

Internationale statistische Mobilitätsanalyse mit Daten von Fahrzeugflotten und Drohnenaufzeichnungen

vorgelegt von
Oliver Stein

Vollständiger Abdruck der von der Fakultät für Bauingenieurwesen und Umweltwissenschaften der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

angenommenen Dissertation.

Gutachter:

1. Univ.-Prof. Dr.-Ing. Silja Hoffmann
2. Prof. Dr.-Ing. Thao Dang

Die Dissertation wurde am 28. Juli 2025 bei der Universität der Bundeswehr München eingereicht und durch die Fakultät für Bauingenieurwesen und Umweltwissenschaften am 3. Dezember 2025 angenommen. Die mündliche Prüfung fand am 10. Februar 2026 statt.

Danksagung

Ich möchte zunächst Professorin Silja Hoffmann herzlich für ihre Betreuung während meiner Doktorarbeit danken. Ihre fachliche Unterstützung hat maßgeblich zum Erfolg meiner Arbeit beigetragen. Die wertvollen Ratschläge und das tiefgehende Fachwissen von Professorin Hoffmann haben mir geholfen, die Herausforderungen meiner Forschung zu bewältigen und meine Ergebnisse stets auf einem hohen Niveau zu präsentieren. Des Weiteren möchte ich mich herzlich bei Professor Thao Dang bedanken, der mein Koreferat übernommen hat und mir während des gesamten Prozesses äußerst hilfreich zur Seite stand. Seine Fachkompetenz hat einen entscheidenden Beitrag zur Qualität meiner Arbeit geleistet und ich schätze sein Engagement und seine Zeit außerordentlich.

Meinen herzlichen Dank spreche ich meinen Kollegen bei Mercedes-Benz AG aus, die mich auf vielfältige Weise in den vergangenen Jahren unterstützt haben. Besonders bedanke ich mich bei Hubert Rehborn, der nicht nur bei der Erstellung meiner Arbeit Hilfe geleistet hat, sondern dessen Engagement und ständige Hilfsbereitschaft bei Fragen und Herausforderungen maßgeblich zum Gelingen dieser Arbeit beigetragen haben.

Ich möchte mich bei meiner Familie bedanken, die mir in dieser Zeit bei jedem Schritt zur Seite stand, immer an mich geglaubt und nie aufgegeben hat. Ebenso möchte ich meinen Freunden danken, die mich durch Höhen und Tiefen begleitet haben. Ohne ihre Unterstützung wäre diese Arbeit nicht möglich gewesen. Diese Arbeit widme ich voller Dankbarkeit meiner Familie und meinen Freunden, die nicht nur Zuschauer, sondern Mitgestalter meines Erfolgs waren.

Kurzzusammenfassung

Das übergeordnete Ziel dieser Arbeit ist ein besseres Verständnis des Fahrverhaltens auf globaler Ebene zu schaffen. Zu diesem Zweck werden weltweite Flottendaten analysiert, um einen Einblick in das Fahrverhalten einer Flotte zu erhalten. Die detaillierte Analyse der Verteilungen von Fahrtparametern gibt Aufschluss über das Fahrverhalten in verschiedenen Regionen. Die Variation der Fahrtparameter zurückgelegte Strecke, Reisezeit und Durchschnittsgeschwindigkeit in verschiedenen Regionen stand im Mittelpunkt der Untersuchung. Durch eine umfassende Analyse konnten nicht nur Unterschiede und Gemeinsamkeiten in den Verteilungen dieser Parameter identifiziert werden, sondern auch Gruppierungen der Länder nach Ähnlichkeiten in den Verteilungen gefunden werden. Diese Gruppierungen sollen Aufschluss darüber geben, welche Faktoren das Verkehrsverhalten in ähnlichen Regionen beeinflussen. Darüber hinaus ist die Frage nach geeigneten Verteilungsfunktionen für die Modellierung der Fahrparameter der Fahrzeugflotte Gegenstand der Untersuchung. Es wird untersucht, welche bereits bekannten Verteilungsfunktionen für zurückgelegte Strecke, Reisezeit und Durchschnittsgeschwindigkeit in Frage kommen und wie diese an real gemessene Daten angepasst werden können. Dieser Schritt ermöglicht nicht nur eine realitätsnahe Abbildung der Fahrtparameter, sondern eröffnet auch die Möglichkeit, Szenarien und Prognosen für die Zukunft zu erstellen.

Allerdings lassen sich aus diesen großflächigen Daten keine detaillierten Rückschlüsse auf lokale Verkehrsdynamiken ziehen, insbesondere auf spezifische Fahrmanöver wie das Einfädelverhalten auf Autobahnen oder Interaktionen zwischen Fahrzeugen. Um diese mikroskopischen Aspekte des Verkehrsflusses zu untersuchen, wurden zusätzlich Drohnenaufnahmen analysiert. Diese hochauflösenden Luftbilder erlauben eine präzise Erfassung von Fahrzeugbewegungen, Zeitlücken zwischen Fahrzeugen sowie Spurwechselstrategien und zeigen auf, wie einzelne Fahrzeuge sich in komplexen Verkehrssituationen verhalten. Diese werden im Detail analysiert, wobei ein besonderes Augenmerk auf mögliche Unterschiede in der Nutzungshäufigkeit der Spurwechselpositionen gelegt wurde. Die Ergebnisse zeigten, welche Bereiche der Autobahneinfahrt von den Fahrern bevorzugt zum Spurwechsel genutzt werden. Ein weiterer Schwerpunkt war die Analyse der Zeitlücken zwischen den Fahrzeugen auf der Einfädelspur während des Einfädelvorgangs. Es wird untersucht, wie diese

Zeitabstände im Bereich des Einfädelvorgangs variieren und welche Rolle sie für das Einfädelverhalten spielen. Die Ergebnisse zeigten die Verteilung der Zeitlücken während des Einfahrtvorgangs und ermöglichten Einblicke in deren Auswirkungen auf das Einfädelverhalten der Fahrzeuge.

Das Geschwindigkeitsverhalten der einfahrenden Fahrzeuge wurde untersucht, um den Einfluss des Einfädelns auf die Geschwindigkeitsprofile der Fahrzeuge auf der Hauptfahrbahn zu analysieren. Die Ergebnisse zeigen, dass sich die Geschwindigkeiten der bereits auf der Autobahn fahrenden Pkw durch die Einfädelvorgänge der neu hinzukommenden Fahrzeuge anpassen. Diese Anpassung äußert sich in einer kurzfristig erhöhten Varianz der Geschwindigkeit auf der Hauptfahrbahn, da sowohl Verzögerungen als auch Beschleunigungen auftreten, um den Einfädelvorgang zu ermöglichen. Insgesamt liefert die Analyse von Drohnendaten wichtige Erkenntnisse über das Einfädelverhalten, die Veränderung von Zeitlücken und die Geschwindigkeitsprofile bei Autobahneinfahrten.

Abstract

The goal of this research is a better understanding of driving behavior on a global scale. This thesis analyzes global fleet data to gain insight into the trip driving behavior of a fleet. The detailed analysis of the distributions of the trip parameters provides insight into the trip driving behavior in different regions. The focus of this work was the variation of the trip parameters distance traveled, travel time and average speed in different regions. Through a comprehensive analysis, not only differences and similarities in the distributions of these parameters could be identified, but also a grouping of countries according to similarities in the distributions could be found. These groupings are intended to give an indication of which factors influence driving behavior in similar regions. The study of traffic parameters in different countries, cities and on different days also provides insight into global traffic patterns and trends. Furthermore, the question of suitable distribution functions for modeling the trip parameters of the vehicle fleet is the subject of the study. It was investigated which distribution functions for distance, travel time and average speed are already known and how they can be adapted to real measured data. This step not only enables a realistic representation of the trip parameters, but also opens up the possibility of creating scenarios and forecasts for the future.

However, these large-scale data do not allow detailed conclusions to be drawn about local traffic dynamics, in particular specific driving maneuvers such as merging behavior on highways or interactions between vehicles. To investigate these microscopic aspects of traffic flow, drone images were also analyzed. These high-resolution aerial images allow precise recording of vehicle movements, time gaps between vehicles, and lane change strategies, and show how individual vehicles behave in complex traffic situations. These were analyzed in detail, with particular attention paid to possible differences in the frequency of use of lane change positions. The results showed which areas of the highway entrance are preferentially used by drivers to change lanes. Another focus was the analysis of time gaps between vehicles in the merging lane during the merging process. It was investigated how these time gaps vary in the area of the merging process and what role they play for the merging behavior. The findings revealed the distribution of time gaps during the merging process and offered insights into their impact on the merging behavior of vehicles.

The behavior of vehicles entering the main carriageway was analyzed to understand how merging affects speed profiles. Our findings indicate that the speeds of already-driving cars adapt to merging processes by newly arriving vehicles, causing a temporary increase in speed variance on the main carriageway. Both decelerations and accelerations due to merging occur, enabling this adaptation. Overall, the analysis of drone data provides important insights into merging behavior, time gaps, and speed profiles at freeway entrances.

Inhaltsverzeichnis

Abbildungsverzeichnis	xi
Tabellenverzeichnis	xii
Mathematische Notationen	xiii
1 Einleitung	1
1.1 Motivation	1
1.2 Aufgabenstellung und Leitfragen	2
1.3 Aufbau der Arbeit	3
2 Stand der Technik & Wissenschaft	5
2.1 Daten	5
2.1.1 Datenrepräsentation	5
2.1.2 Attribute	6
2.1.3 Datenbereinigung und -vorverarbeitung	7
2.2 Clustering	9
2.2.1 Ähnlichkeit	9
2.2.2 K-Means Clustering	10
2.2.3 Hierarchische Clusteranalyse	12
2.2.4 Bestimmung der Clusteranzahl	13
2.3 Methoden zur Erhebung von Verkehrsdaten	15
2.3.1 Detektordaten	15
2.3.2 Floating Car Data	16
2.3.3 Luftbeobachtungen	17
2.4 Fahrzeugstromdaten	19
2.4.1 Aggregierte Verkehrsanalysen	20
2.4.2 Individuelle Fahrzeugbewegungen	20
2.4.3 Dynamik von Fahrzeuggruppen	20
2.5 Straßennetze in Graphendarstellung	21
2.5.1 Graphennotation	21
2.5.2 Repräsentation von Straßennetzen als Graph	22

2.5.3	Zuordnung der Flottendaten zu Straßensegmenten (Map Matching)	24
2.6	Erhebung von Mobilitätsparameter	25
3	Analyse weltweiter FCD einer PKW-Flotte	27
3.1	Datensatz	27
3.1.1	Beschreibung des Datensatzes	28
3.1.2	Datenvorverarbeitung	31
3.1.3	Datenbereinigung	32
3.1.4	Map Matching	33
3.2	Internationale Fahrtparameter	35
3.2.1	Wahl der Fahrtparameter	35
3.2.2	Distanz	36
3.2.3	Fahrtdauer	38
3.2.4	Geschwindigkeit	41
3.2.5	Tagesganglinien	44
3.3	Vergleich zu Mobilitätsumfragen	47
3.4	Clustering von Verteilungen der Fahrtparameter	49
3.4.1	Ähnlichkeitsmetrik	49
3.4.2	Clustering von Ländern	51
3.4.3	Clustering von Städten	57
3.4.4	Clustering von Tagen	62
3.5	Darstellung gemessener Fahrtparameter einer Flotte durch Verteilungsfunktionen	63
3.5.1	Bewertung der Verteilungspassform	67
3.5.2	Distanz	68
3.5.3	Fahrtdauer	71
3.5.4	Geschwindigkeit	74
3.6	Zusammenfassung	77
4	Drohnenbeobachtungen	80
4.1	Datensatz	80
4.1.1	Beschreibung des Datensatzes	80
4.1.2	Betrachtung des einfahrenden freien Verkehrs im Drohnen- datensatz	83
4.2	Einfädelungsverhalten an einer Autobahneinfahrt	84
4.2.1	Wahl der Position für das Einfädeln	84
4.2.2	Zeitlücken auf den Fahrstreifen	85
4.2.3	Geschwindigkeitsverläufe der einzelnen Fahrzeuge	86
4.3	Zusammenfassung	88

5	Zusammenfassung und Ausblick	90
	Literaturverzeichnis	93
A	Anhang	105
A.1	Fahrtparameter verschiedener Städte	106
A.1.1	Randdaten der Fahrten	106
A.1.2	Empirische kumulative Verteilungsfunktionen	107
A.1.3	Betrachtung der Dendrogramme von Clustering der Städte- verteilungen	110

Abbildungsverzeichnis

2.1	Exemplarischer Datensatz mit verschiedenen Darstellungen	8
2.2	Mehrere Iterationen des K-Means Clusteringverfahrens	11
2.3	Iterationen der Verfahren des hierarchischen Clusterings	12
2.4	Übliche Verlinkungsverfahren beim hierarchischen Clustering	13
2.5	Exemplarischer Datensatz mit der Ellenbogen Methode	14
2.6	Doppelinduktionsschleifen an einem Autobahnabschnitt	16
2.7	FCD Datenerhebung und Verkehrsinformationsübermittlung.	17
2.8	Luftbeobachtung mithilfe einer Drohne an einem Autobahnabschnitt.	19
2.9	Darstellung des einfachen Graphen	21
2.10	Darstellung des allgemeinen Graphen	22
2.11	Darstellung des gerichteten Graphen	22
2.12	Straßennetzes als Graph	23
3.1	Heatmap der Positionsdaten im Datensatz	30
3.2	Anzahl von Fahrten je Land im Datensatz	33
3.3	Exemplarische räumliche Positionsdaten vor und nach dem Map- matching	34
3.4	Histogramm der Positions Korrektur durch Mapmatchingverfahren.	34
3.5	Histogramm der global gemessenen Distanzen der Flotte.	38
3.6	Empirische kumulative Verteilungsfunktion der gefahrenen Fahrtdi- stanzen pro untersuchtem Land.	39
3.7	Tagesverlauf der gefahrenen Distanzen in Deutschland.	40
3.8	Empirische kumulative Verteilungsfunktion der gemessenen Fahrt- dauern pro untersuchtem Land.	42
3.9	Tagesverlauf der gefahrenen Reisezeiten in Deutschland.	43
3.10	Verteilungsfunktion der gemessenen mittleren Geschwindigkeiten	44
3.11	Tagesverlauf der gefahrenen Durchschnittsgeschwindigkeit in Deutsch- land.	45
3.12	Durchschnittliche Fahrzeuganfragen der Flottenfahrzeuge	47
3.13	Heatmap der Unähnlichkeitsmatrix für Länder $S_{distance}$	52
3.14	Dendrogramm der Unähnlichkeitsmatrix für Länder $S_{distance}$	53
3.15	Heatmap der Unähnlichkeitsmatrix für Länder $S_{duration}$	54
3.16	Dendrogramm der Unähnlichkeitsmatrix für Länder $S_{duration}$	55

3.17	Heatmap der Unähnlichkeitsmatrix für Länder $S_{velocity}$.	56
3.18	Dendrogramm der Unähnlichkeitsmatrix für Länder $S_{velocity}$.	57
3.19	Dendrogramm der Unähnlichkeitsmatrix für Städte $S_{distance}$.	58
3.20	Dendrogramm der Fahrt dauern für Städte	60
3.21	Dendrogramm der Unähnlichkeitsmatrix für Städte $S_{duration}$.	61
3.22	Dendrogramm der Unähnlichkeitsmatrix für Städte $S_{velocity}$.	62
3.23	Heatmap der Unähnlichkeitsmatrix für die Geschwindigkeit an den Tagen der Messkampagne S_{days} .	64
3.24	Dendrogramm der Unähnlichkeitsmatrix für die Geschwindigkeit an den Tagen der Messkampagne S_{days} .	65
3.25	Empirical Cumulative Distribution Function (ECDF) Schaubild der verschiedenen gemessenen Tage für Stuttgart und den verschiedenen Städten.	66
3.26	Histogramm der gemessenen Distanzen	70
3.27	Histogramm der gemessenen Fahrt dauern	73
3.28	Histogramm der gemessenen Durchschnittsgeschwindigkeiten	76
4.1	Ausschnitt aus zwei der Aufnahmen mit verschiedenen Perspektiven	81
4.2	Von der Drohne erfasster Bereich der Autobahn A8.	82
4.3	Geschwindigkeit je Fahrstreifen über die Zeit der Aufnahme r_1 .	84
4.4	Verteilung der Einfädelpositionen von der Einfahrt. Links als Boxplot und rechts das dazugehörige Histogramm.	85
4.5	Nettozeitlücken in Abhängigkeit zur Position der Einfahrt.	86
4.6	Geschwindigkeitsverläufe der Fahrzeuge auf dem rechten Autobahnfahrstreifen.	87
4.7	Geschwindigkeitsverläufe der Fahrzeuge auf dem mittleren Autobahnfahrstreifen.	88
A.1	ECDF der exemplarischen Distanzen exemplarischer Städte.	107
A.2	ECDF der Fahrt dauern exemplarischer Städte.	108
A.3	ECDF der mittleren Geschwindigkeiten exemplarischer Städte.	108
A.4	Dendrogramm der Fahrt dauern und der sich dadurch resultierenden Cluster von Städten. In (a) mit Schnitt bei einer Unähnlichkeit $d = 0.8$ und (b) bei $d = 0.3$.	111
A.5	Dendrogramm der mittleren Geschwindigkeiten und der sich dadurch resultierenden Cluster von Städten. In (a) mit Schnitt bei einer Unähnlichkeit $d = 0.8$ und (b) bei $d = 0.4$.	113

Tabellenverzeichnis

3.1	Zeiträume der Messkampagne und Anzahl der erfassten Fahrten je Region.	29
3.2	Zusammenfassung der gefahrenen Distanzen pro Fahrt.	37
3.3	Zusammenfassung der Fahrtzeiten pro Fahrt.	40
3.4	Zusammenfassung der mittleren Fahrtgeschwindigkeit pro Fahrt. . .	43
3.5	Vergleich der ermittelten Werte aus der Literatur mit den Ergebnissen dieser Arbeit.	48
3.6	Zusammenfassung der verschiedenen Verteilungsfunktionen für die Distanz und deren Bewertung.	69
3.7	Zusammenfassung der verschiedenen Verteilungsfunktionen für die Fahrdauer und deren Bewertung.	72
3.8	Zusammenfassung der verschiedenen Verteilungsfunktionen für die Durchschnittsgeschwindigkeit und deren Bewertung.	75
4.1	Eigenschaften des Drohnendatensatzes an der A8.	83
A.1	Zusammenfassung der Fahrtparameter im Städtevergleich.	107

Mathematische Notationen

X	Merkmalsmatrix
x^j	Objekt der Merkmalsmatrix X
x_n^j	Attribut von Objekt x^j
$d(x^j, x^k)$	Distanz zwischen Objekten x^j und x^k
C_k	Cluster
μ_k	Centroid des Clusters C_k
s	Silhouettenkoeffizient
a	Intracuster Distanz
b	Extracuster Distanz
$\overline{v_{Ges}}$	mittlere Verkehrsgeschwindigkeit
q	Verkehrsfluss
ρ	Verkehrsdichte
v_x	longitudinale Geschwindigkeit
a_x	longitudinale Beschleunigung
a_y	laterale Beschleunigung
th	Nettozeitlücke
G	Graph
G_{simple}	schlichter Graph
$G_{general}$	allgemeiner Graph
$V(G)$	Menge an Knoten des Graphen G
$E(G)$	Menge an Kanten des Graphen G
D	Digraph
$A(D)$	Arc
p_t	Vektor mit räumlichen Koordinaten und Zeitstempel
Tr	Trajektorie bestehend aus p_t
MR_G	Sequenz aus Kanten e des Graphen G
seg_n	Segment (Teilmenge) von Tr

Abkürzungsverzeichnis

PKW Personenkraftwagen	6
LKW Lastkraftwagen	6
FCD Floating Car Data	20
OSM Open Street Map	23
LCSS Longest Common Subsequence	24
HMM Hidden Markov Model	25
CRF Conditional Random Field	25
WGT Weighted Graph Technique	25
ECDF Empirical Cumulative Distribution Function	xi

1 Einleitung

1.1 Motivation

In einer Welt, die von rasanten technologischen Fortschritten im Mobilitätssektor geprägt ist, wird ein tiefes Verständnis der menschlichen Mobilität zunehmend entscheidend. Dieses Wissen bildet nicht nur die Grundlage für die Optimierung unserer gegenwärtigen Fortbewegungsmöglichkeiten, sondern ebnet auch den Weg für die Entwicklung zukunftsweisender Mobilitätslösungen [1–3]. Doch wie gestaltet sich unsere Fortbewegung eigentlich? Welche Wege legt jeder von uns täglich zurück, und welche Zeiten und durchschnittlichen Geschwindigkeiten sind damit verbunden? Diese Fragen werfen nicht nur ein Licht auf unsere gegenwärtige Mobilität, sondern geben auch Hinweise darauf, wie sie verbessert werden können [4, 5].

Eine Möglichkeit, Informationen über das Mobilitätsverhalten der Bevölkerung zu erfassen, sind Mobilitätsbefragungen. Dabei werden Haushalte zu ihren täglichen Wegen, den genutzten Verkehrsmitteln sowie den Zwecken ihrer Fahrten befragt. Diese Methode liefert wertvolle Erkenntnisse über das Mobilitätsverhalten verschiedener Bevölkerungsgruppen, ist jedoch durch die Stichprobengröße sowie die subjektive Genauigkeit der Angaben begrenzt [6, 7]. Eine andere Herangehensweise zur Datenerhebung basiert auf stationären Verkehrszählungen, beispielsweise durch Induktionsschleifen oder andere Detektoren an Straßenquerschnitten. Diese erfassen den motorisierten Verkehr (Pkw, Lkw, Motorräder), liefern jedoch keine Informationen über nicht-motorisierte Verkehrsteilnehmer wie Fußgänger oder Radfahrer. Zudem sind diese Messungen auf den jeweiligen Installationsort beschränkt und ermöglichen keine umfassende Betrachtung individueller Mobilitätsmuster. Mit der Weiterentwicklung von Telekommunikationstechnologien wie 5G und der zunehmenden Vernetzung von Fahrzeugen und mobilen Endgeräten hat sich die Mobilitätsforschung in den letzten Jahren hin zu einer datengetriebenen Wissenschaft entwickelt. Durch die großflächige Erfassung von Floating Car Data (FCD) [5] sowie durch Sensorik in modernen Fahrzeugflotten ist es möglich, anonymisierte Bewegungsmuster über große Gebiete hinweg zu analysieren. Diese Ansätze ergänzen klassische Methoden der Mobilitätsforschung und ermöglichen eine differenziertere

Betrachtung von Mobilität und Verkehr, indem sie nicht nur Verkehrszahlen, sondern auch individuelle Verhaltensweisen im Straßenraum erfassen [3, 8–11]. Diese Erkenntnisse bergen das Potenzial, nicht nur die aktuelle Mobilität zu verbessern, sondern auch innovative Lösungen für die Zukunft zu gestalten, sei es im Bereich Carsharing, Mitfahrgelegenheiten oder der Elektromobilität [12–14].

Diese Arbeit nutzt diese Möglichkeiten, um ein globales Verständnis des Verkehrsverhaltens zu vertiefen. Die Analyse von Fahrtparametern auf internationaler Ebene, darunter zurückgelegte Entfernungen, Reisezeiten und Durchschnittsgeschwindigkeiten, eröffnet die Chance, Unterschiede und Gemeinsamkeiten in dem Nutzungsverhalten zu identifizieren. Diese Erkenntnisse legen nicht nur lokale Gegebenheiten offen, sondern tragen auch dazu bei, Maßnahmen zur Verbesserung der Verkehrssituation in ähnlichen Regionen zu entwickeln.

1.2 Aufgabenstellung und Leitfragen

Das übergeordnete Ziel dieser Forschungsarbeit ist es, das Verständnis des Verkehrsverhaltens auf globaler Ebene zu vertiefen, indem die Fahrtparameter zurückgelegte Entfernung, Reisezeit und Durchschnittsgeschwindigkeit einer Fahrzeugflotte analysiert und verglichen werden. Mobilität spielt eine entscheidende Rolle in modernen Gesellschaften und beeinflusst die Lebensqualität der Menschen sowie ökonomische und ökologische Aspekte [15–17]. Daher ist es wichtig, die Mobilitätsmuster in verschiedenen Ländern zu untersuchen und zu verstehen. Ein besonderer Schwerpunkt liegt dabei auf dem länderübergreifenden Vergleich von Fahrtparametern, um Unterschiede und Gemeinsamkeiten in den Verhaltensmustern zu identifizieren. Jedes Land hat spezifische Merkmale, die das Verkehrsverhalten beeinflussen können, wie z.B. Infrastruktur, wirtschaftliche Situation, Bevölkerungsdichte und kulturelle Gewohnheiten. Der Vergleich ermöglicht es, nationale und regionale Besonderheiten zu erkennen. Gleichzeitig können Gemeinsamkeiten aufgedeckt werden, die auf universelle Faktoren hinweisen, die das Verkehrsverhalten weltweit beeinflussen. Eine detaillierte Untersuchung dieser Verteilungen hilft, Erkenntnisse über das Verkehrsverhalten in den einzelnen Regionen zu gewinnen und mögliche Ausreißer oder Besonderheiten zu identifizieren. Ein weiteres Ziel der Arbeit ist es, mögliche Gruppierungen von Ländern aufgrund von Ähnlichkeiten oder Unterschieden in der Mobilitätsverteilung zu identifizieren. Eine solche Gruppierung kann wichtige Erkenntnisse darüber liefern, welche Faktoren das Verkehrsverhalten in ähnlichen Regionen beeinflussen und welche Unterschiede möglicherweise auf spezifische lokale Bedingungen zurückzuführen sind. Die Gruppierung erleichtert die Entwicklung gezielter Maßnahmen zur Verbesserung der Verkehrssituation, die den spezifischen

Bedürfnissen und Herausforderungen der jeweiligen Gruppe gerecht werden. Um das Fahrverhalten der untersuchten Fahrzeuge realitätsnah abbilden und für weitere Untersuchungen nutzen zu können, ist die Ermittlung einer geeigneten bekannten Verteilungsfunktion von großer Bedeutung. Diese Funktion ermöglicht es, das Verkehrsverhalten der Fahrzeugflotte in den verschiedenen Ländern abzubilden und somit Szenarien und Prognosen für die Zukunft zu erstellen. Eine genaue Modellierung eröffnet auch Möglichkeiten für den Einsatz von Simulationen und Prognosemodellen zur Optimierung des Verkehrsmanagements und zur Bewältigung zukünftiger Herausforderungen im Bereich der Mobilität. Daraus ergeben sich folgende Forschungsfragen:

Wie variieren die Fahrparameter gefahrene Distanz, Reisezeit und durchschnittliche Geschwindigkeit in verschiedenen Ländern?

Gibt es signifikante Ähnlichkeiten oder Unterschiede zwischen den Verteilungen der Fahrparameter in verschiedenen Ländern, und wie können diese gruppiert werden?

Welche bekannten Verteilungsfunktionen kommen für die Modellierung des Fahrverhaltens der Fahrzeugflotte in Frage und wie lassen sich diese an die analysierten Daten anpassen?

1.3 Aufbau der Arbeit

Die vorliegende Arbeit gliedert sich in mehrere Kapitel, die verschiedene Aspekte der Erfassung und Auswertung von Fahrzeugpositionsdaten auf individueller und aggregierter Ebene beleuchten.

Kapitel 2 gibt einen Überblick, welcher ein Verständnis für diese Arbeit schaffen soll. Dieses Kapitel behandelt die Grundlagen im Umgang mit realen Daten und die verschiedenen Repräsentationsformen. Zusätzlich werden verschiedene Methoden des Data Minings vorgestellt, mit denen es möglich ist eine messbare Ähnlichkeit in den Daten zu ermitteln.

Außerdem wird ein Überblick über den aktuellen Stand der Technik gegeben und der wissenschaftliche Kontext dieser Arbeit dargelegt. Dazu werden einschlägige Forschungsarbeiten präsentiert, deren Methoden und Ergebnisse zusammengefasst und eingeordnet. Anschließend wird erläutert, welcher spezifische Beitrag diese Arbeit zur bestehenden Forschung leistet und wie sie neue Erkenntnisse zur Mobilitätsanalyse liefert.

Kapitel 3 beschäftigt sich mit der Erklärung des Datensatzes, welche die Erhebung, die speziellen Eigenschaften und die Bereinigung der Daten umfasst. Außerdem behandelt dieses Kapitel die Analyse der zurückgelegten Distanz, der Reisezeit und den ermittelten mittleren Geschwindigkeit durch direkte Vergleiche von Verteilungen der einzelnen Länder im Datensatz sowie der Einordnung der Ergebnisse mit den Ergebnissen relevanter Literatur. Zudem wird durch die Verwendung der Wassersteindistanz eine messbare Ähnlichkeit der 20 in dem Datensatz häufigsten Länder errechnet und in Dendrogrammen untersucht. Diese Ähnlichkeit wird für die verschiedenen Fahrtparameter, als auch für Länder, Städte und Tage bestimmt. Am Ende dieses Kapitels wird für jede der gemessenen Metriken eine in der Literatur bekannte angepasste Funktion vorgestellt, welche dem Verhalten der Flotte für die verschiedenen Fahrtparametern am meisten ähnelt.

In Kapitel 4 wird eine mikroskopische Betrachtung des Verkehrsgeschehens präsentiert, die eine detaillierte Analyse des Einfädelverhaltens an Autobahneinfahrten ermöglicht. Während die vorhergehende makroskopische Untersuchung der Flottendaten globale Mobilitätsmuster, Fahrtparameter und regionale Unterschiede beleuchtet, erlauben die Drohnendaten eine hochauflösende Betrachtung individueller Fahrzeuginteraktionen in kritischen Verkehrssituationen.

Im Fokus dieser Analyse steht die Wahl des Einfädelpunkts also die Stelle, an der Fahrzeuge von der Einfädelspur auf die Hauptfahrbahn wechseln. Mithilfe der Drohnendaten wurden verschiedene Situationen analysiert und miteinander verglichen, um Muster und Dynamiken dieses Einfädelprozesses zu identifizieren.

Zusätzlich wurde untersucht, wie sich das Einfädeln auf die Geschwindigkeitsprofile der Fahrzeuge auf der Hauptfahrbahn auswirkt und ob es systematische Veränderungen in der Geschwindigkeitsverteilung gibt. Diese mikroskopische Perspektive ergänzt die makroskopische Analyse, indem sie detaillierte Einblicke in lokale Verkehrsdynamiken liefert, die aus aggregierten Flottendaten allein nicht ableitbar wären.

Abschließend werden in Kapitel 5 die wichtigsten Erkenntnisse zusammengefasst und ein Ausblick auf mögliche Entwicklungen und Potenziale gegeben.

2 Stand der Technik & Wissenschaft

Dieses Kapitel enthält die notwendigen Informationen für ein besseres Verständnis aller weiteren Kapitel. Es bietet einen Überblick über Datensätze, deren Erscheinungsformen, Repräsentation und Bereinigungsmethoden beim Arbeiten mit real gemessenen Datensätzen. Zunächst wird das übergeordnete Thema Data Mining vorgestellt, wobei insbesondere auf die Methoden des Clustering sowie der Klassifizierung eingegangen wird. Zusätzlich werden verschiedene Arten der Verkehrsbeobachtungen und deren Detailebenen vorgestellt. Außerdem wird die Repräsentation von Straßennetzen als mathematische Graphen inklusive der Zuordnung von räumlichen Koordinaten zu Graphenelementen aufgezeigt. Abschließend wird die Literatur der Themen Mobilitätsumfragen und Erkenntnisse aus verschiedenen Drohnenbeobachtungen behandelt.

2.1 Daten

Der folgende Abschnitt behandelt die mathematische Darstellung von Daten. Dabei werden verschiedene Arten von Attributen erläutert sowie Methoden zur Datenbereinigung und -transformation vorgestellt, die für eine sinnvolle Analyse erforderlich sind. Anschließend wird erklärt, wie geeignete Attribute für einen Datensatz ausgewählt werden.

2.1.1 Datenrepräsentation

In der heutigen Zeit gibt es eine Vielzahl von Daten unterschiedlichster Herkunft, die für Auswertungen und zur Informationsgewinnung genutzt werden können. Die Daten können in verschiedenen Formen vorliegen, z.B. als Datenbanken, Logdateien, Bilder oder Videoaufzeichnungen. Eine Form der Darstellung von Daten ist die Verwendung einer Merkmalsmatrix X . Eine exemplarische Darstellung einer

Merkmalsmatrix X ist in 2.1 [18, 19] gegeben.

$$X := \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_n^1 \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \dots & \dots & \dots & \dots \\ x_1^j & x_2^j & \dots & x_n^j \end{bmatrix} = \begin{bmatrix} x^1 \\ x^2 \\ \dots \\ x^j \end{bmatrix} \in \mathbb{R}^{j \times n}$$

mit $j \in \mathbb{N}$ ist die Anzahl der Beobachtungen,

$n \in \mathbb{N}$ ist die Anzahl der Merkmale (Features),

$x_k^i \in \mathbb{R}$ sind die Einträge der Matrix.

(2.1)

Jede Zeile der Matrix X repräsentiert ein Objekt beziehungsweise eine Beobachtung $x \in \mathbb{R}^{1 \times n}$ des Datensatzes mit n Attributen. Jede Spalte beschreibt ein Attribut der Objekte von X [18].

2.1.2 Attribute

Die Attribute von Objekten können in verschiedenen Formen auftreten und erfasst werden. Die Erscheinungsformen von Attributen können in 6 verschiedene Kategorien gruppiert werden [19, 20]:

Nominal: Nominale Attribute ordnen Objekte verschiedenen eindeutigen Klassen zu. Beispiele hierfür sind der Typ eines Objekts (Fußgänger, Personenkraftwagen (**PKW**), Lastkraftwagen (**LKW**), Baum, ...), die Farbe oder auch Herkunftsort eines Objekts.

Binär: Ein Sonderfall der nominalen Attribute mit nur zwei Möglichkeiten (ja/nein, 0/1, trifft zu/trifft nicht zu). Jedes nominale Attribut kann auch in ein binäres Attribut umgewandelt werden, indem nach dem Auftreten einer Kategorie gefragt wird (**PKW** oder nicht **PKW**).

Ordinal: Ordinale Attribute ordnen Objekte wie nominale Attribute verschiedenen Klassen zu, mit dem Unterschied, dass ordinale Attribute in einer sinnvollen Reihenfolge (Small, Medium, Large) angeordnet werden können. Es werden jedoch keine Abstände zwischen den verschiedenen Größen definiert.

Integer: Integre Attribute nehmen ganzzahlige Werte an und haben im Gegensatz zu nominalen Attributen eine arithmetische Bedeutung. Ein Beispiel für ein integriertes Attribut sind die Sitze in einem Flugzeug.

Intervallskala: Intervallskalierte Attribute können reelle Werte annehmen. Sie besitzen keinen absoluten Nullpunkt, der einen völligen Mangel der gemessenen Eigenschaft bedeuten würde. So bedeutet z. B. im Fall der Temperatur 0 Grad Celsius nicht, dass keine Temperatur vorhanden ist.

Verhältnisskala: Bei Attributen mit Verhältnisskala nehmen die Werte reelle Werte an, mit der Besonderheit, dass der Nullpunkt die Abwesenheit anzeigt.

In vielen Praxisbeispielen wird zwischen kategorialen Daten (nominal, binär, ordinal) und kontinuierlichen Daten (ganzzahlig, Intervallskala, Verhältnisskala) unterschieden. Es ist wichtig, die Daten in einer geeigneten Form zu erheben, da z.B. bei nominalen Attributen oft keine sinnvollen Distanzen ermittelt werden können [19].

2.1.3 Datenbereinigung und -vorverarbeitung

Datensätze aus realen Messungen können nicht als fehlerfrei angesehen werden. Die Datensätze enthalten häufig Einträge in falschen oder sogar gemischten Formaten, fehlende Einträge aufgrund von Ausfällen der Messgeräte, Einträge, die sich im Laufe der Zeit nicht ändern, oder Einträge, die außerhalb realistischer Grenzen liegen. Derartige Daten verfälschen die Auswertungen realer Datensätze und müssen daher im Vorfeld aussortiert werden [19, 21]. Jedes Attribut muss für weitere Auswertungen in der Maßeinheit konsistent sein. Bei Datensätzen, die aus mehreren verschiedenen Datensätzen zusammengefügt wurden, können die gleichen Attribute in unterschiedlichen Einheiten auftreten. Beispielsweise kann die Temperatur in °C und in °F angegeben sein, oder das Datum kann in TT/MM/JJJJ und MM/TT/JJJJ angegeben sein. In diesen Fällen sollte der Datensatz in eine einheitliche Einheit oder Form konvertiert werden. Ein Beispiel ist in Abbildung 2.1 dargestellt. Sie zeigt die Darstellung eines Datensatzes, in dem die Werte für die Attribute X und Y in verschiedenen Einheiten angegeben sind. In 2.1(a) sind die Werte für X in Metern und die Werte für Y in Zentimetern dargestellt. Die Punkte liegen dicht beieinander und deuten auf eine zusammenhängende Struktur hin. Erst in Abbildung 2.1(b) mit einer einheitlichen Dimension wird deutlich, dass drei verschiedene Strukturen in den Daten liegen. Häufig kommt es auch vor, dass Attribute im Datensatz teilweise fehlen, weil bei der Messung ein Fehler aufgetreten ist oder die Messgröße nur schwer zu erfassen ist. Unvollständige

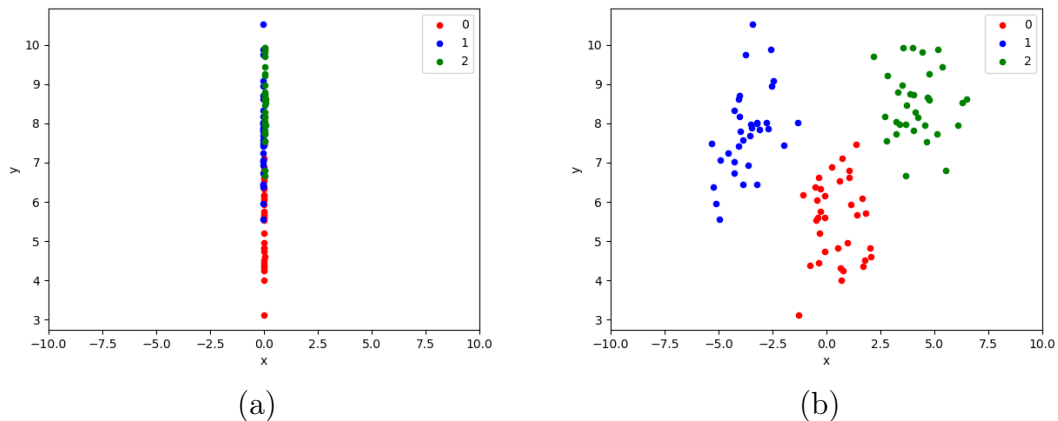


Abbildung 2.1: Exemplarischer Datensatz mit Werten für X und Y (a) X in Meter und Y in Centimeter (b) X und Y in Centimeter.

Attribute können nur dann sinnvoll weiterverwendet werden, wenn die fehlenden Datenreihen z.B. durch Mittelwert, Median oder Modus ergänzt werden. Eine Anpassung der Daten erhöht die Wahrscheinlichkeit irreführender Ergebnisse. Eine Alternative zum Umgang mit unvollständigen Attributen besteht darin, die Daten zu verwerfen und nicht für weitere Auswertungen zu verwenden. Eines der häufigsten Probleme bei realen Messungen sind verrauschte Daten. Diese Daten führen zu einem falschen Verständnis der Daten. Es gibt zwei Möglichkeiten, mit verrauschten Daten umzugehen [22]. Die Daten können durch Datenaufbereitung korrigiert werden, wobei verrauschte Werte durch geeignetere Werte ersetzt werden. Die zweite Möglichkeit, mit verrauschten Daten umzugehen, besteht darin, die Daten mit Hilfe von Filtern aus dem Datensatz zu entfernen. Attribute können sich über den gesamten Datensatz kaum oder gar nicht ändern und weisen daher eine geringe Varianz auf. Diese Attribute bieten bei der Auswertung kaum einen Mehrwert, da sie einen unbedeutenden Einfluss auf die Ergebnisse haben.

Bei großen Datenmengen ist es wichtig, nicht unbedingt alle Daten für die Analysen zu verwenden. Zu viele Daten zu verwenden ist im besten Fall ressourcenintensiv und im schlimmsten Fall kontraproduktiv, da dies dazu führen kann, dass anstelle relevanter Muster vor allem zufällige Schwankungen oder Rauschen gelernt werden und dadurch die Aussagekraft der Analyse verfälscht wird [19]. Es ist sinnvoll, Attribute vorab auszusortieren und den Datensatz so anzupassen, dass nur die für die Ergebnisse relevanten Attribute Verwendung finden.

2.2 Clustering

Clustering ist eine Unterkategorie des Data Mining. Das Clustern von Daten ist hilfreich, um Gruppen zu finden, ohne die Daten kennen zu müssen [23–26]. Das Hauptziel des Clustering ist es, die Daten in sogenannte Cluster zu gruppieren, so dass die Daten innerhalb eines Clusters einander ähnlich sind und gleichzeitig unähnlich zu den anderen gefundenen Clustern. Die Ähnlichkeit der Daten wird beim Clustering über eine Distanz- oder Ähnlichkeitsmetrik bestimmt und entsprechenden Gruppen zugeordnet [26, 27]. Clustering gehört zur Gruppe des unüberwachten Lernens. Das bedeutet, dass die Gruppen während des Clusterings gefunden werden und nicht von Anfang an im Datensatz besser bekannt waren.

2.2.1 Ähnlichkeit

Im Umgang mit Daten wird Ähnlichkeit üblicherweise als räumliche Nähe im n -dimensionalen Merkmalsraum aufgefasst. Das bedeutet, dass Objekte, die nahe beieinander liegen, sich ähnlicher sind als Objekte, die weiter voneinander entfernt sind [25]. Die Ähnlichkeit bzw. Distanz zwischen Objekten ist für das Clustering wesentlich [26, 28]. Wenn die Daten in Form einer Matrix vorliegen, wie in Gleichung 3.3, kann eine Distanz zwischen allen Objekten des Datensatzes bestimmt werden. Die Ähnlichkeit zwischen zwei Objekten $x^j, x^k \in \mathbb{R}^{1 \times n}$ ist die paarweise Distanz zwischen den Attributen von x^j und x^k [19]. Eines der bekanntesten Abstandsmaße ist der euklidische Abstand 2.2.

$$d_{euclidean}(x^j, x^k) = \sqrt{\sum_{i=1}^n x_i^j - x_i^k} \quad (2.2)$$

Die Wahl der Distanzmetrik hat einen großen Einfluss auf die Performance des Clusterings. Eine schlecht gewählte Metrik kann zu unzufriedenstellenden Resultaten führen. Darüber hinaus hängt die Wahl der Metrik auch von den Eingabedaten ab. Für Daten kategorischer Natur ist es beispielsweise sinnvoll, eine Übereinstimmungsrate zu wählen. Hierbei werden die kategorischen Attribute zweier Objekte miteinander verglichen. Je höher die Anzahl der Attribute mit unterschiedlichen

Kategorien, desto weiter liegen die Objekte auseinander.

$$d_{mismatch}(x^j, x^k) = \frac{p - m}{p} = 1 - \frac{m}{p} \in [0,1],$$

wobei m die Anzahl an Übereinstimmungen
und p die Anzahl an Attributen bezeichnet.

(2.3)

2.2.2 K-Means Clustering

Das K-Means Clusteringverfahren gruppiert Daten mithilfe der räumlichen Distanz. Hierfür werden k -Centroide gewählt, welche die zentralsten Punkte eines Clusters darstellen. Der Parameter k ist bei diesem Verfahren obligatorisch und muss im Vorfeld bekannt sein. Zunächst werden zufällig k -Centroide $\mu_1, \mu_2, \dots, \mu_k$ gewählt. Ausgehend von jedem einzelnen Datenpunkt werden die Abstände zu den verschiedenen Centroiden bestimmt, den naheliegendsten Centroiden und damit zu den Clustern $C_1, C_2, \dots, C_k$ zugeordnet [18, 29, 30]. Im Anschluss wird für jeden Cluster, mithilfe des Mittelwerts μ_j der Daten innerhalb des Clusters, der Centroid neu bestimmt. Die Centroide sind im Fall von K-Means keine Datenreihen und somit auch nicht Teil des Datensatzes, sondern nur im gleichen Zahlenraum. In jeder Iteration des K-Means Clusteringverfahrens wandert jeder Centroid zum Zentrum seines Clusters. Das Verfahren wählt seine Centroide so, dass die Inertia oder die Summe der Quadrate minimal ist:

$$\sum_{i=0}^n \min_{\mu_j \in C} (||x_i - \mu_j||^2)$$

n : Anzahl der Datenpunkte

x_i : Datenpunkt i im Merkmalsraum

C : Menge der Clusterzentren

μ_j : Zentrum (Mittelpunkt) des j -ten Clusters, $\mu_j \in C$

$||x_i - \mu_j||^2$: Quadratische euklidische Distanz zwischen x_i und μ_j

(2.4)

Das Verfahren ist beendet, wenn die Centroids gegen ein lokales Minimum konvergieren. Das Ergebnis dieses Verfahrens ist, dass alle Daten des Datensatzes einem der k -Cluster zugeordnet wurden [29, 31].

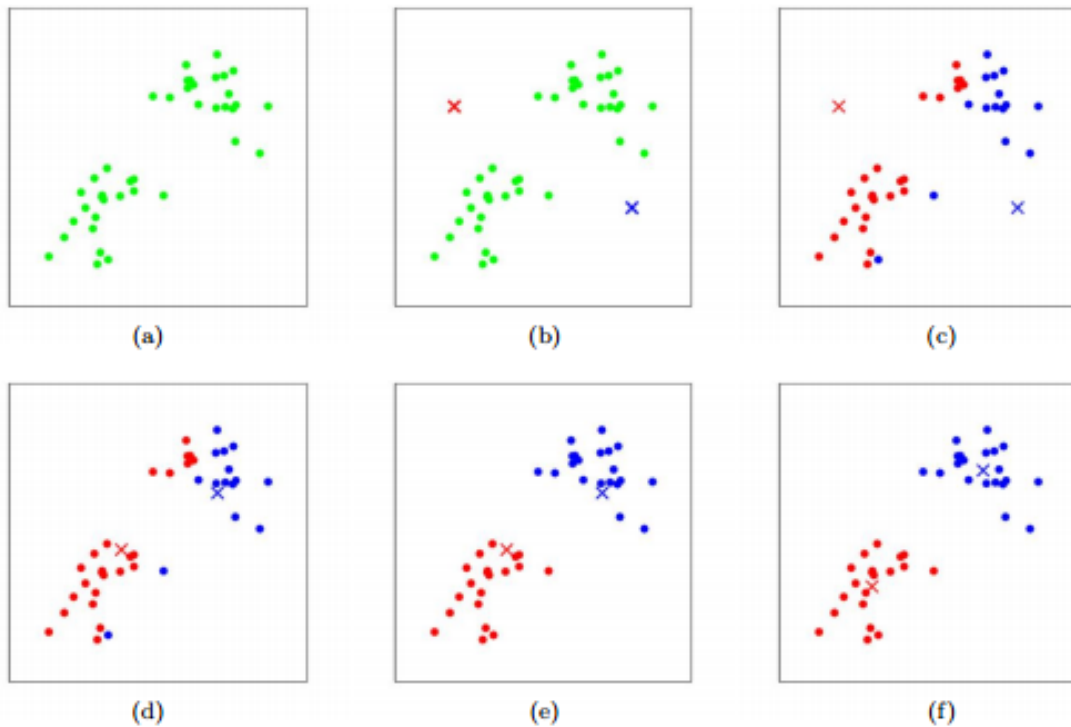


Abbildung 2.2: Mehrere Iterationen des K-Means Clusteringverfahrens [31].

In Abbildung 2.2 sind verschiedene Iterationen des K-Means Clustering zu sehen. In (a) ist der Datensatz in seiner unverarbeiteten Form zu sehen. Nach dem die Wahl für $k = 2$ Cluster a priori getroffen wurde, werden zwei zufällig gewählte Centroiden μ_1 und μ_2 in (b) gewählt. Von den zwei Centroiden ausgehend, werden in (c) die naheliegenden Punkten den zwei Centroiden zugeordnet. Die angenommenen Cluster sind in blau und rot gekennzeichnet. Nach dem jeder Punkt zugeordnet wurde, werden die Positionen für die Centroiden in (d) neu berechnet und die Centroiden wandern zum Zentrum der Cluster dieser Iteration. In (e) wird die Zuordnung der naheliegenden Punkte zu den neu berechneten Centroiden gezeigt. Die letzte Abbildung (f) zeigt die neue und letzte Position der Centroiden, da in jeder weiteren Iteration die Centroiden nicht mehr wandern würden.

Diese Methode des Clusterings liefert gute Ergebnisse bei Daten, die eine konvexe oder isotropische Form abbilden. Darüber hinaus skaliert das K-Means Clustering gut und bietet sich daher auch bei größeren Datenmengen an [29]. Allerdings hat das Verfahren Nachteile, da die Cluster aufgrund von Inertia gefunden werden. Werden die Centroiden $\mu_1, \mu_2, \dots, \mu_k$ in der ersten Iteration zu nah beieinander gewählt, kann das Ergebnis weit entfernt von dem gewünschten Clustern liegen.

Aus diesem Grund werden in diesem Verfahren die Centroide weit voneinander entfernt platziert.

2.2.3 Hierarchische Clusteranalyse

Die Methoden der hierarchische Clusteranalyse geben den Daten eine hierarchische Struktur. Diese hierarchische Aufgliederung wird als Baumstruktur, in Form eines Dendrogramms, abgebildet. Das Dendrogramm zeigt an oberster Stelle einen Cluster, der alle Objekte des Datensatzes enthält und an unterster Stelle einen eigenen Cluster für jedes einzelne Objekt in den Daten. Für den Umgang mit hierarchischen Clusteranalysen gibt es zwei Vorgehen: Das erste ist das *agglomerative Verfahren* (bottom-up) und das zweite ist das *divisive Verfahren* (top-down).

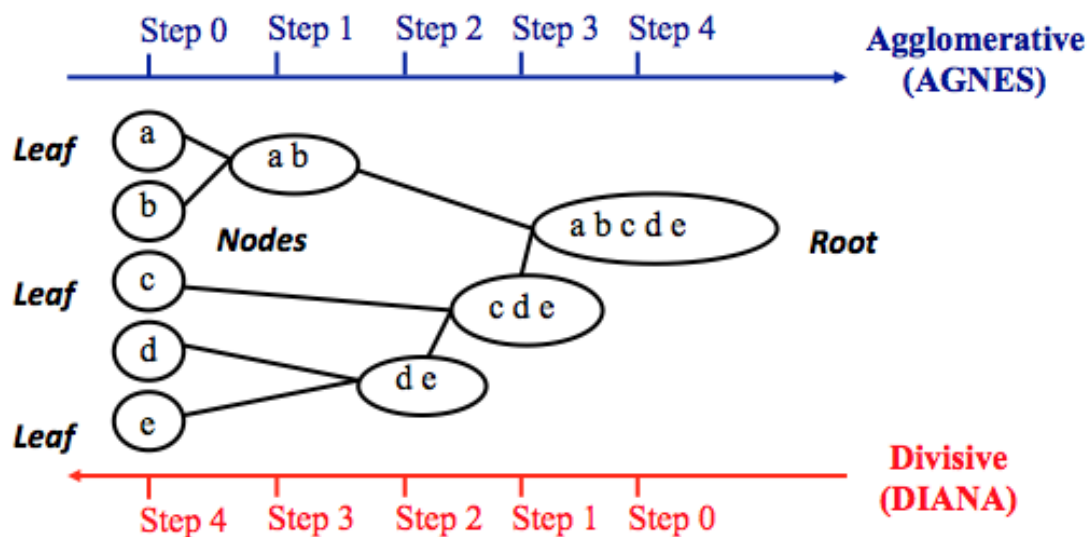


Abbildung 2.3: Iterationen der Verfahren des hierarchischen Clusterings [32].

Abbildung 2.3 zeigt in blau das agglomerative Verfahren. Bei diesem Verfahren wird zu Beginn für jedes Objekt in einem Datensatz ein eigener Cluster vorgesehen [33]. Im ersten Schritt werden die zwei räumlich nächsten Cluster in einen gemeinsamen Cluster eingruppiert. Für den entstandenen Cluster gibt es verschiedene Verlinkungsverfahren die Distanzen zu allen weiteren Clustern zu ermitteln. Üblicherweise werden die Verlinkungsverfahren single-link, complete-link oder average-link für die weitere Zuordnung verwendet [34]. Diese verschiedenen Verfahren und deren Funktionsweisen sind in Abbildung 2.4 aufgezeigt. Das single-link Verfahren verwendet die minimale Distanz zwischen zwei Clustern und wird häufig als Nearest-Neighbour

Clusteralgorithmus bezeichnet. Das complete-link Verfahren setzt anders als beim single-link Verfahren auf den weitesten Abstand zwischen zwei gefundenen Clustern. Dieses Verfahren wird häufig auch als Farthest-Neighbour Clusteralgorithmus bezeichnet. Beim average-link Verfahren wird ein imaginärer Mittelpunkt bestimmt, der sich aus der Intracuster Ähnlichkeit zusammensetzt. Durch diese Mittelpunkte wird der Abstand zu allen weiteren Clustern errechnet. Bei allen drei Verfahren wird trotz der unterschiedlichen Berechnung für die Distanzen der Cluster für eine Zusammenführung von zwei Clustern immer der kürzeste Abstand gewählt. In jedem Schritt werden die zwei räumlich naheliegendsten Cluster zu einem größeren zusammengefasst bis am Ende des Verfahrens ein Cluster entsteht, der alle Objekte in dem Datensatz beinhaltet. Das divisive Verfahren ist in Abbildung 2.3 in rot dargestellt. Es funktioniert genau umgekehrt zu dem agglomerativen Verfahren. Zu Beginn ist hier ein Cluster der alle Objekte enthält. In jeder weiteren Iteration werden die zwei weitest entfernten Objekte in neue Cluster unterteilt bis zum Schluss für jedes Objekt ein individueller Cluster entstanden ist [34].

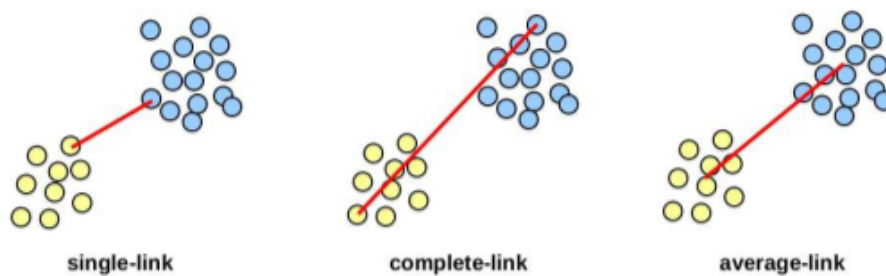


Abbildung 2.4: Übliche Verlinkungsverfahren beim hierarchischen Clustering [32].

2.2.4 Bestimmung der Clusteranzahl

Nachdem die Clusteringmethoden auf den Datensatz angewendet wurden, ist noch unklar, wie das Clustering performt hat und wie valide die geformten Cluster sind. In diesen Fällen können Metriken berechnet werden, welche die Qualität des Clusterings messen. Die bekanntesten Methoden zur Bestimmung der Clusteranzahl sind der Silhouetten Koeffizient und die Ellenbogen Methode (engl. Ellbow Method).

Eine dieser Metriken ist der Silhouetten Koeffizient s siehe 2.5. Dieser bestimmt die Abstände innerhalb eines Clusters im Verhältnis zu den Abständen von Cluster zu Cluster. Im Idealfall sind die Clusterkohäsion (Intracuster Distanz) a gering

und die Clusterseparation (Extracuster Distanz) b groß [18, 22, 29].

$$s := \frac{b - a}{\max(b, a)} \in [-1, 1]$$

mit $b > 0$ und $a \neq 0$

(2.5)

Um den Silhouetten Koeffizienten s zu berechnen ist es notwendig eine Nummer von Clustern $k > 1$ zu wählen. Somit ist die Bedingung erfüllt, dass die Clusterseparation b größer als Null ist. Der Wertebereich des Silhouetten Koeffizienten s geht von $[-1, 1]$. Ein Silhouetten Koeffizient von $s = 0$ bedeutet nach 2.5, dass $a = b$ und somit die Clusterkohäsion und -separation gleich sind. Das kann darauf hindeuten, dass es bei den Clustern eine Überlappung gibt. Ein negativer Silhouetten Koeffizient s deutet darauf hin, dass die Objekte falsch geclustert wurden und in einen anderen Cluster gehören, oder eine andere Anzahl der Cluster gewählt werden sollte. Ein positiver Silhouetten Koeffizient zeigt an, dass die Abstände innerhalb eines Clusters kleiner sind, als die Abstände zu anderen Clustern und damit vermutlich korrekt geclustert worden sind.

Eine weitere Methode zur Bestimmung der Clusteranzahl ist die Ellenbogen Methode. Die Ellenbogen Methode ist eines der ältesten Verfahren für das Bestimmen der Clusteranzahl. Abbildung 2.5 zeigt die Anwendung der Ellenbogen Methode an einem exemplarischen Datensatz.

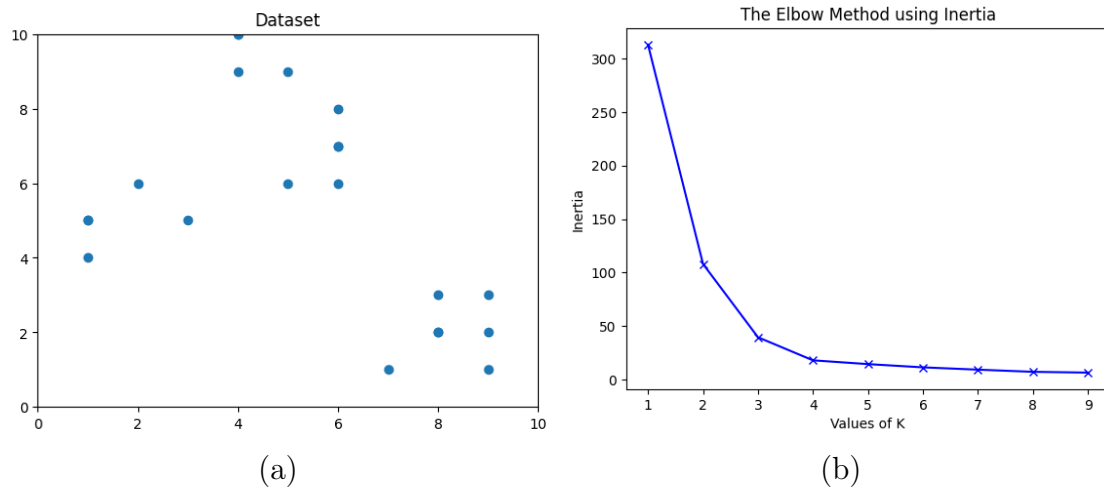


Abbildung 2.5: Exemplarische Anwendung der Ellenbogen Methode. (a) Exemplarischer Datensatz mit Datenpunkten für X und Y (b) Das dazugehörige Schaubild der Ellenbogen Methode (aus [35]) angepasst.

Dabei wird zunächst von einer Clusterzahl von $K = 2$ ausgegangen und schrittweise um einen Cluster erweitert. Zur Berechnung jeder Iteration der Ellenbogen-Methode wird üblicherweise die Summe der quadrierten Abweichungen von einem hypothetischen Mittelwert oder dem tatsächlich zentralsten Datenpunkt des Clusters herangezogen. Diese Summen werden in einem Diagramm dargestellt (siehe 2.5(b)), wobei die Anzahl der Cluster auf der X-Achse und typischerweise die Summe der quadrierten Abweichungen auf der Y-Achse aufgetragen wird. Abbildung 2.5(b) beginnt aus Anschaulichkeit bei $K = 1$, womit aber kein Clustering durchgeführt werden würde sondern der gesamte Datensatz berücksichtigt wird. Die Elbow-Methode basiert auf der Beobachtung, dass die Summe der quadrierten Abweichungen in der Regel mit der Anzahl der Cluster abnimmt. Der entscheidende Punkt in diesem Diagramm, an dem die Abnahme der Abweichung abflacht und die Kurve abflacht, wird als Ellbogenpunkt bezeichnet. Dieser Punkt stellt die optimale Anzahl von Clustern dar, da er den besten Kompromiss zwischen der Modellkomplexität und der Qualität der Clusterbildung darstellt. Die Ellenbogenmethode unterstützt bei der Auswahl der geeigneten Clusteranzahl für die Datenanalyse [35, 36].

2.3 Methoden zur Erhebung von Verkehrsdaten

Ein wichtiger Bestandteil von Verkehrssystemen ist die Messung der aktuellen Verkehrszustände in einem Straßennetz. Nicht nur für Echtzeitanwendungen (z. B. *dynamische Ampelschaltungen, Navigationssysteme mit Echtzeit-Verkehrsinformationen*) sind aktuelle Verkehrszustandsmessungen unverzichtbar, auch für die Optimierung von Straßennetzen werden umfangreiche Verkehrsdaten benötigt. Zur Erfassung der Daten werden unterschiedliche Sensortechnologien eingesetzt. In diesem Abschnitt werden verschiedene Sensortechnologien vorgestellt.

2.3.1 Detektordaten

Häufig verwendete Sensoren sind sogenannte Detektoren, die in zahlreichen Ausführungen vorhanden sind. Darunter gibt es Radar-, Video, Ultraschall-, Wärmebild- und Magnetfelddetektoren. Die am weitest verbreitetste Variante ist allerdings der Induktionsschleifendetektor. Diese fest im Boden verlegten Drahtschleifen erzeugen ein Magnetfeld, welches durch Überfahrt eines Fahrzeugs mit metallischen Elementen gestört wird [37]. In Abbildung 2.6 wird die Erfassung von Fahrzeugen mit Induktionsschleifen verdeutlicht. Dabei sind in jedem Fahrstreifen zwei Induktionsschleifen hintereinander installiert worden. In diesem Fall handelt es sich auch um sogenannte Doppelinduktionsschleifen, welche zusätzlich zu der bloßen Anwesenheit

eines Fahrzeugs und den Zeitlücken zwischen Fahrzeugen ein breiteres Spektrum an Informationen erfassen können. Diese messen im Übrigen auch die Geschwindigkeit, die Länge und Fahrzeugklasse von Fahrzeugen. Doppelinduktionsschleifen finden hauptsächlich außerorts ihre Verwendung, während Einzelinduktionsschleifen in der Regel innerorts ausreichend sind. Die Auswertung der Schleifensignale findet in der Regel über einen Mikrocontroller statt. Erfasste Daten werden in einem lokalen Rechner gespeichert oder an eine Datenzentrale gesendet [37].

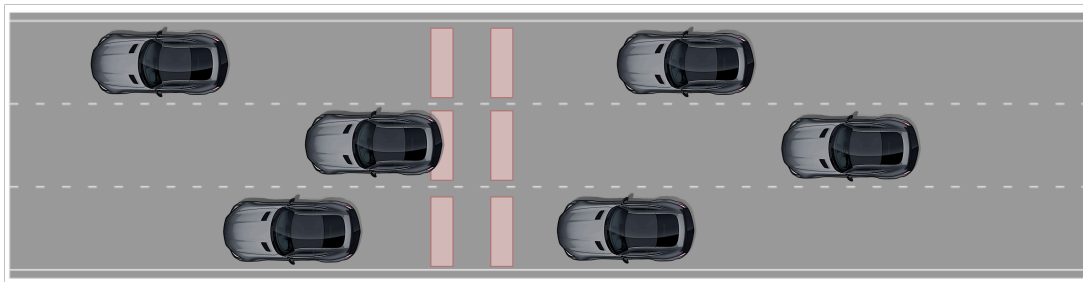


Abbildung 2.6: Vorinstallierte Doppelinduktionsschleifen (rot) an einem exemplarischen drei streifigen Autobahnabschnitt.

2.3.2 Floating Car Data

Eine Möglichkeit zur Erhebung umfangreicher Verkehrsdaten sind sogenannte Floating Car Data (FCD). Dabei werden die Positionsdaten von Fahrzeugen, die sich im Verkehrsfluss bewegen („floaten“), mittels GPS erfasst und in regelmäßigen Abständen an eine zentrale Datenplattform übermittelt [38].

Im Rahmen dieser Arbeit enthalten die FCD-Daten Positionsinformationen, Zeitstempel und eine ID, die einer einzelnen Fahrt zugeordnet werden kann, jedoch keine Rückschlüsse auf ein bestimmtes Fahrzeug zulässt. Mithilfe des Abstands zwischen zwei übermittelten Positionen und der dazugehörigen Zeitspanne kann die Geschwindigkeit einzelner Fahrzeuge berechnet werden.

Zusätzlich werden die übermittelten Daten, Straßenabschnitten zugeordnet und für diese Abschnitte können, je nach Datenlage, die Durchschnittsgeschwindigkeiten ermittelt werden [39].

Abbildung 2.7 zeigt exemplarisch den Datenaustausch zwischen kommunikationsfähigen Fahrzeugen (weiß) in fließendem Verkehr und einer Datenzentrale. Die Fahrzeuge senden ihre letzten Positionsdaten an die Datenzentrale, welche somit die aktuelle Geschwindigkeit an verschiedenen Straßenabschnitten bestimmen kann.

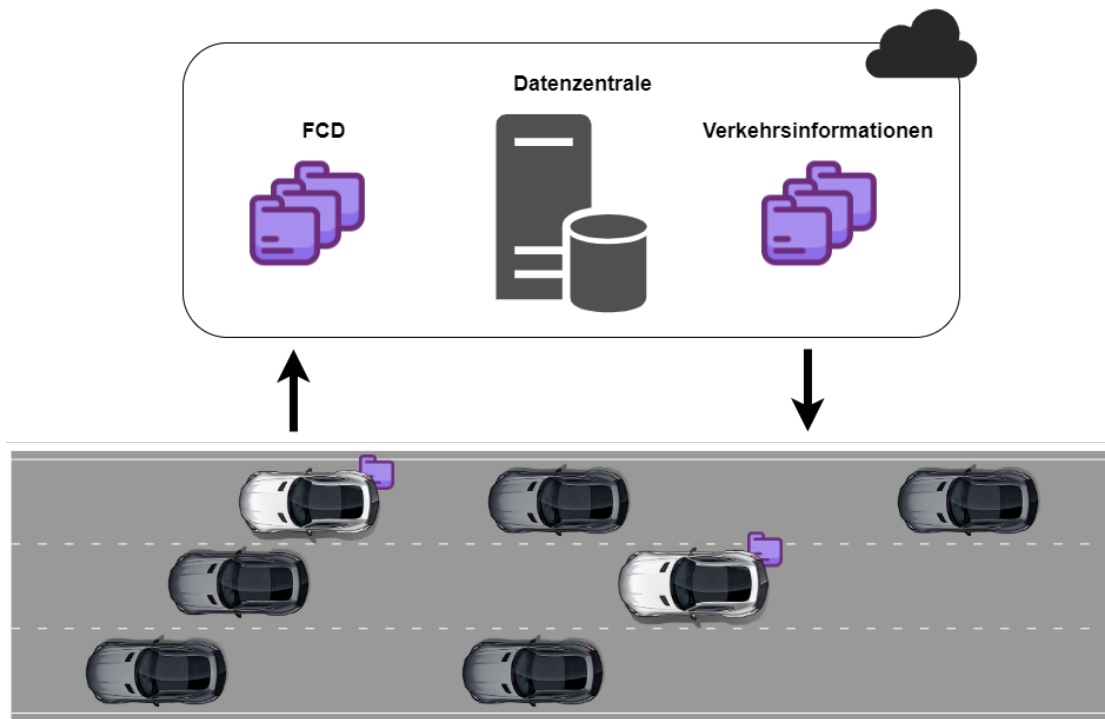


Abbildung 2.7: FCD Datenerhebung und Verkehrsinformationsübermittlung.

Durch Abweichungen von den üblichen gemessenen Geschwindigkeiten können Stauwarnungen ermittelt werden. Bei den heutigen Systemen werden üblicherweise alle 2 Minuten die Fahrzeugpositionen und die Verkehrslage zwischen Fahrzeug und Datenzentrale ausgetauscht [39].

2.3.3 Luftbeobachtungen

Die Untersuchung von Verkehr mithilfe von Drohnen eröffnet neue Perspektiven für die Analyse und Überwachung von Verkehrsströmen. Damit können aus der Vogelperspektive präzise und aktuelle Daten über Straßen, Autobahnen und Verkehrsknotenpunkte gesammelt werden. Die Vorteile dieser Methode liegen in ihrer Flexibilität und Mobilität. Drohnen können schnell zu verschiedenen Standorten bewegt werden und sind in der Lage, Verkehrsengpässe, Unfälle oder ungewöhnliche Verkehrsmuster aus der Luft zu erfassen. Mehrere Forscher haben sich dieser Aufgabe angenommen und deren Auswertungen in Datensätzen, sowie Artikeln veröffentlicht.

Das erste Experiment mit luftgestützter Beobachtung wurde von Treiterer be-

reits 1975 durchgeführt [40]. Hierbei wurden die Zeitlücken der Fahrzeuge, sowie Geschwindigkeit, Verkehrsdichte, Beschleunigungsverhalten, Spurwechsel und das allgemeine Verhalten der Fahrzeuge auf zwei städtischen Autobahnen in Columbus, Ohio untersucht. Dieses erfolgreiche Konzept der vollständigen Erfassung des Verkehrs wurde im Projekt *Next Generation Simulation* fortgesetzt, indem stationäre Kameras den Verkehr auf Autobahnen und innerstädtischen Situationen erfassten und in einem umfangreichen Datensatz zusammengetragen wurden [41]. Ein weiteres großflächiges Experiment ist die pNEUMA Betrachtung. Das durchgeführte Experiment ist eine groß angelegte Sammlung von naturalistischen Fahrzeugtrajektorien im städtischen Verkehr. Für diesen Datensatz überwachten mehrere Drohnen das Stadtzentrum von Athen, Griechenland. Die Drohnen schafften es, über 100 km des innerstädtischen Straßennetzes abzudecken, mit einer Fläche von 1,3 km², um eine eingehende Untersuchung städtischer Verkehrsphänomene zu ermöglichen [42]. Ein ähnliches Experiment wurde von den Erstellern des highD-Datensatzes durchgeführt [43]. Dieser Datensatz wurde von einer Drohne an sechs verschiedenen Standorten entlang deutscher Autobahnen aufgezeichnet. Mit einer Verarbeitungskette verschiedener Algorithmen wurden Merkmale wie Fahrzeugtypen, Trajektorien, Größe und Manöver aus den Videos extrahiert. Die Ersteller erweiterten ihren Datensatz an der Autobahn um weitere Beobachtungen und Kreuzungen (inD) [44] und Kreisverkehren (rounD) [45]. Die Autoren von [46] haben in einem Experiment auf einer Autobahn untersucht, welche Lücken von einfädelnden Fahrzeugen akzeptiert und welche abgelehnt werden. Sie verwendeten zwei Datensätze an verschiedenen Autobahneinfahrten. Die Autoren präsentierten ein stochastisches Modell der Lückenabweisung und -akzeptanz.

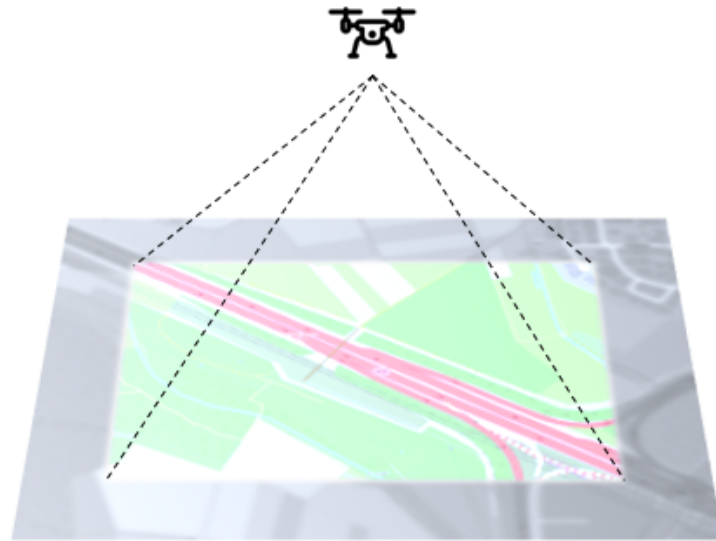


Abbildung 2.8: Luftbeobachtung mithilfe einer Drohne an einem Autobahnabschnitt.

Abbildung 2.8 zeigt exemplarisch eine Beobachtung des Verkehrs an einer Autobahn mithilfe einer Drohne. Je nach Blickwinkel der Kamera auf den Streckenabschnitt ist es möglich Fahrzeuge bis zu einer Entfernung von 600 Metern zu erfassen. Wichtig ist, dass bei besonders schrägem Blickwinkel der Positionsfehler bei besonders großen Entfernungen maximal ist und bei Annäherung an die Kamera abnimmt. Steht die Drohne genau senkrecht zur Straße ist der Positionsfehler nahezu konstant, dabei nimmt aber das Sichtfeld ab und es kann nur noch ein ca. 400 Meter großer Streckenabschnitt beobachtet werden. Zudem ist der Erfassungszeitraum auf die Akkulaufzeit der Drohne und somit meist auf ca. 25 Minuten begrenzt.

2.4 Fahrzeugstromdaten

In diesem Abschnitt werden die verschiedenen Analyseebenen des Verkehrsflusses beschrieben. Dabei wird zwischen aggregierten Verkehrsanalysen, individuellen Fahrzeugbewegungen und der Dynamik von Fahrzeuggruppen unterschieden. Diese Ansätze ermöglichen eine detaillierte Untersuchung des Verkehrsflusses durch die Kombination unterschiedlicher Maßstäbe und Perspektiven der Verkehrsanalyse.

2.4.1 Aggregierte Verkehrsanalysen

Bei diesem Ansatz wird ohne Berücksichtigung einzelner Fahrzeuge oder Fahreigenschaften der Verkehrsfluss als Ganzes in Raum und Zeit betrachtet. Dazu werden über größere Streckenabschnitte oder längere Zeiträume aggregierte Verkehrsdaten erhoben. Daraus ergeben sich typische Kenngrößen wie die mittlere Verkehrsgeschwindigkeit $\bar{v}_{Ges}$, der Verkehrsfluss q , die Verkehrsdichte ρ oder Eigenschaften der Infrastruktur wie die Anzahl der Fahrstreifen [47–50]. Aggregierte Verkehrsanalysen werden in der Regel mit Induktionsschleifen, Kameras und anderen Sensoren durchgeführt, die an strategischen Punkten entlang von Straßen und Autobahnen installiert werden. Diese Beobachtungen sind nützlich für die Analyse von Verkehrsmustern im Zeitverlauf und in großen Gebieten, wie z.B. einer Stadt oder einer Region. Als Datengrundlage für diese Ebene von Verkehrsbeobachtungen können sowohl Floating Car Data (FCD)- als auch Detektordaten dienen.

2.4.2 Individuelle Fahrzeugbewegungen

Bei diesem Ansatz werden einzelne Fahrzeuge und ihre Bewegungen untersucht. Bei der Messung müssen die einzelnen Fahrzeuge räumlich und zeitlich unterschieden werden können. Dabei werden Faktoren wie Geschwindigkeit v_x , Beschleunigung (a_x, a_y) , Nettozeitlücke (th) , Spurwechsel $(pos_{Spurwechsel})$ und Reaktionszeit (τ) berücksichtigt. Beobachtungen individueller Fahrzeuge werden in der Regel mit Videokameras, Lasersensoren oder anderen hochauflösenden Instrumenten durchgeführt, die in der Lage sind, detaillierte Daten über einzelne Fahrzeuge zu erfassen. Zum Teil werden diese Daten mit entsprechend ausgerüsteten Fahrzeugen erhoben. Diese Beobachtungen sind nützlich, um das Fahrerverhalten und die Interaktion zwischen Fahrzeugen zu verstehen. Die Datenbasis lässt sich durch den Einsatz von Versuchsfahrzeugen, deren Verkehrsverhalten durch entsprechende Sensorik aufgezeichnet wird, erweitern. Darüber hinaus eignen sich Luftbilddaufnahmen besonders für Einzelfahrzeugbetrachtungen, da sie die Bewegungen aller Fahrzeuge auf engem Raum flächendeckend verfolgen können [47–50].

2.4.3 Dynamik von Fahrzeuggruppen

Dieser Ansatz liegt zwischen den beiden zuvor genannten Betrachtungsebenen. Er betrachtet den Verkehrsfluss als eine Ansammlung von Fahrzeuggruppen und nicht als einzelne Fahrzeuge. Er basiert auf der Idee, dass der Verkehrsfluss als ein kontinuierlicher Strom von Fahrzeugen angenähert werden kann, jedoch mit Schwankungen in Geschwindigkeit und Dichte. Als Grundlage für mesoskopische

Verkehrsdaten können sowohl Detektordaten als auch FCD herangezogen werden [51].

2.5 Straßennetze in Graphendarstellung

Die Graphentheorie ist ein Teilgebiet der Mathematik, das sich mit den Eigenschaften und dem Verhalten von Graphen beschäftigt. Ein Graph ist eine mathematische Struktur, die aus Knoten besteht, welche durch Kanten verbunden sind. Mit Graphen lassen sich zahlreiche reale Phänomene modellieren, z. B. soziale Netzwerke, Verkehrsnetzwerke und chemische Verbindungen. Dieser Abschnitt behandelt den Aufbau von Graphen, die Repräsentation von Straßennetzwerken als Graphen und abschließend die Zuordnung von real gemessenen Trajektorien und Kanten des Graphen eines Straßennetzes.

2.5.1 Graphennotation

Ein einfacher Graph $G := (V, E)$ besteht aus einer nicht leeren endlichen Menge V , bestehend aus Knoten (engl. Vertices) und einer endlichen Menge E mit $E \subseteq V^2$ gefüllt mit Paaren von Elementen aus V welche die Kanten (engl. Edges) abbilden. Eine Kante $\{v, w\} \in E$ verbindet die Knoten v und w und wird für gewöhnlich mit vw abgekürzt [52, 53].

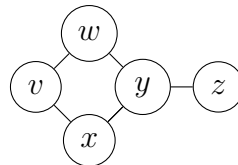


Abbildung 2.9: Exemplarische Darstellung des einfachen Graphen $G_{simple}(V, E)$.

Abbildung 2.9 zeigt die Darstellung des einfachen Graphen $G_{simple}(V, E)$ mit $V(G_{simple}) = \{v, w, x, y, z\}$ und $E(G_{simple}) = \{vw, vx, wy, xy, yz\}$. Zusätzlich zu einfachen Graphen gibt es auch die allgemeinen Graphen oder nur Graphen. Diese unterscheiden sich von den einfachen Graphen, indem in der Menge $E(G_{simple})$ Mehrfacheinträge von Kanten und Schleifen (Verweisung von einem Knoten auf sich selbst vv) erlaubt sind. Somit ist jeder einfacher Graph auch ein Graph aber nicht jeder Graph auch ein einfacher Graph [52, 53].

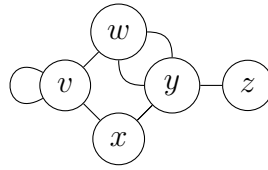


Abbildung 2.10: Exemplarische Darstellung des Graphen $G_{general}(V, E)$.

Abbildung 2.10 zeigt die Darstellung des Graphen $G_{general}(V, E)$ mit $V(G_{general}) = \{v, w, x, y, z\}$ und $E(G_{general}) = \{vw, vx, wy, wy, xy, yz, vv\}$. Dabei beinhaltet der Graph $G_{general}$ die Kanten wy zwei mal und die Schleife vv , was ihn zum allgemeinen Graphen macht und ihn von einfachen Graphen unterscheidet.

Zu der Familie der Graphen gehören auch noch gerichtete Graphen D , auch Digraphen genannt. Ein gerichteter Graph D besteht wie bei den Graphen auch aus einer nicht leeren endlichen Menge $V(D)$ und einer endlichen Menge $A(D)$ bestehend aus Paaren von $V(D)$ [52, 53].

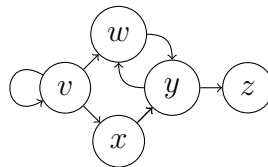


Abbildung 2.11: Exemplarische Darstellung des gerichteten Graphen $D(V, A)$.

Abbildung 2.11 zeigt die Darstellung des gerichteten Graphen $D(V, A)$ mit $V(D) = \{v, w, x, y, z\}$ und $A(D) = \{vw, vx, wy, yw, xy, yz, vv\}$. Die Richtung der Kanten wird dabei durch die Reihenfolge der Elemente von $A(D)$ vorgegeben. Die Kante vw ist eine gerichtete Kante von Knoten v zu Knoten w .

2.5.2 Repräsentation von Straßennetzen als Graph

Während ein Graph eine abstrakte mathematische Darstellung von Elementen und ihren Verbindungen abbildet, kann ein Netzwerk als eine reale Abbildung eines Graphen betrachtet werden. Beispielsweise können soziale Netzwerke als Graph dargestellt werden, indem Menschen durch Knoten und deren Beziehungen als Kanten gezeigt werden. Weitaus komplexer sind Netzwerke, deren Topologie weder völlig regulär noch völlig zufällig ist. Straßennetze sind nur ein Beispiel für komplexe räumliche Netzwerke. Ein räumliches Netzwerk ist planar, wenn zwei Dimensionen ausreichend sind um die Knoten und Kanten des Netzwerks ohne

Kreuzungen abzubilden. Allerdings sind durch Tunnels, Brücken, Autobahnen und Schnellstraßen viele Straßennetze nicht planar abbildbar. Straßennetze können primal (Knoten: Kreuzungen, Kanten: Straßen) als auch dual (Knoten: Straßen, Kanten: Kreuzungen) dargestellt werden [54–56].

In dieser Arbeit werden Straßennetze durch primale Graphen dargestellt. Diese Darstellung berücksichtigt die geografischen, räumlichen und metrischen Informationen, welche durch die duale Darstellung verloren, gehen. Die Darstellung von Straßennetzen der Kartenanbieter Open Street Map (OSM) nutzt Kreuzungen als Knoten, verwendet aber darüber hinaus Krümmungen und infrastrukturelle Veränderungen der Straße als weitere Knoten.

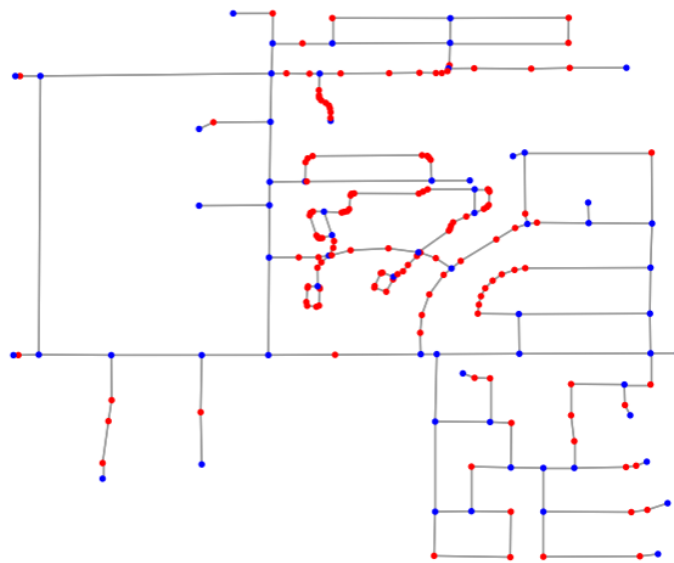


Abbildung 2.12: Exemplarische Abbildung eines Straßennetzes als Graph des Kartenanbieters OSM [54].

Die Abbildung 2.12 zeigt den Graphen von OSM eines exemplarischen Straßennetzes. Dabei sind alle Straßenabschnitte in schwarz abgebildet. Die Kreuzungen und damit echte Kanten sind durch blaue Kreise an den entsprechenden Stellen gekennzeichnet. Die roten Kreise sind durch die Implementierung von OSM hinzugekommen und treten bei jeder Krümmung eines Straßenabschnitts auf.

2.5.3 Zuordnung der Flottendaten zu Straßensegmenten (Map Matching)

Unter dem Map Matching wird das Zuordnen von einer Sequenz von Trajektorien Tr (chronologisch sortierte Koordinaten mit Zeitstempel und gegebenenfalls zusätzlichen Informationen) oder FCD zu einer Sequenz von zusammenhängenden Kanten $MR_G(Tr)$ in dem Graphen $G(V, E)$ des Straßennetzwerks bezeichnet (siehe Gleichung 2.6).

$$\begin{aligned} Tr &= [p_{t_0} \ p_{t_0+\Delta t} \ \dots \ p_{t_0+n\cdot\Delta t}] \in \mathbb{R}^{1 \times n}, \\ p_{t_i} &= [t \ x \ y] \in \mathbb{R}^{1 \times 3} \ \forall i \in 0, \dots, n, \\ MR_G &= [e_1 \ e_2 \ \dots \ e_n] \in E(G)^{1 \times n} \end{aligned} \tag{2.6}$$

Für das Finden eines passenden Kantenkandidaten $e \in E(G)$ von einem Element p_{t_i} der Sequenz Tr gibt es verschiedene Möglichkeiten. Die relevante Literatur nennt vier verschiedene Verfahren für das Map Matching: Ähnlichkeitsmodell, Zustandsübergangsmodell, Kandidatentwicklungsmodell und dem Bewertungsmodell. Diese Arbeit vertieft sich auf das Ähnlichkeitsmodell und das Zustandsübergangsmodell, weshalb diese beiden Verfahren näher vorgestellt werden [57].

Das Ähnlichkeitsmodell bezieht sich auf einen Ansatz, der die Kanten zurückgibt, die der Trajektorie geometrisch und/oder topologisch am nächsten sind. Da die Bewegung eines Fahrzeugs immer der Topologie des zugrunde liegenden Straßennetzwerks folgt und das Fahrzeug niemals von einem Segment zu einem anderen springen kann, sollte die Trajektorie dem wahren Weg auf der Karte ähneln. Der Schwerpunkt dieses Modells basiert auf der Frage, wie Nähe definiert werden kann. Diese Ähnlichkeit kann distanzbasiert ermittelt werden wobei für viele Implementierungen die Fréchetdistanz verwendet wird. Alternativ kann die Trajektorie Tr in kleinere Segmente seg_n zerstückelt werden, sodass $seg_1 \rightarrow seg_2 \rightarrow \dots \rightarrow seg_n = Tr$ gilt. Über die Longest Common Subsequence (LCSS) Bewertung der Pfade zwischen zwei Segmentpaaren wird das resultierende Matching Ergebnis ermittelt [57].

Das Zustandsübergangsmodell erstellt einen gewichteten topologischen Graphen, der alle möglichen Routen enthält, in die das Fahrzeug fahren könnte. Dabei bildet der erstellte Graph die möglichen Zustände, an denen sich das Fahrzeug befinden könnte als Knoten ab und die Kanten sind die Zustandsübergänge zu verschiedenen Zeitpunkten. Anders als bei Graphen von Straßennetzwerken, ist das Gewicht der Zustandsübergangsgraphen die Wahrscheinlichkeit von einem Knoten zu einem nächsten zu wandern. Das Ergebnis des Matchings dieses Modells ergibt sich aus dem

optimalen Pfad im Graphen insgesamt. Dafür gibt es drei bekannte Methoden: das Hidden Markov Model ([HMM](#)), das Conditional Random Field ([CRF](#)) und Weighted Graph Technique ([WGT](#)). Eine der häufigsten Implementierungen ist die [HMM](#) Methode, die sich auf den Fall konzentriert bei dem Zustände nicht beobachtbar sind (hidden) aber aufgrund der aktuell herrschenden Zustände abschätzen lassen. Aufgrund des Fehlers bei der Positionsbestimmung können alle Straßen in der Nähe des Fahrzeugs ein potenzieller Standort (Zustand) des Fahrzeugs sein. Jede dieser Straßen hat eine bestimmte Wahrscheinlichkeit der Standort des Fahrzeugs zu sein. Da sich das Fahrzeug kontinuierlich bewegt, wird der Übergang zwischen zwei aufeinanderfolgenden Zeitstempeln durch die Übergangswahrscheinlichkeit zwischen ihren Kandidatenzuständen bestimmt [58]. Das Ziel besteht darin einen optimalen Pfad zu finden, der einen Kandidaten in jedem Zeitstempel verbindet. Der resultierende Pfad wird mit dem Viterbi-Algorithmus ermittelt [57].

2.6 Erhebung von Mobilitätsparameter

Die Beurteilung oder der Vergleich des internationalen Mobilitätsverhaltens von Fahrzeugen erfordert eine systematische Erfassung und Auswertung umfassender Daten der Teilnehmer im Individualverkehr.

In verschiedenen Umfragen wird das Mobilitätsverhalten der Bevölkerung untersucht. Diese Umfragen werden von Institutionen weltweit untersucht und in Berichten für das Befragungsjahr zusammengetragen. Eine dieser Befragungen wird vom Karlsruher Institute of Technology durchgeführt und jährlich in dem Mobilitätspanel (MOP) zusammengefasst. Dabei werden Haushalte dazu gebeten ein Wegetagebuch zu führen. Die Personen sollen für jeden Weg die genutzten Verkehrsmittel, den Wegzweck, die zurückgelegte Entfernung sowie Start und Ende des Weges notieren. Diese Umfrage zielt darauf ab das gesamte Mobilitätsverhalten und damit das Tür-zu-Tür Verhalten durch Stichproben abzubilden und zu untersuchen [59]. Es gibt zu dem MOP noch die Umfrage zu der Mobilität in Deutschland (MiD). Die letzte Umfrage wurde 2017 erhoben und in einem Ergebnisbericht veröffentlicht. Eine weitere Untersuchung wird gerade noch bis 2024 durchgeführt und analysiert. Hauptziel dieser Erhebung ist es, umfassende und zuverlässige Daten zur soziodemographischen Struktur von Personen und Haushalten sowie zu ihrem täglichen Verkehrsverhalten (z.B. zurückgelegte Wege nach Zwecken und benutzten Verkehrsmitteln) über einen gesamten Jahreszeitraum zu erheben. Die gewonnenen Informationen dienen nach entsprechender Gewichtung und Hochrechnung als Rahmen und Ergänzung zu anderen Verkehrserhebungen. Ähnlich wie beim MOP werden die Daten durch ausgefüllte Tagebücher der beteiligten Haushalte

erhoben [60]. Ein weiteres erwähntes Panel ist das Netherlands Mobility Panel (MPN), eine Haushaltspanel-Umfrage, die Daten zum Reiseverhalten sammelt. Das Panel verwendet ein modernes Reisetagebuch-Design, um Informationen zu einzelnen Reisen und Reiseabschnitten zu erfassen. Die Daten des MPN werden genutzt, um Zusammenhänge zwischen Lebensveränderungen, wie beispielsweise beruflichen Veränderungen, der Geburt eines Kindes oder einer Scheidung, und den daraus resultierenden Mobilitätsmustern zu untersuchen. Hier wird wie bei den zuvor genannten Umfragen ein Tagebuch für jeden Haushalt geführt, in dem nicht nur die zurückgelegten Strecken erfasst werden, sondern auch Informationen zu genutzten Verkehrsmitteln, Mitreisenden, eventuellen Verzögerungen und Parkkosten [61].

3 Analyse weltweiter FCD einer PKW-Flotte

Der Zugang zu internationalen Fahrzeugflottendaten kann Einblicke in Trends, Herausforderungen und Chancen im Verkehrssektor geben. Diese Daten können Unternehmen, politischen Entscheidungsträgern und Forschern als Grundlage für Entscheidungen dienen und zu einer nachhaltigeren Zukunft führen können.

In diesem Kapitel wird die Erhebung eines Datensatzes beschrieben, der die **FCD** einer **PKW** Flotte enthält. Dabei wird die weltweite räumliche Verteilung der Daten aufgezeigt. Zusätzlich werden Fahrtparameter wie Fahrdistanz, Fahrzeit und Durchschnittsgeschwindigkeit pro Land analysiert, um das **PKW**-Nutzungsverhalten der Marke Mercedes-Benz in den jeweiligen Regionen aufzuzeigen. Darüber hinaus werden die für die jeweiligen Regionen gewonnenen Fahrtenkenngrößen unterschiedlichen Mobilitätserhebungen aus ähnlichen Regionen gegenübergestellt. Ziel ist es, die Ähnlichkeit dieses herstellereinspezifischen Datensatzes, der in zufällig ausgewählten Zeiträumen erhoben wurde, mit vergleichbaren Erhebungen zur Mobilität zu vergleichen. Darüber hinaus wird die Ähnlichkeit der Verteilungen der jeweiligen Fahrtparameter in räumlicher Hinsicht (Länder, Städte) und in zeitlicher Hinsicht (Tage) untersucht. Zu diesem Zweck wird mit Hilfe einer Distanzmetrik eine Ähnlichkeit für die räumliche bzw. zeitliche Betrachtung ermittelt und in Form eines Dendrogramms dargestellt. Schließlich wird eine parametrisierte bekannte Verteilung, die das Verhalten der realen Flotte möglichst genau beschreibt, für jeden zuvor beschriebenen Fahrtparameter dargestellt.

3.1 Datensatz

Der folgende Abschnitt beschreibt die Herkunft der Daten für dieses Kapitel, die als Grundlage für die nachfolgende Auswertung dienen. Des Weiteren wird beschrieben, wie die Daten aus den **FCD** zu Fahrteninformationen wie zurückgelegte Wegstrecke, Fahrtdauer und Durchschnittsgeschwindigkeit zusammengefasst werden. Außerdem

wird die Datenbereinigung beschrieben, deren Ziel die Trennung des Datensatzes von fehlerhaften Daten ist.

3.1.1 Beschreibung des Datensatzes

Der Datensatz der vorliegenden Arbeit besteht aus einer anonymisierten Sammlung von FCD einer Flotte vernetzter PKW. Um Informationen über die aktuelle Verkehrssituation in der näheren Umgebung eines Fahrzeugs zu erhalten, werden zyklisch anonymisierte Positionsdaten an eine Datenzentrale gesendet. Diese Datenzentrale aggregiert die von den einzelnen Fahrzeugen übermittelten Daten und ermittelt aus den Abweichungen von den üblicherweise gemessenen Verkehrsgeschwindigkeiten Aussagen über die aktuelle Verkehrslage. Bei typischen Implementierungen in modernen Fahrzeugen wird üblicherweise alle 2 Minuten ein Datenpaket bestehend aus 24 Positionen und den zugehörigen Zeitstempeln an die Datenzentrale gesendet. Bei der Übertragung wird die Herkunft der Daten anonymisiert. Die Daten können also nicht einem einzelnen Fahrzeug, sondern nur einer eindeutigen *SessionID* zugeordnet werden. Einzelne zusammengehörige Positionspunkte und deren Zeitstempel werden durch Gruppierung der Datenpakete mit gleicher *SessionID* zu einer Fahrt T zusammengefasst. Die Gleichung 3.1 zeigt, wie Datenpakete zu einer Fahrt T zusammengefasst werden.

$$\begin{aligned}
 T &= [p_{t_0} \ p_{t_0+\Delta t} \ \dots \ p_{t_0+n \cdot \Delta t}] \in \mathbb{R}^{1 \times n}, \\
 \text{mit } p_{t_i} &= [\textit{SessionID} \ t_i \ \textit{Breitengrad}_{t_i} \ \textit{Laengengrad}_{t_i}], \ \Delta t = 5\text{sec}, \\
 \text{und } n &:= \text{Anzahl von FCD für die Session der Fahrt } T
 \end{aligned}
 \tag{3.1}$$

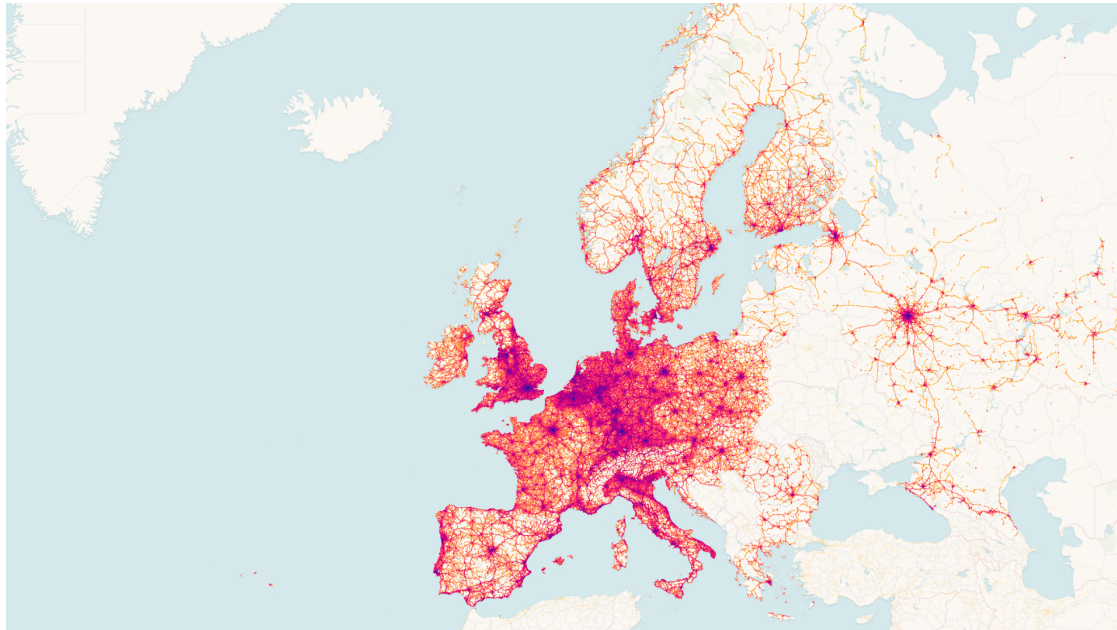
Zudem beziehen sich die Daten dieser Arbeit ausschließlich auf Fahrzeuge der Marke Mercedes-Benz und sind daher nicht unbedingt aussagekräftig für die Gesamtmobilität aller Fahrzeugtypen und -marken. Hinzu kommt, dass die Daten immer nur pro Fahrt erfasst werden, so dass z.B. die Anzahl der Fahrten mit demselben Fahrzeug an einem Tag nicht ermittelt werden kann. Bei den im Datensatz enthaltenen Fahrzeugen handelt es sich überwiegend um PKW, die zum Zeitpunkt der Erstellung des Datensatzes nicht älter als 3 Jahre waren. Der Datensatz wurde in mehreren Messkampagnen von unterschiedlicher Dauer in den Jahren 2021 und 2022 erhoben. Die Zeiträume der Messkampagnen und die Anzahl der erhobenen Fahrten in den jeweiligen Regionen sind in Tabelle 3.1 ersichtlich. Die Tabelle zeigt die Konzentration der Flotte im europäischen Raum mit durchschnittlich 4,5 Mio.

Fahrten pro Tag gegenüber etwas mehr als 1,1 Mio. Fahrten im nordamerikanischen Raum und ca. 1 Mio. Fahrten in China.

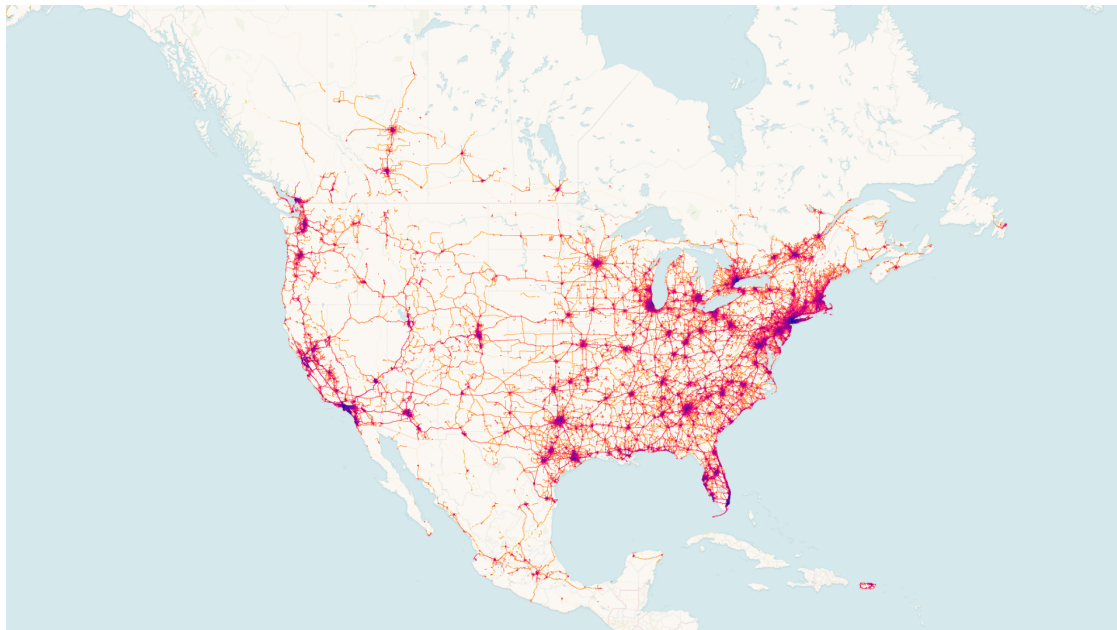
Region	Zeitraum	# Fahrten
Europa	18.02.21 - 23.02.21	25.212.023
	23.04.21 - 30.04.21	36.633.355
	15.06.21 - 26.06.21	55.252.070
	30.07.21 - 02.08.21	11.632.245
	18.03.22 - 25.03.22	43.669.763
	38 Kalendertage	172.399.456
	27 Mo. - Fr.	123.306.389
	9 Sam- & Sonntage	49.093.067
Nord Amerika	23.04.21 - 21.05.21	31.953.912
	29 Kalendertage	31.953.912
	21 Mo. - Fr.	23.403.603
	8 Sam- & Sonntage	8.550.309
China	14.07.21 - 02.08.21	19.976.071
	20 Kalendertage	19.976.071
	14 Mo. - Fr.	13.800.108
	6 Sam- & Sonntage	6.175.963
Total		224.329.439

Tabelle 3.1: Zeiträume der Messkampagne und Anzahl der erfassten Fahrten je Region.

Abbildung 3.1 zeigt alle Positionsdaten in ausgewählten Regionen, die an einem zufällig ausgewählten Tag während der Untersuchungszeiträume erfasst wurden. Als Darstellungsform wurde eine Heatmap gewählt. Dabei werden räumlich naheliegende GPS-Daten zu einem Punkt zusammengefasst. Eine besonders dunkle Darstellung zeigt Ballungsräume mit hoher Datendichte. Häufig befahrene Straßen, sowie Großstädte werden bei den erhobenen Daten der vernetzten Flotte besonders hervorgehoben. Auf dieser beispielhaften Karte sind mehrere Milliarden GPS-Daten aus den Untersuchungszeiträumen dargestellt. Die Einfärbung des europäischen Raumes verdeutlicht nochmals die hohe Konzentration der im Datensatz verfügbaren Flotte. In der Abbildung 3.1 sind einige Bereiche der Karte nicht eingefärbt. In diesen Gebieten steht der Verkehrsinformationsdienst nicht zur Verfügung, so dass keine Positionsdaten verfügbar sind.



(a)



(b)

Abbildung 3.1: Heatmap der Positionsdaten im Datensatz für die Regionen (a) Europa, (b) Nord Amerika [e]¹.

3.1.2 Datenvorverarbeitung

Für den endgültigen Untersuchungsdatensatz werden Informationen wie die zurückgelegte Distanz d_T , die benötigte Zeit t_T und die Durchschnittsgeschwindigkeit $\bar{v}_T$ aus den circa 224 Millionen weltweit erfassten Fahrten T zum Datensatz D zusammengefasst. Die zurückgelegte Distanz wird aus der Summe der einzelnen kürzesten Entfernungen von Position zu Position entlang der Erdkugel (auch als Orthodrome oder als engl. *haversine distance* bezeichnet) der übermittelten Daten einer Fahrt berechnet. Die benötigte Dauer errechnet sich aus der Differenz zwischen dem Zeitpunkt der letzten gesendeten Position und dem Zeitpunkt der ersten Position. Die Durchschnittsgeschwindigkeit wird berechnet, indem das Mittel aus den ermittelten Geschwindigkeiten von FCD-Punkt zu Punkt berechnet wird. Alle Berechnungen für die Zusammenfassung einer Fahrt werden in der Gleichung 3.2 dargestellt.

$$\begin{aligned}
 d_T &= \sum_{i=0}^{n-1} \text{haversine}(p_{t_0+(i+1)*\Delta t}, p_{t_0+i*\Delta t}) \\
 t_T &= t_{0+n*\Delta t, T} - t_{0, T} \\
 \bar{v}_T &= \frac{1}{n-1} \sum_{i=0}^{n-1} v_{t_1+(i+1)*\Delta t} \\
 \text{mit } n &:= \text{Anzahl von FCD für die Session der Fahrt } T
 \end{aligned}
 \tag{3.2}$$

Zusätzlich werden das Herkunftsland $orgCntry_T$ und die Herkunftsstadt $orgCty_T$ des ersten übermittelten FCD Datenpakets einer Fahrt ermittelt. Die Gleichung 3.3 gibt den Inhalt des Datensatzes D an.

1 Kartenmaterial bereitgestellt von: CARTO, OSM
Abbildung erstellt mit: Plotly

$$\begin{aligned}
D^T &= [x_{T_0} \ x_{T_1} \ \dots \ x_{T_n}] \in \mathbb{R}^{1 \times n}, \\
\text{mit } x_T &= [d_T \ t_T \ \bar{v}_T \ \text{orgCntry}_T \ \text{orgCty}_T] \\
n &:= \text{Anzahl an Fahrten}, \\
d_T &: \text{Zurückgelegte Distanz der Fahrt } T, \\
t_T &: \text{Benötigte Zeit für Fahrt } T, \\
\bar{v}_T &: \text{Durchschnittliche Geschwindigkeit der Fahrt } T \\
\text{orgCntry}_T &: \text{Land der Startposition der Fahrt } T \\
\text{orgCty}_T &: \text{Stadt der Startposition der Fahrt } T
\end{aligned} \tag{3.3}$$

3.1.3 Datenbereinigung

Bei der Verwendung von Realdaten ist die Datenbereinigung eine wichtige Aufgabe, um fehlerhafte und verrauschte Daten herauszufiltern. Der Datensatz enthält Fahrzeuge wie Taxis oder parkende Fahrzeuge, die sich während der gesamten Fahrt nicht bewegen. Daher werden Datenreihen mit folgenden Merkmalen aus dem Datensatz entfernt: (I) Fahrten mit einer zurückgelegten Wegstrecke von weniger als 300 m, um parkende und nur umparkende Fahrzeuge aus dem Datensatz auszuschließen. (II) Fahrten mit einem Zeitstempel, der nicht im Zeitraum der Messkampagne liegt und somit die Ergebnisse verfälschen würde. Dazu gehören auch Fahrten, bei denen Zeitsprünge auftreten. (III) Nicht nur die Zeit der einzelnen Fahrzeuge kann springen, sondern auch die Positionsdaten können bei der Übertragung springen. Daher werden Fahrten mit einer Positionsänderung von mehr als 400 m innerhalb von 5 Sekunden herausgefiltert.

Nach der Bereinigung des Datensatzes wurden ca. 4 Millionen Datenreihen verworfen, was 1,7 % des Rohdatensatzes entspricht. Die Tabelle 3.1 zeigt bereits die Anzahl der nach der Bereinigung im Datensatz verbliebenen Fahrten. Nach der Bereinigung verblieben 224 Millionen Fahrten, die für die folgenden Auswertungen verwendet wurden. Abbildung 3.2 zeigt die Häufigkeit der Fahrten pro Land für die 15 Länder, die im bereinigten Datensatz am häufigsten vorkommen. Deutschland sticht mit knapp über 50 Millionen Fahrten aus allen anderen Ländern heraus. In den USA wurden während der Messkampagnen rund 25 Millionen Fahrten aufgezeichnet. Knapp dahinter folgt Großbritannien mit knapp 25 Millionen Fahrten. Diese drei Länder machen zusammen fast 50 % der Fahrten im Datensatz aus.

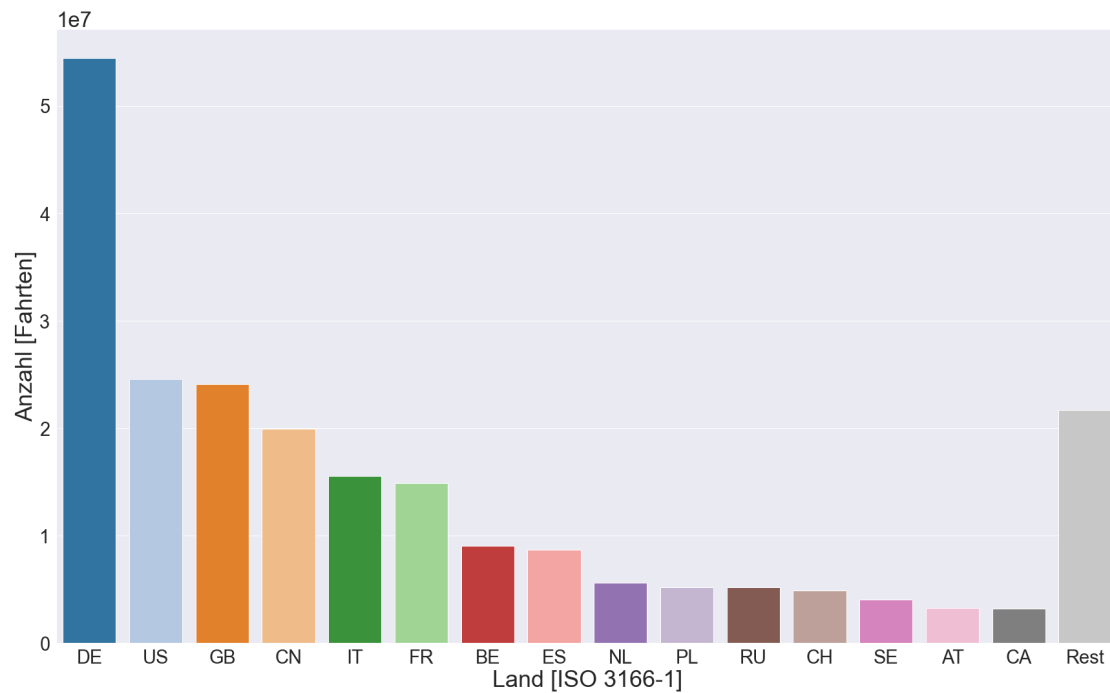


Abbildung 3.2: Anzahl an Fahrten in den 15 häufigsten aufkommenden Ländern im Datensatz.

3.1.4 Map Matching

Durch den Einsatz von Map Matching kann die Genauigkeit von GPS-Koordinaten erheblich verbessert werden. Map Matching ermöglicht es, ungenaue GPS-Koordinaten bekannten Straßenabschnitten zuzuordnen und so die Positionsgenauigkeit zu erhöhen. Dies ist insbesondere in städtischen Gebieten mit komplexem Straßennetz und schwierigen GPS-Empfangsbedingungen von Vorteil. Abbildung 3.3 zeigt die GPS-Rohdaten einer Flotte (a) und das Ergebnis eines Map Matching-Verfahrens (b). Durch die erweiterte Kenntnis der Straßeninfrastruktur wurde in dieser Abbildung ein Teil des Datensatzes den entsprechenden Straßensegmenten zugeordnet. Durch den Abgleich der GPS-Koordinaten mit den verfügbaren Karteninformationen kann die genaue Position eines Fahrzeugs auf der Straße bestimmt werden. Durch die Verfügbarkeit qualitativ hochwertiger Kartendaten können GPS-Koordinaten immer genauer und zuverlässiger auf Straßenpositionen abgebildet werden. Die Abbildung 3.4 zeigt die Positionskorrektur zwischen den unbearbeiteten Positionsdaten der Flotte und den zugeordneten Punkten auf dem Straßenabschnitt. Die meisten Positionskorrekturen liegen unter 10 m, wobei 30 % der Daten eine Ungenauigkeit von unterhalb von 1 m abdecken.

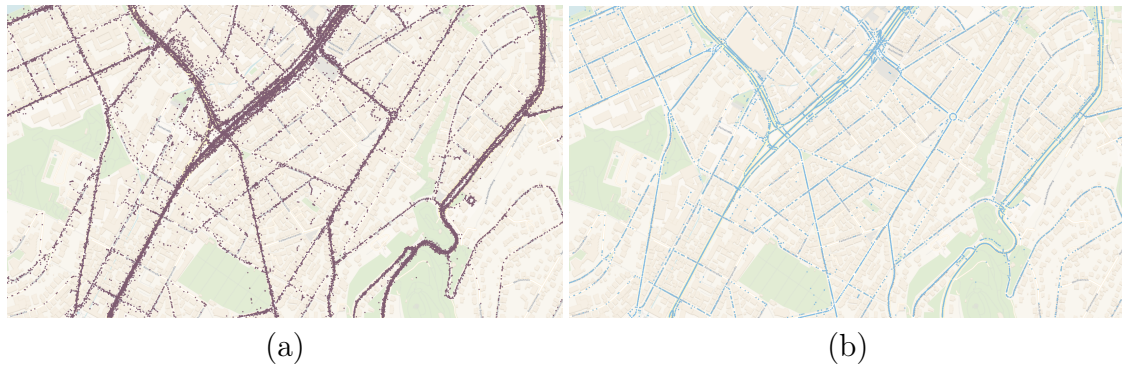


Abbildung 3.3: Exemplarische räumliche Positionsdaten (a) vor dem Mapmatching und (b) nach dem Mapmatching.

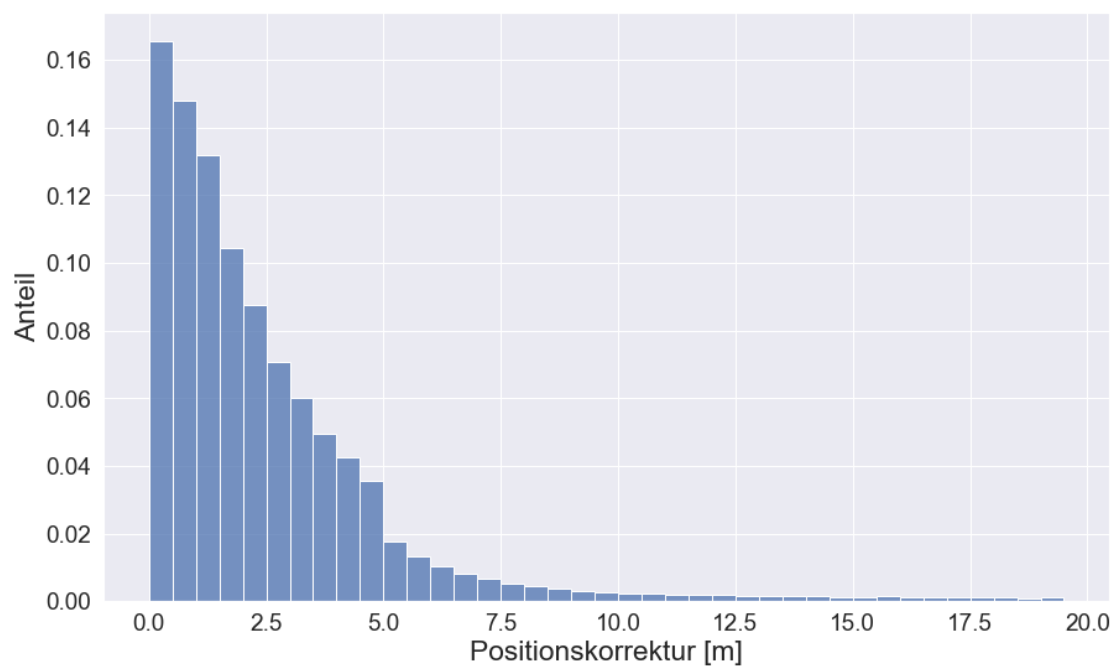


Abbildung 3.4: Histogramm der Positionskorrektur durch Mapmatchingverfahren.

3.2 Internationale Fahrtparameter

Nach der Bereinigung des Datensatzes werden die verbleibenden Daten anhand von drei Fahrtparametern ausgewertet. Dabei werden die fünf Länder mit der höchsten Repräsentanz im Datensatz verglichen. Dieser Abschnitt beschreibt zunächst die Auswahl der Fahrtparameter. Anschließend wird die Auswertung des Datensatzes für die verschiedenen Dimensionen beschrieben: Distanz, Reisezeit und Durchschnittsgeschwindigkeit. Anschließend werden die Tagesganglinien, d.h. die über den Tag verteilten Fahrzeuganfragen, für ausgewählte Länder im Datensatz dargestellt.

3.2.1 Wahl der Fahrtparameter

Mobilitätsmessungen und Erreichbarkeitsparameter lassen sich in der Literatur in fünf Kategorien einteilen: infrastrukturbasierte, standortbasierte, personenbasierte, nutzungsbasierte und tätigkeitsbezogene Messgrößen [62–64]. Aufgrund der Natur der verfügbaren Daten und deren Informationen ist es lediglich möglich die standortbasierten Messgrößen zu betrachten. Standortbezogene Messgrößen untersuchen den Erreichbarkeitsgrad an Standorten, typischerweise auf einer Makroebene zu räumlich verteilten Standorten [65]. Standortbezogene Messgrößen haben sich als sehr nützlich erwiesen, um sowohl die regionale als auch die lokale Erreichbarkeit zu verstehen [63]. Die distanzbasierten Messgrößen sind Teil der standortbezogenen und behandeln die relative Erreichbarkeit zwischen Start und Ziel in den Dimensionen: gefahrene Distanz, die dafür benötigte Zeit und den verbundenen Kosten [66]. Die Wahl der untersuchten Parameter in dieser Arbeit fiel auf die folgenden drei: gefahrene Distanz, benötigte Reisezeit und Durchschnittsgeschwindigkeit.

Die gefahrene Distanz liefert Informationen über tägliche Bewegungsmuster der Flotte. Sie unterstützt, Pendelmuster zu identifizieren und beispielsweise Verkehrsflüsse zwischen urbanen und nicht urbanen Fahrten zu unterscheiden und zu verstehen. Die Reisezeit ist ein entscheidender Indikator für die Effizienz des Verkehrssystems. Durch die Analyse der Reisezeit in Kombination mit der gefahrenen Distanz und dem Standort können Stausituationen und Verkehrsstörungen identifiziert werden, welche die Fahrt dauern verlängern und die Verkehrseffizienz beeinträchtigen. Dieses Wissen ist entscheidend für die Identifizierung von Engpässen und die Planung von Maßnahmen zur Verbesserung des Verkehrsflusses. Die Durchschnittsgeschwindigkeit gibt Aufschluss über die Geschwindigkeit, mit der sich Verkehrsteilnehmer fortbewegen. Sie ermöglicht, die Leistungsfähigkeit von Straßen und Verkehrsnetzen zu bewerten. Eine hohe Durchschnittsgeschwindigkeit deutet auf einen effizienten Verkehrsfluss hin, während niedrige Durchschnittsgeschwindigkeiten auf Engpässe

oder Verkehrsprobleme hinweisen können. Die Analyse der Durchschnittsgeschwindigkeit hilft bei der Bewertung der Verkehrsqualität und der Identifizierung von Bereichen, die Verbesserungen erfordern.

3.2.2 Distanz

Die von Fahrzeugen zurückgelegte Strecke ist ein wesentlicher Parameter für das Verständnis und die Analyse des Verhaltens von Fahrzeugen im Straßenverkehr. Distanzmessungen liefern wichtige Daten für die Ermittlung von Mustern und Trends bei Fahrzeugbewegungen sowie für die Planung, den Entwurf und den Betrieb von Verkehrssystemen. Entfernungsmessungen in Verbindung mit Positionsdaten, die Straßensegmenten zugeordnet sind, können dazu beitragen, Trends in der Fahrzeugnutzung zu erkennen, z. B. die am häufigsten benutzten Routen und die Tageszeiten mit dem höchsten Verkehrsaufkommen, was zur Optimierung von Verkehrsdienstleistungen und -infrastrukturen genutzt werden kann. Die Tabelle 3.2 fasst die gefahrenen Distanzen und die Bewertung der fünf Länder zusammen. Als Referenz sind in der Spalte *Global* die gefahrenen Distanzen für den gesamten Datensatz X mit den untersuchten Ländern und 98 weiteren Ländern aufgeführt. Der Mittelwert ist das arithmetische Mittel aller erfassten Fahrten. Die Standardabweichung gibt die Streuung der jeweiligen Verteilung um den Mittelwert an. Der Median der Distanz ist die Distanz, bei der die Hälfte der Fahrten im Datensatz unter dem angegebenen Wert liegt. In Tabelle 3.2 fällt auf, dass die Fahrzeuge in den verschiedenen Ländern im Durchschnitt sehr ähnliche Distanzen zurücklegen, nämlich etwa 12-14 km. Die kürzesten Strecken werden in China zurückgelegt, die längsten in den Vereinigten Staaten. Der Median verdeutlicht diese Ähnlichkeiten. Der Median der zurückgelegten Distanzen liegt in allen Ländern zwischen 4 und 6 km (mit Ausnahme der Vereinigten Staaten, wo der Median $\sim 6,65$ km beträgt). Die Standardabweichung ist im Vereinigten Königreich (UK) niedriger als in allen anderen Ländern und auch niedriger als der Gesamtwert: Die Fahrzeugflotte scheint von den Nutzern gleichmäßiger ausgelastet zu sein. Im Vergleich zu den Daten der anderen Länder ist China durch kürzere Fahrten und eine dichtere Verteilung der Fahrzeugflotte in städtischen Gebieten gekennzeichnet. In allen untersuchten Ländern sind die Wochenendfahrten länger als die Fahrten unter der Woche: Der Freizeitverkehr scheint zu etwas größeren Entfernungen zu führen als der werktägliche Berufsverkehr. Eine Zusammenfassung für die Distanz und weitere Fahrtparameter für einzelne Städte wird in A.1 aufgeführt.

Die von allen Fahrzeugen der globalen Flotte zurückgelegten Strecken sind in Abbildung 3.5 als Histogramm dargestellt. Jeder Balken zeigt den Anteil des Datensatzes für einen Bereich von fünf Kilometern. Das Histogramm veranschaulicht,

	Deutschland	USA	China	UK	Italien	Global
Anzahl	54,5 Mio	24,6 Mio	20,0 Mio	24,1 Mio	15,5 Mio	224,3 Mio
Mittel km	13,91	14,31	12,20	12,73	13,37	13,85
SD km	29,99	27,32	28,33	24,02	27,92	28,26
Median km	4,92	6,65	4,46	5,15	4,92	5,28
25 % - 75 % Quantil km-km	1,80- 13,21	2,76- 15,34	1,88- 10,90	2,05- 12,88	1,87- 12,81	2,02- 13,65
99 % Quantil km	155,92	130,84	145,42	124,28	144,85	143,90
Mittel Mo. - Fr. km	12,74	13,32	11,48	11,93	12,36	12,71
Mittel Wochenende km	15,00	15,63	12,46	13,28	14,75	14,97

Tabelle 3.2: Zusammenfassung der gefahrenen Distanzen pro Fahrt.

dass sich die gefahrenen Strecken extrem auf kurze Entfernungen konzentrieren, was durch eine positive Skewness (*Global skewness*: ~ 6.3) und eine positive Kurtosis (*Global kurtosis*: ~ 61.7) belegt wird. Der Anteil der kurzen Fahrten macht einen großen Teil des Datensatzes aus.

Abbildung 3.6 zeigt die *ECDF* für jedes der fünf untersuchten Länder, um einen genaueren Überblick zu geben. Die *ECDF* gibt die Wahrscheinlichkeit an, dass ein Messwert aus dem Datensatz höchstens den Wert der untersuchten Variable erreicht hat. In diesem Fall gibt der jeweilige Wert auf der y-Achse die Wahrscheinlichkeit an, mit der eine Fahrt maximal die Entfernung des entsprechenden x-Wertes zurückgelegt hat. Die Kurven der verschiedenen Länder sind auf den ersten Kilometern fast identisch; und auch bei etwa 50 km aufwärts sind die *ECDF*-Kurven der jeweiligen Länder wieder gebündelt. Die USA und China fallen dadurch auf, dass sich ihre *ECDFs* von denen der anderen untersuchten Länder unterscheiden. Während China einen höheren Anteil an kürzeren Distanzen zwischen 5 und 50 km aufweist, haben die Vereinigten Staaten einen höheren Anteil an längeren Distanzen. Die *ECDF*-Kurven für Italien, das Vereinigte Königreich und Deutschland liegen jedoch sehr nahe beieinander. Längere Fahrten von mehr als 100 Kilometern machen in

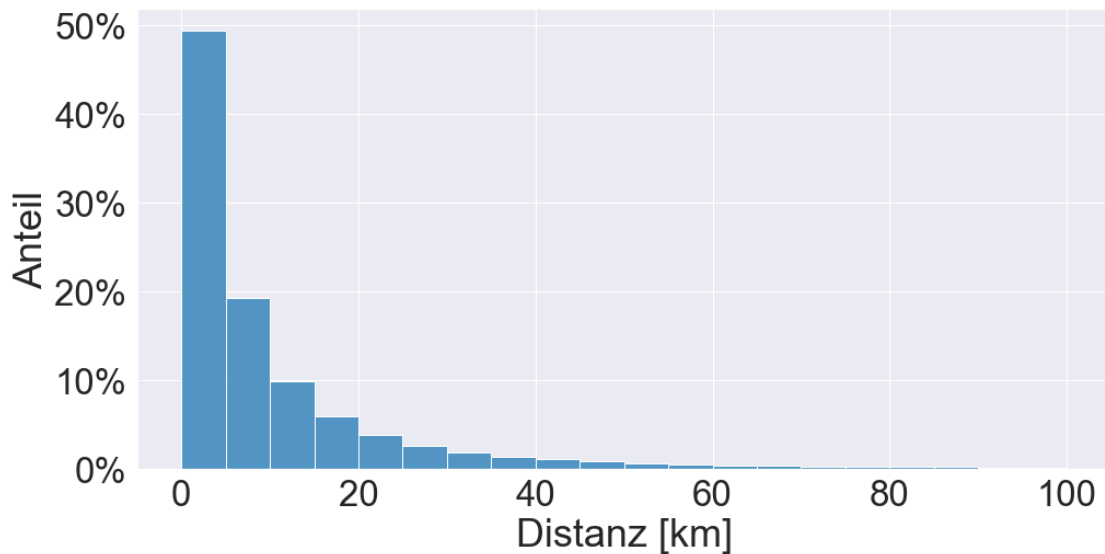


Abbildung 3.5: Histogramm der global gemessenen Distanzen der Flotte.

jedem Land weniger als fünf Prozent aller erfassten Fahrten aus.

Die Abbildung 3.7 zeigt die stündliche Verteilung der zurückgelegten Distanzen für den gesamten Datensatz in Deutschland. Das Bild zeigt deutlich, dass zwischen 8 und 20 Uhr die Verteilungen der zurückgelegten Entfernungen pro Stunde geringere Streuungen aufweisen. Die Boxen, die das 25- und das 75-Prozent-Quantil der Verteilungen markieren, sind in diesem Zeitraum kompakter als die Boxen von 21 Uhr bis 7 Uhr. Die Mediane der zurückgelegten Distanzen in der Nacht sind teilweise doppelt so hoch wie die Mediane tagsüber. Darüber hinaus sind die Ausprägungen der Boxen der zurückgelegten Distanzen zwischen 8 Uhr und 20 Uhr deutlich kürzer als die der Boxen in den übrigen Stunden. Besonders hervorzuheben ist, dass die zurückgelegten Distanzen zwischen 2 Uhr und 6 Uhr im Vergleich zum Rest des Tages die größte Streuung aufweisen. Insgesamt lässt die Abbildung Rückschlüsse auf die Verteilung der zurückgelegten Entfernungen im Tagesverlauf zu. Zwischen 8 und 20 Uhr sind die zurückgelegten Entfernungen tendenziell stabiler und kürzer, während sie zwischen 21 und 7 Uhr länger und variabler sind.

3.2.3 Fahrtdauer

Die Auswertung der Fahrtdauer ergibt ein etwas anderes Bild als die zurückgelegten Entfernungen, wie die Tabelle 3.3 zeigt. Deutschland hat die kürzeste durchschnittliche Reisezeit, dicht gefolgt von Italien. Im Vereinigten Königreich ist die durchschnittliche Reisezeit etwa 2 Minuten länger als in Deutschland. China

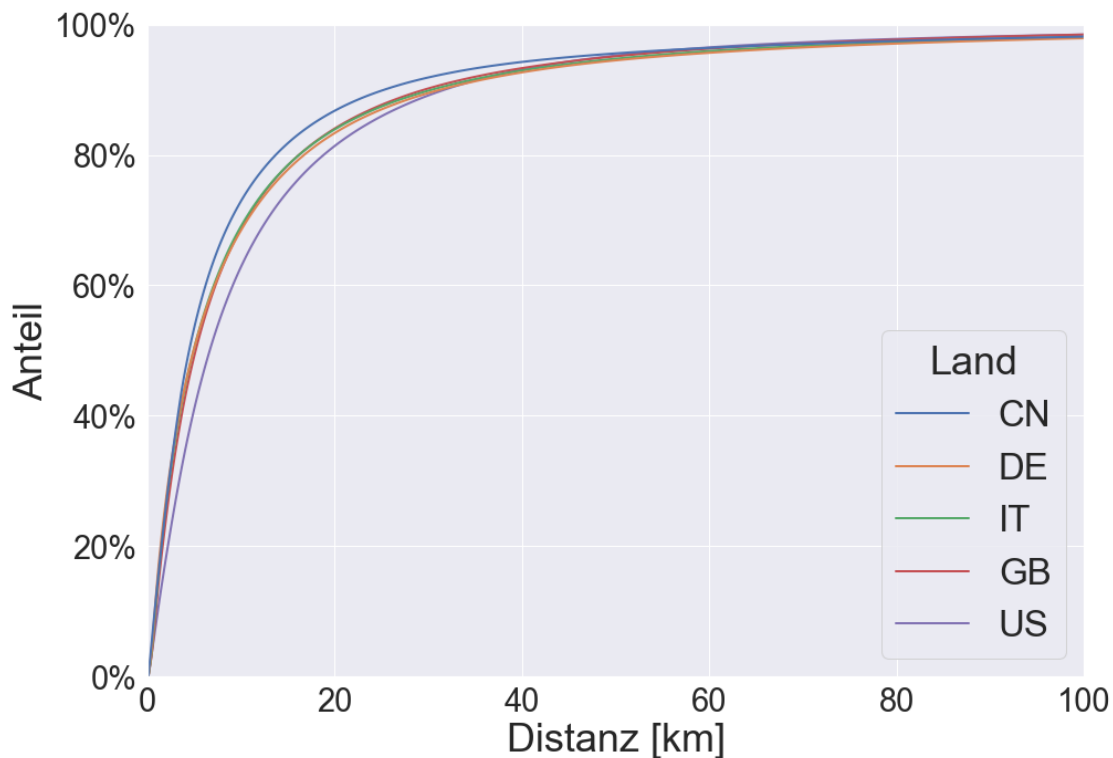


Abbildung 3.6: Empirische kumulative Verteilungsfunktion der gefahrenen Fahrtstrecken pro untersuchtem Land.

und die Vereinigten Staaten weisen die längsten Reisezeiten auf. Bezogen auf die zurückgelegten Entfernungen benötigt China jedoch für die kürzesten Entfernungen die längste Reisezeit. Diese Ergebnisse werden durch die Medianwerte für alle Länder bestätigt. Auch hier ist die Reisezeit in Deutschland am kürzesten, während sie in China und den Vereinigten Staaten etwa 4 Minuten länger ist. Die globalen Reisezeiten für alle Länder sind im Median etwas länger als die Werte für die meisten in dieser Studie untersuchten Länder. Bemerkenswert ist, dass 99 % aller Fahrten weltweit weniger als 02:09 Stunden dauern. In China liegt das 99 %-Quantil 15 Minuten über diesem Wert.

Die [ECDF](#) der Reisezeiten der Länder in [Abbildung 3.8](#) zeigt, dass die Verteilungen für Italien und Deutschland ähnlich sind (siehe [Tabelle 3.3](#)). Die [ECDF](#) der beiden Länder verlaufen sehr ähnlich. Chinas [ECDF](#) für die Fahrtdauer ist am weitesten von den anderen Ländern entfernt und hat die längsten Reisezeiten in dieser Studie. Die verschiedenen [ECDF](#)-Funktionen weisen alle eine durch die Messmethode bedingte Stufenbildung auf. Wie bereits erwähnt, werden bei der Datenübertragung die Fahrzeugpositionen (im Abstand von 5 Sekunden; insgesamt 24 pro Übertragung)

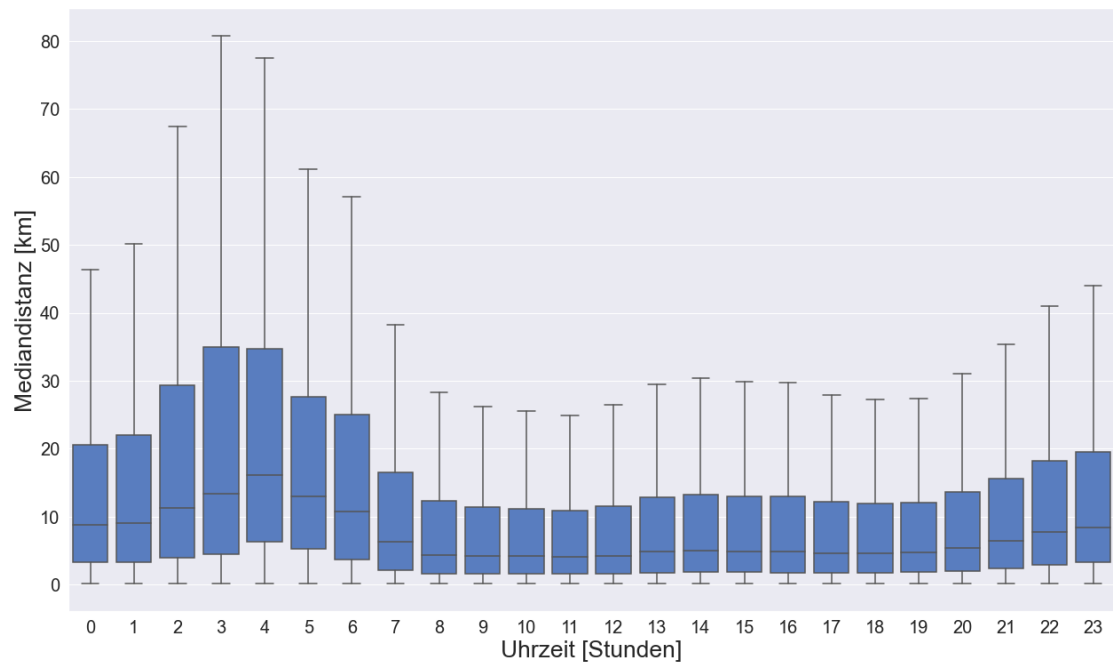


Abbildung 3.7: Tagesverlauf der gefahrenen Distanzen in Deutschland.

	Deutschland	USA	China	UK	Italien	Global
Mittel min:ss	17:56	21:18	23:14	19:50	18:04	19:57
SD min:ss	23:53	24:34	29:03	24:16	22:42	25:19
Median min:ss	09:55	13:53	13:46	11:51	09:57	11:51
25 % - 75 % Quantil min:ss- min:ss	04:26– 21:50	07:50– 26:00	05:50– 28:00	05:50– 24:01	05:49– 21:51	05:50– 23:55
99 % Quantil hh:min:ss	02:02:54	01:59:52	02:23:41	02:02:11	01:55:46	02:08:15

Tabelle 3.3: Zusammenfassung der Fahrtzeiten pro Fahrt.

in der Regel alle 2 Minuten gesendet, was zu einem schrittweisen Anstieg der [ECDF](#) der Reisezeiten führt.

Die dargestellte Abbildung [3.9](#) veranschaulicht die stündliche Verteilung der Reisezeiten für den gesamten Datensatz in Deutschland. Ähnlich wie bei den zurückgelegten Distanzen zeigt sich auch hier, dass zwischen 8 und 20 Uhr die Verteilungen der gefahrenen Durchschnittsgeschwindigkeiten pro Stunde eine gleichmäßige Streuung aufweisen. Die Verteilungen der Reisezeiten in diesem Zeitraum sind nahezu identisch verteilt, mit ähnlichen Medianwerten und nahezu gleichen Streuungen. Die Unterschiede in den Medianwerten zwischen Tag und Nacht betragen im Extremfall weniger als 10 Minuten. Deutlich längere Reisezeiten sind in den frühen Morgenstunden zu beobachten, während sie im Laufe des Tages abnehmen. Die Reisezeiten zeigen eine erhöhte Streuung in der Nacht zwischen 0 Uhr und 6 Uhr. Tagsüber gibt es noch einen leichten Anstieg der Reisezeiten um ca. 14 und 16 Uhr.

3.2.4 Geschwindigkeit

Im direkten Vergleich der Geschwindigkeiten in Tabelle [3.4](#) fallen sowohl beim Mittelwert als auch beim Median zwei Länder auf. Die Durchschnittsgeschwindigkeit in Deutschland beträgt 36,57 km/h und ist damit höher als in allen anderen Ländern. Deutschland weist auch die höchste Standardabweichung auf. China hingegen ist mit einer Durchschnittsgeschwindigkeit von 24,81 km/h und einem Median von 20,61 km/h das Land mit den langsamsten Fahrten in dieser Studie. Im weltweiten Vergleich liegt die mittlere Geschwindigkeit in China, den Vereinigten Staaten und dem Vereinigten Königreich unter dem Durchschnitt. Deutschland liegt mit einem um 7,6 % höheren Mittelwert über dem weltweiten Durchschnitt. Die unter der Woche gemessene Durchschnittsgeschwindigkeit ist fast immer niedriger als die am Wochenende gemessene Geschwindigkeit. In Deutschland ist auch der Unterschied zwischen Wochentagen und Wochenenden am größten. In Italien sind die Geschwindigkeiten an Wochentagen und am Wochenende etwa gleich.

Die [ECDF](#) der Durchschnittsgeschwindigkeiten pro Fahrt zeigt einen Ausreißer [3.10](#). Die Verteilung in China weicht von der Verteilung in allen anderen Ländern ab und liegt im Hinblick auf die jeweiligen Anteile bei den niedrigen Geschwindigkeiten über allen anderen Kurven. Die höchsten Durchschnittsgeschwindigkeiten werden durchweg in Deutschland erreicht. Italien und die USA liegen sehr nahe an der deutschen Kurve, sind aber im Durchschnitt etwas langsamer. In diesem Vergleich sind die Fahrten im Vereinigten Königreich langsamer als die in Deutschland, Italien und den Vereinigten Staaten.

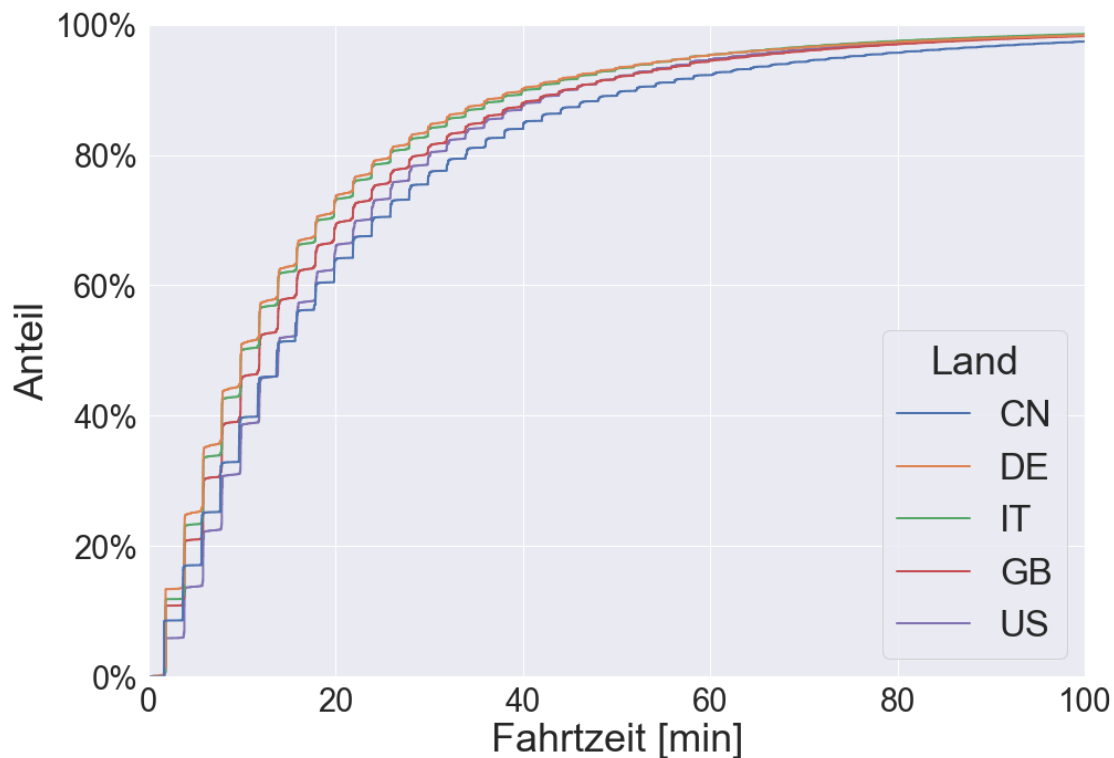


Abbildung 3.8: Empirische kumulative Verteilungsfunktion der gemessenen Fahrtauern pro untersuchtem Land.

Die Abbildung 3.11 zeigt die stündliche Verteilung der gefahrenen Durchschnittsgeschwindigkeiten für den gesamten Datensatz in Deutschland. Das Bild zeigt auch, dass zwischen 8 und 20 Uhr die Verteilungen der gefahrenen Durchschnittsgeschwindigkeiten pro Stunde homogenere Streuungen aufweisen als zu den restlichen Uhrzeiten. Die Boxen, die das 25- und das 75-Prozent-Quantil der Verteilungen markieren, sind in diesem Zeitraum einheitlich. Die Mediane zwischen Tag und Nacht unterscheiden sich im Extremfall um mehr als 10 km/h. Die Ausprägungen sind in den frühen Morgenstunden etwas größer als etwas später am Tag. Auch hier ist wieder bemerkbar, dass die gefahrenen Geschwindigkeiten zwischen 3 Uhr und 6 Uhr eine größere Streuung besitzen. Im Fall der Geschwindigkeit lässt sich auch zeigen, dass zwischen 8 und 20 Uhr, die Durchschnittsgeschwindigkeiten tendenziell einheitlicher und langsamer, während sie zwischen 21 und 7 Uhr schneller und variabler sind.

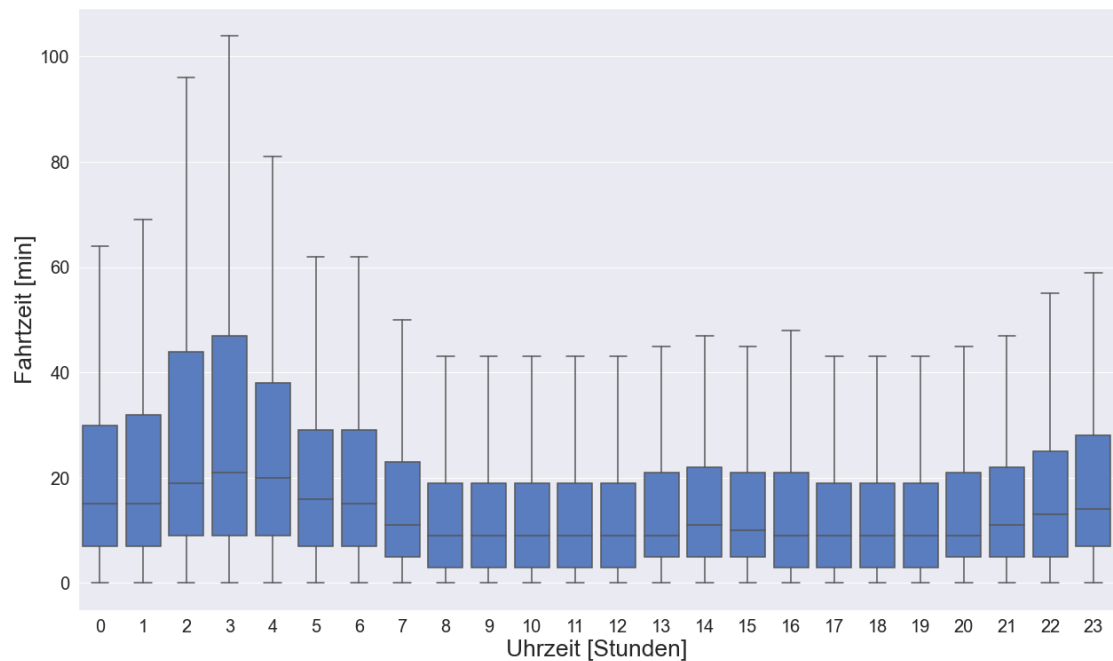


Abbildung 3.9: Tagesverlauf der gefahrenen Reisezeiten in Deutschland.

	Deutschland	USA	China	UK	Italien	Global
Mittel $\frac{km}{h}$	36,57	34,19	24,81	32,43	34,60	33,99
SD $\frac{km}{h}$	23,48	20,65	17,92	19,35	21,28	21,92
Median $\frac{km}{h}$	31,30	30,53	20,61	28,63	30,70	29,10
25 % - 75 % Quantil $\frac{km}{h} - \frac{km}{h}$	19,98– 48,10	19,53– 44,40	13,94– 30,11	19,38– 41,97	18,76– 45,12	18,46– 44,43
99 % Quantil $\frac{km}{h}$	114,49	99,49	100,00	94,15	106,63	106,59
Mittel Mo. - Fr. $\frac{km}{h}$	35,93	33,23	23,86	31,94	34,24	33,03
Mittel Weekende $\frac{km}{h}$	37,64	34,92	25,08	32,73	34,28	34,26

Tabelle 3.4: Zusammenfassung der mittleren Fahrtgeschwindigkeit pro Fahrt.

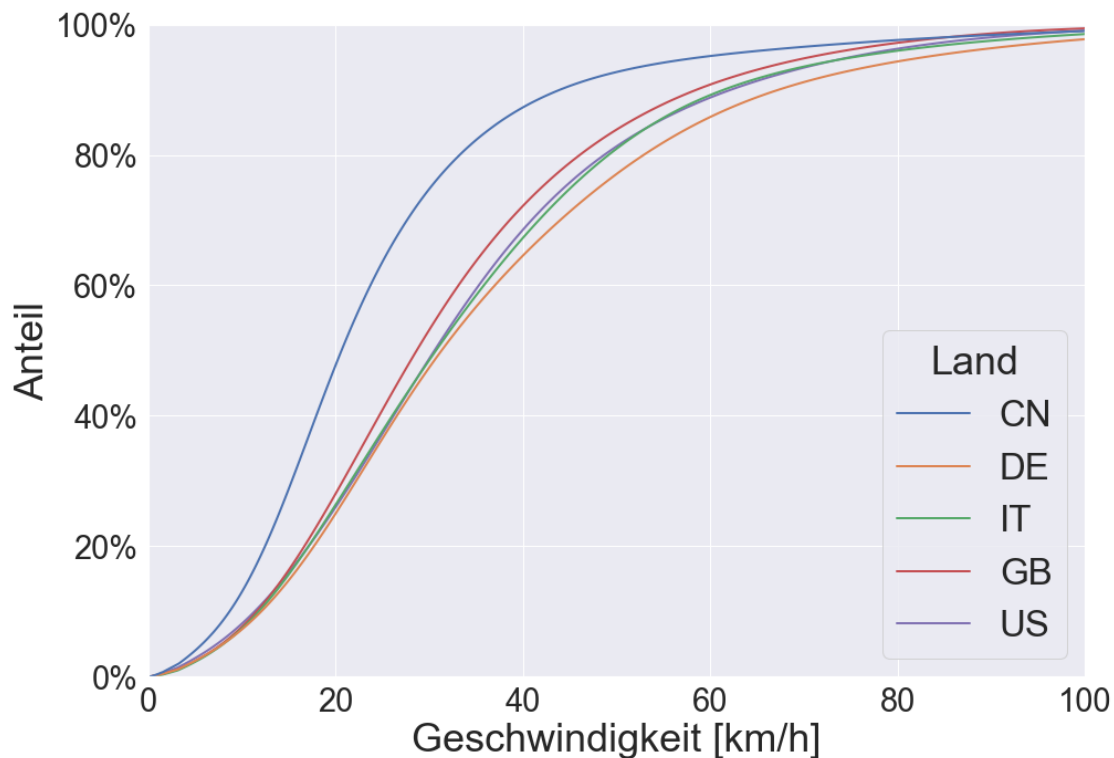


Abbildung 3.10: Empirische kumulative Verteilungsfunktion der gemessenen mittleren Geschwindigkeiten pro untersuchtem Land.

3.2.5 Tagesganglinien

Die Tagesganglinien des Fahrzeugverkehrs beschreiben die Schwankungen des Verkehrsaufkommens im Laufe eines Tages. Diese Muster geben einen Einblick in die Hauptverkehrszeiten, in das Verkehrsaufkommen und in die Herausforderungen, die damit verbunden sind. Die Tagesganglinien des Fahrzeugverkehrs können von Land zu Land, von Stadt zu Stadt und von Region zu Region sehr unterschiedlich sein. In der Abbildung 3.12 sind die Tagesganglinien für eine Reihe von Ländern aus dem Datensatz dargestellt.

Die Abbildung zeigt einige gemeinsame Muster, die weltweit zu beobachten sind. In (a) ist das Verkehrsverhalten von Montag bis Freitag dargestellt. (b) zeigt den zeitlichen Verlauf an Samstagen und Sonntagen. Typischerweise treten Montag bis Freitag morgendliche und abendliche Spitzenzeiten auf. Diese stehen im Zusammenhang mit dem Berufsverkehr und den Pendlerströmen. Während der morgendlichen Hauptverkehrszeit steigt das Verkehrsaufkommen in allen Ländern an, wenn die Pendler anfangen zu arbeiten. In den meisten Ländern liegt der Höhepunkt des

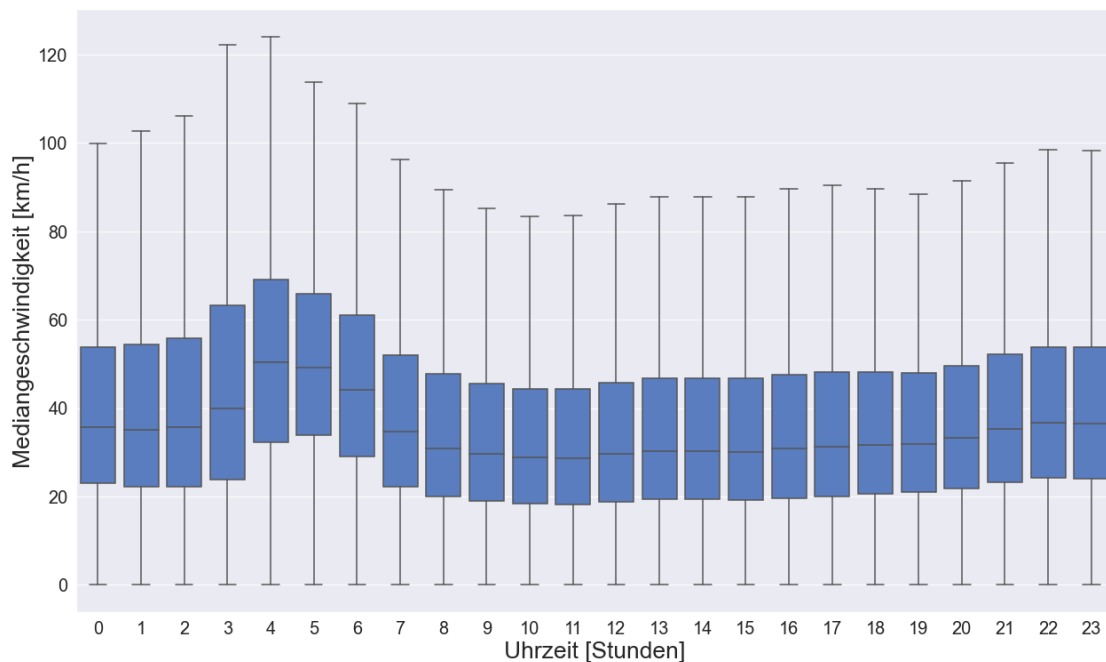
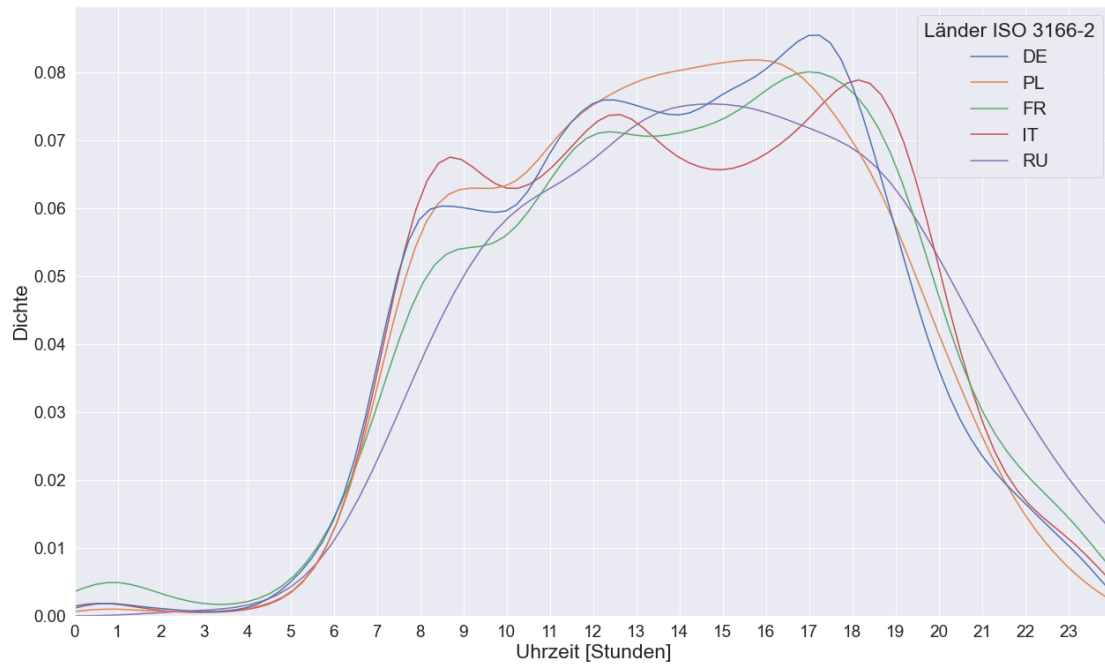


Abbildung 3.11: Tagesverlauf der gefahrenen Durchschnittsgeschwindigkeit in Deutschland.

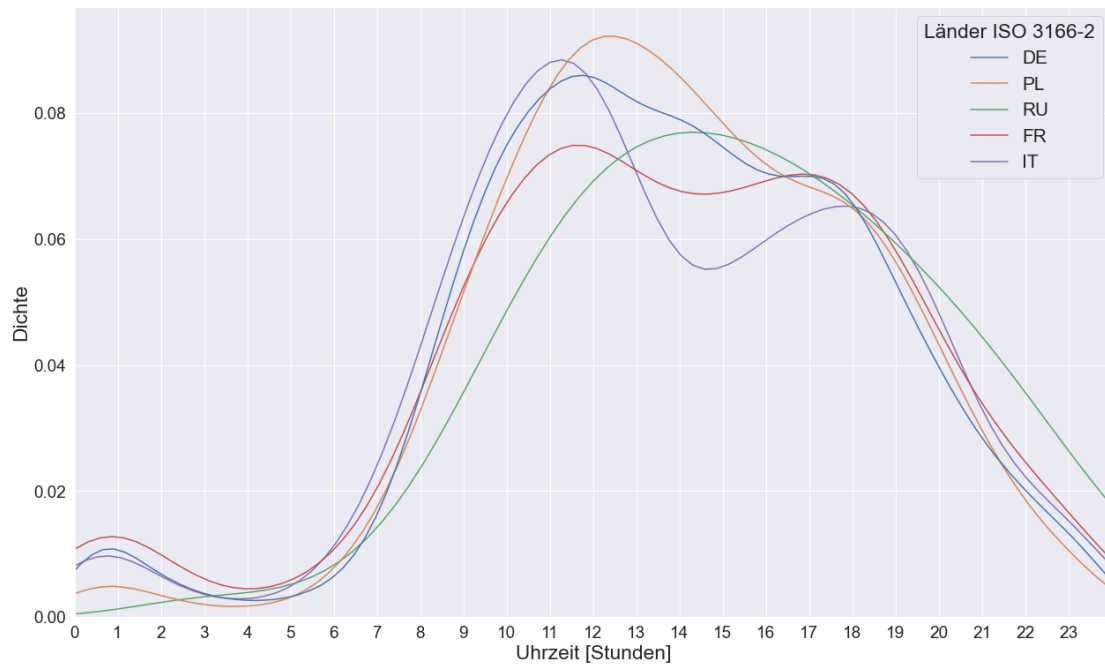
Berufsverkehrs zwischen 7 und 9 Uhr morgens. In [Abbildung 3.12\(a\)](#) ist für die meisten Länder eine Morgenspitze zwischen 8 und 9 Uhr zu erkennen. In Russland ist das Auftreten dieses ersten Plateaus etwas später. Ab etwa 5 Uhr morgens steigt die morgendliche Verkehrsnachfrage in allen Ländern stark an. Zu dieser Zeit sind die Straßen besonders stark befahren und es kann zu Staus und längeren Fahrzeiten kommen. Zwischen 16 und 18 Uhr, wenn sich die Pendler auf den Heimweg machen, treten in der Regel die abendlichen Spitzenzeiten auf. Zu dieser Zeit steigt das Verkehrsaufkommen wieder an, und es kann erneut zu Staus und zu einer Verlängerung der Fahrzeiten kommen. Es ist wichtig zu beachten, dass die Tagesganglinien des PKW-Verkehrs auch durch andere Faktoren beeinflusst werden können. Ereignisse wie Schulzeiten, Einkaufszeiten, Wetterbedingungen, Feiertage und besondere Veranstaltungen können das Verkehrsaufkommen beeinflussen und zu zusätzlichen Spitzenzeiten führen. So ist z. B. in den meisten Ländern ein weiterer Anstieg des Verkehrsaufkommens in der Zeit zwischen 11 Uhr und 13 Uhr zu beobachten. In [3.12\(b\)](#) ändert sich das Nachfrageverhalten der Flottenfahrzeuge. An Samstagen und Sonntagen tritt die erste Spitze zwischen 11 und 13 Uhr auf. Das ist die Zeit, zu der die meisten Benutzeranfragen im Laufe des Tages gestellt werden. In Russland ist die Tagesspitze verschoben und tritt erst gegen 14 Uhr auf. In den meisten Ländern stellt die Mittagszeit auch die höchste Nachfrage des

Tages dar. Lediglich in Italien und Frankreich wird eine zweite Nachfragespitze am Abend gegen 17 und 18 Uhr beobachtet.

Außerdem ist sichtbar, dass die Fahrzeuge international in dem Zeitraum zwischen 0 und 4 Uhr, im Vergleich zum restlichen Tag, kaum genutzt werden.



(a)



(b)

Abbildung 3.12: Durchschnittliche Fahrzeuganfragen der Flottenfahrzeuge der verschiedenen Länder. (a) Montag bis Freitag (b) Samstag & Sonntag

3.3 Vergleich zu Mobilitätsumfragen

In der Tabelle 3.5 werden Mobilitätskennziffern aus bereits veröffentlichten Mobilitätserhebungen mit den Ergebnissen der vorliegenden Untersuchung verglichen. Dabei fällt auf, dass je nach Erhebung eine Vielzahl von Parametern gemessen wurde. In diesem Beitrag wurden nur Ergebnisse aus den Erhebungen verwendet, die auch in der vorliegenden Untersuchung gemessen werden konnten. Teilweise wurden in den Erhebungen Jahres- oder Tageswerte angegeben. In diesen Fällen wurden die Werte auf die hier berechneten Vergleichswerte bezogen. Die entsprechenden Berechnungen sind bei der jeweiligen Quelle angegeben. Für die Befüllung der Spalte Mobilitätserhebungen wurden die aktuellen Erhebungen und Ergebnisse verwendet, die zum Zeitpunkt der Erstellung dieser Arbeit verfügbar waren. Im Falle der USA stammt die jüngste verfügbare Erhebung aus dem Jahr 2017.

In der Tabelle 3.5 werden die Ergebnisse der veröffentlichten Erhebungen den Ergebnissen dieser Untersuchung direkt gegenübergestellt. Da der Fokus der verschiedenen Erhebungen nicht der gleiche war wie in dieser Untersuchung, wurden

Land	Kennzahl	Mobilitätsumfragen	Diese Untersuchung
Deutschland	Distanz [km]	14,75 ¹	13,91
	Fahrtzeit[mm : ss]	20:35 ¹	17:56
USA	Distanz [km]	15,37 ²	14,31
	Fahrtzeit[mm : ss]	-	21:18
UK	Distanz [km]	12,64 ³	12,73
	Fahrtzeit[mm : ss]	-	19:50
Italien	Distanz [km]	12,20 ⁴	13,37
	Fahrtzeit[mm : ss]	21:48 ⁴	18:04
China	Distanz [km]	12,39 ⁵	12,20
	Fahrtzeit[mm : ss]	-	23:14

¹ $\frac{d_{PKW}}{n_{Wege}}$	4-4, 4-6, 4-8 [60]
²	Tab. 3b [67]
³ $\frac{d_{Driver}}{n_{Trips}}$	Chart 7 [68]
⁴	Table 4.24 [69]
⁵	Table 4 [70]

Tabelle 3.5: Vergleich der ermittelten Werte aus der Literatur mit den Ergebnissen dieser Arbeit.

nicht alle Indikatoren gemessen und aufgrund der Messmethode wurde z.B. die Geschwindigkeit nicht berücksichtigt. Außerdem ist zu beachten, dass es sich bei den Vergleichen immer um Erhebungen handelt, deren Ergebnisse auf den Angaben und der Ehrlichkeit der vielen freiwilligen Teilnehmer beruhen. Hinzu kommt, dass die Stichproben deutlich kleiner sind, als dies bei der Erfassung der Positionsdaten einer Flotte möglich ist. Als Vergleichswert wurden beispielsweise im deutschen Mobilitätspanel (MOP) im Zeitraum von September 2020 bis März 2021 63.372 Wege dokumentiert, die sich aus unterschiedlichsten Verkehrsmitteln zusammensetzen. Die landesweite Erhebung National Household Travel Survey des Department of Transportation der Vereinigten Staaten von Amerika aus dem Jahr 2017 hatte für die Auswertung der Wege mit dem Pkw einen Stichprobenumfang von 611.342 Fahrten. Im Durchschnitt wurden in dieser Erhebung in Deutschland fast 2 Millionen Wege pro Tag systematisch erfasst. Beim Vergleich der Mobilitätserhebungen mit der vorliegenden Untersuchung fällt jedoch auf, dass die Ergebnisse dennoch meist in einer ähnlichen Größenordnung liegen. Die Dauer einer Fahrt wird in den anderen Studien oft nicht gemessen und ist daher in der Tabelle nicht immer angegeben. Aufgrund der hohen Standardabweichung der jeweiligen Kennzahlen ist es möglich,

durch eine kleinere Stichprobe und eine andere Verteilung der Teilnehmer von den Durchschnittswerten abzuweichen.

3.4 Clustering von Verteilungen der Fahrtparameter

Die Clusteranalyse ist ein wichtiger Ansatz, um Muster und Zusammenhänge zu identifizieren. Ein Beispiel für die Anwendung von Clusteranalyse in der Verkehrsforschung ist die Gruppierung von Verkehrsteilnehmern nach ihrem Verhaltensmuster. Durch die Analyse von Fahrtdaten wie Distanz, Fahrtdauer und der mittleren Geschwindigkeit werden Länder mit ähnlichen Fahrteigenschaften ermittelt und gruppiert. Aus den verschiedenen ECDF-Diagrammen wird deutlich, dass sich die Verteilungen einiger Länder von den Verteilungen anderer Länder für jede der Fahrmetriken unterscheiden. Einige ECDF-Kurven sind nahezu identisch, während sich andere deutlich von der Mehrheit der anderen Länder unterscheiden. Daher wird die Ähnlichkeit der Verteilungen zwischen den Ländern untersucht. Im folgenden Abschnitt wird das Clustering der Fahrtparameter erläutert. Dabei wird die Wahl der Ähnlichkeitsmetrik beschrieben. Anschließend wird das Clusterverfahren an den Fahrtparameter: Distanz, Fahrtdauer und der mittleren Geschwindigkeit angewendet und präsentiert.

3.4.1 Ähnlichkeitsmetrik

Eine häufig verwendete Methode zur Messung der Ähnlichkeit zwischen zwei Verteilungen ist der Kolmogorov-Smirnov-Test (KS) [71]. Der KS-Test ist ein nichtparametrischer statistischer Test, der den Abstand zwischen den empirischen Verteilungsfunktionen von zwei Stichproben misst. Der KS-Test basiert auf der maximalen Differenz zwischen den kumulativen Verteilungsfunktionen (CDFs) der beiden zu vergleichenden Verteilungen. Eine weitere Methode zur Messung der Ähnlichkeit zwischen zwei Verteilungen ist die Jensen-Shannon-Divergenz (JSD) [72]. Die JSD ist eine symmetrische und geglättete Version der Kullback-Leibler-Divergenz (KLD) und misst den Abstand zwischen zwei Wahrscheinlichkeitsverteilungen [73]. Die JSD ist eine in der Informationstheorie und im maschinellen Lernen häufig verwendete Methode. Eine dritte Methode zur Messung der Ähnlichkeit zwischen zwei Verteilungen ist die Bhattacharyya-Distanz. Der Bhattacharyya-Abstand ist ein Maß für die Überlappung zwischen zwei Verteilungen und basiert auf dem Bhattacharyya-Koeffizienten. Der Bhattacharyya-Abstand wird häufig in der Bildverarbeitung und Mustererkennung verwendet [74]. Andere Methoden zur Messung

der Ähnlichkeit zwischen zwei Verteilungen sind die Wasserstein-Distanz, die Gesamtvariationsdistanz und der Korrelationskoeffizient nach Pearson. Die Wahl der Methode zur Messung der Ähnlichkeit zwischen zwei Verteilungen hängt von dem spezifischen Problem und der Art der zu analysierenden Daten ab.

Um die Ähnlichkeit zweier Verteilungen zu messen, wird die Wasserstein-Distanz [75] verwendet. Für eindimensionale Verteilungen kann die Wasserstein-Distanz auf die folgende Gleichung reduziert werden 3.4, wobei F_{c_1} und F_{c_2} beispielsweise die empirischen kumulativen Verteilungsfunktionen für die Distanzen der exemplarischen Länder $y_{c_1} = [d_{T_{0,c_1}}, d_{T_{1,c_1}}, \dots, d_{T_{n,c_1}}] \in \mathbb{R}^{1 \times n_{c_1}}$ und $y_{c_2} = [d_{T_{0,c_2}}, d_{T_{1,c_2}}, \dots, d_{T_{n,c_2}}] \in \mathbb{R}^{1 \times n_{c_2}}$ darstellen. Diese Methode hat den Vorteil, dass die beiden Vektoren y_{c_1} und y_{c_2} nicht gleich groß sein müssen, damit sie verglichen werden können.

$$d_{c_1,c_2} = W_1(y_{c_1}, y_{c_2}) = \int_{\mathbb{R}} |F_{c_1}(y) - F_{c_2}(y)| dx \in [0, \infty) \quad (3.4)$$

In diesem Fall misst die Wasserstein-Distanz die Fläche zwischen zwei ECDF-Kurven. Für jede Fahrtmetrik wird ein Abstand zwischen jedem Land und jedem anderen Land berechnet und in einer quadratischen Unähnlichkeitsmatrix $S \in \mathbb{R}^{n \times n}$ gespeichert, die in Gleichung 3.5 definiert ist. Um aussagekräftigere Ergebnisse zu erhalten, werden die 20 am stärksten vertretenen Länder des Datensatzes verwendet ($n = 20$). Da es keine Fläche zwischen zwei identischen Kurven gibt, ist der Abstand zwischen der Verteilung eines Landes und sich selbst immer der kürzest mögliche Abstand. Bei der Messung des Abstands kommt es daher immer vor, dass die Unähnlichkeitsmatrix S eine Diagonale mit Nullen aufweist. Die Matrix ist außerdem symmetrisch, da die $W_1(y_{c_1}, y_{c_2}) = W_1(y_{c_2}, y_{c_1})$. Die Obergrenze der Wasserstein-Distanz ist unbekannt, da die Fläche zwischen zwei Kurven unendlich viel Fläche einnehmen kann. Daher wird jeder Eintrag in der Unähnlichkeitsmatrix so normiert, dass er zwischen $[0, 1]$ liegt, wobei 1 den Abstand zwischen den beiden unähnlichsten Ländern darstellt. Da eine Distanz gemessen wird und mit zunehmender Entfernung die Unähnlichkeit auch zunimmt, spricht man bei der Matrix S von einer Unähnlichkeitsmatrix.

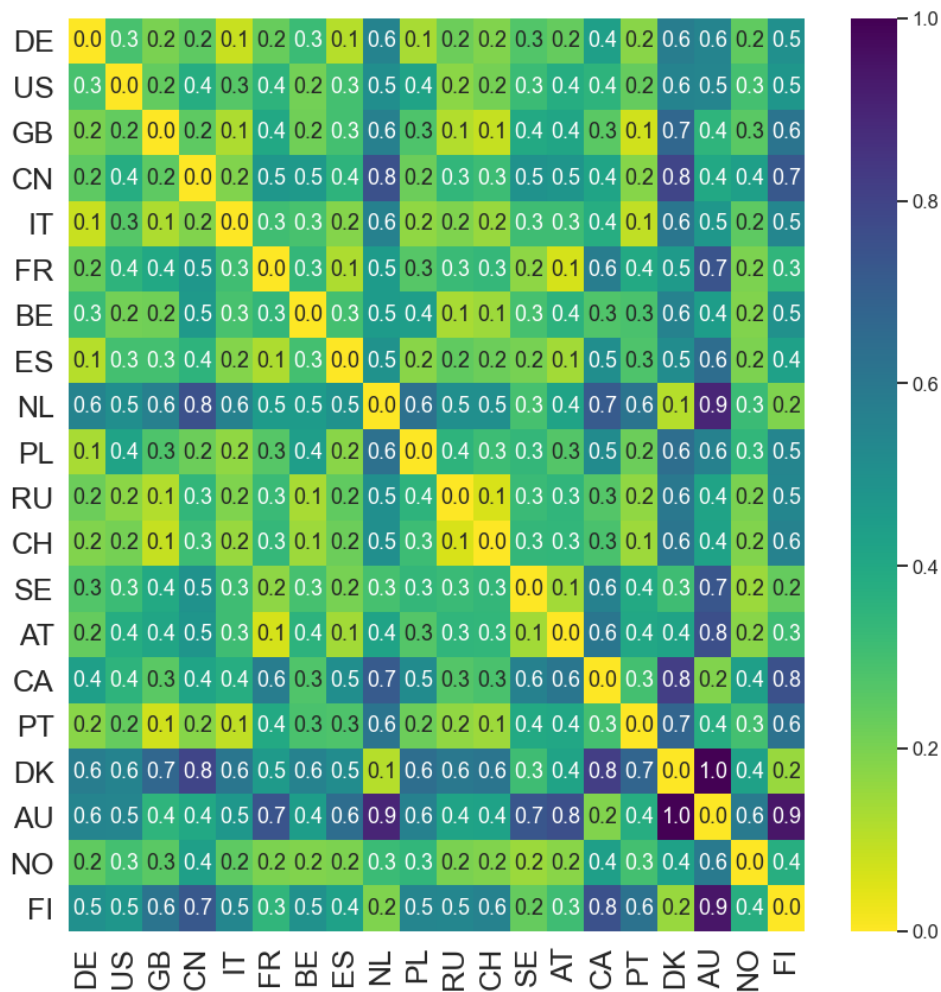
$$S = \begin{bmatrix} 0 & d_{c_1, c_2} & \dots & d_{c_1, c_{n-1}} & d_{c_1, c_n} \\ d_{c_2, c_1} & 0 & \dots & d_{c_2, c_{n-1}} & d_{c_2, c_n} \\ \dots & \dots & \dots & \dots & \dots \\ d_{c_n, c_1} & d_{c_n, c_2} & \dots & d_{c_n, c_{n-1}} & 0 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

mit $c : \text{Land}$
und $d_{c_n, c_n} = 0$

(3.5)

3.4.2 Clustering von Ländern

Die Ähnlichkeitsmetrik wird verwendet, um die Unähnlichkeitsmatrix $S_{distance}$ für die Fahrtmetrik Distanz zu berechnen. Das Resultat dieser Matrix ist in Abbildung 3.13 als Heatmap dargestellt. Jedes Element der Matrix stellt die Unähnlichkeit zweier Länder dar. In dieser Darstellung von $S_{distance}$ sind die Unähnlichkeiten zwischen zwei Ländern sichtbar. Je dunkler der Bereich in der Heatmap ist, desto größer ist der Abstand zwischen den Verteilungen der beiden Länder. Die Niederlande, Dänemark und Finnland sind zum Beispiel die unähnlichsten Länder, da ihre Unähnlichkeit zu fast allen anderen Ländern größer als 0,5 ist. Andere Länder, wie z. B. Deutschland und Italien oder das Vereinigte Königreich und Portugal, sind sich ähnlicher.

Abbildung 3.13: Heatmap der Unähnlichkeitsmatrix für Länder $S_{distance}$.

Ein Dendrogramm kann zeigen, wie verwandt alle Länder für die gegebene Unähnlichkeitsmatrix S sind, wenn die Unähnlichkeitsmatrix S in dieser Form angeordnet ist, wie in Abbildung 3.14 gezeigt. Die Unähnlichkeit wird durch die y-Achse im Dendrogramm dargestellt. Je höher der Punkt ist, an dem zwei Länder ineinander übergehen, desto unähnlicher sind sie. Frankreich und Österreich sind die beiden Länder mit der größten Ähnlichkeit in Bezug auf den Abstand. In einem Dendrogramm kann an jedem beliebigen Punkt eine horizontale Linie gezogen werden, um Elemente von sich kreuzenden Linien zu Clustern zusammenzufassen. Alle Länder werden in einem einzigen Cluster zusammengefasst, wenn eine horizontale Linie am höchsten Punkt des Dendrogramms gezogen wird. Eine horizontale Linie am niedrigsten Punkt ($d = 0$) würde dagegen für jedes Land einen einzelnen Cluster ergeben.

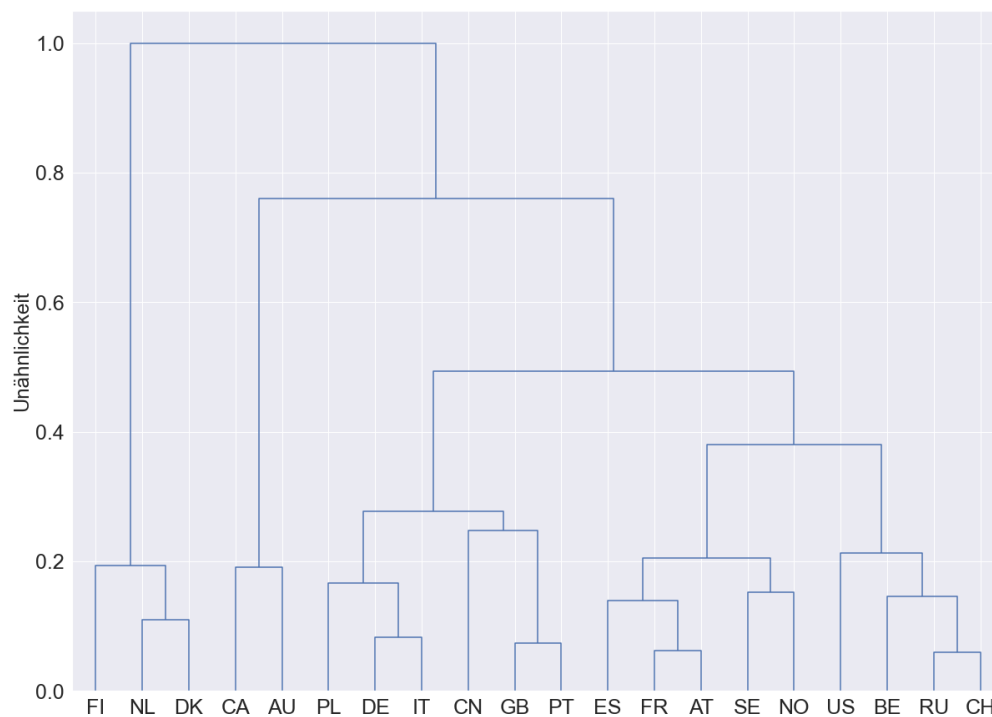
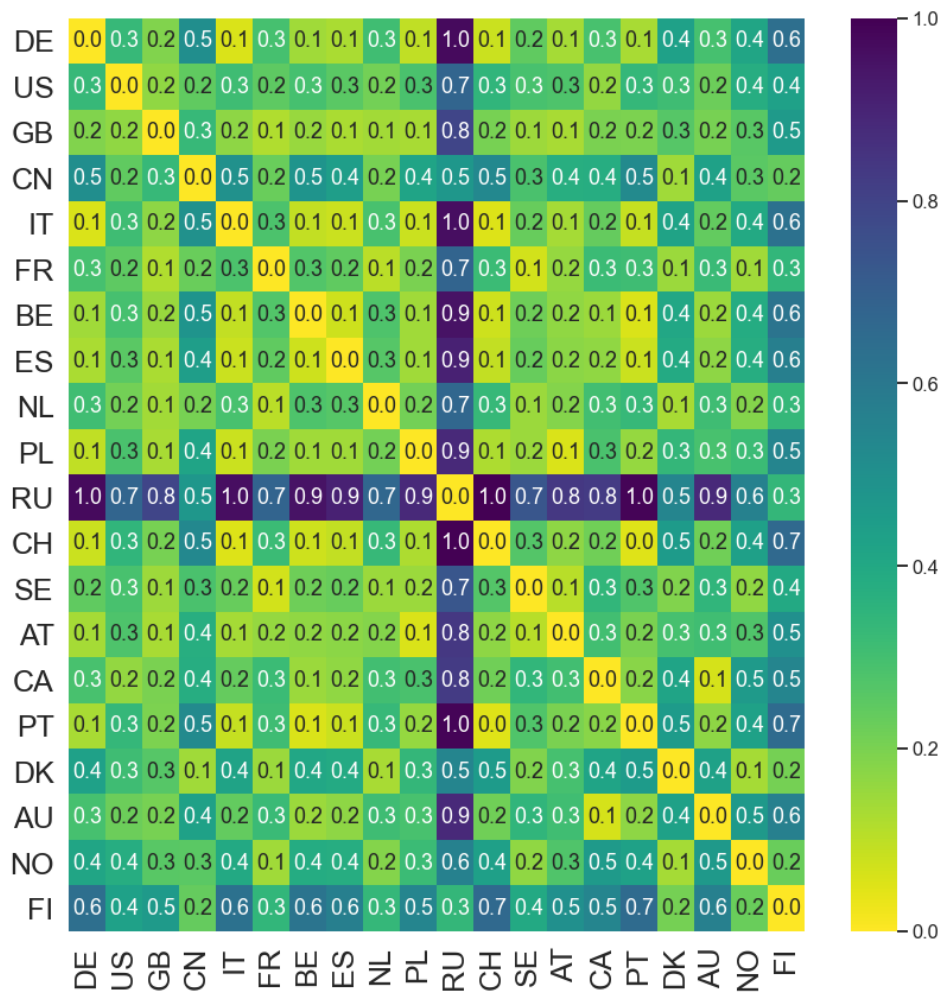


Abbildung 3.14: Dendrogramm der Unähnlichkeitsmatrix für Länder $S_{distance}$.

Wie bei der Distanz wird eine Unähnlichkeitsmatrix $S_{duration}$ berechnet, um die Ähnlichkeiten der Länder in Bezug auf ihre Reisedauer aufzuzeigen. Die vollständige Beschreibung von $S_{duration}$ ist in [Abbildung 3.15](#) dargestellt. Russland ist ein Land, das sich deutlich von den anderen abhebt. Es ist das Land mit der größten Unähnlichkeit in dieser Analyse, mit einer Unähnlichkeit zu den meisten anderen Ländern, die größer als 0,7 ist. Finnland ist jedoch das Land, das der Verteilung in Russland am nächsten kommt.

Abbildung 3.15: Heatmap der Unähnlichkeitsmatrix für Länder $S_{duration}$.

Auch hier zeigt das Dendrogramm in Abbildung 3.16, wie verwandt die Länder sind. In Bezug auf Fahrtdauer sind sich die meisten Länder sehr ähnlich. Viele Länder haben mindestens einen ungleichen Partner mit einer Unähnlichkeit nahe Null. Russland ist die einzige Ausnahme. Das Dendrogramm, das in Abbildung 3.15 dargestellt ist, bestätigt, dass Russland nicht mit allen anderen Ländern fusionieren kann. Es ist das letzte Land, das sich dem Cluster aller anderen Länder anschließt, und es hat eine hohe Unähnlichkeit nahe eins.

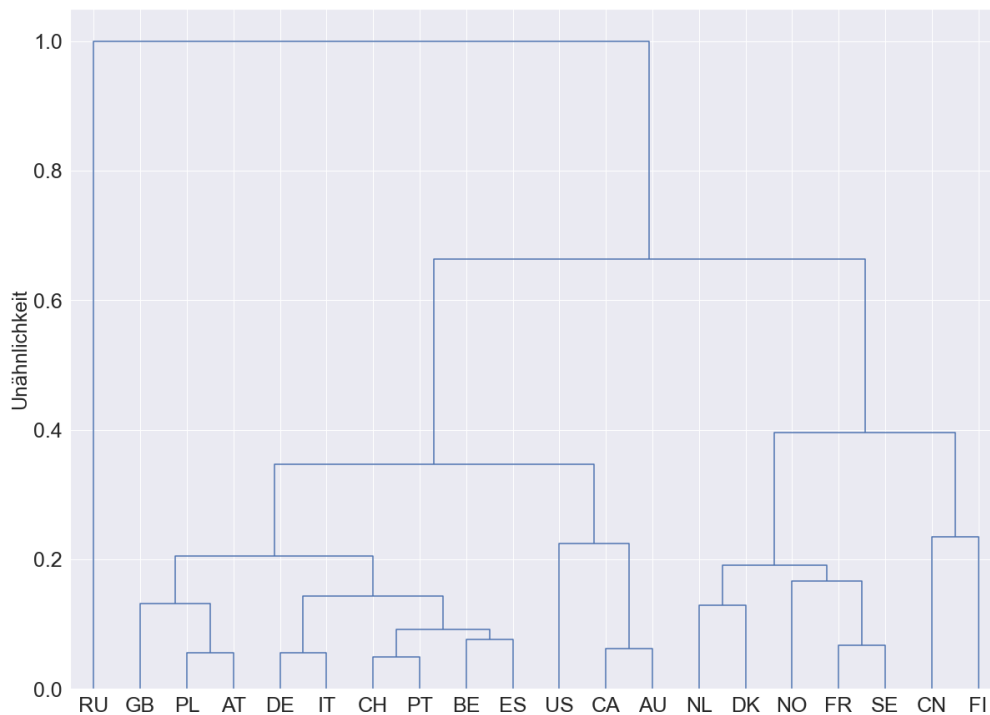
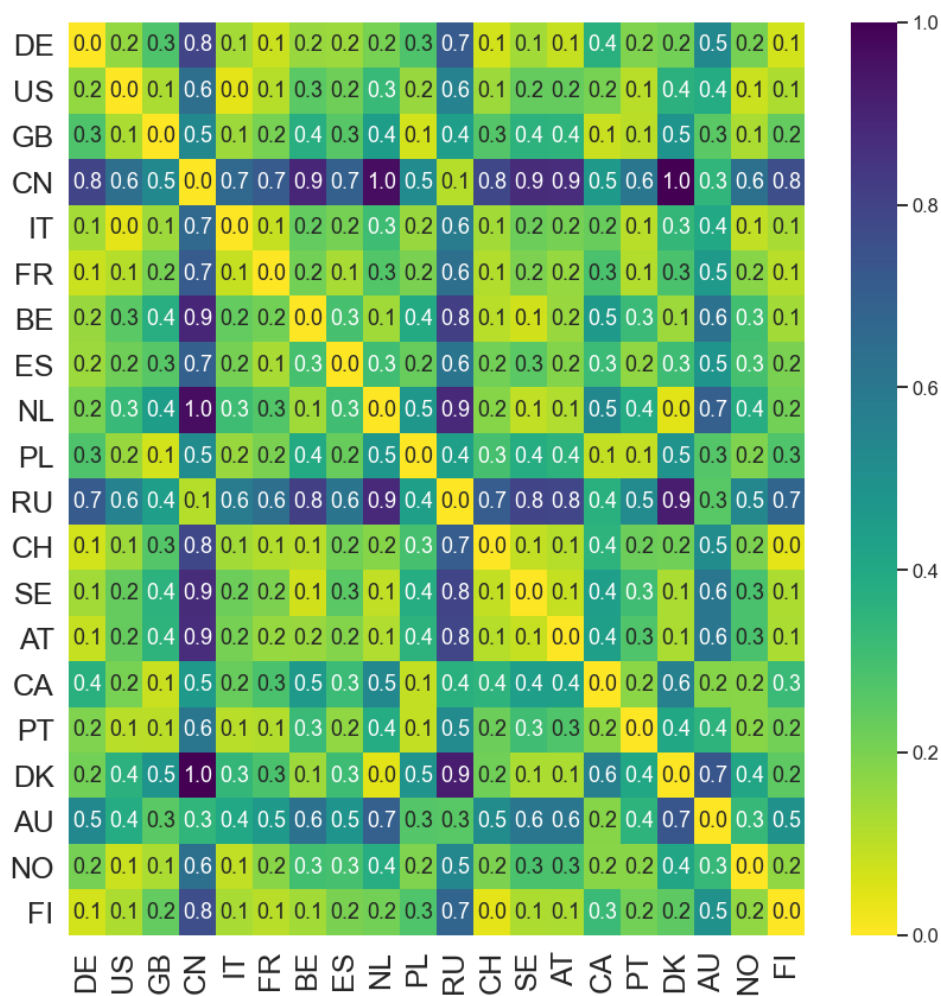


Abbildung 3.16: Dendrogramm der Unähnlichkeitsmatrix für Länder $S_{duration}$.

Anschließend wird die Unähnlichkeitsmatrix $S_{velocity}$ errechnet. Der Inhalt dieser Matrix ist in Abbildung 3.17 dargestellt. Hier stechen die Länder Russland und China am meisten hervor. Beide sind allen anderen Ländern sehr unähnlich, abgesehen von einer sehr engen Verbindung zueinander. Die detaillierten Beziehungen aller Länder für die mittlere Geschwindigkeit sind in Abbildung 3.18 dargestellt. Die Merkmale der Länder China und Russland sind ebenfalls dargestellt, da sie die letzten Länder sind, die in einem Cluster zusammengefasst wurden, der alle Länder mit einer größeren Unähnlichkeit als die meisten anderen Länder enthält.

Abbildung 3.17: Heatmap der Unähnlichkeitsmatrix für Länder $S_{velocity}$.

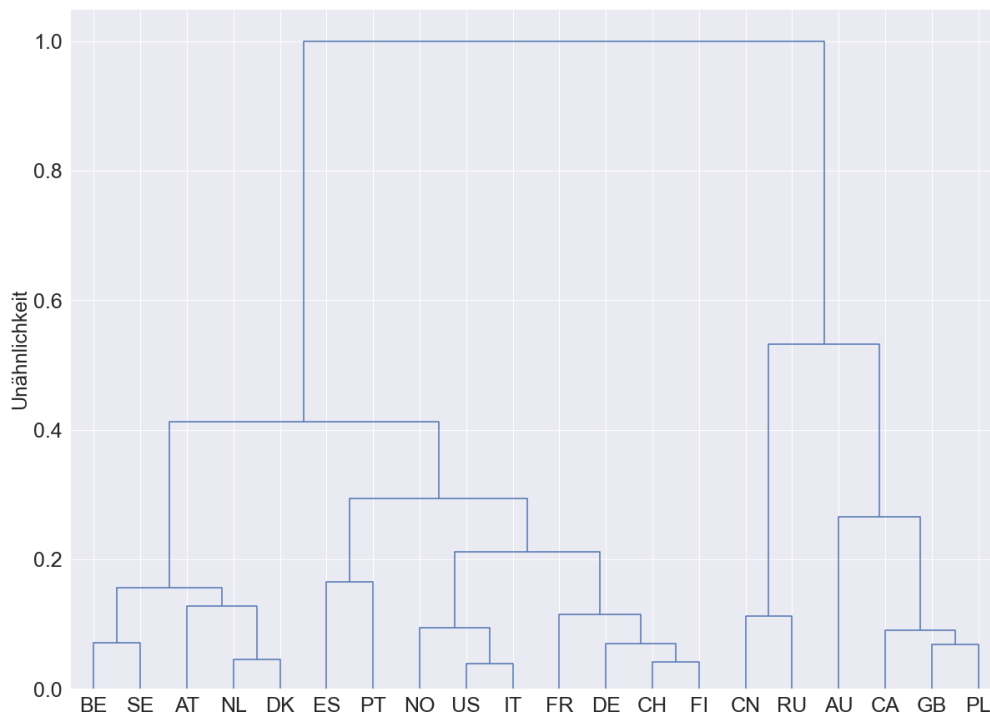


Abbildung 3.18: Dendrogramm der Unähnlichkeitsmatrix für Länder $S_{velocity}$.

3.4.3 Clustering von Städten

Ähnlich wie im vorherigen Vorgehen werden die Metriken für einzelne Städte bestimmt und auf deren Ähnlichkeit untersucht. Die Auswertungen über die gefahrene Distanz, Fahrtdauer und die Geschwindigkeit ähnlich zu den vergangenen Kapiteln werden im Anhang für eine Auswahl von Städten aufgeführt (siehe A.1). Zusätzlich werden die ECDF für die Fahrtparameter aufgezeigt (A.1, A.2, A.3). Die Auswahl für die Städte in der folgenden Untersuchung wurden aus den Räumen Nord Amerika, Europa und China jeweils 10 Städte ausgesucht, die unter den häufigsten in den Datensatz lagen.

Die Abbildung 3.19 zeigt das Dendrogramm der Unähnlichkeit der gefahrenen Distanzen Verteilungen verschiedener Städte. Hier beschreibt das Dendrogramm von oben nach unten eine immer feinere Aufspaltung in verschiedene Cluster aufgrund der Ähnlichkeit der Verteilungen der untersuchten Distanzmetrik. In Abbildung 3.20 wird exemplarisch gezeigt wie die Verteilungen für die gefundenen Cluster $C1$ und $C2$ in 3.20(a) und $C1$ bis $C5$ in (b) aussehen. Zunächst teilt sich bei einer Betrachtung von oben nach unten der Cluster, in dem sich alle Städte befinden in zwei verschiedene Cluster auf (siehe (a)). Der erste gefundene Cluster

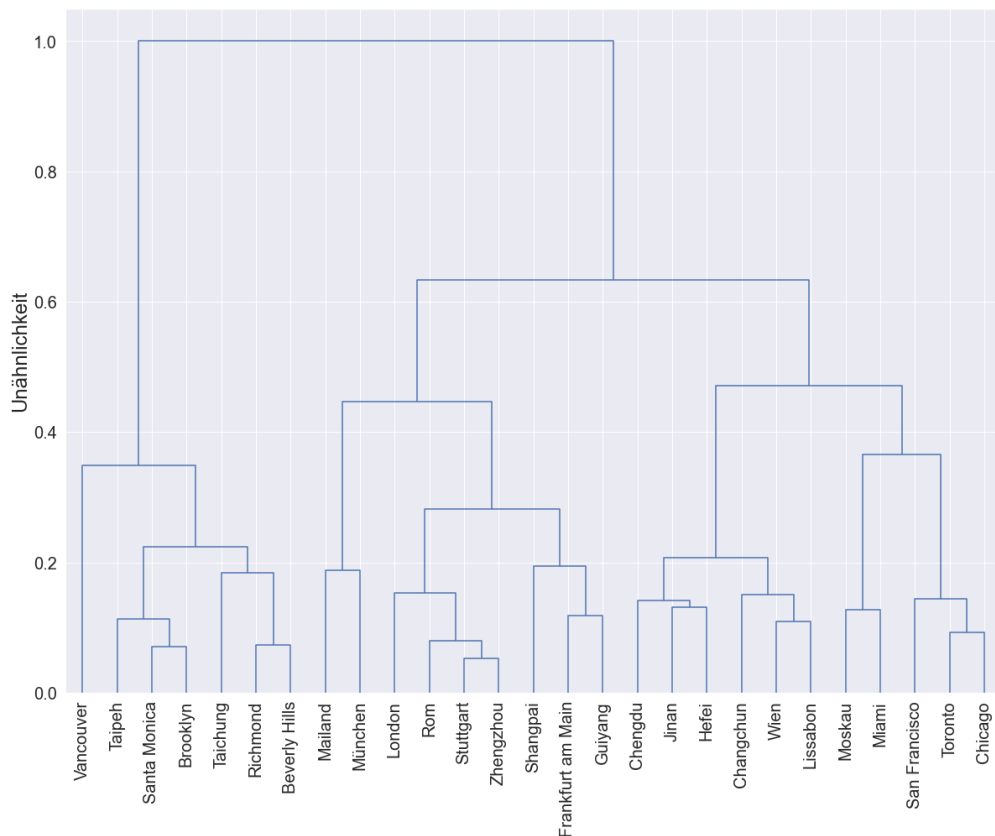
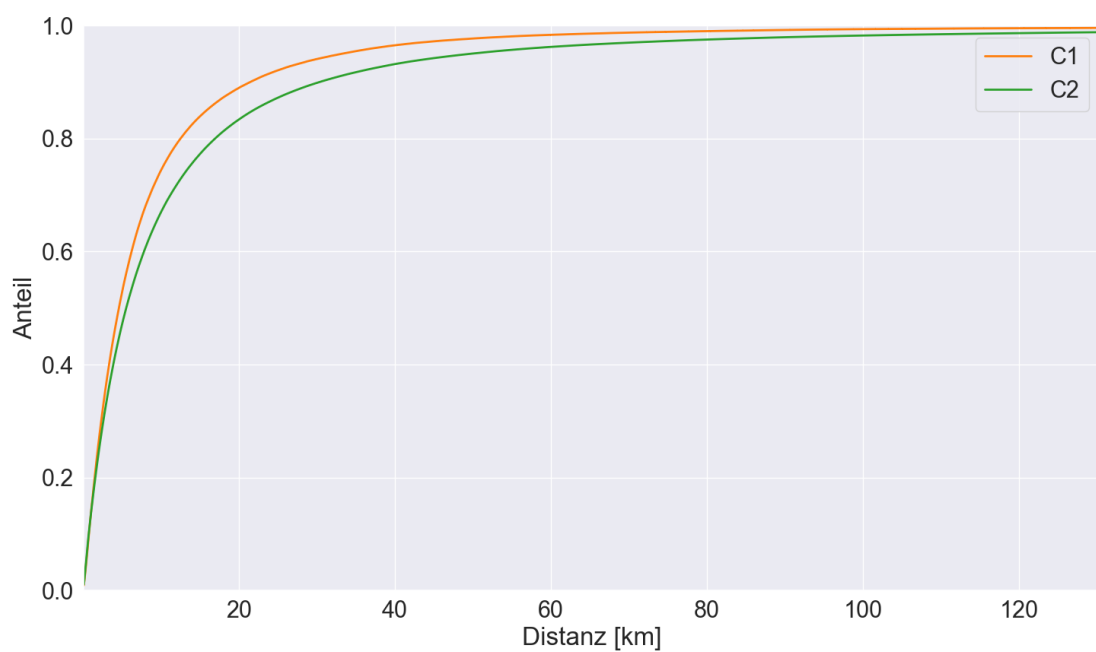
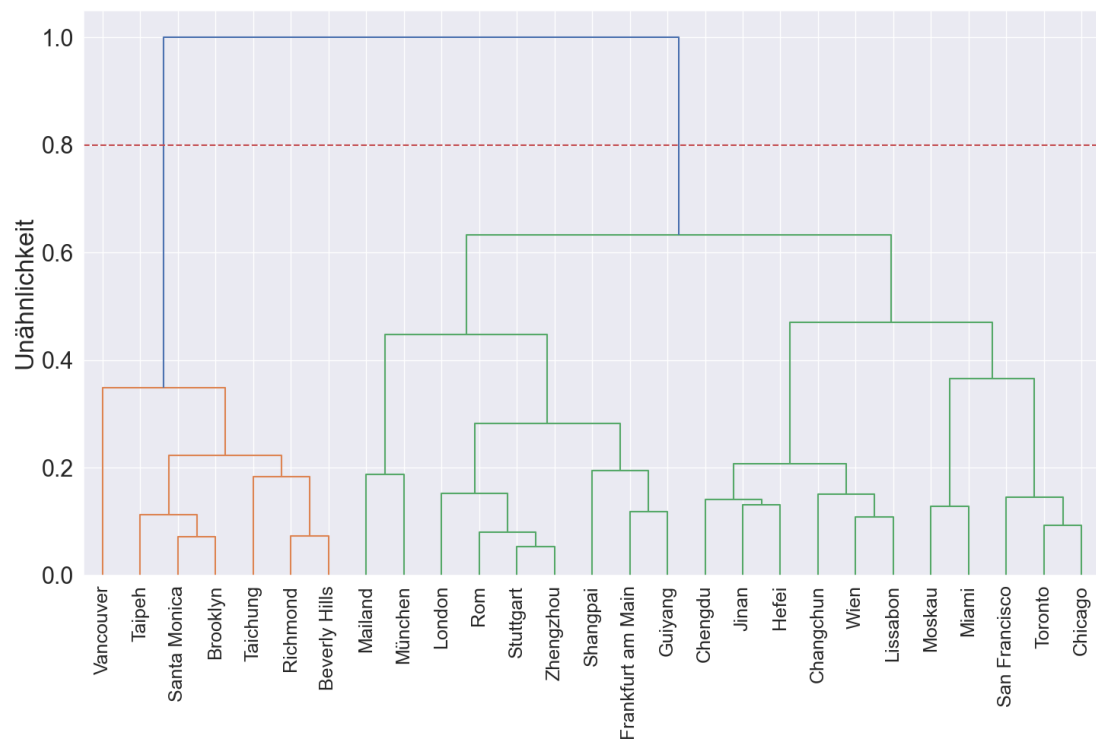


Abbildung 3.19: Dendrogramm der Unähnlichkeitsmatrix für Städte $S_{distance}$.

$C1$ von links beinhaltet Fahrten die kurze Strecken zurücklegen, während die Fahrten in $C2$ die Fahrten zeigen die größere Strecken zurücklegen. Setzt man einen Schnitt durch das Dendrogramm tiefer an, dann entstehen mehrere Cluster mit größerer intracluster Ähnlichkeit. Während der Betrachtung in 3.20(b) der Cluster $C1$ unverändert bleibt, teilt sich der vorherige Cluster $C2$ in weitere vier Cluster auf. Die linken beiden Cluster $C2$ und $C3$ enthalten hierbei die längsten Strecken und $C4$ und $C5$ eher mittelweite Strecken.

Abbildung 3.21 beschreibt die Unähnlichkeit der Verteilungen für die Metrik der Fahrtdauer. In dieser Abbildung fallen bei der Betrachtung von oben nach unten zwei markante Äste auf. Diese kennzeichnen sich dadurch, dass diese Äste deutlich länger sind als alle weiteren Äste. Die Unähnlichkeit zwischen diesen beiden Clustern ist groß, da ein großer Abstand überwunden werden muss, um einen Cluster zu formen, der alle Städte enthält. Darüber hinaus fällt auch auf, dass für den zweiten Cluster die Ähnlichkeiten einzelner Städte sehr groß sind und somit mit einer geringen Unähnlichkeit schon Cluster gefunden werden. Im ersten Cluster passen



(a)

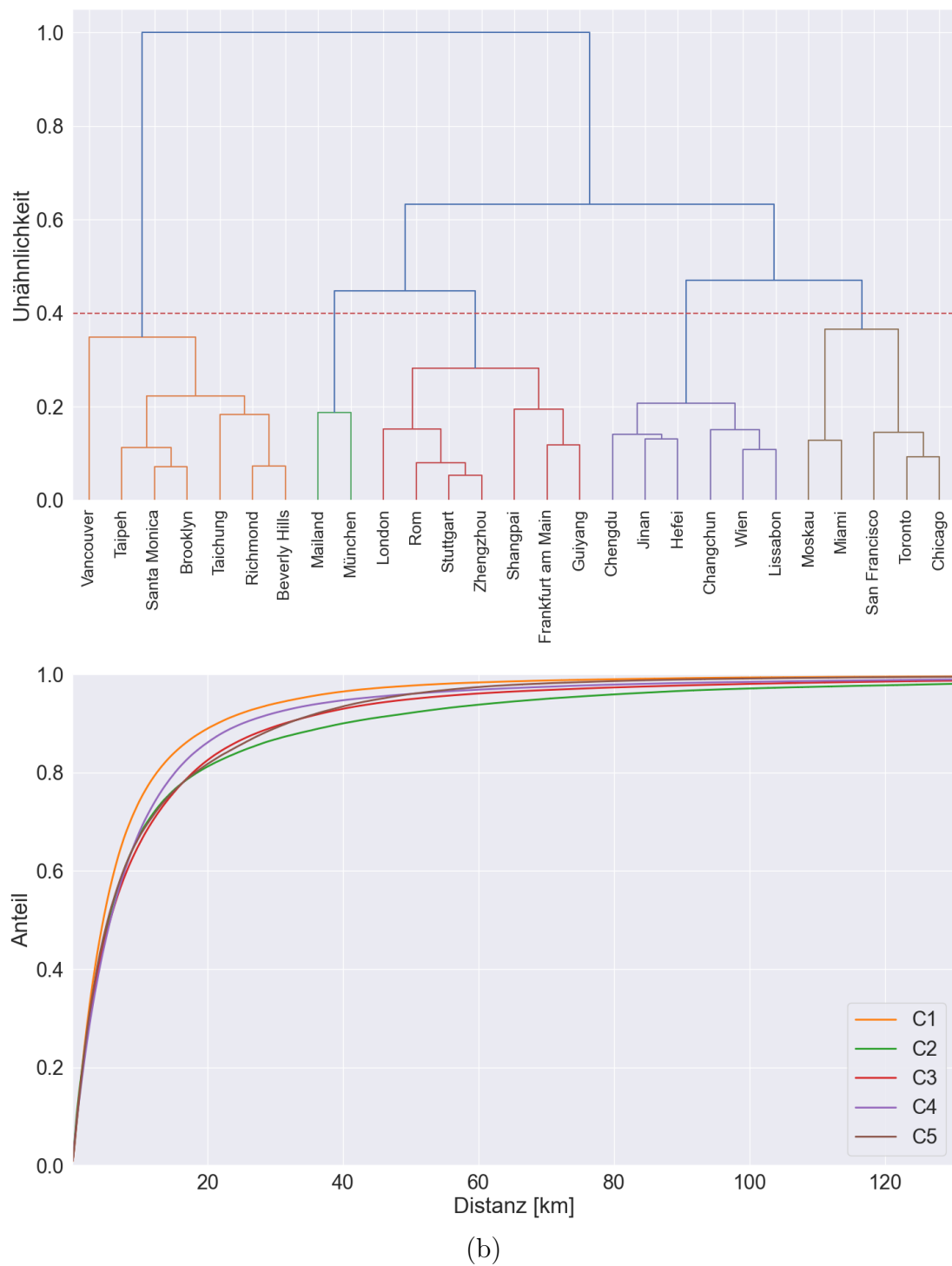


Abbildung 3.20: Dendrogramm der Fahrt dauern für Städte und der sich dadurch resultierenden Cluster von Städten. In (a) mit Schnitt bei einer Unähnlichkeit $d = 0.8$ und (b) bei $d = 0.4$.

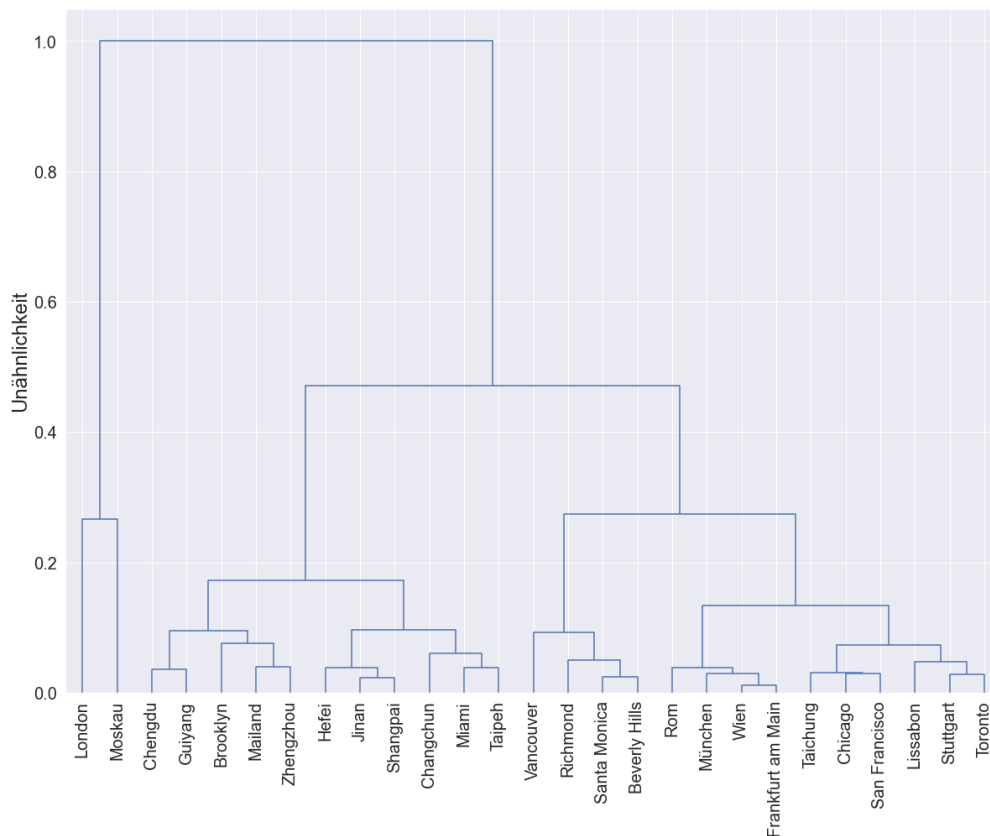


Abbildung 3.21: Dendrogramm der Unähnlichkeitsmatrix für Städte $S_{duration}$.

nur die Städte London und Moskau zusammen haben eine große Unähnlichkeit zu allen weiteren Städten. Eine detaillierte Betrachtung der geformten Cluster wird im Anhang aufgeführt [A.4](#), [A.5](#).

In dem Dendrogramm aus [Abbildung 3.22](#) ist die Unähnlichkeit der Verteilungen der Durchschnittsgeschwindigkeitmetrik zu sehen. Betrachtet man das Dendrogramm wieder von oben nach unten werden zunächst zwei unterschiedliche Gruppen gefunden. Dieser Unterschied ist durch die große Unähnlichkeit bemerkbar, die erreicht werden muss, um die beiden Cluster zusammen zu führen. Im ersten gefundenen Cluster von links befinden sich die Städte mit den langsamsten ermittelten Durchschnittsgeschwindigkeiten. Der zweite gefundene Cluster beinhaltet schnellere Fahrten, die sich in 3 weitere Gruppen aufteilen. Die schnellsten Fahrten sind in der Gruppe [Frankfurt am Main, Stuttgart, München, Lissabon, Miami] enthalten mit der Besonderheit, dass alle deutschen Städte sich in dieser Gruppe befinden. Wie auch bei der Betrachtung der Fahrtdauer der Städte wird eine detaillierte Betrachtung der gefundenen Clusterverteilungen im Anhang gegeben.

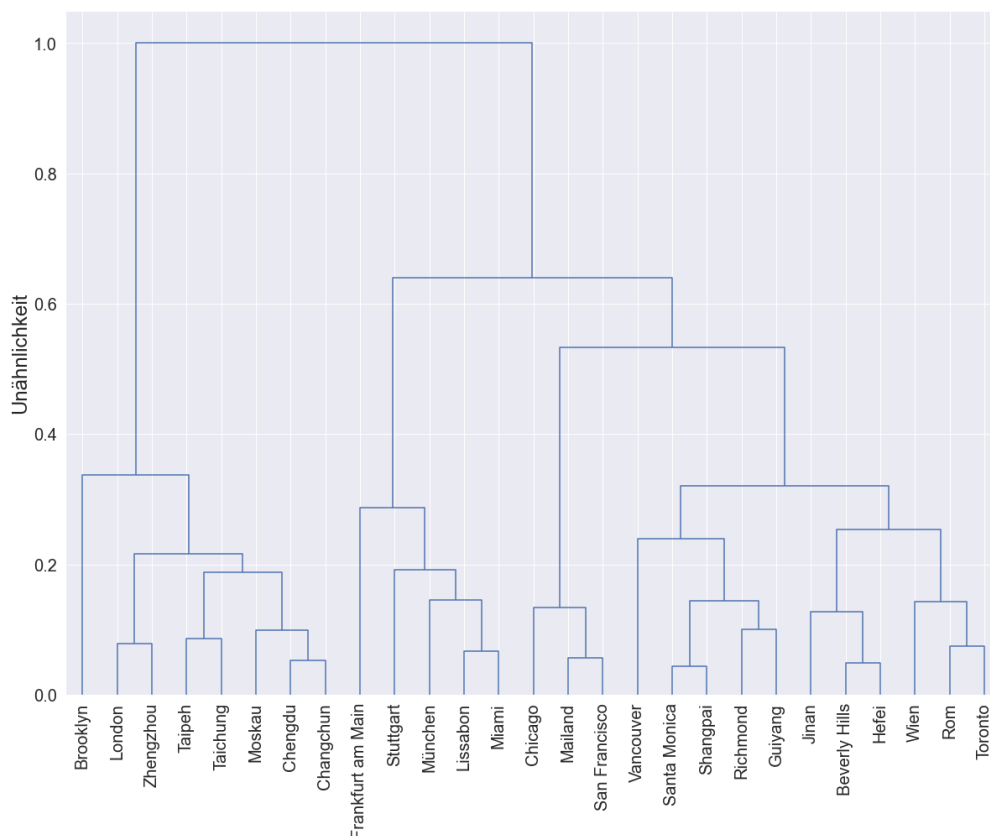


Abbildung 3.22: Dendrogramm der Unähnlichkeitsmatrix für Städte $S_{velocity}$.

3.4.4 Clustering von Tagen

Ähnlich wie bei der Untersuchung der Ähnlichkeit von Ländern und Städten ist auch eine Ähnlichkeitsuntersuchung für die verschiedenen Tage an denen die Messkampagne durchgeführt worden sind. Hierfür wurden die Verteilungen für die 23 verschiedenen Tage der Stadt Stuttgart miteinander verglichen. Die Wahl fiel hier auf Stuttgart, da für den gesamten Zeitraum der Messkampagne die meisten Daten in Stuttgart aufgezeichnet worden sind. Da die Aufnahmen mittags gestartet worden sind, wurde eine Vorauswahl durchgeführt welche Tage verglichen werden. Es wurden daher nur Tage berücksichtigt, an denen eine Menge von mindestens 5000 Fahrten an einem Tag gemessen worden sind. Die Ähnlichkeit der Durchschnittsgeschwindigkeiten der verschiedenen Tage ist in der Heatmap in Abbildung 3.23 dargestellt. Die Abbildung zeigt farblich dargestellt wie weit entfernt oder unähnlich die verschiedenen Tage voneinander sind. Je dunkler der entsprechende Eintrag desto unähnlicher sind die Tage. Bei dieser Darstellung fällt auf, dass viele Tage sehr ähnlich sind. Besonders auffällig sind aber die Tage 21.

Februar 2021, 25. April 2021 und der 20. Juni 2021. Diese Tage sind sehr dunkel dargestellt, da sie unähnlich zu allen weiteren Tagen sind. Diese Einträge sind aber in der Heatmap sehr ähnlich zueinander. Sieht man sich die Wochentage dieser drei Daten an fällt auf, dass es alles Sonntage sind.

Um ein besseres Verständnis zu schaffen, wird die Heatmap für die Unähnlichkeit der Durchschnittsgeschwindigkeiten in 3.24 als Dendrogramm präsentiert. Dabei wurden die verschiedenen Daten durch die entsprechenden Wochentagsnamen ersetzt. Hier lassen sich auch wieder die Sonntage vom 21. Februar, 25. April und der 20. Juni finden. Diese wurden durch das Clustering in einen eigenen Cluster isoliert. Diese unabhängigen Sonntage sind sehr unähnlich zu dem Rest der gemessenen Tage. Hinzu kommt, dass der Samstag als Wochentag auch in einem Cluster von den restlichen Wochentagen isoliert werden konnte. Der Samstags-Cluster ist aber ähnlicher zu Montag bis Freitag als der Sonntags-Cluster. Die übrigen Tage formen einen Werktagscluster, bei dem die feinere Aufspaltung aber nicht zu einer weiteren Isolation führt. Die Werkstage verhalten sich alle sehr ähnlich zueinander.

Um einen Vergleich zu schaffen werden in Abbildung 3.25 die ECDF der Durchschnittsgeschwindigkeiten verschiedener Städte farblich dargestellt und zusätzlich die verschiedenen Verteilungen der gemessenen Tage in Stuttgart in Grau dargestellt. Die blaue Kurve von Stuttgart bildet die ECDF der gesamten gemessenen Verteilung in Stuttgart. Inmitten der grauen Verläufe zeigt sich der gesamte Verlauf von Stuttgart. Dabei fallen drei graue Kurven etwas deutlicher auf, da sie sich weiter entfernt von der blauen, als auch den meisten anderen grauen Kurven befinden. Diese gesonderten Verläufe zeigen die Sonntage, die wie im Schaubild 3.24 isoliert werden konnten. Die Kurve von Lissabon verläuft räumlich sehr nah an der von Stuttgart und auch an den meisten Tagen von Stuttgart. Wie in der vorherigen Untersuchung gezeigt verhalten sich Stuttgart und Lissabon für den Parameter Geschwindigkeit auch ähnlich zueinander. In dieser Abbildung verdeutlicht sich diese Ähnlichkeit, da die Kurve von Lissabon für den größten Teil wie ein weiterer gemessener Tag in Stuttgart aussieht.

3.5 Darstellung gemessener Fahrtparameter einer Flotte durch Verteilungsfunktionen

Die Suche nach einer parametrisierten bekannten Verteilung für eine unbekannte gemessene reale Verteilung ist ein bekanntes Problem bei der statistischen Analyse und Datenmodellierung. Ziel ist es, eine bekannte Verteilung zu finden, die

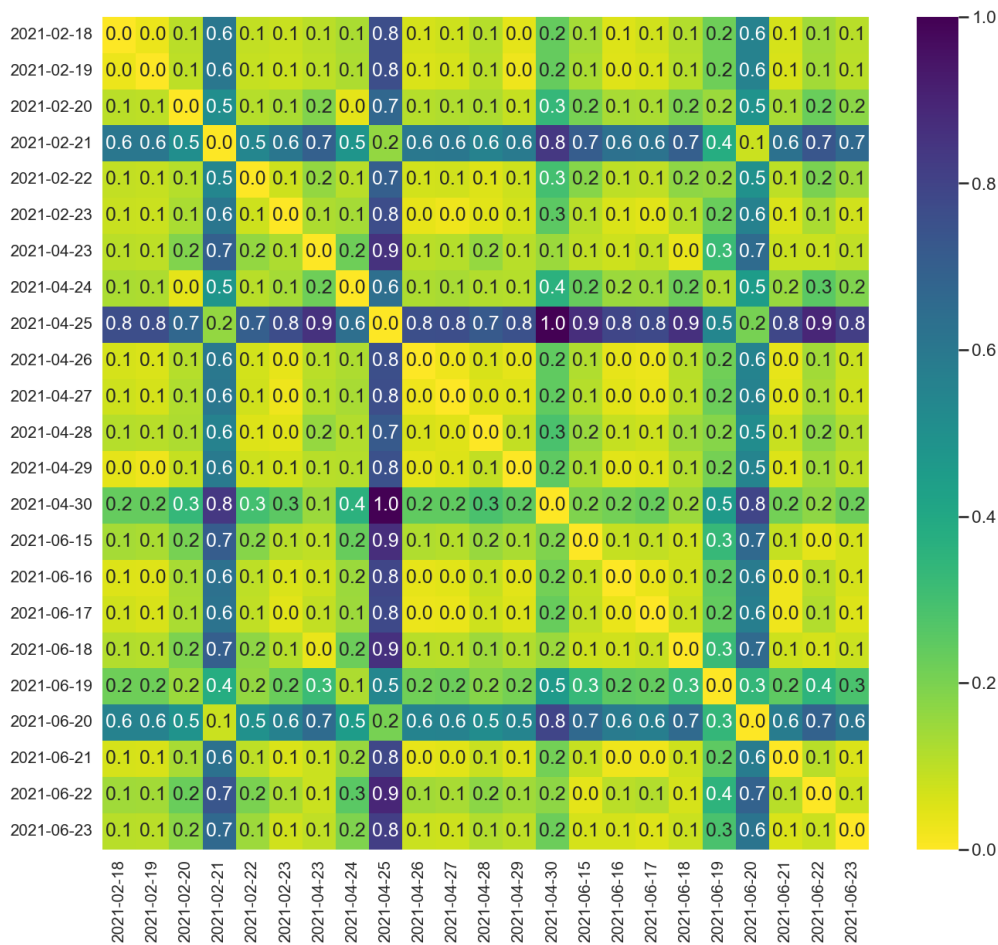


Abbildung 3.23: Heatmap der Unähnlichkeitsmatrix für die Geschwindigkeit an den Tagen der Messkampagne S_{days} .

den Merkmalen der beobachteten Daten nahe kommt, um weitere Analysen und Schlussfolgerungen zu ermöglichen.

Nach der Verarbeitung der Daten besteht der nächste Schritt darin, die Daten zu analysieren und die zugrunde liegende Verteilung zu bestimmen. Dies kann durch verschiedene statistische Verfahren wie Histogramme, Kerndichteschätzung oder Quantil-Quantil-Diagramme (Q-Q) erfolgen. Diese Methoden bieten visuelle Einblicke in die Form, die zentrale Tendenz und die Streuung der Daten und helfen dabei, potenzielle Verteilungsfamilien zu identifizieren, die zu den beobachteten Fahrtmetriken passen könnten. Sobald einige mögliche Verteilungen identifiziert wurden, besteht der nächste Schritt darin, die Parameter dieser Verteilungen zu schätzen. Bei diesem Prozess werden statistische Verfahren wie die Maximum-

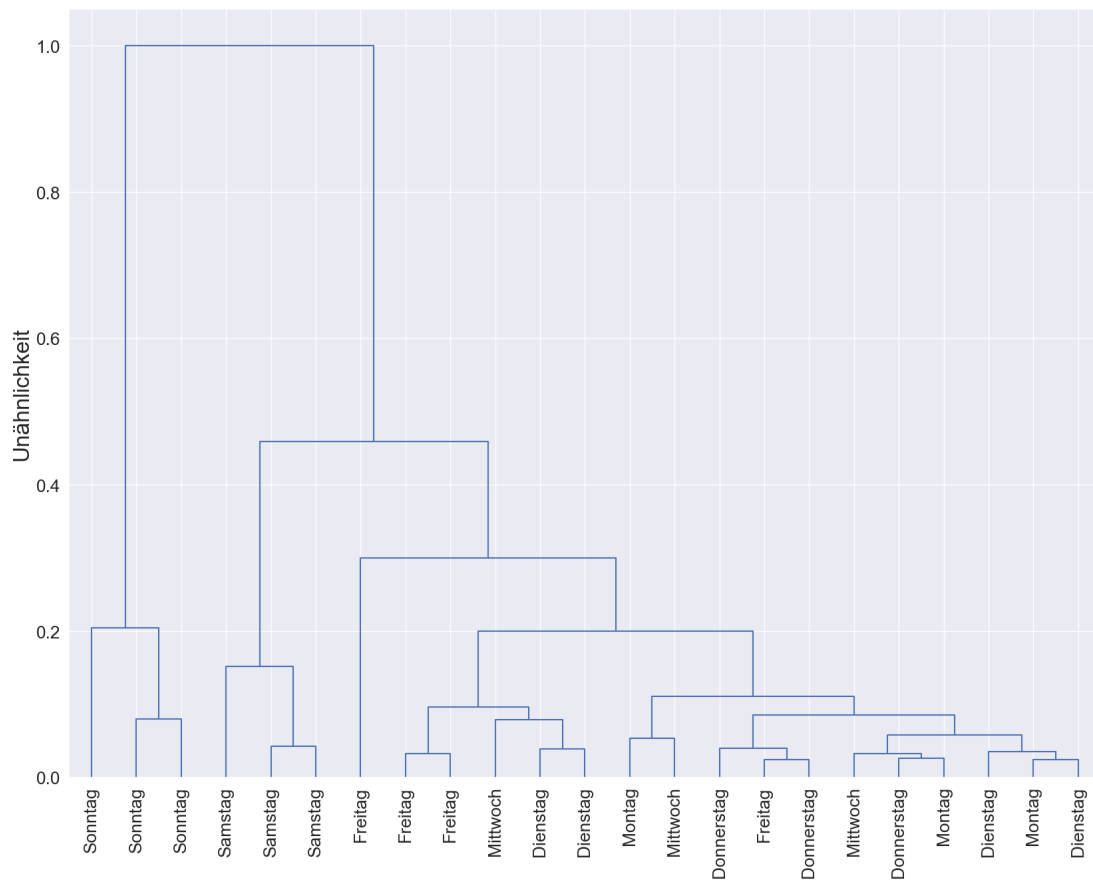


Abbildung 3.24: Dendrogramm der Unähnlichkeitsmatrix für die Geschwindigkeit an den Tagen der Messkampagne S_{days} .

Likelihood-Schätzung (MLE) oder die Methode der Momente eingesetzt, um die Parameterwerte zu ermitteln, die am besten zu den beobachteten Daten passen. Diese geschätzten Parameter definieren die gewählte bekannte Verteilung und ermöglichen ihre Anwendung auf die Daten der Fahrzeugflotte. Um die Güte der Anpassung zwischen der gewählten Verteilung und den beobachteten Daten zu bewerten, werden statistische Tests eingesetzt. Zu den üblichen Tests gehören der Kolmogorov-Smirnov-Test, der Anderson-Darling-Test oder der Chi-Quadrat-Test. Diese Tests vergleichen die empirische Verteilung der beobachteten Daten mit der theoretischen Verteilung, die sich aus der gewählten bekannten Verteilung ergibt, und ermöglichen so eine quantitative Bewertung der Anpassung.

Die Anpassung der Verteilung führt zu einer parametrisierten, bekannten Verteilung, welche die empirisch gemessenen Entfernungen, Dauern und Geschwindigkeiten charakterisiert. Nach der vorhandenen Literatur sind die Verteilungen

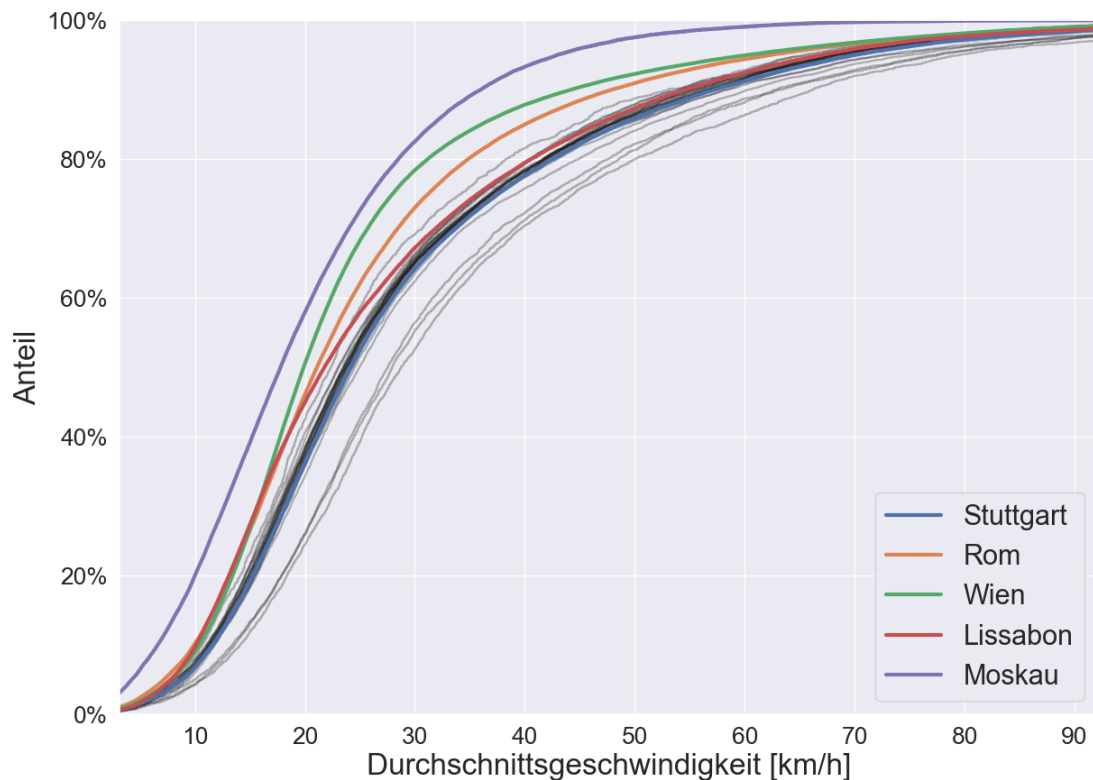


Abbildung 3.25: ECDF Schaubild der verschiedenen gemessenen Tage für Stuttgart und den verschiedenen Städten.

Exponential, (abgeschnittenes) Power-Law, Gamma und Burr für die Fahrtmetriken geeignet. [76–82]. Allerdings passen nicht alle in der Literatur vorhandenen Verteilungen zu den Daten in dieser Untersuchung. Die Verteilungen, die für die jeweilige Fahrtmetrik am besten geeignet sind, werden in den folgenden Abschnitten für die global gemessenen Daten präsentiert. Die Skew- und Kurtosiswerte der Verteilungen der verschiedenen Fahrtmetriken zeigen rechtsschiefe Verteilungen ($skew_{distance} = 6.29$, $skew_{duration} = 4.21$, $skew_{velocity} = 1.31$), wobei die meisten Daten in einem kleinen Bereich konzentriert sind ($kurt_{distance} = 61.72$, $kurt_{duration} = 35.75$, $kurt_{velocity} = 2.28$). Insbesondere die Fahrtmetrik Distanz ist aufgrund des extremen Kurtosis-Wertes dicht in einem kleinen Bereich gepackt. Die 20 am besten passenden kontinuierlichen Verteilungen aus der Python-Bibliothek *scipy.stats* werden in dieser Arbeit verwendet, um das am besten passende Modell zu finden. Darüber hinaus werden die am besten passenden Modelle aus früheren Studien in die Liste der Verteilungen aufgenommen, wenn sie zuvor nicht enthalten waren. Die erforderlichen Parameter für die am besten passenden Modelle werden für die Neuerstellung mit der Python-Bibliothek *scipy* bereitgestellt.

3.5.1 Bewertung der Verteilungspassform

Die Genauigkeit jedes Modells wird anhand der folgenden statistischen Indikatoren bewertet, um das beste Modell für jede Fahrtmetrik auszuwählen: Restquadratsumme (RSS), Bestimmtheitsmaß (R^2), mittlerer absoluter Fehler (MAE) und mittlerer absoluter prozentualer Fehler (MAPE). Für jedes Modell wird ein Konfidenzintervall angegeben, in dem es mit einem Konfidenzniveau an die Verteilung angepasst ist. In dieser Studie beträgt das Konfidenzniveau 95 %. Die untere Grenze des Konfidenzintervalls wird als l_{fit} , während die obere Grenze als u_{fit} bezeichnet wird. Zusätzlich wird die Wahrscheinlichkeit außerhalb des vorhergesagten Intervalls (POPI) als Bewertungsmetrik für die berechneten Konfidenzintervalle dargestellt. Die Berechnung für jede Metrik ist in Gleichung 3.6 angegeben.

$$\begin{aligned}
 RSS &= \sum_{i=1}^n (y_i^{obs} - y_i^{fit})^2 \\
 R^2 &= 1 - \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{fit})^2}{\sum_{i=1}^n (y_i^{obs} - \bar{y}^{obs})^2} \\
 MAE &= \frac{1}{n} \cdot \sum_{i=1}^n |y_i^{obs} - y_i^{fit}| \\
 MAPE &= \frac{100 \%}{n} \cdot \sum_{i=1}^n \left| \frac{y_i^{obs} - y_i^{fit}}{y_i^{obs}} \right| \\
 POPI &= \frac{100 \%}{N} \cdot \sum_{i=1}^N c_i \\
 \text{mit } y_i^{obs} &:= \text{tatsächlich gemessener Wert der } i\text{-ten Beobachtung,} \\
 y_i^{fit} &:= \text{vom Modell vorhergesagter Wert der } i\text{-ten Beobachtung} \\
 n &:= \text{Anzahl an Perzentilen, } N := \text{Stichprobengröße,} \\
 \text{und } c_i &= 0 \text{ wenn } y_i^{obs} \in [l_{fit}, u_{fit}] \text{ sonst } c_i = 1
 \end{aligned} \tag{3.6}$$

Der RSS-Wert misst den quadratischen Abstand zwischen jedem Perzentil der beobachteten und der Modellwerte. Je niedriger der RSS-Wert ist, desto genauer bildet das Modell die Verteilung ab. Das Bestimmtheitsmaß R^2 ist ein statistischer Indikator dafür, wie genau das Modell die beobachteten Ergebnisse wiedergibt. Ein R^2 -Wert nahe 1 weist auf eine starke positive Korrelation hin. Werte nahe Null

zeigen an, dass überhaupt keine Korrelation besteht. Der MAE-Wert berechnet den mittleren absoluten Fehler bei jedem Perzentil, den der MAPE-Wert als Prozentsatz der beobachteten Werte angibt. Auch hier gilt, je niedriger die MAE- und MAPE-Werte sind, desto besser passt das Modell zur Verteilung. Der POPI berechnet die Wahrscheinlichkeit, dass ein beobachteter Wert außerhalb eines 95 %-Konfidenzintervalls liegt. Ein POPI von etwa 5 % modelliert die Verteilung mit einer Genauigkeit von 95 % der beobachteten Verteilung.

3.5.2 Distanz

Die globalen Entfernungsmessungen werden verwendet, um die beobachtete Verteilung mit einem passenden Modell nachzubilden, das die rechtsschiefe Charakteristik des Datensatzes und die dicht besiedelten Werte beschreiben kann. Eine Zusammenfassung findet sich in Tabelle 3.6. Die *lognorm*-Verteilung konnte die gefahrene Distanz detaillierter beschreiben als die anderen Verteilungen gemäß RSS, R^2 , MAE und MAPE. Die *lognorm*-Verteilung hat den niedrigsten RSS-Wert zusammen mit dem niedrigsten MAE. Die MAPE ist bei der *lognorm*-Verteilung nicht die niedrigste, da sie auch Werte über dem maximalen Beobachtungswert im Datensatz mit einer geringen Wahrscheinlichkeit modelliert. Dieser Effekt bewirkt, dass an den Extremen der Verteilung ein Unterschied auftritt. Der POPI liegt etwas über 5 %, was bedeutet, dass das Konfidenzintervall des Verteilungsmodells weniger als 95 % der beobachteten Daten abdeckt. Da die anderen Verteilungen gleichmäßiger verteilt sind und die obere Grenze des Konfidenzniveaus viel später auftritt, ist der POPI für sie niedriger.

Die parametrisierte Verteilung mit ihren Werten für global gemessene Fahrten ist in Abbildung 3.26 dargestellt. Die Wahrscheinlichkeitsdichtefunktion für die *lognorm*-Verteilung ist in Gleichung 3.7 angegeben. Die in dieser Arbeit verwendete Darstellung der Funktion folgt der Implementierung in der Python-Bibliothek `scipy.stats` [83].

$$f(y, s) = \frac{1}{sy\sqrt{2\pi}} \exp\left(-\frac{\log^2(y)}{2s^2}\right)$$

mit $y = \frac{x - loc}{scale}$,

$$y > 0, x > 0, s > 0, scale > 0$$

(3.7)

	RSS	R^2	MAE	MAPE	POPI
<i>lognorm</i>	0.00018	0.98541	0.4	1.56 %	5.23 %
<i>johnsonsu</i>	0.00019	0.99284	0.41	1.67 %	5.23 %
<i>powerlognorm</i>	0.00051	0.98573	0.42	1.46 %	5.25 %
<i>fisk</i>	0.00064	0.72672	5.38	6.41 %	4.1 %
<i>f</i>	0.0009	0.89446	1.35	3.62 %	4.64 %
<i>betaprime</i>	0.0009	0.94475	1.36	3.6 %	4.64 %
<i>invgauss</i>	0.0009	0.99355	0.77	5.01 %	5.42 %
<i>burr</i>	0.00107	0.78260	3.3	5.29 %	4.1 %
<i>mielke</i>	0.00107	0.78847	3.25	5.31 %	4.1 %
<i>fatiguelife</i>	0.00134	0.91464	2.8	12.9 %	5.87 %
<i>genpareto</i>	0.00158	0.92275	0.97	3.96 %	4.36 %
<i>lomax</i>	0.0016	0.91975	1.05	3.85 %	4.33 %
<i>foldcauchy</i>	0.00193	0.59244	10.14	11.44 %	3.43 %
<i>halfcauchy</i>	0.00193	0.59713	9.96	11.36 %	3.43 %
<i>invweibull</i>	0.00304	0.60642	11.24	14.83 %	3.81 %
<i>invgamma</i>	0.00336	0.65049	8.34	13.53 %	4.0 %
<i>halfgennorm</i>	0.00598	0.96809	1.84	12.03 %	4.49 %
<i>loglaplace</i>	0.00625	0.47643	22.71	22.68 %	2.3 %
<i>recipinvgauss</i>	0.00712	0.87336	3.78	25.0 %	5.83 %
<i>johnsonsb</i>	0.01571	0.98829	22.6	68.7 %	1.37 %
<i>powerlaw</i>	0.05366	0.68538	129.98	595.2 %	0.0 %
<i>rayleigh</i>	0.08233	0.57222	12.14	513.54 %	3.73 %
<i>gamma</i>	0.0865	0.90545	13.41	94.91 %	78.41 %

Tabelle 3.6: Zusammenfassung der verschiedenen Verteilungsfunktionen für die Distanz und deren Bewertung.

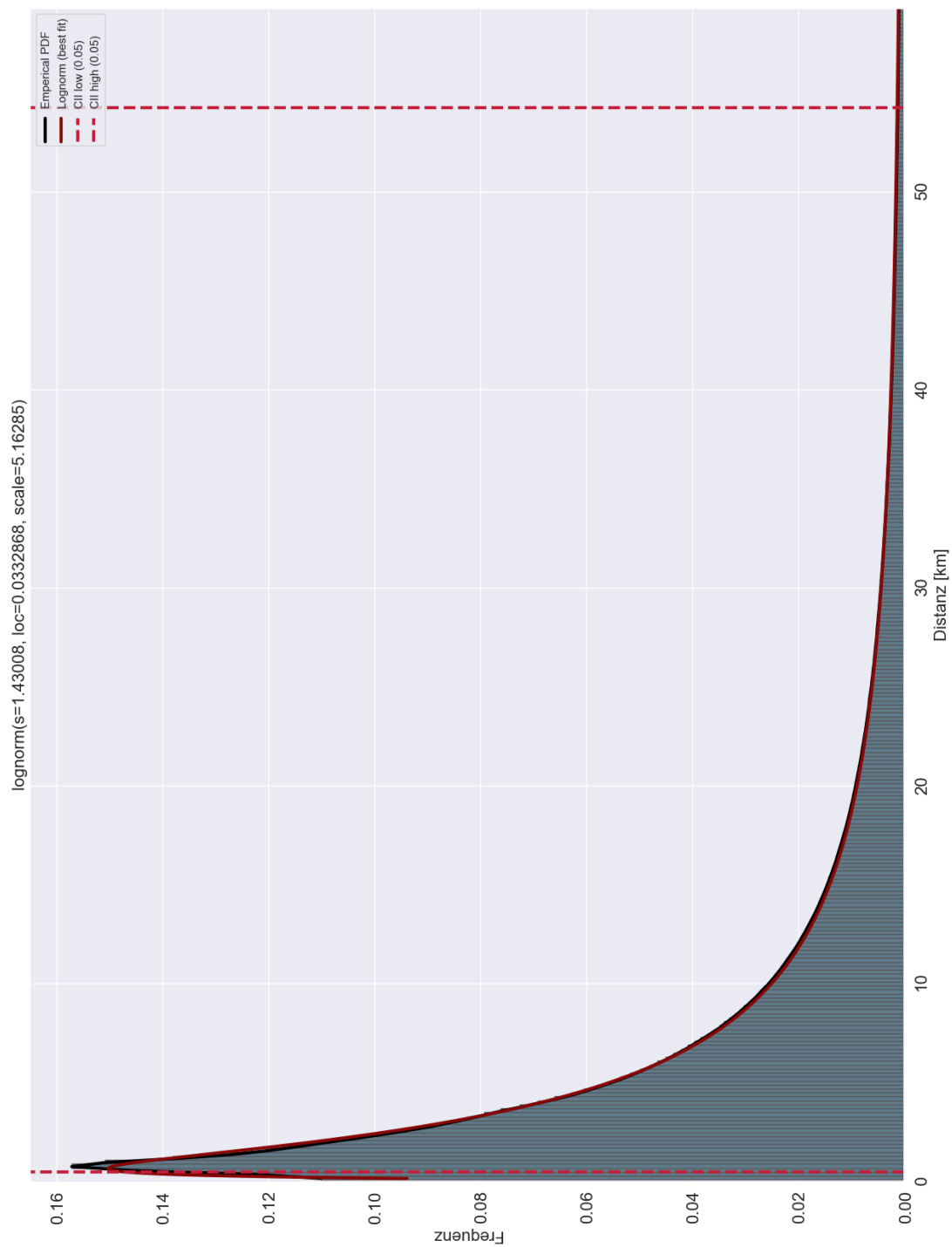


Abbildung 3.26: Histogramm der gemessenen Distanzen mit der parametrisierten Verteilungsfunktion (rot) (angelehnt an [e]).

3.5.3 Fahrtdauer

Die Verteilung für die Fahrtdauer ist aufgrund der Methode zur Messung der Fahrtdauer in Zwei-Minuten-Bins diskretisiert. Die Ergebnisse für die verschiedenen Bewertungsmetriken für die Fahrtdauer sind in Tabelle 3.7 angegeben. Das am besten passende Modell nach dem RSS-Score ist die *halfcauchy*-Verteilung. Der R^2 -Wert von 0,43 deutet jedoch darauf hin, dass die Verteilung keine gute Anpassung für die Gesamtverteilung darstellt. Im Vergleich zu den Werten der anderen Verteilungen sind MAE und MAPE groß. Und das Vertrauen erstreckt sich über einen großen Bereich, der weit über 95 % der beobachteten Verteilung hinausgeht. Infolgedessen wird die *genpareto*-Verteilung als die beste Anpassung für die Bewertungsmetriken in dieser Studie angesehen. Die Verteilung erreicht fast die gleiche Punktzahl wie die Verteilung mit der besten Bewertung und einem niedrigen MAE und MAPE. Die Extremwerte können nicht sehr genau beschrieben werden, da der POPI von einem Zielwert von 5 % abweicht. Die Gleichung für die PDF der *genpareto*-Verteilung ist in 3.8 angegeben. Abbildung 3.27 zeigt die parametrisierte *genpareto*-Verteilung und die empirische Verteilung. Auch hier orientiert sich die Darstellung an der Implementierung in `scipy.stats` [83]:

$$\begin{aligned}
 f(y, c) &= (1 + cy)^{-1-1/c} \\
 \text{mit } y &= \frac{x - loc}{scale}, \\
 y \geq 0, x \geq 0, c > 0, scale > 0
 \end{aligned}
 \tag{3.8}$$

	RSS	R^2	MAE	MAPE	POPI
<i>halfcauchy</i>	7e-05	0.42774	33.65	41.0 %	0.08 %
<i>genpareto</i>	8e-05	0.99172	1.27	12.07 %	2.91 %
<i>lomax</i>	8e-05	0.99251	1.28	12.19 %	2.81 %
<i>halfgennorm</i>	0.00015	0.97042	1.97	13.86 %	3.17 %
<i>foldcauchy</i>	0.00036	0.42963	23.5	26.97 %	0.14 %
<i>recipinvgauss</i>	0.00045	0.96926	1.93	10.97 %	3.5 %
<i>f</i>	0.00069	0.95613	1.25	8.6 %	2.4 %
<i>betaprime</i>	0.00069	0.95738	1.23	8.58 %	2.4 %
<i>johnsonsb</i>	0.00073	0.99856	0.57	7.36 %	2.62 %
<i>burr</i>	0.00073	0.82129	3.34	10.77 %	1.64 %
<i>mielke</i>	0.00073	0.82073	3.36	10.81 %	1.64 %
<i>fisk</i>	0.00078	0.80351	3.7	11.13 %	1.54 %
<i>lognorm</i>	0.00091	0.98897	1.21	8.53 %	2.28 %
<i>fatiguelife</i>	0.00101	0.98360	1.33	8.52 %	2.92 %
<i>loglaplace</i>	0.00123	0.54485	10.89	19.75 %	0.64 %
<i>powerlognorm</i>	0.00123	0.97497	1.46	9.08 %	2.2 %
<i>invgamma</i>	0.00129	0.83521	3.65	12.31 %	1.58 %
<i>invgauss</i>	0.00132	0.99788	0.92	8.78 %	2.46 %
<i>invweibull</i>	0.00132	0.76440	5.04	13.66 %	1.25 %
<i>johnsonsu</i>	0.00154	0.97365	2.01	9.68 %	3.37 %
<i>rayleigh</i>	0.00537	0.77268	8.39	86.17 %	4.52 %
<i>gamma</i>	0.01029	0.84805	19.44	98.42 %	89.17 %
<i>powerlaw</i>	0.01379	0.60774	210.42	1872.67 %	0.0 %

Tabelle 3.7: Zusammenfassung der verschiedenen Verteilungsfunktionen für die Fahrdauer und deren Bewertung.

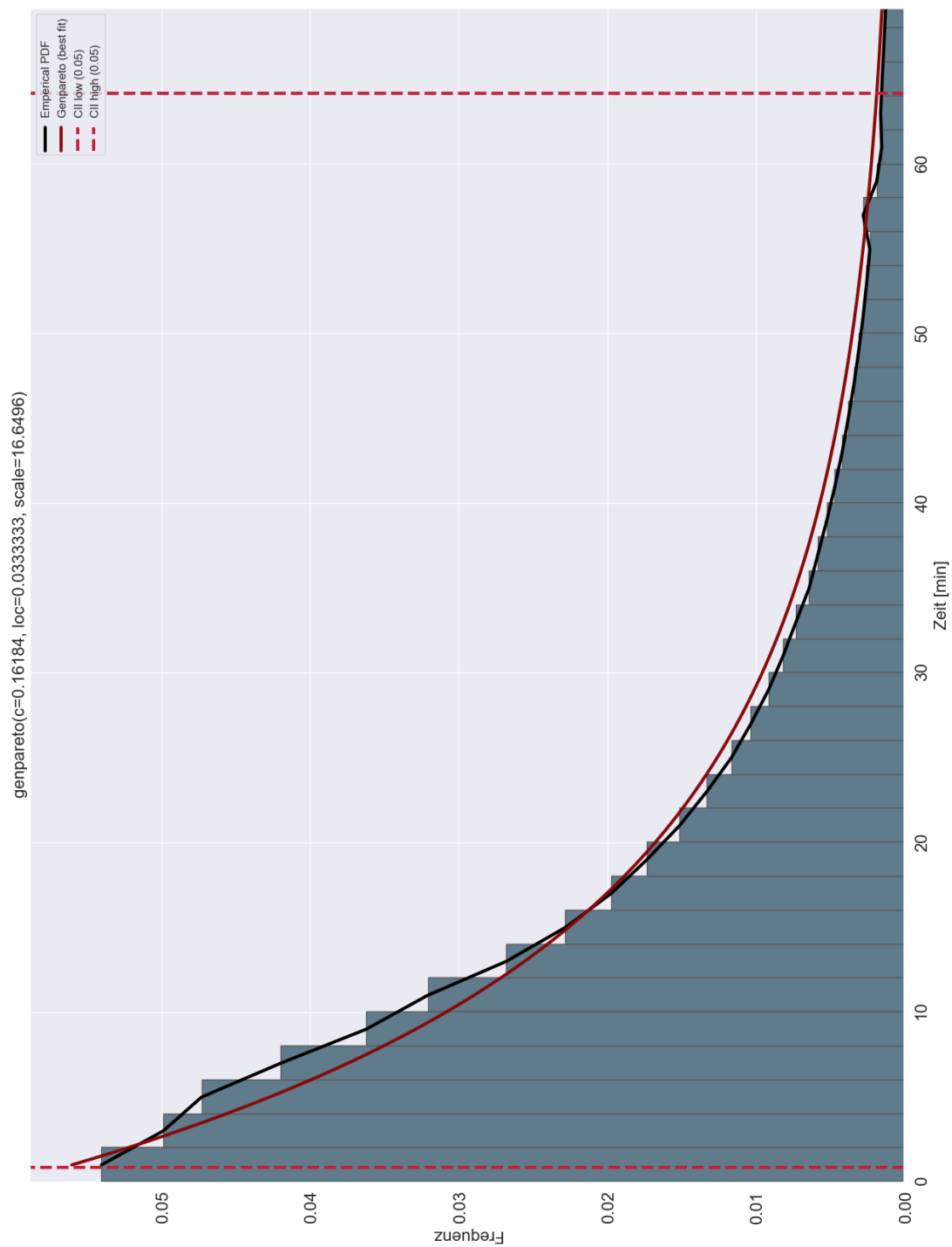


Abbildung 3.27: Histogramm der gemessenen Fahrt dauern mit der parametrisierten Verteilungsfunktion (rot) (angelehnt an [e]).

3.5.4 Geschwindigkeit

Für die mittlere Fahrtdauer wird die beobachtete Verteilung in 1 *km/h* große Bins diskretisiert. Tabelle 3.8 zeigt die Ergebnisse der Anpassung verschiedener gängiger Verteilungen über die Bewertungsmetriken dieser Studie. Im Fall der Fahrtmetrik haben elf der zwanzig angepassten Verteilungen einen gleich niedrigen RSS-Wert, was darauf hindeutet, dass ein breites Spektrum von Verteilungen die allgemeinen Merkmale der Fahrtmetrik Durchschnittsgeschwindigkeit beschreiben könnte. Außerdem haben viele der getesteten Verteilungen eine hohe Bewertung bei den Metriken R^2 , MAE und MAPE. Die *invgauss*-Verteilung hat jedoch eine der besten Bewertungen in der Fahrtmetrik R^2 , die niedrigste MAE/MAPE und einen POPI-Wert von etwa 5 %. Darüber hinaus benötigt die *invgauss*-Verteilung zwei Parameter μ und λ , was sie zu einer der einfacheren Verteilungen macht. Die PDF für die *invgauss*-Verteilung ist in Gleichung 3.9 angegeben [83]. Abbildung 3.28 zeigt die parametrisierte *invgauss*-Verteilung und die empirische Verteilung.

$$f(y, \mu) = \frac{1}{\sqrt{2\pi y^3}} \exp\left(-\frac{(y - \mu)^2}{2y\mu^2}\right)$$

mit $y = \frac{x - loc}{scale}$,

$$y \geq 0, x \geq 0, \mu > 0, scale > 0$$
(3.9)

	RSS	R^2	MAE	MAPE	POPI
<i>johnsonsu</i>	2e-05	0.99638	0.33	1.51 %	5.46 %
<i>powerlognorm</i>	2e-05	0.99690	0.31	1.61 %	5.51 %
<i>invweibull</i>	2e-05	0.99094	0.63	3.05 %	5.11 %
<i>invgauss</i>	3e-05	0.99864	0.28	1.47 %	5.64 %
<i>fatiguelife</i>	4e-05	0.99894	0.31	1.56 %	5.68 %
<i>johnsonsb</i>	4e-05	0.99936	0.36	1.68 %	5.73 %
<i>betaprime</i>	5e-05	0.99866	0.42	1.7 %	5.68 %
<i>f</i>	5e-05	0.99878	0.45	1.78 %	5.69 %
<i>fisk</i>	6e-05	0.92869	1.67	4.3 %	4.03 %
<i>burr</i>	8e-05	0.96071	1.09	2.28 %	4.45 %
<i>mielke</i>	8e-05	0.96130	1.08	2.29 %	4.46 %
<i>gamma</i>	0.00026	0.98756	1.83	9.45 %	5.46 %
<i>rayleigh</i>	0.00061	0.97502	2.9	9.91 %	6.41 %
<i>foldcauchy</i>	0.00065	0.20798	35.16	38.02 %	1.03 %
<i>loglaplace</i>	0.00067	0.67731	6.03	9.58 %	2.79 %
<i>invgamma</i>	0.00405	0.36897	32.03	47.9 %	2.98 %
<i>genpareto</i>	0.00421	0.99391	6.46	27.97 %	1.06 %
<i>halfcauchy</i>	0.00452	0.21094	97.61	120.43 %	0.3 %
<i>lomax</i>	0.00531	0.98119	8.08	32.35 %	0.55 %
<i>powerlaw</i>	0.01084	0.92928	43.78	104.49 %	0.0 %
<i>halfgennorm</i>	0.01552	0.79237	25.87	63.17 %	0.04 %
<i>recipinvgauss</i>	0.01967	0.92404	16.15	54.86 %	0.01 %
<i>lognorm</i>	0.16566	0.62036	27.93	88.98 %	32.25 %

Tabelle 3.8: Zusammenfassung der verschiedenen Verteilungsfunktionen für die Durchschnittsgeschwindigkeit und deren Bewertung.

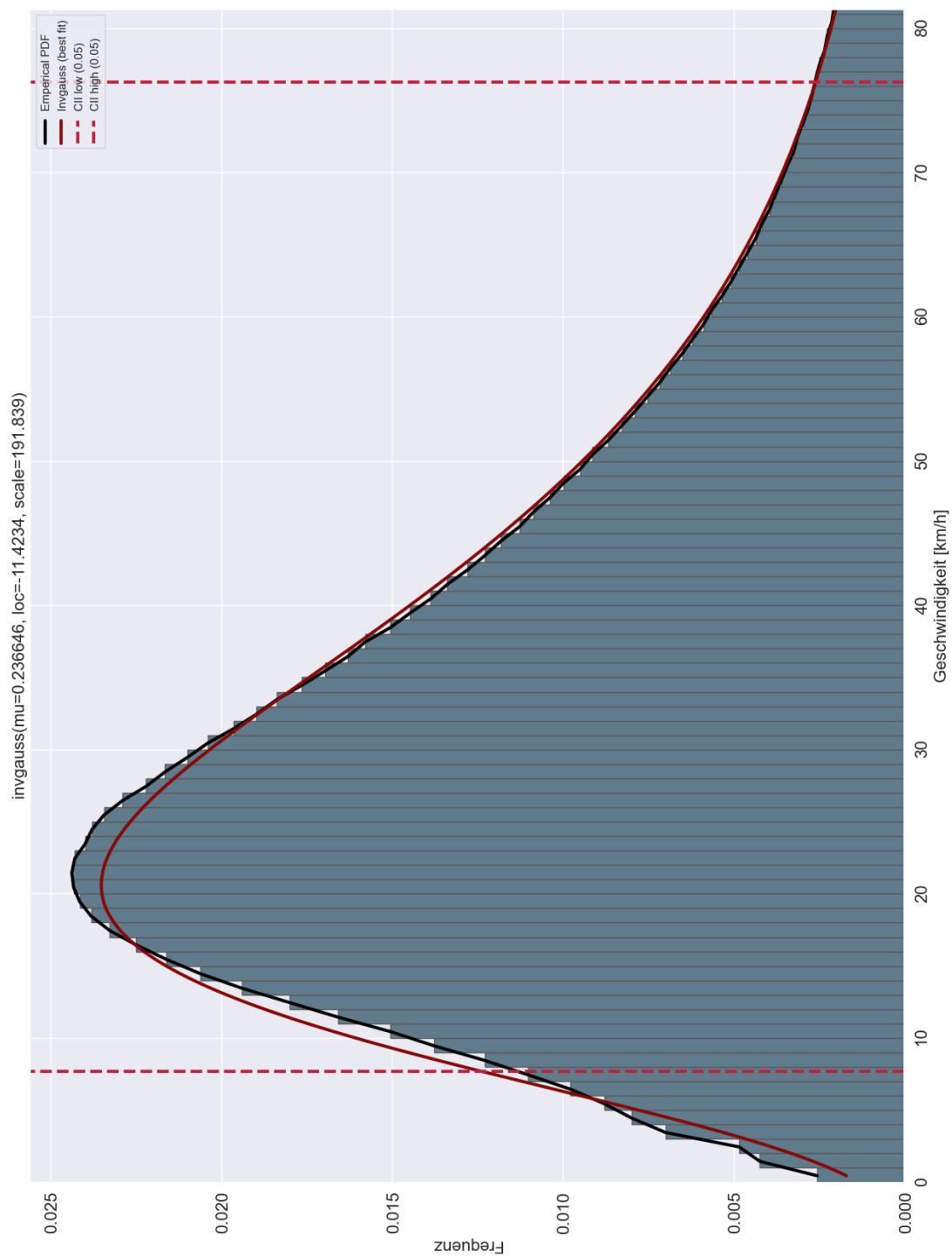


Abbildung 3.28: Histogramm der gemessenen Durchschnittsgeschwindigkeiten mit der parametrisierten Verteilungsfunktion (rot) (angelehnt an [e]).

3.6 Zusammenfassung

Dieses Kapitel beschäftigt sich mit der Erhebung eines Datensatzes von anonymisierten Fahrzeug FCD-Daten einer vernetzten Fahrzeugflotte bestehend aus PKW der Marke Mercedes-Benz und der Bereinigung der fehlerhaften Daten, sowie der ausführlichen Betrachtung der Daten von verschiedenen in Form von direkten Vergleichen, der Tagesganglinien, empirische kumulierte Verteilungsfunktionen, Clusterings und der Bereitstellung von parametrisierten in der Literatur verbreiteten Funktionen zur Bereitstellung der Daten.

Hierfür werden die übermittelten Daten in Messkampagnen aufgezeichnet und mit aus der Literatur bekannten Methoden für das Bereinigen von Datensätzen vorverarbeitet. Außerdem wird für einen Großteil der Daten ein Map Matching durchgeführt und die Genauigkeit der Rohdaten gemessen. Mithilfe der vorverarbeiteten Daten wird ein Datensatz zusammengestellt, der die Fahrdistanz, Fahrtdauer, Durchschnittsgeschwindigkeit, Herkunftsland und -stadt der etwa 224 Millionen Fahrten umfasst.

Im Anschluss wird die Auswahl der Fahrtparameter beschrieben und der Datensatz auf die Fahrtparameter: Fahrdistanz, Fahrtdauer und Durchschnittsgeschwindigkeit untersucht. Dabei stellt sich heraus, dass die Fahrtparameter der einzelnen Länder sich nicht deutlich voneinander unterscheiden. Eine durchschnittliche Fahrt ist zwischen 12 und 14 Kilometer lang und unter den untersuchten Ländern in China am kürzesten. Außerdem kann aus den Daten entnommen werden, dass die Fahrten am Wochenende länger sind als am Wochenende. Bei der Fahrtdauer fallen die Ergebnisse ähnlich aus. Im Mittel sind die Länder vergleichbar. Die Fahrten dauern im Mittel zwischen 18 und 23 Minuten und sind in Deutschland und Italien am kürzesten. In den USA und in China sind bei dieser Untersuchung die längsten Reisezeiten messbar. Ein weiteres Ergebnis dieser Untersuchung ist, dass 99 Prozent der globalen Fahrten kürzer sind als 2 Stunden und 9 Minuten. Die Untersuchung der Durchschnittsgeschwindigkeit zeigte etwas größere Unterschiede in den Extremen. Deutschland hat weltweit die höchste Durchschnittsgeschwindigkeit mit etwa $37 \frac{km}{h}$ und China die niedrigste mit etwa $25 \frac{km}{h}$. Wie auch bei der Distanz sind die Fahrten am Wochenende tendenziell global schneller als zwischen Montag und Freitag. Die Tagesganglinien zeigen das Fahrtverhalten der Fahrzeugnutzer, welches stark mit den Werktagen zusammenhängt. An Werktagen sind morgens und abends Spitzenzeiten zu beobachten, die zu Arbeitsbeginn und Arbeitsende auftreten. Morgens steigt das Verkehrsaufkommen, wenn Pendler zur Arbeit fahren, während abends, wenn sie nach Hause zurückkehren, erneut eine Zunahme des Verkehrs zu verzeichnen ist. Im Gegensatz dazu zeigt sich an Wochenenden ein anderes Bild. Hier gibt es in der Regel eine Spitze zwischen 11 und 13 Uhr. Dies könnte auf

Aktivitäten wie Einkaufen, Ausflüge oder soziale Unternehmungen hinweisen, die eine erhöhte Nachfrage nach Verkehrsdienstleistungen zur Folge haben.

Zusätzlich wurden die Ergebnisse der Fahrtparameter dieser Arbeit mit Mobilitätsuntersuchungen verglichen. Dabei fällt auf, dass je nach Erhebung eine Vielzahl von Parametern gemessen wurde. Es wurden nur Ergebnisse aus den Erhebungen berücksichtigt, die auch in der vorliegenden Untersuchung gemessen werden konnten. Teilweise wurden in den Erhebungen Jahres- oder Tageswerte angegeben, die für den Vergleich auf die hier berechneten Werte bezogen wurden. Die Stichproben in diesen Erhebungen sind oft kleiner als bei der Erfassung von Positionsdaten einer großen Flotte. Dennoch liegen die Ergebnisse in der vorliegenden Untersuchung meist in einer ähnlichen Größenordnung wie in den Vergleichserhebungen. Aufgrund der hohen Standardabweichung der Kennzahlen kann es durch eine kleinere Stichprobe und eine andere Verteilung der Teilnehmer zu Abweichungen von den Durchschnittswerten kommen.

Außerdem wurden die Verteilungen von Ländern, Städten und Tagen der verschiedenen Fahrtparametern geclustert. Um die Ähnlichkeit zwischen verschiedenen Verteilungen zu messen, werden verschiedene Ähnlichkeitsmetriken und deren Einsatzmöglichkeiten vorgestellt. Für die Clusterbildung wird die Wasserstein-Distanz verwendet, um die Unähnlichkeit zwischen den Verteilungen zu messbar zu machen. Eine Unähnlichkeitsmatrix wird erstellt und in Dendrogrammen und Heatmaps visualisiert, um die Ähnlichkeiten zwischen Ländern, Städten und Tagen darzustellen. Die Clusterbildung zeigt, dass einige Länder ähnlichere Mobilitätsmuster haben als andere. Dabei treten Länder wie Russland und China als herausragend unähnlich im Vergleich zu anderen Ländern bei der Untersuchung der Durchschnittsgeschwindigkeit hervor. Die beiden Länder sind zueinander aber sehr ähnlich, was auf die Konzentration der Flotte in überlasteten Großstädten zurückzuführen ist. Auch für Städte werden Cluster gebildet, wobei die Fahrtmetriken Distanz, Fahrtdauer und Geschwindigkeit berücksichtigt werden. Die Ergebnisse zeigen, wie sich Städte mit ähnlichen Fahrtverhaltensweisen gruppieren. Schließlich wird auch die Ähnlichkeit zwischen verschiedenen Tagen untersucht. Dabei werden die Tage anhand von den Fahrtparametern der Stadt Stuttgart miteinander verglichen und Cluster gebildet. Besondere Aufmerksamkeit gilt den Sonntagen, die sich speziell bei der Betrachtung der Verteilungen der Geschwindigkeit als besonders unähnlich zu den anderen Tagen herausstellen. Diese Tage können isoliert werden zusätzlich konnten auch Samstage bei dieser Betrachtung isoliert werden. Die Tage Montag bis Freitag verhalten sich ziemlich ähnlich zueinander, dort unterschied sich kein Tag sonderlich von den anderen. Insgesamt bietet dieser Abschnitt einen Einblick in die Anwendung von Clusteranalysen auf Flottendaten, um Muster und Zusammenhänge zu identifizieren und zu visualisieren.

Zuletzt wird die Darstellung der Fahrtparameter einer globalen Fahrzeugflotte durch parametrisierte Verteilungsfunktionen beschrieben. Dieser Prozess dient dazu, eine bekannte statistische Verteilung zu finden, die den beobachteten Daten nahekommt, um weitere Analysen und Schlussfolgerungen zu ermöglichen und dem Erhebungsaufwand mitsamt dem Einsatz einer gesamten Fahrzeugflotte zu entgehen. Anhand der Literatur wurde eine Reihe an bekannten Verteilungsfunktionen identifiziert, dessen Parameter angepasst werden um eine bestmögliche Passform zu der empirisch gemessenen zu erreichen. Um die bestmöglichen Parameter zu schätzen werden statistische Verfahren verwendet, wie die Maximum-Likelihood-Schätzung (MLE) oder die Methode der Momente. Verschiedene Verteilungen werden auf ihre Eignung für die gemessenen Fahrtdaten hin überprüft. Dabei zeigen sich Unterschiede in verschiedenen statistischen Metriken wie RSS, R², MAE, MAPE und der Wahrscheinlichkeit außerhalb des vorhergesagten Intervalls (POPI). Die Ergebnisse zeigen, dass diese Verteilungen die empirischen Fahrtdaten charakterisieren wie die Lognormal-Verteilung für Entfernungen, die Genpareto-Verteilung für Fahrtauern und die Invgauss-Verteilung für Geschwindigkeiten.

4 Drohnenbeobachtungen

Die Verkehrsbeobachtung mit Drohnen hat in den letzten Jahren stark an Bedeutung gewonnen und gilt als sehr vielversprechend für die Verkehrsüberwachung und -analyse. Mit hochauflösenden Kameras und dem Zusammenspiel verschiedener Sensoren können Drohnen den Verkehr beispielsweise auf Autobahnen an schwer zugänglichen Streckenabschnitten aufzeichnen und überwachen. Durch die Mobilität und die schnelle Inbetriebnahme von Drohnen können diese nahezu beliebig eingesetzt werden, in Bereichen, in denen großflächige Verdeckungen keine Rolle spielen. Die Analyse dieser Videos kann Daten über Verkehrsfluss, -geschwindigkeit und -dichte liefern, die es den Verkehrsbehörden ermöglichen, datengestützte Entscheidungen über Straßenmanagement und -planung zu treffen. Herkömmliche Verfahren zur Datenerfassung im Verkehr, wie Detektorschleifen und Probedaten, erreichen nicht die Präzision, die Drohnenaufnahmen bieten können. Der Einsatz von Drohnen zur Verkehrsüberwachung ist jedoch auch mit Herausforderungen verbunden, insbesondere mit Fragen des Datenschutzes und der Sicherheit, die von den Nutzern dieser Methode der Verkehrsüberwachung sorgfältig berücksichtigt werden müssen. Zusätzlich kommen mit dem Einsatz von Drohnen auch technische Grenzen einher, wie einer durch den Akku beschränkten Aufnahmedauer und die durch die Kamera beschränkte Sicht auf den Verkehr.

4.1 Datensatz

Dieser Abschnitt beschäftigt sich mit der Herkunft, dem Inhalt und dem Aufbau der Daten aus der Drohnenbeobachtung. Es wird beschrieben, wie die Daten erhoben worden sind und wie diese für weitere Auswertungen vorverarbeitet werden.

4.1.1 Beschreibung des Datensatzes

Der hier verwendete Datensatz ist aus einer Drohnenbeobachtung entstanden, die am 14. Juli 2020 zwischen 7 Uhr und 12 Uhr stattgefunden hat. Dabei wurde eine Messkampagne mit einer Drohne durchgeführt, wobei sechs Aufnahmen entstanden

sind, die aufgrund der Akkukapazität alle in etwa 20 Minuten andauern und freien als auch gestauten Verkehr enthalten. Die Beobachtung fand an der Autobahn A8 auf der Höhe von Wendlingen statt. Die Abbildung 4.1 zeigt zwei Szenen aus verschiedenen Blickwinkeln der Aufnahmen. Durch die Nachbearbeitung der Daten werden Fahrzeuge in den Aufzeichnungen erkannt und verschiedenen Fahrzeugklassen (PKW, LKW, Motorrad, Sprinter) zugeteilt. Diese Kategorien sind durch die farblichen Rechtecke um das jeweilige Objekt dargestellt. Jede Farbe steht für eine andere Fahrzeugklasse. Der beobachtete Straßenabschnitt deckt 200 Meter bis 600 Meter der A8 ab und hängt von dem Betrachtungswinkel der Drohne zur Straße ab. Abbildung 4.1(a) zeigt einen größeren Teil der Autobahn, da die Drohne tiefer fliegt und einen steileren Winkel zu dem Streckenabschnitt hat, als beispielsweise 4.1(b).

Die Abbildung 4.2 zeigt eine Skizze der erfassten Fahrstreifen in den Aufnahmen. Die einzelnen Fahrstreifen wurden mit Identifikationsnummern versehen, die in der Abbildung gekennzeichnet sind. Außerdem ist die Fahrtrichtung jedes Fahrstreifens für jeden Streifen markiert. Der Streckenabschnitt beinhaltet den dreistreifigen Autobahnabschnitt (4, 5, 6) von München nach Stuttgart mit einer Einfahrt aus Wendlingen am Neckar (44, 43) und den dreistreifigen Autobahnabschnitt (1, 2, 3) mit einer Ausfahrt nach Wendlingen am Neckar und Nürtingen (11).

Am Ende der Videoverarbeitungskette steht ein Datensatz für jede Aufnahme in der Form von Gleichung 4.1 zur Verfügung. Für alle erkannten Fahrzeuge wird eine

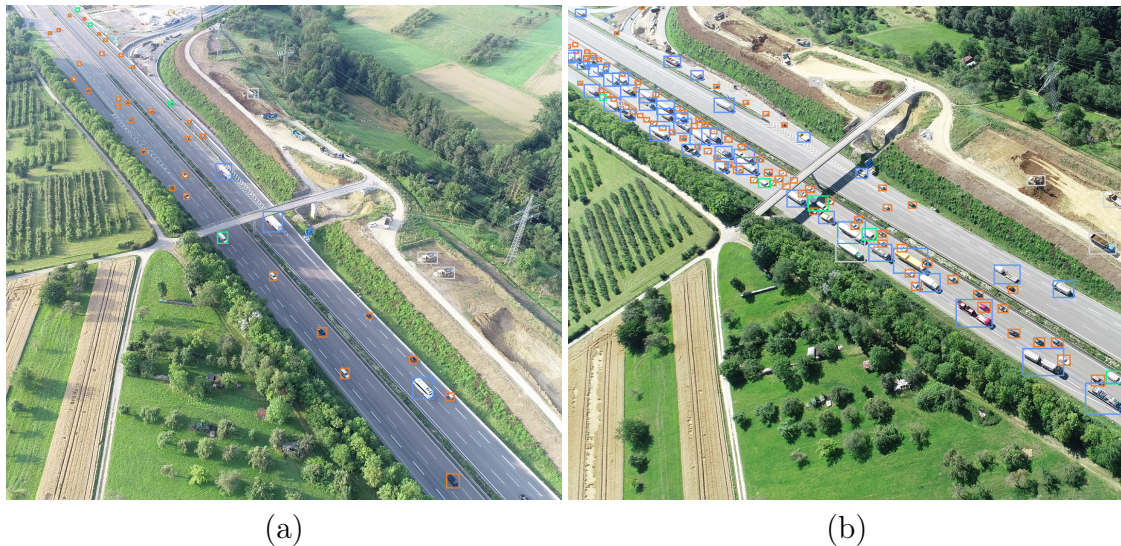


Abbildung 4.1: Ausschnitt aus zwei der Aufnahmen mit verschiedenen Perspektiven (a) Freier Verkehr (b) Gestauter Verkehr.

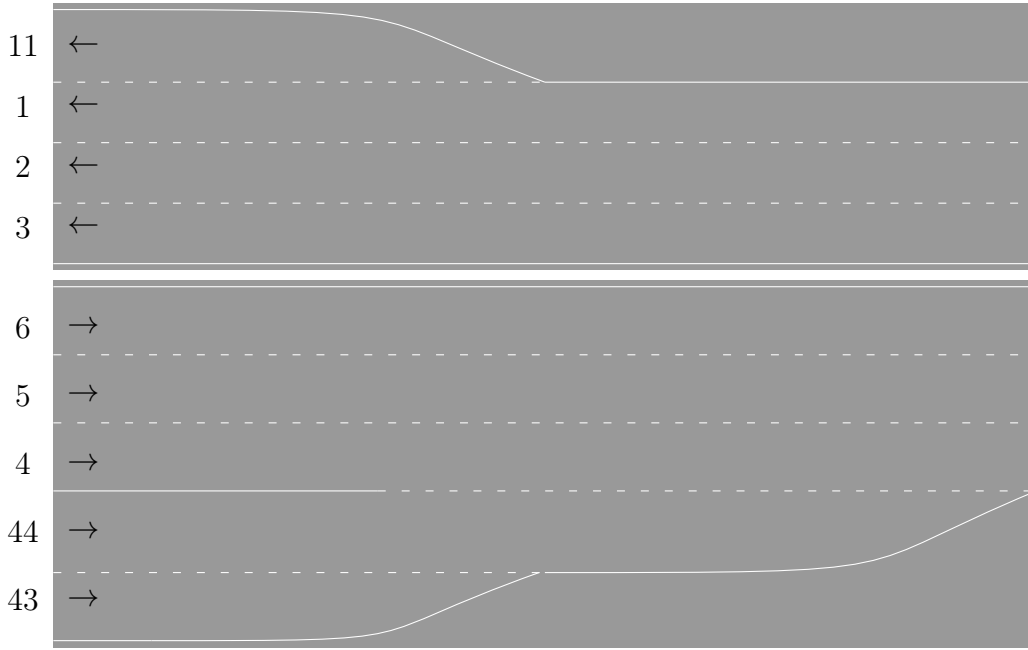


Abbildung 4.2: Von der Drohne erfasster Bereich der Autobahn A8.

eindeutige *id* bestimmt, die sich über die Aufnahme hinweg nicht ändert. Außerdem wird für alle dieser Fahrzeuge zu jedem Zeitpunkt *time* einen Fahrstreifen *lane*, Position entlang der Strecke *position* und Fahrzeuglänge *length* angegeben. Da die Drohne mit etwa 25 Bildern pro Sekunden aufzeichnet liegen die Einträge der Fahrzeuge in der Regel 40 Millisekunden auseinander.

$$D_{drone}^T = [x_0 \ x_1 \ \dots \ x_n] \in \mathbb{R}^{1 \times n},$$

$$\text{mit } x_i^T = [id \ class \ lane \ position \ time \ speed \ length],$$

$$n := \text{Anzahl Bilder} \times \text{Anzahl erfasster Fahrzeuge},$$

$$id : \text{Eindeutige FahrzeugID},$$

$$class : \text{Fahrzeugklasse} \in [Car, Truck, Van, Two - Wheeler],$$

$$lane : \text{Aktueller Fahrstreifen} \in [1, 2, 3, 4, 5, 6, 11, 43, 44],$$

$$position : \text{Position entlang der Strecke},$$

$$time : \text{Zeitstempel},$$

$$speed : \text{Aktuelle Geschwindigkeit},$$

$$length : \text{Fahrzeuglänge},$$

(4.1)

4.1.2 Betrachtung des einfahrenden freien Verkehrs im Drohnendatensatz

Für die Betrachtung des einfahrenden Verkehrs wurden die Fahrstreifen 4, 5, 6, 43 und 44 aus Abbildung 4.2 berücksichtigt. Zusätzlich wurden Aufnahmen extrahiert, die freien Verkehr enthielten mit einer durchschnittlichen Verkehrsgeschwindigkeit von $\bar{v}_{traffic} > 70 \frac{km}{h}$. Bei dieser Betrachtung kommen vier Aufnahmen in Frage deren Eigenschaften in Tabelle 4.1 zusammengefasst sind.

Über die Aufnahme r_1 hinweg wurden insgesamt 1189 Fahrzeuge mit einer durchschnittlichen Geschwindigkeit von ca. $100 \frac{km}{h}$ erfasst. Für jeden Fahrstreifen ergibt das in etwa 900 Fahrzeuge pro Stunde. Damit ist in Aufnahme r_1 auch die größte Anzahl an Fahrzeugen im betrachteten Datensatz aufgezeichnet worden. In dieser Aufnahme bewegen sich auch im Durchschnitt 3150 Fahrzeuge pro Stunde auf der dreistreifigen Autobahn und 1350 Fahrzeuge pro Stunde fahren über die Einfädungsstreifen in den Hauptstrom. Mit einer Aufnahmedauer von 15 Minuten und 50 Sekunden ist die erste Aufnahme die kürzeste von allen. Die Aufnahmen zwei und drei sind bei Betrachtung der Durchschnittsgeschwindigkeit etwas schneller als die anderen und sich in den restlichen Eigenschaften auch ähnlich. In der letzten Aufnahme ist die durchschnittlich gefahrene Geschwindigkeit der Fahrzeuge weit unter den Geschwindigkeiten der anderen Aufnahmen. Auch der Verkehrsfluss q_{on} des zufließenden Stroms ist weit unter den der anderen Aufzeichnungen.

Aufnahme	#Fahrzeuge	+	$\bar{v}[\frac{km}{h}]$	Dauer [min]	q_{in}^*	q_{on}^*
r_1	1189	897,95	98,71	15,88	3145,46	1345,33
r_2	1079	799,39	102,56	16,20	2818,96	1178,02
r_3	1214	821,68	101,03	17,73	2971,31	1141,56
r_4	1002	688,58	76,54	17,46	2632,01	810,91

$$+ \bar{v} = \frac{\#Fahrzeuge}{Stunde * Fahrstreifen}, * \frac{\#Fahrzeuge}{Stunde}$$

Tabelle 4.1: Eigenschaften des Drohnendatensatzes an der A8.



Abbildung 4.3: Geschwindigkeit je Fahrstreifen über die Zeit der Aufnahme r_1 .

Abbildung 4.3 zeigt die gefahrenen Geschwindigkeiten über die Dauer der Aufnahme auf den jeweiligen Fahrstreifen. Bei dieser Darstellung sind 3 verschiedene Geschwindigkeitsverläufe sichtbar. Die Durchschnittsgeschwindigkeit des linken Fahrstreifens ist in grün, die des mittleren in orange und die des rechten Fahrstreifens in blau dargestellt. Dabei fällt auf, dass auf dem rechten Fahrstreifen am langsamsten gefahren wird. Der mittlere Fahrstreifen wird schneller und der linke Fahrstreifen am schnellsten befahren. Zusätzlich fällt auf, dass die Standardabweichung der einzelnen Fahrstreifen trotz verschiedenen Durchschnittsgeschwindigkeiten ziemlich nahe beieinander liegt (Rechts: 10,1, Mitte: 11,4, Links: 14,1).

4.2 Einfädelungsverhalten an einer Autobahneinfahrt

In diesem Abschnitt wird das Einfädelungsverhalten der Fahrzeuge während der Aufnahmen im freien Verkehr betrachtet. Hierbei werden die von ihnen gewählten Einfädelpositionen, sowie die Zeitlücken zwischen den Fahrzeugen untersucht.

4.2.1 Wahl der Position für das Einfädeln

Die Wahl des Einfädelpunkts wird durch die Betrachtung der Positionen beim Fahrstreifenwechsel in den Hauptstrom auf der Autobahn ermittelt. Abbildung 4.4 zeigt die Boxplots der von den Fahrern gewählten Einfädelpunkte, sowie die dazugehörigen Histogramme.

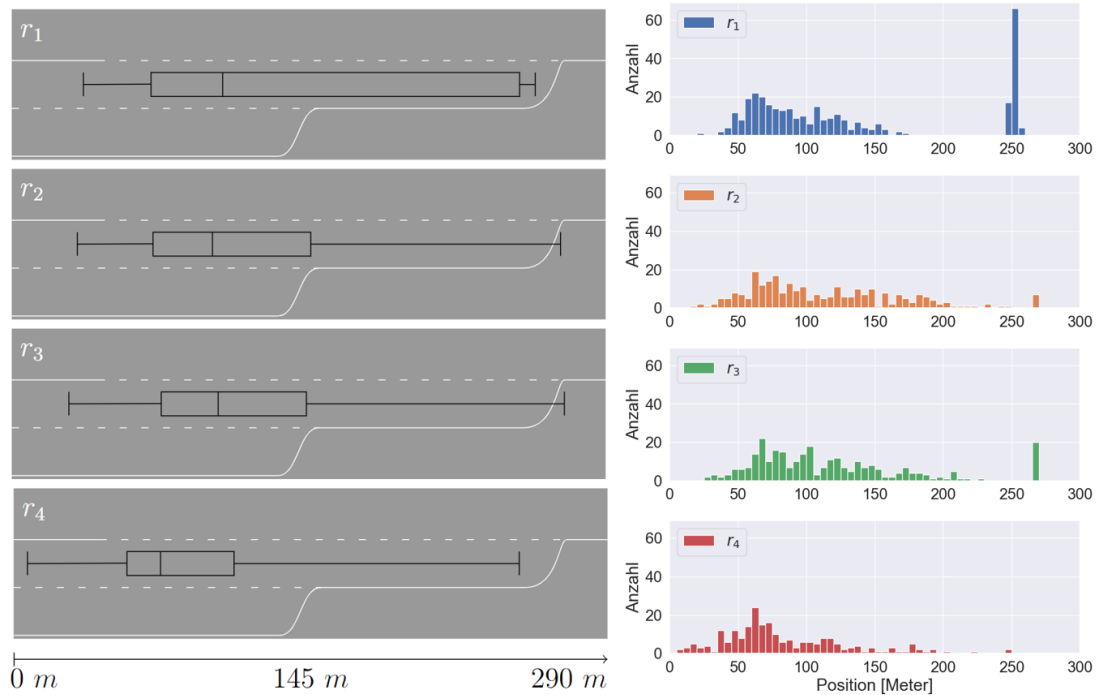


Abbildung 4.4: Verteilung der Einfädelpositionen von der Einfahrt. Links als Boxplot und rechts das dazugehörige Histogramm.

Abbildung 4.4 zeigt in der linken Hälfte die Boxplots der genutzten Einfädelpunkte. Dabei zeigen die Boxen 50% der genutzten Einfädelpunkte und die Ausprägungen der Boxen zeigen die minimal und maximal gemessenen Einfädelpunkte. Der vertikale Strich innerhalb einer Box stellt den Median dar. In der rechten Hälfte ist ein Histogramm der genutzten Einfädelpositionen. Bei der Auswertung aller Aufnahmen sind zwei Bereiche sehr auffällig. Ein häufig genutzter Bereich liegt am Ende der Einfahrt, während am Anfang der Einfahrt auch ein großer Teil der Fahrzeuge einfädelt. Auffällig ist im Übrigen, dass ein Teil der Einfahrt selten als Einfädelpunkt gewählt wird.

4.2.2 Zeitlücken auf den Fahrstreifen

Im nächsten Schritt der Analyse wurden die Zeitlücken auf dem rechten Fahrstreifen (4) untersucht. In Abbildung 4.5 sind die gemessenen Zeitlücken in Abhängigkeit zu der Position relativ zur Einfahrt dargestellt.

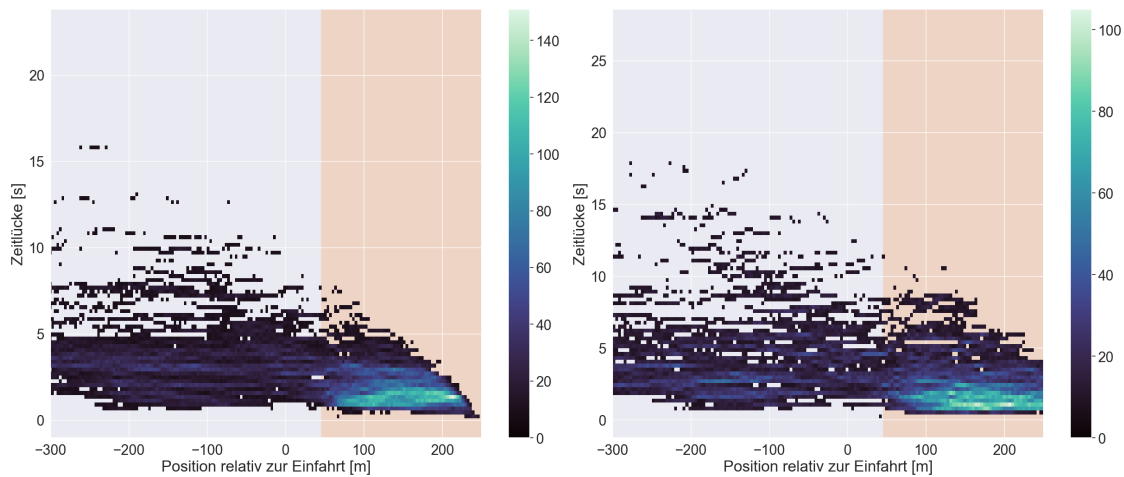


Abbildung 4.5: Nettozeitlücken in Abhängigkeit zur Position der Einfahrt.

Die Abbildung 4.5 zeigt die gemessenen Zeitlücken in Abhängigkeit zur Position der Einfahrt für die Aufnahmen r_1 und r_2 . Zusätzlich wird durch die Farbe die Häufigkeit dargestellt. Ein hell dargestellter Bereich stellt eine höhere Menge an Zeitlücken dar. Außerdem wurde der Bereich in orange gekennzeichnet, in dem ein legaler Fahrstreifenwechsel in den Hauptstrom möglich ist. Bei dieser Darstellung ist durch die begrenzte Sichtweite der Drohne zu jeder Position eine Obergrenze für die Zeitlücken sichtbar. Ist beispielsweise ein Fahrzeug am Anfang des erfassten Bereichs und ein weiteres am Ende, dann ist die Zeitlücke abhängig von der relativen Geschwindigkeit der beiden Fahrzeuge. Je weiter am Ende des Erfassungsbereichs zwei Fahrzeuge liegen, desto kleiner können die maximalen Zeitlücken sein. Dennoch fällt auf, dass die Häufigkeit an kleinen Zeitlücken in dem Bereich der Einfahrt stark zunimmt.

4.2.3 Geschwindigkeitsverläufe der einzelnen Fahrzeuge

Zuletzt wurden die Geschwindigkeitsverläufe aller Fahrzeuge auf dem rechten Fahrstreifen, sowie dem mittleren Fahrstreifen untersucht. Abbildung 4.6 zeigt die Geschwindigkeitsverläufe aller Fahrzeuge auf dem rechten Fahrstreifen. Dabei enden oder beginnen die Verläufe nicht immer von der kleinsten Position, sondern können auch auf Grund von Fahrstreifenwechsel auftauchen und verschwinden.

Hierfür wurden in Abbildung 4.6 die Geschwindigkeitsverläufe der Fahrzeuge, die ursprünglich auf dem rechten Fahrstreifen waren in blau dargestellt, während die der Fahrzeuge, die durch die Einfahrt auf den rechte Fahrstreifen gekommen sind in orange dargestellt sind. Der orangefarbene Hintergrund stellt erneut den Bereich

der Einfahrt dar, ab dem ein legaler Fahrstreifenwechsel möglich ist. Dabei ist vor dem orangenen Bereich ein Bündel an Geschwindigkeitsverläufen, die eine Geschwindigkeit von ca. 80 km/h haben, zu sehen. Ab dem orangenen eingefärbten Bereich nimmt die Standardabweichung der Geschwindigkeit zu (blau: 10,5, orange: 13,2).

Analog zu der Abbildung des rechten Fahrstreifens wird in Abbildung 4.7 der Geschwindigkeitsverlauf der Fahrzeuge auf dem mittleren Fahrstreifen abgebildet. Wieder sind die Fahrzeuge des Einfädelungstreifens orange eingefärbt. Diese haben allerdings einen geringeren Einfluss auf die Varianz der Geschwindigkeiten der Fahrzeuge im Hauptstrom.

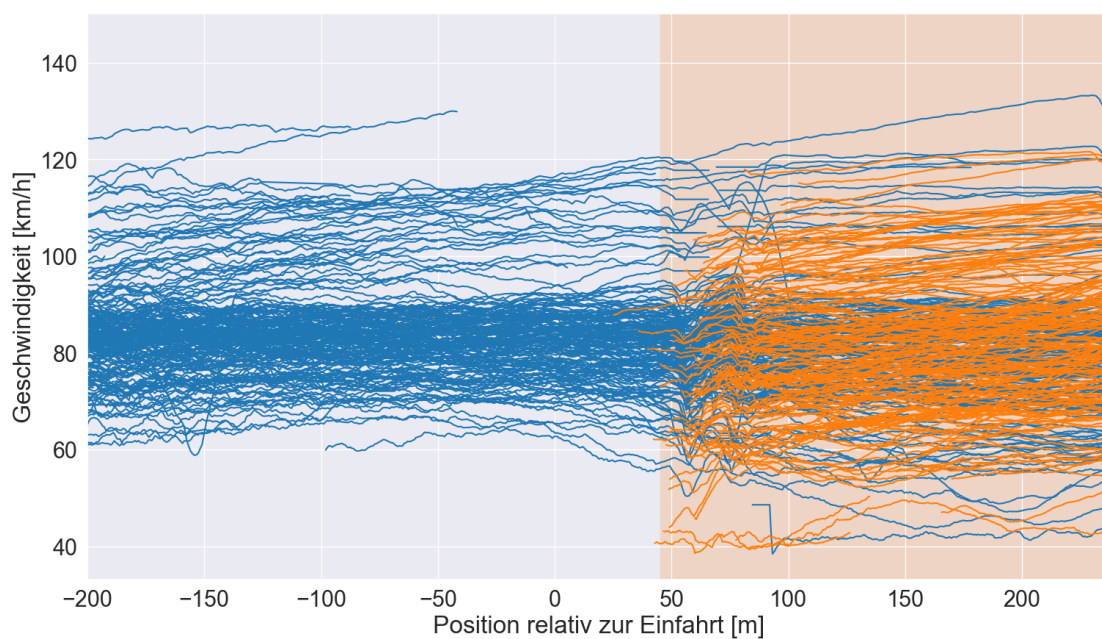


Abbildung 4.6: Geschwindigkeitsverläufe der Fahrzeuge auf dem rechten Autobahnfahrstreifen.

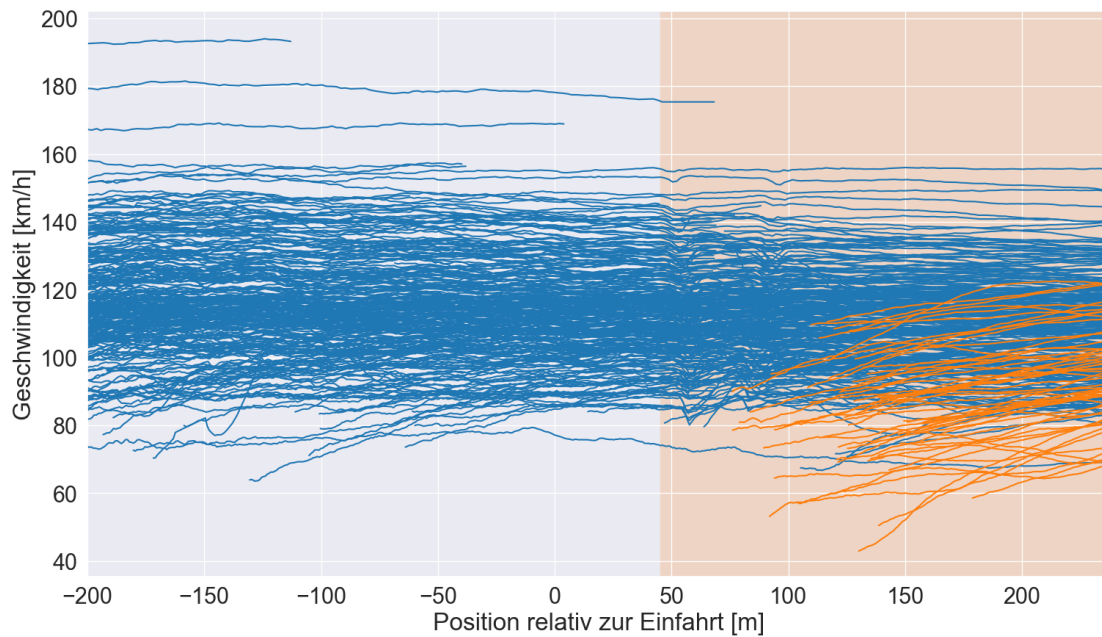


Abbildung 4.7: Geschwindigkeitsverläufe der Fahrzeuge auf dem mittleren Autobahnfahrstreifen.

4.3 Zusammenfassung

Dieses Kapitel beschreibt die Betrachtung der Fahrzeuge eines Autobahnabschnitts mithilfe einer Drohne. Dabei wurden im Rahmen einer Messkampagne mehrere Aufnahmen erstellt und mithilfe einer Nachverarbeitung der Daten ein Datensatz erstellt. Die Datenverarbeitung resultierte in einem Datensatz für jede Aufnahme mit eindeutigen IDs für erkannte Fahrzeuge, begleitet von Zeitstempeln, Fahrstreifeninformationen, Positionen entlang der Strecke und Fahrzeuglängen. Die Betrachtung des einfahrenden Verkehrs berücksichtigte spezifische Fahrstreifen, und es wurden Aufnahmen ausgewählt, die freien Verkehr mit einer Durchschnittsgeschwindigkeit von mehr als 70 km/h enthielten. Die Beobachtungen umfassten einen Abschnitt von 200 Metern bis 600 Metern der A8, abhängig vom Winkel, aus dem die Drohne die Straße betrachtete. Es wird ein grober Überblick über die Eigenschaften der verschiedenen Aufnahmen geboten. Beispielsweise wurden in der ersten Aufnahme insgesamt 1189 Fahrzeuge mit einer durchschnittlichen Geschwindigkeit von etwa 100 km/h erfasst. Die Geschwindigkeiten der Fahrstreifen variieren, wobei der rechte Fahrstreifen die niedrigsten Geschwindigkeiten aufweist und der linke die höchsten. Bei beiden Fahrstreifen ist die Standardabweichung jedoch ähnlich.

Mithilfe des Datensatzes wurde das Einfädelverhalten der einfahrenden Fahrzeuge untersucht. Die Untersuchung der Positionen für das Einfädeln erfolgt durch die Analyse der Positionen während des Fahrstreifenwechsels auf die Autobahn. In den Histogrammen für die verschiedenen Aufnahmen wird deutlich, dass es Einfädelpunkte gibt, die von allen Fahrern messbar häufig gewählt werden. Ein häufig genutzter Bereich befindet sich am Ende der Einfahrt, während auch am Anfang der Einfahrt eine signifikante Anzahl von Fahrzeugen einschert. Bemerkenswert ist, dass ein bestimmter Abschnitt der Einfahrt selten als Einfädelpunkt gewählt wird.

Zusätzlich werden die Zeitlücken des rechten Fahrstreifens untersucht. Dabei wird festgestellt, dass die Zeitlücken im Bereich der Einfahrt immer kleiner werden und oftmals im nur Bereich von 1 und 2 Sekunden liegen. In den untersuchten Aufnahmen schien die Größe der Zeitlücken im Bereich der Einfahrt den Einfädelprozess zu beeinflussen. Je größer die durchschnittlichen Zeitlücken werden, desto früher neigen die Fahrzeuge dazu einzuscheren.

Die Untersuchung der Geschwindigkeitsverläufe aller Fahrzeuge auf den verschiedenen Fahrstreifen zeigt, dass die Varianz der Geschwindigkeiten zunimmt, wenn Fahrzeuge durch die Einfahrt auf den rechten Fahrstreifen gelangten. Die Geschwindigkeitsverläufe vor der Einfahrt laufen als Bündel im Bereich von $80 \frac{\text{km}}{\text{h}}$. Die einfahrenden Fahrzeuge lösen dieses Bündel und verursachen Geschwindigkeitsanpassungen.

5 Zusammenfassung und Ausblick

Ziel dieser Forschungsarbeit war es, durch die Analyse und den Vergleich von Fahrparametern wie Fahrstrecke, Fahrzeit und Durchschnittsgeschwindigkeit einer Fahrzeugflotte das Verkehrsverhalten auf globaler Ebene besser zu verstehen. Zu diesem Zweck wurde ein umfangreicher Datensatz von Fahrdaten einer Flotte zusammengestellt und deren Eigenschaften untersucht. Ein besonderer Schwerpunkt der Arbeit wurde auf die detaillierte Analyse der Verteilungen der Fahrparameter gelegt, da sich die Verteilungen der zurückgelegten Entfernungen, der Reisezeiten und der Durchschnittsgeschwindigkeiten in Abhängigkeit von den spezifischen Bedingungen in den einzelnen Ländern unterscheiden können. Auf diese Weise konnten Erkenntnisse über das Fahrtverhalten in den einzelnen Regionen gewonnen und mögliche Ausreißer oder Besonderheiten identifiziert werden. Der länderübergreifende Vergleich von Fahrparametern half, Unterschiede und Gemeinsamkeiten in den Fahrtmustern zu identifizieren. Jedes Land weist spezifische Merkmale auf, die das Verkehrsverhalten beeinflussen können, darunter Infrastruktur, wirtschaftliche Situation, Bevölkerungsdichte und kulturelle Gewohnheiten. Der Vergleich ermöglicht es, nationale, regionale und zeitliche Besonderheiten zu erkennen, während gleichzeitig Gemeinsamkeiten auf universelle Faktoren hinweisen, die das Verkehrsverhalten weltweit in ähnlicher Weise beeinflussen. Ein weiteres Ziel der Arbeit war es, mögliche Gruppierungen von Ländern aufgrund von Ähnlichkeiten oder Unterschieden in der Fahrparameterverteilungen zu identifizieren. Eine solche Gruppierung könnte wichtige Erkenntnisse darüber liefern, welche Faktoren das Verkehrsverhalten in ähnlichen Regionen beeinflussen und welche Unterschiede möglicherweise auf spezifische lokale Bedingungen zurückzuführen sind. Unterschiedliche länderspezifische Infrastrukturen, gesetzliche Rahmenbedingungen sowie kulturelle Fahrweisen könnten tiefere Einblicke auf die Ähnlichkeit und damit in die Einflussfaktoren auf das Fahrverhalten geben. In einem letzten Schritt wurden die Fahrparameter einer weltweiten Fahrzeugflotte durch parametrisierte Verteilungsfunktionen beschrieben. Im Rahmen der Arbeit wurde intensiv untersucht, welche bekannten Verteilungsfunktionen für die real gemessenen Verteilungen von Parametern wie zurückgelegte Strecke, Reisezeit und Durchschnittsgeschwindigkeit in Frage kommen und wie diese an die analysierten Daten angepasst werden können. Unter Berücksichtigung bekannter Verteilungsfunktionen wurden die Parameter so angepasst, dass eine bestmögliche Übereinstimmung mit den empirisch

gemessenen Flottendaten erzielt wurde. Verschiedene Verteilungen wurden auf ihre Eignung getestet, wobei unterschiedliche statistische Metriken verwendet wurden. Die Ergebnisse zeigen, dass spezifische Verteilungen das empirische Fahrtverhalten charakterisieren können.

Die Analyse von Autobahnabschnitten mit Hilfe von Drohnenaufnahmen ermöglicht einen detaillierten Einblick in das Fahrzeugverhalten. In einer Messkampagne wurden mehrere Aufnahmen erstellt und anschließend zu einem umfangreichen Datensatz verarbeitet. Dieser enthielt eindeutige IDs für die detektierten Fahrzeuge, Zeitstempel, Fahrstreifeninformationen, Positionen entlang der Strecke und Fahrzeuglängen. Die Untersuchung konzentrierte sich auf das Einfädelverhalten der einfahrenden Fahrzeuge, wobei charakteristische Einfädelpunkte anhand der Häufigkeit von Fahrstreifenwechsel identifiziert werden konnten. Es zeigt sich, dass bestimmte Bereiche der Autobahneinfahrt häufig als Einfädelpunkte gewählt werden, während andere selten genutzt werden. Die Zeitlücken auf dem rechten Fahrstreifen wurden ebenfalls untersucht, wobei eine Tendenz zu kleineren Zeitlücken im Bereich der Einfahrt festgestellt wurde. Weiterhin wurde das Geschwindigkeitsverhalten aller Fahrzeuge auf den verschiedenen Fahrstreifen analysiert. Es zeigte sich, dass die Varianz der Geschwindigkeiten zunahm, wenn die Fahrzeuge über die Einfahrt auf den rechten Fahrstreifen wechselten.

Der Inhalt und die Ergebnisse dieser Arbeit beziehen sich hauptsächlich auf die Daten einer Fahrzeugflotte. Die Eckdaten der Fahrtparameter im Datensatz wurden mit verschiedenen Quellen aus der einschlägigen Literatur verglichen. Für weitere Arbeiten dieser Art wäre es wichtig, diese Daten mit Daten anderer Fahrzeughersteller in ähnlicher Größenordnung zu vergleichen und Ähnlichkeiten und Unterschiede herauszuarbeiten. Außerdem bezieht sich dieser Datensatz nur auf [PKW](#). Der Bezug zu [LKW](#) und der Vergleich mit den Ergebnissen dieser Arbeit sollte ebenfalls hergestellt werden.

Die Drohnenbeobachtung fand nur an einem Knotenpunkt der Autobahn statt. Durch mehrere Aufnahmen zu unterschiedlichen Zeiten konnten zwei charakteristische Einfädelungspunkte identifiziert werden. Es ist wichtig, weitere Drohnenbeobachtungen an einer Vielzahl von Knotenpunkten durchzuführen, um die Ergebnisse miteinander vergleichen zu können. Zusätzlich wäre eine Langzeitbeobachtung eines Knotenpunktes interessant, um den Einfluss weiterer externer Faktoren auf das individuelle Verhalten wie z.B. Fahrstreifenwahl, Position des Fahrstreifenwechsels, Einfädelpunkt zu messen. Aufgrund der eingeschränkten Sicht der Drohne auf die Autobahn wäre es wichtig, Streckenabschnitte unmittelbar vor und nach dem Bereich für das Einfädeln zu erfassen. Dieser Sichtbereich ermöglicht es, die Auswirkungen der Einflüsse durch die Einfahrt länger zu erfassen und somit zu messen, ob sich die Zeitlücken wieder vergrößern sowie die Geschwindigkeiten wieder

vereinheitlichen. Darüber hinaus haben die gewonnenen Erkenntnisse potenziellen Nutzen für die Weiterentwicklung autonomer Fahrzeugsysteme. Die Analyse von Einfädelprozessen und Interaktionen zwischen Fahrzeugen kann dazu beitragen, Algorithmen für automatisierte Fahrzeuge zu verbessern, insbesondere im Hinblick auf eine reibungslose Integration in den fließenden Verkehr. Ein präziseres Verständnis von Zeitlücken, Geschwindigkeitsanpassungen und typischen Verhaltensmustern beim Einfädeln könnte helfen, sicherere und effizientere Strategien für autonomes Fahren zu entwickeln.

Die Kombination beider Analyseansätze ermöglicht eine umfassende Untersuchung der Verkehrsdynamik. Durch die aggregierte Betrachtung können zunächst Problembereiche identifiziert werden, z.B. Streckenabschnitte mit wiederkehrenden Stauerscheinungen oder Bereiche mit auffälligen Geschwindigkeitsanpassungen. Die detaillierte Analyse einzelner Fahrzeugbewegungen an genau diesen kritischen Stellen kann dann helfen, die Ursachen dieser Verkehrsmuster besser zu verstehen. So können spezifische Faktoren wie abrupte Spurwechsel, unzureichende Einfädelspuren oder ungewöhnliches Fahrverhalten gezielt untersucht werden, um fundierte Maßnahmen zur Optimierung des Verkehrsflusses abzuleiten. Umgekehrt kann die Untersuchung von Streckenabschnitten, die besonders effizient funktionieren, wertvolle Erkenntnisse liefern. Durch die Analyse solcher Strecken können bestehende verkehrsplanerische Maßnahmen und Infrastrukturmerkmale identifiziert werden, die zu einem reibungslosen Verkehrsablauf beitragen. Diese Erkenntnisse können genutzt werden, um bestehende Problemstellen gezielt zu verbessern und erfolgreiche Konzepte auf andere Streckenabschnitte zu übertragen.

Literaturverzeichnis

- [1] MAHREZ, Zineb ; SABIR, Essaid ; BADIDI, Elarbi ; SAAD, Walid ; SADIK, Mohamed: Smart Urban Mobility: When Mobility Systems Meet Smart Data. In: *IEEE Transactions on Intelligent Transportation Systems* 23 (2022), Juli, Nr. 7, 6222–6239. <http://dx.doi.org/10.1109/TITS.2021.3084907>. – DOI 10.1109/TITS.2021.3084907. – ISSN 1524–9050, 1558–0016
- [2] ZOOK, Matthew ; KRAAK, Menno-Jan ; AHAS, Rein: Geographies of mobility: applications of location-based data. In: *International Journal of Geographical Information Science* 29 (2015), November, Nr. 11, 1935–1940. <http://dx.doi.org/10.1080/13658816.2015.1061667>. – DOI 10.1080/13658816.2015.1061667. – ISSN 1365–8816. – Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13658816.2015.1061667>
- [3] Laboratory of Emerging Transportation Systems at Chalmers University of Technology [ITS Research Lab]. In: *IEEE Intelligent Transportation Systems Magazine* 13 (2021), Nr. 4, 284–288. <http://dx.doi.org/10.1109/MITS.2021.3105974>. – DOI 10.1109/MITS.2021.3105974. – ISSN 1939–1390, 1941–1197
- [4] AJMAR, A. ; ARCO, E. ; BOCCARDO, P. ; PEREZ, F.: FLOATING CAR DATA (FCD) FOR MOBILITY APPLICATIONS. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2-W13 (2019), Juni, 1517–1523. <http://dx.doi.org/10.5194/isprs-archives-XLII-2-W13-1517-2019>. – DOI 10.5194/isprs-archives-XLII-2-W13-1517-2019. – ISSN 1682–1750. – Conference Name: ISPRS Geospatial Week 2019 (Volume XLII-2/W13) - 10–14 June 2019, Enschede, The Netherlands Publisher: Copernicus GmbH
- [5] ALTINTASI, Oruc ; TUYDES-YAMAN, Hediye ; TUNCAY, Kagan: Detection of urban traffic patterns from Floating Car Data (FCD). In: *Transportation Research Procedia* 22 (2017), 382–391. <http://dx.doi.org/10.1016/j.trpro.2017.03.057>. – DOI 10.1016/j.trpro.2017.03.057. – ISSN 23521465

- [6] ECKE, Lisa ; CHLOND, Bastian ; MAGDOLEN, Miriam ; VALLÉE, Jan ; VORTISCH, Peter: Deutsches Mobilitätspanel (MOP) – Wissenschaftliche Begleitung und Auswertungen Bericht 2020/2021: Alltagsmobilität und Fahrleistung / Karlsruher Institut für Technologie (KIT). Version: 2021. <http://dx.doi.org/10.5445/IR/1000140958>. 2021. – Forschungsbericht
- [7] HU, Ps: Summary of Travel Trends 2001 National Household Travel Survey. Version: Januar 2005. <http://dx.doi.org/10.2172/885762>. 2005 (ORNL/TM-2004/297, 885762). – Forschungsbericht. – ORNL/TM-2004/297, 885762 S.
- [8] HU, Yucheng ; LI, Minwei ; LIU, Hao ; GUO, Xiaolu ; WANG, Xiaowei ; LI, Tiejun: City traffic forecasting using taxi GPS data: A coarse-grained cellular automata model. In: *arXiv:1612.02540 [physics]* (2016), Dezember. <http://arxiv.org/abs/1612.02540>. – arXiv: 1612.02540
- [9] FOULADVAND, M. E. ; DAROONEH, A. H.: Statistical analysis of floating-car data: an empirical study. In: *The European Physical Journal B* 47 (2005), September, Nr. 2, 319–328. <http://dx.doi.org/10.1140/epjb/e2005-00325-x>. – DOI 10.1140/epjb/e2005-00325-x. – ISSN 1434-6028, 1434-6036
- [10] *Extracting travel patterns from floating car data to identify electric mobility needs: A case study in a metropolitan area.* <https://www.tandfonline.com/doi/epdf/10.1080/15568318.2021.2004629?needAccess=true&role=button>. – ISSN: 1556-8318
- [11] YONG-CHUAN, Zhang ; XIAO-QING, Zuo ; TING, Zhang li ; ZHEN-TING, Chen: Traffic Congestion Detection Based On GPS Floating-Car Data. In: *Procedia Engineering* 15 (2011), 5541–5546. <http://dx.doi.org/10.1016/j.proeng.2011.08.1028>. – DOI 10.1016/j.proeng.2011.08.1028. – ISSN 18777058
- [12] NIGRO, Marialisa ; FERRARA, Marina ; DE VINCENTIS, Rosita ; LIBERTO, Carlo ; VALENTI, Gaetano: Data Driven Approaches for Sustainable Development of E-Mobility in Urban Areas. In: *Energies* 14 (2021), Juli, Nr. 13, 3949. <http://dx.doi.org/10.3390/en14133949>. – DOI 10.3390/en14133949. – ISSN 1996-1073
- [13] HE, Wen ; LI, Deyi ; ZHANG, Tianlei ; AN, Lifeng ; GUO, Mu ; CHEN, Guisheng: Mining regular routes from GPS data for ridesharing recommendations. In: *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. Beijing China : ACM, August 2012. – ISBN 978-1-4503-1542-5, 79–86

- [14] AUER, Markus: *Empirical Analysis of Driving and Route Choice Behavior in Traffic Networks Based on Connected Vehicle Data*. München, Universität der Bundeswehr München, Doktors der Ingenieurwissenschaften (Dr.-Ing.), März 2019. <https://athene-forschung.unibw.de/doc/131488/131488.pdf>
- [15] FRANKE, Thea ; SIMS-GOULD, Joanie ; CHAUDHURY, Habib ; WINTERS, Meghan ; MCKAY, Heather: 'It makes your life worthwhile. It gives you a purpose in living': mobility experiences among active older adults with low income. In: *Ageing & Society* 39 (2019), August, Nr. 8, 1639–1666. <http://dx.doi.org/10.1017/S0144686X18000181>. – DOI 10.1017/S0144686X18000181. – ISSN 0144–686X, 1469–1779. – Publisher: Cambridge University Press
- [16] MEIJERING, Louise: Towards meaningful mobility: a research agenda for movement within and between places in later life. In: *Ageing and Society* 41 (2021), April, Nr. 4, 711–723. <http://dx.doi.org/10.1017/S0144686X19001296>. – DOI 10.1017/S0144686X19001296. – ISSN 0144–686X, 1469–1779
- [17] MUSSELWHITE, Charles ; HADDAD, Hebba: Mobility, accessibility and quality of later life. In: *Quality in Ageing and Older Adults* 11 (2010), März, Nr. 1, 25–37. <http://dx.doi.org/10.5042/qiaoa.2010.0153>. – DOI 10.5042/qiaoa.2010.0153. – ISSN 1471–7794
- [18] RASCHKA, Sebastian ; MIRJALILI, Vahid: *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, 3rd Edition*. 3. Birmingham Mumbai : Packt Publishing, 2019. – ISBN 978–1–78995–575–0
- [19] BRAMER, Max: *Principles of Data Mining*. London : Springer London, 2016 (Undergraduate Topics in Computer Science). <http://dx.doi.org/10.1007/978-1-4471-7307-6>. <http://dx.doi.org/10.1007/978-1-4471-7307-6>. – ISBN 978–1–4471–7306–9 978–1–4471–7307–6
- [20] LERMAN, Israël C.: *Foundations and Methods in Combinatorial and Statistical Data Analysis and Clustering*. London : Springer London, 2016 (Advanced Information and Knowledge Processing). <http://dx.doi.org/10.1007/978-1-4471-6793-8>. <http://dx.doi.org/10.1007/978-1-4471-6793-8>. – ISBN 978–1–4471–6791–4 978–1–4471–6793–8
- [21] CUNNINGHAM, Solveig A. ; MUIR, Jonathan A.: Data Cleaning. Version: 2023. <http://dx.doi.org/10.1017/9781009010054.022>. In: NICHOLS, Austin L. (Hrsg.) ; EDLUND, John (Hrsg.): *The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences: Volume 1: Building a Program of Research*. Cambridge : Cambridge University Press, 2023

- (Cambridge Handbooks in Psychology). – DOI 10.1017/9781009010054.022. – ISBN 978-1-316-51852-6, 443–467
- [22] TENG, Choh M.: A Comparison of Noise Handling Techniques. In: *FLAIRS Conference* (2001), Mai, S. 269–273
- [23] GUHA, Sudipto ; RASTOGI, Rajeev ; SHIM, Kyuseok: CURE: an efficient clustering algorithm for large databases. In: *ACM SIGMOD Record* 27 (1998), Juni, Nr. 2, 73–84. <http://dx.doi.org/10.1145/276305.276312>. – DOI 10.1145/276305.276312. – ISSN 0163-5808
- [24] NATHIYA, G ; PUNITHA, Mrs S C.: An Analytical Study on Behavior of Clusters Using K Means, EM and K* Means Algorithm. 7 (2010), Nr. 3
- [25] XU, Dongkuan ; TIAN, Yingjie: A Comprehensive Survey of Clustering Algorithms. In: *Annals of Data Science* 2 (2015), Juni, Nr. 2, 165–193. <http://dx.doi.org/10.1007/s40745-015-0040-1>. – DOI 10.1007/s40745-015-0040-1. – ISSN 2198-5812
- [26] EZUGWU, Absalom E. ; IKOTUN, Abiodun M. ; OYELADE, Olaide O. ; ABUALIGAH, Laith ; AGUSHAKA, Jeffery O. ; EKE, Christopher I. ; AKINYELU, Andronicus A.: A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. In: *Engineering Applications of Artificial Intelligence* 110 (2022), April, 104743. <http://dx.doi.org/10.1016/j.engappai.2022.104743>. – DOI 10.1016/j.engappai.2022.104743. – ISSN 0952-1976
- [27] HASTIE, Trevor J. ; TIBSHIRANI, Robert ; FRIEDMAN, Jerome H.: *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. New York : Springer, 2009 (Springer series in statistics). – ISBN 978-0-387-84858-7
- [28] SHIRKHOORSHIDI, Ali S. ; AGHABOZORGI, Saeed ; WAH, Teh Y.: A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data. In: *PLoS ONE* 10 (2015), Dezember, Nr. 12, e0144059. <http://dx.doi.org/10.1371/journal.pone.0144059>. – DOI 10.1371/journal.pone.0144059. – ISSN 1932-6203
- [29] 2.3. Clustering. <https://scikit-learn/stable/modules/clustering.html>
- [30] HARTIGAN, J. A. ; WONG, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1979), Nr. 1, 100–108. <http://dx.doi.org/10.2307/2346830>. – DOI 10.2307/2346830. – ISSN 0035-9254. – Publisher: [Wiley, Royal Statistical Society]

- [31] PIECH, Chris: *CS221*. <https://stanford.edu/~cpiech/cs221/handouts/kmeans.html>
- [32] W., Samuel ; A., Evans ; AHISHAKIYE, Emmanuel: Consumer Segmentation and Profiling using Demographic Data and Spending Habits Obtained through Daily Mobile Conversations. In: *International Journal of Computer Applications* 181 (2018), August, Nr. 9, 33–42. <http://dx.doi.org/10.5120/ijca2018917609>. – DOI 10.5120/ijca2018917609. – ISSN 09758887
- [33] PATEL, Sakshi ; SIHMAR, Shivani ; JATAIN, Aman: A study of hierarchical clustering algorithms. In: *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2015, 537–541
- [34] DABHI, Dipak P. ; PATEL, Mihir R.: Extensive Survey on Hierarchical Clustering Methods in Data Mining. 03, Nr. 11
- [35] *Elbow Method for optimal value of k in KMeans*. <https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/>. Version: Juni 2019. – Section: Machine Learning
- [36] SHI, Congming ; WEI, Bingtao ; WEI, Shoulin ; WANG, Wen ; LIU, Hai ; LIU, Jialei: A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. In: *EURASIP Journal on Wireless Communications and Networking* 2021 (2021), Februar, Nr. 1, 31. <http://dx.doi.org/10.1186/s13638-021-01910-w>. – DOI 10.1186/s13638-021-01910-w. – ISSN 1687–1499
- [37] UNGUREANU, Tudor ; ILIC, Mario ; RADON, Sebastian ; ROTHE, Lisa ; REICHERT, Max ; SCHÖBER, Christin ; STAMATAKIS, Ioannis ; HEINRICH, Thomas: *Vergleich der Detektoren für die Verkehrserfassung an signalisierten Knotenpunkten: = Comparison of detectors for traffic detection at signalized intersections*. Bremen : Fachverlag NW in Carl Ed. Schünemann KG, 2020 (Berichte der Bundesanstalt für Strassenwesen V, Verkehrstechnik Heft V 336). – ISBN 978–3–95606–530–9
- [38] BROCKFELD, Elmar: Benefits and limits of recent floating car data technology - an evaluation study.
- [39] TREIBER, Martin ; KESTING, Arne: *Traffic Flow Dynamics: Data, Models and Simulation*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2013. <http://dx.doi.org/10.1007/978-3-642-32460-4>. <http://dx.doi.org/10.1007/978-3-642-32460-4>. – ISBN 978–3–642–32459–8 978–3–642–32460–4

- [40] TREITERER, J. ; CLEAR, D. ; HOWARTH, D. ; LEE, Jong J. ; MYERS, John J.: INVESTIGATION OF TRAFFIC DYNAMICS BY AERIAL PHOTOGRAMMETRY TECHNIQUES. (1975), Nr. OHIO-DOT-09-75, FCP 40T2-052, 224 p. <https://trid.trb.org/view/115205>. – Number: Ees-278-3
- [41] U.S. DEPARTMENT OF TRANSPORTATION FEDERAL HIGHWAY ADMINISTRATION: *Next Generation Simulation (NGSIM) Vehicle Trajectories and Supporting Data*. <http://dx.doi.org/10.21949/1504477>. Version: 2017
- [42] BARMPOUNAKIS, Emmanouil ; GEROLIMINIS, Nikolas: On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment. In: *Transportation Research Part C: Emerging Technologies* 111 (2020), Februar, 50–71. <http://dx.doi.org/10.1016/j.trc.2019.11.023>. – DOI 10.1016/j.trc.2019.11.023. – ISSN 0968090X
- [43] KRAJEWSKI, Robert ; BOCK, Julian ; KLOEKER, Laurent ; ECKSTEIN, Lutz: The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. Maui, HI : IEEE, November 2018. – ISBN 978-1-72810-321-1 978-1-72810-323-5, 2118–2125
- [44] BOCK, Julian ; KRAJEWSKI, Robert ; MOERS, Tobias ; RUNDE, Steffen ; VATER, Lennart ; ECKSTEIN, Lutz: *The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections*. <http://arxiv.org/abs/1911.07602>. Version: November 2019. – arXiv:1911.07602 [cs, eess]
- [45] KRAJEWSKI, Robert ; MOERS, Tobias ; BOCK, Julian ; VATER, Lennart ; ECKSTEIN, Lutz: The roundD Dataset: A Drone Dataset of Road User Trajectories at Roundabouts in Germany. In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. Rhodes, Greece : IEEE, September 2020. – ISBN 978-1-72814-149-7, 1–6
- [46] MARCZAK, Florian ; DAAMEN, Winnie ; BUISSON, Christine: Merging behaviour: Empirical comparison between two sites and new theory development. In: *Transportation Research Part C: Emerging Technologies* 36 (2013), November, 530–546. <http://dx.doi.org/10.1016/j.trc.2013.07.007>. – DOI 10.1016/j.trc.2013.07.007. – ISSN 0968090X
- [47] KERNER, Boris: *Introduction to Modern Traffic Flow Theory and Control: The Long Road to Three-Phase Traffic Theory*. 2009. <http://dx.doi.org/10.1007/978-3-642-02605-8>. <http://dx.doi.org/10.1007/978-3-642-02605-8>. – ISBN 978-3-642-02604-1

- [48] MAY, A. D.: *TRAFFIC FLOW FUNDAMENTALS*. 1990 <https://trid.trb.org/view/356201>. – ISBN 978-0-13-926072-8
- [49] *Highway Capacity Manual 7th Edition: A Guide for Multimodal Mobility Analysis*. Washington, D.C. : National Academies Press, 2022. <http://dx.doi.org/10.17226/26432>. <http://dx.doi.org/10.17226/26432>. – ISBN 978-0-309-27566-8
- [50] GARTNER, Nathan ; MESSER, C. ; RATHI, A.: *Traffic Flow Theory: A State-of-the-Art Report*. 2001
- [51] WANG, Yimin ; HE, Zhaocheng: Mesoscopic modelling and analysis of traffic flow based on stationary observations. In: *Procedia Computer Science* 151 (2019), 800–807. <http://dx.doi.org/10.1016/j.procs.2019.04.109>. – DOI 10.1016/j.procs.2019.04.109. – ISSN 18770509
- [52] WILSON, Robin J.: *Introduction to graph theory*. 4. ed., [Nachdr.]. Harlow Munich : Prentice Hall, 2009. – ISBN 978-0-582-24993-6
- [53] DIESTEL, Reinhard ; DIESTEL, Reinhard (Hrsg.): *Graph Theory*. Berlin, Heidelberg : Springer, 2017 (Graduate Texts in Mathematics). http://dx.doi.org/10.1007/978-3-662-53622-3_1. http://dx.doi.org/10.1007/978-3-662-53622-3_1. – ISBN 978-3-662-53622-3
- [54] *OSMnx: Python for Street Networks*. <https://geoffboeing.com/2016/11/osmnx-python-street-networks/>. Version: November 2016
- [55] BOEING, Geoff: OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. In: *Computers, Environment and Urban Systems* 65 (2017), September, 126–139. <http://dx.doi.org/10.1016/j.compenvurbsys.2017.05.004>. – DOI 10.1016/j.compenvurbsys.2017.05.004. – ISSN 01989715
- [56] PORTA, Sergio ; CRUCITTI, Paolo ; LATORA, Vito: The Network Analysis of Urban Streets: A Primal Approach. In: *Environment and Planning B: Planning and Design* 33 (2006), Oktober, Nr. 5, 705–725. <http://dx.doi.org/10.1068/b32045>. – DOI 10.1068/b32045. – ISSN 0265-8135, 1472-3417
- [57] CHAO, Pingfu ; XU, Yehong ; HUA, Wen ; ZHOU, Xiaofang: *A Survey on Map-Matching Algorithms*. <http://arxiv.org/abs/1910.13065>. Version: Oktober 2019. – arXiv:1910.13065 [cs]
- [58] NEWSON, Paul ; KRUMM, John: Hidden Markov Map Matching Through Noise and Sparseness, 2009, 336–343

- [59] ECKE, Lisa ; CHLOND, Bastian ; MAGDOLEN, Miriam ; VORTISCH, Peter: Deutsches Mobilitätspanel (MOP) – Wissenschaftliche Begleitung und Auswertungen Bericht 2019/2020: Alltagsmobilität und Fahrleistung / Karlsruher Institut für Technologie (KIT). Version: 2020. <http://dx.doi.org/10.5445/IR/1000126557>. 2020. – Forschungsbericht. – Medium: PDF
- [60] NOBIS, Claudia ; KUHNIMHOF, Tobias: *Mobilität in Deutschland - MiD: Ergebnisbericht*. <https://www.mobilitaet-in-deutschland.de>. Version: Dezember 2018
- [61] HOOGENDOORN-LANSER, Sascha ; SCHAAP, Nina T. W. ; OLDEKALTER, Marie-José: The Netherlands Mobility Panel: An Innovative Design Approach for Web-based Longitudinal Travel Data Collection. In: *Transportation Research Procedia* 11 (2015), Januar, 311–329. <http://dx.doi.org/10.1016/j.trpro.2015.12.027>. – DOI 10.1016/j.trpro.2015.12.027. – ISSN 2352–1465
- [62] BUSBY, Jeffrey R.: Accessibility-Based Transit Planning.
- [63] WARADE, Ritesh K.: The Accessibility and Development Impacts of New Transit Infrastructure: The Circle Line in Chicago.
- [64] DUCAS, Caroline R.: Incorporating Livability Benefits into the Federal Transit Administration New Starts Project Evaluation Process through Accessibility-Based Modeling.
- [65] GEURS, Karst T. ; WEE, Bert van: Accessibility evaluation of land-use and transport strategies: review and research directions. In: *Journal of Transport Geography* 12 (2004), Juni, Nr. 2, 127–140. <http://dx.doi.org/10.1016/j.jtrangeo.2003.10.005>. – DOI 10.1016/j.jtrangeo.2003.10.005. – ISSN 0966–6923
- [66] SILVA, Cecília: *Comparative accessibility for mobility management : the structural accessibility layer*. Porto, Portugal, University of Porto, Diss., Juli 2008
- [67] MCGUCKIN, N. ; FUCCI, A.: Summary of Travel Trends: 2017 National Household Travel Survey / Federal Highway Administration. Version: Juli 2018. <https://trid.trb.org/view/2239026>. Rockville, MD, Juli 2018 (FHWA-PL-18-019). – Forschungsbericht. – 148 S.
- [68] *National Travel Survey: 2020*. <https://www.gov.uk/government/statistics/national-travel-survey-2020/national-travel-survey-2020>

- [69] EUROPEAN COMMISSION. JOINT RESEARCH CENTRE. INSTITUTE FOR PROSPECTIVE TECHNOLOGICAL STUDIES.: *Analysis of national travel statistics in Europe: OPTIMISM WP2 : harmonisation of national travel statistics in Europe*. LU : Publications Office, 2013 <https://data.europa.eu/doi/10.2788/59474>
- [70] OU, Shiqi ; YU, Rujie ; LIN, Zhenhong ; REN, Huanhuan ; HE, Xin ; PRZESMITZKI, Steven ; BOUCHARD, Jessey: Intensity and daily pattern of passenger vehicle use by region and class in China: estimation and implications for energy use and electrification. In: *Mitigation and Adaptation Strategies for Global Change* 25 (2020), März, Nr. 3, 307–327. <http://dx.doi.org/10.1007/s11027-019-09887-0>. – DOI 10.1007/s11027-019-09887-0. – ISSN 1381–2386, 1573–1596
- [71] FABBRI, Renato ; DE LEÓN, Fernando G.: *A Statistical Distance Derived From The Kolmogorov-Smirnov Test: specification, reference measures (benchmarks) and example uses*. <http://arxiv.org/abs/1711.00761>. Version: Oktober 2017. – arXiv:1711.00761 [physics]
- [72] LEVENE, Mark ; KONONOVICIUS, Aleksejus: Empirical Survival Jensen-Shannon Divergence as a Goodness-of-Fit Measure for Maximum Likelihood Estimation and Curve Fitting. In: *Communications in Statistics - Simulation and Computation* 50 (2021), November, Nr. 11, 3751–3767. <http://dx.doi.org/10.1080/03610918.2019.1630435>. – DOI 10.1080/03610918.2019.1630435. – ISSN 0361–0918, 1532–4141. – arXiv:1809.11052 [physics, q-fin, stat]
- [73] *Goodness-of-fit tests based on Kullback-Leibler information / IEEE Journals & Magazine / IEEE Xplore*. <https://ieeexplore.ieee.org/document/1331678>
- [74] EUBANK, R. L.: Testing Goodness of Fit with Multinomial Data. In: *Journal of the American Statistical Association* 92 (1997), Nr. 439, 1084–1093. <http://dx.doi.org/10.2307/2965573>. – DOI 10.2307/2965573. – ISSN 0162–1459. – Publisher: [American Statistical Association, Taylor & Francis, Ltd.]
- [75] PANARETOS, Victor M. ; ZEMEL, Yoav: Statistical Aspects of Wasserstein Distances. In: *Annual Review of Statistics and Its Application* 6 (2019), März, Nr. 1, 405–431. <http://dx.doi.org/10.1146/annurev-statistics-030718-104938>. – DOI 10.1146/annurev-statistics-030718-104938. – ISSN 2326–8298, 2326–831X. – arXiv:1806.05500 [stat]
- [76] TAYLOR, Michael A.: Fosgerau’s travel time reliability ratio and the Burr distribution. In: *Transportation Research Part B: Methodological* 97 (2017),

- März, 50–63. <http://dx.doi.org/10.1016/j.trb.2016.12.001>. – DOI 10.1016/j.trb.2016.12.001. – ISSN 01912615
- [77] TANG, Jinjun ; LIU, Fang ; WANG, Yinhai ; WANG, Hua: Uncovering urban human mobility from large scale taxi GPS data. In: *Physica A: Statistical Mechanics and its Applications* 438 (2015), November, 140–153. <http://dx.doi.org/10.1016/j.physa.2015.06.032>. – DOI 10.1016/j.physa.2015.06.032. – ISSN 03784371
- [78] ZHANG, Shen ; TANG, Jinjun ; WANG, Haixiao ; WANG, Yinhai ; AN, Shi: Revealing intra-urban travel patterns and service ranges from taxi trajectories. In: *Journal of Transport Geography* 61 (2017), Mai, 72–86. <http://dx.doi.org/10.1016/j.jtrangeo.2017.04.009>. – DOI 10.1016/j.jtrangeo.2017.04.009. – ISSN 09666923
- [79] WANG, Wenjun ; PAN, Lin ; YUAN, Ning ; ZHANG, Sen ; LIU, Dong: A comparative analysis of intra-city human mobility by taxi. In: *Physica A: Statistical Mechanics and its Applications* 420 (2015), Februar, 134–147. <http://dx.doi.org/10.1016/j.physa.2014.10.085>. – DOI 10.1016/j.physa.2014.10.085. – ISSN 03784371
- [80] JIANG, Shixiong ; GUAN, Wei ; ZHANG, Wenyi ; CHEN, Xu ; YANG, Liu: Human mobility in space from three modes of public transportation. In: *Physica A* (2017), S. 12
- [81] SONG, Ha Y. ; LEE, Jun S.: Finding a simple probability distribution for human mobile speed. In: *Pervasive and Mobile Computing* 25 (2016), Januar, 26–47. <http://dx.doi.org/10.1016/j.pmcj.2015.10.015>. – DOI 10.1016/j.pmcj.2015.10.015. – ISSN 1574–1192
- [82] SHI, Chaoyang ; LI, Qingquan ; LU, Shiwei ; YANG, Xiping: Modeling the Distribution of Human Mobility Metrics with Online Car-Hailing Data—An Empirical Study in Xi'an, China. In: *ISPRS International Journal of Geo-Information* 10 (2021), April, Nr. 4, 268. <http://dx.doi.org/10.3390/ijgi10040268>. – DOI 10.3390/ijgi10040268. – ISSN 2220–9964
- [83] SCIPY DEVELOPERS: *Statistical functions (scipy.stats) — SciPy v1.15.3 Manual*. <https://docs.scipy.org/doc/scipy/reference/stats.html>. Version: 2024

Publikationen

- [a] Oliver Stein, Jonas Paczia, Hubert Rehborn *Deep data analysis of a drone data set: moving queues at intersections*, 2021 ITS World Congress Hamburg, Hamburg, 2021.
- [b] Oliver Stein, Stefan Kaufmann, Michael Watzko, Hubert Rehborn, Jonas Paczia *Learning from drone observations on highways: characteristics of vehicular traffic at an on-ramp*, 2021 IEEE 11th International Conference on Consumer Electronics ICCE-Berlin 2021, Berlin, Deutschland, 2021.
- [c] Oliver Stein, Silja Hoffmann, Jonas Paczia, Hubert Rehborn, Micha Koller *Mobilitätskennwerte für Elektro- und Verbrennerfahrzeuge aus Flottendaten*, Straßenverkehrstechnik Ausgabe 12/2021, 2021.
- [d] Oliver Stein, Silja Hoffmann, Jonas Paczia, Hubert Rehborn *Internationale Mobilitätskennwerte aus Flottendaten*, Straßenverkehrstechnik Ausgabe 09/2022, 2022.
- [e] Oliver Stein, Hubert Rehborn, Jonas Paczia, Silja Hoffmann *Similarity of Mobility Metrics of Connected Vehicles around the Globe*, Transportation Research Board Annual Meeting, Washington, USA, 2023.
- [e] Oliver Stein, Hubert Rehborn, Jonas Paczia, Silja Hoffmann *Similarity of Mobility in Different Cities with the aid of Fleeting Car Data*, 15th ITS European Congress, Lissabon, Portugal, 2023.
- [A] Michael Buchholz et al. *Handling Occlusions in Automated Driving Using a MEC Server-based Environment Model From Infrastructure Sensors*, IEEE Intelligent Transportation Systems Magazine, Vol. 14, Nr. 3, Seite 106-120, 2022,
- [B] Jonas Paczia et al. *Congested Pattern Analysis in the Framework of Kerner's Three-Phase Traffic Theory: A Comparison Between U.S. and German Probe Vehicle Data* , Transportation Research Board Annual Meeting, 2023.
- [C] Jonas Paczia et al. *Connected Vehicle Data: Precise Congestion Recognition in Traffic Simulations and Empirical Measurements* , Transportation Research Board Annual Meeting, 2023.

Patente

- [I] Oliver Stein, Hubert Rehborn, Micha Koller, Frank Starman, *Verfahren zur Routenführung eines Fahrzeugs und Fahrzeug*, Deutsches Patent- und Markenamt, 2021.
- [II] Hubert Rehborn, Oliver Stein, Jonas Paczia, *Methode zur energieeffizienten Fahrweise von automatisierten Fahrzeugen mit Reinforcement-Lernen* Deutsches Patent- und Markenamt, 2021.
- [III] Hubert Rehborn, Oliver Stein, Frank Starman, *Verfahren zum verbesserten Aktualisierungsverhalten von Verkehrsdiensten* Deutsches Patent- und Markenamt, 2021.

Anhang A

A Anhang

A.1 Fahrtparameter verschiedener Städte

A.1.1 Randdaten der Fahrten

	Stuttgart	Wien	Rom	Los Angeles	Chicago	Moskau
Anzahl	159 Tsd	203 Tsd	117 Tsd	66 Tsd	100 Tsd	58 Tsd
Distanz						
Mittel <i>km</i>	13,02	11,41	12,69	9,52	11,62	14,39
SD <i>km</i>	24,78	22,76	25,84	17,09	21,06	19,86
Median <i>km</i>	5,60	5,22	5,42	4,72	4,42	7,80
25 % - 75 % Quantil <i>km – km</i>	2,15- 14,08	2,1- 11,42	2,06- 12,47	2,04- 8,90	1,91 - 12,01	3,2 - 17,44
99 % Quantil <i>km</i>	131,21	104,10	127,27	80,33	82,24	92,87
Fahrtdauer						
Mittel <i>min : ss</i>	21:02	23:58	23:47	18:51	13:21	43:59
SD <i>min : ss</i>	21:18	27:14	25:49	20:00	23:23	45:27
Median <i>min : ss</i>	15:50	15:57	15:51	13:50	15:52	29:49
25 % - 75 % Quantil <i>min : ss – min : ss</i>	07:50- 27:42	07:51- 29:50	07:50- 30:04	06:35- 23:51	14:00- 30:07	05:50- 56:30
99 % Quantil <i>hh : mm : ss</i>	01:29:51	02:27:54	02:12:29	01:38:25	01:49:50	03:01:36

	Stuttgart	Wien	Rom	Los Angeles	Chicago	Moskau
Geschwindigkeit						
Mittel $\frac{km}{h}$	29,91	24,57	26,07	26,73	24,27	20,34
SD $\frac{km}{h}$	19,53	16,66	17,72	16,73	18,07	12,34
Median $\frac{km}{h}$	24,20	20,03	21,30	23,33	18,33	17,95
25 % - 75 % Quantil $\frac{km}{h} - \frac{km}{h}$	16,77– 37,64	14,67– 28,35	14,70 – 31,77	16,32– 32,03	12,66 – 29,64	11,44 – 26,39
99 % Quantil $\frac{km}{h}$	97,67	92,59	92,59	88,27	87,50	62,39

Tabelle A.1: Zusammenfassung der Fahrtparameter im Städtevergleich.

A.1.2 Empirische kumulative Verteilungsfunktionen

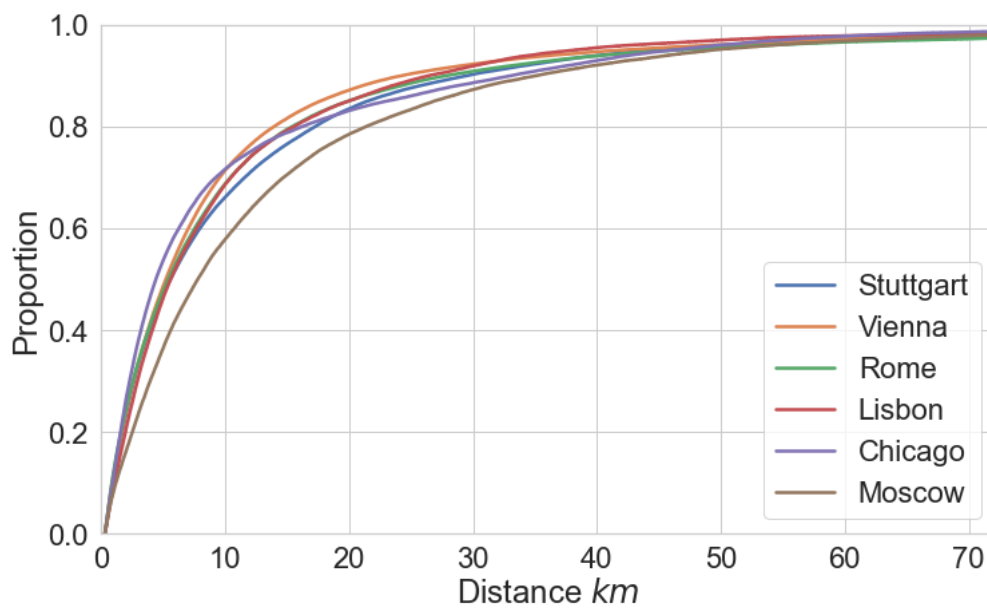


Abbildung A.1: ECDF der exemplarischen Distanzen exemplarischer Städte.

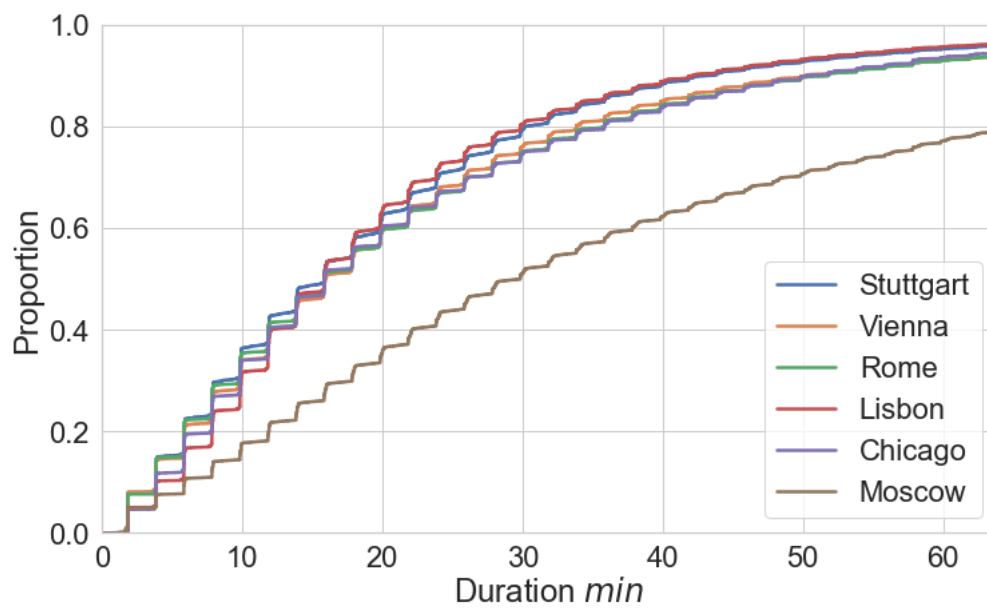


Abbildung A.2: ECDF der Fahrtauern exemplarischer Städte.

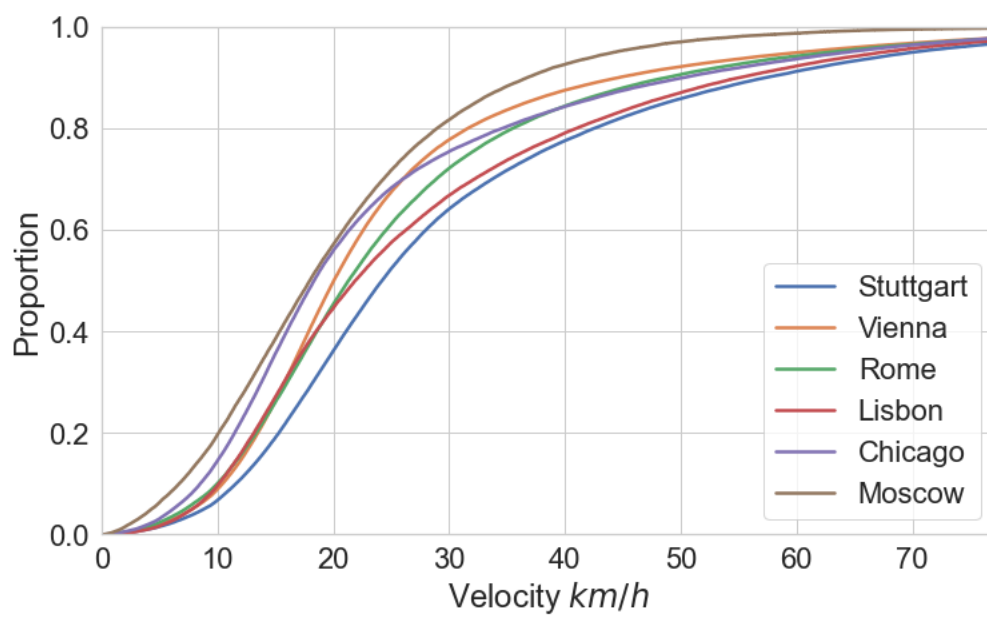
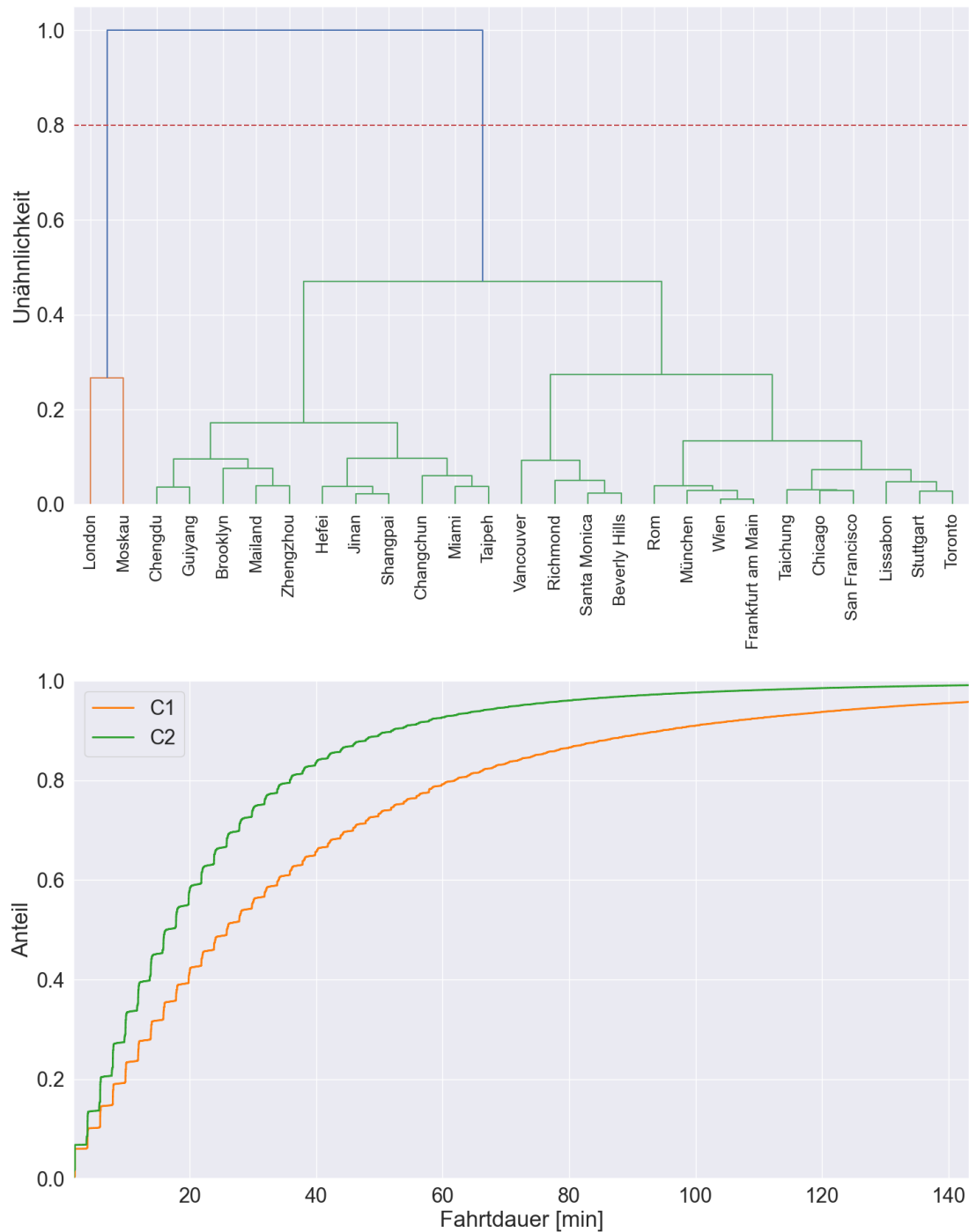


Abbildung A.3: ECDF der mittleren Geschwindigkeiten exemplarischer Städte.

A.1.3 Betrachtung der Dendrogramme von Clustering der Städteverteilungen



(a)

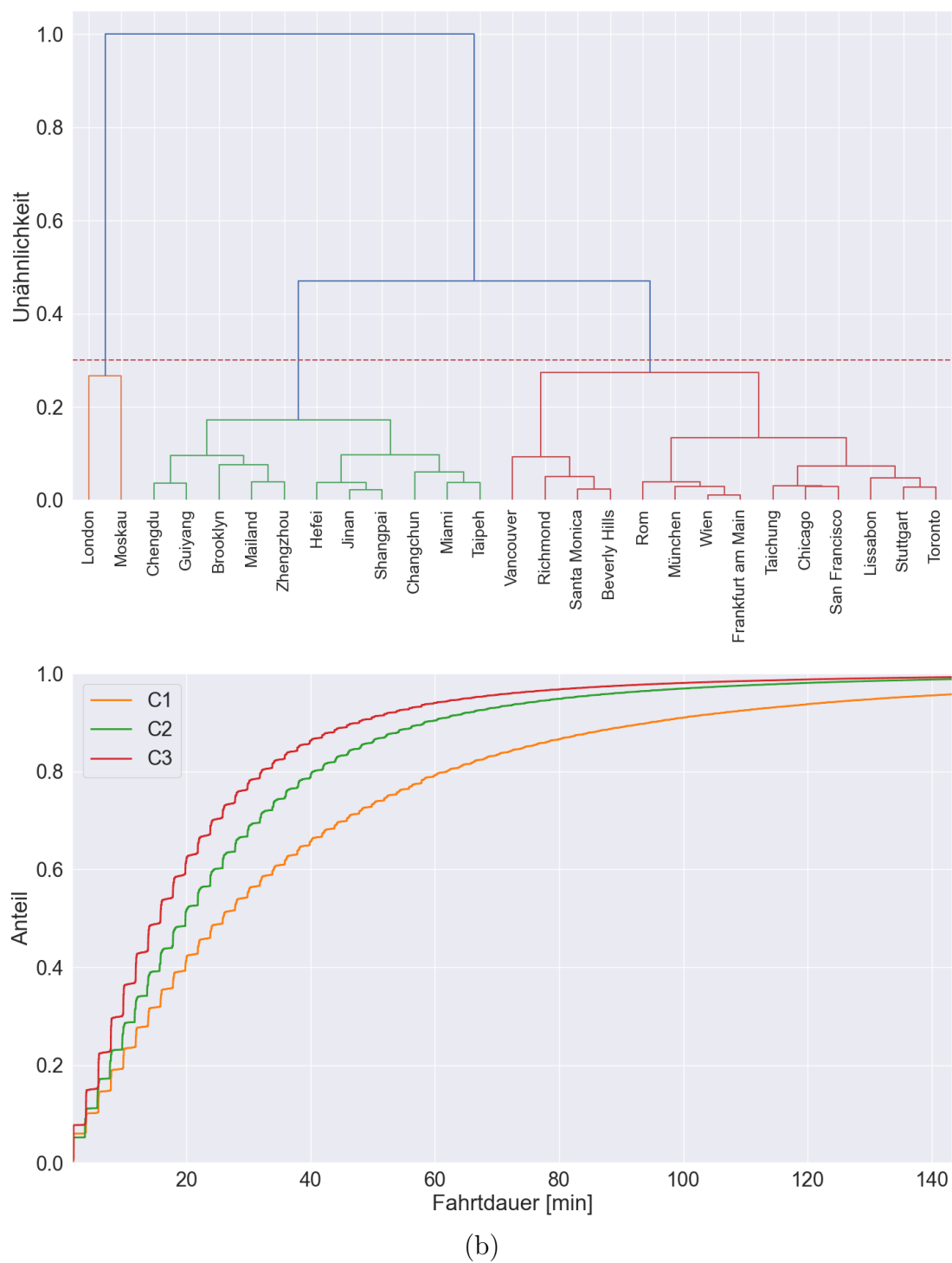
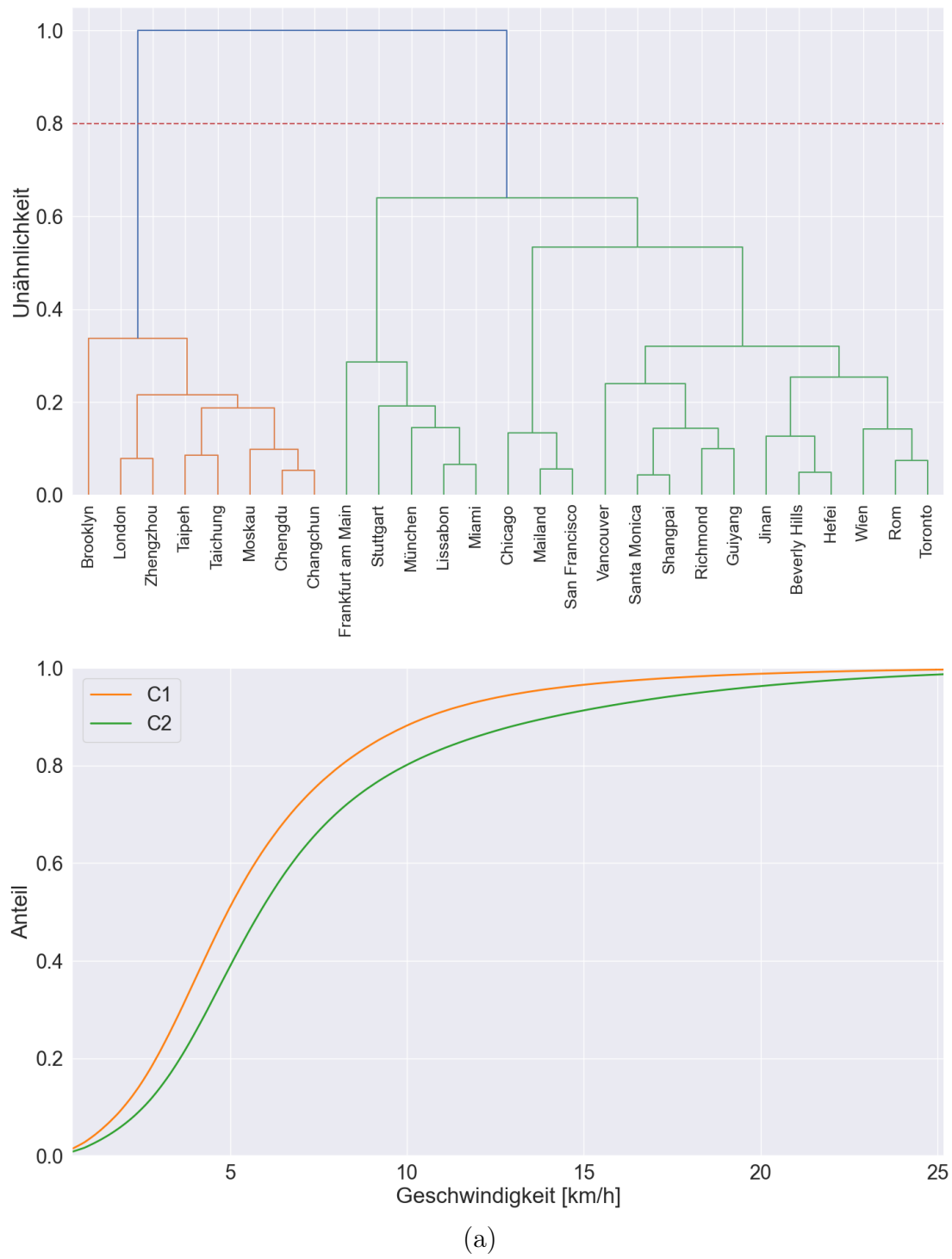


Abbildung A.4: Dendrogram der Fahrtzeiten und der sich dadurch resultierenden Cluster von Städten. In (a) mit Schnitt bei einer Unähnlichkeit $d = 0.8$ und (b) bei $d = 0.3$.



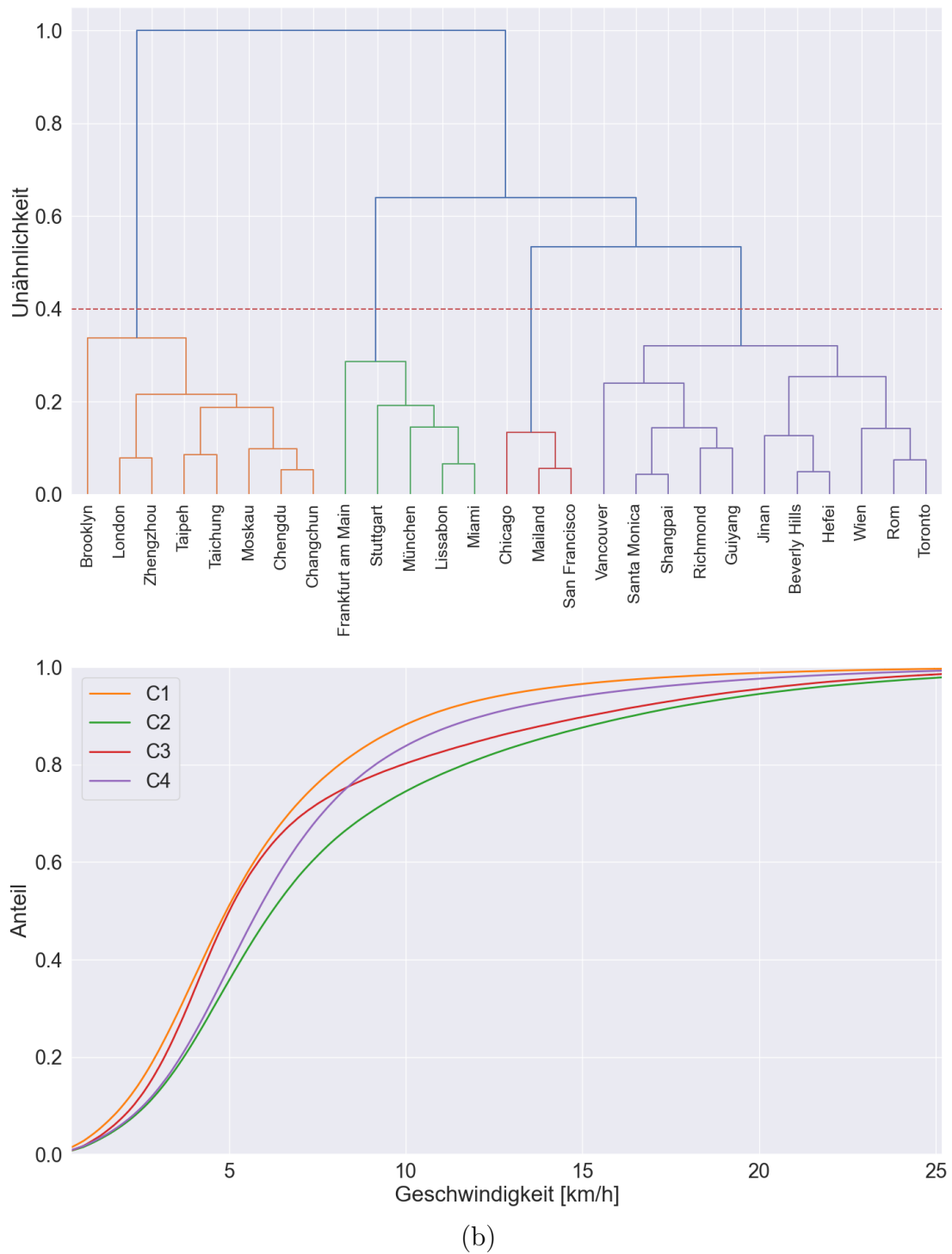


Abbildung A.5: Dendrogram der mittleren Geschwindigkeiten und der sich dadurch resultierenden Cluster von Städten. In (a) mit Schnitt bei einer Unähnlichkeit $d = 0.8$ und (b) bei $d = 0.4$.