verkehrsteilnehmer-Kategorie [Eigene Darstellung]

Für eine geeignete Darstellung der verwendeten Signale für die Clustermethode, dient das Diagramm 6.7, indem die einzelnen Trajektorien, auf Basis der Signale zu den Positionen, der Verkehrsteilnehmer dargestellt werden. Dabei ist anzumerken, dass die Trajektorien der Verkehrsteilnehmer im relativen Bezug zum Ego-Fahrzeug stehen. So verlaufen zum Beispiel die rot gefärbten Trajektorien unmittelbar vor dem Ego-Fahrzeug, während die schwarz gefärbten Trajektorien rechts, also auf der rechten Nebenspur, neben und vor dem Ego-Fahrzeug verlaufen.

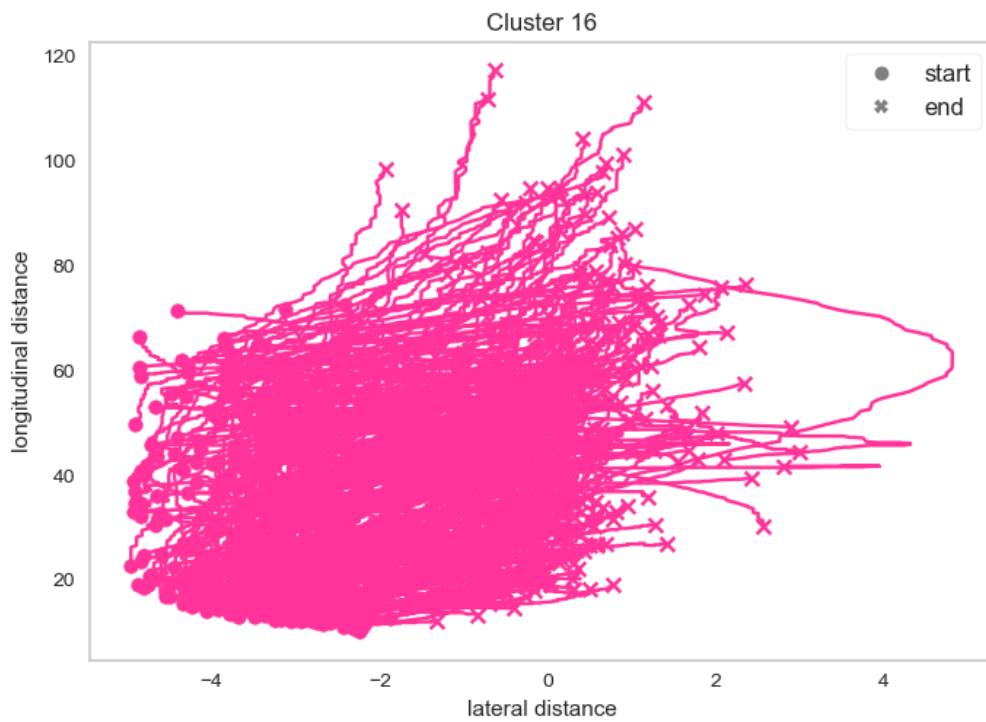


**Diagramm 6.7:** Trajektorien der Verkehrsteilnehmer [Eigene Darstellung]

Um einen Einblick in die einzelnen Cluster zu erhalten und damit die Interpretation von Clustern zu ermöglichen, wird im Folgenden eine Auswahl von Clustern untersucht, die übrigen Darstellungen der anderen Cluster finden sich im Anhang. So zum Beispiel sind in dem Diagramm 6.8 Trajektorien zu einem Muster für einen Spurwechsel von der rechten Seite in die mittlere Fahrbahn bzw. linke Fahrbahn festzustellen. Ein weiteres Beispiel zeigt ein Muster eines Überholvorgangs. Aufgrund der Endpunkte der Trajektorien, die oberhalb ihrer Startpunkte münden, handelt es sich hierbei um das Überholen des Ego-Fahrzeugs von Verkehrsteilnehmern (6.9). Es ist festzustellen, dass nicht alle Trajektorien der vorgestellten Cluster zu einem gleichen Szenario zugeordnet werden können. Solche Abweichungen entstehen durch das verwendete Cluster-Verfahren.

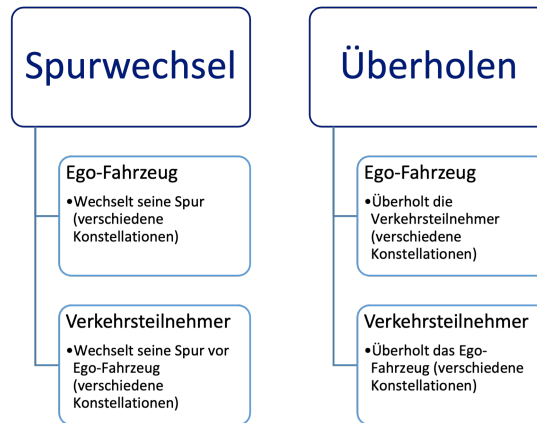


**Diagramm 6.8:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 19 [Eigene Darstellung]



**Diagramm 6.9:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 16 [Eigene Darstellung]

Die Abbildung 6.1 zeigt die Gesamtheit gefundener Szenarien auf Basis der Interpretation von Clustern der Verkehrsteilnehmer-Kategorie. Es lassen sich zwei übergeordnete Szenarien, nämlich Spurwechsel und Überholen, definieren. Diese beiden Szenarien können in speziellere Szenarien verfeinert werden.



**Abbildung 6.1:** Erkannte potenzielle Szenarien für die Verkehrsteilnehmer-Kategorie [Eigene Darstellung]

## 6.2 Kategorie 2: Ego

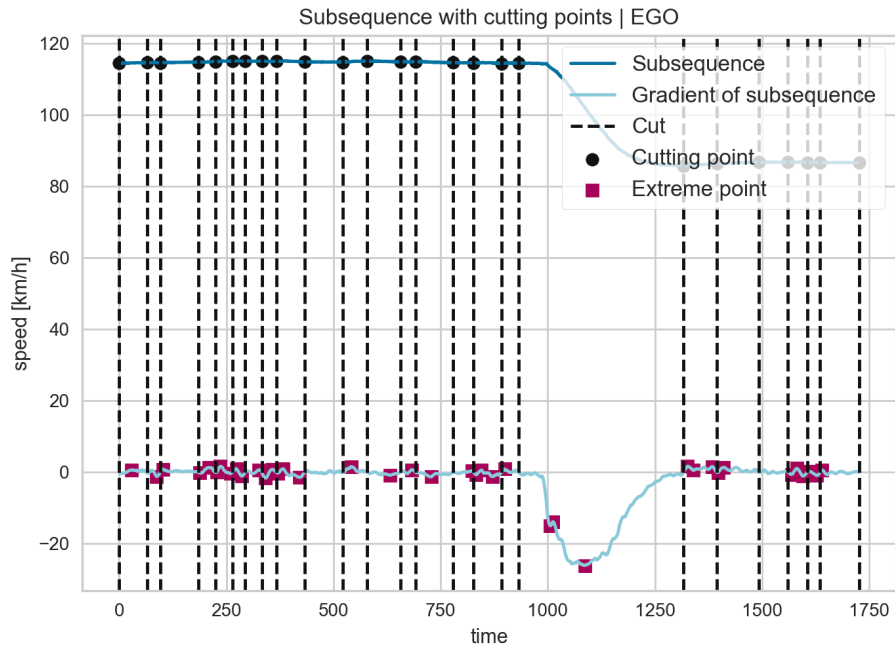
Analog zum Abschnitt zur Verkehrsteilnehmer-Kategorie werden in diesem Abschnitt die Ergebnisse für die Kategorie Ego vorgestellt. Zunächst werden die verwendeten Signale sowie Features aufgezählt. Des Weiteren werden weitere Beispiel zur Anwendung des Schnittalgorithmus mit unterschiedlichen Prominenzwerten untersucht. Folglich wird von den Qualitätsmaßen die optimale Clusteranzahl bestimmt, die im Weiteren für den Schnitt des Dendrogramms verwendet wird. Anhand weiterer Grafiken werden die einzelnen Cluster zur Interpretation dargestellt. Anschließend werden die verwendeten Features in einem Boxplot dargestellt. Letztlich werden alle erkannten Szenarien in einer Abbildung dargestellt.

Für die Auswertung in der Ego-Kategorie wird die Geschwindigkeit des Ego-Fahrzeugs als Signal untersucht (siehe Tabelle 6.2).

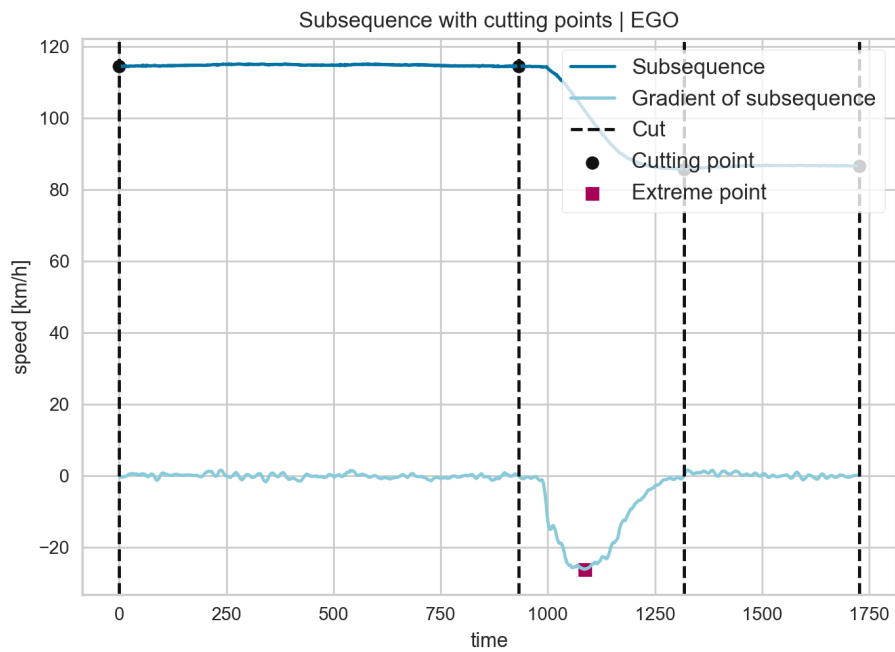
Signal	Features
Geschwindigkeit	g_min
	g_max
	g_med

**Tabelle 6.2:** Überblick über die verwendeten Signale und Features für die Ego-Kategorie [Eigene Darstellung]

Beispiele für die Auswirkung der Wahl des Prominenzwertes finden sich in den Diagrammen 6.10 und 6.11 wieder. Hierbei ist ein höher Prominenzwert sinnvoll, um zusammenhängende Abschnitte mit Mustern mit konstanter Geschwindigkeit von Musterabschnitten mit fallender Geschwindigkeit zu unterscheiden.



**Diagramm 6.10:** Ein Beispiel für das Schneiden des Signals „Geschwindigkeit“ der Ego-Kategorie mit Prominenzwert 1 [Eigene Darstellung]

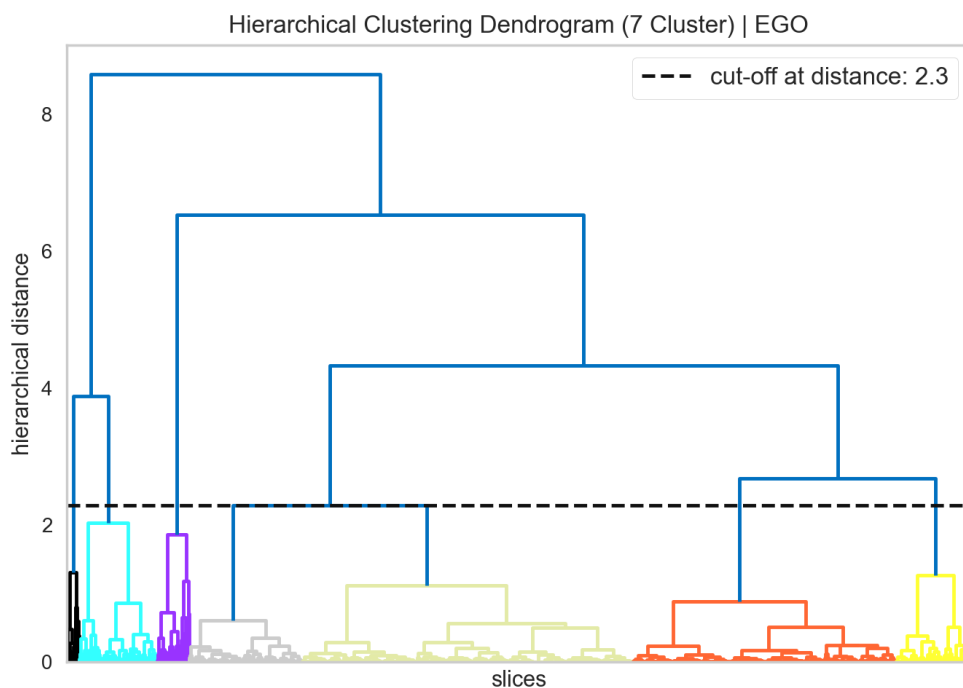


**Diagramm 6.11:** Ein Beispiel für das Schneiden des Signals „Geschwindigkeit“ der Ego-Kategorie mit Prominenzwert 8 [Eigene Darstellung]

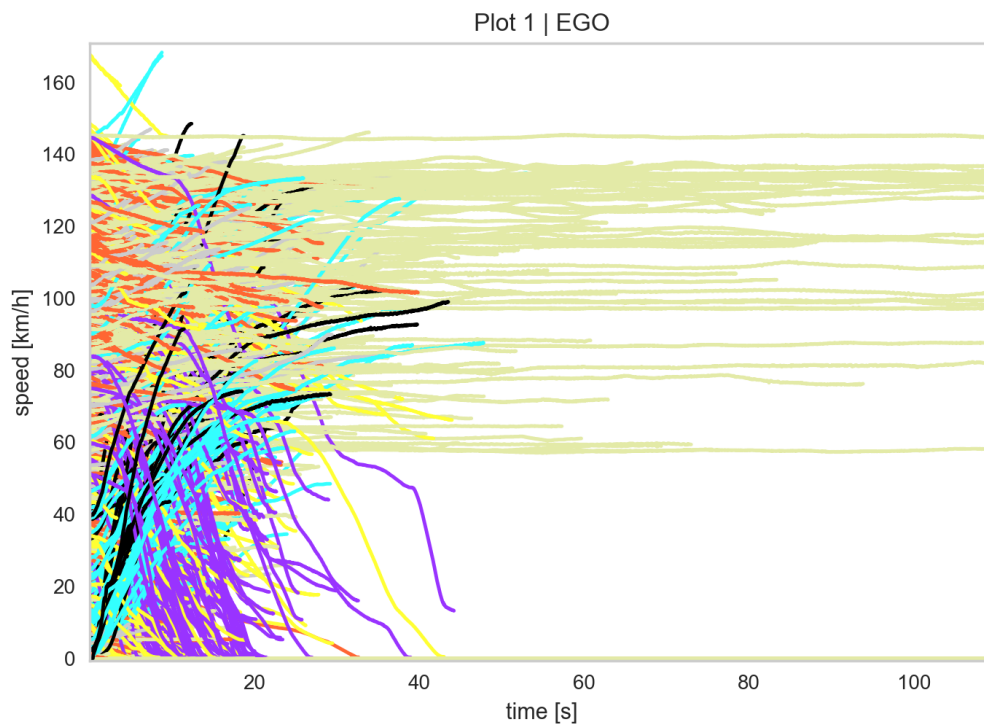
Die Tabelle 6.3 zählt die Vorschlagswerte von den Qualitätsmaßen für die optimale Clusteranzahl auf. In diesem Fall wird die von Calinski-Herabsz vorgeschlagene Zahl übernommen. Entsprechend wird ein Dendrogramm mit einer Clusteranzahl von 7 erstellt (siehe Diagramm 6.12).

Qualitätsmaß	Optimale Clusteranzahl
Distortion	6
Calinski-Herabsz	7
Silhouette	2
Manuelle Bestimmung (Calinski-Herabsz)	7

**Tabelle 6.3:** Ego-Kategorie: Überblick über die vorgeschlagenen Werte für die optimale Clusteranzahl [Eigene Darstellung]



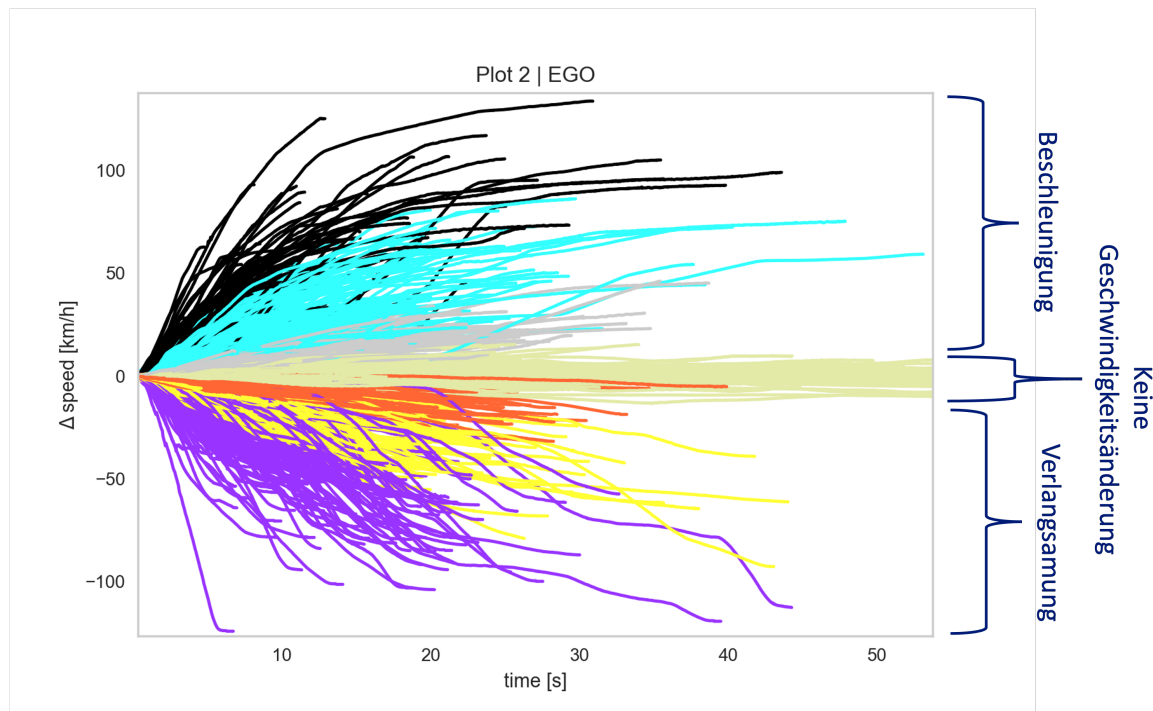
**Diagramm 6.12:** Ein Dendrogramm mit 7 Clustern als Ergebnis des Clusters von Signalen der Ego-Kategorie [Eigene Darstellung]



**Diagramm 6.13:** Alle geclusterten Signale der Ego-Kategorie [Eigene Darstellung]

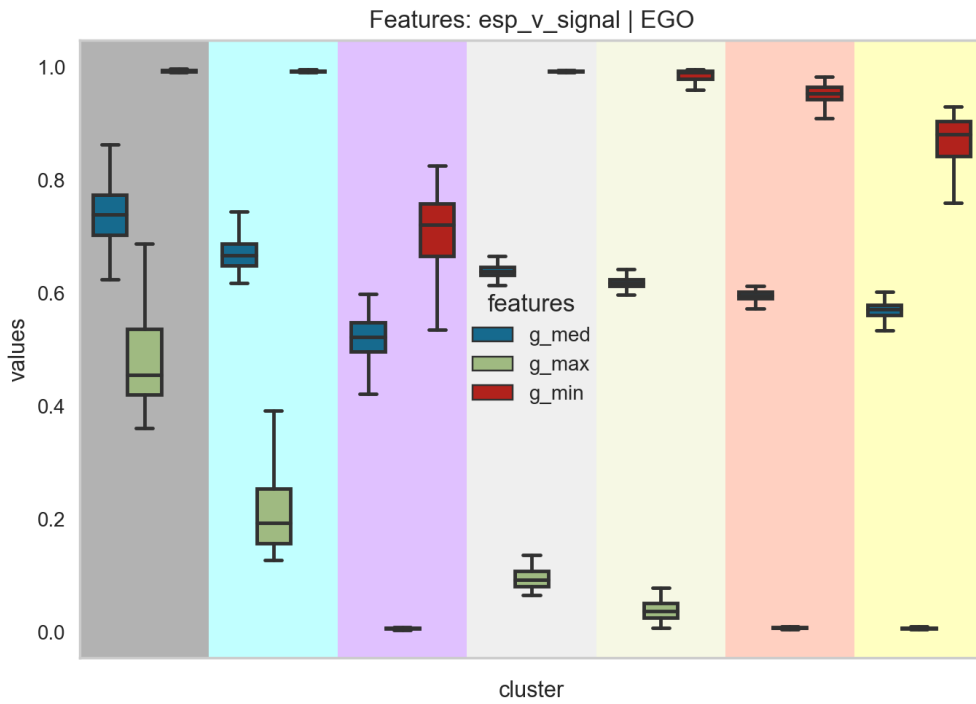
Zur Abbildung der geclusterten Geschwindigkeitssignale werden diese in dem Diagramm 6.13 zusammengefasst. Auf den ersten Blick fällt auf, dass aufgrund der Überlappungen der gesamten Signale, es kompliziert ist diese zu interpretieren. Abhilfe schafft das nächste Diagramm, indem die Änderung der Geschwindigkeitswerte dargestellt werden (siehe Diagramm 6.14). Dabei kann man anhand der Signale Muster als Beschleunigungs-, Verlangsamungsszenarien sowie Szenarien, bei dem sich die Geschwindigkeit des Fahrzeugs nicht ändert bzw. konstant bleibt, deuten. Es ist aber auch festzustellen, dass das Cluster mit dem zugeordneten Szenario „keine Geschwindigkeitsänderung“ auch Signale aus anderen Clustern enthält, sodass hier noch eine weitere Trennung dieses Clusters realisiert werden kann, indem der Baum weiter unten geschnitten wird.





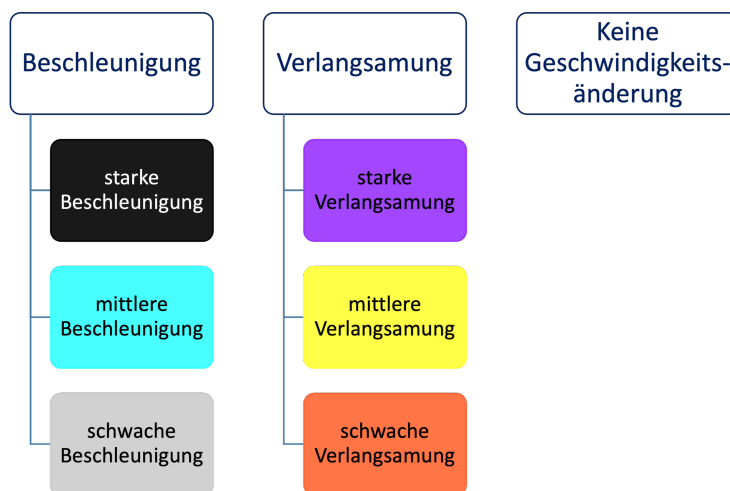
**Diagramm 6.14:** Alle geclusterten Signale der Umwelt-Kategorie mit Referenz zur Null-Linie [Eigene Darstellung]

Um zu sehen, welche Features einen Einfluss auf die entstandenen Cluster haben, bietet sich ein Boxplot an. Hierfür werden auf der X-Achse die Cluster aus dem vorherigen Dendrogramm, entsprechend der Reihenfolge, farblich hinterlegt. Auf der Y-Achse werden die Werte der normierten Features angezeigt. Innerhalb der Cluster befinden sich die Werte zu den verwendeten Features. So hat beispielsweise das als Beschleunigungs-Szenario gedeutete Cluster in schwarzer Farbe, deutlich höhere Werte für das Feature „g\_max“. Im Gegenzug dazu besitzen die als Verlangsamungs-Szenario gedeutete Cluster niedrigere Werte für „g\_min“ als die Beschleunigungs-Szenarien.



**Diagramm 6.15:** Ein Boxplot mit Feature-Werten des Signals „Geschwindigkeit“ für alle Cluster der Ego-Kategorie [Eigene Darstellung]

Die Abbildung 6.2 zählt die möglichen Szenarien auf, die anhand vorheriger Darstellungen zu den entstandenen Clustern interpretiert werden können. Somit entstehen drei Hauptszenarien, während Szenarien zur Beschleunigung und Verlangsamung in weitere Szenarien eingeteilt werden können.



**Abbildung 6.2:** Erkannte potenzielle Szenarien für die Ego-Kategorie [Eigene Darstellung]

## 6.3 Kategorie 3: Umwelt

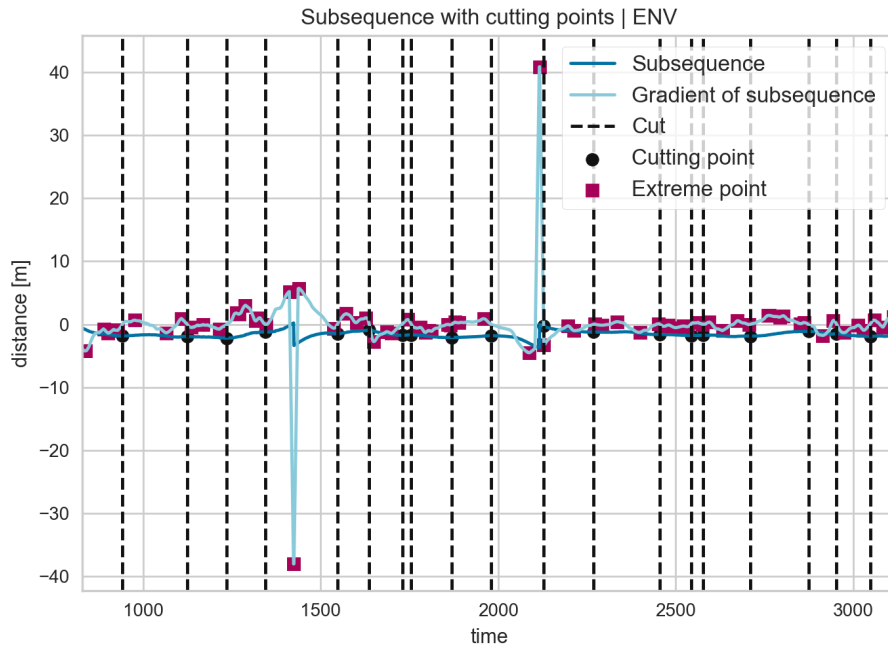
Die Ergebnisse für die Kategorie „Verkehrsteilnehmer“ werden in diesem Abschnitt vorgestellt. Zu diesem Zweck werden die hierfür verwendeten Signale und die dazugehörigen ausgewählten Features in der Tabelle 6.4 aufgezählt. Es folgen Grafiken zum Schneiden von Signalen mit unterschiedlichen Prominenzwerten. Für den optimalen Schnitt des Dendrogramms werden die Werte verschiedener Qualitätsmaße sowie die manuelle Wertzuweisung dargestellt. Anschließend wird ein Dendrogramm mit den resultierenden Clustern vorgestellt. In diesem Zusammenhang werden diese Cluster durch zusätzliche Diagramme weiter untersucht. Abschließend folgt eine Darstellung der interpretierten Szenarien auf Basis der Clusterbildung.

Signal	Feature
Linie rechts	s_end
	g_min
	g_max
Linie links	s_end
	g_min
	g_max

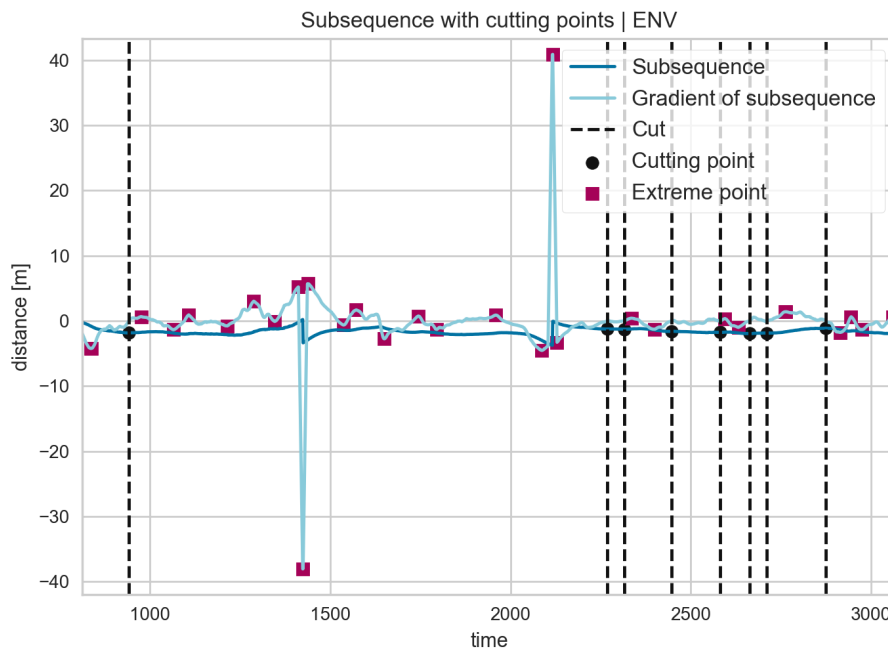
**Tabelle 6.4:** Überblick über die verwendeten Signale und Features für die Umwelt-Kategorie [Eigene Darstellung]

Die Diagramme 6.16 und 6.17 unterscheiden sich in der Wahl des Wertes für die Prominenz. Sodass im Diagramm 6.16 mit dem Prominenzwert 1 Signalabschnitte für die Muster eines Spurwechsels zur rechten Seite und im weiteren Abschnitt eines Spurwechsels zur linken Seite. Die beiden Muster sind anhand der großen Sprünge des Signals bzw. des Gradienten ersichtlich, wobei diese sich auch stark von anderen Mustern abgrenzen. Liegt der Wert bei 1 für die Prominenz, kann ein Signalabschnitt mit einem Muster für einen Mehrfachspurwechsel erzeugt werden (siehe Diagramm 6.17).

Daraus folgt, dass für dieses Beispiel der Prominenzwert wahrscheinlich eine entscheidende Rolle zur Findung von Mustern für kombinierte Spurwechsel spielt.



**Diagramm 6.16:** Ein Beispiel für das Schneiden des Signals „Linie rechts“ der Umwelt-Kategorie mit Prominenzwert 0.1 [Eigene Darstellung]



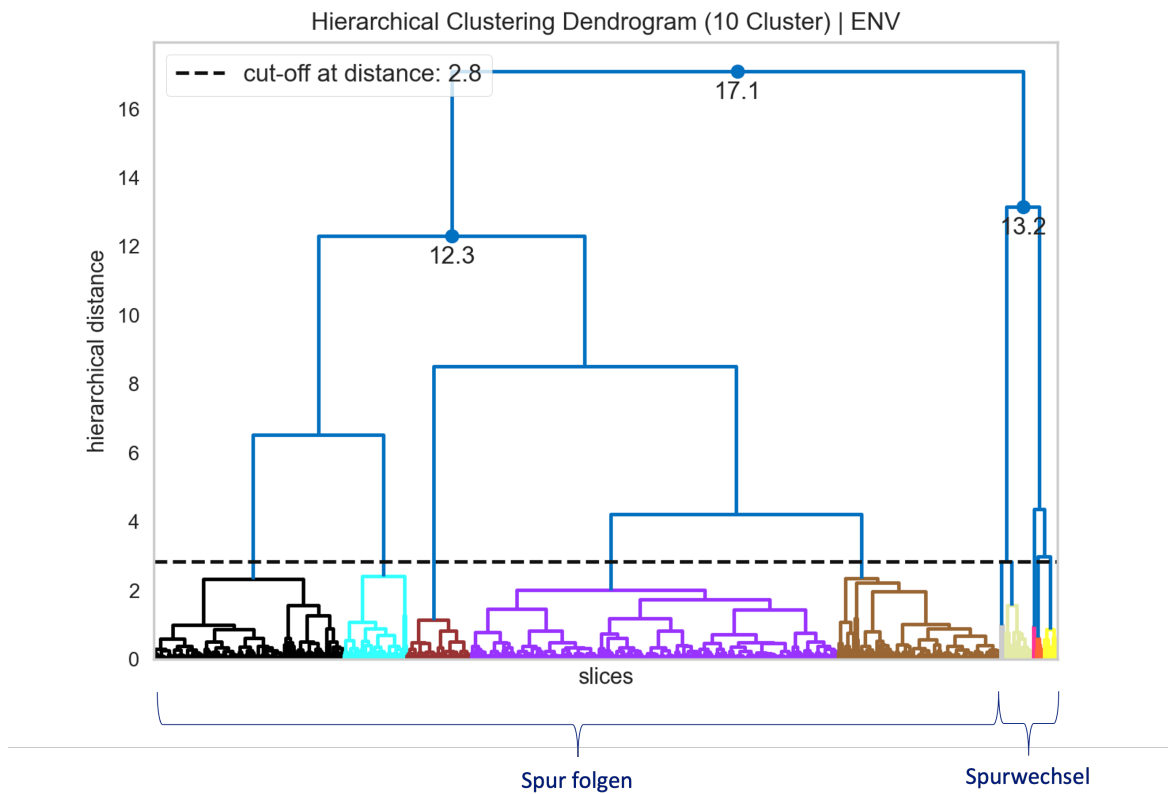
**Diagramm 6.17:** Ein Beispiel für das Schneiden des Signals „Linie rechts“ der Umwelt-Kategorie mit Prominenzwert 1 [Eigene Darstellung]

Als optimale Clusteranzahl für das Dendrogramm in der Umwelt-Kategorie schlagen die Qualitäts-

maße Distortion und Calinski-Herabsz die Zahl 5 vor. Lediglich Silhouette weicht mit der Zahl von 2 stark von den anderen ab. Zur Erleichterung der Interpretierbarkeit weiterer Darstellungen hinsichtlich entstandener Cluster, wird die optimale Clusteranzahl von 10 bestimmt (siehe Diagramm 6.17).

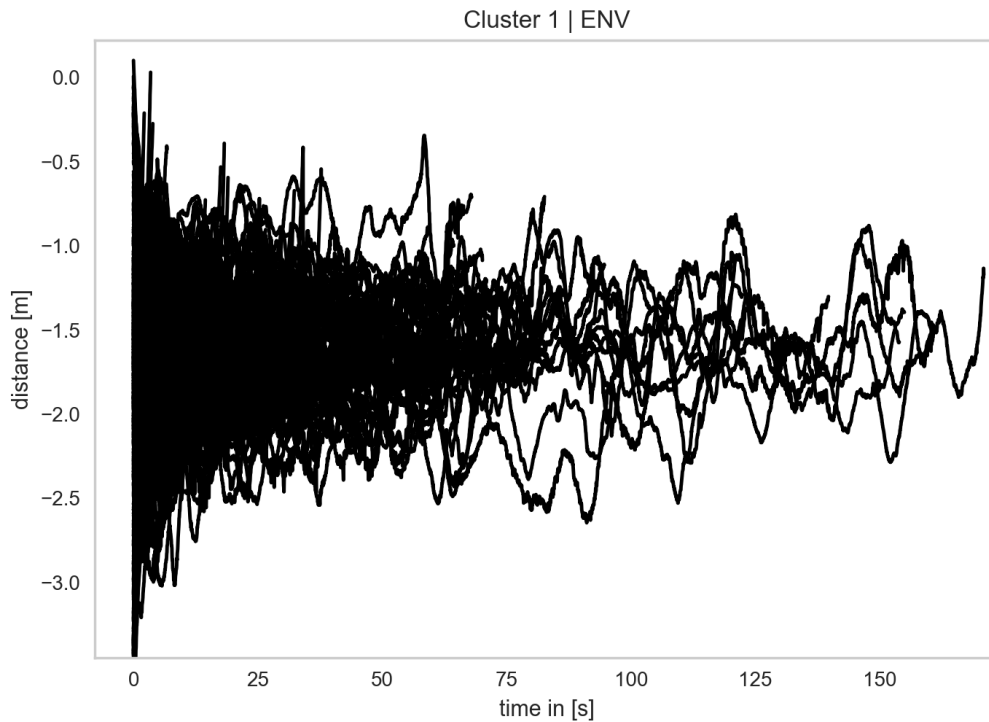
Qualitätsmaß	Optimale Clusteranzahl
Distortion	5
Calinski-Herabsz	5
Silhouette	2
Manuelle Bestimmung	10

**Tabelle 6.5:** Umwelt-Kategorie: Überblick über die vorgeschlagenen Werte für die optimale Clusteranzahl [Eigene Darstellung]



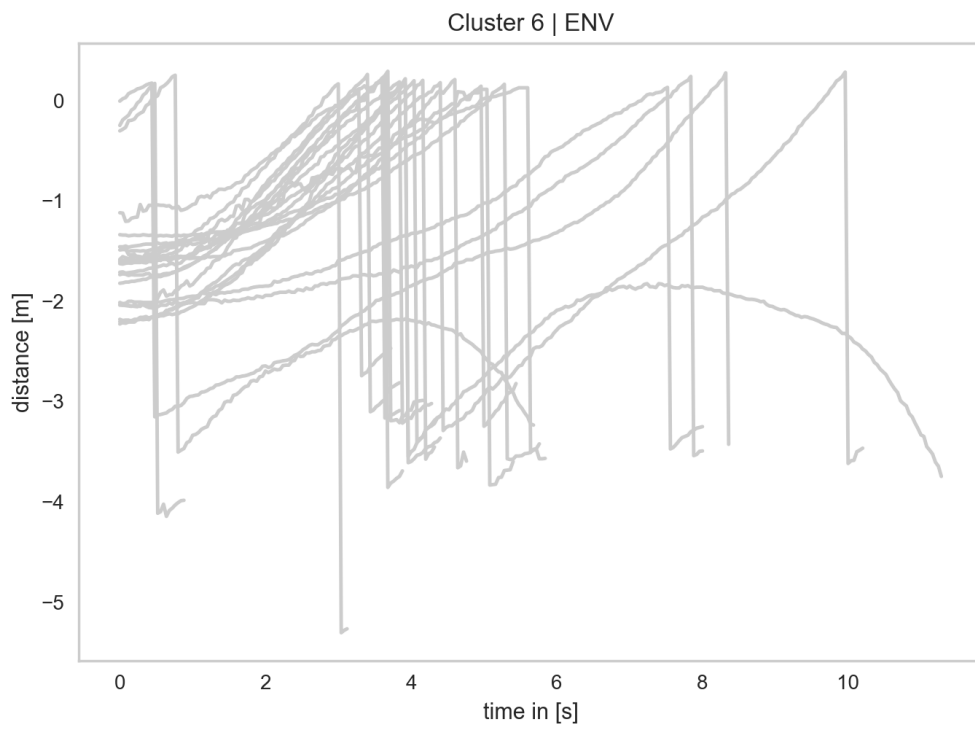
**Diagramm 6.18:** Ein Dendrogramm mit zehn Clustern als Ergebnis des Clusters von Signalen der Umwelt-Kategorie [Eigene Darstellung]

Für ein mögliches Szenario „Spur folgen“ kann das Diagramm 6.19 für diesen Zweck gedeutet werden, da die meisten Signale in dem mittleren Wertebereich pendeln. Weitere Cluster zu diesem Szenario befinden sich im Anhang.

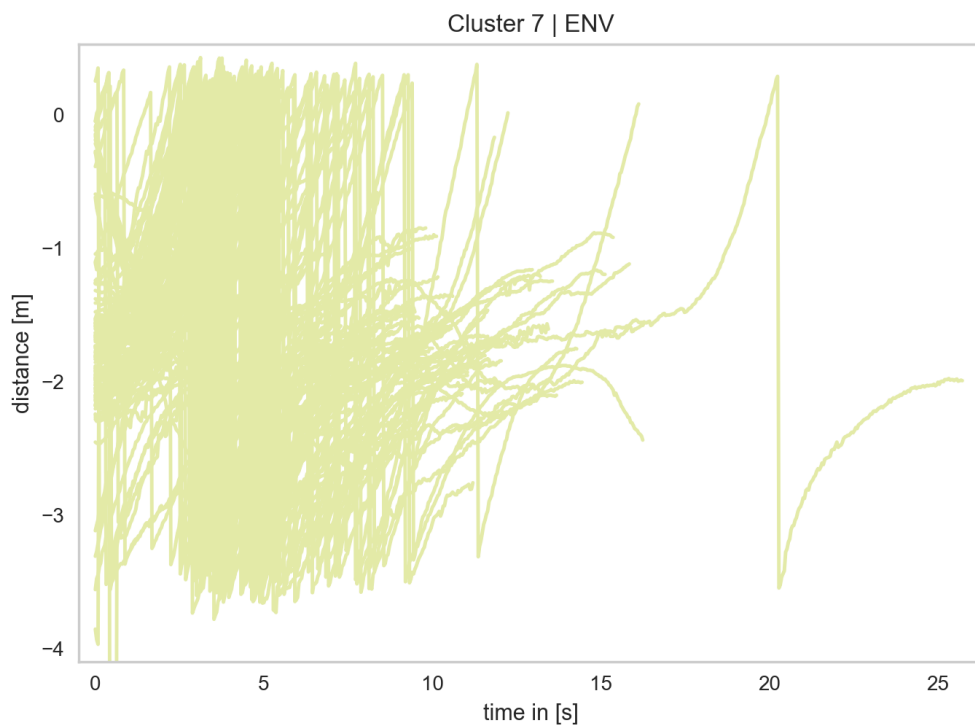


**Diagramm 6.19:** Signale der Umwelt-Kategorie aus Cluster 1 [Eigene Darstellung]

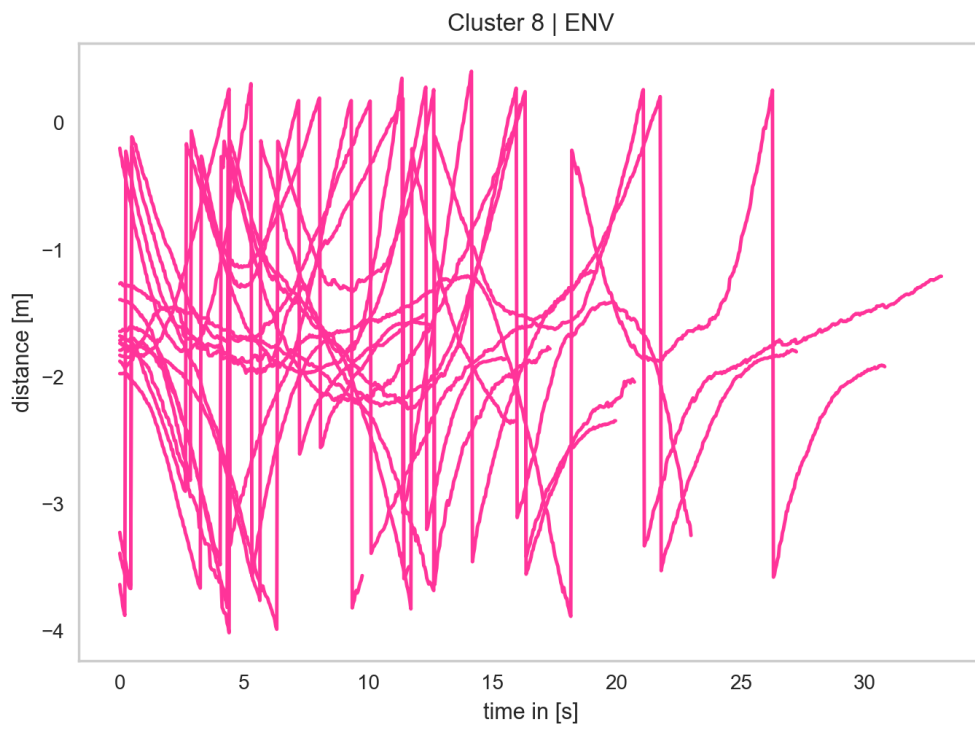
Das Diagramm 6.20 zeigt ein weiteres potenzielles Szenario auf. Es deutet auf einen „Spurwechsel nach rechts“ an, da die Signale sich der Null-Linie nähern, das heißt der Abstand des Ego-Fahrzeugs zur rechten Spurlinie geht gegen Null und sobald das Fahrzeug über der rechten Spurlinie ist, wird die neue rechte Spurlinie als rechte Spurlinie erkannt. Dadurch entsteht auch der gewaltige Sprung im Signal, da beim Überqueren der rechten Spurlinie der Abstand zur neuen rechten Spurlinie maximal ist. Ein Szenario für einen möglichen „Doppelspurwechsel nach rechts“ ist im Diagramm 6.21 zu finden. Weiterhin stellt das Diagramm 6.22 Muster für das Szenario eines „Mehrfachspurwechsels“ dar. Ein Szenario mit „Spurwechsel nach links“ wird im Diagramm 6.23 abgebildet. Das Diagramm 6.24 enthält Muster für das Szenario „Doppelspurwechsel nach links“.



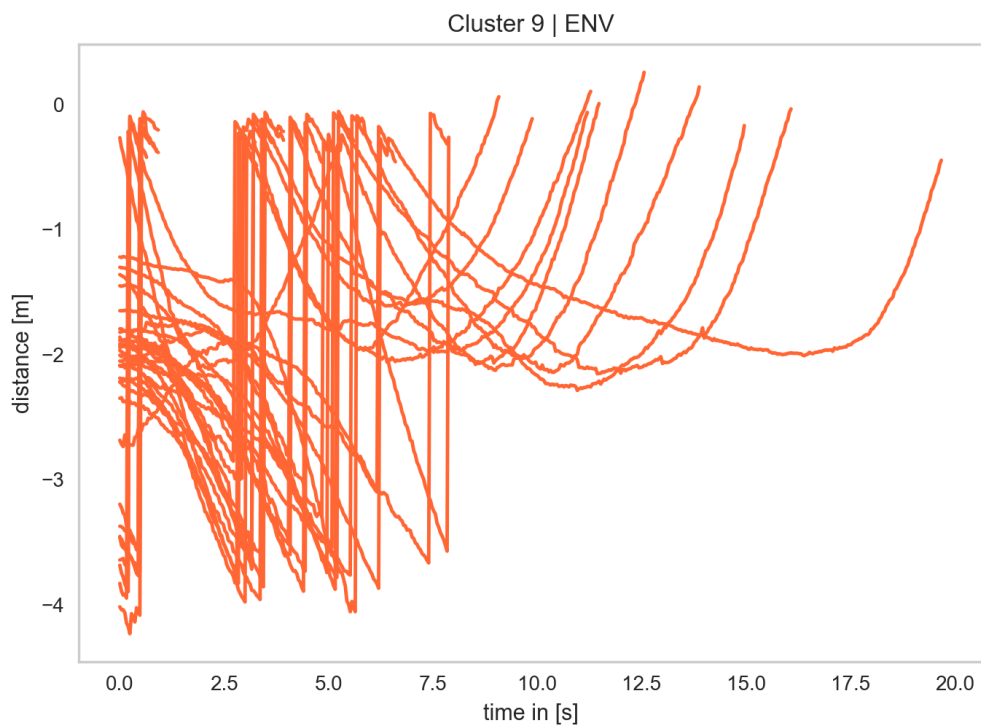
**Diagramm 6.20:** Signale der Umwelt-Kategorie aus Cluster 6 [Eigene Darstellung]



**Diagramm 6.21:** Signale der Umwelt-Kategorie aus Cluster 7 [Eigene Darstellung]

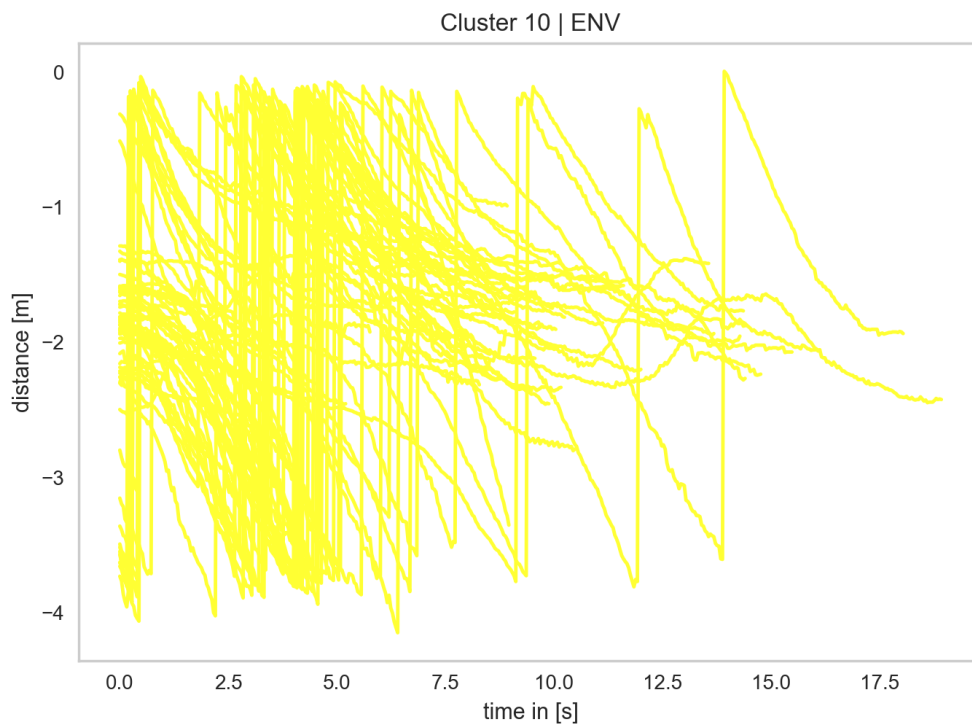


**Diagramm 6.22:** Signale der Umwelt-Kategorie aus Cluster 8 [Eigene Darstellung]



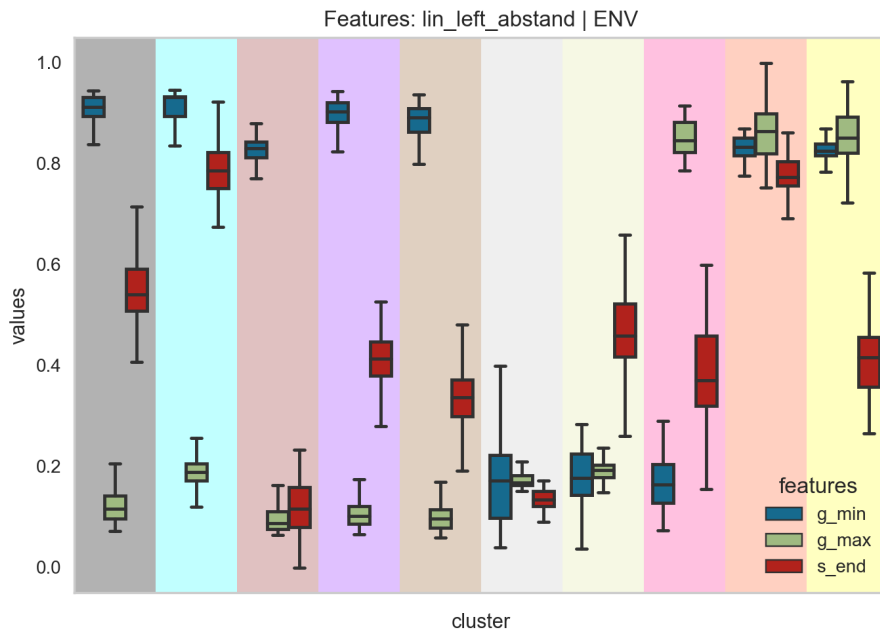
**Diagramm 6.23:** Signale der Umwelt-Kategorie aus Cluster 9 [Eigene Darstellung]



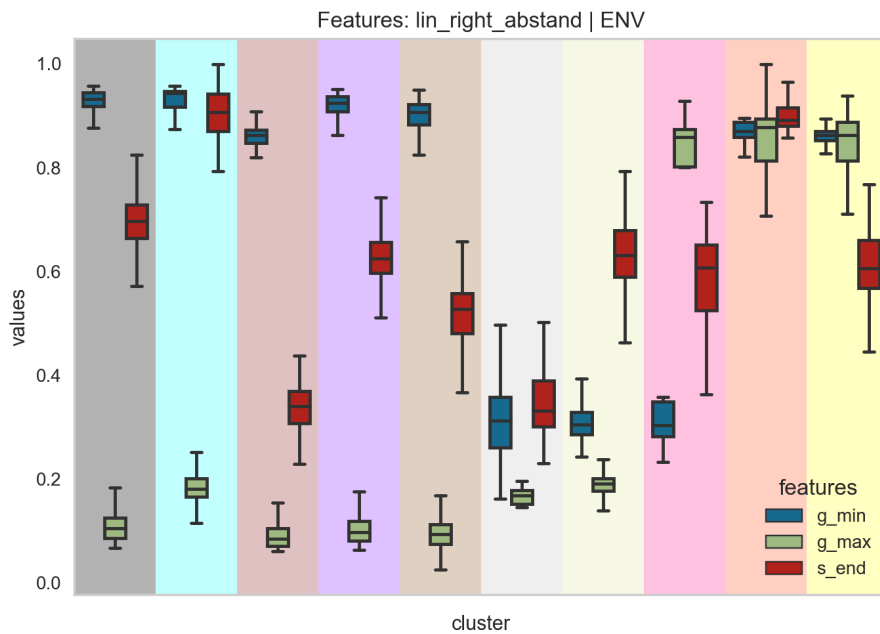


**Diagramm 6.24:** Signale der Umwelt-Kategorie aus Cluster 10 [Eigene Darstellung]

Die verantwortlichen Features für die Entstehung der Cluster können dem Boxplot aus Diagramm 6.25 für das Signal „Linie links“ und dem Boxplot aus Diagramm 6.26 für das Signal „Linie rechts“ entnommen werden. So haben zum Beispiel die Varianten zum Spurwechsel deutlich höhere Werte des Features „g\_max“.

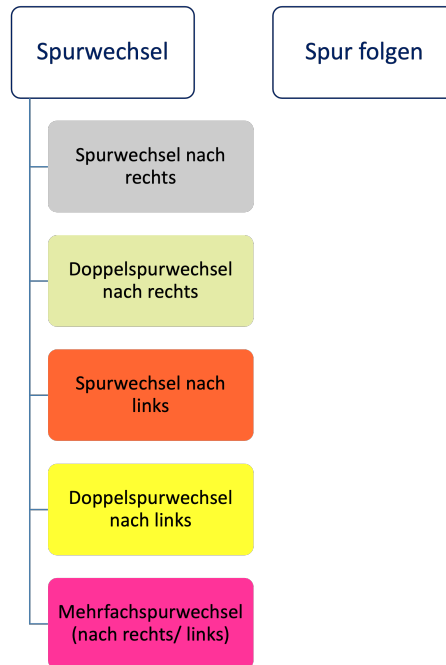


**Diagramm 6.25:** Ein Boxplot mit Feature-Werten des Signals „Linie links“ für alle Cluster der Umwelt-Kategorie [Eigene Darstellung]



**Diagramm 6.26:** Ein Boxplot mit Feature-Werten des Signals „Linie rechts“ für alle Cluster der Umwelt-Kategorie [Eigene Darstellung]

Für die Umwelt-Kategorie können die gewonnen potenzielle Szenarien durch die Interpretation der Darstellungen zu den entstandenen Clustern in der Abbildung 6.3 dargestellt werden. Grundsätzlich werden zwei Hauptszenarien gedeutet, wobei das Spurwechsel-Szenario in weitere Variationen gespaltet werden kann.



**Abbildung 6.3:** Erkannte potenzielle Szenarien für die Umwelt-Kategorie [Eigene Darstellung]

## 6.4 Grenzen der Methode

Die implementierte Methode zeigt interessante Ergebnisse, hat aber auch einige Schwächen, die im Folgenden aufgeführt werden: Die Methode dieser Arbeit verwendet Parameter wie die Prominenz, die für das Schneiden von Signalen notwendig ist, um die Schnittpunkte zu bestimmen. Die Wahl des Prominenzwertes beeinflusst stark die Anzahl der gefundenen Signalschnittpunkte sowie die Qualität der Signalschnitte und wirkt sich somit auf die Ergebnisse der Clustermethode aus.

Außerdem ist die signalspezifische Wahl des Prominenzwertes keine Garantie für aussagekräftige Ergebnisse, da ein individueller Wert für die Prominenz auch ggf. für die Gesamtmenge der Signalabschnitte eines Signaltyps ermittelt werden muss.

Die Dendrogramme sind bei großen Datensätzen leicht unübersichtlich und schwer zu interpretieren, sodass weitere Darstellungen notwendig sind, um zu einem Ergebnis zu kommen. Wobei es auch bei den weiteren Darstellungen zu Verwirrungen bei der Interpretation kommen kann. Außerdem muss oft eine manuelle Auswahl der optimalen Anzahl von Clustern für den Ausschnitt des Baumes gewählt werden, da sonst die Grafiken für die einzelnen Cluster zur Interpretation ungeeignet sind.

Da das Ergebnis aus der manuellen Interpretation resultiert, können leicht Fehler bei der Interpretation der Muster entstehen.

## 6.5 Diskussion

Grundsätzlich kann davon ausgegangen werden, dass die umgesetzte Methodik dieser Arbeit, aufgebaut auf einer Methode aus der Literatur, das Ziel zum Finden von weiteren Fahrscenarien in den Messdaten durch Einsatz maschineller Lernalgorithmen und von Erweiterungen erfüllt.

In der Tabelle 6.6 wird die Methode dieser Arbeit seiner Ausgangsmethode nach gegenübergestellt, indem die Eingangsdaten für die Clustermethode und die Endergebnisse, also die als potenziell gedeutete Szenarien, präsentiert werden. Dabei ist es anhand der zahlreichen Ergebnisse ersichtlich, dass durch die Erweiterungen weitere potenzielle Szenarien in den realen Messdaten identifiziert werden können. Weiterhin ist zu erwähnen, dass alle aufgelisteten Features zu Testzwecken verwendet werden und nur eine Teilmenge für die Darstellungen dieser Arbeit (siehe Abbildung 5.2) in Betracht gezogen wird.

	Methode nach [Montanari, German und Djanatliev 2020]	Methode dieser Arbeit	Kategorie
Signal	Krümmung links	Position x Position y	<i>Verkehrsteilnehmer</i>
	Geschwindigkeit	Geschwindigkeit	<i>Ego</i>
	Linie links	Linie rechts Linie links	<i>Umwelt</i>
Feature	s_med, s_max, s_min, g_med, g_max, g_min	s_med, s_max, s_min, g_med, g_max, g_min s_start, s_end, g_start, g_end, g_big	
Datenmenge [Fahrzeit]	ca. 30 Minuten	ca. 20 Stunden	
<b>Resultate</b>			
Potenzielle Szenarien		- Spurwechsel >vom Ego-Fahrzeug >vom Verkehrsteilnehmer - Überholen >vom Ego-Fahrzeug >vom Verkehrsteilnehmer	<i>Verkehrsteilnehmer</i>
		- Beschleunigung >starke, mittlere, schwache - Verlangsamung >starke, mittlere, schwache - konst. Geschwindigkeit	<i>Ego</i>
	- Spurwechsel >links/ rechts, Mehrfachspurwechsel - Kein Spurwechsel/ spezielle Eckfälle	- Spurwechsel >links/ rechts, Doppelspurwechsel links/ rechts, Mehrfachspurwechsel - Spur folgen/ kein Spurwechsel	<i>Umwelt</i>

**Tabelle 6.6:** Vergleich Francescos Methode mit meiner [Eigene Darstellung]

## 7 Zusammenfassung und Ausblick

In diesem Kapitel werden die Ziele, der Aufbau und der Ansatz der Arbeit zusammengefasst. Es schließt sich eine Betrachtung der möglichen Anwendungen der ausgearbeiteten Methodik an.

### 7.1 Zusammenfassung

Die Zielsetzung dieser Arbeit ist die Entwicklung eines datenbasierten Verfahrens, um automatisiert Fahrscenarien aus realen Messdaten von Fahrzeugen zu extrahieren.

Zu Beginn der Arbeit wird im zweiten Kapitel ein Überblick über das automatisierte Fahren gegeben. In diesem Abschnitt werden grundlegende Definitionen, unterschiedliche Arten von Szenarien und verschiedene Grade der Automatisierung erläutert.

Anschließend werden Grundlagen zum maschinellen Lernen vermittelt. Neben den vielen Untermethoden des maschinellen Lernens, wird insbesondere hierbei Wert auf das agglomerative Verfahren als Untergebiet der hierarchischen Cluster-Verfahren gelegt, welches für die vorliegende Arbeit zum Clustern von Signalabschnitten verwendet wird.

Weiterhin werden Forschungsprojekte samt ihrer Projekte und der aktuelle Stand der Technik vermittelt, indem unterschiedliche Ansätze zur Generierung und Identifizierung von Fahrscenarien dargestellt werden.

Als Nächstes wird die zu entwickelnde Methode erklärt. Zu diesem Zweck wird die Idee sowie die Vorgehensweise dieser Methode, ausgehend von einer Methode aus der Publikation von den Autoren Montanari, German und Djanatliev zur Basisbildung sowie zur Einführung von Erweiterungen, erläutert.

Folglich werden die Ergebnisse der umgesetzten Methodik vorgestellt und bewertet. Hierzu werden unterschiedliche potenzielle Szenarien für jede der drei Kategorien Verkehrsteilnehmer, Ego und Umwelt gefunden. Außerdem werden die Schwächen sowie Zielerfüllung dieser Methode diskutiert. Eines der größten Schwächen beruht auf den großen Datenmengen, die für das Clustering verwendet werden. Es führt zur Unübersichtlichkeit in den Darstellungen und erschwert das Interpretieren von möglichen Szenarien. Zudem wird die Methode dieser Arbeit mit seiner Ausgangsbasis, sprich dem Ansatz nach Montanari, German und Djanatliev, zum Vergleich gestellt. Es zeigt sich, dass die Einführung von Erweiterungen und somit die Methodik dieser Arbeit die Identifikation weiterer Szenarien ermöglicht.

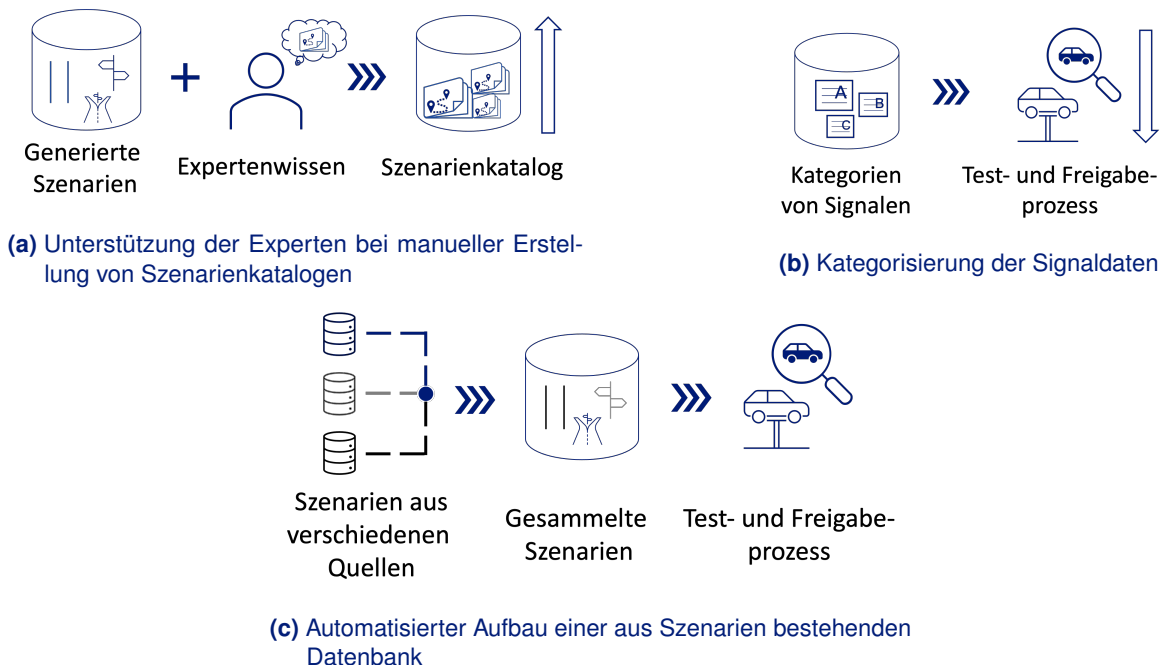
Diese Arbeit will einen Grundstein für die automatisierte Identifikation von Fahrscenarien in realen Messdaten mittels Clustern legen. Die resultierenden Cluster aus den drei definierten Kategorien Verkehrsteilnehmer, Ego und Umwelt können als potenzielle Fahrscenarien für ein effizientes Testen der Fahrfunktionen eingesetzt werden.

## 7.2 Ausblick

Die in dieser Arbeit entwickelte Methode kann zur automatisierten Generierung von Fahrscenarien und damit z. B. als Unterstützung für die Ersteller von Szenarienkatalogen zur Fahrzeugprüfung eingesetzt werden. Dabei stärken die gewonnenen Szenarien nicht nur das Wissensspektrum der Experten, sie können auch bestehende Kataloge erweitern beziehungsweise auf Vollständigkeit prüfen oder sogar bei der Konzeption neuer Kataloge dienen (siehe Abbildung 7.1a).

Um den Aufwand für den Test- und Freigabeprozess zu minimieren, können die zahlreichen Signale aus den Fahrdaten für die Weiterverarbeitung nicht nur nach den Perspektiven Verkehrsteilnehmer, Ego sowie Umgebung, sondern auch nach der Art des Verkehrswegs in Autobahn, Landstraße und Stadtstraße klassifiziert werden. Ein möglicher Ansatz hierfür ist, eine automatisierte Fahrfunktion des Parkassistenten der Kategorie Stadtstraße zuzuordnen, um nur die Signale dieser Kategorie für die weitere Verarbeitung oder Auswertung zu berücksichtigen (siehe Abbildung 7.1b).

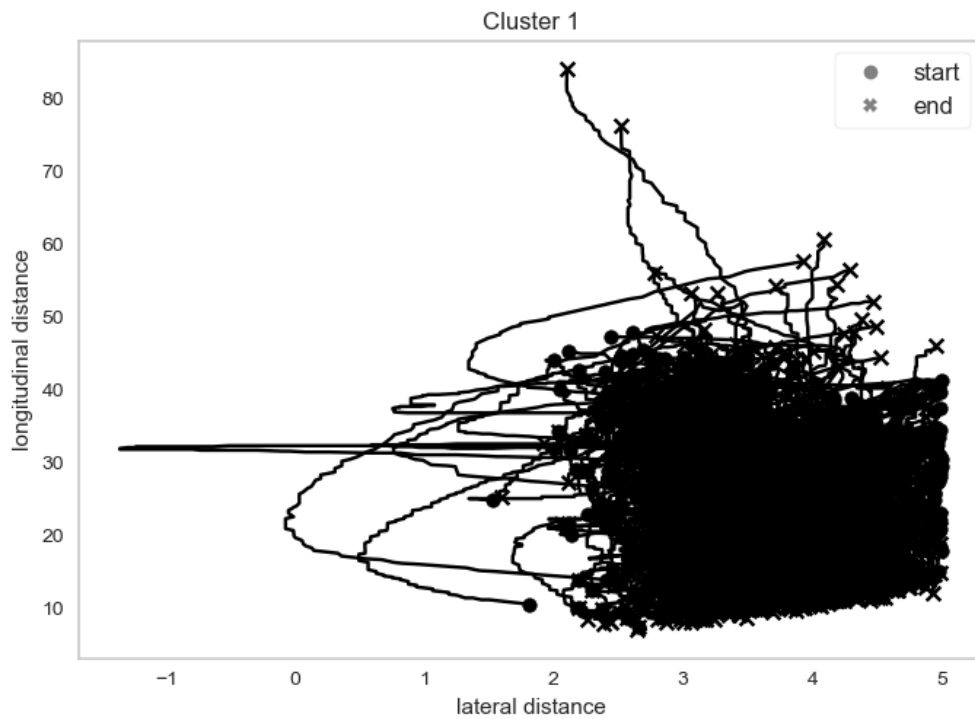
Mit der in dieser Arbeit vorgeschlagenen Methode können zukünftig die gewonnenen Szenarien über eine Schnittstelle zur Datenbank entsprechend aufbereitet und in der Datenbank archiviert werden. Geeignete Algorithmen erkennen dabei neue Szenarien und reichern die Datenbank mit diesen an. Zudem können die Szenarien auch aus weiteren, sowie unterschiedlichen Quellen hervorgehen. So kann die Datenbank im Laufe der Zeit mit neuen Szenarien ergänzt und zum Testen und Freigeben von automatisierten Fahrfunktionen genutzt werden (siehe Abbildung 7.1c).



**Abbildung 7.1:** Vorschläge für die möglichen Anwendungen der ausgearbeiteten Methodik [Eigene Darstellung]

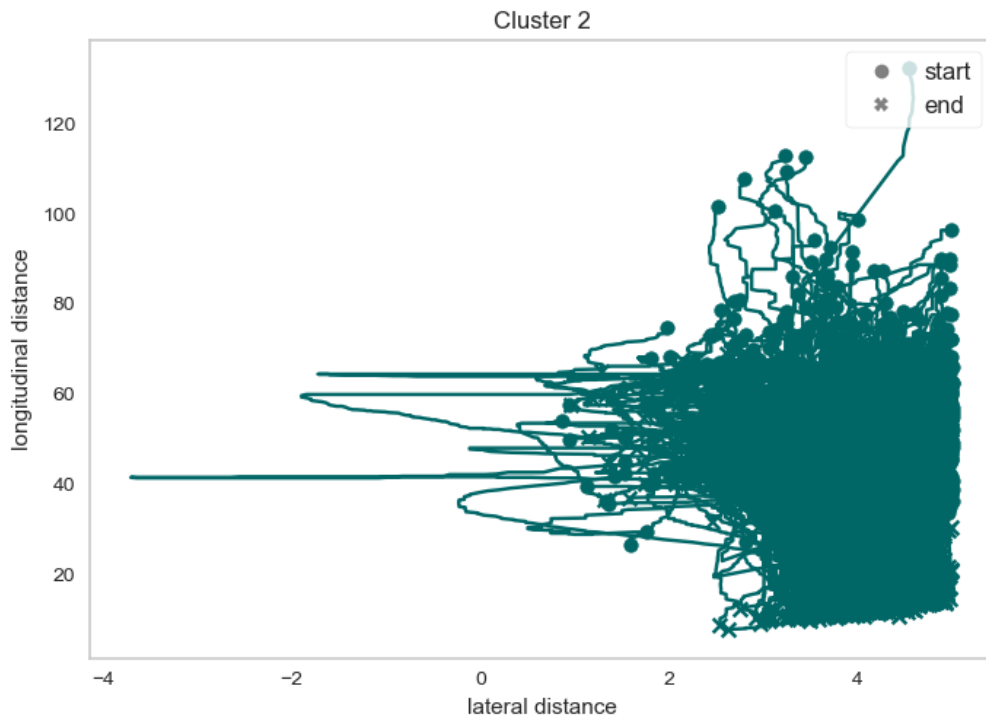
# Anhang

## Verkehrsteilnehmer-Kategorie

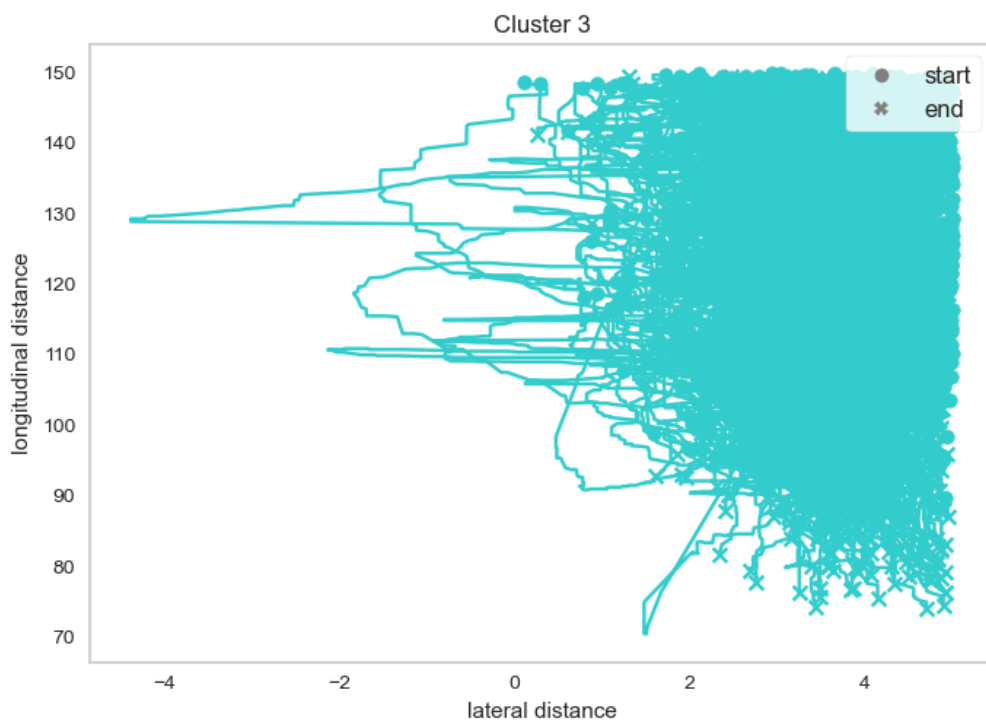


**Diagramm A.1:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 1 [Eigene Darstellung]

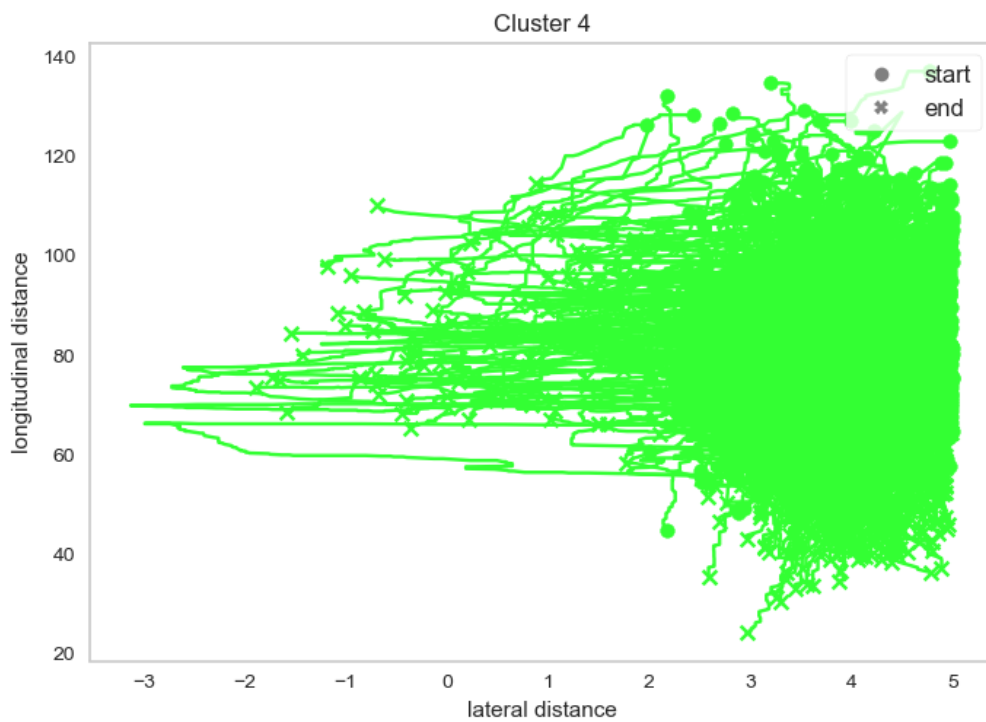




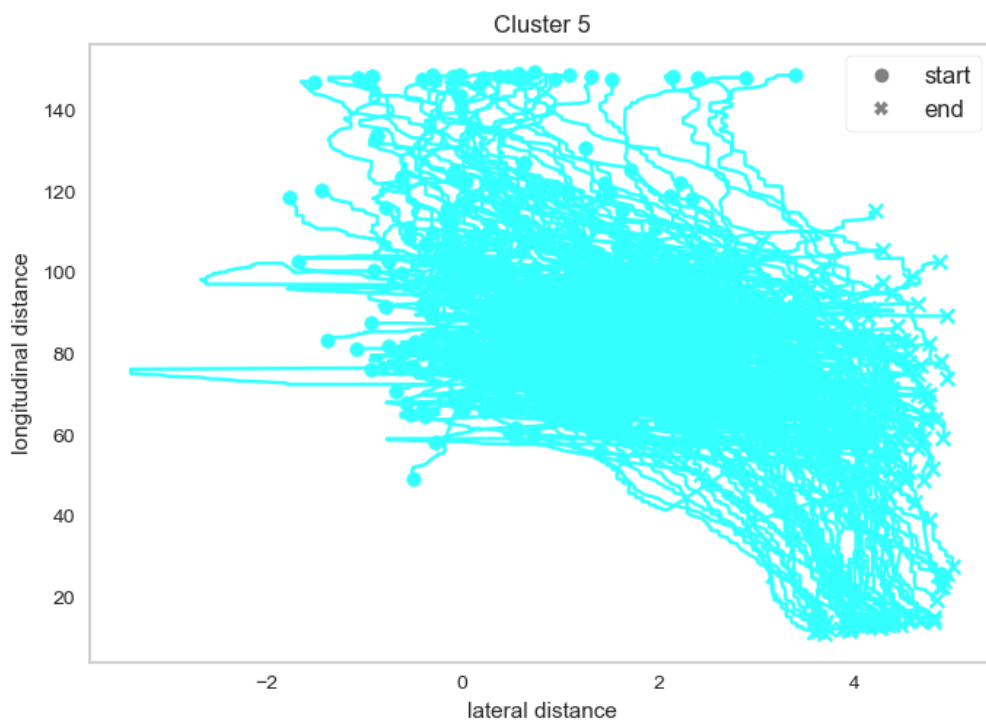
**Diagramm A.2:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 2 [Eigene Darstellung]



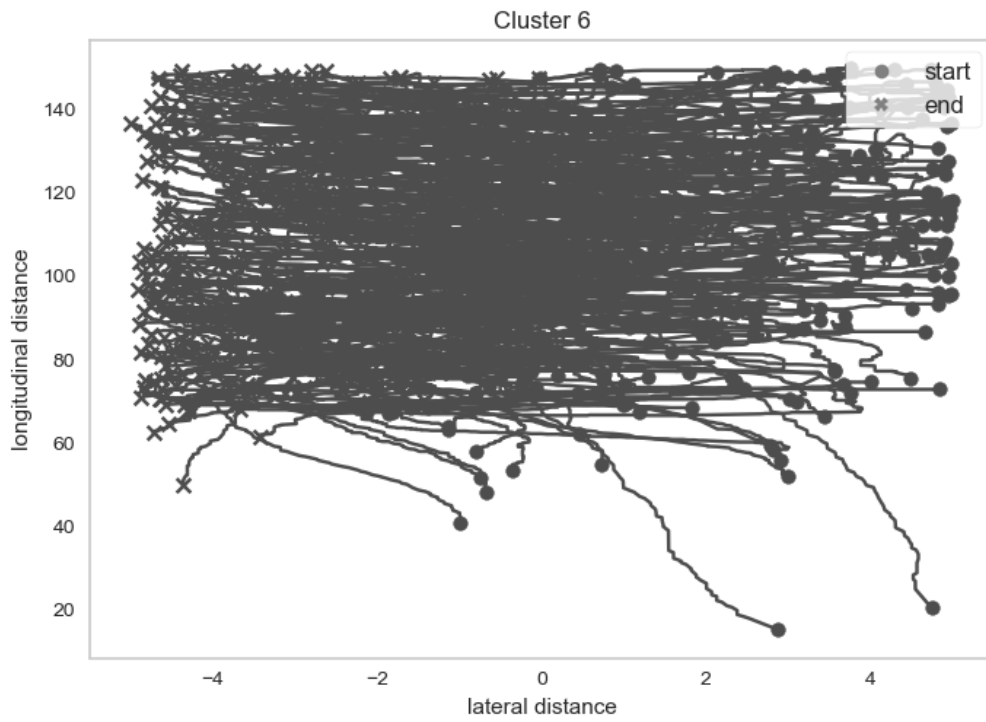
**Diagramm A.3:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 3 [Eigene Darstellung]



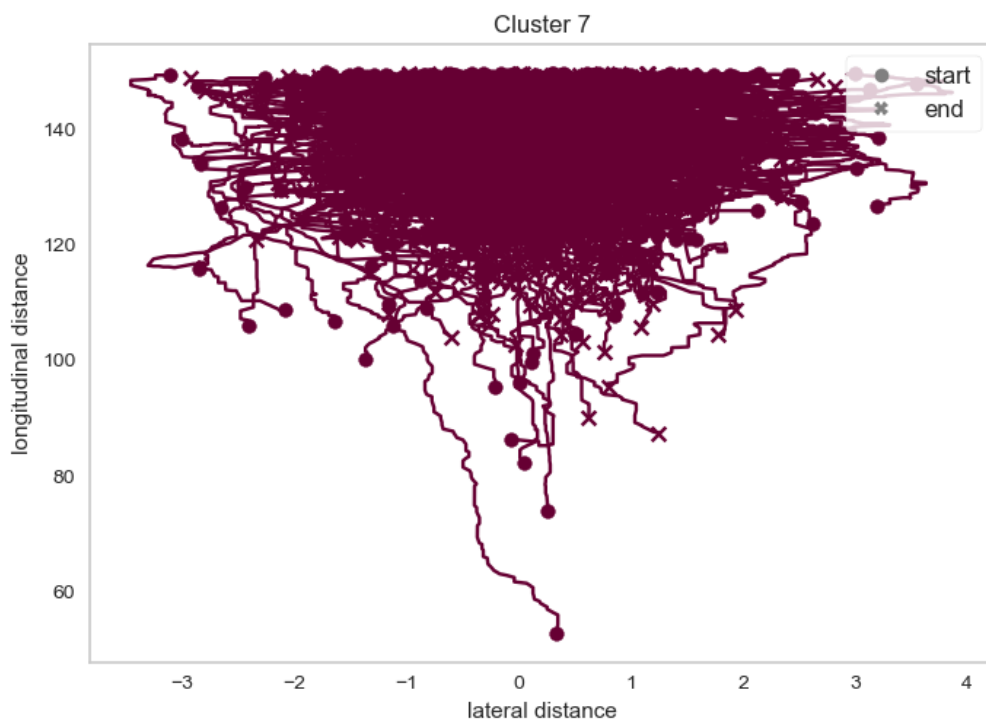
**Diagramm A.4:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 4 [Eigene Darstellung]



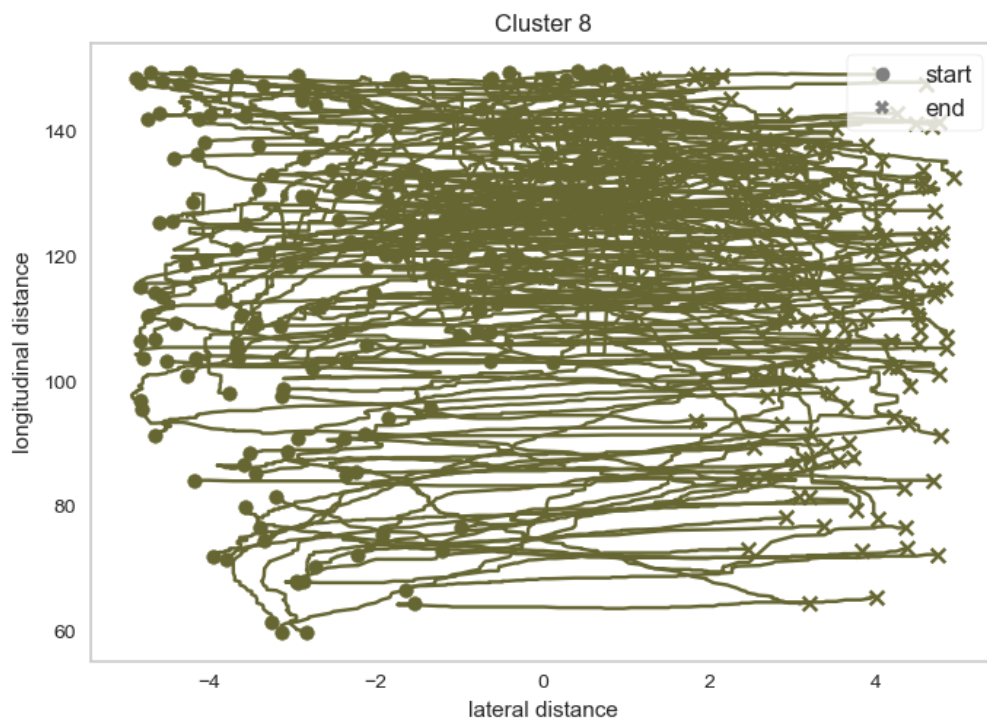
**Diagramm A.5:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 5 [Eigene Darstellung]



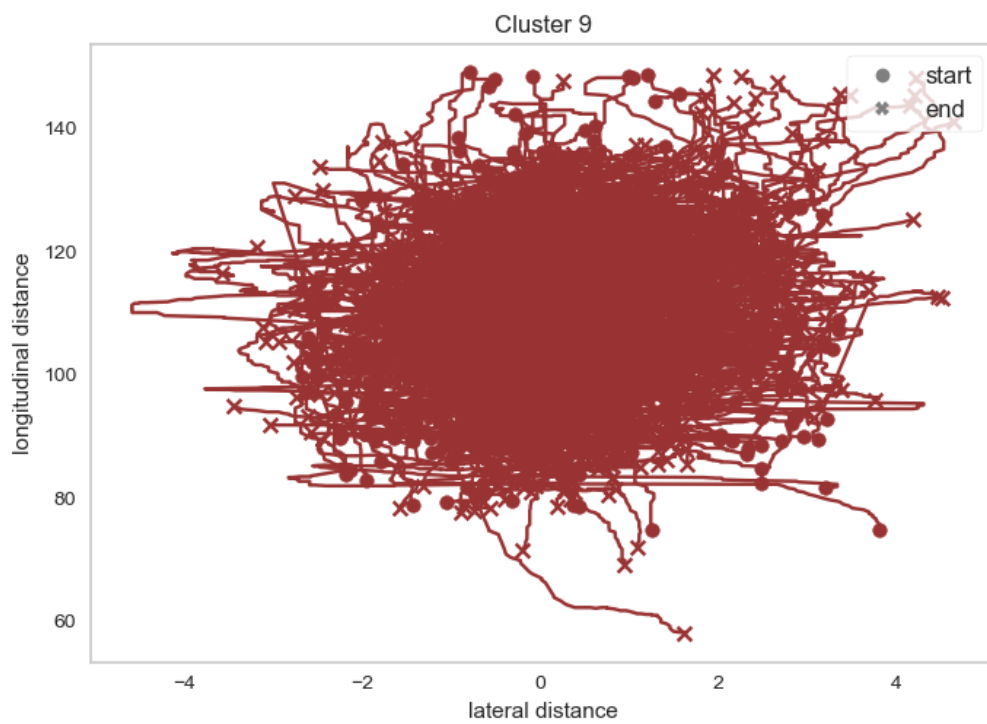
**Diagramm A.6:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 6 [Eigene Darstellung]



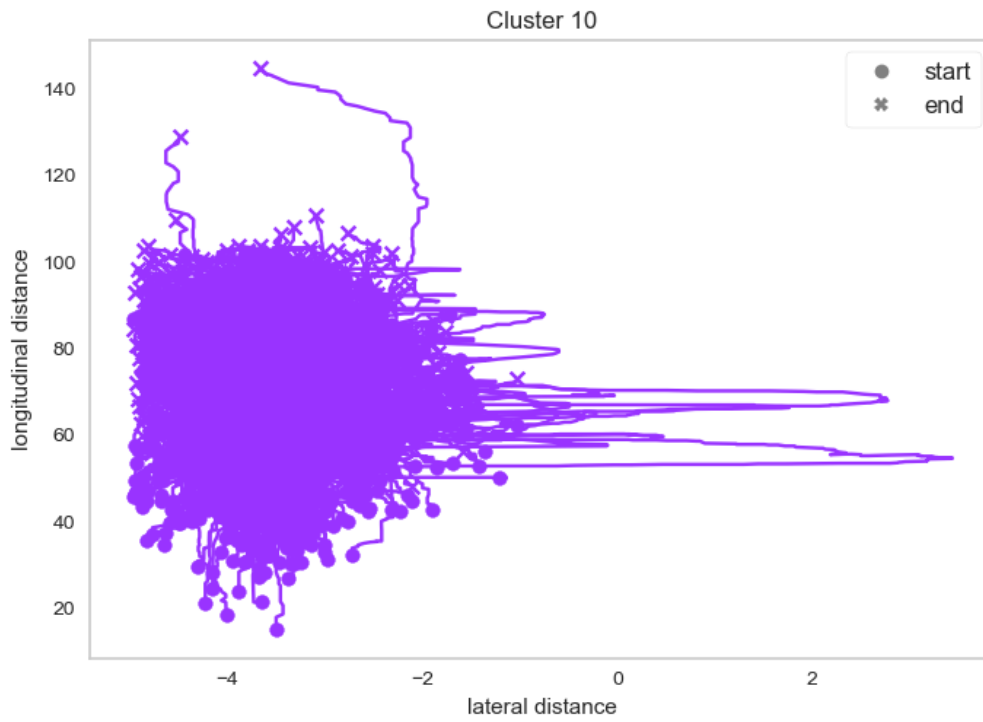
**Diagramm A.7:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 7 [Eigene Darstellung]



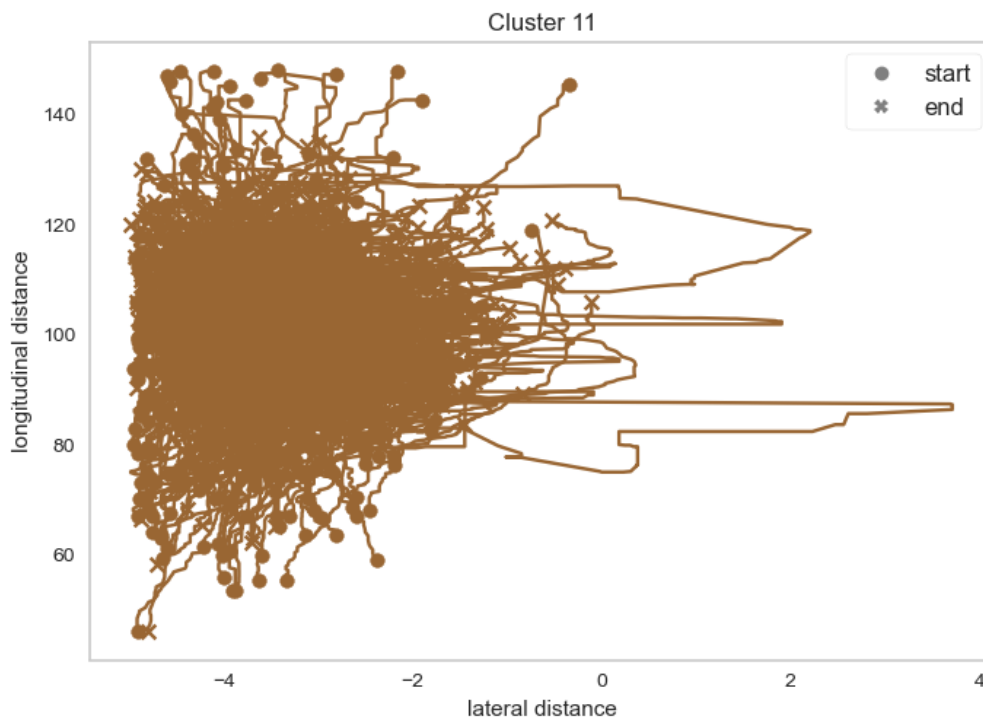
**Diagramm A.8:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 8 [Eigene Darstellung]



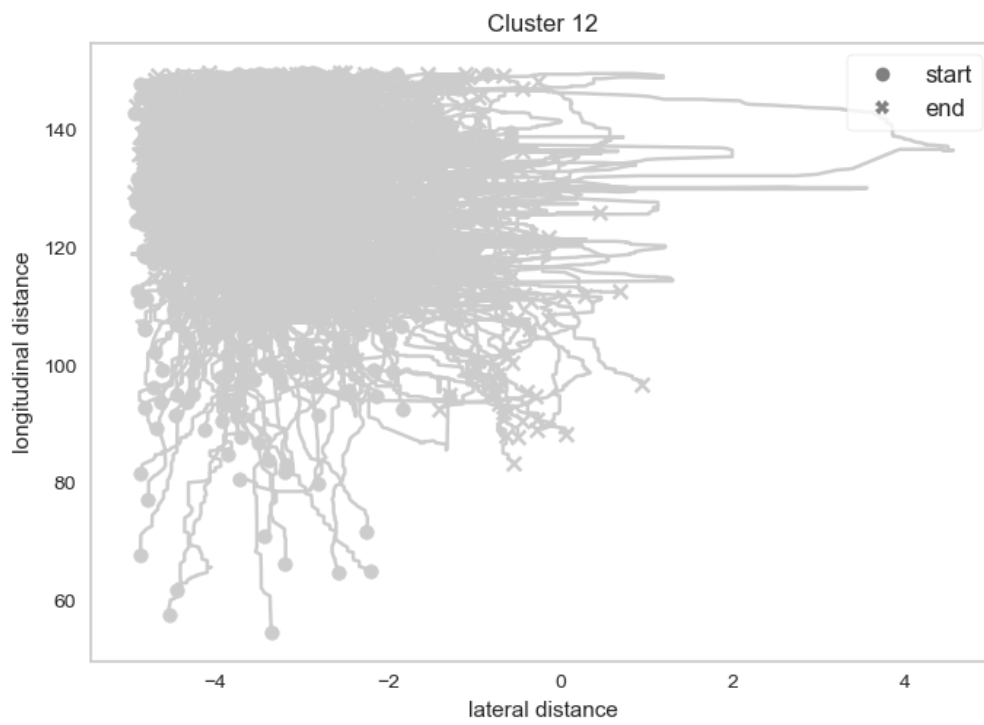
**Diagramm A.9:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 9 [Eigene Darstellung]



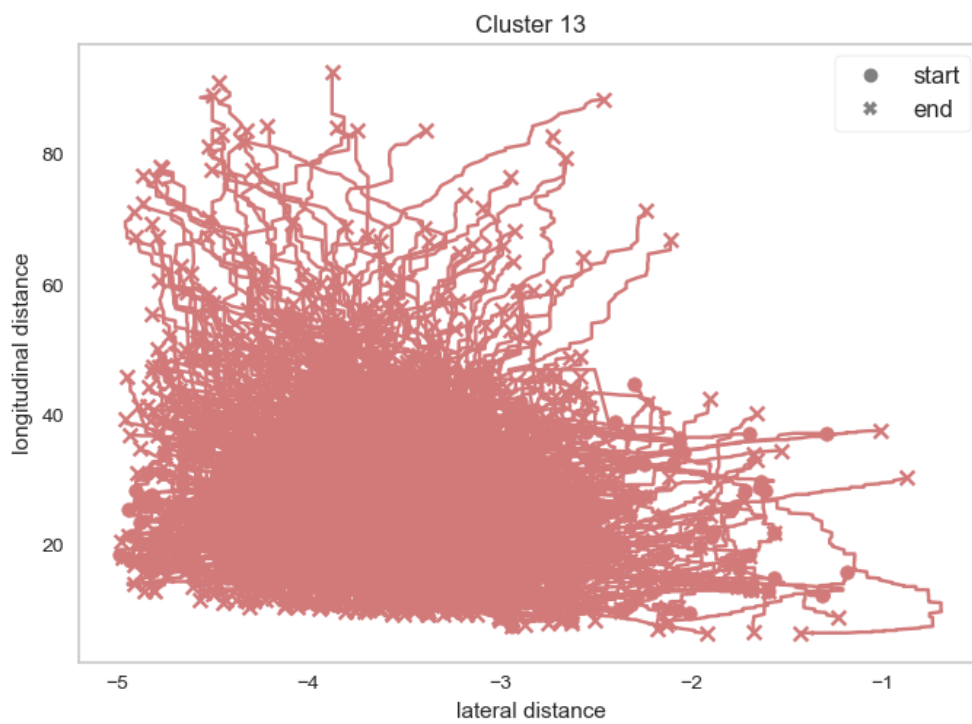
**Diagramm A.10:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 10 [Eigene Darstellung]



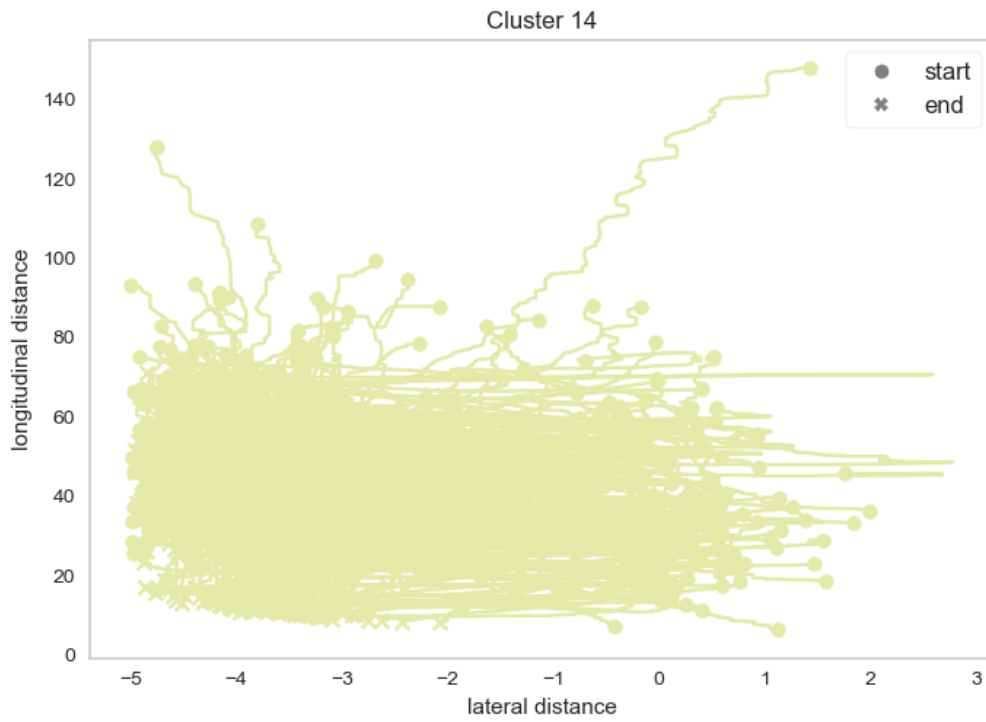
**Diagramm A.11:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 11 [Eigene Darstellung]



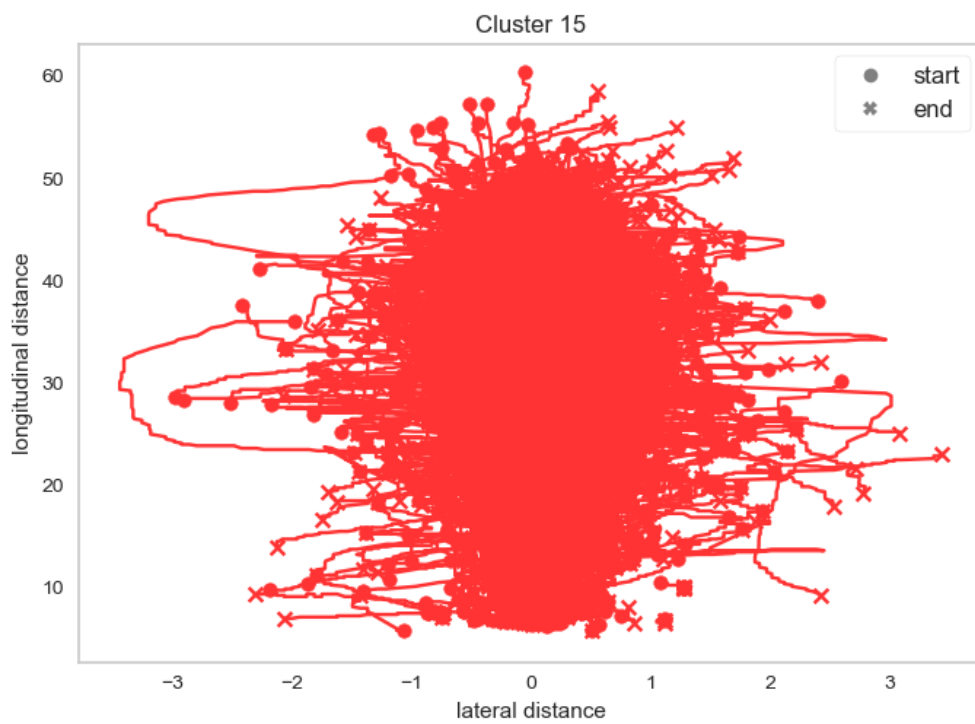
**Diagramm A.12:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 12 [Eigene Darstellung]



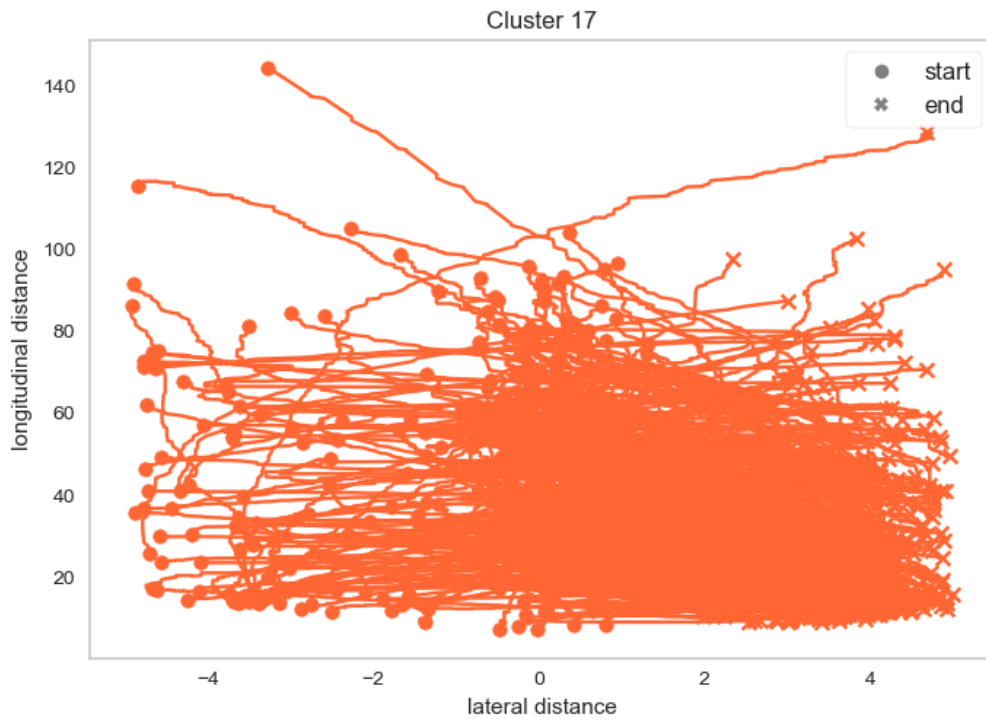
**Diagramm A.13:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 13 [Eigene Darstellung]



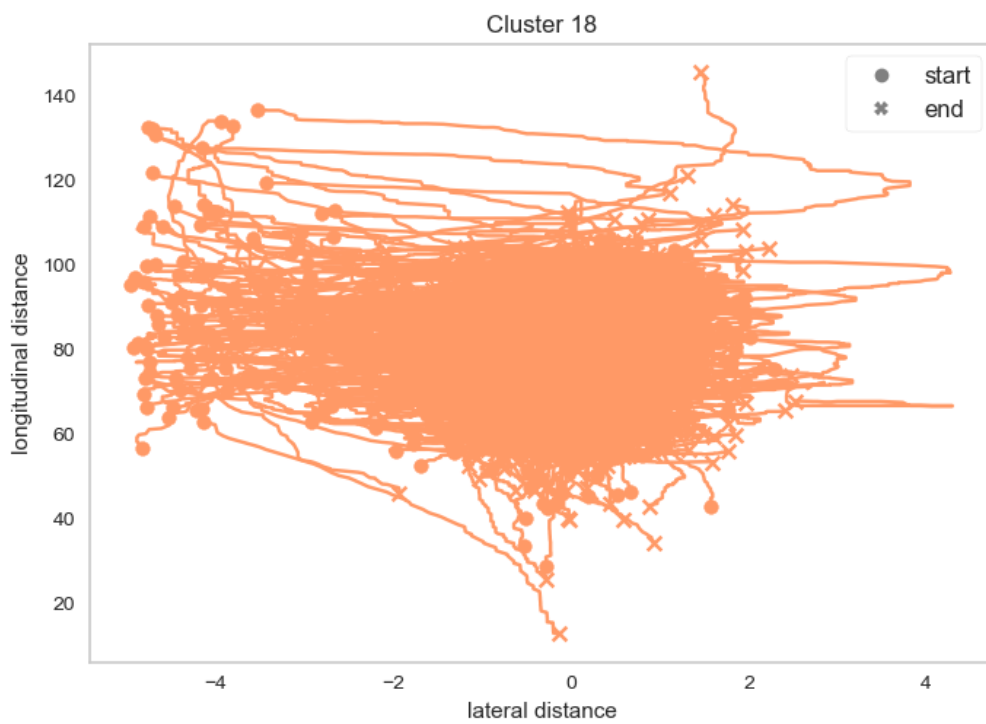
**Diagramm A.14:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 14 [Eigene Darstellung]



**Diagramm A.15:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 15 [Eigene Darstellung]



**Diagramm A.16:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 17 [Eigene Darstellung]



**Diagramm A.17:** Signale der Verkehrsteilnehmer-Kategorie aus Cluster 18 [Eigene Darstellung]



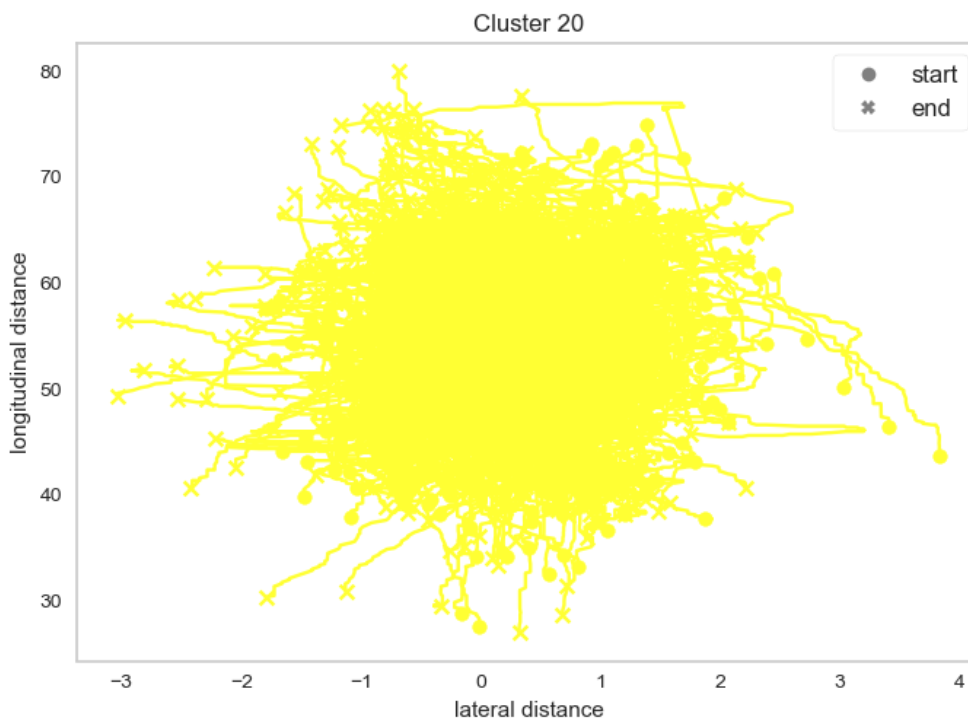


Diagramm A.18: Signale der Verkehrsteilnehmer-Kategorie aus Cluster 20 [Eigene Darstellung]

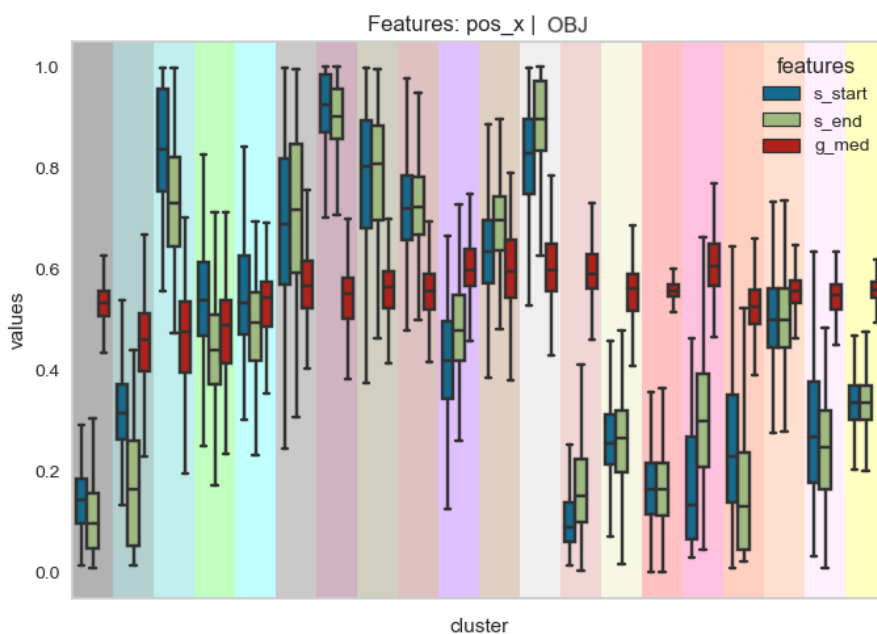
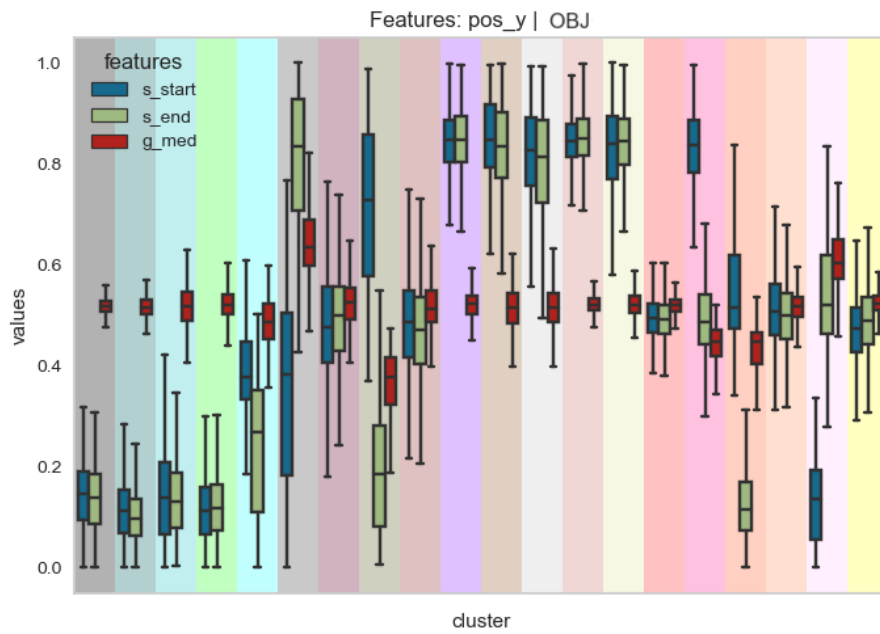
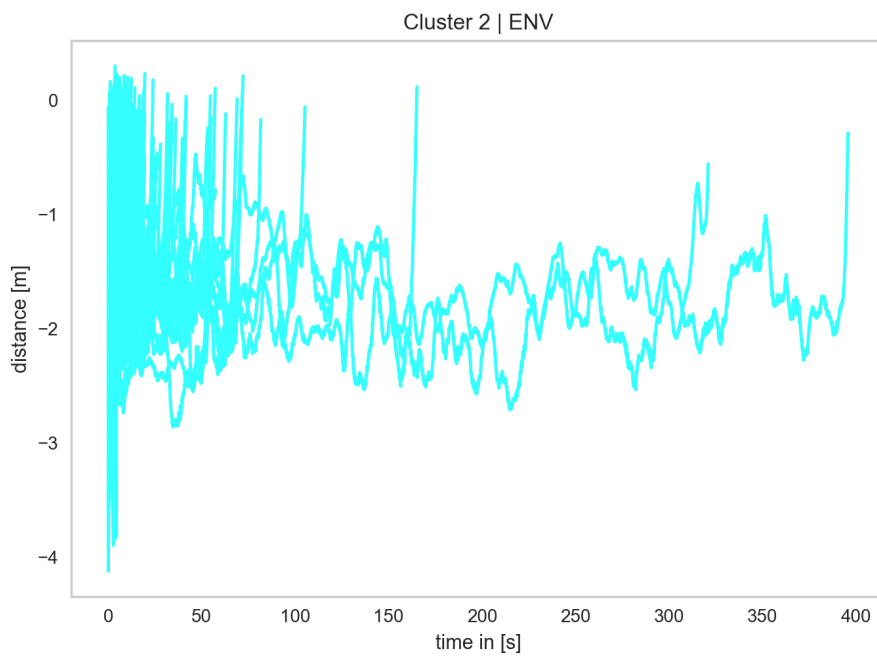


Diagramm A.19: Ein Boxplot mit Feature-Werten des Signals „Position x“ für alle Cluster der Verkehrsteilnehmer-Kategorie [Eigene Darstellung]

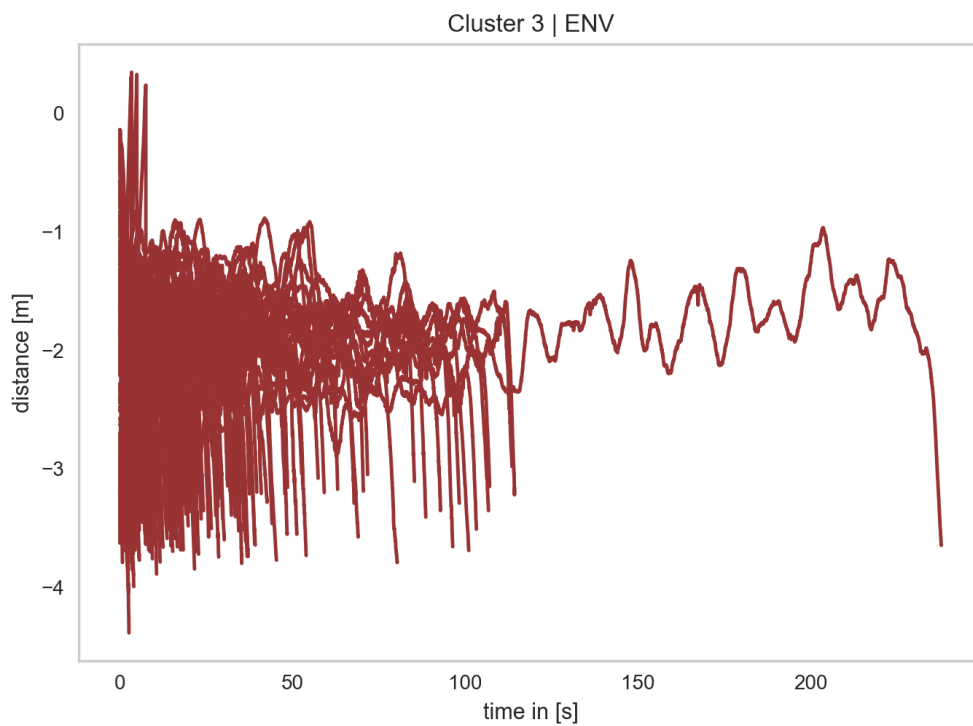


**Diagramm A.20:** Ein Boxplot mit Feature-Werten des Signals „Position y“ für alle Cluster der Verkehrsteilnehmer-Kategorie [Eigene Darstellung]

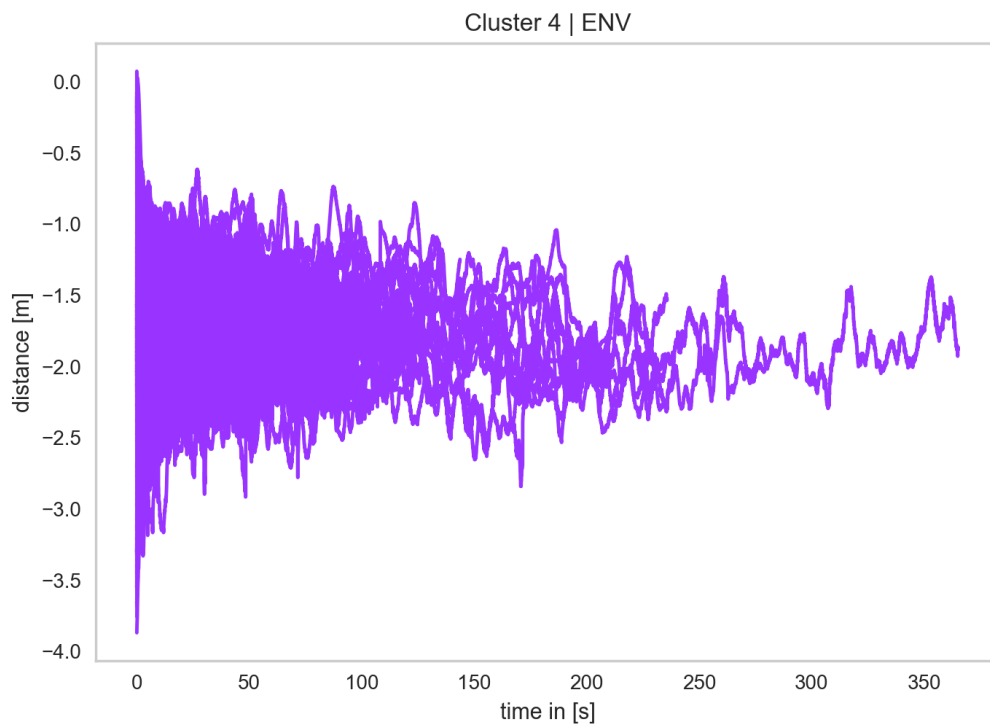
## Umwelt-Kategorie



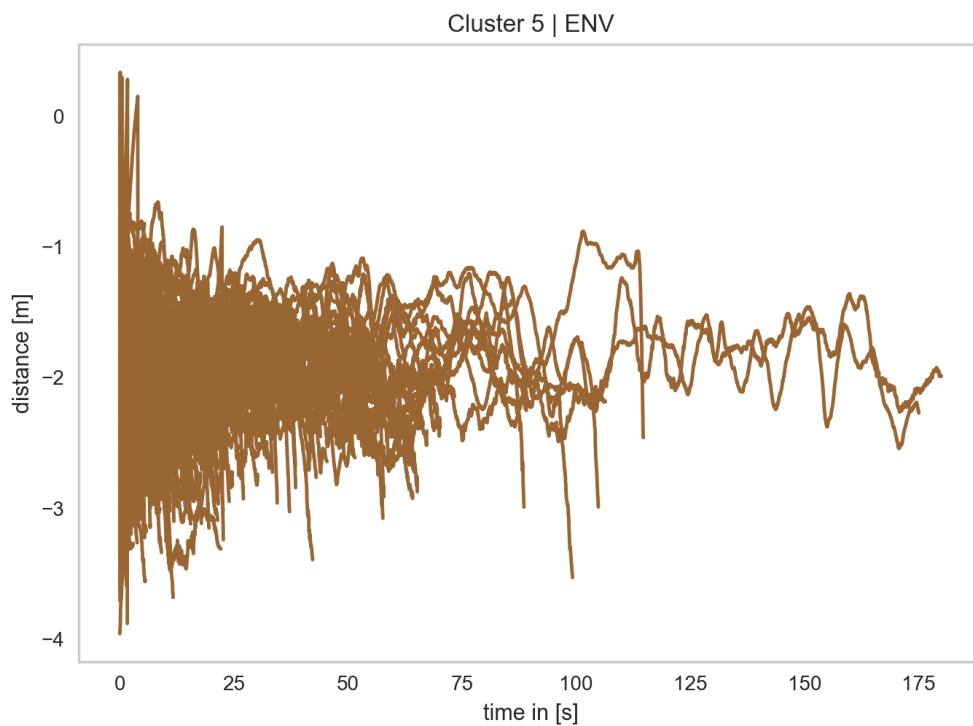
**Diagramm A.21:** Signale der Umwelt-Kategorie aus Cluster 2 [Eigene Darstellung]



**Diagramm A.22:** Signale der Umwelt-Kategorie aus Cluster 3 [Eigene Darstellung]



**Diagramm A.23:** Signale der Umwelt-Kategorie aus Cluster 4 [Eigene Darstellung]



**Diagramm A.24:** Signale der Umwelt-Kategorie aus Cluster 5 [Eigene Darstellung]

## CD-ROM

In der folgenden Ordnerstruktur befinden sich die Dateien:

### **Ausarbeitung**

- Diese Arbeit im PDF-Format.

### **Internetquellen**

- Eine Sammlung aller öffentlich verfügbaren und genutzten Internetdokumente im PDF-Format.

### **Auswertungsergebnisse**

- Eine Sammlung weiterer Auswertungsgrafiken für die Kategorien „Verkehrsteilnehmer, Umwelt und Ego mit den Qualitätsmaßen“ „Distortion, Galinski-Herabsz und Silhouette“.

# Literaturverzeichnis

- Amorim, Renato Cordeiro de (Apr. 2015). „Feature Relevance in Ward’s Hierarchical Clustering Using the L<sub>p</sub> Norm“. In: *Journal of Classification* 32.1, Seiten 46–62. ISSN: 0176-4268, 1432-1343. DOI: 10.1007/s00357-015-9167-1. URL: <http://link.springer.com/10.1007/s00357-015-9167-1> (zuletzt aufgerufen 22. 08. 2021).
- Artificial Intelligence Network Ingolstadt gGmbH (2021). *Forschungsschwerpunkte*. Ingolstadt. URL: <https://www.ainin.de/aktuelle-forschungsschwerpunkte/> (zuletzt aufgerufen 29. 07. 2021).
- Bagschik, Gerrit, Till Menzel und Markus Maurer (Juni 2018). „Ontology based Scene Creation for the Development of Automated Vehicles“. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. 2018 IEEE Intelligent Vehicles Symposium (IV). Changshu: IEEE, Seiten 1813–1820. ISBN: 978-1-5386-4452-2. DOI: 10.1109/IVS.2018.8500632. URL: <https://ieeexplore.ieee.org/document/8500632/> (zuletzt aufgerufen 24. 07. 2021).
- Beglerovic, Halil u. a. (12. Dez. 2018). „Deep Learning Applied to Scenario Classification for Lane-Keep-Assist Systems“. In: *Applied Sciences* 8.12, Seite 2590. ISSN: 2076-3417. DOI: 10.3390/app8122590. URL: <http://www.mdpi.com/2076-3417/8/12/2590> (zuletzt aufgerufen 04. 08. 2021).
- Bengler, Klaus, Herausgeber (2017). *11. Workshop Fahrerassistenzsysteme FAS 2017*. OCLC: 1129595953. ISBN: 978-3-00-055656-2.
- BMVI (2021). *SAVeNoW – Funktions- und Verkehrs-Sicherheit für Automatisierte und Vernetzte Mobilität – Nutzen für die Gesellschaft und oekologische Wirkung*. Berlin. URL: <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/AVF-projekte/savenow.html> (zuletzt aufgerufen 29. 07. 2021).
- Botsch, Michael und Wolfgang Utschick (2020). *Fahrzeugsicherheit und automatisiertes Fahren: Methoden der Signalverarbeitung und des maschinellen Lernens*. München: Hanser. 448 Seiten. ISBN: 978-3-446-45326-5.
- Defays, D. (1. Apr. 1977). „An efficient algorithm for a complete link method“. In: *The Computer Journal* 20.4, Seiten 364–366. ISSN: 0010-4620, 1460-2067. DOI: 10.1093/comjnl/20.4.364. URL: <https://academic.oup.com/comjnl/article-lookup/doi/10.1093/comjnl/20.4.364> (zuletzt aufgerufen 23. 08. 2021).
- Ebner, Adrian (13. Nov. 2014). „Referenzszenarien als Grundlage für die Entwicklung und Bewertung von Systemen der Aktiven Sicherheit“. In: Unter Mitarbeit von Technische Universität Berlin, Technische Universität Berlin und Werner Huber. Publisher: Technische Universität Berlin. DOI: 10.14279/DEPOSITONCE-4248. URL: <https://depositonce.tu-berlin.de/handle/11303/4545> (zuletzt aufgerufen 16. 08. 2021).
- Eckstein, Lutz, Herausgeber (2020). *13. Workshop Fahrerassistenz und automatisiertes Fahren FAS 2020*. OCLC: 1129595953. ISBN: 978-3-941543-47-8. URL: <https://www.uni-das.de/images/pdf/fas-workshop/2020/FAS2020-Tagungsband-150dpi.pdf> (zuletzt aufgerufen 28. 07. 2021).
- Ester, Martin und Jörg Sander (2000). *Knowledge discovery in databases: Techniken und Anwendungen*. Berlin Heidelberg: Springer. 281 Seiten. ISBN: 978-3-540-67328-6.
- Furtado, Andre Salvaro, Laercio Lima Pilla und Vania Bogorny (Mai 2018). „A branch and bound strategy for Fast Trajectory Similarity Measuring“. In: *Data & Knowledge Engineering* 115, Seiten 16–31. ISSN: 0169023X. DOI: 10.1016/j.datak.2018.01.003. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0169023X17302021> (zuletzt aufgerufen 22. 07. 2021).

- Gasser, Tom M., Herausgeber (2012). *Rechtsfolgen zunehmender Fahrzeugautomatisierung: gemeinsamer Schlussbericht der Projektgruppe*. Berichte der Bundesanstalt für Strassenwesen. F, Fahrzeugtechnik Heft F 83. Bremerhaven: Wirtschaftsverlag NW. 124 Seiten. ISBN: 978-3-86918-189-9.
- ISO/IEC (2018). *3.1.18: Information technology — Internet of Things (IoT) — Vocabulary — Entity*. URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:20924:ed-1:v1:en/> (zuletzt aufgerufen 22. 07. 2021).
- Kruber, Friedrich u. a. (Juni 2019). „Unsupervised and Supervised Learning with the Random Forest Algorithm for Traffic Scenario Clustering and Classification“. In: *2019 IEEE Intelligent Vehicles Symposium (IV)*. 2019 IEEE Intelligent Vehicles Symposium (IV). Paris, France: IEEE, Seiten 2463–2470. ISBN: 978-1-72810-560-4. DOI: 10.1109/IVS.2019.8813994. URL: <https://ieeexplore.ieee.org/document/8813994/> (zuletzt aufgerufen 04. 08. 2021).
- Luft- und Raumfahrt e. V. (DLR), Deutsches Zentrum für (28. Feb. 2020). *Schlussbericht für das Gesamtprojekt PEGASUS*. Braunschweig. URL: [https://www.pegasusprojekt.de/files/tmp/pdf/PEGASUS\\_Abschlussbericht\\_Gesamtprojekt.PDF](https://www.pegasusprojekt.de/files/tmp/pdf/PEGASUS_Abschlussbericht_Gesamtprojekt.PDF) (zuletzt aufgerufen 29. 07. 2021).
- Maurer, Markus u. a., Herausgeber (2015). *Autonomes Fahren*. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-662-45853-2. DOI: 10.1007/978-3-662-45854-9. URL: <http://link.springer.com/10.1007/978-3-662-45854-9> (zuletzt aufgerufen 08. 03. 2021).
- Menzel, Till, Gerrit Bagschik und Markus Maurer (Juni 2018). „Scenarios for Development, Test and Validation of Automated Vehicles“. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. 2018 IEEE Intelligent Vehicles Symposium (IV). Changshu: IEEE, Seiten 1821–1827. ISBN: 978-1-5386-4452-2. DOI: 10.1109/IVS.2018.8500406. URL: <https://ieeexplore.ieee.org/document/8500406/> (zuletzt aufgerufen 23. 07. 2021).
- Mockenhaupt, Andreas und Jürgen Rüttgers (2021). *Digitalisierung und Künstliche Intelligenz in der Produktion: Grundlagen und Anwendung*. Lehrbuch. Wiesbaden: Springer Vieweg. 347 Seiten. ISBN: 978-3-658-32772-9. DOI: 10.1007/978-3-658-32773-6.
- Montanari, F., R. German und A. Djanatljev (Okt. 2020). „Pattern Recognition for Driving Scenario Detection in Real Driving Data“. In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. 2020 IEEE Intelligent Vehicles Symposium (IV), Seiten 590–597. DOI: 10.1109/IV47402.2020.9304560.
- National Highway Traffic Safety Administration (NHTSA) (2021). *Automated Vehicles for Safety*. New Jersey Avenue. URL: <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety> (zuletzt aufgerufen 27. 07. 2021).
- Riedmaier, Stefan u. a. (2020). „Survey on Scenario-Based Safety Assessment of Automated Vehicles“. In: *IEEE Access* 8, Seiten 87456–87477. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2993730. URL: <https://ieeexplore.ieee.org/document/9090897/> (zuletzt aufgerufen 27. 07. 2021).
- SAE International (30. Apr. 2021). *J3016C: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International*. URL: [https://www-sae-org.thi.idm.oclc.org/standards/content/j3016\\_202104/](https://www-sae-org.thi.idm.oclc.org/standards/content/j3016_202104/) (zuletzt aufgerufen 22. 07. 2021).
- Schuldt, Fabian (Apr. 2017). „Ein Beitrag für den methodischen Test von automatisierten Fahrfunktionen mit Hilfe von virtuellen Umgebungen“. Dissertation. DOI: 10.24355/dbbs.084-201704241210. URL: [https://publikationsserver.tu-braunschweig.de/receive/dbbs\\_mods\\_00064747](https://publikationsserver.tu-braunschweig.de/receive/dbbs_mods_00064747).
- Sibson, R. (1. Jan. 1973). „SLINK: An optimally efficient algorithm for the single-link cluster method“. In: *The Computer Journal* 16.1, Seiten 30–34. ISSN: 0010-4620, 1460-2067. DOI: 10.1093/comjnl/

- 16.1.30. URL: <https://academic.oup.com/comjnl/article-lookup/doi/10.1093/comjnl/16.1.30> (zuletzt aufgerufen 23. 08. 2021).
- Steimle, Markus u. a. (2018). „Ein Beitrag zur Terminologie für den szenarienbasierten Testansatz automatisierter Fahrfunktionen“. In: *2018 AAET Automatisiertes und vernetztes Fahren*. Braunschweig. URL: [https://www.tu-braunschweig.de/fileadmin/Redaktionsgruppen/Institute\\_Fakultaet\\_5/IFR/Dateien\\_EFS/Publikationen/EinBeitragZurTerminologie.pdf](https://www.tu-braunschweig.de/fileadmin/Redaktionsgruppen/Institute_Fakultaet_5/IFR/Dateien_EFS/Publikationen/EinBeitragZurTerminologie.pdf) (zuletzt aufgerufen 18. 07. 2021).
- Technische Hochschule Ingolstadt (2021). *CARISSMA*. Ingolstadt. URL: <https://www.thi.de/forschung/carissma> (zuletzt aufgerufen 29. 07. 2021).
- The scikit-yb developers (2021). *Elbow Method*. URL: <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html> (zuletzt aufgerufen 27. 07. 2021).
- Ulbrich, Simon u. a. (Sep. 2015). „Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving“. In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. 2015 IEEE 18th International Conference on Intelligent Transportation Systems - (ITSC 2015). Gran Canaria, Spain: IEEE, Seiten 982–988. ISBN: 978-1-4673-6596-3. DOI: 10.1109/ITSC.2015.164. URL: <http://ieeexplore.ieee.org/document/7313256/> (zuletzt aufgerufen 18. 07. 2021).
- Ward, Joe H. (März 1963). „Hierarchical Grouping to Optimize an Objective Function“. In: *Journal of the American Statistical Association* 58.301, Seiten 236–244. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1963.10500845. URL: <http://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845> (zuletzt aufgerufen 23. 08. 2021).
- Weber, Hendrik u. a. (12. Juni 2019). „A framework for definition of logical scenarios for safety assurance of automated driving“. In: *Traffic Injury Prevention* 20 (sup1), S.65–S.70. ISSN: 1538-9588, 1538-957X. DOI: 10.1080/15389588.2019.1630827. URL: <https://www.tandfonline.com/doi/full/10.1080/15389588.2019.1630827> (zuletzt aufgerufen 04. 08. 2021).
- Winner, Hermann, Herausgeber (2015). *Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort*. 3., überarb. u. erg. Aufl. ATZ/MTZ-Fachbuch. Wiesbaden: Springer Vieweg. 1216 Seiten. ISBN: 978-3-658-05734-3. DOI: 10.1007/978-3-658-05734-3.



# Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit ohne fremde Hilfe angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Teile, die wörtlich oder sinngemäß einer Veröffentlichung entstammen, sind als solche erkennbar. Die Arbeit wurde noch nicht veröffentlicht und auch noch keiner anderen Prüfungsbehörde vorgelegt.

Ingolstadt, 31. August 2021

.....  
Springer Eduard