

**Schriftenreihe
Verkehr**

Heft 35

Dominik Bieland

**Ableitung von Quelle-Ziel-Matrizen im ÖPNV
auf Basis von WLAN- und Bluetooth-Daten**

Institut für Verkehrswesen

**Fachgebiet Verkehrsplanung
und Verkehrssysteme**

Universität Kassel

kassel
university



press

Schriftenreihe Verkehr der Universität Kassel

Herausgeber:

Institut für Verkehrswesen

Dominik Bieland

**Ableitung von Quelle-Ziel-Matrizen im ÖPNV auf Basis
von WLAN- und Bluetooth-Daten**

Die vorliegende Arbeit wurde vom Fachbereich Bauingenieur- und Umweltingenieurwesen an der Universität Kassel als Dissertation zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften (Dr.-Ing.) angenommen.

Erster Gutachter: Univ.-Prof. Dr.-Ing. Carsten Sommer, Universität Kassel
Zweiter Gutachter: Univ.-Prof. Dr.-Ing. Markus Friedrich, Universität Stuttgart

Tag der mündlichen Prüfung

27. Oktober 2022



Diese Veröffentlichung – ausgenommen Zitate und anderweitig gekennzeichnete Teile – ist unter der Creative-Commons-Lizenz Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-SA 4.0: <https://creativecommons.org/licenses/by-sa/4.0/deed.de>) lizenziert.

Zugl.: Kassel, Univ., Diss. 2022

ISBN: 978-3-7376-1120-6

DOI: <https://doi.org/10.17170/kobra-202306018138>

Schriftenreihe Verkehr

Heft 35 – Juli 2023

Herausgeber:

Institut für Verkehrswesen

Universität Kassel

34109 Kassel

© 2023, kassel university press, Kassel

<https://kup.uni-kassel.de>

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen

Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über

<http://dnb.dnb.de> abrufbar

Vorwort des Herausgebers

Der öffentliche Personennahverkehr (ÖPNV) in Deutschland steht vor großen Herausforderungen: Demografische und sozioökonomische Veränderungen führen zu einem Rückgang der an den ÖV-gebundenen Zwangskunden und neuen, i. d. R. höheren Anforderungen an das ÖPNV-System. Darüber hinaus hat die Corona-Pandemie das Mobilitätsverhalten verändert, so dass bis heute die Fahrgastnachfrage unter dem Niveau des Vor-Corona-Jahres 2019 liegt. Der ÖPNV gilt jedoch als ein Baustein für die Einhaltung der verkehrsspezifischen Klimaschutzziele und soll dementsprechend seinen Marktanteil gegenüber dem motorisierten Individualverkehr (MIV) erheblich steigern.

Parallel bietet die Digitalisierung neue Möglichkeiten bei Planung, Betrieb und Marketing des ÖPNV. Vor diesem Hintergrund stehen viele Aufgabenträger und Verkehrsunternehmen vor der Aufgabe, einerseits neue Kundengruppen durch ein qualitativ hochwertiges Angebot zu gewinnen und andererseits dabei die Kosten für die Leistungserbringung im Rahmen zu halten. Die Lösung dieser komplexen Aufgabe lässt sich nur durch eine Vielzahl von Strategien und Maßnahmen in sämtlichen Bereichen der Aufgabenträger bzw. Verkehrsunternehmen umsetzen. Ein wesentlicher Bereich ist dabei die Nahverkehrs- und Angebotsplanung, die aktuelle Daten und valide Prognoseergebnisse zur Verkehrsnachfrage benötigt.

In diesem Kontext steht die Dissertation von Herrn Bieland. In der vorliegenden Arbeit wird ein Verfahren vorgestellt, das auf Basis von in Bussen erfassten WLAN-, Bluetooth- und Daten aus automatischen Fahrgastzählssystemen (AFZS) Quelle-Ziel-Matrizen für Linienfahrten ermittelt. Neben der Entwicklung des Verfahrens bildet vor allem die Kalibrierung, Validierung und Evaluation des Verfahrens anhand realer Mess- und Referenzdaten ein Schwerpunkt der Arbeit. Das Verfahren wurde exemplarisch anhand mehrerer Linienfahrten in Nordhessen getestet und hinsichtlich ihrer Güte evaluiert. Die Dissertation entstand im Rahmen des vom Bundesministerium für Digitales und Verkehr (BMDV) geförderten Forschungsprojektes MobileDataFusion, bei dem Herr Bieland für die Entwicklung und Validierung des Verfahrens verantwortlich war.

Herr Bieland liefert mit seiner interdisziplinären Forschungsarbeit nicht nur einen bedeutenden Beitrag im Forschungsfeld der Datenerfassung und des Datenmanagements im Verkehr, sondern darüber hinaus ein Verfahren für die Praxis, das die Datengrundlagen bei der Nahverkehrs- und Angebots-

planung verbessern kann. Er knüpft mit seiner Dissertation an das Themenfeld der integrierten Verkehrsplanung an, das einen Forschungsschwerpunkt des Fachgebiets darstellt, und ergänzt damit die Kompetenz des Fachgebiets bei der IT-gestützten Erfassung und Verarbeitung großer Datenmengen im Verkehr.

Kassel, im Mai 2023

Carsten Sommer

Danksagung

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Fachgebiet Verkehrsplanung und Verkehrssysteme der Universität Kassel und darüber hinaus. Mit der Fertigstellung der vorliegenden Dissertation geht ein intensiver und lehrreicher Lebensabschnitt zu Ende.

An erster Stelle möchte ich meinen herzlichen Dank an meinen Doktorvater Herrn Univ.-Prof. Dr.-Ing. Carsten Sommer richten. Die stetige Förderung und Unterstützung, die intensiven methodischen Diskussionen in allen von mir bearbeiteten Forschungsprojekten und das von mir gefühlte unerschütterliche Vertrauen in meine Fähigkeiten haben maßgeblich dazu beigetragen, dass diese Dissertation überhaupt entstanden ist. Ebenfalls möchte ich mich herzlich bei Herrn Univ.-Prof. Dr.-Ing. Markus Friedrich für die fachlichen Anregungen, die positive Zusprache und insbesondere für die Bereitschaft zur Übernahme des Zweitgutachtens bedanken.

Ein großer Dank gebührt auch allen Kolleginnen und Kollegen des Fachgebiets Verkehrsplanung und Verkehrssysteme der Universität Kassel, die ich in meiner Zeit als Wissenschaftlicher Mitarbeiter kennenlernen durfte, für die stets ehrliche und offene Kritik, die Hilfsbereitschaft und nicht zuletzt die angenehme, freundschaftliche Arbeitsatmosphäre am Lehrstuhl.

Auch den Projektpartnern von MobileDataFusion gilt mein Dank für die kollegiale und konstruktive Zusammenarbeit, die intensiven Diskussionen sowie die schnellen Reaktionen bei Rückfragen und Weiterentwicklungswünschen der Software.

Für die fortwährende und tatkräftige Unterstützung meines bisherigen beruflichen und privaten Lebensweges gilt mein besonderer Dank meiner Familie, die mir sowohl Freiräume für das Schreiben der Dissertation als auch für das neue Familienleben zu Dritt geschaffen hat.

Kurzfassung

Demografische und sozioökonomische Entwicklungen sowie Veränderungen im Mobilitätsverhalten führen zu weniger Zwangskunden im ÖPNV. Gleichzeitig steigt das Kundenpotential, das sich situationsabhängig für ein bestimmtes Verkehrsmittel entscheidet (multimodales Verhalten) und im Zuge der Verkehrswende werden tarifliche Maßnahmen wie Nulltarif oder Bürgertickets diskutiert, die das Potential zu einer hohen und kurzfristigen Steigerung der Fahrgastnachfrage besitzen. Informationen zur Nachfrage erhalten damit einen Bedeutungsgewinn, um die bestehenden Potentiale zu nutzen und auf ggf. kurzfristige Nachfragesteigerungen reagieren zu können. Gleichzeitig steigt der Smartphone-Besitz in Deutschland seit Jahren an. Vor dem Hintergrund, dass die meisten Smartphones regelmäßig und ohne das Zutun des Nutzers Signale aussenden, um WLAN-Netze (Wireless Local Area Network) oder Bluetooth-Geräte zu finden, fallen bei entsprechenden Routern in Reichweite entlang des realisierten Weges Registrierungs- und ggf. Nutzungsdaten an. Diese Daten besitzen das Potential in Verknüpfung mit anderen Datenquellen (AFZS) ein Gesamtbild der Verkehrsnachfrage im ÖPNV-Netz aufzubauen, das für die ÖPNV-Planung und Marketing verwendet werden kann. Diese stets aktuelle Datengrundlage kann insbesondere in den Bereichen der (Angebots-)Planung, Einnahmenaufteilung sowie Tarifkonzeption einen Mehrwert bieten.

Das zentrale Ziel dieser Dissertation stellt daher die Entwicklung und Anwendung eines Verfahrens dar, das basierend auf AFZS-Daten, die in vielen Verkehrsunternehmen und -verbänden vorliegen, sowie automatisch erfassbaren Daten unterschiedlicher Quellen (WLAN- und Bluetooth-Daten) die Fahrgastnachfrage in Form von linienspezifischen Quelle-Ziel-Matrizen automatisch, mit ausreichender Genauigkeit und für lange Erhebungszeiträume, ermittelt. WLAN-Daten bezeichnen hier WLAN Probe Requests, die bei der Suche von Endgeräten nach Access Points entstehen. Bluetooth-Daten bezeichnen Inquiry-Response-Nachrichten, die als Antwort auf eine Inquiry-Nachricht durch (aktive und sichtbare) Endgeräte in der Reichweite versandt werden.

Linien- und tagesscharfe Quelle-Ziel-Matrizen ermöglichen ÖPNV-Betreibern nach Wochentagen und Tageszeiten differenzierte Auswertungen, deren Ergebnisse in den Planungs- und Betriebsprozess eingebracht werden können. Gleichzeitig ist die benötigte Infrastrukturausrüstung kostengünstig. Verkehrsunternehmen können die Daten somit eigenständig erfassen, sodass

sie sich weder in eine Abhängigkeit von anderen Datenzulieferern (etwa Netzbetreiber für Mobilfunkdaten) begeben noch durch eine andauernde finanzielle Belastung für den (wiederkehrenden) Kauf von Daten belastet werden.

Die Datenerfassung erfolgt innerhalb der Linienbusse verschiedener Linien im Untersuchungsraum des Nordhessischen Verkehrsverbundes (Linien 54, 100, 500). Die Linienfahrzeuge wurden dazu mit der technischen Hardware zur Datenerfassung ausgestattet. Die datenschutzrechtlichen Belange wurden bereits bei der Erfassung der Daten berücksichtigt sowie bei allen weiteren Verfahrensschritten. Alle durch die Fahrzeuge erfassten WLAN- und Bluetooth-Daten, die sich aus Signalen von Fahrgästen und Nicht-Fahrgästen (Fußgänger, Radfahrer, Pkw-Fahrer etc.) zusammensetzen, stellen die Datengrundlage für das Verfahren dar.

Da lediglich die Daten der Fahrgäste für die Ermittlung der Quelle-Ziel-Relationen von Relevanz sind, werden die erfassten Daten zunächst in zwei Stufen gefiltert. Stufe 1 filtert unplausible und nicht nutzbare Daten, wie etwa Daten mit fehlerhafter Zuordnung von Fahrtinformationen. Die zweite Stufe der Filterung hat das Ziel, mit hoher Wahrscheinlichkeit Daten von Nicht-Fahrgästen herauszufiltern, etwa auf Basis der Luftlinienentfernung zwischen der ersten und letzten Erfassung eines Endgeräts. Dabei können Daten sowohl „false positive“ (Daten von Nicht-Fahrgästen, die als Fahrgäste interpretiert werden) und „false negative“ (Daten von Fahrgästen, die als Nicht-Fahrgäste interpretiert werden) sein, da die zweite Stufe der Filterung mit Unsicherheiten verbunden ist. Aus diesem Grund wurden im Rahmen der Verfahrensentwicklung verschiedene Filtereinstellungen in Szenarien untersucht und hinsichtlich ihrer Güte verglichen.

Die verbliebenen Daten werden in Quelle-Ziel-Matrizen aufbereitet. Die Zeilen und Spalten bestehen aus den Haltestellen der betrachteten Linienfahrt, die Matrixelemente ergeben sich aus der Zuordnung des ersten und letzten erfassten Signals eines Endgeräts zu Haltestellen, die den potentiellen Ein- und Ausstiegsort markieren. Das Ergebnis dieses Modellschritts ist die Ausgangsmatrix und bildet die Basis für die Hochrechnung. Es ist zu berücksichtigen, dass Unsicherheiten bei der Festlegung der Ein- und Ausstiegshaltestelle vorliegen. Diese sind auf das heterogene Sendeverhalten der verschiedenen Endgeräte zurückzuführen. Das Sendeverhalten ist von mit zahlreichen Faktoren abhängig, allen voran dem Hersteller, der Betriebssystemversion sowie der Nutzung des Endgeräts. Dies führt dazu, dass die erste Erfassung eines Endgeräts sowohl unmittelbar mit, als auch mehrere Minuten nach

dem Einstieg erfolgen kann. Um diese Unsicherheiten bei der Festlegung der Ein- und Ausstiegshaltestelle zu berücksichtigen, wurde die Fuzzy-Set-Theorie angewandt, die eine scharfe Zuteilung von Ein- und Ausstiegsort vermeidet und stattdessen unscharfe Mengen für einzelne Quelle-Ziel-Relationen ermittelt. Die Zugehörigkeitsfunktion berücksichtigt dabei alle Haltestellen, die bereits zwanzig Minuten vor Erfassung des ersten Signals eines Endgeräts angefahren wurden, als potentielle Einstiegsorte. Dies gilt analog für alle kommenden Haltestellen nach Erfassung des letzten Signals eines Endgeräts für den Ausstiegsort.

Mit der Hochrechnung wird die Ausgangsmatrix mittels eines angepassten Steigerungsfaktorenmodells, das die mathematische Schwäche der Modelle bei strukturellen Nullwerten berücksichtigt, schrittweise so korrigiert, dass die Randsummen eingehalten werden. Randsummen bilden hier die Ein- und Aussteigerzahlen je Haltestelle, die mit hoher Genauigkeit durch das AFZS vorliegen. Im Ergebnis liegt eine Bewertungsmatrix vor. Es ist zu beachten, dass Steigerungsfaktorenmodelle eine starke Abhängigkeit von der Ausgangsmatrix besitzen (Prinzip der Informationsminimierung). In der Ausgangsmatrix vorhandene Verzerrungen können sich daher negativ auf die Ergebnisse auswirken. Zur Überprüfung des Mehrwerts der WLAN- und Bluetooth-Daten werden zwei Hochrechnungsverfahren betrachtet, die keine Eingangsdaten benötigen: Das Verfahren nach Li und Cassidy, das lediglich auf den Randsummen (AFZS-Daten) basiert und das Fratar-Verfahren unter Zugrundelegung einer Einser-Matrix (alle Quelle-Ziel-Relationen erhalten den Wert 1). Mit der Hochrechnung wird die Wahrscheinlichste unter den möglichen Matrizen, die die Randsummen einhalten, bestimmt. Dabei ist zu beachten, dass unter allen möglichen Lösungen nur eine die Realität abbildet – dies muss jedoch nicht die wahrscheinlichste Lösung sein.

Die Bewertung der Verfahrensergebnisse erfolgt unter Zuhilfenahme verschiedener Gütemaße, die die Unterschiede der Modellschätzung (Bewertungsmatrix) im Bezug zu der realen Verteilung der Nachfrage auf die Quelle-Ziel-Relationen der betrachteten Linienfahrten (Referenzmatrix) quantifizieren. Die Referenzmatrizen basieren auf Fahrgastbefragungen, die als Vollerhebung einzelner Linienfahrten der ausgestatteten Buslinien durchgeführt wurden. Zur Bewertung der Ergebnisse werden der GEH-Wert, die Fehlerquadratsumme, die Abweichung der Verkehrsleistung in verschiedenen Entfernungsklassen sowie das sogenannte „Treffermaß“ und der „Überfüllungsfaktor“ genutzt. Das Treffermaß gibt kurzgefasst an, welcher Anteil der Zellen

sowohl in der Bewertungsmatrix als auch in der Referenzmatrix mit einem Wert größer 0 belegt ist. Mit dem Überfüllungsfaktor wird gezeigt, ob in der Bewertungsmatrix mehr, weniger oder gleich viele Zellen gegenüber der Referenzmatrix belegt sind.

Der Mehrwert des Verfahrens wird auf Basis des Vergleichs der erzielten Ergebnisse unter Berücksichtigung der WLAN- und Bluetooth-Daten mit einer gleichverteilten Matrix, die keine Informationen zur Verteilung der Quelle-Ziel-Relationen in der Realität enthält, abgeleitet.

Zur Erstellung der Bewertungsmatrizen für insgesamt 29 vorliegende Referenzfahrten werden verschiedene Filter- und Hochrechnungssets verwendet. Insgesamt wurden so über 30.000 Bewertungsmatrizen für die Auswertung ermittelt – mehr als 1.100 je Referenzfahrt. Die Auswertung der Ergebnisse der betrachteten Parameterkombinationen erfolgte sowohl differenziert für die einzelnen Gütemaße als auch in Form einer gütemaßübergreifenden Bewertung. Darüber hinaus wird die gütemaßübergreifende Bewertung für verschiedene Klasseneinteilungen der Referenzfahrten – nach Linie und nach Fahrgastnachfrage – durchgeführt. Für die Bewertung der Ergebnisse wurde auf eine Normierung zurückgegriffen. Dazu wurden die erzielten Ergebnisse in den betrachteten Gütemaßen in vier Merkmalen zu Punktwerten zusammenfasst, unter Berücksichtigung ihrer Bedeutung gewichtet und aufsummiert.

Auf Basis dieses Vorgehens wurden sieben Parameterkombinationen ermittelt, die die höchsten Punktwerte erzielen. Diese Parameterkombinationen greifen sowohl auf WLAN- als auch auf Bluetooth-Daten als Quelle zurück. Die Eingangsdaten zur Erstellung der Ausgangsmatrix werden hier durch eine Filterung auf Basis der GPS-Luftlinienentfernung reduziert. Bei differenzierter Betrachtung der Referenzfahrten nach Linien und Nachfrageklassen hat sich insbesondere die Parameterkombination aus WLAN- und Bluetooth-Daten mit Filterset P51_GPS200 mit Fuzzifizierung und Hochrechnung mittels Fratar-Verfahren (Hilfswert 1,0000, Reduktionsgrenzwert 0,5) als robust erwiesen: Diese Parameterkombination liegt in jeder betrachteten Klasse mindestens unter den 10% der besten Parameterkombinationen.

Darüber hinaus schneiden die ermittelten Parameterkombinationen stets besser ab als das beste Verfahren ohne Ausgangsdaten, ohne Filterung der Eingangsdaten sowie ohne Hochrechnung der Ausgangsmatrizen. Diese Erkenntnisse bestätigen den Mehrwert des Verfahrens sowie der entwickelten Verfahrensschritte.

Gleichzeitig ist zu berücksichtigen, dass einzelne betrachtete Parameterkombinationen insgesamt jedoch nur bei maximal 13 der 29 betrachteten Referenzfahrten (etwa 45%) die GEH-Grenzwerte einhalten. Auf einzelnen Referenzfahrten konnten die Grenzwerte durch keine der berücksichtigten Parameterkombinationen eingehalten werden. Die erforderliche Genauigkeit wird somit bei einem Großteil der betrachteten Referenzfahrten nicht erreicht. Daraus ergibt sich die Folgerung, dass das Verfahren in der praktischen Anwendung eine hohe Wahrscheinlichkeit besitzt, dass die Zuordnung der Fahrgäste auf die Quelle-Ziel-Relationen Fehler oberhalb der gesetzten Grenzwerte aufweist.

Die Gründe für diese Ergebnisse wurden ausführlich untersucht: Insgesamt hat sich gezeigt, dass eine Vielzahl an Datenpunkten erfasst wird, die zu großen Teilen nicht von Fahrgästen, sondern von Störquellen stammen. Die Informationen, die innerhalb der erfassten Datenpakete übermittelt werden, sind darüber hinaus nicht geeignet, Daten von Fahrgästen eindeutig von Stördaten trennen zu können. Die damit bestehenden Verzerrungen können nicht allein auf Basis der Hochrechnung und damit dem Ausgleich der Matrixelemente auf die Randsummen ausgeglichen werden. Trotz dieser Verzerrungen konnte ein Mehrwert der Nutzung von WLAN- und Bluetooth-Daten gegenüber Vergleichsverfahren erreicht werden, die nur auf den Randsummen basieren.

Dabei ist zu berücksichtigen, dass das Verfahren bislang für 29 Referenzfahrten betrachtet wurde. Die Stichprobe für die Bewertung der Verfahrensergebnisse ist damit gering, sodass der Standardfehler steigt. Zufällige Ereignisse haben damit einen sehr großen Einfluss auf das Ergebnis. Das Verfahren sollte daher auf weiteren Linienfahrten sowie ggf. Linien geprüft werden.

Bei der Bewertung der Verfahrensergebnisse vor dem Hintergrund der angestrebten Einsatzbereiche (Angebotsplanung, Tarifplanung, Einnahmenaufteilung, Kalibrierung von Verkehrsmodellen) ist zu beachten, dass die Anforderungen an die räumliche und zeitliche Genauigkeit abhängig vom Anwendungszweck sind. Darüber hinaus ist die Qualität bisher genutzter Quelle-Ziel-Daten zu berücksichtigen. Quelle-Ziel-Daten stammen bislang aus Verkehrsmodellen und basieren auf Haushalts- und/oder Fahrgastbefragungen. Da eine repräsentative Erhebung von Quelle-Ziel-Verflechtungen mit einem erheblichen Aufwand verbunden wäre, sind Hochrechnungen notwendig, die zu bislang nicht quantifizierten Verzerrungen und Ungenauigkeiten in den bestehenden Daten führen. Demgegenüber wurden im Rahmen der

Verfahrensbewertung die Ungenauigkeiten der Verfahrensergebnisse bereits für 29 Referenzfahrten anhand verschiedener Gütemaße ermittelt.

Mit dem aufgezeigten Verfahren besteht die Möglichkeit, dauerhaft Daten zu erfassen und somit eine bislang nicht vorhandene Datengrundlage zu bilden. Den ÖPNV-Betreibern wird es damit u. a. ermöglicht Informationen zur Nachfrage wiederkehrender Events (etwa Fußballspiele, Konzerte o. ä.) zu sammeln und das Angebot entsprechend anzupassen, um etwa Kapazitätsüberschreitungen zu vermeiden. Gleichzeitig können die Daten auf Tarifzonenebene aggregiert werden und in die Entwicklung von Tarifen eingehen.

Das in dieser Dissertation entwickelte Verfahren kann als Grundbaustein genutzt werden, um Quelle-Ziel-Daten abzuleiten. Zur Optimierung der Ergebnisse können weitere oder zusätzliche Datenquellen integriert werden, die eine höhere Qualität als die in dieser Arbeit genutzten WLAN- und Bluetooth-Daten besitzen. Dazu zählen Datenquellen mit eindeutigen Zuordnungsmerkmal – etwa durch Nutzung von E-Ticket-Daten oder Daten einer eigens für die Erfassung von Quelle-Ziel-Relationen der Fahrgäste entwickelten App. Da das Verfahren offen für eine Integration zusätzlicher Datenquellen ist, können auch weitere Datenquellen wie Daten aus Verbindungsanfragen, Bus-WLAN, einem stadtweiten WLAN oder Vertriebsdaten integriert werden. Die Eignung der Daten sowie deren Nutzbarkeit sind jedoch methodisch zu prüfen. Bei Nutzung mehrerer und zusätzlicher Datenquellen ist insbesondere zu prüfen, ob es sich um disjunkte Mengen an Daten einzelner Fahrgäste bzw. Endgeräte handelt. Bei der Verwendung von WLAN- und Bluetooth-Daten ist zu berücksichtigen, dass Maßnahmen zum Schutz der Privatsphäre der Smartphone-Nutzer seitens der OEM zunehmen, etwa durch die MAC-Randomisierung. Dies führt zu einer sinkenden Anzahl verwertbarer Daten.

Aus Sicht des Autors besitzt bereits das bestehende Verfahren großes Potential – insbesondere für Anwendungsfälle, in denen lange Fahrtabschnitte (>30 Minuten zwischen Haltestellen) dominieren. Mit steigendem zeitlichen Abstand zwischen Haltestellen ist darüber hinaus davon auszugehen, dass die Erfassungsquote der Fahrgäste zwischen den Haltestellen zunimmt und damit die Genauigkeit der Zuordnung von Ein- und Ausstiegshaltestelle steigt. Weiterhin ist offen, ob eine zeitliche Aggregation, d. h. die Aggregation mehrerer Linienfahrten des gleichen Tagtyps und der gleichen Zeitschicht dazu führen können, die Güte der Verfahrensergebnisse zu verbessern. Auch

die Verbesserung der Verfahrensgüte durch räumliche Aggregation – ggf. in Kombination mit der zeitlichen Aggregation – sind detailliert zu prüfen.

Derzeit zielt das Verfahren darauf ab, Quelle-Ziel-Relationen auf Linienfahrebene zu betrachten. Für die Angebotsplanung, wie auch für die Einnahmeverteilung ist aber die Betrachtung von Verbundfahrten von Relevanz. Daher ist weiterhin auf Basis zusätzlicher Fahrzeugdaten oder zusätzlicher Datenquellen die Güte des Verfahrens zur Ermittlung von Umstiegen, bzw. Verbundfahrten zu prüfen.

Abstract

Demographic and socio-economic developments as well as changes in the mobility behaviour led to fewer mandatory customers in public transport. At the same time, the potential of customers who choose a certain means of transport depending on the situation (multimodal behaviour) is increasing. Furthermore, tariff measures, such as zero fares or citizen tickets, are being discussed, which have the potential for a short-term increase in passenger demand. Information on the demand is thus becoming more important in order to utilize the existing potential and to be able to react to any short-term increases in demand. At the same time, smartphone ownership in Germany has been on the rise for years. Given that most smartphones regularly send out signals to find WLAN (Wireless Local Area Network) networks or Bluetooth devices without the user's involvement, the corresponding routers within range along the realized route receive registration and usage data. In combination with other data sources (APC - automatic passenger counting) these data have the potential to provide insight into the overall traffic demand in the public transport network, which can be used for public transport planning and marketing. This constantly updated data basis offers added value, especially in the areas of (service) planning, revenue sharing, and fare design.

The central objective of this dissertation is therefore the development and application of a procedure which, based on already existing data in the transport companies and transport associations (APC data) as well as automatically collectable data from different sources (WLAN and Bluetooth data), automatically determines passenger demand in the form of line-specific origin-destination matrices with sufficient accuracy and for long data collection periods. WLAN data refers to WLAN probe requests that are generated when devices search for access points. Bluetooth data refers to inquiry response messages sent in response to an inquiry message by (active and visible) devices within range.

Line- and day-based origin-destination matrices enable public transport operators to perform detailed evaluations broken down by day of the week and time of day. The results of which can be incorporated into the planning and operating process. At the same time, the required infrastructure equipment is cost-effective. Transport companies are thus able to collect the data independently, so that they are neither dependent on other data suppliers nor burdened by a permanent financial burden for the (recurring) purchase of data.

The data capture is carried within the vehicles of different lines in the study area of the North Hessian Transport Association (lines 54, 100, 500). The buses were equipped with the technical hardware for data collection. Data privacy concerns were already taken into account when the data was collected and in all subsequent steps of the process. All WLAN and Bluetooth data collected by the vehicles, which are composed of signals from passengers and non-passengers (pedestrians, cyclists, automobile drivers, etc.), form the data basis for the approach.

Since only passenger data is of relevance for determining origin-destination relations, the collected data is initially filtered in two steps. Step 1 filters out implausible and unusable data, such as data with incorrect assignment of journey information. The second step of filtering aims to filter out data from non-passengers with a high confidence level, such as based on the linear distance between the first and last capture of a device. In this process, false positives (data from non-passengers interpreted as passengers) and false negatives (data from passengers interpreted as non-passengers) cannot be excluded. The second step of filtering is accordingly associated with uncertainties, so that various filter settings were examined in scenarios and compared in terms of their quality as part of the process development.

The remaining data are processed in origin-destination matrices. The rows and columns consist of the stops of the line trip under consideration, the matrix elements result from the assignment of the first and last detected signal of a device to the respective stops marking the potential boarding and alighting location. The result of this model step is the initial matrix and forms the basis for the extrapolation. It should be noted that there are uncertainties in the definition of the boarding and alighting stops. These are due to the heterogeneous transmission behaviour of the different devices. The transmission behaviour varies with numerous factors, first and foremost the manufacturer, the operating system version and the use of the device. As a result, the first detection of a device can occur both immediately upon boarding and several minutes after boarding. To account for these uncertainties in determining the boarding and alighting stops, fuzzy set theory was applied, which avoids sharp assignment of boarding and alighting locations and instead determines fuzzy sets for individual origin-destination relations. The membership function takes into account all stops that were already approached twenty minutes before the first signal of a terminal was detected as potential boarding

locations. This applies analogously to all upcoming stops after the detection of the last signal of a terminal for the alighting location.

With the extrapolation, the initial matrix is corrected gradually by means of an adapted iterative proportional fitting model, which takes the mathematical weakness of the models with structural zero values into account, so that the sums of the rows and columns are met. The marginal totals here are the boarding and alighting counts per stop, which are available with a high degree of accuracy from the APC. The result is an evaluation matrix. It must be taken into account that iterative proportional fitting models have a strong dependence on the initial matrix (principle of information minimisation). Distortions present in the initial can therefore have a negative effect on the results. To verify the added value of the WLAN and Bluetooth data, two extrapolation procedures are considered that do not require any input data: The Li and Cassidy method, which is based only on the marginal sums (AFZS data), and the Fratar method, which is based on an matrix where all origin-destination relations are assigned the value 1. The extrapolation determines the most probable among the possible matrices that meet the marginal sums. It should be noted that among all possible solutions, only one represents reality - but this does not necessarily have to be the most probable solution.

The evaluation of the results of the approach is carried out with the help of different quality measures, which quantify the differences of the model estimation (evaluation matrix) in relation to the real distribution of the origin-destination relations of the considered regular journeys (reference matrix). The reference matrices are based on passenger surveys that were conducted as full surveys of individual line trips of the equipped bus routes. To evaluate the results, the GEH value, the sum of error squares, the deviation of the transport performance in different distance classes as well as the so-called "hit measure" and the "overcrowding factor" are used. The hit measure briefly indicates which proportion of the cells in the evaluation matrix is also occupied (value greater than 0) in the reference matrix. The overcrowding factor shows whether more, fewer or the same number of cells are occupied in the evaluation matrix compared to the reference matrix.

The added value of the approach is derived based on the comparison of the obtained results considering the WLAN and Bluetooth data with an equilibrium matrix, that does not contain any information about the distribution of the origin-destination relations in reality.

Various filter and extrapolation sets are used to create the evaluation matrices for a total of 29 available reference trips. In total, more than 30.000 evaluation matrices were determined for the evaluation - more than 1.100 per reference trip. The evaluation of the results of the considered parameter combinations was carried out both differentiated for the individual quality measures and in the form of a comprehensive evaluation of the quality measures. In addition, the comprehensive evaluation is carried out for different classifications of the reference trips - according to the bus Line and according to passenger demand. For the evaluation of the results, a normalisation of the individual parameter combinations was used on the basis of the results for all reference trips. For this purpose, the results obtained in the quality measures under consideration were combined into scores in four characteristics, weighted according to their significance and summed up.

Based on this procedure, seven parameter combinations were determined that achieved the highest scores. These parameter combinations use both WLAN and Bluetooth data as a source. The input data for creating the output matrix are reduced here by filtering on the basis of the GPS airline distance. In a differentiated analysis of the reference trips broken down by lines and demand classes, the parameter combination of WLAN and Bluetooth data with filter set P51_GPS200 with fuzzification and extrapolation using the Fratar method (auxiliary value 1.0000, reduction threshold 0.5) proved to be particularly robust: This parameter combination is at least among the 10% of the best parameter combinations in each class considered.

Furthermore, the determined parameter combinations always perform better than the best approach without initial data, without filtering of the input data and without extrapolation of the initial matrices. These findings confirm the added value of the approach and the process steps developed. At the same time, it must be taken into account that some of the parameter combinations only comply with the GEH limits in a maximum of 13 of the 29 reference trips considered (approx. 45%). For some reference trips, the limit values could not be met by any of the parameter combinations included in the analysis. This leads to the conclusion that the approach in practical application has a high probability of errors in the allocation of passengers to the origin-destination relations above the set limit values.

The reasons for these results were investigated in detail: Overall, it has been shown that a huge number of data points are collected, many of which do not originate from passengers, but from sources of interference. Furthermore, the

information contained in the collected data packets is not suitable for clearly separating passenger data from interference data. The existing distortions cannot be compensated for solely on the basis of the extrapolation and the equalisation of the matrix elements to the marginal sums. Despite these distortions, it was possible to achieve added value by using WLAN and Bluetooth data compared to comparison methods that are only based on the marginal totals.

Moreover, the approach has been evaluated for 29 reference trips. The sample for the evaluation of the results of the approach is therefore small, so that the standard error increases. Random events thus strongly influence the result. The approach should therefore be tested on further line trips and, potentially, on other lines.

Regarding the results of the approach against the background of the intended areas of application (service planning, tariff planning, revenue distribution, calibration of traffic models), it must be considered that the requirements for spatial and temporal accuracy depend on the purpose of the application. Furthermore, the quality of previously used origin-destination data has to be taken into account. Origin-destination data have so far been derived from traffic models and are based on household and/or passenger surveys. Since a representative survey of origin-destination links would require considerable effort, extrapolations are necessary that lead to unquantified distortions and inaccuracies in the existing data. In contrast, the inaccuracies of the approach results have already been determined for 29 reference trips within the framework of the procedure evaluation using various quality measures.

With the approach described above, it is possible to collect data on a permanent basis and thus create a previously non-existent data basis. This enables public transport operators to collect information on the demand for recurring events (such as football matches, concerts, etc.) and to adjust the service accordingly in order to avoid capacity overruns or underruns. At the same time, the data can be aggregated at tariff zone level and used in the development of tariffs.

The approach developed in this dissertation can be used as a foundation to derive origin-destination data. To optimise the results, other or additional data sources can be integrated that have a higher quality than the WLAN and Bluetooth data used in this work. Such data sources include data with a unique mapping feature – for example, e-ticket data or data from an app specifically designed to capture passenger origin-destination relations. As the approach

is open to the integration of additional data sources, other data sources such as data from connection requests, bus WLAN, a city-wide WLAN or sales data can also be integrated. However, the suitability of the data as well as its usability must be methodically checked. When using multiple and additional data sources, it must be checked in particular whether disjunctive amounts of data of individual passengers or end devices are involved. When using WLAN and Bluetooth data, it must be taken into account that measures to protect the privacy of smartphone users on the part of the OEM are increasing, for example through MAC randomisation. This leads to a decreasing number of usable data.

From the author's point of view, the existing approach already has great potential for applications in which long journey segments (>30 minutes between stops) dominate. With increasing time intervals between stops, it can be assumed that the detection rate of passengers between stops increases and thus the accuracy of the assignment of boarding and alighting stops increases. Furthermore, it remains to be seen whether a temporal aggregation, i.e. the aggregation of several line trips of the same day type and the same time shift, can lead to an improvement in the quality of the process results. The improvement of the process quality through spatial aggregation - possibly in combination with temporal aggregation - must also be examined in detail.

Currently, the approach aims to consider origin-destination relations at the level of line trips. However, for service planning, as well as for revenue sharing, the consideration of the entire trip of the passengers (possibly using several lines) is relevant. Therefore, the quality of the approach for determining transfers or entire trips must continue to be tested on the basis of additional vehicle data or additional data sources.

Inhalt

Vorwort des Herausgebers	v
Danksagung	vii
Kurzfassung	viii
Abstract	xv
Inhalt	xxi
Abbildungsverzeichnis	xxiv
Tabellenverzeichnis	xxvii
Abkürzungsverzeichnis	xxx
1 Einführung	1
1.1 Problemstellung	1
1.2 Ziel der Arbeit.....	3
1.3 Einsatzbereiche von Quelle-Ziel-Daten	5
1.4 Aufbau der Arbeit	10
2 Technische und methodische Grundlagen	13
2.1 Einführung.....	13
2.2 Nachfragedatenquellen im ÖPNV	13
2.2.1 Fahrgastzählung	14
2.2.2 Fahrgastbefragung	15
2.2.3 Vertriebsdaten.....	16
2.2.4 Verbindungsauskunftsdaten	17
2.2.5 Mobilfunk-, WLAN- und Bluetooth-Daten.....	18
2.2.6 Übersicht von Vor- und Nachteilen der Nachfragedatenquellen	26
2.3 (Netzwerk-)Technische Grundlagen	28
2.3.1 ISO/OSI-Modell.....	28
2.3.2 Grundlagen WLAN.....	29
2.3.3 Grundlagen Bluetooth	43
2.4 System zur Datenerfassung.....	50
2.4.1 Kurzbeschreibung des technischen Systems der Datenerfassung.....	50
2.4.2 Datensatzbeschreibung	53
2.4.3 Privatsphäre und Datenschutz.....	57

2.5	Datenanalyse und Bewertung.....	58
2.5.1	Prüfung der WLAN-Daten.....	59
2.5.2	Prüfung der Bluetooth-Daten.....	71
2.5.3	Beurteilung der Erfassungsgenauigkeit.....	74
3	Verfahren zur Ermittlung von Quelle-Ziel-Relationen auf Basis der Input-Daten.....	78
3.1	Einführung.....	78
3.2	Filterung der Rohdaten.....	81
3.2.1	Filterung Stufe 1: Ermittlung relevanter Daten.....	82
3.2.2	Filterung Stufe 2: Identifizierung potentieller Fahrgastdaten..	83
3.2.3	Kritische Würdigung.....	87
3.3	Erstellung von Quelle-Ziel-Matrizen.....	89
3.3.1	Definition: Quelle-Ziel-Matrizen.....	89
3.3.2	Bildung von Quelle-Ziel-Matrizen auf Basis der gefilterten Daten.....	90
3.3.3	Bildung von Quelle-Ziel-Matrizen unter Berücksichtigung von Unschärfen.....	99
3.4	Beschreibung von Verfahren zum Matrixausgleich.....	118
3.4.1	Verfahren auf Basis von Ausgangsmatrizen.....	119
3.4.2	Verfahren ohne Ausgangsmatrix.....	129
3.4.3	Kritische Würdigung der Verfahren zur Hochrechnung.....	135
3.4.4	Auswahl von Verfahren zur Hochrechnung.....	138
4	Ergebnisse des Verfahrens in der praktischen Anwendung.....	141
4.1	Einführung.....	141
4.2	Auswahl Linien und Haltestellen im Untersuchungsgebiet.....	142
4.3	Erfassung von Referenzdaten.....	145
4.4	Ableitung von Gütemaßen zur Bewertung der Verfahrensqualität.....	148
4.4.1	Definition von Gütemaßen zur Bewertung der Verfahrensqualität.....	148
4.4.2	Definition von Grenzwerten der Gütemaße.....	161
4.4.3	Auswahl der Gütemaße für die Bewertung der Verfahrensergebnisse.....	164
4.4.4	Kritische Würdigung.....	165
4.5	Parametrisierung des Verfahrens für die exemplarische Anwendung.....	167

4.5.1	Berücksichtigte Parameter der Filterung	167
4.5.2	Berücksichtigte Parameter der Hochrechnung	170
4.6	Ermittlung der Verfahrensergebnisse	172
4.6.1	Prüfung der Eingangsdaten auf Vollständigkeit.....	173
4.6.2	Vorbereitung der Datenanalyse	176
4.6.3	Vorgehen zur Bewertung der Parameterkombinationen	177
4.6.4	Darstellung der Ergebnisse aus der Erprobungsphase	179
4.6.5	Bewertung der Ergebnisse	202
4.7	Prüfung von Ursachen für die Güte der Verfahrensergebnisse.....	206
5	Weiterentwicklung des Verfahrens (räumliche Aggregation)	219
5.1	Einführung.....	219
5.2	Vorgehen räumliche Aggregation	219
5.3	Ergebnisse räumliche Aggregation.....	221
5.4	Bewertung der Ergebnisse	224
6	Zusammenfassung, Fazit und weiterer Forschungsbedarf.....	227
6.1	Zusammenfassung	227
6.2	Fazit und Einsatzbereiche des Verfahrens	240
6.3	Ausblick und Forschungsbedarf	241
	Literaturverzeichnis.....	247
	Anhang	274

Abbildungsverzeichnis

Abbildung 1-1:	Überblick über den Forschungsansatz	4
Abbildung 2-1:	Kanäle im 2,4 GHz Frequenzspektrum (OFDM) (Digi-Key 2018)	32
Abbildung 2-2:	Kanalwechsellmuster gesendeter Probe Request Bursts von ausgewählten Endgeräten (Waltari und Kangasharju 2016)	38
Abbildung 2-3:	Aufbau einer MAC-Adresse (Martin et al. 2017)	39
Abbildung 2-4:	Schematische Darstellung des Inquiry-Prozess (Bluetooth SIG 2019, S. 1324)	47
Abbildung 2-5:	Übersicht des technischen Systems zur Datenerfassung (Bieland 2020)	51
Abbildung 2-6:	Einbaubeispiel der Erfassungsinfrastruktur an Haltestellen des NVV (Nordhessischer Verkehrsverbund (NVV) 2020)	51
Abbildung 2-7:	Versuchsaufbau zur Ermittlung der Signaldämpfung (Skizze)	66
Abbildung 2-8:	RSSI-Verteilung, $n = 714.994$	71
Abbildung 2-9:	Modellhafte Darstellung von erfassten Signalen im Fahrzeug während einer Linienfahrt	75
Abbildung 3-1:	Überblick zum methodischen Vorgehen (Verfahren)	78
Abbildung 3-2:	Skizze: Reduzierung der Rohdaten durch Filterstufen	82
Abbildung 3-3:	Aufbau einer Quelle-Ziel-Matrix	89
Abbildung 3-4:	Skizze zur Verdeutlichung des Bezugszeitpunkts zur Zuordnung der Einstiegshaltestelle	92
Abbildung 3-5:	Skizze zur Verdeutlichung des Bezugszeitpunkts zur Zuordnung der Ausstiegshaltestelle	92
Abbildung 3-6:	Bestimmung der Quelle-Ziel-Relation einer MAC-Adresse	93
Abbildung 3-7:	Summe der Quelle-Ziel-Relationen einer Linienfahrt über alle MAC-Adressen nach Filterung	93
Abbildung 3-8:	Modellhafte Darstellung von erfassten Signalen während einer Linienfahrt im Fahrzeug sowie an Haltestellen	95
Abbildung 3-9:	Skizze zur Präzisierung von Ein- und Ausstiegshaltestelle auf Basis von weiteren Fahrzeugdaten	96

Abbildung 3-10: Scharfe Abgrenzung in der klassischen Mengentheorie und graduelle Zugehörigkeiten in der Fuzzy-Set-Theorie (Nissen 2007).....	101
Abbildung 3-11: Beispiele geläufiger Formen von Zugehörigkeitsfunktionen (Nissen 2007)	101
Abbildung 3-12: Beispiel für eine klassische Menge (links) und eine unscharfe Menge (rechts) (Kahlert 1995)	101
Abbildung 3-13: Ermittlung möglicher Quelle-Ziel-Relationen mit Fuzzifizierung	103
Abbildung 3-14: Häufigkeitsverteilung des zeitlichen Abstands zwischen Bluetooth- (grau, n = 13.862) und WLAN-Signalen (schwarz, n = 14.761) aggregiert in Minuten-Klassen (Dataset Bieland/Mobile Data Fusion 2022)	105
Abbildung 3-15: Skizze zur Verdeutlichung des Bezugszeitpunkts zur Zuordnung von relevanten Haltestellen zu den korrekten Zeitintervallen.....	106
Abbildung 3-16: Skizze zur Verdeutlichung der Auswahl relevanter Zeitintervalle	108
Abbildung 3-17: Bestimmung der Anzahl relevanter Haltestellen (Beispiel: Einstiegshaltestellen)	110
Abbildung 3-18: Ermittlung der Zugehörigkeitswerte aller möglichen Quelle-Ziel-Relationen	114
Abbildung 3-19: Ermittlung möglicher Quelle-Ziel-Relationen unter Berücksichtigung von zusätzlichen Daten	117
Abbildung 3-20: Abhängigkeit zwischen Fehler und Modellkomplexität (Ortúzar und Willumsen 2004, S. 72).....	136
Abbildung 4-1: Überblick zum methodischen Vorgehen (Bewertung)....	142
Abbildung 4-2: GEH-Werte für Wertepaare im Bereich von 0 bis 30	156
Abbildung 4-3: Beispiel zur Berechnung des Anteils der Treffer.....	160
Abbildung 4-4: Beispiel Quelle-Ziel-Matrix mit vollständiger Erfassung (Fahrt der Linie 100, 22.09.2020, 15:33 Uhr - WLAN- und Bluetooth-Daten, keine Filterung) (Dataset Bieland/Mobile Data Fusion 2022)	175
Abbildung 4-5: Beispiel Quelle-Ziel-Matrix mit unvollständiger Erfassung (Fahrt der Linie 100, 11.10.2020, 15:35 Uhr - WLAN- und Bluetooth-Daten, keine Filterung) (Dataset Bieland/Mobile Data Fusion 2022)	175

Abbildung 4-6:	Vorgehen zur Ermittlung der optimalen Parameterkombination für das entwickelte Verfahren ...	178
Abbildung 4-7:	Bewertung des Übereinstimmungsmerkmals unter Berücksichtigung verschiedener Ergebnisse des Treffermaß und Überfüllungsfaktors in Prozent (Schrittweite: 0,1)	195
Abbildung 4-8:	Gegenüberstellung von Fahrgastzahl und Signalrate (Anzahl WLAN- und Bluetooth-Signale pro Fahrtminute) auf ausgewählten Linienfahrten der Linien 54, 100 und 500	209
Abbildung 4-9:	Entwicklung des Anteils randomisierter MAC-Adressen an allen erfassten WLAN-Signalen im Zeitraum vom 21.11.2019 bis 04.11.2021	217

Tabellenverzeichnis

Tabelle 2-1:	Übersicht der Vor- und Nachteile der behandelten Nachfragedatenquellen	26
Tabelle 2-2:	Schichten des OSI-Referenzmodells (Baun 2018; der-wirtschaftsingenieur.de 2010; Heinemann Verlag GmbH; AIRNET 2015; Badertscher et al. 2006)	29
Tabelle 2-3:	Reichweite und Leistung der Bluetooth-Klassen (Axel Springer SE 2021).....	44
Tabelle 2-4:	Übertragene Daten für die Verfahrensentwicklung aus Probe Request Frames	54
Tabelle 2-5:	Übertragene Daten für die Verfahrensentwicklung aus Inquiry Nachrichten	55
Tabelle 2-6:	Übertragene Daten für die Verfahrensentwicklung aus dem AFZS-System.....	57
Tabelle 2-7:	Smartphone-Testflotte.....	60
Tabelle 2-8:	Übersicht zur Umrechnung von Hexadezimalwerten in Binärwerte	62
Tabelle 2-9:	Ergebnisübersicht: Sendeeigenschaften der Smartphone-Testgeräte	63
Tabelle 2-10:	Messwerte der Signalstärke [dBm]: Xiaomi Redmi Note 4	67
Tabelle 2-11:	WLAN-Signale: Häufigkeitsverteilung der Anzahl erfasster Signale	68
Tabelle 2-12:	WLAN-Signale: mittlerer Zeitabstand zwischen erfassten Signalen	69
Tabelle 2-13:	Bluetooth-Signale: Häufigkeitsverteilung der Anzahl erfasster Signale	72
Tabelle 2-14:	Bluetooth-Signale: mittlerer Zeitabstand zwischen erfassten Signalen	73
Tabelle 3-1:	Übersicht der Verfahrensschritte sowie zugehöriger „Tools“	80
Tabelle 3-2:	Kriterien zur Identifikation von Fahrgastdaten	85
Tabelle 3-3:	Häufigkeitsverteilung der GPS- Luftlinienentfernung je MAC-Adresse und Linienfahrt	86
Tabelle 3-4:	Häufigkeitsverteilung des mittleren RSSI der WLAN-Signale je Linienfahrt.....	86

Tabelle 3-5:	Häufigkeitsverteilung der Anzahl an Linienfahrten je MAC-Adresse	87
Tabelle 3-6:	False Positives, False Negatives bei der Identifizierung von Fahrgastdaten	88
Tabelle 4-1:	Übersicht der ausgestatteten Haltestellen mit Datenerfassung (init innovation in traffic systems SE 2022)	145
Tabelle 4-2:	Klasseneinteilung der Referenzfahrten nach Verwertbarkeit	147
Tabelle 4-3:	Übersicht zu den uneingeschränkt verwertbaren Fahrgastbefragungen.....	147
Tabelle 4-4:	Einteilung und Beschreibung der Entfernungsklassen ..	150
Tabelle 4-5:	Beispiel zur Relativierung der WFQS	152
Tabelle 4-6:	Beispiele zu absolutem und relativem Fehler	154
Tabelle 4-7:	Absolute und relative Abweichung in Abhängigkeit des GEH-Grenzwerts (gerundet, Fahrgastzahlen je Relation)	162
Tabelle 4-8:	Definition von Grenzwerten des GEH-Werts zur Bewertung auf Basis der Personenfahrten	163
Tabelle 4-9:	Filterindikatoren und Parameterausprägungen für die Verfahrenskalibrierung	168
Tabelle 4-10:	Zuordnung zwischen Filterset und Parameterausprägungen	169
Tabelle 4-11:	Differenzierungsmerkmale und Ausprägungen für die Kalibrierung des Verfahrens	170
Tabelle 4-12:	Parametrisierung Fratar-Verfahren	171
Tabelle 4-13:	Parametrisierung Li-Cassidy-Verfahren	171
Tabelle 4-14:	Referenzfahrten mit Datenausfällen	176
Tabelle 4-15:	Übersicht der Anzahl von Grenzwert-Einhaltungen bei allen untersuchten Parameterkombinationen	180
Tabelle 4-16:	Ergebnisse der Vergleichsverfahren (Li-Cassidy, Fratar mit Filterset 99 (FullFilter)) unter Hervorhebung (graue Markierung) der besten gütemaßspezifischen Ergebnisse	182
Tabelle 4-17:	Parameterkombinationen mit maximaler Einhaltung des Gütemaßes GEH 1,5.....	184
Tabelle 4-18:	Parameterkombinationen mit maximaler Einhaltung des Gütemaßes GEH 0,7.....	185

Tabelle 4-19:	Gegenüberstellung der Ergebnisse je Parameterkombination zur Einhaltung der GEH-Grenzwerte (1,5 und 0,7)	186
Tabelle 4-20:	Parameterkombinationen mit maximaler Einhaltung des Gütemaßes Abweichung der Verkehrsleistung (ohne Li-Cassidy)	188
Tabelle 4-21:	Parameterkombinationen mit maximaler Einhaltung des Gütemaßes Abweichung der Verkehrsleistung (Abweichungen der Verkehrsleistung in den Entfernungsklassen)	189
Tabelle 4-22:	Parameterkombinationen mit maximaler Einhaltung des Gütemaßes Treffermaß (Anteil Treffer)	190
Tabelle 4-23:	Vorgehen zur Normierung und Gewichtung der parameterkombinationsspezifischen Ergebnisse	194
Tabelle 4-24:	Beste Parameterkombinationen auf Basis der Normierung und Gewichtung nach Tabelle 4-23	197
Tabelle 4-25:	Beste Parameterkombinationen mit originären Gütemaßen	197
Tabelle 4-26:	Übersicht (Ausschnitt) der besten Parameterkombinationen auf Basis der Normierung und Gewichtung (Tabelle 4-23) nach betrachteter Linie	199
Tabelle 4-27:	Übersicht (Ausschnitt) der besten Parameterkombinationen auf Basis der Normierung und Gewichtung (Tabelle 4-23) nach Nachfrageklassen	201
Tabelle 4-28:	Ergebnisse der Anzahl und des Anteils übereinstimmender Zellen zwischen Ausgangs- und Referenzmatrix (Dataset Bieland/Mobile Data Fusion 2022)	214
Tabelle 5-1:	Ausgewählte Linienfahrten für die räumliche Aggregation (Linie 100, Linie 500)	220
Tabelle 5-2:	Ergebnisübersicht ausgewählter Referenzfahrten zur räumlichen Aggregation (ohne Aggregation)	223
Tabelle 5-3:	Ergebnisübersicht ausgewählter Referenzfahrten zur räumlichen Aggregation (mit Aggregation)	223

Abkürzungsverzeichnis

°C	Grad Celsius
2D	zweidimensional
3D	dreidimensional
A	Ampere
AFZS	Automatische Fahrgastzählsysteme
AG	Aktiengesellschaft
AP	Access Point
App	Application / Anwendungssoftware
ASCII	American Standard Code for Information Interchange
BDSG	Bundesdatenschutzgesetz
BIBO	Be-In/Be-Out
Bio.	Millionen
BLE	Bluetooth Low Energy (ab Bluetooth Version 4.0)
BMVI	Bundesministerium für Verkehr und digitale Infrastruktur
BT	Bluetooth
bzw.	beziehungsweise
CAC	Channel Access Code
CAPI	Computer Assisted Personal Interview
CICO	Check-In/Check-Out
CID	Company-ID
CLK	Clockcounter, Systemzeit
CO ₂	Kohlenstoffdioxid
d. h.	das heißt
DAC	Device Access Code
dbm	Dezibel Milliwatt
DC	Direct Current (Gleichstrom)
DIAC	Dedicated Inquiry Access Code
DSGVO	Datenschutz-Grundverordnung der Europäischen Union
dt	delta time - zeitlicher Abstand
e. V.	eingetragener Verein
EDR	Enhanced Data Rate
EFM	Elektronisches Fahrgeldmanagement
EOTD	Enhanced Observed Time Difference
etc	et cetera
E-Tickets	Elektronische Fahrausweise
FG VPVS	Fachgebiet Verkehrsplanung und Verkehrssysteme der Universität Kassel
FGSV	Forschungsgesellschaft für Straßen- und Verkehrswesen
FHS	Frequency Hopping Synchronisation
FHSS	Frequency Hopping Spread Spectrum
FIS	Forschungs-Informations-System
FPD	Floating Phone Data
FQS	Fehlerquadratsumme
FTP	File Transfer Protocol
GB	Gigabyte

GEH-Wert	Gütemaß zur Bewertung von Modellschätzungen
ggf.	gegebenenfalls
GHz	Gigahertz
GIAC	General Inquiry Access Code
GmbH	Gesellschaft mit beschränkter Haftung
GPS	Global Positioning System
GSM	Global System for Mobile Communications
h	Stunde
HDSIG	Hessisches Datenschutz- und Informationsfreiheitsgesetz
HH:MM:SS	Angabe Zeitformat in Stunden (HH), Minuten (MM), Sekunden (SS)
HR-Sets	Hochrechnungsset (im Rahmen des Verfahrens)
HT	High Througput
HTML	Hypertext Markup Language
HTTP	HyperText Transfer Protocol
HVZ	Hauptverkehrszeit
i. d. R.	in der Regel
i. W.	im Wesentlichen
IAC	Inquiry Access Code
IAO	Institut für Arbeitswirtschaft und Organisation
ID	Identifikator
IE	Information Elements
IEEE	Institute of Electrical and Electronic Engineers
INIT SE	Init Innovation in Traffic Systems SE (Societas Europaea)
iOS	Internetwork Operating System - Betriebssystem des Unternehmens Apple
IoT	Internet of Things
IP	Internet Protocol (Standard: Version 4)
IPF	Iterative Proportional Fitting
IPv6	Internet Protocol Version 6
iris	Infrared & Intelligent Sensors
IRMA	InfraRed Motion Analyser
ISM-Band	Industrial, Scientific and Medical Band
ISO	International Standards Organisation
kBit	Kilobit
km	Kilometer
kum.	kumuliert
LBF	Linienbeförderungsfälle
Lfd. Nr.	Laufende Nummer
LMSC	Local Network Standards Committee
LSA	Lichtsignalanlage
m	Meter
MAC	Media Access Control
MAN	Metropolitan Area Network
MBit	Megabit
MDF	Bezeichnung Datensatz im Forschungsprojekt "Mobile Data Fusion"
MDS	Modulation and Coding Scheme
MHz	Megahertz

Mio.	Millionen
ms	Millisekunden
MVV	Münchener Verkehrs- und Tarifverbund
N. V.	Niederländische Aktiengesellschaft (naamloze vennootschap)
NFC	Near-Field-Communication
nHS	Anzahl relevanter Haltestellen
NO _x	Stickstoffoxide
Nr.	Nummer
NS	Nederlandse Spoorwegen
NVV	Nordhessischer Verkehrsverbund
NVZ	Nebenverkehrszeiten
o. ä.	oder ähnliches
o. g.	oben genannt
o.J.	ohne Jahr
OD-Matrix	Origin-Destination-Matrix
OEM	Original Equipment Manufacturer
ÖPNV	Öffentlicher Personennahverkehr
OS	Operating System
OSI	Open Systems Interconnection
OUA	Organisationally Unique Address
OUI	Organisationally Unique Identifier
ÖV	Öffentlicher Verkehr
PAPI	Paper and Pencil Interviewing
PC	Personal Computer
PDU	Protocol Data Unit
Pkm	Personenkilometer
Pkw	Personenkraftwagen
PNL	Preferred Network List
POI	Point of Interest
PR	Probe Requests
QZ-Matrix	Quelle-Ziel-Matrix
RFID	Radio-frequency Identification
RSSI	Received Signal Strength Indicator
SD	Secure Digital Memory Card
SIG	Special Interest Group
SMTP	Simple Mail Transfer Protocol
SN	Sequence Number
SSID	Service Set Identifier
SVZ	Schwachverkehrszeiten
TCP	Transmission Control Protocol
TfL	Transport for London
TSF	Timing Synchronization Function
u. a.	unter anderem
U-Bahn	U-Bahn
UDP	User Datagram Protocol
UTC	koordinierte Weltzeit

V	Volt
v. a.	vor allem
VBF	Verbundbeförderungsfälle
VDV	Verband Deutscher Verkehrsunternehmen e.V.
VDV-KA	VDV-Kernapplikation
WFQS	Wurzel der Fehlerquadratsumme
WGS-84	World Geodetic System 1984 (geodätisches Referenzsystem)
Wifi	Kunstbegriff, bezeichnet u. a. WLAN-Netze nach IEEE 802.11 Standard
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network
xMND	Extended Mobile Network Data
z. B.	zum Beispiel
z. T.	zum Teil
ZI	Zeitintervalle

1 Einführung

1.1 Problemstellung

Der Öffentliche Personennahverkehrs (ÖPNV) in Deutschland steht vor großen Herausforderungen. Demografische und sozioökonomische Entwicklungen sowie Veränderungen im Mobilitätsverhalten führen zu weniger Zwangskunden im ÖPNV, aber zu einem steigenden Kundenpotential, das sich situationsabhängig für ein bestimmtes Verkehrsmittel entscheidet (multimodales Verhalten), wie auch die zunehmende Bedeutung und Anzahl an Smartphone-Apps für die multimodale Wegeplanung zeigt (Mobilitäts-Apps) (PTV Planung Transport Verkehr AG 2018; ifmo - Institut für Mobilitätsforschung 2011; Kuhnimhof 2012; Nobis 2013; VCD Verkehrsclub Deutschland e.V. 2021). In einigen Städten und Ballungsräumen wuchs die Nachfrage nach Beförderungsleistungen im ÖPNV bis zum Frühjahr 2020 stärker als das Angebot (Sommer und Bieland 2018). Im Zuge der Covid-19-Pandemie brach die ÖPNV-Nachfrage im Frühjahr um bis zu 90% ein (Verband Deutscher Verkehrsunternehmen e. V. (VDV) 2020). Die Wirkungen der Pandemie auf das Mobilitätsverhalten sind Untersuchungen zu Folge eine verstärkte Nutzung von Individualverkehrsmitteln durch wahlfreie Personen, bei gleichzeitiger Verringerung der Anzahl durchgeführter Wege aufgrund von Homeoffice oder der generellen Vermeidung von Kontakten (Nobis und Eisenmann 2020; Zehl et al. 2020; Rothengatter et al. 2021; Bucsky 2020; Schrader 2020). Darüber hinaus werden im Zuge der Verkehrswende tarifliche Maßnahmen wie Nulltarif oder Bürgertickets diskutiert, die zu starken Fahrpreissenkungen für die Kunden des öffentlichen Verkehrs (ÖV) führen und somit das Potential zu einer hohen Steigerung der Fahrgastnachfrage besitzen (Bieland und Sommer 2021).

Bei einem zunehmenden Potential von wahlfreien und multimodalen Kunden werden Planung und Marketing des ÖPNV noch wichtiger als sie es ohnehin bereits sind. Eine zielgerichtete und kundenorientierte Planung des ÖPNV erfordert jedoch genaue und aktuelle Daten zum Mobilitätsverhalten und zur Fahrgastnachfrage. Insbesondere die Angebotsplanung braucht daher eine verbesserte Datengrundlage, um der rasch verändernden Entwicklung der Nachfrage künftig gerecht werden zu können.

Dabei ist zu berücksichtigen, dass die Datengrundlage zur Modellierung von Quelle-Ziel-Verflechtungen in einem ÖPNV-Netz sowie zu den Umsteigevorgängen an einzelnen Haltestellen heutzutage in der Regel über

personalintensive und damit aufwändige Fahrgasterhebungen erfasst werden. Diese Erhebungen werden aus Kostengründen nur als Querschnittserhebung für einen Stichtag im Abstand mehrerer Jahre punktuell oder flächendeckend für ein Bedienungs- bzw. Verbundgebiet durchgeführt. Sommer (2002) hat in seiner Dissertation den notwendigen Stichprobenanteil für die Abbildung der Verkehrsnachfrage unter Verwendung einer einfachen Zufallsauswahl mit einem relativen Fehler von 20% für die Stadt Braunschweig mit 81,6% angegeben. Dies entsprach zum Bezugszeitpunkt über 200.000 Personen. Aufgrund der finanziellen Restriktionen in der Planungspraxis sind Auftraggeber in der Regel nicht bereit, den finanziellen Aufwand in dieser Größenordnung für lediglich temporär brauchbare Daten aufzubringen. Entsprechend ist davon auszugehen, dass bei derzeit vorliegenden Quelle-Ziel-Relationen auf Basis von (Haushalts-)Befragungen unbekannte Verzerrungen vorliegen.

Weitergehende Befragungen, in denen das intrapersonelle Nutzungsverhalten über einen längeren Zeitraum von Wochen oder Monaten erfasst wird, sind unüblich und aus Gründen der Belastung für die Befragten praktisch nicht möglich. Die mangelnde Aktualität der erfassten Daten führt dazu, dass diese im täglichen Betrieb nur eine geringe Rolle spielen. Aktuelle Daten zum Mengengerüst der Fahrgastnachfrage (Ein- und Aussteiger) werden bereits bei zahlreichen Verkehrsunternehmen mit Hilfe von automatischen Fahrgastzählssystemen (AFZS) erfasst.

Gleichzeitig steigt der Smartphone-Besitz in Deutschland seit Jahren an. Vor dem Hintergrund, dass die meisten Smartphones regelmäßig Signale ausstrahlen, um WLAN-Netze (Wireless Local Area Network) oder Bluetooth-Geräte zu finden, fallen viele Daten ohne Zutun des Nutzers an, die für ÖPNV-Planung und Marketing verwendet werden können. Das ständige Mitführen eines Smartphones ermöglicht es dem Nutzer, mobil und spontan Verbindungen aus der Fahrplanauskunft der ÖPNV-Betreiber (Verkehrsunternehmen, Verkehrsverbände, Aufgabenträger) abzufragen. Die bei der Abfrage erzeugten Daten können ebenfalls als Datenquelle genutzt werden. Darüber hinaus bieten die ÖPNV-Betreiber zunehmend die WLAN-Nutzung in Fahrzeugen als kostenlosen Service an (Hutter 2016; Sebayang 2019; KVG Kieler Verkehrsgesellschaft mbH; Niedersächsisches Ministerium für Wirtschaft, Arbeit, Verkehr und Digitalisierung 2020). Für die Nutzung dieses Dienstes loggt sich der ÖPNV-Kunde ein, sodass auch hier Daten erzeugt werden, die bislang in der Regel lediglich zur Abrechnung mit dem Provider genutzt werden.

Sofern Kunden mit einem Smartphone den ÖPNV nutzen, fallen bei eingeschaltetem WLAN und/oder Bluetooth in den in Reichweite gelegenen WLAN- und Bluetooth-Routern entlang des realisierten Weges Registrierungs- und ggf. Nutzungsdaten an. Diese Daten sollen verwendet werden, um in der Verknüpfung mit anderen Datenquellen (AFZS) ein Gesamtbild der Verkehrsnachfrage im ÖPNV-Netz aufzubauen. Diese neue und stets aktuelle Datengrundlage bietet einen erheblichen Mehrwert insbesondere in den Bereichen der (Angebots-)Planung, Einnahmenaufteilung sowie Tarifkonzeption.

1.2 Ziel der Arbeit

In dieser Dissertation wird ein Verfahren beschrieben, das basierend auf

- AFZS-Daten, die in vielen Verkehrsunternehmen und -verbänden vorliegen sowie
- automatisch erfassbaren Daten unterschiedlicher Quellen

die Fahrgastnachfrage in Form von linienspezifischen Quelle-Ziel-Matrizen automatisch für lange Erhebungszeiträume ermittelt. Dazu sollen neben AFZS-Daten auch Daten genutzt werden, die durch die Suche von WLAN-Netzen oder Bluetooth-Geräten ohne Zutun des Smartphone-Nutzers entstehen (Probe Requests [WLAN], Inquiry-Response-Nachrichten [Bluetooth]) (passive Datenerzeugung). Diese Daten registrieren die Anwesenheit der Endgeräte – und damit auch die der Fahrgäste – zu verschiedenen Zeitpunkten auf einem Abschnitt bzw. der gesamten Linienfahrt.

Das Forschungsvorhaben führt also Daten aus verschiedenen, zum Teil noch nicht durch den ÖPNV erschlossenen Quellen zusammen und entwickelt darauf aufbauend die methodischen Grundlagen zur Nutzbarmachung ebendieser Daten für den praktischen Einsatz. Vorrangig geht es darum, den ÖPNV-Betreibern genaue und aktuelle Informationen über die Nachfrage in ihrem Verkehrsnetz bereit zu stellen. Durch die Nutzung der WLAN- und Bluetooth-Daten sollen zusätzlich zum bekannten Mengengerüst der Fahrgastnachfrage aus den AFZS-Daten die Quelle-Ziel-Verflechtungen von Linienfahrten abgebildet werden. Dies bietet eine kostengünstige Alternative zur bisher üblichen Erfassung dieser Merkmale im Rahmen aufwändiger Fahrgastbefragungen. Bestehende Verfahren zur Fahrgasterhebung werden somit auf eine neue methodisch-technologische Basis gestellt. Dabei werden die datenschutzrechtlichen Belange bereits bei der Erfassung und Aufbereitung der Daten berücksichtigt. Abbildung 1-1 gibt einen Überblick des Forschungsansatzes im Hinblick auf die angestrebten Datenquellen sowie den Status Quo.

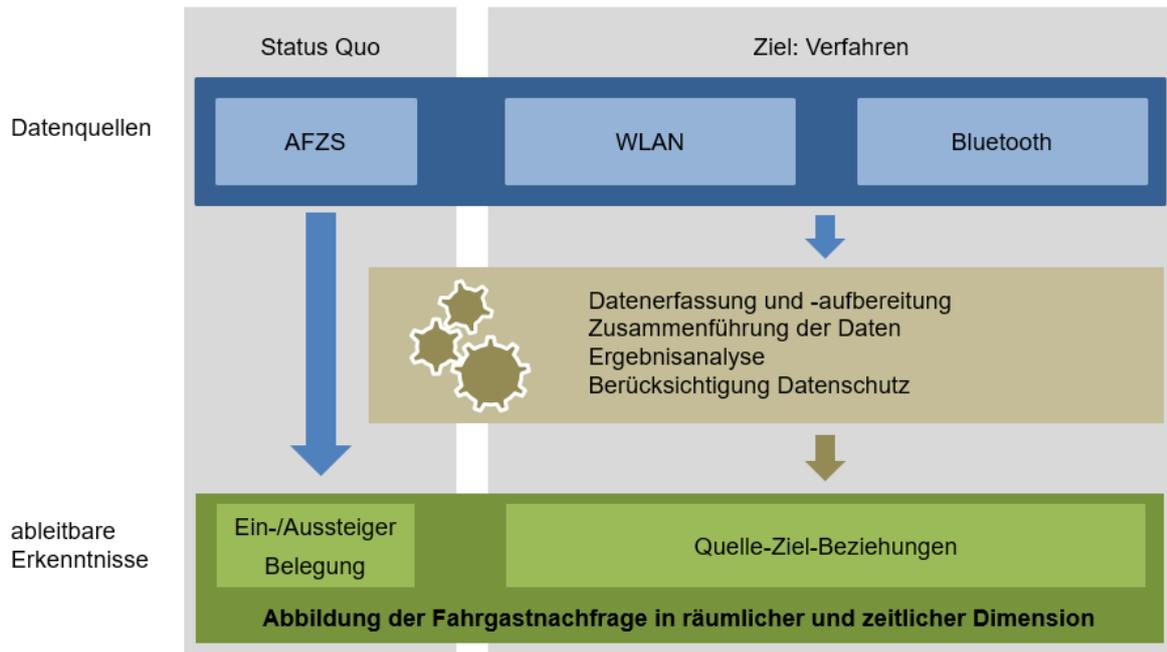


Abbildung 1-1: Überblick über den Forschungsansatz

Ein Verfahren, welches monatlich oder sogar tagesaktuell die Quelle-Ziel-Verflechtungen abbilden kann, stellt eine Verbesserung der Datenverfügbarkeit dar: Quelle-Ziel-Verflechtungen könnten zukünftig auf Grund der umfassenden Datengrundlagen nach Wochentagen und Tageszeiten differenziert ausgewertet und in den Planungs- und Betriebsprozess (u. a. Angebotsplanung, Betriebsplanung, Verkehrsmanagement) eingebracht werden. Der Erkenntnisgewinn kann in der Planung und im Marketing des ÖPNV genutzt werden. Mit einer daraus resultierenden stärkeren Nutzung des ÖPNV durch bestehende sowie neue Kunden kann die Nachhaltigkeit des (ÖPNV-)Verkehrssystems gesteigert werden. Das zu entwickelnde Verfahren bietet ÖPNV-Betreibern die Möglichkeit, ihre Planungen anhand der zusätzlichen Nachfragedaten auszurichten und noch nicht genutzte Potentiale bei der Kundenbindung und -gewinnung durch ein attraktives, kundenorientiertes Angebot auszuschöpfen. Falls durch das neue Verfahren Einsparungen bei den Erhebungskosten möglich werden und gleichzeitig steigende Erlöse durch neue Kunden generiert werden können, steigt der Kostendeckungsgrad. Werden die eingesparten Kosten für eine Attraktivitätssteigerung des Angebots eingesetzt, verbessern sich die Erreichbarkeit zentraler Einrichtungen und die Mobilitätsoptionen der Menschen. Insbesondere Pkw-Fahrten der wahlfreien und multimodalen Verkehrsteilnehmer sollen durch ein attraktiveres Angebot auf den ÖV verlagert werden, mit dem Ziel, die verkehrsbedingten Umweltbelastungen zu reduzieren (u. a. NO_x -, CO_2 -Emissionen).

Da die Ausrüstung von Fahrzeugen mit der für die Erfassung notwendigen WLAN- und Bluetooth-Infrastruktur vergleichsweise kostengünstig ist, können Verkehrsunternehmen die Daten eigenständig erfassen, sodass sie sich weder in eine Abhängigkeit von anderen Datenzulieferern (etwa Netzbetreiber für Mobilfunkdaten) begeben, noch eine andauernde finanzielle Belastung für den (wiederkehrenden) Kauf von Daten entsteht. Darüber hinaus ist zu berücksichtigen, dass Verkehrsunternehmen und -verbände ihre Fahrzeuge (Busse, Tram) zunehmend mit WLAN ausstatten und somit (unbeabsichtigt) bereits einen Teil der notwendigen Infrastruktur beschaffen.

1.3 Einsatzbereiche von Quelle-Ziel-Daten

Quelle-Ziel-Matrizen (engl.: Origin-Destination-Matrix) sind in der Verkehrsplanung von hoher Bedeutung, da sie neben dem reinen Mengengerüst der Nachfrage auch die Quelle und das Ziel der Fahrgastfahrten auf Relationsebene darstellen. Das Verfahren zielt darauf ab, die Datengrundlage zu Quelle-Ziel-Verflechtungen zu verbessern, sodass im Idealfall Auswertungen nach Wochentagen und Tageszeiten differenziert ermöglicht werden. Der Informationsgehalt der bislang genutzten AFZS-Daten wird somit durch Informationen über die Quelle und das Ziel der ÖV-Wege der Fahrgäste ergänzt. Aufbauend auf diesen zusätzlichen Informationen soll das Verfahren insbesondere für die Angebotsplanung, aber auch für weitere Anwendungsbereiche – etwa Tarifgestaltung, Kundenlenkung oder Maßnahmenevaluation – einen Mehrwert leisten.

Grundsätzliche Anforderungen an die Verfahrensergebnisse sind dabei zunächst, dass diese die Bearbeitung des Anwendungsgebiets mit ausreichender bzw. gegenüber dem Status Quo höherer Qualität ermöglichen. Die dargestellten Anwendungsgebiete haben unterschiedliche Anforderungen an die Genauigkeit der Daten. Unter „Genauigkeit“ werden an dieser Stelle lediglich die räumliche Auflösung und zeitliche Aggregation der Verfahrensergebnisse zusammengefasst. Die räumliche Auflösung bezeichnet den Detailgrad der Darstellung von Ein- und Ausstiegsort (Quelle und Ziel der Fahrt). Dies kann sich auf den gesamten Weg (Verbundbeförderungsfall, VBF) oder auf einzelne Wegetappen (Linienbeförderungsfall, LBF) beziehen. Die räumliche Auflösung kann zwischen einer haltstellenscharfen (höchste Auflösung) bis hin zu einer groben Auflösung, beispielsweise auf Ebene von Tarifzonen, liegen. Die zeitliche Aggregation variiert dabei zwischen einer linienfahrtscharfen Abbildung der Quelle-Ziel-Relationen (höchste Auflösung) bis hin zur Aggregation auf Tages-, Monats- oder Jahresebene, differenziert nach

Tagestypen und Zeitschichten. Anforderungen an den Datenschutz werden gesondert betrachtet (Kapitel 2.4.3).

Im Folgenden werden die genannten Anwendungsgebiete sowie die Anforderungen an die räumliche und zeitliche Auflösung qualitativ beschrieben.

Angebotsplanung: Seitens der ÖPNV-Aufgabenträger sind aktuelle und zuverlässige Nachfragedaten zur systematischen Planung des Verkehrsangebots im Bedienungsgebiet notwendig, um die Vorgaben aus strategischen Planungen (z. B. Nahverkehrsplan) zu erreichen. Diese Daten stammen derzeit vorwiegend aus Nachfragemodellen (modellierte Matrizen), da die Erhebung realer Quelle-Ziel-Verflechtungen aufgrund der notwendigen Stichprobengröße einen erheblichen Aufwand bedeuten (Sommer 2002). Die reale Zielwahl aller Verkehrsteilnehmer ist im Planungsprozess unbekannt. Daher wird in der Regel auf Ergebnisse von Verkehrsnachfragemodellen zurückgegriffen, die wiederum empirische Reiseweitenverteilungen zur Kalibrierung und Validierung nutzen (Sommer 2022).

Die Angebotsplanung umfasst die Streckennetzplanung bestehend aus Haltestellen und Fahrwegen, die Linienplanung, die Einsatzplanung sowie die Kapazitätsplanung. Sie ist eng mit der Fahrplanung verknüpft. Detaillierte und stets aktuelle Daten zu den Quelle-Ziel-Verflechtungen im Untersuchungsraum liefern eine verbesserte Datengrundlage, die als Eingangsgröße für Verkehrsmodelle genutzt werden kann und somit zur Optimierung der Angebotsplanung beiträgt. Diese Daten bieten beispielsweise einen Mehrwert bei der Einrichtung einer Expressbuslinie, die nur ausgewählte Haltestellen bedient, bei der Neuordnung der Linienwege oder auch bei der Angebotsplanung bei Großveranstaltungen. Die automatische und dauerhafte Erfassung ermöglicht es darüber hinaus, zukünftig jahreszeitliche Schwankungen der Nachfrage differenziert nach ÖV-Linien bei der Planung zu berücksichtigen. Insbesondere für die Streckennetz- und Linienplanung sind Quelle-Ziel-Matrizen auf Verbundebene (inkl. Umstiege) notwendig.

Bei Anwendung von ÖPNV-Nachfragedaten für die Angebotsplanung bleibt zu berücksichtigen, dass die Daten lediglich die Nachfrage der Nutzer des ÖPNV abbilden. Die Nachfrage von Nicht-Nutzern bleibt unbekannt und kann entsprechend nicht in die Angebotsplanung eingehen.

Anforderungen: Für eine systematische Angebotsplanung (Streckennetz-, Linienplanung, Einsatzplanung, Kapazitätsplanung) sind Quelle-Ziel-Beziehungen auf Verbundfahrtenebene in hoher räumlicher und zeitliche Auflösung

ideal. Wie beschrieben, liegen diese Daten in der Regel nicht vor, sodass hier räumlich und/oder zeitlich aggregierte Daten zur realen Nachfrage bereits zu einer Verbesserung der Datengrundlage führen. Für die Streckennetz- und Linienplanung können beispielsweise aggregierte Zeiträume – etwa eine Differenzierung zwischen Schulzeit und Schulferien und nach Tagestypen (z. B. Normalwerktag, Samstag, Sonn- und Feiertag) – ausreichen (zeitliche Aggregation). Für die Planung der Bedienungshäufigkeit und der Kapazitäten reichen dagegen die häufig bereits vorliegenden AFZS-Daten aus.

Tarifgestaltung: Im Rahmen der Tarifgestaltung im ÖPNV werden die Preise zwischen Start- und Zielhaltestelle der Fahrgäste festgelegt (Sommer und Sauer 2016). Bei der Tarifgestaltung kann dabei u. a. räumlich, zeitlich oder zielgruppenspezifisch (z. B. Seltennutzer, Gelegenheitskunde, Stammkunde) differenziert werden. Unter Verwendung der zusätzlichen Datengrundlage wird ermöglicht, die Nachfrage auf verschiedenen Quelle-Ziel-Beziehungen im Verbundraum differenziert nach Wochentag und Uhrzeit auszuwerten. Auf Basis der Nachfragedaten (Quelle-Ziel-Matrizen) stehen beispielsweise umfangreiche Daten für die Entwicklung von entfernungsbasierten, leistungsgerechten Tarifen zur Verfügung (Luftlinie, tatsächliche Fahrdistanz). In Verbindung mit Daten zu den genutzten Fahrausweisen sind darüber hinaus Prognosen zur Erlösentwicklung infolge der Tarifanpassung möglich. Angaben zu den Fahrausweisen (Bartarif, Zeitkarten) können nicht durch das vorgestellte Verfahren abgeleitet werden.

Gleichzeitig können Tarife mit dem Ziel der Kundenlenkung entwickelt werden. Unter Kundenlenkung werden dabei Maßnahmen bezeichnet, durch die die Kunden dazu angeregt werden, ihre Wege außerhalb der Spitzenstunde durchzuführen, sodass die Nachfrage auf schwächer ausgelastete Linien oder Zeitschichten verteilt wird. Ziel der Kundenlenkung ist es, die Kapazität der bestehenden Infrastruktur effizienter zu nutzen. Maßnahmen der Kundenlenkung umfassen beispielsweise Anreize in Form von Preisnachlässen bei Fahrgastfahrten außerhalb der Spitzenstunde, um Nachfragespitzen abzuschwächen.

Anforderungen: Die Anforderungen variieren auch hier mit dem spezifischen Anwendungsfall bzw. dem Ziel der Tarifgestaltung. Bei der Konzeption entfernungsbasierter Tarife etwa ist eine haltstellenscharfe Erfassung der Quelle-Ziel-Beziehungen zielführend, da nur so der die Reiseweiten der durchgeführten Fahrgastfahrten ermittelbar ist. Soll dem Tarif die tatsächliche Weglänge zu Grunde gelegt werden, ist auch die Berücksichtigung der

Fahrtroute (Informationen zur genutzten ÖV-Linie durch die Erfassung oder auch per Routing) notwendig. Sofern der Tarif nach Tageszeit differenzierte Kosten (z. B. höhere Kosten während der Hauptverkehrszeit) einbeziehen soll, ist eine entsprechende zeitliche Aggregation zielführend. Werden Tarifreformen (ohne Berücksichtigung entfernungsbasierter Tarife) geplant, sind dagegen räumliche Aggregationen auf Ebene von Tarifzonen ausreichend. Dabei ist zu berücksichtigen, dass es in den Übergangszonen zu Unschärfen kommt.

Einnahmenaufteilung: Im Verbundraum eines Verkehrsverbundes agieren verschiedene Verkehrsunternehmen, die wechselseitig die von den Verbundpartnern ausgegebenen Fahrausweise akzeptieren, um den Fahrgästen eine einfach verständliche Tarifstruktur anbieten zu können. Die Zuordnung der Erlöse an die beteiligten Verkehrsunternehmen obliegt dem Verkehrsverbund und benötigt eine belastbare Datengrundlage. Aufgrund der Heterogenität der durch die verschiedenen Verkehrsverbünde genutzten Einnahmenaufteilungsverfahren (etwa nachfrageorientierte oder angebotsorientierte Verfahren) ist im Einzelfall zu prüfen, in welchem Umfang die anvisierten Verfahrensergebnisse dazu geeignet sind, bisherige Fahrgasterhebungen zu ergänzen oder gar zu ersetzen. Durch eine Erhebung der Fahrgastströme in unterjährigen Zeiträumen kann eine leistungsgerechte Einnahmenaufteilung auch bei Veränderungen des Liniennetzes im Laufe eines Kalenderjahres, etwa bei der Eröffnung neuer Streckenabschnitte, gewährleistet werden. Als Datengrundlage für die Einnahmenaufteilung werden i. d. R. neben Nachfragedaten je Linie und Umsteigeinformationen auch Angaben zu den genutzten Fahrausweisen benötigt. Angaben zu den Fahrausweisen (Bartarif, Zeitkarten) können nicht durch das vorgestellte Verfahren abgeleitet werden. Darüber hinaus ist der Mehrwert gegenüber AFZS-Systemen (Investitionskosten, Datengüte) zu prüfen.

Anforderungen: Die Anforderungen sind in Abhängigkeit des vereinbarten Einnahmenaufteilungsverfahrens im Verbundraum zu wählen. Die räumliche Auflösung kann zwischen einer haltestellenscharfen Zuordnung der Verkehrsleistung auf einzelnen Streckenabschnitten (in Personenkilometern) bis hin zu einer Zuordnung basierend auf Gebieten von Verkehrsunternehmen oder Aufgabenträgern reichen. Zeitlich können die Daten auf ein Jahr aggregiert werden, da die Einnahmen (bislang) auf Jahresebene verteilt werden. Grundsätzlich bieten die Daten auch die Möglichkeit zu einer unterjährigen (Zwischen-)Abrechnung.

Verkehrsnachfragemodelle: Verkehrsnachfragemodelle ermitteln die Verkehrsnachfrage auf Basis der Raum- und Siedlungsstruktur, dem Verkehrsangebot und dem individuellen Mobilitätsverhalten. Sie bilden die wesentlichen Ursache-Wirkungs-Zusammenhänge zwischen der Verkehrsnachfrage und ihren Determinanten ab. Als Grundstruktur der in der Planungspraxis eingesetzten Modelle hat sich der so genannte Vier-Stufen-Algorithmus durchgesetzt, bei dem die Verkehrsnachfrage in vier aufeinander aufbauenden Teilmodellen ermittelt wird:

- Verkehrserzeugung (Ermittlung des Quell- und Zielverkehrsaufkommens einer Verkehrszelle),
- Verkehrszielwahl (Ermittlung der Verkehrsströme zwischen den Verkehrszellen),
- Verkehrsmoduswahl (Aufteilung der Verkehrsströme nach den einzelnen Verkehrsmitteln),
- Routenwahl und Umlegung (Ermittlung der Verkehrsbelastungen im Verkehrsnetz).

Je nach Teilmodell, verfügbaren Daten etc. werden unterschiedliche mathematische Ansätze genutzt. Neben Ansätzen aus der Physik (u. a. Gravitations-, Entropiemodell) werden insbesondere ökonometrische Entscheidungsmodelle (u. a. Logit-, Probit-Modell) eingesetzt. Verkehrsnachfragemodelle können auf Basis von Raumstrukturgrößen, Personengruppen oder Einzelpersonen das Verkehrsgeschehen simulieren. Detaillierte Informationen zu den einzelnen Modellen und Modelltypen sind u. a. bei (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2022), Friedrich et al. (2019), Ortúzar und Willumsen (2011) und Wermuth (2005) zu finden.

In Abhängigkeit der Komplexität des gewählten Modellierungsansatzes werden Variablen und Funktionen genutzt (etwa Widerstandsfunktionen), um die Zielwerte zu ermitteln. Mit dem Vorliegen vollständiger Quelle-Ziel-Matrizen der ÖPNV-Nachfrage in einem Untersuchungsgebiet stehen umfangreiche Daten zur Kalibrierung und Validierung verschiedener Wahlmodelle und deren Einflussgrößen (etwa Reisezeit, Preis etc.) zu Verfügung, die in Verkehrsmodellen eingesetzt werden. Ebenso besteht die Möglichkeit – sofern vorhanden – strukturelle oder linienspezifische Unterschiede zu ermitteln und in der Modellierung zu berücksichtigen. Diese neue Datengrundlage bietet insbesondere für den zweiten Modellschritt des Vier-Stufen-Algorithmus, der Verkehrsverteilung (Zielwahl), einen Mehrwert. Unter der Voraussetzung,

dass der Quellverkehr korrekt ermittelt wurde, können auch die Nutzenfunktionen für die Verkehrsmittelwahl kalibriert werden.

Anforderungen: Die Anforderung variieren mit der Komplexität und dem Detailgrad des Modellierungsverfahrens. Die räumliche Auflösung hängt von der Verkehrszelleneinteilung ab und kann daher zwischen einer haltestellenscharfen Zuordnung bis hin zu einer Zuordnung auf Stadtteilen reichen. Die zeitliche Auflösung liegt zwischen linienfahrtscharfen Daten und einer Aggregation über größere Zeiträume (etwa ein Jahr). Mikroskopische Modelle führen zu hohen Anforderungen an die räumliche und zeitliche Auflösung (etwa Modelle auf Linienfahrebene), während makroskopische Modelle eine räumliche und / oder zeitliche Aggregation zulassen.

Neben den beschriebenen, wesentlichen Anwendungsbereichen für die ÖPNV-Betreiber eröffnen die Verfahrensergebnisse weitere Anwendungsfelder:

Maßnahmenevaluation: Infolge der Umsetzung von Maßnahmen, etwa Anpassungen bzw. Neuordnung des Liniennetzes oder tarifliche Anpassungen, sind meist umfangreiche Evaluationen begleitet von Fahrgastbefragungen notwendig, um die Wirkungen der Maßnahme auf die Fahrgastnachfrage zu ermitteln. Auf Grundlage des Verfahrens können die Wirkungen auf Linienfahrebene ermittelt werden. Dabei ist insbesondere eine Betrachtung über längere Zeiträume möglich, um die Wirkung von Angebots- oder Tarifmaßnahmen analysieren zu können.

Fahrgastverhalten bei Störungen im System: Störungen im System (Ausfall, Umbau/Modernisierung) führen dazu, dass sich Fahrgastströme (temporär) verändern. Bei einer Vollausstattung der ÖV-Linien in einem Untersuchungsgebiet mit der Hardware zur Erfassung der Fahrgäste kann im Fall von Störungen eine Analyse durchgeführt werden, wie sich die Fahrgäste verhalten und welche alternativen Routen sie wählen. Mit diesem Wissen kann das Störungsmanagement etwa durch gezielten Einsatz von Ersatzverkehren oder Empfehlungen von Alternativverbindungen verbessert werden.

1.4 Aufbau der Arbeit

Die vorliegende Arbeit gliedert sich in sechs Kapitel. Das erste Kapitel thematisiert die Problemstellung und das Ziel der Arbeit: Das entwickelte Verfahren soll auf Basis neuer Eingangsdaten ÖPNV-Betreibern stets aktuelle Quelle-Ziel-Matrizen zur Abbildung der ÖPNV-Nachfrage liefern. Darüber

hinaus werden die Einsatzmöglichkeiten dieser neuen Datengrundlage dargestellt und – in diesem Unterkapitel – der Aufbau der Arbeit vorgestellt.

Kapitel 2 beginnt mit einer Beschreibung des Status Quo der Erfassung von Nachfragedaten im ÖPNV. Dazu zählen insbesondere Fahrgastzählungen, Befragungen sowie die Auswertung von ohnehin anfallenden Daten aus dem Fahrkartenverkauf (e-Ticketing), Verbindungsauskunft sowie Methoden des Trackings (Mobilfunk, WLAN, Bluetooth).

Da die Erfassung von WLAN- und Bluetooth-Signalen einen essentiellen Bestandteil des im Rahmen dieser Dissertation entwickelten Verfahrens bilden, werden im nächsten Schritt die (netzwerk-)technischen Grundlagen beschrieben. Ziel ist es, auch für Fachfremde ein grundlegendes Verständnis für die Funktionsweise von WLAN und Bluetooth sowie den erfassten Probe Requests (WLAN) und Inquiry-Response-Nachrichten (Bluetooth) zu schaffen. Dabei werden sowohl Erkenntnisse aus der Literatur ermittelt und dargestellt als auch eigene Untersuchungen durchgeführt, die der Operationalisierung der erfassten Daten im Rahmen der Verfahrensentwicklung dienen. Darüber hinaus wird das technische System zur Datenerfassung von WLAN-, Bluetooth- und AFZS-Daten vorgestellt.

Es folgt Kapitel 3, das die Entwicklungsphase des Verfahrens in drei wesentlichen Schritten aufzeigt: Filterung der Daten, Erstellung von Quelle-Ziel-Matrizen auf Basis der erfassten Daten und Hochrechnung der Daten durch Matrixausgleichsverfahren. Die durch das technische System aufgezeichneten WLAN- und Bluetooth-Daten umfassen neben den Signalen von Fahrgästen zu großen Teilen auch Signale von Nicht-Fahrgästen und bedürfen daher einer Aufbereitung. Die Datenaufbereitung beschreibt insbesondere die Filterung der Daten (unplausible Daten, Stördaten) sowie die Aufbereitung der Daten in Form von linienspezifischen Quelle-Ziel-Matrizen. Die aufbereiteten Quelle-Ziel-Matrizen werden anschließend auf Basis der erfassten AFZS-Daten, die die Anzahl der ein- und aussteigenden Fahrgäste je Haltestelle mit hoher Genauigkeit ermitteln, über sogenannte Steigerungsfaktorenmodelle hochgerechnet.

In Kapitel 4 werden die Ergebnisse der exemplarischen Anwendung des Verfahrens (Erprobungsphase) im Untersuchungsgebiet des Nordhessischen Verkehrsverbunds (NVV) dargestellt. Dazu wurden verschiedene Linien und Haltestellen im NVV ausgewählt und mit der technischen Hardware zur Datenerfassung ausgestattet. Zur Bewertung des Verfahrens ist es notwendig, Unterschiede der Modellschätzung im Bezug zu der aus

Fahrgastbefragungen bekannten realen Verteilung der Quelle-Ziel-Relationen der betrachteten Linienfahrt zu quantifizieren. Dazu werden Gütemaße definiert sowie das Vorgehen der Fahrgastbefragung kurz dargestellt.

Zur Berücksichtigung von Unsicherheiten bei der Modellschätzung werden die Eingangsparameter (i. W. Filtereinstellungen) in verschiedenen Szenarien variiert. Die Ergebnisse der Modellschätzungen in Abhängigkeit der szenarienbasierten Parametereinstellungen werden zur Ableitung der optimalen Parametereinstellungen des Modells genutzt (Kalibrierung). Ziel ist es, die Parametereinstellungen zu ermitteln, die zu Ergebnissen mit der höchstmöglichen Güte führen.

Der Mehrwert der erfassten WLAN- und Bluetooth-Daten für die Hochrechnung wird darüber hinaus über Vergleichsverfahren geprüft, das lediglich auf den Randsummen (AFZS-Daten) basiert. Die Ergebnisse werden aufbereitet und in aggregierter Form vorgestellt und anhand der definierten Gütemaße bewertet. Das Kapitel schließt mit einer Bewertung der Ergebnisse des entwickelten Verfahrens unter Einbezug der im ersten Kapitel beschriebenen Einsatzgebiete.

Mit Kapitel 5 wird eine Form der Weiterentwicklung des Verfahrens auf Basis der bestehenden Input-Daten exemplarisch untersucht. Dabei kommt es zu einer räumlichen Aggregation der Daten, sodass die Anzahl möglicher Relationen abnimmt. Die Ergebnisse der Aggregation werden hier dargestellt und den nicht-aggregierten Ergebnissen gegenübergestellt.

Im letzten Kapitel 6 werden die zentralen Ergebnisse der Dissertation zusammengefasst. Darüber hinaus erfolgt ein Ausblick und eine Vorstellung des weiteren Forschungsbedarfs.

2 Technische und methodische Grundlagen

2.1 Einführung

In diesem Kapitel werden die gängigen Verfahren zur Erfassung von Nachfragedaten im ÖPNV beschrieben, die die Grundlage für die Angebots- und Betriebsplanung sowie für weitergehende verkehrswirtschaftliche Entscheidungen darstellen. Neben der klassischen Fahrgastbefragung und -zählung werden insbesondere neue Datenquellen beschrieben, die eine aktuelle Datenbasis liefern können (AFZS, Elektronisches Fahrgeldmanagement (EFM), Mobilfunk, WLAN, Bluetooth).

Ferner wird mit den (netzwerk-)technischen Grundlagen ein Verständnis für die technische Basis des in dieser Arbeit entwickelten Verfahrens zur Ableitung von Quelle-Ziel-Matrizen auf Basis von WLAN-, Bluetooth- und AFZS-Daten vermittelt. Hierzu werden die grundlegenden Funktionen von WLAN und Bluetooth sowie die Spezifika von WLAN Probe Requests und Bluetooth Inquiry-Nachrichten beschrieben.

Es folgt die Vorstellung des technischen Systems zur Erfassung der WLAN- und Bluetooth-Daten. Die durch das technische System erfassten und gespeicherten Daten bilden die Grundlage für die weiteren Verfahrensschritte (Kapitel 2.4). Neben der Hardware zur Erfassung wird auch die Übermittlung der Daten sowie die Beschreibung der erfassten Daten thematisiert. Weiterhin wird auf die Anforderungen des Datenschutzes bzw. der Privatsphäre bei der Datenerfassung eingegangen.

In Kapitel 2.5 wird die Nutzung der durch das technische System erfassten Daten bewertet. Dabei werden die Eigenschaften der erfassten WLAN- und Bluetooth-Daten anhand einer quantitativen Datengrundlage aus dem Realbetrieb analysiert und hinsichtlich einer Operationalisierung für die Ableitung von Quelle-Ziel-Matrizen diskutiert.

2.2 Nachfragedatenquellen im ÖPNV

Zuverlässige und aktuelle Daten über die Nachfrage im Verkehrsnetz sind die Grundlage für die Angebots- und Betriebsplanung im ÖPNV, für viele verkehrswirtschaftliche Fragestellungen (z. B. Tarifgestaltung, Einnahmenaufteilung) sowie für Verkehrsmanagement und Fahrgastinformation. Diese Nachfragedaten werden bislang aufwändig i. d. R. durch Fahrgasterhebungen und Haushaltsbefragungen (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2012) generiert. Trotz des relativ großen Aufwandes sind die

erhobenen Daten häufig begrenzt hinsichtlich der zeitlichen Aussagekraft (i. d. R. Verhalten an einem Stichtag) und der Aktualität (Erhebungsrhythmus alle 3 bis 10 Jahre). Daher werden zunehmend technikbasierte Geräte und Verfahren entwickelt und genutzt, die diese Datengrundlage erweitern, wie etwa Automatische Fahrgastzählsysteme (AFZS) oder Vertriebsdaten aus dem elektronischen Fahrgeldmanagement (Verband Deutscher Verkehrsunternehmen e. V. (VDV) 2018; Forschungs-Informationen-System (FIS) 2020). In den folgenden Kapiteln werden die verschiedenen Verfahren für die Ermittlung von Nachfragedaten im Kontext des ÖPNV kurz beschrieben.

2.2.1 Fahrgastzählung

Die Zählung von Personen im Öffentlichen Verkehr dient der Ermittlung von Kennwerten zur Fahrgastmenge von Linienfahrten oder an Haltestellen und damit der Bemessung des Leistungsangebots. Als wesentliche Verfahren zur Fahrgastzählung können

- manuelle Fahrgastzählung durch mitfahrendes Personal und
- Automatische Fahrgastzählsysteme (AFZS)

unterschieden werden. Bei einer Zählung werden die ein- und aussteigenden Fahrgäste eines Linienfahrzeugs gezählt. Daher kann als Standort für eine Zählung sowohl das Fahrzeug selbst als auch die Haltestelle dienen. Die zu ermittelnden Kennwerte einer Zählung bilden u. a. Haltestellen-, Querschnitts-, Linien- und Streckenbelastungen (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2012).

2.2.1.1 Manuelle Fahrgastzählung

Bei der manuellen Zählung erfasst das Zählpersonal mittels Formblättern, Handzählgeräten oder computergestützt (z. B. Smartphone-App) die Fahrgastmengen. In der Regel werden manuelle Fahrgastzählungen mit Fahrgastbefragungen kombiniert. Um etwaige Zählfehler zu korrigieren, wird neben den Einsteigern häufig auch die Besetzung des Fahrzeugs während der Linienfahrt gezählt. Die Auswahl und Schulung des Erhebungspersonals sowie die Einsatzplanung insbesondere bei stark frequentierten Linienfahrten haben eine hohe Bedeutung für die Qualität der Ergebnisse. Fahrgastzählungen können im Rahmen der Stichprobenplanung hinsichtlich der Erhebungsart (Teilerhebung / Vollerhebung) sowie verschiedener Merkmale zur Schichtung (Jahreszeit, Tagestyp, Zeitschicht) differenziert werden (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2012).

2.2.1.2 Automatische Fahrgastzählsysteme

Automatische Fahrgastzählsysteme (AFZS) bezeichnen eine Form der Fahrgastzählung, bei der die ein- und aussteigenden Fahrgäste mit Hilfe technischer Geräte gezählt werden. AFZS-Geräte werden in der Regel im Eingangsbereich der Fahrzeugtüren angebracht und zählen damit die Fahrgäste bezogen auf eine Linienfahrt. Auf Grundlage der erfassten Daten kann die Fahrzeugbelegung sowie die Verkehrsmenge (Linienbeförderungsfälle) und Verkehrsleistung (Personenkilometer, Pkm) abgeleitet werden. Da die einzelnen Fahrtweiten der Fahrgäste nicht bekannt sind, können keine Aussagen zur Fahrtweitenverteilung oder zu Quelle-Ziel-Relationen getroffen werden.

Zur Gewährleistung einer Vergleichbarkeit der erfassten AFZS-Daten zwischen Verkehrsunternehmen und vor dem Hintergrund der Nutzung dieser Daten für die Verfahren der Einnahmenaufteilung wurden (hohe) Anforderungen an die Messgenauigkeit von AFZS-Geräten durch die VDV-Schrift 457 (Verband Deutscher Verkehrsunternehmen e. V. (VDV) 2018) definiert. Unter anderem darf die Summe der Ein- und Aussteiger aller Halte nur um 1% von dem Realwert (gemessen durch Referenzzählungen) abweichen, gleichzeitig dürfen nicht mehr als 5% der Halte fehlerhafte Daten erfassen. Unter „fehlerhaft“ werden hier Abweichungen um mehr als ein Drittel verstanden.

Der Vorteil von AFZS gegenüber manuellen Fahrgastzählungen ist der geringe Erhebungsaufwand und die Verfügbarkeit von Echtzeitdaten. Demgegenüber stehen vergleichsweise hohe Investitionskosten durch die Ausstattung (Schnieder 2018). Zu den AFZS zählen – unabhängig von der Einhaltung der Anforderungen des VDV – Drucksensoren, Lichtschranken, Infrarotsysteme und Kameras (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2012; Rakebrandt 2013).

2.2.2 Fahrgastbefragung

Fahrgastbefragungen eignen sich zur Ermittlung diverser Kennzahlen zur Nachfrage sowie von Einstellungen und Präferenzen der Nutzer. Anhand von Fahrgastbefragungen können Informationen wie z. B. Quelle, Ziel und Zweck des Weges sowie der genutzte Tarif und persönliche Einstellungen (etwa Häufigkeit der Nutzung des ÖPNV im Monat) ermittelt werden. Die Befragungen werden überwiegend in den Fahrzeugen oder an Haltestellen durchgeführt. Dabei führt das Erhebungspersonal mithilfe von Fragebögen in analoger (PAPI – Paper and pencil interviewing) oder digitaler Form (CAPI – computer assisted personal interviewing) ein persönliches Interview mit den

Fahrgästen. Um repräsentative Ergebnisse ableiten zu können, ist eine Kombination mit Fahrgastzählungen zur Hochrechnung unvollständiger Befragungsdaten erforderlich. Je nach Erhebungsziel und Aufgabenstellung variieren die zu ermittelnden Kennwerte und dementsprechend auch die Inhalte der Fahrgastbefragung. Sowohl in der Vorbereitung als auch in der Durchführung und Auswertung sind Fahrgastbefragungen sowohl zeitlich als auch finanziell aufwändig. Darüber hinaus bilden die Ergebnisse lediglich eine zeitliche begrenzte Momentaufnahme der Nachfragesituation (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2012).

2.2.3 Vertriebsdaten

2.2.3.1 Klassischer Vertrieb

Aus dem Verkauf von Fahrausweisen können je nach Ticketart unterschiedliche Informationen gewonnen werden. Im Wesentlichen können Tarifprodukte für Gelegenheitskunden (Einzelfahrkarte, Kurzstrecke, Tageskarte) und Zeitkarten (Wochen-, Monats-, Jahreskarte inkl. Schüler- und Studententicket) unterschieden werden. Während im Bartarif zum Teil Informationen zur Start- und Zielhaltestelle enthalten sind, wird bei Zeitkarten in der Regel lediglich der räumliche und zeitliche Geltungsbereich erfasst. Unter Zuhilfenahme von Nutzungshäufigkeiten für die verschiedenen Tarifprodukte (z. B. aus repräsentativen Fahrgastbefragungen) können die Fahrgastzahlen abgeschätzt werden (Bieland und Sommer 2021).

2.2.3.2 Elektronisches Fahrgeldmanagement (EFM)

Elektronische Fahrausweise (E-Tickets) im ÖPNV sind Bestandteil des sogenannten EFM (Elektronisches Fahrgeldmanagement) und werden durch die Verkehrsunternehmen und Verbünde in Deutschland stufenweise eingeführt.

In der ersten Stufe der VDV-Kernapplikation (VDV-KA) (eBezahlen) wird es dem Kunden ermöglicht, die Fahrtberechtigung in Form eines Papierscheins bargeldlos zu erwerben. Dazu stehen verschiedene Bezahlmöglichkeiten beim Verkehrsunternehmen bzw. -verbund zur Verfügung: Prepaid-Abrechnung, Postpaid-Abrechnung oder On-Trip-Abrechnung (VDV-KA KG 2012).

Mit der zweiten Stufe der VDV-KA (eTicket) wird kein klassischer Papierfahrschein mehr benötigt. Stattdessen erfolgt die Fahrtberechtigung elektronisch, sodass ein virtueller Fahrschein auf einem elektronischen Medium – auf der Chipkarte selbst oder in einem Hintergrundsystem – abgelegt wird. Bei

sogenannten „Handy-Tickets“ wird als elektronisches Medium das Smartphone des Fahrgastes genutzt (Forschungs-Informationssystem (FIS) 2020).

Die dritte Stufe des EFM liegt in der automatischen Fahrpreisberechnung: Hierbei wird der Fahrpreis auf Basis der zurückgelegten Strecken in einem definierten Zeitraum (ggf. unter Berücksichtigung weiterer Bedingungen) automatisch ermittelt. Voraussetzung für die Abrechnung sowie neue Tarifmodelle (entfernungsbasierte Tarife) ist die genaue Erfassung des Ein- und Ausstiegsorts mittels In-Out-Systemen. Beispielhaft seien hier CICO (Check-In/Check-Out) oder BIBO (Be-In/Be-Out) genannt (Forschungs-Informationssystem (FIS) 2019).

Beim CICO-Ansatz muss der Fahrgast selbst tätig werden und sich aktiv bei Eintritt in das Fahrzeug mit seinem Nutzermedium anmelden und beim Aussteigen entsprechend abmelden. Die Anmeldung kann sowohl durch das Vorhalten eines NFC-fähigen Smartphones (NFC - Near-Field-Communication) oder durch Vorhalten der RFID-Chipkarte (RFID – radio-frequency identification) an ein Terminal durchgeführt werden (VDV eTicket Service GmbH & Co. KG o.J.). Im Gegensatz dazu wird mit dem BIBO-Ansatz die Anwesenheit des elektronischen Nutzermediums automatisch und berührungslos registriert, sodass keine Interaktion mit dem Fahrgast notwendig ist, um Fahrtbeginn und -ende genau zu bestimmen (Forschungs-Informationssystem (FIS) 2019). Während CICO-Systeme in vielen Städten bereits eingesetzt werden (z. B. Schwäbisch-Hall (ALMEX GmbH 2021), London (Oyster Card, (Transport for London (TfL) 2016), wurden BIBO-Systeme bisher nur in Pilotprojekten getestet, wie „EasyRide“ in den Regionen Genf und Basel (Forschungs-Informationssystem (FIS) 2019; BLIC GmbH und KCW GmbH 2014).

Die für das E-Ticket und insbesondere für die automatische Fahrpreisberechnung erforderlichen Daten beinhalten Informationen zu Quelle-Ziel-Beziehungen. Weiterhin können ggf. Umsteigebeziehungen durch Routingalgorithmen abgeleitet werden. Da für die Nutzung von E-Tickets eine Registrierung mit Namen, Adresse und Geburtsdatum vorausgesetzt wird, können neben fahrtenbezogenen Daten auch Informationen über die demographische Struktur der E-Ticket-Kunden erfasst werden (Messe Berlin 2020).

2.2.4 Verbindungsauskunftsdaten

Mithilfe von Datenauswertungen der Verbindungsabfragen elektronischer Fahrplanauskünfte, wie z. B. der App der Deutschen Bahn, können

Reiseketten nachgebildet und speziell Umsteigebeziehungen an Knotenpunkten erschlossen werden. Neben Start- und Zielhaltestelle der Fahrt können auch die vor- oder nachgeschalteten Verkehrsmittel mit einbezogen werden. Bei der Nutzung elektronischer Verbindungsanfragen zur Ableitung der Verkehrsnachfrage sind die Zusammensetzung der Fahrgäste (Stammkunden mit routinisierten Wegen, Gelegenheitskunden, Touristen und Selten-Nutzer) und das Nutzungsverhalten – wie etwa der Häufigkeit der Anfragen für eine Verbindung, der Genauigkeit der Anfrage (Ein- und Ausstiegsort) oder der generellen Nutzung von Anfragen (regelmäßigen Fahrten erfordern gegenüber unregelmäßigen Fahrten keine Anfragen) – zu berücksichtigen. Aufgrund der (unbekannten) Verzerrungen sind die Verbindungsauskunftsdaten nur schwer zu interpretieren.

Aktuell nimmt die Operationalisierung von Verbindungsauskunftsdaten im ÖPNV zu Planungs- und Informationszwecken zu. Dazu werden u. a. Verbindungsauskunftsdaten mit weiteren Datensätzen (etwa Daten aus dem elektronischen Ticketing) kombiniert, um Belegungsprognosen für die einzelnen Linienfahrten zu erstellen (HaCon Ingenieurgesellschaft mbH 2020; Sommer et al. 2019).

2.2.5 Mobilfunk-, WLAN- und Bluetooth-Daten

Mit der Verbreitung mobiler Geräte in der Bevölkerung stehen Akteuren der Stadt- und Verkehrsplanung eine Vielzahl von Daten („Floating Phone Data“) zur Erfassung von Mobilitätsmustern zur Verfügung. Die zunehmende Nutzung von GPS-, WLAN-, und Bluetooth-Technologien auf mobilen Geräten kann für die Verfolgung von Objekten oder Personen, das sogenannte „Tracking“, genutzt werden.

Der Nutzen der gewonnenen Daten ist vielfältig und reicht von Analysen des Einkaufsverhaltens (et solutions, 2020) über Besucherstromzählungen (Jarvis et al. 2019; Deutscher Bundestag 2017) bis hin zur Ableitung von Informationen zur Auslastung von öffentlichen Räumen und Einrichtungen (Universität Kassel 2020). In der Verkehrsplanung werden Bluetooth- und WLAN-Sensordaten der Verkehrsteilnehmer unter anderem für die Echtzeit-Erfassung von Reisezeiten, die Kalibrierung von Verkehrsmodellen und den Wirkungsnachweis von Maßnahmen genutzt (c.c.com Moser GmbH 2020). Im Kontext des ÖPNV stellen Trackingverfahren mit Hilfe von Mobilfunk-, E-Ticket-, Bluetooth- und WLAN-Daten potentielle Alternativen zur manuellen Fahrgasterhebung dar. Die Tracking-Daten sollen im ÖPNV beispielsweise

für die Planung und Optimierung der öffentlichen Verkehrssysteme und für Echtzeit-Fahrgastinformation eingesetzt werden.

Während Floating-Car-Data bereits kommerziell bereitgestellt werden, ist die Nutzung von Floating-Phone-Data und insbesondere Mobilfunkdaten in der Regel auf Pilotprojekte begrenzt, wenngleich diesen Daten ein hohes Potential zugeschrieben wird. Schwächen werden insbesondere in der fehlenden Möglichkeit zur Beschreibung – und damit auch zur Interpretation oder eingeschränkten Nutzung – der (großen) Stichprobe gesehen (Schmidt und Männel 2017).

Tracking-Daten werden dabei als ergänzende Datenquelle zusätzlich zu Zähldaten, empirisch ermittelten Werten aus Verkehrserhebungen oder Live-Daten der Verkehrsunternehmen eingesetzt (Thorun et al. 2018).

2.2.5.1 Mobilfunk-Daten (GSM-Daten)

Mobilfunkdaten basieren auf der Kommunikation von Mobiltelefonen und Basisstationen bzw. Funkzellen und werden durch Telekommunikationsdienstleister erfasst. Diese Daten ermöglichen erst die funktionierende Nutzung von Endgeräten im Mobilfunknetz. Die Zuordnung von Signalen einzelner Endgeräte zu Mobilfunkzellen ermöglicht die zeitliche und räumliche Nachverfolgung einzelner Endgeräte.

Dabei ist zu berücksichtigen, dass eine funkzellenscharfe Zuordnung lediglich bei aktiven Ereignissen (Nutzung des Endgeräts zum Telefonieren, Internet, SMS) möglich ist. Bei passiver Nutzung werden Daten bei Übertritt eines Endgeräts in eine anderen Mobilfunkzelle oder zur Regelkommunikation des Endgeräts mit dem Netzwerk einem Mobilfunkzellenverbund (Location Area) zugewiesen (Schmidt und Männel 2017).

Die Genauigkeit der Standortbestimmung ist dabei von verschiedenen Einflussfaktoren, wie etwa der Ausstattung des Raums mit Basisstationen, gerätespezifischen Eigenschaften des Mobilfunktelefons sowie dem genutzten Verfahren zur Positionsbestimmung abhängig. Zur Positionsbestimmung können unter anderem die

- Cell-ID-Methode und Timing Advance oder das
- EOTD-Verfahren (Enhanced Observed Time Difference)

eingesetzt werden. Die Cell-ID-Methode basiert auf der Ortung der Funkzelle, in der das Mobiltelefon eingebucht ist. Die Reichweite der Funkzellen und damit auch die Genauigkeit der Ortung variieren von ca. 100 bis 200 m im

städtischen Raum bis hin zu über 30 km in ländlichen Regionen. Dabei gilt: Je kleiner der Radius der Funkzelle, desto genauer kann die Ortung stattfinden. Zur Erhöhung der Genauigkeit kann der Parameter Timing Advance genutzt werden. Der Parameter Timing Advance (Laufzeit) gibt Aufschluss über den Abstand zwischen dem Mobiltelefon und der Basisstation¹. Das EOTD-Verfahren nutzt Signale von mehreren Basisstationen und ermöglicht auf Grundlage der Zeitdifferenzen (Ankunftszeiten der Signale von mehreren Basisstationen) eine genauere Positionsbestimmung. Die Genauigkeit der Ortung kann dadurch auf bis zu 30 m verbessert werden. Darüber hinaus ist die Möglichkeit zur Kommunikation mit mehreren Basisstationen in vielen ländlichen Gebieten nicht gegeben. Um die Genauigkeit der Ortung zu erhöhen, wäre ein Ausbau der Netzinfrastruktur notwendig (Schelewsky 2014; Elektronik-Kompendium.de 2018).

Mobilfunknetze sind inzwischen überall verfügbar (Schelewsky 2014), sodass durchgehend Daten von mobilen Endgeräten anfallen – bei Telefónica Deutschland werden beispielsweise täglich rund fünf Milliarden Datenpunkte von über 44 Millionen Mobilfunkanschlüssen erzeugt (Telefónica Germany NEXT GmbH 2018; Telefónica Germany GmbH & Co. OHG 2021). Somit liegen zeitlich hochaufgelöste Daten mit einer räumlichen Differenzierung auf Postleitzahlebene vor. Durch diese Datenbasis ist es möglich, verschiedene Analysen durchzuführen und somit Erkenntnisse hinsichtlich von Bewegungsströmen, Tagesganglinien oder auch Einzugsbereichen von Großveranstaltungen zu erlangen.

Telefónica Germany GmbH (ehem. Telefónica NEXT) kooperiert in verschiedenen Projekten mit Anwendungs- und Forschungspartnern, um die Nutzbarkeit der Daten, u. a. im Verkehrssektor, zu verdeutlichen. Beispielhaft wurden

- in Kooperation mit dem Fraunhofer IAO (Institut für Arbeitswirtschaft und Organisation) und Teralytics der Beitrag von Mobilfunkdaten für die Verkehrsplanung für den Raum Stuttgart untersucht (Schmidt und Männel 2017),
- in Kooperation mit der flinc GmbH die Auswirkungen der kompletten Ersetzung der Privatfahrzeuge durch ein Shuttle-System für den Raum Hamburg untersucht (flinc GmbH 2016),

¹ Infrastruktur zur Übertragung von Funksignalen, z.B. von Mobilfunknetzen.

- in Kooperation u. a. mit HaCon, IVU Traffic Technologies AG im Projekt ProTrain (BMVI) eine bessere Verteilung der Fahrgäste im SPNV in Brandenburg angestrebt. Bestehende Kapazitäten sollten besser genutzt werden, indem die erwartete und tatsächliche Nachfrage detaillierter bestimmt wurde.

Daneben werden Mobilfunkdaten, sogenannte „Floating Phone Data“ (FPD), im Rahmen des Forschungsprojekts „xMND“ (extended Mobile Network Data, xMND) in Leipzig und München zusammen mit den GPS-Daten von Smartphones freiwilliger Teilnehmer gesammelt und u. a. zum Monitoring der Verkehrsnachfrage im ÖPNV ausgewertet (Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) 2021). Dabei ist zu beachten, dass dieses Vorgehen die aktive Teilnahme und Zustimmung der Smartphonebesitzer erfordert. Darüber hinaus werden FPD im Routing von Navigationssystemen (Echtzeit-Navigation) oder als Datengrundlage für Verkehrsanalysen (u. a. Reisezeitanalysen) kommerziell genutzt (INRIX Inc. 2020; TomTom International BV 2021).

Pinelli et al. (2016) nutzen Mobilfunkdaten der Firma Orange in Abidjan (Elfenbeinküste) mit dem Ziel, die Reisezeiten der Nutzer zu verringern. Dazu werden Quelle-Ziel-Matrizen auf Basis der vorliegenden Daten erstellt und durch Routing-Algorithmen ein theoretisches Einsparpotential von 27% der Reisezeiten durch Ermittlung des Nutzergleichgewichts errechnet. Huang et al. (2019) zeigen die Schwächen von Studien zur Operationalisierung von Mobilfunkdaten mit besonderen Augenmerk auf die Abgrenzbarkeit von Verkehrsmitteln. Insbesondere wird hier bemängelt, dass zwar Methoden zur Operationalisierung von Mobilfunkdaten ermittelt werden, jedoch eine Evaluation der Ergebnisse aufgrund der fehlenden Datengrundlage ausbleibt.

Die Eignung von Mobilfunkdaten für die Angebotsplanung im lokalen und regionalen ÖPNV sowie für die Nutzung in Informationsdiensten für Reisende (Fahrgastlenkung) kann aus verschiedenen Gründen als problematisch betrachtet werden:

- **Abhängigkeit von Netzbetreibern / laufende Kosten für die Datenbeschaffung:** Die Daten fallen bei den Netzbetreibern an. Der Zugang zu den Mobilfunkdaten für Forschungseinrichtungen, Verkehrsunternehmen oder Verkehrsverbände kann nur durch eine Kooperation mit dem entsprechenden Netzbetreiber erfolgen oder durch den Kauf der Daten. Gerade Verkehrsunternehmen benötigen für die Angebotsplanung und

die Wirkungskontrolle von Maßnahmen der Angebotsplanung Längsschnittdaten. Unabhängig davon, ob die Daten gekauft oder durch eine Kooperation bereitgestellt werden, besteht eine Abhängigkeit von dem jeweiligen Netzbetreiber.

- **Genauigkeit der Mobilfunkdaten:** Mobilfunkdaten können bis auf 30 m genau den Standort des Mobiltelefons bestimmen. Nach Schelewsky (2014) liegt die Genauigkeit einer Positionsbestimmung über Mobilfunk nicht unter 200 m. Die Genauigkeit nimmt dabei in Abhängigkeit der Anzahl und Verteilung der Basisstationen im Raum ab, sodass gerade in ländlichen Regionen Ungenauigkeiten von mehr als 1.000 m auftreten können. In Abhängigkeit des Untersuchungsziels, der Planungsebene sowie des Verkehrszellen- und Netzmodells bestehen unterschiedliche Anforderungen an die Genauigkeit der Positionsdaten. Die Ermittlung von Quelle-Ziel-Matrizen auf Bundesebene (Bundesverkehrswegeplanung) hat geringere Ansprüche an die Genauigkeit der Ortung als vergleichbare Untersuchungen auf regionaler oder kommunaler Ebene (Sommer 2002).
- **Abgrenzung verschiedener Verkehrsmittel:** Infolge der Größe der Funkzellen sind für eine Zuordnung der Verkehrsteilnehmer zu bestimmten Verkehrsmitteln (Bus, Pkw, Fahrrad etc.) weitergehende Analyse-tools notwendig (Dialog Publishers Verlagsgesellschaft 2017). Eine valide Zuordnung der erfassten Mobilfunkdaten zu ÖV-Wegen (oder anderen Verkehrsmitteln) ist somit (noch) nicht möglich, jedoch essentiell für die zielgerichtete Angebotsplanung der Verkehrsunternehmen.

2.2.5.2 WLAN- und Bluetooth-Daten

Im Hinblick auf die große Verbreitung „smarter“ Mobilgeräte hat die Nutzung von Bluetooth- und WLAN-Daten in den letzten Jahren zunehmend an Aufmerksamkeit gewonnen (Baeta et al. 2017). Gleichzeitig bieten immer mehr (deutsche) Verkehrsunternehmen einen (kostenlosen) WLAN-Zugang für ÖPNV-Kunden in den Fahrzeugen an. Aktuell sind bereits verschiedene Regionalbuslinien beispielsweise in den Bedienungsgebieten des Münchner Verkehrs- und Tarifverbundes (MVG) und des Nordhessischen Verkehrsverbunds (NVV) mit der entsprechenden Technik ausgestattet.

Bislang werden anfallende WLAN-Daten in Deutschland zur Analyse des Einkaufsverhaltens genutzt, etwa um Kundenfrequenzen, Bewegungsprofile im Ladengeschäft, Verweildauern oder Konversionsraten zu bestimmen (Ipsos

Retail Performance 2016; Lukaß 2016). Möglichkeiten der Nutzung von Bluetooth- und WLAN-Daten im ÖPNV wurden durch verschiedene Forschungsarbeiten untersucht und umfassen im Wesentlichen die Bereiche:

- Abschätzung von Fahrgastzahlen (Myrvoll et al. 2017; Oransirikul et al. 2014; Bai et al. 2017; Pattanusorn et al. 2016; Mikkelsen et al. 2016; Song und Wynter 2017; Baeta et al. 2017; Oransirikul et al. 2019; Nitti et al. 2020; Pu et al. 2019; Paradedada et al. 2019)
- Nachverfolgung von Wegen (Hidayat et al. 2020; Dunlap et al. 2016; Sapiiezynski et al. 2015; Liebig und Wagoum 2012; Versichele et al. 2012; Haase und Handy 2004; Ji et al. 2017)

Dabei ist zu berücksichtigen, dass bei den aufgeführten Quellen die Validierung der entwickelten Methoden anhand von quantitativen Messungen im realen Testumfeld fehlen. Es werden lediglich Pilotprojekte, etwa zur generellen Tauglichkeit des Trackings beschrieben. Das Tracking umfasst Anwendungen im ÖPNV, in Innenräumen sowie während Veranstaltungen im Freien. Dabei werden keine Anforderungen an die Genauigkeit der Ergebnisse definiert und entsprechend keine realen Vergleichsdaten erfasst.

Das Ausbleiben der Analyse zur Genauigkeit ist auf den hohen Aufwand zur Erhebung wahrer Quelle-Ziel-Matrizen zurückzuführen (Pu et al. 2019). In Arbeiten, die die Ergebnisse der entwickelten Verfahren realen Daten (Quelle-Ziel-Matrix, Ein- und Aussteigerzahlen) gegenüberstellen, konnten Differenzen zwischen den realen Daten sowie den Verfahrensergebnissen nicht vermieden werden (Hidayat et al. 2018; Ji et al. 2017; Oransirikul et al. 2014). Dies ist nach Pu et al. auf die harten (unflexiblen) Filterkriterien und Spannweiten zurückzuführen. Einen Überblick über bisherige Arbeiten sowie deren Schwächen in Bezug auf die Verarbeitung von Daten, die im realen Umfeld erhoben werden, bietet Paradedada et al. (2019). Dieser macht insbesondere auf die Überschätzung der Prognosequalität aufgrund einer starken Aggregation der Ergebnisse aufmerksam. So werden beispielsweise lediglich die Gesamtzahl an Ein- und Aussteigern einer Linienfahrt berücksichtigt, statt Angaben je Haltestelle zu vergleichen.

Aktuelle Veröffentlichungen beschäftigen sich insbesondere mit der Entwicklung geeigneter Algorithmen, um die Genauigkeit der WLAN-Tracking Ergebnisse zu verbessern (Nitti et al. 2020; Paradedada et al. 2019; Ryu et al. 2020). Nitti et al. (2020) haben dabei Tests unter Laborbedingungen durchgeführt, in dem mit einem Pkw und acht Endgeräten die Erzeugung von Quelle-Ziel-

Matrizen über 13 Haltestellen nachgewiesen wird. Dazu wurden die Endgeräte bei einem (theoretischen) Einstieg angeschaltet, bei einem Ausstieg ausgeschaltet. Darüber hinaus war es hier möglich, die (theoretischen) Insassen auf Basis der empfangenen Signalstärke eindeutig zu identifizieren. Der Ansatz von Ryu et al. (2020) basiert auf der Erfassung von WLAN-Signalen lediglich an Haltestellen mit einem Untersuchungsgebiet, das aus zwei Haltestellen mit geringer räumlicher Entfernung als Einstiegshaltestellen sowie zwei Haltestellen mit geringer räumlicher Entfernung als Ausstiegshaltestellen besteht.

Zwar hat das Forschungsinteresse zur Nutzung von WLAN- und Bluetooth in den letzten Jahren zugenommen, jedoch fehlt bislang die Anwendung von Verfahren zur Bestimmung von Quelle-Ziel-Relationen für den ÖPNV in der Praxis und die Validierung der Ergebnisse. Ein Großteil der genannten Publikationen fokussiert die Belegung der Fahrzeuge und behandelt die Ableitung von Quelle-Ziel-Relationen nicht prioritär. Daher werden Verfahren zur Abgrenzung von Fahrgastdaten gegenüber Stördaten auf Basis der Signalstärke vorgestellt, deren Anwendbarkeit in der Praxis jedoch anzuzweifeln sind (vgl. Kapitel 2.5.1). Weiterhin fehlt aus Sicht des Autors der Umgang mit Unschärfen bei der Erfassung von Signalen in Fahrzeugen, die zur Zuordnung von Ein- und Ausstiegsorten genutzt werden. Diese Unschärfen sind auf Basis des Sendeverhalten der Endgeräte zurückzuführen und werden in Kapitel 2.5.1 erläutert. Die fehlende Berücksichtigung des Senderverhaltens ist aus Sicht des Verfassers auf die Ausrichtung der Projekte zurückzuführen, vornehmlich die Belegung zu ermitteln. Hinsichtlich der Ableitung von Quelle-Ziel-Matrizen werden durch Hidayat et al. (2020) nur grobe Verflechtungen zwischen Haltestellen aufgezeigt. Pu et al. (2019) dagegen ermitteln eine Quelle-Ziel-Matrix und vergleichen diese lediglich mit den realen Ein- und Ausstiegszahlen an den Haltestellen (Randsummen, AFZS-Daten). Die Integration dieser Datenquelle zur Erhöhung der Genauigkeit der Schätzung wird jedoch nicht berücksichtigt.

Im internationalen Umfeld wird das WLAN-Tracking bereits in der Praxis angewandt.

- In den Zügen der niederländischen Eisenbahngesellschaft Nederlandse Spoorwegen N.V. wird seit 2010 freies WLAN angeboten (NL Times 2013). Die Passagiere stimmen mit der Nutzung des WLANs der anonymisierten Nutzung ihrer MAC-Adresse zu. Die Daten geben den Einblick in die Gehströme, Gehrouten, Aufenthaltszeiten und Wartebereiche der

Fahrgäste und anderer Bahnfansnutzer (NS Groep N.V. 2020a). In den Zügen zwischen Arnheim, Nijmegen und Den Bosch wird zusätzlich eine Seatfinder App getestet, die auf Basis von WLAN-Verbindungen im Zug anzeigt, in welchem Abteil freie Plätze zu finden sind (NS Groep N.V. 2020b). Seit dem 07. Januar 2021 hat Nederlandse Spoorwegen die Erfassung von WLAN-Daten eingestellt (NS Groep NV 2021). Die Gründe dafür werden nicht genannt.

- Die Londoner Verkehrsbehörde Transport for London (TfL) analysiert WLAN-Daten in Londoner U-Bahn-Stationen, um den Betrieb, die Planung und die Informationsbereitstellung an die Kunden zu optimieren. Nach einer vierwöchigen Pilot-Phase im Jahr 2016 wird das Tracking seit 2017 permanent in den U-Bahn Stationen durchgeführt (Transport for London (TfL) 2017, 2019). Seit der Festeinführung sind 97% der Londoner U-Bahn-Stationen mit WLAN Access Points (AP) ausgestattet. Ziel der Datenerfassung ist primär die Ermittlung der Zeit von Fahrgästen, um von den Stationen bis zu den U-Bahnen zu gelangen. Die Reisezeitinformationen auf der Webseite und der App der Behörde werden an die ermittelte Dauer der Fußwege angepasst. Daneben konnten aus den WLAN-Daten Erkenntnisse zur Auslastung der einzelnen Stationen abgeleitet werden. Diese Informationen werden den Kunden zugänglich gemacht. Darüber hinaus werden die Daten genutzt, um Werbeinnahmen zu steigern: Durch Daten, die die Laufwege der Fahrgäste innerhalb der Stationen zeigen, konnten prominente Flächen für Werbung identifiziert und vermarktet werden (Transport for London (TfL) 2019, 2021; Weinstein 2016). Dabei sind weitergehend die besonderen Eigenschaften der U-Bahn-Stationen (weitgehende Abschirmung gegenüber Stördaten) zu berücksichtigen. Die Ableitung von Quelle-Ziel-Verflechtungen im Verkehrsnetz ist nicht vorgesehen.

Obwohl das Potential der Daten in der Forschung als sehr hoch angesehen wird, mangelt es bisher an Anwendungen in der Praxis in Deutschland außerhalb von Pilotprojekten. Die fehlende Umsetzung kann zumindest teilweise auf die Anforderungen des Datenschutzes zurückgeführt werden: In Deutschland werden beispielsweise zahlreiche Diskussionen darüber geführt, ob die MAC-Adresse als personenbezogenes Datum gewertet wird, sodass die Erfassung und Speicherung datenschutzrechtliche Handlungen erfordert.

Darüber hinaus wird die Nachverfolgbarkeit einzelner Endgeräte – wie es in den o. g. Forschungsberichten durchgeführt wurde – durch den zunehmenden Schutz der Privatsphäre seitens der Hersteller erschwert. Diese verändern softwareseitig die bislang „festen“ Identifikationsmerkmale einzelner Endgeräte (Kapitel 2.3.2.5).

2.2.6 Übersicht von Vor- und Nachteilen der Nachfragedatenquellen

In Tabelle 2-1 werden die beschriebenen Nachfragedatenquellen kurz hinsichtlich ihrer Vor- und Nachteile bewertet. Darüber hinaus wird auf die Nutzbarkeit für Anwendungen im ÖPNV eingegangen.

Tabelle 2-1: Übersicht der Vor- und Nachteile der behandelten Nachfragedatenquellen

Nachfragedatenquelle	Vorteile	Nachteile	Nutzbarkeit für Anwendungen im ÖPNV
Mobilfunk-, Bluetooth- und WLAN-Daten	<ul style="list-style-type: none"> geringer Investitionsaufwand für Erfassungsinfrastruktur stets aktuelle Daten über lange Zeiträume durch dauerhafte Erfassung 	<ul style="list-style-type: none"> Maßnahmen zum Schutz der Privatsphäre durch OS-Hersteller (OS = Operating System) nehmen zu, wodurch die Nachverfolgbarkeit beeinträchtigt wird Filterung von Stördaten notwendig (unbekannte) Verzerrungen der Stichprobe: Nur eine Teilmenge der Grundgesamtheit wird erfasst, mehrfache Erfassung einzelner Personen. Es sind zusätzliche Informationen zur Stichprobe notwendig, insbesondere zur Hochrechnung Unschärfen aufgrund von Funkzellen sowie dem Sendeverhalten keine Information zu genutztem Tarif 	<p>Diese Datenart bietet grundsätzlich großes Potential für Anwendungen im ÖPNV. So könnten die Auslastung, Dauer des Aufenthalts an Bahnhöfen oder anderen Standorten, Quelle-Ziel-Matrizen sowie Umstiege ermittelt werden. Derzeit ist dies jedoch aufgrund von Unschärfen sowie Stördaten und Verzerrungen nicht gelungen.</p>
Zählraten AFZS	<ul style="list-style-type: none"> hohe Genauigkeit durch Anforderungen an AFZS stets aktuelle Daten über lange Zeiträume durch dauerhafte Erfassung langfristige Kostenvorteile Objekterkennung möglich 	<ul style="list-style-type: none"> hoher Investitionsaufwand Notwendigkeit von Planungen für den Fahrzeugeinsatz in Abhängigkeit der Ausstattungsrate, um ein "Gesamtbild" zu erhalten keine Information zu genutztem Tarif, Nutzungshäufigkeit etc. 	<p>Diese Datenart misst bei jeder Linienfahrt der ausgestatteten Fahrzeuge die Ein- und Aussteiger. Die Daten liegen somit über einen langen Zeitraum vor. Quelle-Ziel-Matrizen können nicht abgeleitet werden, Umstiege können nicht verfolgt werden.</p>

Nachfrage- datenquelle	Vorteile	Nachteile	Nutzbarkeit für Anwen- dungen im ÖPNV
Zähl- daten manuell	<ul style="list-style-type: none"> • Personalplanung und -einsatz flexibler als Fahrzeugdisposition • Möglichkeit zur Ergänzung einer Befragung 	<ul style="list-style-type: none"> • hoher finanzieller Aufwand für Personal • hoher zeitlicher Aufwand für die Durchführung • Aussagekraft der Daten ist i.d.R. zeitlich begrenzt • Zählungen durch Personal sind fehleranfällig • keine Information zu genutztem Tarif, Nutzungshäufigkeit etc. 	<p>Mit der manuellen Erfassung von Zähl- und Belegungsdaten für den Zeitraum der Erhebung generiert. Ohne zusätzliche Befragungen sind hier - wie auch bei den automatischen Zähl- und Belegungsdaten - keine weiteren Aussagen zu Quelle und Ziel der Fahrgastfahrt oder genutzten Tarifen möglich.</p>
Befragung	<ul style="list-style-type: none"> • Personalplanung und -einsatz flexibel • Erfassung umfassender Informationen (u. a. Tarif, Quell, Ziel, Soziodemographie, Einstellungen) • Repräsentativität durch Vollerhebung / Zufallsauswahl 	<ul style="list-style-type: none"> • hoher finanzieller Aufwand für Personal • hoher zeitlicher Aufwand für die Durchführung • Aussagekraft der Daten ist i.d.R. zeitlich begrenzt • Befragungen durch Personal sind fehleranfällig. Schulungen notwendig, Einfluss des Interviewers auf Ergebnisse • Kenntnisse zur Grundgesamtheit notwendig zur Hochrechnung 	<p>Befragungen bieten die Möglichkeit, Informationen zu den Fahrgästen zu erfassen, die durch automatische Erfassungen nicht ermittelt werden können. Ähnlich wie die manuelle Erfassung von Zähl- und Belegungsdaten stellen die Ergebnisse eine zeitlich begrenzte Momentaufnahme dar.</p>
Verbin- dungs- aus- kunft	<ul style="list-style-type: none"> • stets aktuelle Daten über lange Zeiträume durch dauerhafte Erfassung • Daten besitzen einen Quelle-Ziel-Bezug (Ein- und Ausstiegshaltestelle) • Daten liegen für das gesamte Bedienungsgebiet des Verkehrsunternehmens bzw. -verbunds vor • geringer Investitionsaufwand (Verbindungsaus-kunft in der Regel schon vorhanden) 	<ul style="list-style-type: none"> • keine Möglichkeit zur Trennung relevanter (Daten, die zu einer Fahrt geführt haben) von nicht relevanten Daten (Daten die zu keiner Fahrt geführt haben) • ggf. Kooperation mit Dritten notwendig (Betreiber der Verbindungsaus-kunft) • keine Daten zur genutzten Route bzw. genutztem Fahrzeug 	<p>Verbindungsaus-kunftsdaten bieten gegenüber weiteren Datenquellen den Vorteil, dass Informationen über die Quelle und das Ziel einer ggf. geplanten Fahrgastfahrt übermittelt werden. Gleichzeitig gehen entsprechend der Überlegungen des Autors eine Vielzahl an Daten ein, die nicht zu einer Fahrgastfahrt führen, bzw. mehrfach abgefragt werden. Konkrete Informationen zur genutzten Route bzw. Umstiege liegen somit nicht vor. Auch die Ableitung der Belegung von Fahrzeugen ist nur durch weitere Aufbereitungsschritte und ggf. nur für einen Teil der Linien möglich.</p>

Nachfrage- datenquelle	Vorteile	Nachteile	Nutzbarkeit für Anwen- dungen im ÖPNV
Vertriebsda- ten: Klassi- scher Ver- trieb	<ul style="list-style-type: none"> Daten besitzen ggf. einen Quelle-Ziel-Bezug (Ein- und Ausstiegshaltestelle) bzw. Informationen zu der Tarifzone der Fahrgastfahrt (nur bei Fahrausweisen des Gelegenheitsstarrs) 	<ul style="list-style-type: none"> Quelle-Ziel-Bezug bzw. Anzahl an Fahrgastfahrten für Zeitkartenkunden nicht gegeben (dauerhafte Fahrtberechtigung in den gewählten Tarifzonen) 	Vertriebsdaten ermöglichen es, Informationen zur Quelle und Ziel von Fahrgastfahrten einer Teilgruppe der Fahrgäste abzuleiten. Darüber hinaus besteht die Möglichkeit, Informationen zu den Kunden - insbesondere bei den elektronischen Tickets – zu nutzen. CICO bzw. BIBO bieten das Potential, haltestellenscharfe Quelle-Ziel-Matrizen zu erfassen. Es ist zu berücksichtigen, dass bis zur flächendeckenden und verpflichtenden Einführung eine Hochrechnung notwendig ist.
Vertriebsda- ten: elektro- nisches Fahrgeld- manage- ment (EFM)	<ul style="list-style-type: none"> CICO / BIBO enthält Informationen zu Quelle und Ziel der Fahrt hohe Genauigkeit durch Anforderungen an Preisermittlung stets aktuelle Daten über lange Zeiträume durch dauerhafte Erfassung zusätzliche Informationen zum Kunden liegen vor (zur Abrechnung) 	<ul style="list-style-type: none"> hoher Investitionsaufwand (unbekannte) Verzerrungen der Stichprobe: Nur eine Teilmenge der Grundgesamtheit wird erfasst. Es sind zusätzliche Informationen zur Stichprobe und Grundgesamtheit notwendig, um eine Hochrechnung durchzuführen 	

2.3 (Netzwerk-)Technische Grundlagen

Die Kommunikation zwischen technischen Endgeräten, wie beispielsweise Computern und Smartphones, erfolgt durch den Austausch von Informationen in Form binärer Daten. Grundlage für den Signalaustausch sind dabei analoge Signale.

2.3.1 ISO/OSI-Modell

Das OSI-Referenzmodell (Open Systems Interconnection) wurde 1984 von der ISO (International Standards Organisation) als Standard eingeführt (ISO 7498:1984). Die Standardisierung durch das OSI-Referenzmodell verfolgt das Ziel, einen allgemein verbindlichen Rahmen für den elektronischen Datenaustausch zu schaffen. Durch die Standardisierung in Form von Verfahren und Regeln soll die Netzwerkkommunikation zwischen verschiedenen Endgeräten verschiedener Hersteller sichergestellt werden. Durch den Standard wird demzufolge verhindert, dass einzelne Hersteller eigene Realisierungen des Datentransfers entwickeln, die lediglich mit Geräten des spezifischen Herstellers untereinander, nicht aber (offen) mit Geräten anderer Hersteller kommunizieren können (Badertscher et al. 2006).

Mit dem OSI-Referenzmodell wurde die Kommunikation in sieben Schichten (engl. „Layer“) aufgeteilt. In jeder Schicht werden die Regeln für die Datenverarbeitung durch Protokolle definiert, sodass die Signale von allen Geräten richtig „verstanden“ werden. Die Struktur des OSI-Modells ist hierarchisch

aufgebaut, sodass jede Schicht genau definierte Aufgaben ausführt und die Daten über Schnittstellen in die jeweils darüber- oder darunterliegende Schicht übermittelt. Eine Kurzbeschreibung der Funktion sowie Beispielsprotokolle der Schichten sind in Tabelle 2-2 dargestellt. Jede Schicht ist dabei in sich geschlossen, sodass einzelne Protokolle ersetzt werden können, ohne den Kommunikationsprozess nachhaltig zu beeinflussen.

Tabelle 2-2: Schichten des OSI-Referenzmodells (Baun 2018; der-wirtschaftsingenieur.de 2010; Heinemann Verlag GmbH; AIRNET 2015; Badertscher et al. 2006)

Schicht	Funktion	Protokolle (Beispiel)
Anwendungsschicht	Schnittstelle zwischen Mensch und Computer-Netzwerk (bspw. Browser oder E-Mail-Client), dient der Datenein- und -ausgabe.	HTTP, SMTP
Darstellungsschicht	Einheitliche Interpretation bzw. Kodierung der Daten in eine systemunabhängige Form (Zahlendarstellung, Zeichensatz etc.).	ASCII, HTML
Sitzungsschicht	Organisation (Aufbau, Aufrechterhaltung und Abbau der Kommunikation) und Synchronisation des Datenaustauschs. Darüber hinaus werden Kontrollpunkte eingebaut, sodass bei einem Verbindungsabbruch die Übertragung ab dem letzten Kontrollpunkt aufsetzen kann.	Telnet*, FTP* (können auch der Anwendungsschicht zugeordnet werden)
Transportschicht	Bereitstellung und Aufteilung der Daten in Segmente. Kontrolle der Datenübertragung.	TCP, UDP
Vermittlungsschicht	Bestimmung von Paketeleitwegen vom Ursprungs- zum Bestimmungsort (Routing) sowie die Bestimmung eines optimalen Weges durch ein oder mehrere Netzwerke (Routing zwischen Netzwerkknoten).	IP, IPv6
Sicherungsschicht	Aufbereitung der Daten für die Übertragung über das physikalische Medium zu Datenframes und physikalische Adressierung der Netzwerkkarten (MAC-Adresse). Fehlererkennung u. a. über Prüfsummen.	MAC
Bitübertragungsschicht	Umsetzung der Daten in (digitale) Signale und Realisierung der Signalübertragung. Die Protokolle definieren u. a. die Übertragungsgeschwindigkeit.	Ethernet, WLAN, Bluetooth

2.3.2 Grundlagen WLAN

2.3.2.1 IEEE 802.11

Der IEEE 802.11 Standard standardisiert und regelt die Kommunikation innerhalb von Funknetzwerken wie WLAN. Der Standard gehört zu der Normenfamilie der IEEE 802 Standards, die durch das Local Network Standards Committee (LMSC) entworfen wurde. Während die Arbeit des LMSC im Jahr 1980 noch auf die Entwicklung von Standards für kabelgebundene Netzwerke (LAN) abzielte, wurde das Arbeitsspektrum in den folgenden Jahren u. a. um LAN-Sicherheit, Metropolitan Area Networks (MAN) und höhere Datenraten ergänzt (IEEE 802 LAN/MAN Standards Committee 2004). Ein MAN

(Metropolitan Area Network) bezeichnet ein Netzwerk, bei dem mehrere Netzwerke in einer Stadt oder Region zu einem größeren Gesamt-Netzwerk für ebendiese Stadt oder Region zusammengeschlossen werden und deckt damit einen größeren räumlichen Bereich als lokale Netzwerke ab (Rouse 2005).

Die erste Version des IEEE 802.11 wurde 1997 verabschiedet. Dieser wurde durch die Standards 802.11a, b, g, h, n, ac weiterentwickelt und die Übertragungsraten und -technik für die Frequenzbereiche 2,4 GHz und 5 GHz optimiert. Mit dem IEEE 802.11ad wurde darüber hinaus ein Standard für die Bitübertragung im 60 GHz Frequenzspektrum definiert. Die Standards unterscheiden sich i. W. hinsichtlich der Datenübertragungsrate, der verwendeten Frequenzblöcke und den Kanalbreiten (Welotec GmbH 2017; Baun 2018).

Neben dem IEEE 802.11 Standard, der oft als Synonym für WLAN genutzt wird, definiert die Normenfamilie IEEE 802 mit dem IEEE 802.3 Standard auch die Kommunikation in kabelgebundenem LAN (Ethernet) sowie mit dem IEEE 802.15 Standard die Kommunikation über WPAN (Wireless Personal Area Network), zu denen auch Bluetooth zählt.

Im Rahmen dieser Dissertation wird mit WLAN stets ein Funknetzwerk nach dem IEEE 802.11 Standard bezeichnet.

2.3.2.2 Funktionsweise von WLAN-Netzen

Smartphones bieten die Möglichkeit, überall und jederzeit mit (fast) jeder Person kommunizieren bzw. Daten und Informationen austauschen zu können. Folge war und ist die rasante Verbreitung mobiler Endgeräte in der Bevölkerung (i. W. Smartphone, Tablet, Laptop) (Bitkom 2018). Darüber hinaus besteht in Deutschland ein nahezu ubiquitärer Zugang zu Funknetzwerken – ob Zuhause, bei der Arbeitsstelle oder während der Freizeit (Sportverein, Gastronomie, Bahnreise, öffentliches WLAN in Städten) – und damit häufig einhergehend auch der Internet-Zugang.

Ein WLAN bezeichnet umgangssprachlich drahtlose lokale Funknetzwerke, die i. d. R. auf den Standards der Normenfamilie der IEEE 802 Standards basieren und somit auch als Wi-Fi bezeichnet werden. Die Funktechnologie überträgt die Daten dabei mithilfe elektromagnetischer Wellen. Dies setzt Antennen bei allen Teilnehmern (Sender und Empfänger) voraus (Badertscher et al. 2006).

In Deutschland sind die nutzbaren Frequenzbereiche für WLAN in Allgemeinzuteilungen der Bundesnetzagentur geregelt und liegen im Bereich zwischen

2,4000 bis 2,4835 GHz (2,4 GHz) und 5,150 bis 5,725 GHz (5GHz) (Bundesnetzagentur 2020). Die nutzbaren Frequenzbereiche in anderen Ländern können davon abweichen. Ebenso differieren die erlaubten Sendeleistungen, wodurch die Reichweite beeinflusst wird (Elektronik-Kompendium.de 2019c).

Aufgrund der unterschiedlichen Wellenlänge im 2,4 GHz bzw. 5 GHz Frequenzblock ergeben sich spezifische Vor- und Nachteile: Ein Funknetz im 5 GHz Bereich kann bei gleicher Sendeleistung zwar mehr Daten übertragen, wird jedoch aufgrund der kürzeren Wellenlänge stärker durch Hindernisse beeinflusst als ein Netz im Bereich von 2,4 GHz. Störungen bei der Ausbreitung elektromagnetischer Wellen können durch Brechung, Beugung oder Absorption sowie Interferenzen auftreten. Bei Hindernissen (Signaldämpfung) ist dabei v.a. das Material entscheidend – Decken und Wände aus Stahlbeton können elektromagnetische Strahlung abschirmen (Faraday'scher Käfig) und reflektieren. Bei der Betrachtung von elektromagnetischen Wellen, die durch Smartphones versendet werden, sind insbesondere Interferenzen zwischen Quellen, wie beispielsweise mehreren WLAN-Netzen oder auch Bluetooth zu nennen, da beide Netzwerktechnologien die gleichen Frequenzbänder nutzen. Interferenzen treten darüber hinaus in Form von elektromagnetischem Rauschen auf, die durch Überlagerung von elektromagnetischer Strahlung verschiedener Strahlungsquellen (Motor, Mikrowelle, Haushaltsgeräte etc.) entstehen (Baun 2018; Mietke 2019; Schreiner 2019; Dionicio 2016).

Das 2,4 Ghz Frequenzspektrum liegt im ISM-Band (Industrial, Scientific and Medical Band). Mit dem ISM-Band werden Frequenzbereiche bezeichnet, die international für Hochfrequenzanwendungen von Industrie, Wissenschaft, Medizin sowie in häuslichen Bereichen ausgewiesen sind und lizenz- und genehmigungsfrei genutzt werden dürfen. Voraussetzung ist jedoch, dass die entsprechenden Geräte für den Betrieb in der Bundesrepublik Deutschland gekennzeichnet sind. Daher wird es auch von Funktechniken, wie u. a. WLAN und Bluetooth verwendet (Bundesnetzagentur 2015).

Zur Sicherstellung einer fehlerfreien Übertragung und zur Vermeidung von Interferenzen beim Datenaustausch durch mehrere Netzwerke und Access Points in räumlicher Nähe nutzen die Access Points verschiedene Kanäle innerhalb der Frequenzspektren: Im 2,4 GHz-Frequenzspektrum stehen grundsätzlich 14 Kanäle mit einem Abstand von 5 MHz zueinander zu Verfügung. Dabei ist zu berücksichtigen, dass in Europa lediglich die Kanäle 1 bis 13 genutzt werden dürfen. Die Trägerfrequenzen in Deutschland liegen im Bereich von 2412 MHz (Kanal 1) bis 2472 MHz (Kanal 13) (Schreiner 2019).

Bei einer benötigten Kanalbreite von 20 MHz (802.11g, 802.11n) ergeben sich vier überlappungsfreie Kanäle 1, 5, 9, 13 (Abbildung 2-1). Somit können mit den aktuellen Standards bis zu vier Netzwerke im gleichen räumlichen Bereich funken, ohne sich gegenseitig zu stören. Im älteren Standard 802.11b war ein Abstand zwischen den Trägerfrequenzen von 22 MHz gefordert, sodass sich hier nur drei überlappungsfreie Kanäle (1, 6, 11 bzw. 1, 7, 13) ergeben. Die Kanalaufteilung auf drei Kanäle wird auch heute (Stand 12/2019) noch verwendet, um die Kompatibilität zu Geräten zu gewährleisten, die noch den IEEE 802.11b Standard nutzen (Somolinos 2009; Gessler und Krause 2015; Schreiner 2019).

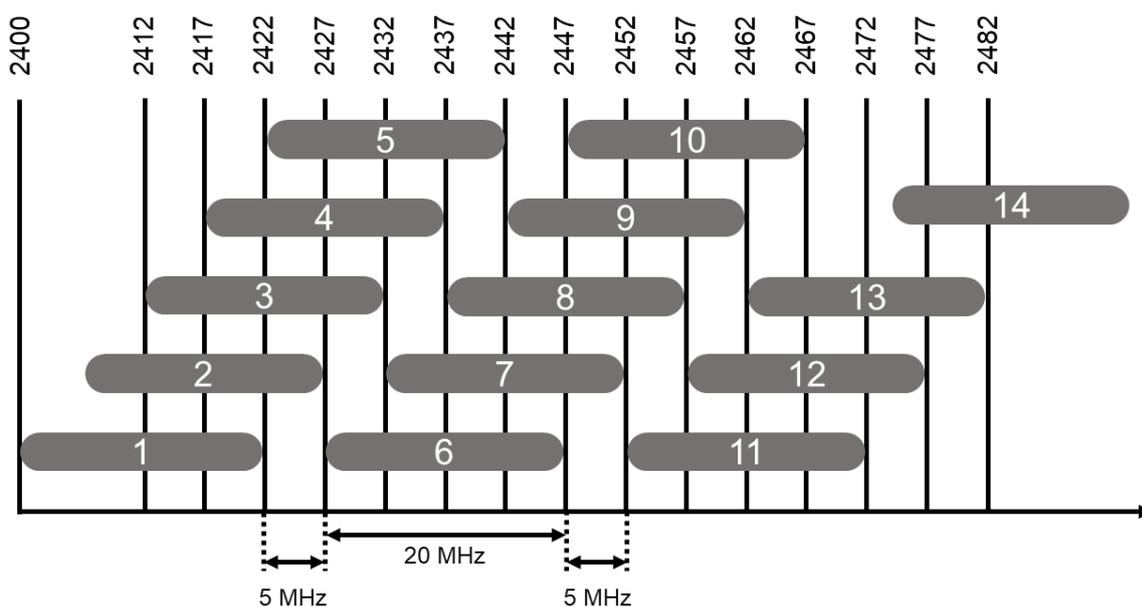


Abbildung 2-1: Kanäle im 2,4 GHz Frequenzspektrum (OFDM) (Digi-Key 2018)

Das 5 GHz Frequenzspektrum wurde 1999 erstmals mit dem IEEE 802.11a für die Nutzung von WLAN erschlossen, um eine Alternative zum stark ausgelasteten 2,4 GHz Frequenzspektrum zu bieten. Die Frequenzbereiche im 5 GHz Frequenzspektrum sind weltweit nicht einheitlich geregelt. In Deutschland dient es der Nutzung durch die Allgemeinheit zur breitbandigen Datenübertragung (WLAN) (Bundesnetzagentur 2018). Dazu sind die Bereiche 5,150 GHz – 5,350 GHz sowie 5,470 – 5,725 GHz freigegeben. Aufgrund der kürzeren Wellenlänge sind hier höhere Übertragungsraten im Vergleich zum 2,4 GHz Frequenzspektrum möglich. Mit der kürzeren Wellenlänge gehen jedoch die o. g. Nachteile (v. a. geringere Sendereichweite) einher (Schreiner 2019; Elektronik-Kompendium.de 2019c).

2.3.2.3 Probe Request: Definition und Aufbau

Die Kommunikation zwischen (mobilen) Endgeräten (Smartphone, Tablet, Laptop) in einem WLAN findet in der Regel über einen oder mehrere Access Points, wie beispielsweise Router, statt. Damit ein Endgerät Daten über einen Access Point senden oder empfangen kann, muss es sich entsprechend dem IEEE 802.11 zunächst authentifizieren und anschließend dem Access Point zugeordnet bzw. assoziiert werden (Cisco Systems 2018).

Vor dem Verbindungsaufbau zwischen einem Access Point und einem Endgerät muss das Endgerät zunächst über das Vorhandensein des Access Points informiert sein. Endgeräte haben zwei Möglichkeiten einen Access Points – und damit ein WLAN-Netzwerk – zu erkennen. Zum einen kann das Endgerät alle Kanäle der Frequenzspektren 2,4 GHz und 5 GHz nach Beacon Frames absuchen (passive service discovery). WLAN-Beacons sind Management-Pakete, mit denen Access Points u. a. Endgeräte in Reichweite über ihre Anwesenheit informieren und gleichzeitig Informationen zur Identifikation, wie etwa die SSID² oder den Verschlüsselungsmodus, übermitteln (Elektronik-Kompendium.de 2019b; Schreiner 2019). Bei dieser Methode kann es vorkommen, dass das Endgerät vergleichsweise lange Zeit die verschiedenen Kanäle „abhören“ muss, um einen Beacon Frame zu empfangen. Durch das dauerhafte und aktive Abhören wird entsprechend viel Energie verbraucht (Demir 2013).

Da dieses Vorgehen vergleichsweise ineffizient und langsam ist, können Endgeräte auch aktiv verfügbare Access Points in der Nähe anfragen. Dazu versendet das Endgerät selbst sogenannte „Probe Requests Frames“ (kurz Probe Requests) auf allen Kanälen des jeweiligen Frequenzbereichs (2,4 GHz, 5 GHz) (Cisco Systems 2014; Waltari und Kangasharju 2016; Kyne 2018). Zum Auffinden von Netzwerken werden die Probe Requests durch die Endgeräte in sogenannten „Bursts“ auf den verschiedenen Kanälen versendet. Bursts zeichnen sich durch eine undefinierte Anzahl von Probe Requests aus, die in sehr kurzer Zeit – meist unter 500 ms – auf verschiedenen Kanälen versendet werden (Matte 2017).

Die Probe Requests werden an die Broadcast-MAC-Adresse (FF:FF:FF:FF:FF:FF, s. Kapitel 2.3.2.5) geschickt (undirected Probe Request), sodass alle Access Points in Reichweite diese empfangen (Sauter 2013). Daneben fragen sogenannte „directed Probe Requests“ direkt und

² SSID steht für „Service Set Identifier“ und bezeichnet den (frei wählbaren) Funknetzwerknamen eines WLANs.

ausschließlich bekannte (gespeicherte) Netzwerke anhand der SSID an. Diese Art von Probe Request ist notwendig, um eine Verbindung mit sogenannten „hidden networks“ (versteckte Netzwerke) herstellen zu können, da diese selbst keine Beacon-Management-Frames versenden und somit für unbekannte Geräte unsichtbar sind (Schreiner 2019). Daher besitzt jedes Endgerät eine Liste von bekannten Netzwerken: die preferred network list (PNL). Der Großteil der versendeten Probe Requests von mobilen Endgeräten wird an die Broadcast-Adresse geschickt (Waltari und Kangasharju 2016).

Innerhalb eines Probe Requests Frames (Management Frame) werden verschiedene Daten übertragen. Dazu gehören unter anderem:

- **Sequence Number:** Die Sequence Number (SN) bezeichnet einen Zähler, der mit jedem versendeten Probe Requests um den Wert 1 erhöht wird. Bei Erreichen eines Maximalwerts (4096) wird der Zähler zurückgesetzt. Die SN dient dazu, fehlerhafte oder fehlgeschlagene Übertragungen erneut zu senden bzw. zu empfangen und Duplikate zu filtern.
- **Information Elements:** Information Elements (IE) umfassen u. a. SSID, unterstützte Frequenzen, Hardwareeigenschaften sowie weitere Merkmale, die für den Verbindungsaufbau und Datenaustausch zwischen dem mobilen Endgerät und dem Access Point von Relevanz sind. Detaillierte Informationen zu den Information Elements werden in Gast (2021) gegeben.

Eine vollständige Liste des Aufbaus und der Informationen von Management Frames ist dem Standard der IEEE (IEEE Computer Society 2016) zu entnehmen.

Mit den in den Probe Requests übermittelten Informationen kann der Access Point überprüfen, ob ein Datenaustausch entsprechend den Spezifikationen des Endgeräts grundsätzlich möglich ist. Im weiteren Verbindungsaufbau sendet der Access Point – sofern eine Datenübertragung möglich ist – eine Probe Response, die wiederum Informationen zu seinen Spezifikationen sowie seine SSID enthält. Aus diesen Probe Responses kann das scannende Endgerät eine Liste mit verfügbaren Access Points (bzw. Netzwerken) in der Umgebung erstellen. Anschließend findet eine Authentifizierung (Authentication) und Assoziierung (Association) des mobilen Endgeräts an dem entsprechenden Access Point statt. Erst nachdem das Endgerät mit dem Access Point assoziiert ist, kann die Datenübertragung starten (Cisco Systems 2018).

2.3.2.4 Sendeverhalten von Probe Requests

Das Sendeverhalten von Probe Requests wird in dieser Arbeit durch die Anzahl an gesendeten Probe Requests in einem definierten Zeitintervall unter verschiedenen Rahmenbedingungen beschrieben. Es wurde bereits durch verschiedene Wissenschaftler untersucht. Die Anzahl versendeter Probe Requests hängt demnach von folgenden Faktoren ab (Freudiger 2015; Rütermann et al. 2019):

- **Hersteller und Betriebssystem(-version):** Hersteller von mobilen Endgeräten greifen aktuell nahezu vollständig auf das Google-Betriebssystem Android (77%) oder das Apple-Betriebssystem iOS (23%) zurück (Stand 01/2020). Der Marktanteil der Betriebssysteme Windows, Symbian und Blackberry an der Internetnutzung mit Mobiltelefonen liegt in Summe bei 0,3% (StatCounter 2020). Zwischen den verschiedenen Betriebssystemversionen der einzelnen Hersteller bzw. Updates der Versionen konnten Unterschiede im Sendeverhalten der Probe Requests (Anzahl der Probe Requests, Zeitabstände zwischen Probe Requests) festgestellt werden (Freudiger 2015; Rütermann et al. 2019; Demir 2013).

Android hat seit der Veröffentlichung von Android Version 5 (11/2014) insgesamt 15 verschiedene Betriebssystem-Versionen umgesetzt (inklusive Update-Versionen, wie Version 5.1.2) (Medicus 2020). *iOS* hat seit der Veröffentlichung von iOS Version 12 (09/2018) bereits 27 verschiedene Betriebssystem-Versionen entwickelt (inkl. Update-Versionen) (Apple Inc. 2021a, 2021b). Sowohl bei Android als auch bei iOS-Geräten besitzen die aktuellen Versionen nach kurzer Zeit die größten Marktanteile (van Genabith 2019; Medicus 2020).

- **Anzahl bekannter Netzwerke:** Die Anzahl versendeter Probe Requests eines Endgeräts ist von der Zahl der gespeicherten Netzwerke abhängig. Grund dafür ist, dass versteckte Netzwerke – wie oben beschrieben – selbst keine Beacon-Frames aussenden und der Verbindungsaufbau somit von dem Endgerät selbst ausgeht. Je mehr Netzwerke auf einem Endgerät gespeichert sind, desto mehr Probe Requests werden von dem Endgerät versendet, um das Vorhandensein ebendieser Netzwerke in Reichweite zu prüfen (Freudiger 2015).

- **WLAN-Verbindungsstatus des Endgeräts:** Die Häufigkeit, in der das Endgerät Probe Requests aussendet, hängt auch vom Verbindungsstatus ab. Ein mobiles Endgerät kann
 - mit einem Access Point verbunden sein,
 - mit keinem Access Point verbunden sein und sich nicht in der Nähe von einem oder mehreren Access Points befinden oder
 - mit keinem Access Point verbunden sein und sich in der Nähe von einem oder mehreren Access Points befinden.

Ist ein Endgerät mit einem Access Point verbunden – d. h. besteht eine aktive Verbindung zu einem lokalen Netzwerk – nimmt die Anzahl an Probe Requests stark ab (Rütermann et al. 2019; Freudiger 2015). Die weiteren Probe Requests dienen in diesem Fall vor allem dem WLAN-Roaming: Wird beispielsweise ein Netzwerk über mehrere Access Points betrieben, z. B. bei stadtwitem WLAN oder bei Einsatz von WLAN-Repeatern, wird so sichergestellt, dass stets die Verbindung zu dem Access Point des Netzwerks mit dem besten Empfang bzw. der stärksten Signalübertragung besteht (WLAN Roaming) (Elektronik-Kompendium.de 2021c).

- **WLAN-spezifische Einstellungen des Endgeräts:** Dem Nutzer stehen verschiedene Möglichkeiten zur Verfügung, die WLAN-Konnektivität seines Endgeräts zu beeinflussen. Am bekanntesten ist die Einstellung zum Ein- oder Ausschalten von WLAN-Verbindungen in den Einstellungen der Endgeräte. Das Abschalten des WLAN über den „WLAN-Schiebeschalter“ führt jedoch nicht dazu, dass das Endgerät keine Probe Requests mehr versendet. Android-Geräte verschicken weiterhin Funknachrichten, u. a. für Google Standortdienste, Apple-Geräte nutzen Funknachrichten für wichtige Prozesse, u. a. location-based services oder für andere Apple Komponenten (Apple Inc. 2020; Matte et al. 2018). Solange diese Prozesse nicht separat abgestellt werden, versendet das jeweilige Endgerät zumindest sporadisch Probe Requests. Daneben besteht die Möglichkeit, die Funknachrichten mittels „Flugmodus“ abzustellen. Auch hier konnte nachgewiesen werden, dass diese Einstellung nur wenig Effekt auf die Anzahl versendeter Probe Requests besitzt (Freudiger 2015).

- **Nutzung des Endgeräts:** Neben den WLAN-spezifischen Einstellungen hat die Nutzung des Endgeräts einen starken Einfluss auf die Anzahl versendeter Probe Requests. Die Komplexität der Nutzungsmöglichkeiten von Smartphones oder anderen mobilen Endgeräten können in dieser Arbeit nicht umfassend dargestellt werden. Zur Vereinfachung werden die drei folgenden typischen Nutzungszustände unterschieden:
 - Entsperrten des Geräts durch Anschalten des Bildschirms („Aufwecken“) führt zum sofortigen Versenden von Probe Requests (Rütermann et al. 2019; Freudiger 2015).
 - Bei Nichtnutzung des Endgeräts befindet es sich typischerweise im „Standby“-Modus. Trotz der fehlenden Nutzung werden sporadisch Probe Requests versendet (Freudiger 2015).
 - Das Starten von Apps kann dazu führen, dass Probe Requests versendet werden. Dies ist beispielhaft beim Start von Google Maps der Fall, sofern die Standortdienste aktiviert sind (Matte et al. 2018; Waltari und Kangasharju 2016). Dabei ist zu berücksichtigen, dass installierte Apps die Anzahl versendeter Probe Requests sowohl bei aktiver Nutzung als auch im Standby beeinflussen können. Einen Spezialfall stellt das Öffnen des WLAN-Menüs und die Darstellung der umliegenden WLAN-Netzwerke dar. Das Aufrufen dieses Menüpunkts führt zu sofortigem Versenden von Probe Requests (Waltari und Kangasharju 2016).
- **Energiemodus bzw. Batteriezustand des Endgeräts:** Mobile Endgeräte verfügen in Abhängigkeit der Ladung des Akkus über einen Energiesparmodus. Im Energiesparmodus werden meist neben der Reduzierung der Displayhelligkeit auch die Funkprozesse stark verringert, um die Laufleistung des Endgeräts vor dem automatischen Abschalten zu erhöhen. Die Anzahl an Probe Requests wird daher verringert. Ist das Endgerät hingegen vollgeladen bzw. befindet es sich an einer Ladestation gibt es keine derartigen Einschränkungen (Rütermann et al. 2019; Demir 2013).

Auf Basis dieser Erkenntnisse kann festgehalten werden, dass eine starke Abhängigkeit zwischen der Anzahl gesendeter Probe Requests und den Rahmenbedingungen – insbesondere der Nutzung des Smartphones und des WLAN-Verbindungsstatus – besteht. Diese Erkenntnisse stützen auch die

Ergebnisse der von Paradedda et al. (2019) durchgeführten Testläufe, der in drei Szenarien die Anwesenheit einer Gruppe von Smartphones in einem Raum untersucht hat. Im Ergebnis kommen die Autoren zu dem Schluss, dass nicht garantiert werden kann, alle bzw. jedes einzelne Smartphone zu erfassen. Unter Berücksichtigung der weiteren beeinflussenden Faktoren kann darüber hinaus festgehalten werden, dass keine Sendemuster einzelner Endgeräte ableitbar sind.

Bei der Interpretation ist zu berücksichtigen, dass die Erkenntnisse der Autoren auf der Aufzeichnung von Probe Requests durch einen WLAN-Adapter basieren. Da ein WLAN-Adapter nur die Funknachrichten jeweils eines Kanals zur gleichen Zeit erfassen kann, können Funknachrichten auf allen weiteren Kanälen im 2,4 GHz- und im gesamten 5 GHz-Frequenzbereich nicht in die Analyse eingehen.

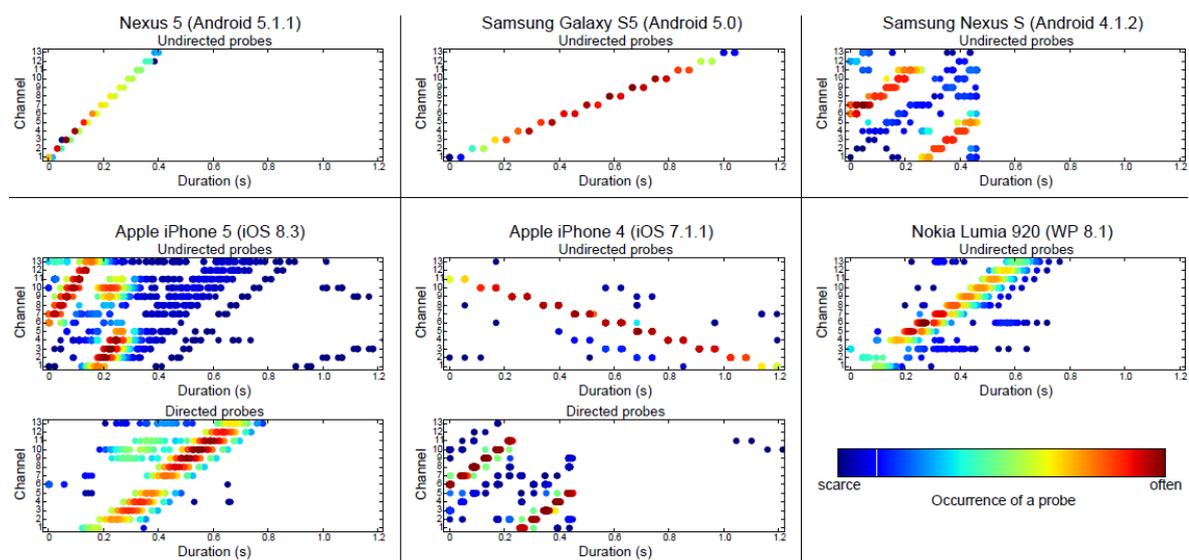


Abbildung 2-2: Kanalwechsellmuster gesendeter Probe Request Bursts von ausgewählten Endgeräten (Waltari und Kangasharju 2016)

Waltari und Kangasharju (2016) haben die Wirkung fehlender Adapter, d. h., wenn die Anzahl der Adapter kleiner ist als die Anzahl zu erfassender Kanäle, untersucht. In ihrem Versuch wurden Probe Requests unter kontrollierten Bedingungen versandt und über vier Stunden durch 13 WLAN-Adapter – ein Adapter für jeden Kanal im 2,4 GHz-Frequenzbereich – aufgezeichnet. Die Ergebnisse zeigen erwartungsgemäß, dass mit Verringern der Anzahl aufzeichnender Adapter ein annähernd linearer Rückgang der Erfassungsquote gegenüber der Vollerfassung verbunden ist. Bei Aufzeichnung mit nur einem

WLAN-Adapter³ werden daher nur rund 1/13 (rund 8 %) der auf allen Kanälen versendeten Probe Requests empfangen. Die Ergebnisse ihrer Untersuchung (Abbildung 2-2) zeigen auch, dass die untersuchten Smartphones nacheinander auf den verschiedenen Kanälen im 2,4-GHz-Spektrum Probe Requests aussenden.

Abbildung 2-2 verdeutlicht nochmals die Bedeutung des Herstellers und Betriebssystems für die Sendeeigenschaften von Probe Requests bei der aktiven Suche nach Netzwerken auf verschiedenen Kanälen. Auf Grundlage der Farbe sind darüber hinaus Unterschiede in der Anzahl der versendeten Probe Requests auf den einzelnen Kanälen festzustellen.

2.3.2.5 MAC-Adresse und MAC-Randomisierung

MAC-Adresse

Die Kommunikation zwischen Endgeräten in verschiedenen Netzwerken wird durch die entsprechende Hardware (Netzwerkkarte, WLAN-Chip) ermöglicht. Für einen zielgerichteten Austausch von Daten oder Informationen ist es notwendig, dass die verbaute Hardware in jedem Gerät eindeutig identifiziert und adressiert werden kann. Dazu erhält jede Netzwerkkarte bzw. Chip eine weltweit eindeutige, durch den Hersteller vergebene MAC-Adresse (Media Access Control Address) (Badertscher et al. 2006; Baun 2018).

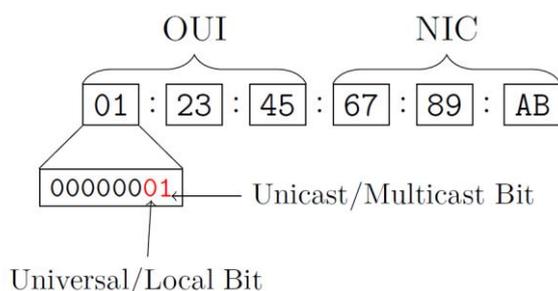


Abbildung 2-3: Aufbau einer MAC-Adresse (Martin et al. 2017)

Eine MAC-Adresse besteht aus sechs Bytes im Hexadezimalsystem (bzw. 48 Bits im Binärsystem), die mit Doppelpunkten oder Bindestrichen voneinander getrennt geschrieben werden. Die Zeichenfolge „AB:CD:EF:12:34:56“ ist ein Beispiel für eine MAC-Adresse. Insgesamt stehen also 2^{48} MAC-Adressen (rund 281,5 Bio.) zur Verfügung.

Die ersten drei Bytes der MAC-Adresse werden als Organisationally Unique Identifier (OUI) bezeichnet und bilden die Herstellerkennung. Die

³ Der Adapter hat die Daten mit einem Channel Hopping Algorithmus erfasst.

Zeichenfolge des OUI wird gegen eine Gebühr von dem Institute of Electrical and Electronics Engineers (IEEE)⁴ an die verschiedenen Hardware-Hersteller vergeben. Eine Liste der Zuordnung von OUI zu Herstellern ist in Institute of Electrical and Electronics Engineers (IEEE) (2019) gegeben. Daneben werden die ersten drei Bytes auch für die Company-ID (CID) genutzt. Auch die CID wird einmalig an Unternehmen (beispielsweise Google Inc.) vergeben, so dass keine Doppelungen mit der OUI auftreten können. Die verbleibenden drei Bytes werden als Organisationally Unique Address (OUA) bezeichnet und vom Hersteller festgelegt. Der Hersteller ist dafür verantwortlich, dass die OUA-Bitfolge nur einmal vergeben wird (Elektronik-Kompodium.de 2019a).

Den ersten beiden übertragenen Bits der MAC-Adresse – die Übertragung beginnt bei dem niedrigwertigsten Bit eines Oktetts – kommt eine besondere Bedeutung zu (Abbildung 2-3): Ist das erste Bit gesetzt (=1), handelt es sich um eine Multi-Cast-Adresse, d. h. es wird eine Gruppe von Endgeräten angesprochen. Ein Beispiel dafür ist die Broadcast-MAC-Adresse (FF:FF:FF:FF:FF:FF). An diese Adresse gesendete Pakete werden von allen Geräten im Netzwerk verarbeitet. Das zweite Bit („lokaler Bit“) gibt an, ob es sich um eine universelle, d. h. weltweit eindeutige MAC-Adresse, handelt (=0), oder ob die MAC Adresse lokal durch Software verändert wurde (=1). Wird eine CID verwendet, ist der lokale Bit grundsätzlich gesetzt (Elektronik-Kompodium.de 2019a; Badertscher et al. 2006; IEEE Computer Society 2016). Als Beispiel dafür kann der CID der Google Inc. genutzt werden: „DA:A1:19“ (Martin et al. 2017). Wird die Hexadezimalzahl „DA“ in das binäre Zahlensystem umgerechnet, ergibt dies den Wert 11011010. Das zweite Bit ist unterstrichen und gibt an, dass es sich um eine lokal veränderte MAC-Adresse handelt.

Im Kommunikationsprozess zwischen Endgeräten wird die physische Adresse der jeweiligen Netzwerkschnittstellen (Netzwerkkarten, WLAN-Chip etc.) des Senders und Empfängers durch die Protokolle der Sicherungsschicht (Layer 2 des OSI-Modells) im Datenpaket (Frame) ergänzt. Bei Übereinstimmung der Zieladresse mit der eigenen physikalischen Adresse des empfangenden Endgeräts werden die Daten an die höheren Schichten weitergegeben (Baun 2018).

⁴ Das IEEE entwickelt unter anderem Standards für Techniken, Hardware und Software.

MAC-Randomisierung

Um eine WLAN-Verbindung einzugehen, benötigt das Smartphone Informationen über die sich in Reichweite befindenden, verfügbaren WLAN-Access Points. Dazu werden Probe Requests ohne das Zutun des Anwenders ausgesendet, bei denen verschiedene Informationen – unter anderem die MAC-Adresse – versendet werden. Durch Erfassung dieser Probe Requests an verschiedenen Orten ist es grundsätzlich möglich, einzelne MAC-Adressen und damit auch die Nutzer zeitlich und räumlich zu verfolgen (passives Tracking). Das Tracking kann dabei auf Ladengeschäfte begrenzt sein (insoft GmbH 2019), aber auch ganze Stadtgebiete und größer skalierte Regionen umfassen (Basalamah 2016). Darüber hinaus können bekannte SSID des Endgeräts in den Probe Request angefragt und übertragen werden. Ein Verschneiden der bekannten SSID mit Standortkarten von WLAN-Netzwerken (WiGLE.net 2019; Cunche et al. 2014) ermöglicht auch hier Rückschlüsse über die Aufenthaltsorte des jeweiligen Besitzers.

Um die Privatsphäre der Nutzer zu schützen, wird daher zunehmend eine MAC-Randomisierung von Herstellern implementiert (Cunche und Matte 2016). Das Betriebssystem iOS verfügt über die Randomisierung ab Version 8 (Skinner und Novak 2015), Android ab Version 6 (sofern die verbaute Hardware dies unterstützt) (Android Developers 2021), Windows ab Version 10 (sofern die verbaute Hardware dies unterstützt) (Tkachenko 2016). Mit der MAC-Randomisierung wird bei jedem Probe Request Burst eine neue, lokal erzeugte MAC-Adresse verwendet. Bei diesem von Gruteser und Grunwald entwickelten Verfahren werden die randomisierten MAC-Adressen als „Wegwerf-Identifizier“ (disposable interface identifiers) bezeichnet (Gruteser und Grunwald 2005) und dienen dazu, das passive Tracking zu verhindern. Durch die MAC-Randomisierung ist es dem Beobachter nicht möglich, die Signale einzelner Geräte anhand der MAC-Adresse zu verschiedenen Zeitpunkten bzw. an verschiedenen Orten zu identifizieren bzw. von den Signalen anderer Geräte zu differenzieren. Aus der Beziehung von Endgerät zu MAC-Adresse im Verhältnis 1:1 wird mittels Randomisierung ein Verhältnis von 1:N, da jedes Endgerät unendlich viele MAC-Adressen erzeugen kann. Somit werden Verfahren, die auf diesem Merkmal basieren, stark eingeschränkt bzw. unmöglich (Martin et al. 2017; Rütermann et al. 2019). Gleichzeitig berücksichtigt die MAC-Randomisierung zunächst nur die MAC-Adresse des Endgeräts. Das heißt, dass die weiteren übertragenen Daten erhalten bleiben, jedoch kein separates Identifizierungsmerkmal besitzen. Neben der MAC-Adresse

werden i. W. Daten zu den Hardware-Eigenschaften des Endgeräts übertragen – etwa unterstützte Modulationsverfahren, Frequenzen und Übertragungsraten – die für die Kommunikation zwischen dem Endgerät und dem Access Point relevant sind. Diese Daten können ggf. in Kombination miteinander als neue Identifier eingesetzt werden. Endgeräte mit gleichen Eigenschaften sind damit jedoch nicht zu unterscheiden. Eine Übersicht der in Probe Requests übertragenen Informationen gibt O'Reilly Online Learning, (2022). Für die MAC-Randomisierung gibt es keine Vorgaben, sodass die Verfahren nach Hersteller, Betriebssystem und Smartphone-Modell variieren (Vanhoef et al. 2016; Martin et al. 2017).

Die MAC-Randomisierung ist bei aktuellen Smartphones standardmäßig aktiviert, wenn nach Netzwerken gesucht wird. Sobald jedoch eine Verbindung mit einem Access Point besteht, wurde die wahre MAC-Adresse genutzt. Mit der Einführung von Android 11 wird es den Nutzern ermöglicht, die MAC-Adresse von Probe Requests auch bei bestehender Verbindung zu einem Access Point zu randomisieren, sodass hier nicht mehr die wahre MAC-Adresse angezeigt wird (Wallen 2021). iOS generiert für jede WLAN-Verbindung neue randomisierte MAC-Adressen, behält aber die MAC-Adresse bei Verbindung zu einem Netzwerk bei (Stand 02/2021) (Apple Inc. 2022). An der Stelle ist zu berücksichtigen, dass die Maßnahmen zum Datenschutz und zum Schutz der Privatsphäre sich schnell weiterentwickeln.

Gleichzeitig reduziert sich die Frequenz der Probe Requests jedoch stark. Dabei ist zu berücksichtigen, dass Nutzer in der Regel keine Verbindung zu einem Access Point besitzen, wenn diese mobil sind (Ausnahmen: Stadtweites WLAN, WLAN im ÖPNV). Bereits heute besteht die Möglichkeit, die Einstellungen so zu verändern, dass auch bei Verbindung mit einem Access Point die MAC-Randomisierung bei Probe Requests weiter durchgeführt wird. Dabei ist jedoch zu beachten, dass die randomisierte MAC-Adresse in dem Fall für die Dauer der Verbindung mit dem Access Point gleichbleibt. Mit der Veröffentlichung und Verbreitung von Android 10 soll die MAC-Randomisierung auch bei Verbindung mit einem Access Point standardmäßig durchgeführt werden (Android Source 2020).

Da die Einführung der MAC-Randomisierung einen sehr starken Einfluss auf die Datenqualität von Trackingverfahren hat, kommt Verfahren, die der Randomisierung entgegenwirken, eine hohe Bedeutung zu. In der Forschung gab es erste Versuche, die MAC-Randomisierung zu umgehen. Zur Zuordnung

von Probe Requests (PR) mit randomisierten MAC-Adressen zu einzelnen Endgeräten (Derandomisierung) kam es u. a. zur Operationalisierung

- der in den Probe Requests mitgesendeten zusätzlichen Informationen (Informationen Elements, Sequence Number) (Freudiger 2015; Vanhoef et al. 2016; Nitti et al. 2020),
- der Reihenfolge der Kanäle, auf denen die Probe Requests versendet werden (Waltari und Kangasharju 2016),
- des zeitlichen Abstands zwischen Probe Request Bursts (Loh et al. 2008; Matte et al. 2016),
- radiometrischer (=elektromagnetische Strahlung) Merkmale der emittierten Signale (Brik et al. 2008) oder
- des Datenaustauschs von Apps im Mobilfunk (Stöber et al. 2013).

als mögliche Identifikatoren. Keines der Verfahren bietet jedoch die Möglichkeit, randomisierte MAC-Adressen unter Realbedingungen einzelnen Endgeräten fehlerfrei zuzuordnen.

2.3.3 Grundlagen Bluetooth

2.3.3.1 Bluetooth-Versionen

Bluetooth bezeichnet eine standardisierte (IEEE 802.15) Funktechnik zur kabellosen Übertragung von Daten zwischen unterschiedlichen Geräten (Peer-to-peer oder in Piconetzen⁵). Bluetooth wurde 1994 durch die Mobilfunkhersteller Ericsson und Nokia entwickelt, um Geräte über kurze Distanzen verbinden zu können. Im Jahr 1998 wurde die Interessengemeinschaft Bluetooth Special Interest Group (SIG) gegründet, der inzwischen mehr als 30.000 Unternehmen angehören, darunter Smartphone- und Chiphersteller (u. a. Ericsson, Nokia, Microsoft, Intel, Apple). Derzeit aktuell ist die fünfte Version von Bluetooth (Bluetooth 5, Stand 01/2021). Seit der Weiterentwicklung von Bluetooth in der Version 4.0 im Jahr 2010 wurde der Fokus auf das Stromsparen für IoT-Anwendungen (Internet of Things), wie etwa Fitnessstracker mit begrenzter Akkukapazität, gesetzt (Hofmann 2020).

Bluetooth ist u. a. durch die Anwendung auf (mobilen) Endgeräten (PC, Smartphones) zur Übertragung von Sprache und Daten (kontinuierlicher Datenstrom) beispielsweise auf Kopfhörer oder Headsets bekannt. Dazu nutzt

⁵ Bei der Verbindung zwischen einem Master und mindestens einem Slave entsteht ein Piconetz – ein Ad-hoc-Netzwerk zur Kommunikation unter Bluetooth-Geräten.

Bluetooth – wie auch WLAN – im 2,4 GHz-Spektrum, das ISM-Band in den Bereichen zwischen 2,402 und 2,480 GHz, sodass es weltweit lizenzfrei betrieben werden darf. In dem Frequenzbereich stehen 79 Kanäle mit einer Breite von 1 MHz zur Verfügung. Zur Steigerung der Stabilität aufgrund möglicher Interferenzen durch andere Übertragungen (WLAN, Mikrowellen) nutzt Bluetooth ein Frequenzsprungverfahren (Frequency Hopping Spread Spectrum, FHSS), auch Frequenzhopping genannt, wobei die Kanäle zur Datenübertragung bis zu 1.600-mal pro Sekunde gewechselt werden. Der Ablauf des Frequenzwechsels ist dabei den verbundenen Geräten bekannt und gründet u. a. auf der Bluetooth-Device Address (MAC-Adresse der Bluetooth-Einheit) des Masters (Universität Rostock o.J.; Bluetooth SIG 2019).

Mit der Weiterentwicklung von Bluetooth seit der ersten Version (1.0) wurden u. a. die Übertragungsraten gesteigert und der Energieverbrauch minimiert. Während die Nutzung von Bluetooth bis zur Version 3.0 grundsätzlich mit einem hohen Stromverbrauch verbunden war, da der Fokus auf einer kontinuierlichen Verbindung mit der Möglichkeit zum Austausch großer Datenmengen lag, hat Bluetooth Low Energy (BLE) einen anderen Fokus: Geräte, die BLE nutzen, sind auf einen geringen Stromverbrauch optimiert. Daher sind sie in der Regel nur kurz aktiv und übertragen geringe Datenmengen bei einer geringen Sendeleistung (Elektronik-Kompendium.de 2021b). Bluetooth Low Energy und Bluetooth Version 5 werden insbesondere von gängigen Smartphones, Tablets und Wearables (Smartwatch, Fitnessband) verwendet (Elektronik-Kompendium.de 2021a; Hofmann 2020).

Die Reichweite von Bluetooth hängt neben der Sendeleistung von weiteren Parametern, wie beispielsweise Hindernissen auf dem Ausbreitungsweg, ab. Bluetooth unterscheidet drei Klassen (Tabelle 2-3), die sich hinsichtlich der Leistung und damit auch der Reichweite unterscheiden. Die Bluetooth-Klasse wird in Abhängigkeit der angestrebten Nutzung gewählt – Smartphones, Bluetooth-Kopfhörer oder MP3-Player verwenden in der Regel Klasse 3. Größere als in der Tabelle angegebene Distanzen zwischen Sender und Empfänger führen zum Verbindungsabbruch.

Tabelle 2-3: Reichweite und Leistung der Bluetooth-Klassen (Axel Springer SE 2021)

Klasse	Leistung (mW)	Reichweite allgemein
Klasse 1	100	ca. 100 m
Klasse 2	2,5	ca. 50 m
Klasse 3	1	ca. 1 m

Mit Bluetooth Version 5.1 wurden darüber hinaus insbesondere Funktionen zur Positionserfassung und Richtungserkennung erweitert, sodass ein Signal bis auf 10 Zentimeter genau geortet werden kann. Diese Funktion dient als Grundlage für die Navigation in Gebäuden oder auch für die Information zu bestimmten POIs in Museen oder anderen Standorten (Elektronik-Kompendium.de 2021a).

2.3.3.2 Bluetooth-Profile

Über sogenannte Bluetooth-Profile wird eine an die Erfordernisse der jeweiligen Geräte angepasste Auswahl an Bluetooth-Funktionen ermöglicht. Bluetooth Profile sind Schnittstellenspezifikationen und definieren Protokolle und Regeln, die Bluetooth-Geräten herstellerunabhängig eine Kommunikation auf Anwendungsebene erlauben. Durch ein oder mehrere Profile erhält das Gerät lediglich auf die notwendigen Funktionen Zugriff, womit letztlich die Effizienz und Laufzeit des Geräts erhöht wird. Headsets (u. a. Audiokanäle, Lautstärkeregelung) erhalten also andere Funktionen als Drucker. Eine Übersicht der standardisierten Bluetooth-Profile ist in Elektronik-Kompendium.de (2021a) gegeben.

2.3.3.3 Verbindungsaufbau

Für die Verwaltung von Verbindungen, deren korrekten Aufbau, Erhalt und Abbau werden verschiedene Bluetooth-Zustände genutzt. Ein Endgerät kann i. W. die Zustände „Standby“ oder „Connection“ aufweisen. Das Gerät befindet sich standardmäßig im Standby, in dem nur die Uhr (Clock) betrieben wird, sodass der Energieverbrauch gering ist. Der Standby-Zustand wird verlassen, um eine Verbindung einzugehen. Dazu sucht das Gerät nach Inquiry- oder Paging-Nachrichten bzw. versendet diese selbst (Bluetooth SIG 2019). Neben den beiden Hauptzuständen werden sieben Unterzustände unterschieden (Inquiry, Inquiry Scan, Inquiry Response, Page, Page Scan, Master Response und Slave Response). Im Folgenden wird auf die Zustände im Inquiry- und Paging-Prozess eingegangen.

Die bei einer Datenübertragung versendeten Pakete bestehen aus Access Code (Zugriffscod), PDU (Protocol Data Unit) Header (Paketkopf) sowie PDU Payload (Nutzlast). Der Access Code bestimmt dabei, ob

- die Daten an ein bekanntes Gerät übermittelt werden (DAC, Device Access Code),
- die Daten an alle Geräte innerhalb eines Piconetzes versendet werden (CAC, Channel Access Code) oder

- in der Nähe befindliche Geräte während des Inquiry Prozesses gesucht werden (IAC, Inquiry Access Code). Der IAC wird weiter in GIAC (General IAC) und DIAC (Dedicated IAC) unterschieden. GIAC wird genutzt, um alle Geräte anzusprechen, während DIAC nur Geräte mit bestimmten Charakteristiken anspricht (Name, Hands-free Devices) (Karlsson und Persson 2002).

Im Rahmen des Verbindungsaufbaus (Inquiry, Paging) werden sogenannte ID-Pakete (ID = Identifikator) übermittelt, die keinen Payload besitzen. Unter „Payload“ werden die eigentlichen Nutzdaten – etwa der zu übertragende Text – verstanden, die innerhalb eines Paketes übertragen werden. Der Payload beinhaltet keine Informationen zu dem Ziel des Pakets im Netzwerk.

Der erste Schritt zur Etablierung einer Verbindung zwischen bislang unbekanntem Geräten ist das sogenannte „Inquiring“ entsprechend den Vorgaben der Bluetooth-Spezifikationen (Bluetooth SIG 2019). Bestand im Vorfeld bereits eine Verbindung, kann der Inquiry-Prozess übersprungen werden, so dass die Verbindung zwischen den bekannten Geräten bei eingeschaltetem Bluetooth automatisch stattfindet. Die Bluetooth-Kommunikation wird über Piconetze organisiert, die aus einem Master und bis zu sieben Slaves bestehen (Bluetooth SIG 2019).

Im Verbindungsaufbau mit Bluetooth sind darüber hinaus drei Sicherheitsstufen zu unterscheiden (DATACOM Buchverlag GmbH 2021):

- **Stufe 1 (niedrige Sicherheit):** Geräte können sich gegenseitig erkennen und ohne Authentifizierung miteinander kommunizieren. Zum Erschweren des Abhörens findet Frequenz-Hopping statt.
- **Stufe 2 (anwendungsbezogene Sicherheit):** Die Sicherheitsmechanismen werden von der Anwendungsschicht (Layer 7) durchgesetzt. Dienste können hier für alle Geräte offen sein, eine Authentifizierung einfordern oder zusätzliche Sicherheitsmechanismen erfordern.
- **Stufe 3 (verbindungsbezogene Sicherheit):** Die Sicherheitsmechanismen werden auf der Verbindungsschicht durchgeführt. Diese Stufe ist die sicherste Stufe, Geräte können sich gegenseitig nur durch Freigabe erkennen und benötigen für eine Verbindung eine Authentifizierung.

Inquiry-Prozess

Zur vereinfachten Beschreibung des Inquiry Prozesses werden die beteiligten Geräte als Master und Slave bzw. Slaves beschrieben. Diese Rollen sind **vor**

einer Verbindungsaufnahme grundsätzlich noch nicht definiert. Das Gerät, das den Inquiry-Prozess anstößt und Inquiry-Nachrichten aussendet, wird als Master bezeichnet. Das Gerät, das nach Inquiry-Nachrichten scannt, wird als Slave bezeichnet (Bluetooth SIG 2019).

Der Inquiry-Prozess ist schematisch in Abbildung 2-4 dargestellt. Er dient dem Master dazu, eine Liste von Bluetooth-Geräten zu erstellen, die sich in räumlicher Nähe befinden. Dazu wechselt der Master vom Standby- oder Connection-Zustand in den Inquiry-Substate und versendet mehrfach auf festgelegten Kanälen (Frequenzen) Inquiry Requests aus, die im Paketkopf die notwendigen Daten zum weiteren Aufbau einer Verbindung enthalten. Dazu gehören der Clockcounter (CLK, Systemzeit) und die MAC-Adresse des Masters.

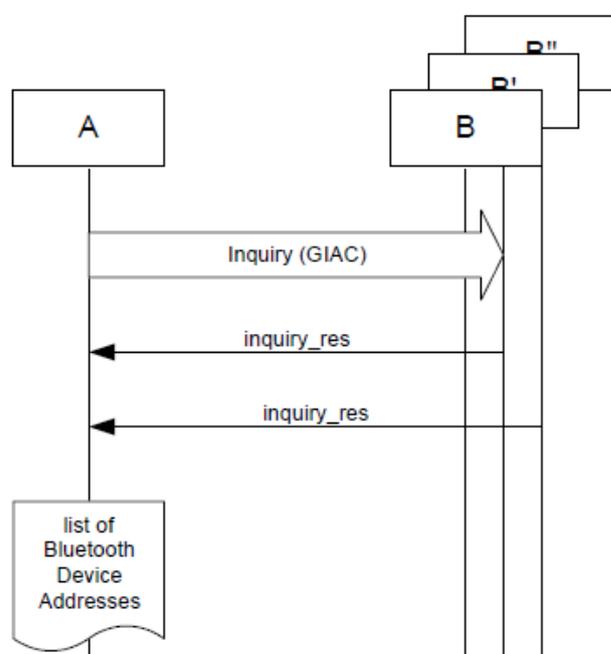


Abbildung 2-4: Schematische Darstellung des Inquiry-Prozess (Bluetooth SIG 2019, S. 1324)

Slaves, die gefunden werden wollen – z.B. durch Aktivieren der Suche nach Bluetooth-Geräten – wechseln in den Inquiry-Scan-Substate und scannen ebendiese Kanäle und Frequenzen nach Inquiry-Requests. Damit sich die Frequenzen von Master und Slave überschneiden, ist die Frequenzwechselrate im Scanning Modus kleiner als im Inquiring Modus. In der Abbildung befinden sich die Geräte B' und B'' im Inquiry-Scan-Substate. Das Gerät B ist nicht sichtbar.

Empfängt der Slave ein ID Paket eines Masters, wechselt er in den Inquiry Response Substate und versendet ein FHS-Paket (Frequency Hopping

Synchronisation), das u. a. seinem lokalen Namen sowie Geräte-Adresse, Clockwerte, unterstützte Funktionen und weitere Parameter beinhaltet. Das FHS-Paket dient der Synchronisation zwischen dem Master und dem Slave. Detaillierte Informationen dazu bieten die Bluetooth Core Spezifikationen (Bluetooth SIG 2019).

Auf Grundlage der durch den Master empfangenen Inquiry-Responses kann eine Liste mit Geräten erstellt werden, mit denen der Master sich im Rahmen des Paging Prozess verbinden kann. Nähere Informationen bei (Bluetooth SIG 2019), (Karlsson und Persson 2002) oder (Schmietendorf 2011).

Paging-Prozess

Zum Aufbau einer Verbindung zwischen Bluetooth-fähigen Geräten folgt nach dem Inquiry-Prozess der Page-Prozess mit dem Ziel Daten auszutauschen. Bei der Herstellung einer Verbindung über den Page-Prozess benötigt der Master nur die Bluetooth-Adresse des bzw. der Slaves, mit dem/denen eine Verbindung eingegangen werden soll. Alle weiteren Informationen liegen dem Master entweder aus dem Inquiry-Prozess und/oder aus einer vorhergegangenen Verbindung vor. Analog zum Inquiry-Prozess wird auch hier das Gerät als Master bezeichnet, das den Page-Vorgang initiiert (Bluetooth SIG 2019).

Der Page-Prozess erfolgt ähnlich wie der Inquiry-Prozess. Dazu wechselt der Master in den Page-Substate und versendet auf festgelegten Kanälen (Frequenzen) mehrere Paging-Nachrichten mit seiner Bluetooth-Adresse an die Slaves (DIAC). Dies ist für die Abstimmung der Hopping Frequenz zwischen Master und Slave notwendig.

Slaves, die bereits mit einem Master verbunden waren, wechseln eigenständig und wiederkehrend in den Page-Scan-Substate, um Paging-Nachrichten auf den festgelegten Kanälen von einem Master zu empfangen, der sich verbinden möchte. Sobald der Slave den Master erkannt hat, werden weitere Informationen zur Synchronisation (Channel Hopping Sequenz im Piconetz, Clock-Wert) ausgetauscht (Slave Page Response, Master Page Response) (Bluetooth SIG 2019; Schmietendorf 2011).

Zur Übertragung von Datenpaketen im Piconetz wird von beiden Devices derselbe Channel Access Code genutzt (Bluetooth SIG 2019).

Verbindung

Sind die Geräte miteinander verbunden, können diese durch Datenübertragung oder -empfang an der Kommunikation im Piconetz teilnehmen. Die aktiven Geräte im Piconetz erhalten durch den Master eine aus drei Bits bestehende AMA (Active Member Address), die die Größe eines Piconetzes auf maximal sieben Slaves festlegt (Binärsystem: $111 = 7$) (Schmietendorf 2011).

Aufgrund der Anwendung auf mobilen Endgeräten kommt der Energieeffizienz im Rahmen der Bluetooth-Kommunikation eine hohe Bedeutung zu. Zu diesem Zweck können einzelne Geräte in einen Energiesparzustand wechseln. Im Energiesparzustand bleiben die Geräte im Piconetz, nehmen jedoch nicht durchgehend an der Kommunikation teil. Der Energiesparzustand kann nach Bedarf aufgehoben werden (Schmietendorf 2011).

Zusätzliche Parameter in Piconetzen

Sowohl beim Aufbau als auch bei der Organisation bzw. Aufrechterhaltung von Verbindungen besteht die Möglichkeit, zusätzliche Informationen vom Endgeräten im bestehenden Piconetz oder während des Inquiry-Prozesses von Geräten in der Nähe einzuholen. Dazu zählen sogenannte „informational parameters“ und „status parameters“ (Bluetooth SIG 2019). Diese Parameter sind durch den Chip-Hersteller festgelegt und können nicht durch Software überschrieben werden. Abrufbar sind dabei unter anderem der Verbindungsstatus des Endgeräts, unterstützte Anwendungen, die MAC-Adresse, der Name des Geräts sowie die empfangene Signalstärke (RSSI).

Zusätzlich können Informationen zur sogenannten Bluetooth-Class ermittelt werden. Die Bluetooth-Class geht auf die Bluetooth-Profile zurück und gibt sowohl Informationen zu

- der Funktion des Geräts, wie etwa zur Positionierung / Navigation, Aufnahme (Scanner, Mikrofon), Wiedergabe (Drucker, Lautsprecher), Telefonie und
- dem Endgerättyp.

Der Endgerättyp wird dabei unterschieden in die sogenannte

- Major Device Class, wie etwa Computer, Telefon, Peripheriegerät (Maus, Tastatur), Wearable und
- Minor Device Class zur Spezifizierung der Major Class.

Durch die Minor Device Class kann entsprechend in der Major Class „Telefon“ ein Smartphone von anderen Telefon-Klassen unterschieden werden (Bluetooth SIG 2009).

2.3.3.4 MAC-Adresse und MAC-Randomisierung

Die Bluetooth MAC-Adresse besitzt den gleichen Aufbau wie die WLAN-MAC-Adresse (Kapitel 2.3.2.5). Die Randomisierung der MAC-Adresse ist aus den bereits genannten Gründen der Privatsphäre und zur Vermeidung eines Tracking seit Bluetooth Version 4.0 (Low Energy) möglich. Nach Issoufaly und Tournoux (2017) nutzt jedoch nur ein kleiner Teil der Geräte die Möglichkeit zur Randomisierung der MAC-Adresse.

Bei Einsatz der MAC-Randomisierung gibt es keine Vorgaben, sodass die Intervalle, in denen die MAC-Adresse geändert wird, durch die Hersteller der Bluetooth-Chips festgelegt werden (Bluetooth SIG 2015). Eine neue MAC-Adresse wird erst dann generiert, wenn der Adapter ausgeschaltet und danach erneut eingeschaltet wird (Bluetooth SIG 2019).

Gleichzeitig ist anzumerken, dass bereits Forschung hinsichtlich der Nachverfolgbarkeit von Bluetooth Geräten besteht, die ihre MAC-Adresse randomisieren. Beispielfhaft seien hier Becker et al. (2019) und Martin et al. (2019) genannt, die auf Spezifika der übertragenen Pakete aufbauen sowie auf dem Sendeverhalten.

2.4 System zur Datenerfassung

2.4.1 Kurzbeschreibung des technischen Systems der Datenerfassung

Die bestehende Hardware in den Fahrzeugen des NVV bzw. der angehörigen Verkehrsunternehmen ist bereits in der Lage, u. a.

- ein- und aussteigende Fahrgäste automatisch zu zählen (AFZS),
- den Fahrweg mit GPS-Koordinaten aufzuzeichnen,
- den Linienweg und Haltestellen zu differenzieren sowie
- die Daten über Hardware der Firma INIT bzw. der Firma IVU an einen Server zu senden.

Das im Rahmen dieser Dissertation genutzte technische System zur Datenerfassung von WLAN- und Bluetooth-Signalen setzt auf diesem System auf. Die Datenerfassung in den Fahrzeugen erfolgt durch Einbau eines zusätzlichen Systems (Raspberry Pi). Diese Lösung wurde konzipiert und umgesetzt,

da eine zeitnahe Erfassung von WLAN- und Bluetooth-Daten durch kostengünstige Hardware erfolgen sollte, die keine zusätzlichen Antennen im Bus benötigt (Abbildung 2-5). Darüber hinaus werden mit dieser Hardware auch WLAN- und Bluetooth-Daten an Haltestellen erfasst (Abbildung 2-6).

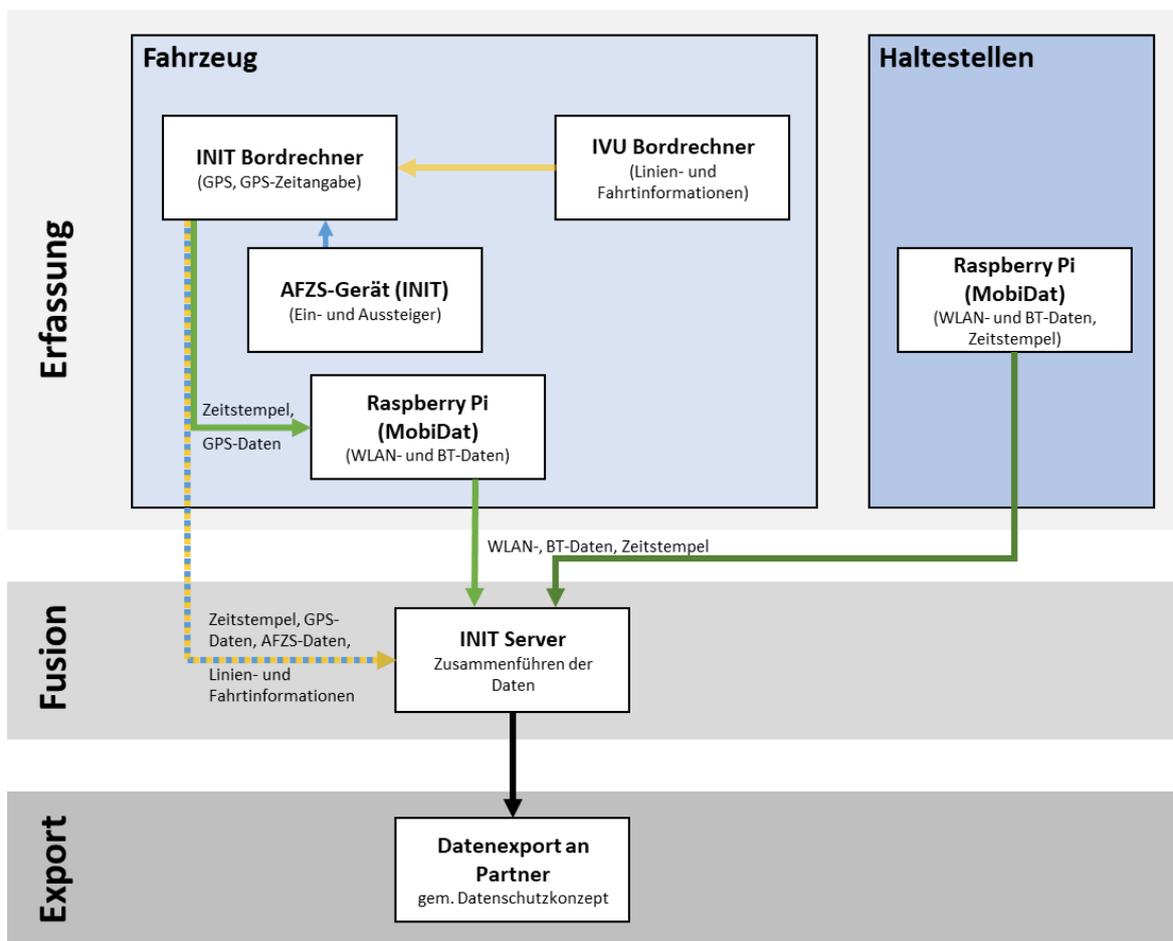


Abbildung 2-5: Übersicht des technischen Systems zur Datenerfassung (Bieland 2020)



Abbildung 2-6: Einbaubeispiel der Erfassungsinfrastruktur an Haltestellen des NVV (Nordhessischer Verkehrsverbund (NVV) 2020)

Bei Nachweis der Brauchbarkeit der Daten für das Ziel dieser Arbeit kann das System, bestehend aus Raspberry Pi Geräten, in den Fahrzeugen aufgehoben und die Komponenten zur Datenerfassung in die bestehende Hardware der Firma INIT (Bordrechner „CoPILOT“) integriert werden.

Für die Erfassung der WLAN- und Bluetooth-Daten wurde folgende Hardware in den ausgewählten Fahrzeugen und Haltestellen (Kapitel 4.2) verbaut:

- Raspberry Pi 3 Modell B+, Computer-Board inkl. Gehäuse,
- TPLINK TL-WN722N WLAN Adapter,
- Sandisk 16GB Micro-SD Speicherkarte,
- wasserdichte DC-DC-Abwärtswandler 12V/24V auf 5V 3A,
- GSM Modul.

Die Kosten pro Fahrzeug bzw. Haltestelle lagen bei ca. 110,- € zuzüglich Kosten für das benötigte monatliche Datenvolumen der SIM-Karten sowie den Einbau. Die Software auf den Raspberry Pi 3 wurde durch die INIT GmbH aufgespielt und erfasst

- die WLAN- (Probe Requests) und
- Bluetooth-Signale (Bluetooth-Inquiry-Nachrichten, keine BLE-Nachrichten (BLE = Bluetooth Low Energy))

in der Umgebung. Die Erfassung von WLAN-Signalen erfolgt statisch auf Kanal 1 des 2,4 GHz Frequenzspektrums. Da ein WLAN-Adapter nur die Funknachrichten jeweils eines Channels zur gleichen Zeit erfassen kann, kommt es zum Datenverlust der Funknachrichten auf allen weiteren Kanälen im 2,4 GHz- und im gesamten 5 GHz-Frequenzbereich.

Aufgrund der erläuterten Eigenschaften des 5 GHz-Spektrums (Durchdringung von Hindernissen, Störanfälligkeit) ist davon auszugehen, dass auch in Zukunft Verbindungen zwischen Smartphones und Access Points vornehmlich im 2,4 GHz-Spektrum durchgeführt werden. Neben dem statischen Abhören eines Kanals besteht grundsätzlich auch die Möglichkeit des sogenannten „Channel Hoppings“. Beim Channel Hopping wird der Kanal, in dem die Signale erfasst werden, in kurzen Zeitintervallen (mehrfach pro Sekunde) gewechselt. Die Erkenntnisse von Waltari und Kangasharju (2016) haben gezeigt, dass Endgeräte in Probe Request Bursts alle Kanäle abfragen, um so ein vollständiges Bild der vorhandenen WLAN-Netze zu erhalten. Statisches Abhören eines Kanals hat daher den Vorteil gegenüber dem dynamischen Ansatz, dass eine höhere Wahrscheinlichkeit besteht, eine größere Anzahl

verschiedener Endgeräte zu erfassen. Beim dynamischen Abhören dagegen besteht die Möglichkeit, dass einzelne Endgeräte nicht erfasst werden, weil keine Überschneidung zwischen abhörendem Kanal der Empfangsstation und Sendekanal des Endgeräts stattfindet.

Zur Erfassung der Bluetooth-Daten wird der Bluetooth-Chip in den Inquiry-Modus versetzt (Kapitel 2.3). Die Software sieht vor, dass alle 10 Sekunden ein Scan nach in der Nähe befindlichen Bluetooth-Geräten durchgeführt wird.

Die aufgezeichneten WLAN- und Bluetooth-Daten werden täglich über Mobilfunk an die INIT SE gesendet, vor der Speicherung anonymisiert und mit weiteren Fahrzeugdaten – i. W. Daten zum Fahrzeug, der Linie, GPS-Koordinaten und die zuletzt angefahrne Haltestelle – ergänzt (Abbildung 2-5). Anschließend erhalten die Beteiligten, unter anderem das Fachgebiet VPVS, Zugriff auf die Daten (Abbildung 2-5). Die AFZS-Daten werden gesondert zur Verfügung gestellt und lassen sich auf Basis einer eindeutigen Fahrt-ID mit den weiteren erfassten Daten verbinden.

2.4.2 Datensatzbeschreibung

Mit Hilfe des technischen Systems werden WLAN-, Bluetooth- und AFZS-Daten erfasst und weitergeleitet. In diesem Kapitel werden die übermittelten Datensätze, die als Grundlage für das Verfahren genutzt werden, beschrieben.

2.4.2.1 WLAN

Unter WLAN-Daten werden WLAN Probe Requests verstanden. Durch die Erfassung und Auswertung von WLAN Probe Request Daten sollen Rückschlüsse über den Ein- und Ausstiegsort einzelner Fahrgäste gezogen werden.

Das technische System erfasst und übermittelt WLAN- und Bluetooth-Daten seit September 2019. Aufgrund verschiedener Anpassungen der Software (i. W. Ergänzung von Daten, Erweiterungen des Systems zur Korrektur fehlender Daten), die mit umfassenden Prüfungen der Daten verbunden waren, ist die Nutzbarkeit der Daten jedoch verzögert. Seit August 2020 liegt ein stabil laufendes System zur Erfassung und Übermittlung der Daten vor. In der Verfahrensentwicklung werden lediglich Daten betrachtet, die ab August 2020 erfasst wurden.

Ein Ausschnitt der erfassten und übermittelten Daten ist in Tabelle 2-4 dargestellt. Die erfassten Daten werden unterschieden in

- Daten, die in den ausgesendeten Probe Request erzeugt werden (Smartphone)
- Daten, die durch das ÖV-Fahrzeug ergänzt werden (ÖV-Fahrzeug) sowie
- Daten, die die Erfassung durch die Hardware an einer Haltestelle kennzeichnen.

Tabelle 2-4: Übertragene Daten für die Verfahrensentwicklung aus Probe Request Frames

Daten- quelle	Information / Merk- mal	Kurzbeschreibung
Smartphone	EVENT_ID	Einzigartige ID des aufgezeichneten Events.
	TIME_STAMP	Zeitangabe des gesendeten Probe Request. Die Zeitangaben werden zwischen Access point und den verbundenen mobilen Endgeräten durch die Timing Synchronization Function (TSF) synchronisiert.
	MAC_TYPE	Unterscheidung der Art der MAC-Adresse nach „WIFI“ (Probe Request) oder „BLUETOOTH“ (Inquiry Response). Die WLAN- und Bluetooth-MAC-Adresse sind innerhalb eines Endgeräts unterschiedlich, da sie auf dem genutzten Chip basieren.
	HASHED_MAC_ ADDRESS	Die MAC-Adresse dient zur eindeutigen Identifikation eines Endgeräts. Auf Grundlage des Datenschutzkonzepts wird diese durch einen Hash-Wert anonymisiert. Der Hash-Algorithmus wird zu Beginn eines Kalendermonats angepasst.
	LOCALLY_ ADMINISTERED	Gibt den Wert des "least significant Bits" zurück. Dieser gibt Aufschluss darüber, ob die MAC-Adresse lokal erzeugt wurde (MAC-Randomisierung).
	SIGNAL_STRENGTH	Bei der Erfassung eines Probe Request Frames misst der Access Point die Signalstärke des empfangenen Signals. Diese wird durch den recieved signal strength indicator (RSSI) angezeigt.
ÖV-Fahrzeug	SENSOR_ID	Einzigartige ID der Datenerfassungseinheit
	VEHICLE_ID	Einzigartige ID des Fahrzeugs
	JOURNEY_ID	ID der Linienfahrt des Fahrzeugs, in dem sich der Sensor (Raspberry Pi) befindet. Es findet ein Abgleich mit den Planungsdaten des NVV statt (VDV452-Export).
	OPERATING_DAY	Einsatztag des Fahrzeugs
	PREVI- OUS_CALLED_ STOP_PLACE_ID	ID der Haltestelle, an der das Fahrzeug zuletzt gehalten hat oder gerade hält.
	PREVIOUS_CALL_ INDEX	Index des Halts innerhalb der Fahrt, beginnend bei 0.
	LONGITUDE	Längengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs
LATITUDE	Breitengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs	
Halte- stelle	STOP_PLACE_ID	Einzigartige ID der Haltestelle, an der sich der Raspberry Pi/Copilot-PC befindet (nur bei Haltestellendaten)

Eine vollständige Übersicht der übertragenen Daten befindet sich im Anhang 1. Die hier dargestellten Daten stellen die Datengrundlage des im Rahmen dieser Dissertation entwickelten Verfahrens dar. Eine detaillierte Beschreibung der genannten Informationen sowie aller weiteren Merkmale die in

Probe Requests übertragen werden, sind (IEEE Computer Society 2016) zu entnehmen.

2.4.2.2 Bluetooth

Unter Bluetooth-Daten werden im Folgenden Bluetooth Inquiry Nachrichten entsprechend dem Bluetooth 3.0 Standard (kein BLE) verstanden, die bei der Suche nach Bluetooth-Geräten in der Nähe sowie bei dem Verbindungsaufbau anfallen. Durch die Erfassung und Auswertung von BT-Daten sollen analog zum Vorgehen der WLAN-Daten Rückschlüsse über den Ein- und Ausstiegsort einzelner Fahrgäste getroffen werden.

Im Gegensatz zu den WLAN-Daten werden bei Bluetooth lediglich die MAC-Adresse sowie der Zeitstempel der Inquiry-Nachricht in den ausgestatteten Fahrzeugen und Haltestellen erfasst. Anschließend werden die fahrspezifischen Daten des Fahrzeugs (Linien ID, Journey ID, etc.) hinzugefügt.

Analog zu den Informationen und der Datensatzbeschreibung der WLAN-Daten werden in Tabelle 2-5 die erfassten und übertragenen Bluetooth-Daten dargestellt.

Tabelle 2-5: Übertragene Daten für die Verfahrensentwicklung aus Inquiry Nachrichten

Datenquelle	Information / Merkmal	Kurzbeschreibung
Smartphone	EVENT_ID	Einzigartige ID des aufgezeichneten Events.
	TIME_STAMP	Zeitangabe des gesendeten Probe Request. Die Zeitangaben werden zwischen Access Point und den verbundenen mobilen Endgeräten durch die Timing Synchronization Function (TSF) synchronisiert.
	MAC_TYPE	Unterscheidung der Art der MAC-Adresse nach „WIFI“ (Probe Request) oder „BLUETOOTH“ (Inquiry Message). Die WLAN- und Bluetooth-MAC-Adresse sind innerhalb eines Endgeräts unterschiedlich, da sie auf dem genutzten Chip basieren.
	HASHED_MAC_ADDRESS	Die MAC-Adresse dient zur eindeutigen Identifikation eines Endgeräts. Auf Grundlage des Datenschutzkonzepts wird diese durch einen Hash-Wert anonymisiert. Der Hash-Algorithmus wird zu Beginn eines Kalendermonats angepasst.
ÖV-Fahrzeug	Analog zu WLAN-Daten	
Haltestelle		

2.4.2.3 AFZS-Daten

AFZS-Daten beschreiben Daten zu der Anzahl von Ein- und Aussteigern, die das automatische Fahrgastzählsystem zählt (Kapitel 2.2.1). Im Verbundgebiet des NVV verfügen die Fahrzeuge aller regionalen Buslinien des öffentlichen Verkehrs (Aufgabenträger NVV) über AFZS. Die verbauten AFZS-Geräte halten die Messgenauigkeiten der VDV-Schrift 457 (Verband Deutscher Verkehrsunternehmen e. V. (VDV) 2018) ein, sodass die Differenz der Ein- und Aussteiger zwischen Messwerten und realen Zählwerten über alle Haltestellen 1% nicht übersteigt.

Diese Daten werden als Export für ausgewählte Linien bereitgestellt, die im Rahmen des Projekts Mobile Data Fusion mit der Hardware zur Erfassung von WLAN- und BT-Daten ausgestattet werden.

Eine Übersicht der durch das technische System erfassten und übermittelten Daten ist ausschnittsweise in Tabelle 2-6 dargestellt. Die AFZS-Daten liegen seit September 2019 vor.

Tabelle 2-6: Übertragene Daten für die Verfahrensentwicklung aus dem AFZS-System

Information / Merkmal	Kurzbeschreibung
OPERATING_DAY	Einsatztag des Fahrzeugs
VEHICLE_ID	Einzigartige ID des Fahrzeugs
METERS	Gefahrene Distanz in Meter zwischen Einträgen (Haltestellen)
STOP_LONGNAME	Vollständiger Haltestellenname
STOP_SHORTNAME	vierstellige Kennung der Haltestelle (Buchstaben)
STOP_CODE	vierstellige Kennung der Haltestelle (Zahlen)
PLANNED_ARRIVAL	Ankunftszeit an der Haltestelle gemäß Fahrplan
PLANNED_DEPARTURE	Abfahrtszeit an der Haltestelle gemäß Fahrplan
ACTUAL_ARRIVAL	Reale Ankunftszeit an der Haltestelle
ACTUAL_DEPARTURE	Reale Abfahrtszeit an der Haltestelle
GPS_LATITUDE	Breitengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs
GPS_LONGITUDE	Längengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs
STOP_TYPE	Grund für den Halt an der Haltestelle (u. a. geplanter Halte, Halt auf Fahrgastwunsch, Durchfahrt, Störfall)
PASSENGER_IN	Zählwert der Einsteiger nach Datenkorrektur durch internen Algorithmus
PASSENGER_OUT	Zählwert der Aussteiger nach Datenkorrektur durch internen Algorithmus
PASSENGER_LOAD	Belegung des Fahrzeugs als Differenz zwischen korrigierten Zählwerten der Ein- und Aussteiger
LINE_ID	Linienkennung der Linienfahrt des Fahrzeugs.
JOURNEY_ID	ID der Linienfahrt des Fahrzeugs.
APC_QUALITY	Kennzahl zur Genauigkeit der AFZS-Daten (Ein- und Aussteigerzahlen) auf Basis des Umfangs der notwendigen Korrekturen in Prozent.

2.4.3 Privatsphäre und Datenschutz

Das Smartphone hinterlässt im Rahmen dieser vielfältigen Verbindungsmöglichkeiten (Mobilfunk, WLAN, Bluetooth, Web-Anwendungen) Daten zur Verbindung sowie zu den ausgetauschten Informationen, ohne dass der Nutzer dies unmittelbar bemerkt. Diese Daten können zum Teil räumlich und zeitlich nachverfolgt werden (Tracking). Daneben greifen verschiedene Apps (etwa Google Maps) regelmäßig und mit dem Einverständnis des Nutzers auf den Standort des Smartphones zu.

Der Nutzen der gewonnenen Daten ist vielfältig und ermöglicht, wie in Kapitel 2.2.5 beschrieben, Auswertungen zum Einkaufsverhalten, Beziehungen zwischen Menschen oder auch der räumlichen Nachverfolgung.

Um diesem aus Nutzersicht oft ungewollten Tracking entgegenzuwirken, implementieren die Betriebssystemhersteller die sog. MAC-Randomisierung.

Rechtlich ist umstritten, ob MAC-Adressen als personenbezogene Daten nach Artikel 4 Nr. 1 DSGVO (Europäische Union 2021) zu werten sind, da im Gegensatz zu IP-Adressen oder Mobiltelefonnummern kein Verzeichnis mit einer direkten oder indirekten Zuordnung von MAC-Adressen zu Personen besteht. Sofern die Daten sofort anonymisiert werden, sehen die Wissenschaftlichen Dienste des Deutschen Bundestags keinen Personenbezug für Anwendungen gemäß Artikel 2 Absatz 1 DSGVO. Dabei wird der Standpunkt vertreten, dass selbst bei längerer Speicherdauer die Identifizierung einer natürlichen Person mittels einer MAC-Adresse nicht oder nur mit erheblichem Aufwand gelingen dürfte. Die Ausarbeitung kommt schließlich zu der Schlussfolgerung, dass der Personenbezug bei MAC-Adressen nicht komplett ausgeschlossen werden kann. Daher wird eine gesonderte Interessenabwägung im Einzelfall empfohlen (Deutscher Bundestag 2017).

Das in dieser Arbeit verwendete Datenschutz-Verfahren gründet auf dem vom BMVI geförderten Projekt „Automatische Ermittlung der Fahrgastnachfrage aus AFZS-, WLAN-, Bluetooth- und Verbindungsdaten“ (Mobile Data Fusion). Die Anforderungen des Datenschutzes werden von Beginn an bei der Datenerfassung berücksichtigt. Durch den Projektpartner BLIC wurde in Abstimmung mit dem Landesdatenschutzbeauftragten des Landes Hessen ein Datenschutzkonzept für die geplante Projektumsetzung entwickelt. Die Datenverarbeitung wird unter Berücksichtigung der geltenden gesetzlichen Vorschriften analysiert und Rahmenvorgaben für eine datenschutzkonforme Projektumsetzung festgelegt. Dabei sind in Ergänzung zur DSGVO auf Bundes- und Landesebene insbesondere die ergänzenden Regelungen zur Datenverarbeitung für Forschungszwecke relevant (§24 HDSIG bzw. §27 BDSG). Mit der Anonymisierung bzw. Pseudonymisierung der MAC-Adressen bei der Datenerfassung im Projekt wird ein ausreichendes Schutzniveau erreicht. Detaillierte Ausführungen zum Datenschutzkonzept sind den Projektberichten von Mobile Data Fusion zu entnehmen (Dontsova et al. 2020).

2.5 Datenanalyse und Bewertung

Das technische System erfasst eine Vielzahl verschiedener Informationen zu jedem empfangenen WLAN- bzw. Bluetooth-Signal. In Kapitel 2.3 wurden die netzwerktechnischen Grundlagen erläutert. Ziel dieses Kapitels ist es, die Ergebnisse der Literaturrecherche anhand der durch das technische System erfassten Daten zu verifizieren bzw. falsifizieren sowie anhand von zusätzlichen Untersuchungen ein Verständnis für das Sendeverhalten von Smartphones zu schaffen. Das Sendeverhalten sowie die Faktoren, die das

Sendeverhalten beeinflussen und beschreiben sind essentiell, um im weiteren Verlauf der Datenaufbereitung eine Filterung von „Stördaten“ (Daten von Nicht-Fahrgästen) vorzunehmen (Kapitel 3.2) und darauf basierend Quelle-Ziel-Matrizen abzuleiten (Kapitel 3.3).

Im Verlauf der Dissertation wurden durch automatische Updates der Betriebssysteme von Smartphones (Operating System, OS) deren Sendeeigenschaften nachhaltig gegenüber dem Zustand zu Beginn des Projekts bzw. dem Zustand vor dem Update geändert (u. a. Marktdurchdringung von MAC-Randomisierung). Die erfolgten Änderungen zielen dabei auf die Stärkung der Privatsphäre der Nutzer ab, womit gleichzeitig die Erfassbarkeit und Nachverfolgbarkeit einzelner Smartphones im Rahmen der Dissertation (erheblich) erschwert werden.

In diesem Kapitel werden zunächst die erfassten WLAN und Bluetooth-Daten untersucht. Auf Basis der Analyse werden Schlussfolgerungen für die Operationalisierbarkeit der Daten festgehalten. Dabei wird insbesondere auf die Erfassungsgenauigkeit im Hinblick auf den Ein- und Ausstiegsort einzelner Personenfahrten auf Basis der empfangenen Signale eingegangen.

2.5.1 Prüfung der WLAN-Daten

Im Rahmen dieser Dissertation werden nur solche WLAN-Daten berücksichtigt, deren lokaler Bit auf eine feste MAC-Adresse hinweist. Die Nutzbarkeit des lokalen Bits zur Identifikation ebendieser Daten wurde mittels Tests unter Laborbedingungen nachgewiesen (Kapitel 2.5.1.1). Die durchgeführten Tests wurden hinsichtlich Sendeeigenschaften der getesteten Smartphones ausgewertet.

WLAN-Daten mit randomisierter MAC-Adresse haben zum Zeitpunkt der Analyse einen geringen Anteil an allen empfangenen Daten dargestellt. Der Aufwand zur Operationalisierung von randomisierten WLAN-Daten durch die Entwicklung eines neuen Identifiers wird vor dem Hintergrund der geringen Erfolgsaussichten (Kapitel 2.3.2.5) nicht weiterverfolgt.

Darüber hinaus fand eine quantitative Überprüfung von erfassten WLAN-Daten aus dem Realbetrieb statt, um Erkenntnisse zur Erfassungsgenauigkeit sowie zur Nutzbarkeit der übermittelten Signalstärke (Received Signal Strength Indicator (RSSI)) als Indikator für Endgeräte innerhalb bzw. außerhalb des Erfassungsfahrzeugs abzuleiten.

2.5.1.1 Prüfung der WLAN-Daten aus Labortests

Zur Prüfung der Ergebnisse aus der Literaturrecherche sowie zur Prüfung auf etwaige Neuerungen im Sendeverhalten, die noch nicht in Veröffentlichungen dargestellt sind, wurden Untersuchungen unter Laborbedingungen durchgeführt.

Zur Untersuchung des Sendeverhaltens wurden die Endgeräte entsprechend den Ergebnissen der Literaturrecherche hinsichtlich der Einflussfaktoren auf das Sendeverhalten in verschiedenen Zuständen nach Nutzung (aktiv, inaktiv) und Verbindungsstatus (aktive Verbindung zu einem Access Point, Access Point in der Nähe (ohne Verbindung), kein Access Point in der Nähe) für die Dauer von jeweils 15 Minuten (je Smartphone und je Zustand) getestet. Stand der Untersuchung ist April 2020.

Aufgrund der vorherrschenden MAC-Randomisierung sowie dem Umstand, dass WLAN-Signale verschiedener Geräte in nahezu allen Räumen vorhanden sind, finden die Tests in einer gegen elektrische Strahlung abgesicherten Messkammer statt. Das bedeutet, dass elektrischen Wellen nicht aus der Kammer herausstrahlen können und ebenso keine elektrischen Wellen von außerhalb in die Messkammer gelangen. In dieser Messkammer werden die verschiedenen Smartphones der Testflotte (Tabelle 2-7) einzeln getestet, so dass die erfassten Probe Requests mit verschiedenen randomisierten MAC-Adressen eindeutig dem Testsmartphone (einziges Smartphone in der Kammer) zugeordnet werden können. Die Smartphones besitzen jeweils zum Testbeginn eine Akkuladung von mindestens 80%.

Tabelle 2-7: Smartphone-Testflotte

Hersteller	Typbezeichnung	Betriebssystem
Samsung	Galaxy S7	Android 7
Apple	6S	iOS 12.0.1
CAT	S61 (A9)	Android 9
CAT	S61 (A10)	Android 10

Die Untersuchung zielt darauf ab, Unterschiede im Sendeverhalten zwischen Smartphones bzw. zwischen Zuständen zu ermitteln. Dazu werden betrachtet:

- Anzahl von Probe Request Bursts und zeitliche Abstände zwischen Probe Request Bursts,

- Nutzbarkeit des lokalen Bit zur Identifikation von randomisierten MAC-Adressen.

Darüber hinaus werden Informationen zu der übermittelten MAC-Adresse sowie ausgewählten Informationen Elements (HT-Capabilities und Extended HT-Capabilities) gegeben.

Untersuchung zum lokalen Bit

Die erfassten Daten aus den Tests in der Messkammer wurden hinsichtlich der Nutzbarkeit des lokalen Bits (Kapitel 2.3.2.5) als Indikator für nicht-randomisierte MAC-Adressen geprüft. Die aktive Randomisierung konnte durch den Versuchsaufbau (Erfassung in einer gegen elektromagnetische Strahlung abgeschirmten Messkammer) festgestellt werden: Da lediglich die Signale von einem Smartphone erfasst werden konnten, liegt eine aktive Randomisierung dann vor, wenn Probe Requests von verschiedenen MAC-Adressen erfasst wurden.

Zur Prüfung des lokalen Bits wurde das erste übertragene Byte jeder MAC-Adresse in das Binärsystem umgewandelt und geprüft, ob der lokale Bit (zweites übertragenes Bit) bei der Randomisierung entsprechend den Vorgaben des IEEE (IEEE Computer Society 2016) konstant angibt, dass es sich um eine lokal veränderte (=randomisierte) MAC-Adresse handelt und somit den Wert „1“ besitzt.

Von den geprüften randomisierten MAC-Adressen besaß das lokale Bit in allen Fällen den Wert „1“. Darüber hinaus wurden die MAC-Adressen der Fälle geprüft, in denen keine Randomisierung erfolgte (bei Verbindung mit einem AP). In diesen Fällen nahm das lokale Bit stets den Wert „0“ ein.

Es ist somit davon auszugehen, dass das lokale Bit als Indikator für feste MAC-Adressen genutzt werden kann. Eine genauere Betrachtung des ersten Bytes der MAC-Adresse zeigt, dass diese bei der Randomisierung

- des iPhone 6S stets einen der Werte [2, 6, A, E],
- des Samsung S7 statisch den Wert [2] sowie
- des CAT S 61 mit der Google-CID (DA:A1:19) den Wert [A] annimmt.

Diese Werte haben erwartungsgemäß gemeinsam, dass sie nach Umformung in das Binärsystem eine „1“ beim zweitniedrigsten Bit aufweisen (Tabelle 2-8).

Bei der Verwendung des lokalen Bits zur zeitlich-räumlichen Nachverfolgung ist zu berücksichtigen, dass das Randomisierungsverhalten in drei Gruppen untergliedert werden kann (Stand 11/2021):

- Endgeräte, die durchweg ihre wahre MAC-Adresse aussenden
- Endgeräte, die eine randomisierte MAC-Adresse versenden, jedoch bei Eingehen einer aktiven Verbindung zu einem Access Point die wahre MAC-Adresse aussenden
- Endgeräte, die eine randomisierte MAC-Adresse verwenden, die bei Eingehen einer aktiven Verbindung zu einem Access Point eine neue randomisierte MAC-Adresse für Probe Requests verwenden.

Tabelle 2-8: Übersicht zur Umrechnung von Hexadezimalwerten in Binärwerte

Hexadezimalwert	Binärwert	Hexadezimalwert	Binärwert
0	0000	8	1000
1	0001	9	1001
2	0010	A	1010
3	0011	B	1011
4	0100	C	1100
5	0101	D	1101
6	0110	E	1110
7	0111	F	1111

Untersuchung zum Sendeverhalten

In Tabelle 2-9 sind die Ergebnisse der durchgeführten Tests (Stand 04/2020) dargestellt. Der minimale und maximale zeitliche Abstand ist auf die Zeit zwischen zwei aufeinander folgende Bursts bezogen. Der mittlere zeitliche Abstand bezeichnet das arithmetische Mittel aus allen Abstandswerten einer Messreihe.

Tabelle 2-9: Ergebnisübersicht: Sendeeigenschaften der Smartphone-Testgeräte

Zustand	Aktiv	x	x	x			
	Inaktiv / Standby				x	x	x
Verbindung zu AP	AP in der Nähe		x			x	
	kein AP in der Nähe	x			x		x
iPhone 6S	Anzahl Bursts	1	17	68	4	1	0
	Minimaler Abstand	zu wenig Daten	00:00:06	00:00:02	00:02:45	zu wenig Daten	zu wenig Daten
	Maximaler Abstand	zu wenig Daten	00:04:05	00:01:50	00:04:00	zu wenig Daten	zu wenig Daten
	Mittlerer Abstand	zu wenig Daten	00:00:55	00:00:13	00:03:15	zu wenig Daten	zu wenig Daten
Samsung S7	Anzahl Bursts	21	11	11	1	0	0
	Minimaler Abstand	00:00:05	00:00:03	00:00:04	zu wenig Daten	zu wenig Daten	zu wenig Daten
	Maximaler Abstand	00:04:03	00:02:50	00:03:51	zu wenig Daten	zu wenig Daten	zu wenig Daten
	Mittlerer Abstand	00:02:25	00:01:18	00:01:27	zu wenig Daten	zu wenig Daten	zu wenig Daten
CAT S61 (A9)	Anzahl Bursts	8	55	26	0	32	34
	Minimaler Abstand	00:00:20	00:00:08	00:00:10	zu wenig Daten	00:00:18	00:00:10
	Maximaler Abstand	00:03:01	00:02:01	00:03:01	zu wenig Daten	00:03:00	00:02:00
	Mittlerer Abstand	00:01:08	00:00:59	00:01:14	zu wenig Daten	00:01:01	00:00:57
CAT S61 (A10)	Anzahl Bursts	10	15	8	1	20	19
	Minimaler Abstand	00:00:13	00:00:03	00:01:58	zu wenig Daten	00:00:05	00:00:07
	Maximaler Abstand	00:03:32	00:03:17	00:02:07	zu wenig Daten	00:01:43	00:01:06
	Mittlerer Abstand	00:01:33	00:01:04	00:02:01	zu wenig Daten	00:00:45	00:00:51

Insgesamt zeigt sich unter den betrachteten Smartphones, dass hinsichtlich der Anzahl der Probe Requests Bursts sowie der zeitlichen Abstände Unterschiede zwischen den Smartphones sowie zwischen den Zuständen bestehen. Es ist davon auszugehen, dass durch Berücksichtigung weiterer Smartphones oder längerer Testzeiträume die Inhomogenität der Ergebnisse zunimmt.

Anzahl Probe Request Bursts

Die Ergebnisse der durchgeführten Tests spiegeln die Erkenntnisse aus der Literaturanalyse (Stand 2017 – 2020) wider. Die Anzahl und Frequenz von Probe Request Bursts variiert vor allem mit dem Smartphone sowie dem Zustand und Verbindungsstatus des Smartphones:

- Bei aktiver Nutzung werden in 8 von 12 Fällen gegenüber dem reinen Mitführen des Smartphones im „Standby“ (=inaktiv) mehr Probe Request Bursts versendet.
- Bei aktiver Verbindung zu einem Access Point wird die Anzahl der ausgesendeten Probe Requests Bursts in 11 von 12 Fällen stark reduziert.

Die Anzahl der versendeten bzw. erfassten Probe Request Bursts hat direkten Einfluss auf die zeitlichen Abstände zwischen den erfassten Probe Request Burst. Als Auswertungsmerkmal wurden Probe Request Bursts gewählt, da diese einen gestarteten Datenaustausch durch das Endgerät indizieren.

Zeitlicher Abstand

Der minimale zeitliche Abstand zwischen zwei Bursts liegt bei den betrachteten Smartphones und unter den Bedingungen der Labortests bei unter 20 Sekunden (Ausnahme: iPhone 6S, Zustand: inaktiv mit Verbindung zur einem AP). Dabei ist zu berücksichtigen, dass Probe Requests, die einen zeitlichen Abstand von unter einer Sekunde zueinander besitzen, als Burst zusammengefasst wurden. Die maximale Zeit zwischen zwei Bursts ist unbekannt, da in einigen Zuständen (i.d.R. bei Inaktivität) sehr wenig, bis gar keine Probe Requests im Testzeitraum versendet wurden. Werden die Tests betrachtet, in denen mehr als zwei Bursts empfangen wurden, liegt in Abhängigkeit des Smartphones und des Zustands der maximale zeitliche Abstand zwischen zwei Bursts etwa zwischen zwei und vier Minuten. Der mittlere Zeitabstand (arithmetisches Mittel) zwischen zwei Bursts liegt zwischen 13 Sekunden und über 3 Minuten in Abhängigkeit des Smartphones und Zustands.

MAC-Adresse und Informationen Elements (HT-Capabilities und Extended HT-Capabilities)

Bei der MAC-Adresse ist zu beobachten, dass diese bei einer Verbindung zu einem Access Point konstant bleibt. In diesem Fall wird bei der deutlich reduzierten Anzahl an Probe Request die wahre MAC-Adresse des Geräts zur Suche weiterer Netzwerke bzw. zur Suche weiterer Access Points des gleichen Netzwerks (etwa WLAN-Verstärker) verwendet. Besteht keine aktive Verbindung zu einem Access Points wird die MAC-Adresse mit jedem Burst geändert.

Bei allen für die Tests genutzten Smartphones bleiben sowohl die HT Capabilities als auch die Extended HT-Capabilities unabhängig vom Verbindungsstatus und Nutzungsstatus eines Smartphones konstant.

Eine quantitative Betrachtung der im Feld erfassten Probe Requests eines Fahrzeugsensors im Zeitraum vom 07.08.2020 bis 09.08.2020 ($n = 157.155$, Dataset Bieland/Mobile Data Fusion 2020) zeigt, dass die in Probe Requests übermittelten

- HT Capabilities bei rund 6% der erfassten einzigartigen MAC-Adressen nicht konstant sind und
- Extended HT Capabilities bei rund 4% der einzigartigen MAC-Adressen nicht konstant sind.

Dabei ist zu berücksichtigen, dass die Extended HT-Capabilities im Vergleich zu den HT Capabilities häufiger in Probe Requests nicht mitversendet werden. Die Abweichung tritt dabei zum Teil auch innerhalb der Probe Requests eines Bursts auf. Daher muss festgehalten werden, dass die HT Capabilities lediglich bei dem Großteil der Endgeräte konstant sind, jedoch ein (geringer) Teil der Endgeräte hier verschiedene Kombinationen der Werte innerhalb eines Bursts versendet. Ein Grund für die unterschiedlich übermittelten Eigenschaften kann in der Kompatibilität zu Access Points vermutet werden, die auf den älteren Standards, etwa 802.11 a/g, basieren (Tektronix 2013).

Untersuchung zur Signalstärke (RSSI)

In einer weiteren Untersuchung wurde die Wirkung der Fahrzeugkarosserie auf den RSSI gemessen. Ziel der Untersuchung war es, eine klare Abgrenzbarkeit zwischen erfassten Probe Requests aus dem Innenraum des Fahrzeugs (Endgeräte von Fahrgästen) und außerhalb der Fahrzeuge (nicht relevante Daten) zu ermitteln. Dabei war es aufgrund der Randomisierung der MAC-Adressen notwendig, jeden Probe Request hinsichtlich der Position in Bezug zum Fahrzeug zuordnen zu können.

Für die Durchführung wurden die Smartphones (mindestens zwei unterschiedliche Smartphones, vollgeladen, eingeschaltetes WLAN, keine Verbindung zu einem AP) innerhalb und außerhalb des Fahrzeugs (Standardlinienbus) in verschiedenen Entfernungen in einer Höhe von rund 1m aktiviert, so dass diese Probe Request Bursts aussenden, die unmittelbar mit dem Laptop bzw. der Empfangseinheit im Fahrzeug empfangen wurden (Abbildung 2-7). Je Position wurden drei Probe Request Burst (Messungen) erzeugt. Die Tests wurden mit geschlossenen und geöffneten Türen und Fenstern durchgeführt.

Bei der Interpretation ist zu berücksichtigen, dass die Position des Empfängers in den Fahrzeugen zur Datenerfassung nicht einheitlich erfolgen konnte. Der Versuchsaufbau bezieht die verschiedenen Positionen des Empfängers durch die Positionen des Smartphones im Fahrzeug (vorne, mittig, hinten) ein.

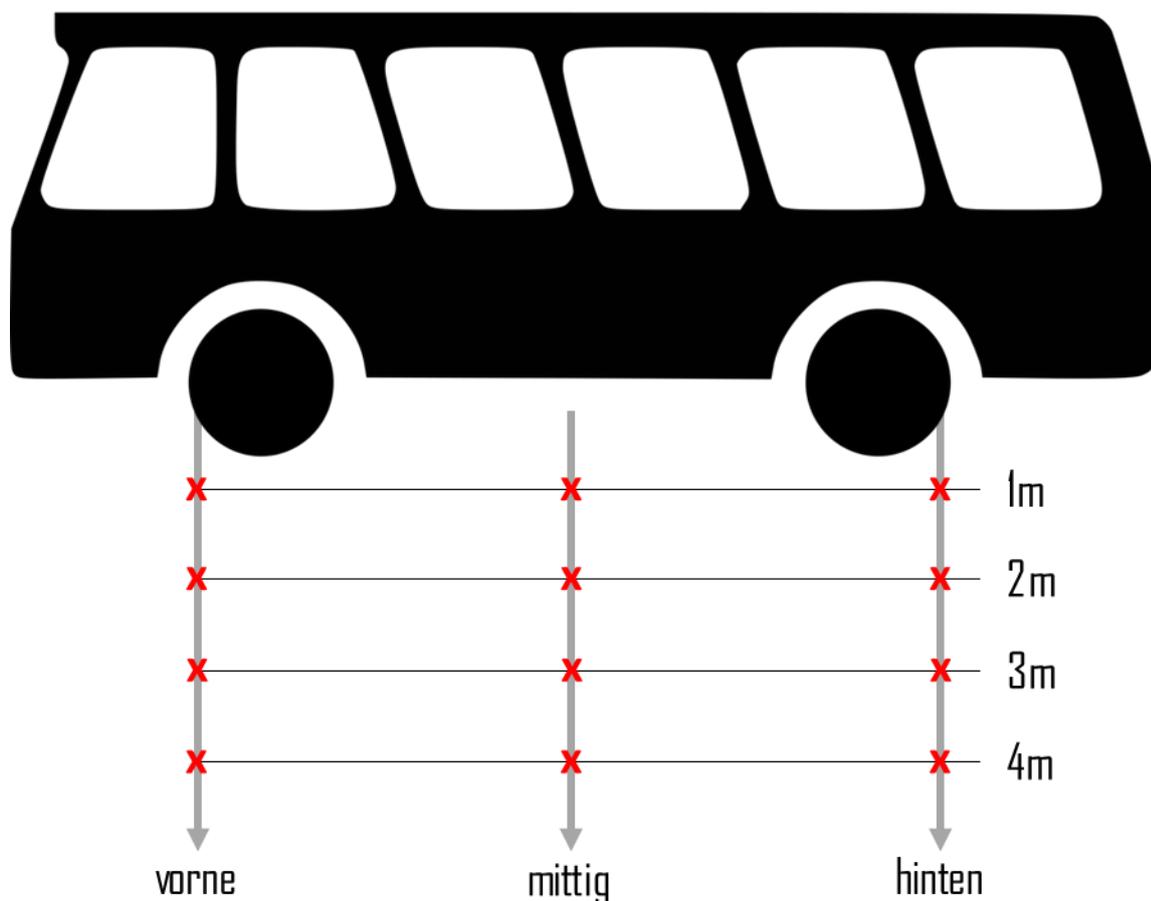


Abbildung 2-7: Versuchsaufbau zur Ermittlung der Signaldämpfung (Skizze)

Ein Ausschnitt der Ergebnisse der Messungen der Signalstärke (RSSI) innerhalb und außerhalb des Standardlinienbusses sind in Tabelle 2-10 dargestellt. Die Signalstärke wird dabei in dBm (Dezibel Milliwatt) angegeben. Der Leistungspegel in dBm beschreibt das Verhältnis der Leistung im Vergleich zur Bezugsleistung von 1 mW. Da der Leistungspegel durch die Distanz zwischen Sender und Empfänger abgeschwächt wird, ist dieser bei den Messungen stets negativ. Je näher der Wert also an 0 ist, desto höher ist der empfangene Leistungspegel. Dies kann auf eine geringere Distanz zwischen Sender und Empfänger hindeuten. Eine weitere Möglichkeit sind Unterschiede der Sendeleistung beim Aussenden der Signale des Senders.

Die Betrachtung der Signalstärke einzelner Endgeräte zeigt, dass keine eindeutige Unterscheidung möglich ist. Die Spannweiten der RSSI-Werte von Signalen einzelner Smartphones, die sowohl innerhalb (bei geschlossenen Türen) als auch außerhalb des Fahrzeugs in verschiedenen Abstandsklassen getestet wurden, überschneiden sich zu großen Teilen. Es liegen keine signifikanten Unterschiede (F-Test) zwischen den Varianzen der Stichproben (innerhalb / außerhalb des Fahrzeugs) vor.

Tabelle 2-10: Messwerte der Signalstärke [dBm]: Xiaomi Redmi Note 4

		Messung	Xiaomi Redmi Note 4					
			Türen geöffnet			Türen geschlossen		
			vorne	mittig	hinten	vorne	mittig	hinten
innen		1	-46	-38	-48	-45	-45	-42
		2	-55	-31	-54	-45	-37	-44
		3	-43	-29	-49	-48	-34	-46
außen	1m	1	-48	-47	-64	-52	-37	-50
		2	-47	-44	-51	-51	-41	-47
		3	-50	-49	-53	-59	-41	-47
	4m	1	-53	-48	-47	-59	-59	-48
		2	-52	-55	-51	-66	-59	-47
		3	-67	-51	-49	-66	-64	-50

Beispielhaft seien hier die Ergebnisse des Xiaomi Redmi Note 4 genannt: Die RSSI-Werte im Innenraum liegen zwischen -29 und -55 dBm und außerhalb des Fahrzeugs zwischen -37 und -67 dBm. Darüber hinaus konnten Unterschiede in der Spannweite und dem RSSI-Wert zwischen verschiedenen Smartphones der Testflotte festgestellt werden. In dem Zuge ist zu berücksichtigen, dass es sich hier um Tests unter Laborbedingungen handelt, wodurch die einzelnen erfassten Probe Requests eindeutig einem Smartphone zugeordnet werden können. In der Realität müsste zunächst eine Zuordnung einzelner und ggf. randomisierter Probe Requests zu einem Endgerät stattfinden.

2.5.1.2 Prüfung der WLAN-Daten aus dem Realbetrieb

Zur Analyse der erfassten Daten wurde ein Datensatz erstellt, bestehend aus den erfassten Daten eines Fahrzeugs bzw. Sensors im Zeitraum vom 06.10.2020 bis 01.11.2020 (Dataset Bieland/Mobile Data Fusion 2022). Der Datensatz umfasst insgesamt über 714.000 WLAN-Signale von 56.131 nicht randomisierten MAC-Adressen.

Auf dieser Grundlage wurden exemplarische Untersuchungen durchgeführt

- zur Anzahl der erfassten Signale,
- zur mittleren zeitlichen Verteilung der Signale von einzelnen Endgeräten pro Linienfahrt (MAC-Adresse) und
- zur erfassten Signalstärke.

Die Auswertungen wurden tagesscharf durchgeführt und anschließend über alle Betrachtungstage aggregiert. Die Einzigartigkeit einer MAC-Adresse pro

Tag wurde durch die Verbindung von Datum und gehashter MAC-Adresse erreicht. Diese Kombination wird in den folgenden Tabellen und Beschreibungen als „MAC-Adresse*“ bezeichnet. Da einzelne (feste) MAC-Adressen an mehreren Tagen erfasst wurden, ist die Anzahl der in der Tabelle betrachteten MAC-Adressen* (Kombination aus Datum und gehashter MAC-Adresse zur tagesscharfen Auswertung) mit 72.658 größer als die Anzahl aller einzigartigen MAC-Adressen über den Betrachtungszeitraum (56.131).

Untersuchung zur Anzahl erfasster Signale

Tabelle 2-11 stellt die Häufigkeitsverteilung der Anzahl erfasster Signale von Endgeräten in Klassen dar. Aus der Tabelle geht hervor, dass von mehr als 26.000 MAC-Adressen* - knapp ein Drittel – lediglich ein Signal erfasst wurde. Für diese MAC-Adressen* können entsprechend keine zeitlichen Abstände zwischen Signalen ermittelt werden. Von den verbleibenden rund 46.300 MAC-Adressen* wurde mehr als ein Signal erfasst. Die mittlere Anzahl erfasster Signale pro MAC-Adresse* liegt bei zehn Signalen. Werden nur MAC-Adressen* betrachtet, die mehrfach erfasst wurden, steigt der Mittelwert auf 15 Signale an. In wenigen Ausnahmen werden mehr als 100 Signale erfasst. Hier ist davon auszugehen, dass es sich nicht um Fahrgäste, sondern um störende Signale handelt (Nicht-Fahrgäste), wie etwa das Endgerät des Busfahrers.

Tabelle 2-11: WLAN-Signale: Häufigkeitsverteilung der Anzahl erfasster Signale

Anzahl Signale von MAC-Adressen*	Anzahl relevanter MAC-Adressen*	Rel. Häufigkeit an allen Geräten	Kumulierte Häufigkeit
≤ 1	26.387	36%	36%
>1 bis 5	32.522	45%	81%
>5 bis 10	6.466	9%	90%
>10 bis 50	5.323	7%	97%
>50 bis 100	950	1%	99%
>100	1.010	1%	100%
Gesamt	72.658		

Hinweis: MAC-Adressen = Kombination aus Datum und gehashter MAC-Adresse zur tagesscharfen Auswertung*

Untersuchung zur zeitlichen Verteilung der Signale

Die Ergebnisse der Auswertung des mittleren zeitlichen Abstands zwischen erfassten Signalen einer MAC-Adresse* sind in Tabelle 2-12 dargestellt. Dabei wurden nur die MAC-Adressen* berücksichtigt, von denen mehr als ein Signal erfasst werden konnte. Darüber hinaus wurden Signale, die weniger als eine Sekunde auseinanderliegen, aggregiert und Signale, die mehr als

eine Stunde und zwanzig Minuten (= Umlaufzeit des Fahrzeugs) auseinanderliegen, als neue Fahrgastfahrt interpretiert. Wurde ein Signal als eine neue Fahrgastfahrt interpretiert, wurde der Zeitabstand zum vorherigen Signal (>01:20h) nicht berücksichtigt und die folgenden Signale der MAC-Adresse* weiterhin entsprechend den genannten Rahmenbedingungen ausgewertet.

Tabelle 2-12: WLAN-Signale: mittlerer Zeitabstand zwischen erfassten Signalen

Mittlerer Zeitabstand zwischen Signalen einer MAC-Adresse*	Anzahl relevanter MAC-Adressen*	Rel. Häufigkeit an relevanten MAC-Adressen*	Kumulierte Häufigkeit
≤ 1 min	9.711	64%	64%
>1 bis 5 min	2.887	19%	83%
>5 bis 10 min	848	6%	88%
>10 bis 30 min	1.138	7%	96%
>30 bis 45 min	378	2%	98%
>45 bis 60 min	140	1%	99%
>60 bis 80 min	155	1%	100%
Gesamt	15.257		

Hinweis: MAC-Adressen = Kombination aus Datum und gehashter MAC-Adresse zur tagesscharfen Auswertung*

Von den ursprünglich 72.658 MAC-Adressen* bleiben entsprechend der geschärften Rahmenbedingungen 15.257 relevante MAC-Adressen* (21%) zur Analyse. Dabei ist anzumerken, dass ein Großteil der Signale der Kategorie „bis 5“ innerhalb eines Bursts und damit in weniger als einer Sekunde empfangen wurde. Diese Daten wurden nicht weiter betrachtet. Anschließend wurde je MAC-Adresse* der mittlere Abstand zwischen den erfassten Signalen einer Fahrgastfahrt gebildet. Insgesamt wurde von 26.387 MAC-Adressen* nur ein Signal erfasst, weitere 31.014 MAC-Adressen* hatten einen Zeitabstand von unter einer Sekunde und wurden daher als einzelner Burst ohne Zeitabstand verworfen.

Ein Großteil (64%) der Endgeräte, die mehr als ein Signal ausgesendet haben, werden im Mittel bereits innerhalb von einer Minute erneut erfasst. Wird der Zeitraum auf fünf Minuten erweitert, steigt der Anteil auf 83%. Bei 11% der MAC-Adressen*, von denen mehr als ein Signal erfasst wurde, liegt eine Pause von im Mittel über 10 Minuten zwischen den erfassten Signalen vor. Das bedeutet, dass 89% der Geräte, von denen mehr als ein Signal erfasst wurde, im Mittel innerhalb von 10 Minuten mehrfach erfasst wurden. Bei der Interpretation ist zu beachten, dass etwaige „Störsender“ (etwa Dauersender oder wiederkehrend erfasste stationäre WLAN-Sender auf dem Linienweg) nicht herausgefiltert wurden. Insgesamt, d. h. über alle betrachteten mehrfach

erfassten MAC-Adressen* liegt der mittlere zeitliche Abstand zwischen Signalen vom gleichen Endgerät bei mehr als vier Minuten (00:04:24). Im Mittel wurden rund zwölf Werte (zeitliche Differenz) je MAC-Adresse* berücksichtigt.

Aufbauend auf den Analysen lässt sich festhalten, dass bei den WLAN-Daten kein festes Muster zur Wiedererkennung abgeleitet werden kann. Gleichzeitig wird deutlich, dass bei dem überwiegenden Teil der Endgeräte, von denen mehr als ein Signal erfasst wurde, der zeitliche Abstand zwischen zwei erfassten Bursts im geringen Minutenbereich (bis 5 Minuten) liegt.

Sofern ein Programm entwickelt wird, das auch randomisierte MAC-Adressen mit hoher Wahrscheinlichkeit einzelnen Endgeräten zuordnen kann – beispielsweise über die Information Elements – sind die Auswertungen zu überarbeiten.

Untersuchung zur Signalstärke (RSSI)

Nach Oransirikul et al. (2019) und Nitti et al. (2020) konnten Unterschiede zwischen der empfangenen Signalstärke (RSSI) von Geräten innerhalb und außerhalb eines Fahrzeugs bzw. Raums festgestellt werden. Nitti et al. haben daraufhin die Spannweiten des RSSI für Signale innerhalb (> -51 dBm) bzw. außerhalb (< -61 dBm) des Fahrzeugs festgelegt. Diese Spannweite basiert auf einem Versuch, der in Räumen einer Universität durchgeführt wurde. Über das Material der Wände bzw. die Bauweise sind keine Informationen bekannt. Die Übertragbarkeit auf Fahrzeuge ist nicht nachgewiesen.

Zur Überprüfung der Brauchbarkeit der empfangenen Signalstärke wurden daher zum einen die in Kapitel 2.5.1.1 beschriebene Untersuchung unter Laborbedingungen durchgeführt. Darüber hinaus wurden Daten aus dem Realbetrieb analysiert. Die Analyse zeigt bei über 714.000 aufgezeichneten Probe Requests in der Zeit zwischen dem 06.10.2020 und 01.11.2020 entgegen den Erkenntnissen von Nitti et al. eine Verteilung der RSSI-Werte (Abbildung 2-8), die keine klare Differenzierung in die Gruppen Fahrgäste und Nicht-Fahrgäste zulässt. Dies bestätigt die Erkenntnisse aus dem durchgeführten Labortest. Auffällig bei den Daten ist die große Häufigkeit von knapp 17% der erfassten Probe Request Signalstärken bei -95 dBm. Dieses Phänomen konnte auch nach Rücksprache mit dem Softwareentwickler für die Erfassungsinfrastruktur nicht geklärt werden. Es wird vermutet, dass geringere empfangene Signalstärken auf den Wert von -95 dBm aggregiert werden.

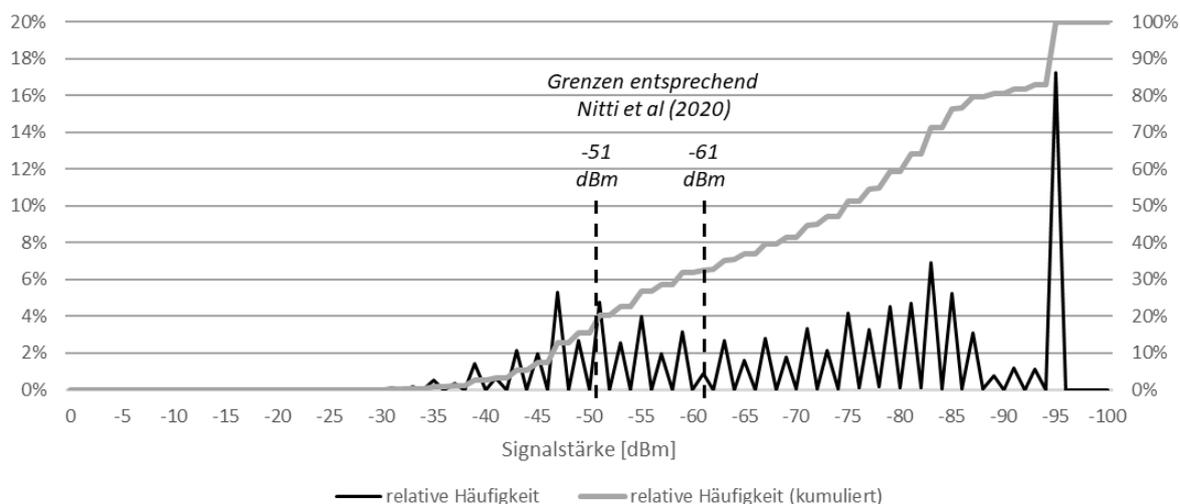


Abbildung 2-8: RSSI-Verteilung, n = 714.994

2.5.2 Prüfung der Bluetooth-Daten

Zur Analyse der erfassten Daten wurde ein Datensatz erstellt, bestehend aus den erfassten Daten eines Fahrzeugs bzw. Sensors (MDF 13) im Zeitraum vom 06.10.2020 bis 01.11.2020 (Dataset Bieland/Mobile Data Fusion 2022). Der Datensatz umfasst insgesamt über 253.000 Bluetooth-Signale von 48.385 einzigartigen MAC-Adressen. Auf dieser Grundlage wurden exemplarische Untersuchungen durchgeführt

- zur mittleren zeitlichen Verteilung der Signale von einzelnen Endgeräten pro Linienfahrt (einzigartige MAC-Adresse) und
- zur Anzahl der erfassten Signale.

Der Inquiry-Prozess wird softwareseitig alle 10 Sekunden durch die Erfassungseinheit angestoßen. Grundsätzlich liegt daher die Annahme nahe, dass Signale von einzelnen Geräten, die sich über einen längeren Zeitraum in der Nähe des Inquirers (Sender) befinden, alle 10 Sekunden erfasst werden. Zur Überprüfung dieser Annahme sowie der zeitlichen Abstände zwischen mehreren Erfassungen des gleichen Endgeräts (= gleiche MAC-Adresse) wurde die o. g. Datengrundlage analysiert.

Die Auswertungen wurden tagesscharf durchgeführt und anschließend über alle Betrachtungstage aggregiert. Die Einzigartigkeit einer MAC-Adresse pro Tag wurde durch die Verbindung von Datum und gehashter MAC-Adresse erreicht. Diese Kombination wird in den folgenden Tabellen und Beschreibungen als „MAC-Adresse*“ bezeichnet. Da einzelne (feste) MAC-Adressen an mehreren Tagen erfasst wurden, ist die Anzahl der in der Tabelle betrachteten MAC-Adressen* (Kombination aus Datum und gehashter MAC-Adresse

zur tagesscharfen Auswertung) mit 53.873 größer als die aller einzigartigen MAC-Adressen über den Betrachtungszeitraum (48.385).

Tabelle 2-13 stellt die Häufigkeitsverteilung der Anzahl erfasster Signale von Endgeräten in Klassen dar. Aus der Tabelle geht hervor, dass von mehr als 36.000 MAC-Adressen* - knapp zwei Drittel – lediglich ein Signal erfasst wurde. Für diese MAC-Adressen* können entsprechend keine zeitlichen Abstände zwischen Signalen ermittelt werden. Von den verbleibenden rund 17.800 MAC-Adressen* wurde mehr als ein Signal erfasst. Die mittlere Anzahl erfasster Signale pro MAC-Adresse* liegt bei rund 5 Signalen. Werden nur MAC-Adressen* betrachtet, die mehrfach erfasst wurden, liegt der Mittelwert bei rund zwölf Signalen. Wenige Ausnahmen werden mehr als 100 Mal erfasst. Hier ist davon auszugehen, dass es sich nicht um Fahrgäste, sondern um störende Signale handelt (Nicht-Fahrgäste), wie etwa das Endgerät des Busfahrers.

Tabelle 2-13: Bluetooth-Signale: Häufigkeitsverteilung der Anzahl erfasster Signale

Anzahl Signale von MAC-Adressen*	Anzahl relevanter MAC-Adressen*	Rel. Häufigkeit an allen Geräten	Kumulierte Häufigkeit
≤ 1	36.118	67%	67%
>1 bis 5	10.188	19%	86%
>5 bis 10	2.518	5%	91%
>10 bis 50	4.199	8%	98%
>50 bis 100	850	2%	100%
Gesamt	53.873		

Hinweis: MAC-Adressen = Kombination aus Datum und gehashter MAC-Adresse zur tagesscharfen Auswertung*

Die Ergebnisse der Auswertung des mittleren zeitlichen Abstands zwischen erfassten Signalen einer MAC-Adresse* sind in Tabelle 2-14 dargestellt. Dabei wurden nur die MAC-Adressen* berücksichtigt, von denen mehr als ein Signal erfasst werden konnte. Darüber hinaus wurden Signale, die weniger als eine Sekunde auseinanderliegen, aggregiert und Signale, die mehr als eine Stunde und zwanzig Minuten (= Umlaufzeit des Fahrzeugs, in dem der betrachtete Sensor verbaut ist) als neue Fahrgastfahrt interpretiert.

Von den ursprünglich 53.873 MAC-Adressen* bleiben entsprechend der geschärften Rahmenbedingungen 14.761 relevante MAC-Adressen* (27%) zur Analyse. Anschließend wurde je MAC-Adresse* der mittlere Abstand zwischen den erfassten Signalen einer Fahrgastfahrt gebildet.

Tabelle 2-14: Bluetooth-Signale: mittlerer Zeitabstand zwischen erfassten Signalen

Mittlerer Zeitabstand zwischen Signalen einer MAC-Adresse*	Anzahl relevanter MAC-Adressen*	Rel. Häufigkeit an relevanten MAC-Adressen*	Kumulierte Häufigkeit
≤ 1 min	10.426	71%	71%
>1 bis 5 min	2.633	18%	89%
>5 bis 10 min	326	2%	91%
>10 bis 30 min	845	6%	97%
>30 bis 45 min	213	1%	98%
>45 bis 60 min	267	2%	100%
>60 bis 80 min	51	0%	100%
Gesamt	14.761		

Hinweis: MAC-Adressen = Kombination aus Datum und gehashter MAC-Adresse zur tagesscharfen Auswertung*

Bei der Interpretation wird deutlich, dass die Endgeräte nicht in der Häufigkeit der versendeten Inquiry-Requests (alle 10 Sekunden) erfasst werden. Es ist davon auszugehen, dass das Inquiry-Response-Verhalten von weiteren, an dieser Stelle unbekanntem, Faktoren abhängt.

Ein Großteil (71%) der Endgeräte, die mehr als ein Signal ausgesendet haben, werden im Mittel bereits innerhalb von einer Minute erneut erfasst. Wird der Zeitraum auf fünf Minuten erweitert, steigt der Anteil auf 89%. Bei 9% der MAC-Adressen*, von denen mehr als ein Signal erfasst wurde, liegt eine Pause von im Mittel über 10 Minuten zwischen den erfassten Signalen vor. Das bedeutet, dass 91% der Geräte, von denen mehr als ein Signal erfasst wurde, im Mittel innerhalb von 10 Minuten mehrfach erfasst wurden. Bei der Interpretation ist zu beachten, dass etwaige „Störsender“ (etwa Dauersender oder wiederkehrend erfasste stationäre Bluetooth-Sender auf dem Linienweg) nicht herausgefiltert wurden. Insgesamt, d. h. über alle betrachteten mehrfach erfassten MAC-Adressen*, liegt der mittlere zeitliche Abstand zwischen Signalen vom gleichen Endgerät bei mehr als drei Minuten (00:03:39). Im Mittel wurden rund sieben Werte (zeitliche Differenz) je MAC-Adresse* berücksichtigt.

Aufbauend auf den Analysen lässt sich festhalten, dass auch bei den Bluetooth-Daten kein festes Muster zur Wiedererkennung abgeleitet werden kann. Gleichzeitig wird deutlich, dass bei dem überwiegenden Teil der Endgeräte, von denen mehr als ein Signal erfasst wurde, der zeitliche Abstand zwischen zwei erfassten Signalen im geringen Minutenbereich (bis 5 Minuten) liegt.

Weitere Merkmale (lokaler Bit, Vendor, RSSI) werden bei den erfassten Bluetooth-Daten nicht mit übertragen.

2.5.3 Beurteilung der Erfassungsgenauigkeit

Die Erfassungsgenauigkeit ist in hohem Maße von dem Zeitpunkt des ersten und letzten erfassten Signals eines Fahrgast-Endgeräts während einer Linienfahrt abhängig. Die Erfassung der WLAN- und Bluetooth-Signale basiert nicht auf einer festen Frequenz, d. h. einem wiederkehrenden Sendemuster, sondern ist von äußeren Faktoren (Nutzerinteraktion, Verbindungsstatus) abhängig, die nicht im Rahmen einer passiven Datenerfassung beeinflusst werden können (Kapitel 2.3.2.4, 2.5.1, 2.5.2).

2.5.3.1 Genauigkeit des Ein- und Ausstiegs

Aus den durchgeführten Tests ging hervor, dass bei dem Großteil der Endgeräte im Abstand von wenigen Minuten wiederholt WLAN- oder Bluetooth-Signale erfasst werden können. Gleichzeitig ist jedoch zu berücksichtigen, dass bei einzelnen Endgeräten zum Teil der zeitliche Abstand zwischen zwei Signalen mehr als 10 Minuten beträgt.

Sowohl das Sendeverhalten der Endgeräte als auch das Nutzungsverhalten der Fahrgäste sind unbekannt und führen dazu, dass eine Zuordnung von Start- und Zielhaltestelle auf Basis des ersten und letzten erfassten Signals (Kapitel 3.2.3) mit Unsicherheiten verbunden ist. Ebenso besteht die Möglichkeit, dass einzelne Fahrgäste trotz Mitführen eines Endgeräts, während einer (kurzen) Linienfahrt nicht erfasst werden. Konkret führt dieser Umstand dazu, dass

- der **Eintrittszeitpunkt** (Quelle der Fahrt) nicht genau bestimmt werden kann: Grundsätzlich ist es möglich, dass das Smartphone eines Fahrgasts den ersten Probe Requests sofort beim Einstieg in das Fahrzeug, nach einer Minute, nach fünf Minuten, nach zehn Minuten oder nach einem größeren Zeitraum verschickt und
- der **Austrittszeitpunkt** (Ziel der Fahrt) nicht genau bestimmt werden kann: Analog zum Eintrittszeitpunkt besteht die Möglichkeit, dass das Smartphone eines Fahrgasts den letzten im Fahrzeug erfassten Probe Request versendet, der Fahrgast jedoch noch mehrere Minuten weiterfährt und das Fahrzeug verlässt, ohne dass ein weiterer Probe Request versendet wird.

Abbildung 2-9 dient zur Verdeutlichung. Hier wird der zeitliche Verlauf einer Linienfahrt sowie drei Szenarien für die Erfassung von Signalen (WLAN /

Bluetooth, schwarz) skizziert. Anhand des Beispiels können folgende Rahmenbedingungen zur Erfassung festgehalten werden:

- Es können keine Signale vor dem Einstieg bzw. nach dem Ausstieg des Fahrgasts im Fahrzeug erfasst werden.
- In Abhängigkeit des Geräts bzw. insbesondere bei WLAN-Signalen des Verbindungszustandes und des Grads der Nutzerinteraktion werden zu verschiedenen Zeitpunkten während der Fahrgastfahrt Signale erfasst.
- Die Verzögerung zwischen Ein- bzw. Ausstieg und ersten bzw. letztem erfassten Signal ist unbekannt.
- Der Verbindungszustand und der Grad der Nutzerinteraktion ist unbekannt.

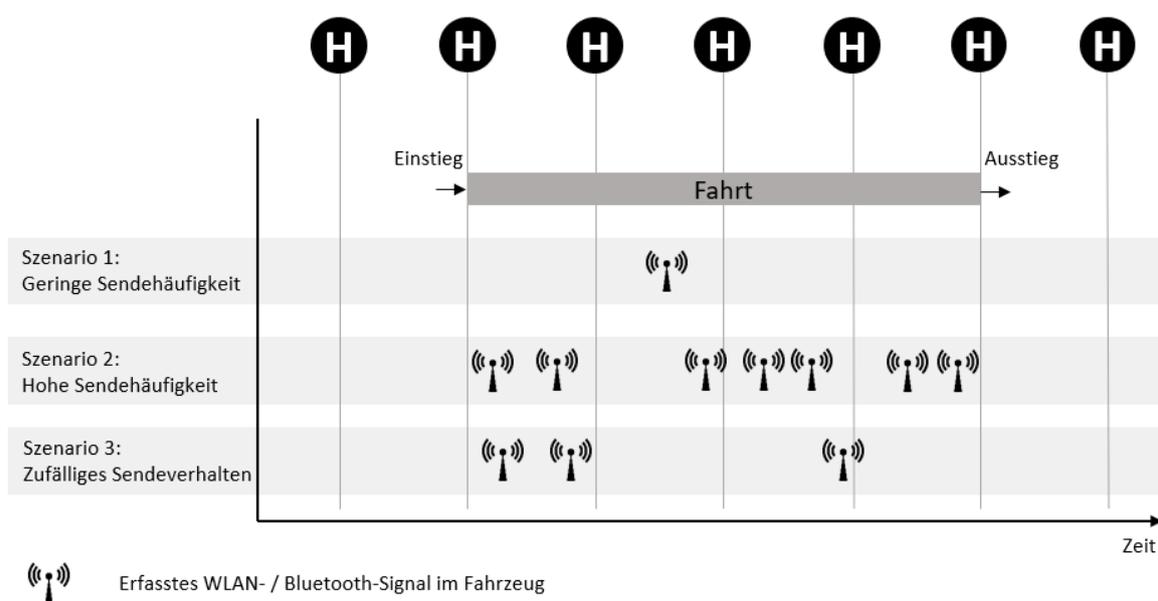


Abbildung 2-9: Modellhafte Darstellung von erfassten Signalen im Fahrzeug während einer Linienfahrt

Mit der Ausstattung von Haltestellen kann insbesondere die Erfassungsgenauigkeit für die Einstiegshaltestelle gesteigert werden, da die Wartezeit der Fahrgäste ebenfalls genutzt werden kann, um Signale der Endgeräte zu empfangen. Nähere Informationen dazu werden in Kapitel 3.3 beschrieben, das die Herleitung der Quelle-Ziel-Matrizen beschreibt.

2.5.3.2 Kurzstreckenfahrer

Kurzstreckenfahrer zeichnen sich durch geringe Reisedistanzen und damit kurze Aufenthaltsdauern innerhalb des ÖPNV-Fahrzeugs aus. Das bedeutet zum einen, dass diese Fahrgäste nur über einen kurzen Zeitraum erfassbar

sind und zum anderen, dass zwischen Ein- und Ausstiegshaltestellen nur wenige weitere Haltestellen liegen.

Vor dem Hintergrund des Sendeverhaltens der Endgeräte ist somit davon auszugehen, dass Kurzstreckenfahrer teilweise nicht bzw. nur zu einem Zeitpunkt erfasst werden (erstes Signal = letztes Signal). Die Wahrscheinlichkeit zur (mehrfachen) Erfassung von Fahrgästen mit langer Reisezeit (>15 Minuten) ist entsprechend höher.

Dies führt zu einem systematischen Fehler bei der Erstellung der Quelle-Ziel-Matrizen auf Basis der erfassten Daten, da die Kurzstreckenfahrten unterschätzt werden.

2.5.3.3 Anzahl Endgeräte der Fahrgäste

Darüber hinaus ist zu berücksichtigen, dass

- einige Fahrgäste über kein WLAN- bzw. Bluetooth-fähiges Endgerät verfügen,
- einige Fahrgäste über ein WLAN- bzw. Bluetooth-fähiges Endgerät verfügen,
- einige Fahrgäste über mehrere WLAN- bzw. Bluetooth-fähige Endgeräte verfügen.

Verfügt ein Fahrgast über kein WLAN- bzw. Bluetooth-fähiges Endgerät, kann dieser Fahrgast und seine Linienfahrt nicht in den erfassten Signalen der Haltestellen und Fahrzeuge nachvollzogen werden, wurde aber dennoch durch das AFZS-System beim Ein- und Ausstieg erfasst. Führt ein Fahrgast ein WLAN- bzw. Bluetooth-fähiges Endgerät mit sich (Annahme: Standardfall) führt es dazu, dass die Signale ebendieses Endgeräts grundsätzlich erfasst werden können. Bei dem Mitführen von mehreren WLAN- bzw. Bluetooth-fähigen Endgeräten (Smartphone, Laptop o. ä.) ergibt sich das Problem, dass zwei oder mehr Geräte zu unterschiedlichen Zeitpunkten Signale aussenden, sodass der Datenbestand verzerrt wird, da die Signale einzelner Endgeräte zunächst als jeweils eine Fahrgastfahrt interpretiert werden. Daraus ergibt sich potentiell ein Fehlerwert für jede Quelle-Ziel-Beziehung, auf der ein Fahrgast auf Basis der WLAN- bzw. Bluetooth-Daten identifiziert wurde.

Ein Herausfiltern dieser Mehrfacherfassung einer Person ist unter den gegebenen Rahmenbedingungen nicht möglich, da in den erfassten Daten keine diesbezüglich nutzbaren Informationen enthalten sind.

Gleichzeitig können einzelne Geräte sowohl WLAN- als auch Bluetooth-Signale aussenden. Da sowohl der WLAN-Chip als auch der Bluetooth-Chip eigene MAC-Adressen besitzen, ergibt sich hier ein ähnliches Problem, sofern WLAN- und Bluetooth-Daten gemeinsam betrachtet werden.

3 Verfahren zur Ermittlung von Quelle-Ziel-Relationen auf Basis der Input-Daten

3.1 Einführung

In diesem Kapitel erfolgt die Beschreibung des Verfahrens, das auf den durch das technische System zur Datenerfassung (Kapitel 2.4) erfassten Daten aus den Fahrzeugen und Haltestellen basiert. Der grundlegende Verfahrensablauf, beginnend bei der automatischen Datenerfassung von WLAN- und Bluetooth-Signalen sowie den Fahrt- und AFZS-Daten mit dem Ziel der Erstellung von linienfahrtspezifischen Quelle-Ziel-Matrizen, wird in Abbildung 3-1 dargestellt.

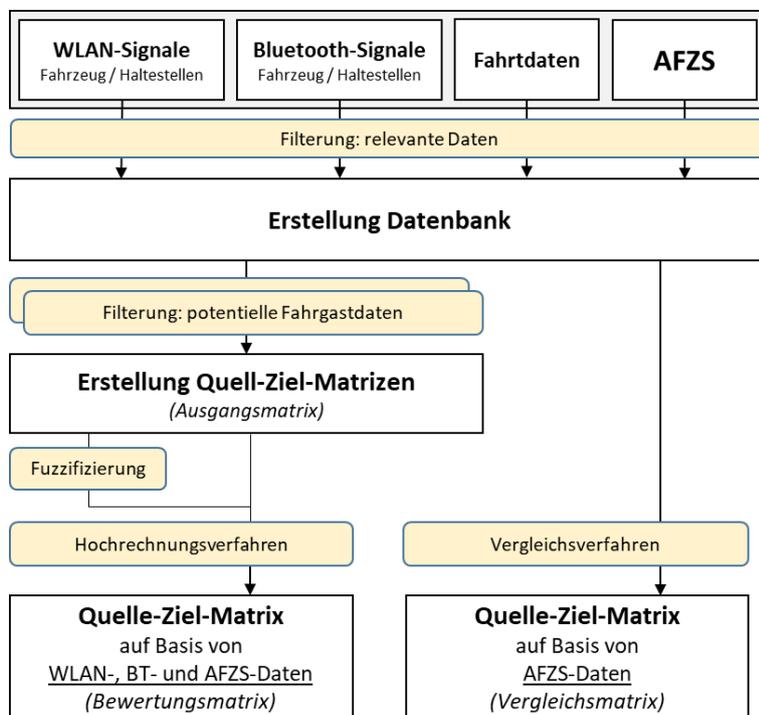


Abbildung 3-1: Überblick zum methodischen Vorgehen (Verfahren)

Die Datenaufzeichnung der WLAN- und Bluetooth-Daten beginnt in Fahrzeugen ab dem Zeitpunkt, zu dem das Fahrzeug gestartet wird, während Haltestellen durchgehend die Daten erfassen. Dabei werden alle Signale in Empfangsreichweite gespeichert. Die erfassten Signale werden in zwei Stufen gefiltert:

Stufe 1 filtert unplausible und nicht nutzbare Daten. Dazu gehören

- erfasste Daten, zu denen fehlerhafte Fahrtdaten zugeordnet wurden (Fahrtdaten, etwa die Journey-ID, sind für die Zuordnung von erfassten WLAN- und Bluetooth-Daten zu den korrekten Linienfahrten notwendig),
- erfasste Daten zu Zeitpunkten, in denen sich kein Fahrgast im Fahrzeug befindet,
- erfasste Daten mit randomisierter MAC-Adresse,
- Daten von Fahrten, in denen das AFZS die Ein- und Aussteiger nicht, bzw. nur mit geringer Qualität (hohe Wahrscheinlichkeit von Messfehlern) gezählt hat.

Die verbleibenden Daten werden in einer Datenbank gespeichert. Da alle Signale in Empfangsreichweite aufgezeichnet werden, ist es nicht zu vermeiden, dass neben den Signalen von Fahrgästen zu großen Teilen auch Signale von Nicht-Fahrgästen – etwa Fußgänger, Radfahrer, Pkw-Fahrer und –Mitfahrer oder auch Signale von Bewohnern oder Beschäftigten in der umliegenden Bebauung der Haltestelle bzw. des Linienwegs ebenfalls erfasst werden.

Aufbauend auf der Literaturrecherche und den Erkenntnissen aus den durchgeführten Untersuchungen (Kapitel 2.2.5.2, 2.3.2.3, 2.3.2.4) erfolgt in **Stufe 2** eine weitere Datenfilterung. Diese Filterung hat das Ziel, mit hoher Wahrscheinlichkeit Daten von Nicht-Fahrgästen herauszufiltern, etwa auf Basis der Luftlinienentfernung zwischen den Positionsdaten des Busses bei der ersten und letzten Erfassung eines Endgeräts. Dabei können false positives (Daten von Nicht-Fahrgästen, die als Fahrgäste interpretiert werden) und false negatives (Daten von Fahrgästen, die als Nicht-Fahrgäste interpretiert werden) nicht ausgeschlossen werden (Kapitel 3.2).

Die verbliebenen Daten werden in einer Quelle-Ziel-Matrix aufbereitet (Kapitel 3.3.2). Die Zeilen und Spalten bestehen aus den Haltestellen des entsprechenden Linienwegs, die Matrixelemente ergeben sich aus der Zuordnung des ersten und letzten erfassten Signals eines Endgeräts zu Haltestellen, die den potentiellen Ein- und Ausstiegsort markieren. Das Ergebnis dieses Modellschritts ist die Ausgangsmatrix und bildet die Basis für die Hochrechnung. Um Unsicherheiten bei der Festlegung der Ein- und Ausstiegshaltestelle zu berücksichtigen, wird darüber hinaus die Fuzzy-Set-Theorie angewandt, wodurch eine scharfe Zuteilung von Ein- und Ausstiegsort vermieden wird (Kapitel 3.3.3).

Da die zweite Stufe der Filterung mit Unsicherheiten verbunden ist, werden die gewählten Filter im Rahmen der Verfahrensbewertung (Kapitel 4.5) variiert. Ziel ist es, die Filtereinstellungen zu ermitteln, die zu Modellergebnissen mit der höchstmöglichen Güte führen.

Mit der Hochrechnung werden zunächst Verfahren zur Schätzung der Verkehrsströme zwischen Quellen und Zielen eines Untersuchungsgebiets auf Grundlage der vorliegenden Input-Daten (AFZS, WLAN, Bluetooth) vorgestellt und diskutiert (Kapitel 3.4). Im Ergebnis soll die Ausgangsmatrix mittels eines Steigerungsfaktorenmodells bzw. eines angepassten Steigerungsfaktorenmodells⁶ schrittweise so hochgerechnet werden, dass die Randsummen eingehalten werden. Randsummen bilden hier die Ein- und Aussteigerzahlen je Haltestelle, die mit hoher Genauigkeit durch das AFZS vorliegen. Das Ergebnis der Hochrechnung auf Basis der Ausgangsmatrix und der AFZS-Daten wird in dieser Arbeit als Bewertungsmatrix bezeichnet.

Der Forschungsansatz basiert auf der Verwendung von (ohnehin) anfallenden WLAN und Bluetooth-Daten der Fahrgäste. Diese Daten bilden die Eingangsgrößen der Ausgangsmatrix, die mittels AFZS-Daten und dem entsprechenden Verfahren hochgerechnet werden. Zur Überprüfung der Bedeutung der Eingangsdaten werden parallel zwei Hochrechnungsverfahren als Vergleichsverfahren verwendet:

- Verfahren nach Li-Cassidy, das lediglich auf den Randsummen (AFZS-Daten) basiert (Kapitel 3.4.2.1) und
- Fratar-Verfahren, dem eine Einser-Matrix⁷ zugrunde liegt.

Die softwaretechnische Umsetzung des Verfahrens wurde auf Grundlage der Vorgaben aus diesem Kapitel (sowie Kapitel 4.4) weitestgehend von der WVI GmbH umgesetzt. Die verschiedenen Verfahrensschritte werden durch vier sogenannte „Tools“ abgebildet (Tabelle 3-1).

Tabelle 3-1: Übersicht der Verfahrensschritte sowie zugehöriger „Tools“

Verfahrensschritt	Tool
Erstellung der Datenbank	<i>Import Tool</i>
Erstellung Quelle-Ziel-Matrizen (Ausgangsmatrix) mit verschiedenen Filtersets	<i>StauFi-Tool</i>
Quelle-Ziel-Matrix auf Basis von WLAN-, BT- und/oder AFZS-Daten (Bewertungsmatrix, Vergleichsmatrix)	<i>HR-Tool</i>

⁶ Das angepasste Steigerungsfaktorenmodell berücksichtigt die mathematische Schwäche von Steigerungsfaktorenmodellen bei Nullwerten.

⁷ Alle Quelle-Ziel-Relationen erhalten den gleichen Wert, z. B. 1.

Die Tools (Software Mobile Data Fusion 2022) sind dabei so programmiert, dass einzelne, für die Verfahrensschritte relevante Parameter (Fahrtauswahl, Auswahl Input-Datenquelle, Filtereinstellungen, weitere Einstellungen) verändert werden können. Da die Umsetzung auf einer relationalen Datenbank basiert, können die Ergebnisse bzw. benötigte Informationen zur Ausgabe flexibel zusammengestellt werden. Relational bedeutet dabei, dass verschiedene Datensätze über einen „Schlüssel“, z. B. eine eindeutige ID, die in allen Datensätzen geführt wird, miteinander verknüpft werden können⁸. Die Darstellung der Ergebnisse sowie deren Bewertung wird in Kapitel 4.6 erläutert.

3.2 Filterung der Rohdaten

Die Notwendigkeit zur Filterung der Rohdaten (WLAN- und Bluetooth-Daten) geht aus Abbildung 3-2 hervor. Die Gesamtmenge der erfassten Daten bezeichnet den Rohdatenbestand, der durch das technische System übermittelt wird. Hier sind alle WLAN- und Bluetooth-Signale aufgeführt, die während der Fahrzeit des Busses bzw. über den betrachteten Zeitraum durch die Haltestellen aufgezeichnet wurden. Dazu zählen die Signale von allen Endgeräten in der Reichweite des Empfängers. Somit werden sowohl die Daten von Endgeräten, die sich im Fahrzeug befinden (Fahrgastdaten), als auch von Endgeräten, die sich nicht im Fahrzeug befinden („Stördaten“), erfasst. Die Erfassung von unerwünschten Signalen („Stördaten“) kann nicht vermieden werden. Die Stördaten können durch Endgeräte von

- anderen Verkehrsteilnehmern, insb. wenn das Erhebungsfahrzeug steht (Haltestelle, Lichtsignalanlage (LSA) o. ä.),
- wartenden Personen bspw. an Haltestellen sowie
- Personen in angrenzenden Gebäuden der Fahrtroute des Fahrzeugs

ausgesendet werden.

Die folgenden Schritte zur Filterung beziehen sich auf die durch Fahrzeuge erfassten Daten. Die Ergebnisse der Filterung (Fahrgastdaten) bilden die Grundlage für die Erstellung der Quelle-Ziel-Matrizen und haben damit eine hohe Bedeutung für die weiteren Verfahrensschritte und das Ergebnis

⁸ Bekanntes Beispiel für relationale Datenbanken ist die Freeware *MySQL*.

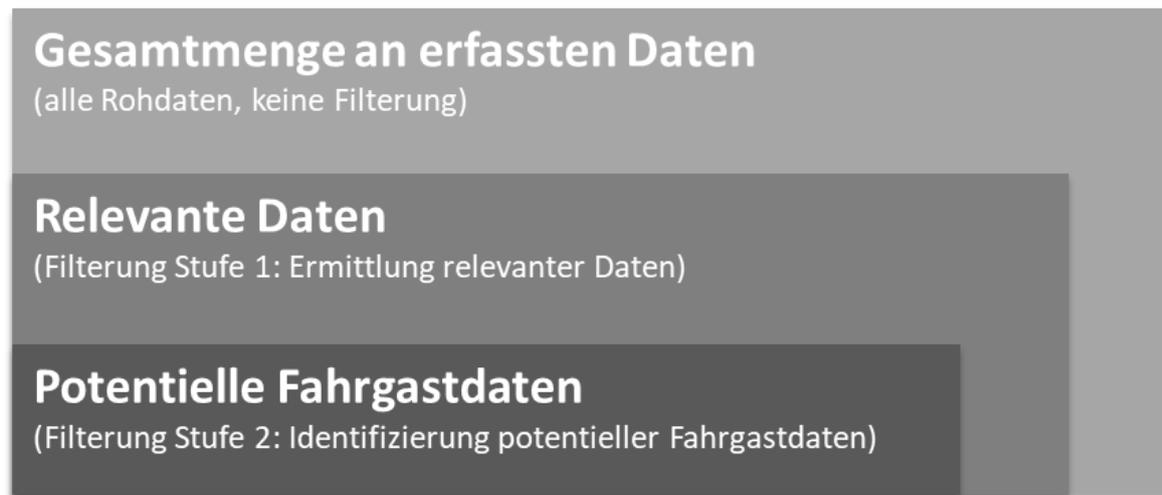


Abbildung 3-2: Skizze: Reduzierung der Rohdaten durch Filterstufen

3.2.1 Filterung Stufe 1: Ermittlung relevanter Daten

Aus der Gesamtmenge der durch die Fahrzeuge erfassten Daten sind in der ersten Stufe der Filterung zunächst die relevanten Rohdaten von den nicht nutzbaren Daten zu trennen. Nicht nutzbare Daten sind im Kontext des Verfahrens solche Daten, die auf Basis der Erfassungszeit, der Qualität der gespielten Daten (Fahrtinformationen, AFZS-Daten) oder weiterer Merkmale im Verfahren nicht genutzt werden können. Dazu zählen:

- **Fehlerhafte Zuordnung von Fahrtinformationen:** Eine Zuordnung der erfassten Daten zu einer Linienfahrt ist nicht möglich. Die Daten sind unbrauchbar.
- **Erfassung von Daten in Zeiträumen, in denen sich kein Fahrgast im Fahrzeug befindet (Basis: AFZS-Daten):** Die AFZS liefern den Belegungsgrad des Fahrzeugs mit hoher Genauigkeit. Befindet sich kein Fahrgast im Fahrzeug, können in logischer Konsequenz keine Daten zu durchgeführten Fahrgastfahrten erfasst werden. Die Daten zu dieser MAC-Adresse können somit für die gesamte betrachtete Linienfahrt verworfen werden.
- **Fehlende und fehlerhafte AFZS-Daten bzw. AFZS-Daten mit geringer Qualität:** Das AFZS überträgt die erfassten Daten und gibt eine Abschätzung der Genauigkeit ebendieser Daten auf einer Skala von 0 bis 100 an, die auf dem Umfang der notwendigen Korrekturen basiert. Liegen keine AFZS-Daten vor, fehlen für die späteren Verfahrensschritte die essentiellen Randsummen für die Hochrechnung (Kapitel 3.4). Haben die AFZS-Daten eine geringe Qualität, waren viele Korrekturen

notwendig, die mit einer erhöhten Unsicherheit einhergehen. Daher werden erfasste Daten zu Linienfahrten mit einer AFZS-Qualität von unter 80% verworfen.

- **Signale mit aktiver MAC-Randomisierung:** Es besteht keine sichere Methode, Signale mit aktiver MAC-Randomisierung (identifiziert am lokalen Bit, Kapitel 2.3.2.5) wiederholt einzelnen Endgeräten zuzuweisen. Der Anteil randomisierter MAC-Adressen an allen erfassten Signalen liegt zum Zeitpunkt der Softwareentwicklung bei weniger als 10% und ist somit vernachlässigbar. Daher werden nur die Signale betrachtet, bei denen keine MAC-Randomisierung vorliegt.

3.2.2 Filterung Stufe 2: Identifizierung potentieller Fahrgastdaten

Mit dem zweiten Schritt der Filterung werden aus den relevanten Rohdaten solche Daten extrahiert, die mit hoher Wahrscheinlichkeit eine durchgeführte Fahrgastfahrt abbilden (Fahrgastfahrt Daten).

Die Differenzierung von Fahrgastdaten gegenüber Stördaten wurde bereits in unterschiedlichen Studien untersucht und greift in der Regel auf Filtermethoden zurück, die die Fahrgastfahrt Daten anhand von Spannweiten einzelner Indikatoren identifizieren (Dunlap et al. 2016; Mishalani et al. 2016). Zu diesen Indikatoren gehören unter anderem

- die Dauer der Erfassung der MAC-Adressen,
- die empfangene Signalstärke,
- die Anzahl an empfangenen WLAN- bzw. Bluetooth-Datenpunkten (Probe Requests, Inquiry-Response-Nachrichten) oder
- die Entfernung zu der letzten bzw. nächsten Haltestelle für die erste und letzte Erfassung eines Datenpunkts einer MAC-Adresse.

Die Schwächen der bisherigen Forschungsarbeiten wurden bereits in Kapitel 2.3 beschrieben. Differenzen zwischen Modellergebnissen und realer Verteilung sind nach Pu et al. (2019) unter anderem auf harte (unflexible) Filterkriterien und Spannweiten zurückzuführen. Nach Paradedda et al. besteht daher weiterhin die Notwendigkeit einer anwendbaren Filtermethode, um Störsignale zu erkennen (Paradedda et al. 2019; Pu et al. 2019).

Basierend auf den bisherigen Erkenntnissen wurden Kriterien abgeleitet, die der Identifikation von Fahrgastfahrt Daten anhand der erfassten Daten einer Linienfahrt dienen (Tabelle 3-2). Die dargestellten Filter werden gemeinsam angewandt (logischer Operator: Und-Verknüpfung). Bei der Identifizierung

von Fahrtdaten wird jedoch (noch) nicht auf die durch Haltestellen sowie weitere Fahrzeuge erfassten Daten zurückgegriffen (Kapitel 3.3.3.3).

Die Wertebereiche sind dabei als flexibel zu betrachten und werden im Rahmen der Kalibrierung angepasst. Die in Tabelle 3-2 angegebenen Werte stellen den ersten Schritt der iterativen Anpassung dar. Zur Einordnung der Werte dienen die Häufigkeitsverteilungen in Tabelle 3-3 bis Tabelle 3-5. Diese Häufigkeitsverteilungen basieren auf einem Datensatz (relevante Daten, nach Filterung Stufe 1) mit allen durch die Fahrzeuge erfassten WLAN- und Bluetooth-Signalen im Zeitraum zwischen dem 15.09.2020 und 31.12.2020 (Dataset Bieland/Mobile Data Fusion 2022).

Tabelle 3-2: Kriterien zur Identifikation von Fahrgastdaten

Kriterium	Wertebereich	Beschreibung
GPS-Luftlinienentfernung je Linienfahrt	>100m	<p>Um eine Luftlinienentfernung >0m auf Basis der GPS-Koordinaten zu bestimmen, ist es zunächst notwendig, dass mindestens zwei Signale des gleichen Endgeräts zu unterschiedlichen Zeitpunkten und an unterschiedlichen Positionen empfangen wurden.</p> <p>Anhand der miterfassten GPS-Daten⁹ kann die Luftlinienentfernung zwischen dem ersten erfassten Signal und dem letzten erfassten Signal des Endgeräts bestimmt werden. Liegt die ermittelte Luftlinienentfernung (Reiseweite) oberhalb des Grenzwerts, ist die Wahrscheinlichkeit hoch, dass es sich hier um eine durchgeführte Fahrgastfahrt handelt.</p> <p>Der Grenzwert sollte dabei mindestens 100m betragen und bleibt als Parameter veränderbar, sodass auch Distanzen über 100m berücksichtigt werden können. Bei der Festlegung des Grenzwerts und der Interpretation ist zu beachten, dass die GPS-Koordinaten des Fahrzeugs zum Zeitpunkt der Erfassung eines Signals in der Datenbank abgelegt werden.</p> <p>Dabei ist nicht auszuschließen, dass die Signale durch einen anderen Verkehrsteilnehmer (Pkw) verursacht wurden, der parallel mit dem Erhebungsfahrzeug fährt (<i>false positive</i>).</p> <p>Dieses Vorgehen ermöglicht es grundsätzlich auch, stationäre Geräte – d. h. Geräte, die durchgehend in demselben abgegrenzten räumlichen Bereich empfangen werden – zu identifizieren und ggf. für die Gültigkeitsdauer der Hash-Funktion (Änderung je Kalendermonat) aus den Daten zu entfernen (Datenreduktion).</p>
Anzahl Linienfahrten mit der gleichen Linie pro Tag	<5	Die Durchführung von mehr als vier Linienfahrten mit der gleichen Linie wird als nicht realistisch angenommen. Werden MAC-Adressen bei mehr als vier Linienfahrten der gleichen Linie pro Tag erfasst, ist die Wahrscheinlichkeit hoch, dass es sich hier um keinen Fahrgast handelt, sondern von Signalen durch fest installierte Hardware in der Umgebung (etwa PC / Smartphones in der umgebenden Wohnbebauung des Linienwegs) oder verbaute Hardware im Fahrzeug.
Nur WLAN: Mittlere Signalstärke (RSSI) einzelner Endgeräte auf der Linienfahrt	> -90 dBm	<p>Die große Streuung des RSSI bei konstanter Entfernung zwischen Sender und Empfänger lässt keine eindeutige Spannweite des RSSI zur Abgrenzung von Signalen innerhalb bzw. außerhalb des Fahrzeugs zu. Dennoch haben Untersuchungen gezeigt (Kapitel 2.5.1.1), dass weder innerhalb noch außerhalb des Fahrzeugs Signale einzelner Endgeräte mit einem RSSI von weniger als -90 dBm erfasst wurden. Daher kann anhand dieses Extremwerts der Signalstärke gefiltert werden.</p> <p>Zu berücksichtigen ist dabei der Mittelwert (arithmetisches Mittel), sodass einmalige Ausreißer mit geringer Signalstärke eines Endgeräts nicht dazu führen, alle Signale derselben MAC-Adresse zu verwerfen.</p> <p>Der Grenzwert beträgt zunächst unter -90 dBm und bleibt als Parameter veränderbar, sodass auch engere Grenzwerte genutzt werden können.</p> <p>Dieses Kriterium ist nur für WLAN-Daten anwendbar, da bei Bluetooth-Daten keine Signalstärke erfasst wird.</p>
Zusätzliche Erfassung in weiterem Fahrzeug bzw. an Haltestelle (abhängig von Fahrplan) (nicht softwaretechnisch umgesetzt)	-	<p>Bei der Identifikation von Fahrgastfahrtdaten ist es notwendig, den Gesamtpool an möglichen Datenbeständen zu betrachten, um insbesondere false negatives infolge von nur einem erfassten Signal auf einer Linienfahrt zu vermeiden.</p> <p>Dazu sind die an Haltestellen sowie in weiteren Fahrzeugen erfassten Daten zu prüfen. Eine detaillierte Beschreibung zur Operationalisierung dieser Daten erfolgt im Rahmen der Erläuterung zur Erstellung von Quelle-Ziel-Matrizen in Kapitel 3.3.3.3.</p>

⁹ Die GPS-Daten weisen eine sehr hohe Qualität in Kassel auf. Sowohl dem NVV (Verkehrsunternehmen) als auch INIT (Hardware) sind keine Zeiträume mit geringer GPS Qualität bzw. fehlerhaften GPS-Daten bekannt (Rohde 2021).

Tabelle 3-3: Häufigkeitsverteilung der GPS- Luftlinienentfernung je MAC-Adresse und Linienfahrt

GPS-Luftlinienentfernung	Anzahl in Klasse	Anteil in Klasse	Anteil in Klasse (Werte >0m)
0 m	5.749.593	77%	
> 0 bis < 100 m	1.127.820	15%	65%
100 bis < 200 m	186.806	2%	11%
200 bis < 300 m	84.688	1%	5%
300 bis < 400 m	50.901	1%	3%
400 bis < 500 m	44.757	1%	3%
500 bis < 600 m	29.730	0%	2%
600 bis < 700 m	23.549	0%	1%
700 bis < 800 m	18.003	0%	1%
800 bis < 900 m	13.019	0%	1%
900 bis < 1000 m	13.342	0%	1%
>= 1000 m	140.481	2%	8%
Summe	7.482.689	100%	100%

Tabelle 3-4: Häufigkeitsverteilung des mittleren RSSI der WLAN-Signale je Linienfahrt

Mittlerer RSSI	Anzahl in Klasse	Anteil in Klasse	Anteil (kum.)
>= 0 dBm	78	0%	0%
< 0 bis > -10 dBm	3	0%	0%
-10 bis > -20 dBm	1.195	0%	0%
-20 bis > -30 dBm	7.736	0%	0%
-30 bis > -40 dBm	106.542	2%	2%
-40 bis > -50 dBm	223.466	4%	6%
-50 bis > -60 dBm	331.735	6%	13%
-60 bis > -70 dBm	329.412	6%	19%
-70 bis > -80 dBm	1.058.052	20%	39%
-80 bis > -90 dBm	1.871.351	36%	75%
-90 bis > -100 dBm	1.341.722	25%	100%
<= -100 dBm	1	0%	100%
Summe	5.271.293	100%	

Tabelle 3-5: Häufigkeitsverteilung der Anzahl an Linienfahrten je MAC-Adresse

Anzahl Fahrgastfahrten je Linie und Tag	Anzahl in Klasse	Anteil in Klasse	Anteil (kum.)
1	4.462.551	60%	60%
2	772.790	10%	70%
3	353.535	5%	75%
4	246.371	3%	78%
5	187.509	3%	80%
6	155.276	2%	83%
7	134.470	2%	84%
8	114.813	2%	86%
9	102.788	1%	87%
>9	952.586	12%	100%
Summe	7.482.689	100%	

Aus den Tabellen ist ablesbar, wie stark einzelne Filter den Datensatz reduzieren. Die große Anzahl an MAC-Adressen mit mehr als neun Linienfahrten pro Tag sind auf stationäre Geräte zurückzuführen, deren Signale von den Erfassungsfahrzeugen ggf. mehrfach pro Tag empfangen werden. Werte des mittleren RSSI „ ≥ 0 dBm“ deuten auf empfangene Signale von Geräten mit sehr hoher Sendeleistung hin.

Daten mit randomisierten MAC-Adressen werden bereits bei der ersten Stufe der Filterung entfernt. Sofern ein Tool besteht, das die randomisierten MAC-Adressen einzelnen Endgeräten zuweist und damit nutzbar macht, sind die beschriebenen Filter (Filterung Stufe 2) auch auf diese Daten anzuwenden.

3.2.3 Kritische Würdigung

Aufgrund unvollständiger Informationen bei der Erfassung der Daten können die Kriterien nicht dem Anspruch gerecht werden, alle Fahrtdaten bzw. ausschließlich Fahrtdaten auszuwählen. Daher besteht Möglichkeit von false negatives und false positives (Tabelle 3-6). False negatives bezeichnen hier Daten von Endgeräten, die in der Realität eine durchgeführte Fahrgastfahrt darstellen, jedoch aufgrund der Filter aus dem Datensatz entfernt werden. False positives bezeichnen analog Daten von Endgeräten, die keine Fahrgastfahrt durchgeführt haben, aber als Fahrgastfahrt interpretiert wurden und somit im Datensatz verbleiben.

Tabelle 3-6: False Positives, False Negatives bei der Identifizierung von Fahrgastdaten

	als Fahrgastdatensatz eingestuft	als Stördatensatz eingestuft
tatsächlicher Fahrgastdatensatz	True positive	False negative
tatsächlicher Stördatensatz	False positive	True negative

False positives: Eine falsche Zuordnung von Daten zu den Fahrtdaten kann insbesondere durch Verkehre auftreten, die parallel mit der betrachteten Linienfahrt erfolgen bzw. durch Endgeräte, die sich im Verlauf der Linienfahrt wiederholt – ggf. auch nur für kurze Dauern – in der Nähe der Erfassungsinfrastruktur befinden und dabei erfasst werden. Dies betrifft vor allem Verkehrsteilnehmer in Pkw sowie Radfahrer. Auch Fußgänger können bei entsprechender Verkehrssituation (lange Haltezeiten an Lichtsignalanlagen, geringe Geschwindigkeit des ÖPNV-Fahrzeugs aufgrund einer hohen Belastung der Verkehrsinfrastruktur) false positives auslösen.

False negatives: Die fehlerhafte Entfernung von Datensätzen, die grundsätzlich eine Fahrgastfahrt repräsentieren, kann insbesondere durch Geräte erfolgen, die nur einmal erfasst werden bzw. deren GPS-Luftlinienentfernung unterhalb des Grenzwerts liegt. Entsprechend den Analysen des Sendeverhaltens kann dies insbesondere für Kurzstreckenfahrten auftreten, aber auch für längere Linienabschnitte, sofern das Endgerät nur in großen Zeitabständen Signale aussendet oder nach dem ersten ausgesendeten Signal abgeschaltet wird. Für diese Endgeräte beträgt die GPS-Luftlinienentfernung 0 m, sodass sie entsprechend eines beispielhaften Grenzwerts von 100 m aus dem Datensatz gefiltert werden. Die Erfassung von nur einem Signal eines Endgeräts im Fahrzeug bietet insgesamt zu wenig Informationen, um bei der Erstellung von Quelle-Ziel-Matrizen berücksichtigt zu werden. Darüber hinaus ist die Wahrscheinlichkeit hoch, dass es sich hier um Stördaten anderer Verkehrsteilnehmer handelt.

Die Ergebnisse dieses Schritts beeinflussen maßgeblich die weiteren Verfahrensschritte (Erstellung der Quelle-Ziel-Matrix) und damit die Ergebnisse der Hochrechnung. Eine funktionierende Filterung führt dazu, relevante Daten von nicht relevanten Daten für die weiteren Verfahrensschritte zu trennen.

3.3 Erstellung von Quelle-Ziel-Matrizen

3.3.1 Definition: Quelle-Ziel-Matrizen

Das Vorliegen von Quelle-Ziel-Matrizen (QZ-Matrix, engl.: Origin-Destination-Matrix, OD-Matrix) ist in der Verkehrsplanung von hoher Bedeutung. In Quelle-Ziel-Matrizen werden neben dem Mengengerüst der betrachteten Fahrgastfahrten sowohl die Quelle (Verkehrserzeugung), als auch das Ziel der einzelnen Fahrgastfahrten (Zielwahl), meist aggregiert auf Ebene von Verkehrszellen, dargestellt. Unter Berücksichtigung dieser Informationen werden in der weiteren Modellierung die Verkehrsmittel- und Routenwahl durchgeführt.

Eine Quelle-Ziel-Matrix zeigt alle Verkehrsbeziehungen zwischen Quellen und Zielen der jeweiligen Verkehrsart (Personen, ÖPNV-Fahrgäste, MIV-Fahrer etc.) für einen Raum (Abrahamsson 1998). Der Untersuchungsraum wird in der Modellierung häufig in Verkehrszellen unterteilt, kann auf Grundlage der Eingangsdaten insbesondere für ÖPNV-Fahrgäste aber auch auf Ebene der Haltestellen untergliedert werden. In der Regel sind Quelle-Ziel-Matrizen quadratisch (N x N), das heißt, dass N Quellen und N Ziele im Untersuchungsraum vorhanden sind (Abbildung 3-3).

nach Ziel D von Quelle O	1	2	3	...	d	...	n	$\sum_{d=1}^n d_{od} = O_o$
1	d_{11}	d_{12}	d_{13}	...	d_{1d}	...	d_{1n}	O_1
2	d_{21}	d_{22}	d_{23}	...	d_{2d}	...	d_{2n}	O_2
3	d_{31}	d_{32}	d_{33}	...	d_{3d}	...	d_{3n}	O_3
...
o	d_{o1}	d_{o2}	d_{o3}	...	d_{od}	...	d_{on}	O_o
...
n	d_{n1}	d_{n2}	d_{n3}	...	d_{nd}	...	d_{nn}	O_n
$\sum_{o=1}^n d_{od} = D_d$	D_1	D_2	D_3	...	D_d	...	D_n	$\sum_{o=1}^n \sum_{d=1}^n d_{od}$

Abbildung 3-3: Aufbau einer Quelle-Ziel-Matrix

In den Matrixelementen d_{od} werden die Personenfahrten von der Quelle o (origin) zum Ziel d (destination) angegeben. Die Summe einer Zeile ergibt die Anzahl an Personenfahrten mit der Quelle o, die Summe der Spalten beschreibt die Anzahl der Personenfahrten mit dem Ziel d. Die Matrizen können

bei Vorliegen entsprechender Daten nach verschiedenen Kriterien differenziert werden, wie z. B. Verkehrsmittel, Wegezweck und Personengruppe.

3.3.2 Bildung von Quelle-Ziel-Matrizen auf Basis der gefilterten Daten

Aufbauend auf den Ergebnissen der Filterung liegt ein Datensatz vor, aus dem entsprechend der gewählten Filtereinstellungen die Einträge entfernt wurden, die als Stördaten eingestuft wurden (z. B. einmalige Erfassung eines Endgeräts, sehr kurze GPS-Distanz o.ä.). Im vorliegenden Datensatz sind somit allein die Daten enthalten, auf deren Grundlage die Quelle-Ziel-Matrix der Linienfahrt abgeleitet werden soll.

3.3.2.1 Ermittlung des Ein- und Ausstiegsortes je Linienfahrt

Die Bildung von Quelle-Ziel-Matrizen auf Basis der gefilterten Daten ordnet die Einstiegs- und Ausstiegshaltestelle entsprechend dem Zeitpunkt der Erfassung des ersten bzw. letzten Signals eines Endgeräts (=MAC-Adresse) zu. Dabei wird für jede einzelne Linienfahrt eine Quelle-Ziel-Matrix erstellt. Die Zuordnung der Daten erfolgt dabei über die „JOURNEY_ID“, durch die jede durchgeführte Linienfahrt im Untersuchungsraum eindeutig identifiziert werden kann.

Der Ablauf zur Bildung der Quelle-Ziel-Matrix erfolgt in drei Schritten:

(1) Erstellung einer leeren $N \times N$ -Matrix (N = Anzahl an Haltestellen der Linienfahrt) über alle Haltestellen der entsprechenden Linienfahrt. Die AFZS-Daten dienen hier als Datengrundlage für die Erstellung der leeren Matrix. In den AFZS-Daten sind sowohl der Linienverlauf sowie die SOLL- und IST-Fahrplandaten, d. h. die wahren Zeitpunkte der Haltestellenan- und -abfahrten, gespeichert.

Es ist zu berücksichtigen, dass bei Betrachtung aller Linienfahrten einer Buslinie innerhalb eines definierten Zeitintervalls sowohl

- unterschiedliche Start- und Endpunkte, als auch
- Abweichungen des Linienwegs auftreten können (z. B. Anfahrt zusätzlicher Haltestellen während Spitzenstunden oder keine Bedienung einzelner Haltestellen).

Daher ist die Matrix für jede Linienfahrt neu zu erstellen bzw. zu prüfen.

(2) Die Zuordnung der Ein- und Ausstiegshaltestelle erfolgt auf Basis des Zeitpunkts der ersten und letzten Erfassung eines Signals eines Endgeräts. Aus logischen Gründen kann die Ausstiegshaltestelle immer nur nach der

Einstiegshaltestelle liegen, d. h. der Fahrgast bzw. das Endgerät hat mindestens eine Fahrgastfahrt zwischen zwei Haltestellen im Fahrzeug zurückgelegt.

Bei der Zuordnung der Haltestellen ist zu beachten, dass der Halt eines Fahrzeugs an einer Haltestelle in der Regel aus der Ankunft des Fahrzeugs, der Haltezeit und der Abfahrt besteht. Ausnahmen bilden hier sogenannte „Durchfahrten“, bei denen kein Halt stattfindet, da kein Zu- oder Ausstieg (leere Haltestelle und keine Anforderung eines Fahrzeugstopps) erfolgt. Die Haltezeit variiert in Abhängigkeit des Fahrplans sowie der Fahrgastnachfrage. Die Haltezeit beträgt im NVV über alle Linien im Mittel acht Sekunden, für die Linie 500 im Mittel 22 Sekunden (Dataset Mobile Data Fusion 2020). In Einzelfällen kann die Haltezeit mehr als eine Minute betragen (etwa zum Ausgleich einer Verfrühung).

Für Zeitpunkte des ersten oder letzten Signals außerhalb der Haltezeit wird die Zuordnung der Ein- und Ausstiegshaltestelle in trivialer Weise vorgenommen:

- Als potentielle **Einstiegshaltestelle** wird die letzte angefahrene Haltestelle vor der ersten Erfassung festgehalten (Abbildung 3-4 (oben)). Später angefahrene Haltestellen kommen als Einstiegsort nicht in Betracht. Andersherum besteht jedoch die Möglichkeit, dass der Fahrgast bzw. das Endgerät bereits früher in das Fahrzeug gestiegen ist.
- Als potentielle **Ausstiegshaltestelle** wird die nächste angefahrene Haltestelle nach der letzten Erfassung festgehalten (Abbildung 3-5 (oben)). Früher angefahrene Haltestellen kommen als Ausstiegsort nicht in Betracht. Andersherum besteht jedoch die Möglichkeit, dass der Fahrgast bzw. das Endgerät noch längere Zeit im Fahrzeug war und erst an einer späteren Haltestelle ausgestiegen ist.

Sofern der Zeitpunkt des ersten oder letzten Signals jedoch in der Haltezeit liegt, müssen die Bedingungen anhand der An- und Abfahrtszeit des Fahrzeugs konkretisiert werden. Dazu werden folgende zusätzlichen Annahmen getroffen (Abbildung 3-4 (unten), Abbildung 3-5 (unten)):

- Für die Bestimmung der Einstiegshaltestelle wird angenommen, dass bei Erfassung während eines Halts die gesamte Haltezeit für einen möglichen Einstieg berücksichtigt wird (Abbildung 3-4 (unten), Formel 3-1) (Ankunftszeitpunkt relevant).

Unsicherheiten ergeben sich dadurch, dass aufgrund der Reichweite von Signalen bereits während der Anfahrt (bzw. vor dem Öffnen der Türen) ein Signal erfasst werden kann, obwohl das Endgerät sich noch nicht im Fahrzeug befindet. In dem Fall würde die vorherige Haltestelle als Einstiegshaltestelle zugeordnet werden.

- Für die Bestimmung der Ausstiegshaltestelle wird angenommen, dass bei Erfassung während eines Halts die gesamte Haltezeit für einen möglichen Ausstieg berücksichtigt wird (Abbildung 3-5 (unten), Formel 3-2) (Abfahrtszeitpunkt relevant).

Unsicherheiten ergeben sich dadurch, dass aufgrund der Reichweite von Signalen auch nach der Abfahrt (bzw. nach Schließen der Türen) noch ein Signal erfasst werden kann, obwohl das Endgerät das Fahrzeug bereits verlassen hat. In dem Fall würde die nächste Haltestelle als Ausstiegshaltestelle zugeordnet werden.

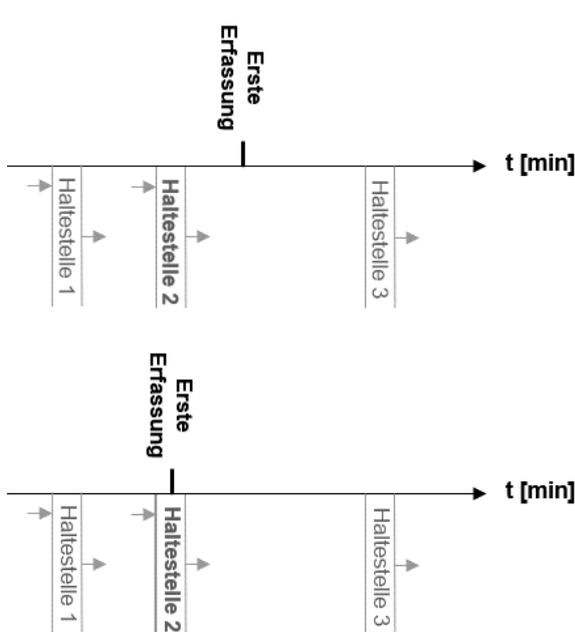


Abbildung 3-4: Skizze zur Verdeutlichung des Bezugszeitpunkts zur Zuordnung der Einstiegshaltestelle

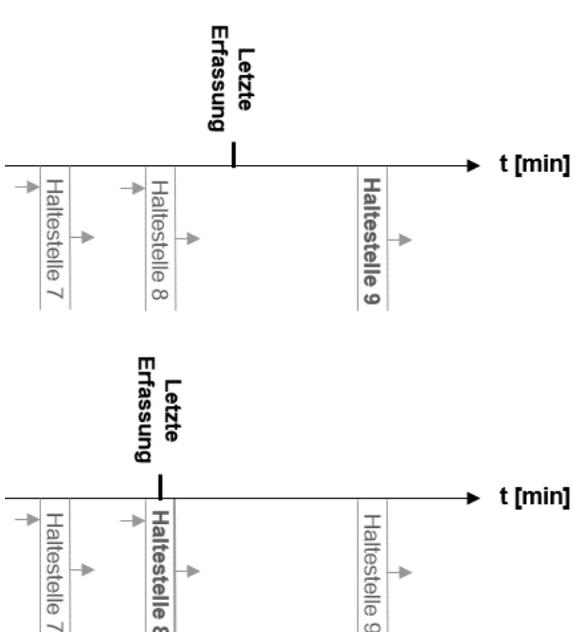


Abbildung 3-5: Skizze zur Verdeutlichung des Bezugszeitpunkts zur Zuordnung der Ausstiegshaltestelle

$$t(S^Q) \geq t(HS_{An,o}) \wedge t(S^Q) < t(HS_{An,o+1}) \tag{Formel 3-1}$$

mit

$t(S^Q)$ Zeitpunkt des ersten Signals für die Bestimmung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
 $t(HS_{An,o})$, Zeitpunkt der Ankunft (An) an der aktuellen (o) bzw. nächsten (o+1) Haltestelle
 $t(HS_{An,o+1})$

$$t(S^Z) \leq t(HS_{Ab,d}) \wedge t(S^Z) > t(HS_{Ab,d-1}) \tag{Formel 3-2}$$

mit

$t(S^Z)$ Zeitpunkt des letzten Signals für die Bestimmung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
 $t(HS_{Ab,d})$, Zeitpunkt der Abfahrt (Ab) an der aktuellen (d) bzw. vorherigen (d-1) Haltestelle
 $t(HS_{Ab,d-1})$

(3) Im dritten Schritt werden die Einzelauswertungen von MAC-Adressen (Abbildung 3-6) aufsummiert, sodass im Ergebnis die Quelle-Ziel-Matrix der Linienfahrt vorliegt (Abbildung 3-7).

	Haltestelle 1								
	Haltestelle 2								
	Haltestelle 3								
	Haltestelle 4								
	Haltestelle 5								
	Haltestelle 6								
	Haltestelle 7								
	Haltestelle 8								
	Haltestelle 9								
Haltestelle 1	-								
Haltestelle 2	-	-							
Haltestelle 3	-	-	-						
Haltestelle 4	-	-	-	1					
Haltestelle 5	-	-	-	-					
Haltestelle 6	-	-	-	-	-				
Haltestelle 7	-	-	-	-	-	-			
Haltestelle 8	-	-	-	-	-	-	-		
Haltestelle 9	-	-	-	-	-	-	-	-	

Abbildung 3-6: Bestimmung der Quelle-Ziel-Relation einer MAC-Adresse

	Haltestelle 1								
	Haltestelle 2								
	Haltestelle 3								
	Haltestelle 4								
	Haltestelle 5								
	Haltestelle 6								
	Haltestelle 7								
	Haltestelle 8								
	Haltestelle 9								
Haltestelle 1	-	4	3	1	0	0	2	1	0
Haltestelle 2	-	-	0	5	3	0	1	2	8
Haltestelle 3	-	-	-	0	1	1	0	8	0
Haltestelle 4	-	-	-	-	1	0	1	0	3
Haltestelle 5	-	-	-	-	-	1	0	5	0
Haltestelle 6	-	-	-	-	-	-	2	3	1
Haltestelle 7	-	-	-	-	-	-	-	-	1
Haltestelle 8	-	-	-	-	-	-	-	-	0
Haltestelle 9	-	-	-	-	-	-	-	-	-

Abbildung 3-7: Summe der Quelle-Ziel-Relationen einer Linienfahrt über alle MAC-Adressen nach Filterung

3.3.2.2 Theoretische Überlegungen zur Präzisierung der Ein- und Ausstiegshaltestelle

In diesem Unterkapitel werden theoretische Überlegungen beschrieben, die auf Basis der vorliegenden Daten die Genauigkeit der Bestimmung der Ein- und Ausstiegshaltestelle erhöhen können. Die Überlegungen sind theoretisch, da diese nicht in der softwaretechnischen Umsetzung, die dieser

Dissertation zugrunde liegt, berücksichtigt wurden. Sie stellen damit mögliche Weiterentwicklungen dar.

Einbezug von Haltestellendaten

Das dargestellte Vorgehen zur Bestimmung der Quelle-Ziel-Relationen auf Basis der gefilterten Daten berücksichtigt lediglich die durch die Fahrzeuginfrastruktur erfassten Daten. Unberücksichtigt sind dabei Daten, die durch die Ausstattung der Haltestellen erfasst werden. Haltestellendaten bieten grundsätzlich die Möglichkeit, sowohl die Einstiegshaltestelle als auch die Ausstiegshaltestelle zu präzisieren:

Einstiegshaltestelle: Wird eine MAC-Adresse

- an einer Haltestelle erfasst, bevor das betrachtete Linienfahrzeug diese Haltestelle erreicht hat sowie
- im späteren Zeitverlauf erneut durch das Fahrzeug erfasst

kann daraus die Schlussfolgerung gezogen werden, dass der Fahrgast mit sehr hoher Wahrscheinlichkeit an ebendieser Haltestelle zugestiegen ist. Die Einstiegshaltestelle ist entsprechend der zusätzlichen Informationen anzupassen.

Ausstiegshaltestelle: Wird eine MAC-Adresse, nachdem diese in einem Linienfahrzeug das letzte Mal erfasst wurde, danach an einer der folgenden Haltestellen der Linienfahrt erfasst, ist zu prüfen, ob es sich hier um einen potentiellen späteren Ausstieg handelt. Dazu sind im Idealfall die IST-Fahrplandaten der Linienfahrt zu berücksichtigen: Fand die Erfassung an der Haltestelle in einem engen Zeitintervall (etwa 30 Sekunden) um die An- und Abfahrtszeit des Linienfahrzeugs an ebendieser Haltestelle statt, kann daraus die Schlussfolgerung gezogen werden, dass der Fahrgast mindestens bis zu dieser Haltestelle weitergefahren ist. Die Ausstiegshaltestelle ist ggf. entsprechend der zusätzlichen Informationen anzupassen.

Zur Verdeutlichung der Zusatzinformation durch die Ausstattung von Haltestellen dient Abbildung 3-8.

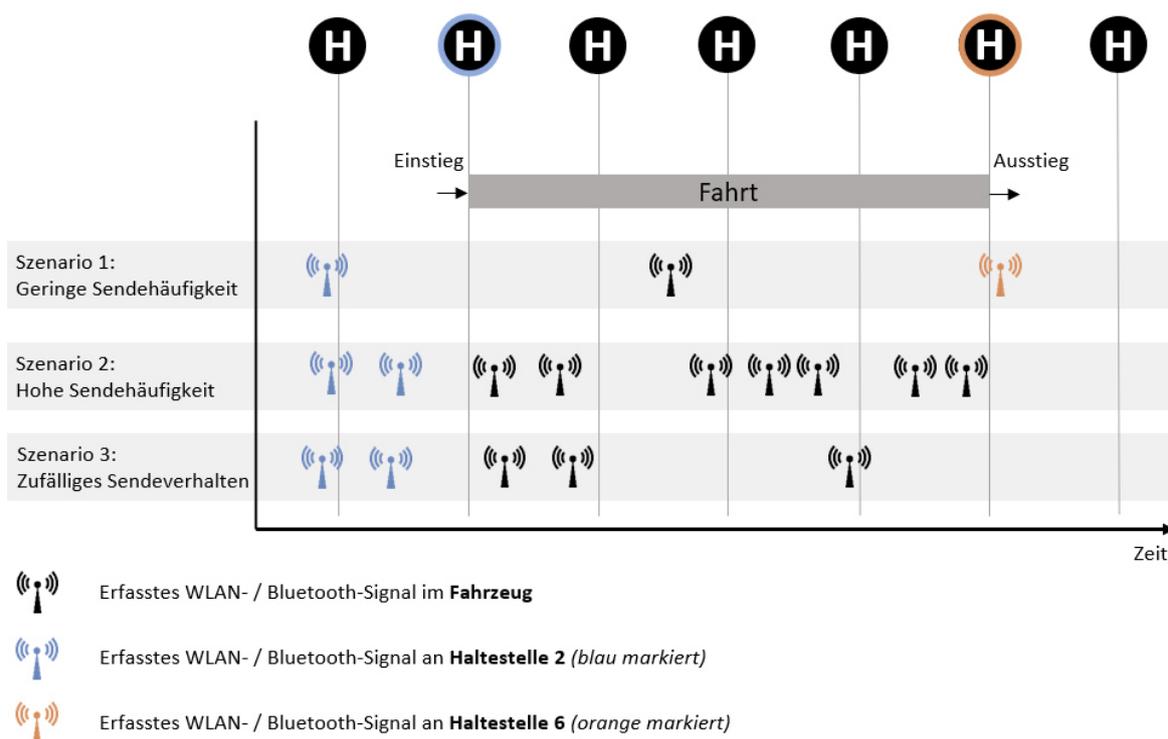


Abbildung 3-8: Modellhafte Darstellung von erfassten Signalen während einer Linienfahrt im Fahrzeug sowie an Haltestellen

Mit der Ausstattung von Haltestellen kann die Wartezeit der Fahrgäste genutzt werden, um Signale der Endgeräte zu empfangen und damit insbesondere die Erfassungsgenauigkeit für die Einstiegshaltestelle zu steigern. Dargestellt sind hier neben den durch das Fahrzeug erfassten Signalen (schwarz) auch die an der zweiten Haltestelle erfassten Signale (blau). Sofern Signale des gleichen Endgeräts erst an einer Haltestelle und im späteren Verlauf im Fahrzeug erfasst wurden, kann die Einstiegshaltestelle genau angegeben werden. Dabei sind die Fahrplandaten zu beachten, sodass der Erfassungszeitpunkt an der Haltestelle und die Wiedererfassung im Fahrzeug eine Linienfahrt durch einen Fahrgast abbilden. Daneben besteht die Möglichkeit, dass die Signale an Haltestelle und im Fahrzeug durch andere Verkehrsteilnehmer, wie etwa parallel fahrende Radfahrer, ausgelöst werden.

Die Erfassungsgenauigkeit der Ausstiegshaltestelle kann ebenfalls durch die Ausstattung der Haltestellen gesteigert werden: Wird ein Signal eines Endgeräts an einer Haltestelle des Linienwegs nach der letzten Erfassung durch das Fahrzeug erfasst (orange), kann die Angabe der Zielhaltestelle verfeinert werden. Dabei ist jedoch zu beachten, dass Fahrgäste an der Zielhaltestelle – im Gegensatz zur Starthaltestelle – keine Wartezeiten haben und sich ggf.

schnell aus der Empfangsreichweite entfernen. Auch hier muss ein Abgleich mit den Fahrplandaten stattfinden.

Einbezug von weiteren Fahrzeugdaten

Das dargestellte Vorgehen zur Bestimmung der Quelle-Ziel-Relationen auf Basis der gefilterten Daten berücksichtigt lediglich die durch das Fahrzeug der betrachteten Linienfahrt selbst erfassten Daten. Unberücksichtigt sind dabei noch Daten, die durch weitere Fahrzeuge erfasst werden.

Signaldata aus weiteren Fahrzeugen bieten die Möglichkeit, die Ein- und Ausstiegshaltestelle aller betrachteten Fahrgastfahrten zu präzisieren. Zur Verdeutlichung der folgenden Ausführungen dient Abbildung 3-9. In der Abbildung sind zwei fiktive Linien (Linie 1 und Linie 2) sowie die Haltestellen mit Nummerierung entsprechend des Fahrplans dargestellt. Die Umstiegshaltestelle (Linie 1: Haltestelle 5 / Linie 2: Haltestelle 4) ist gesondert markiert. Darüber hinaus sind die erste und letzte Erfassung des Signals eines Endgeräts auf der jeweiligen Linienfahrt dargestellt.

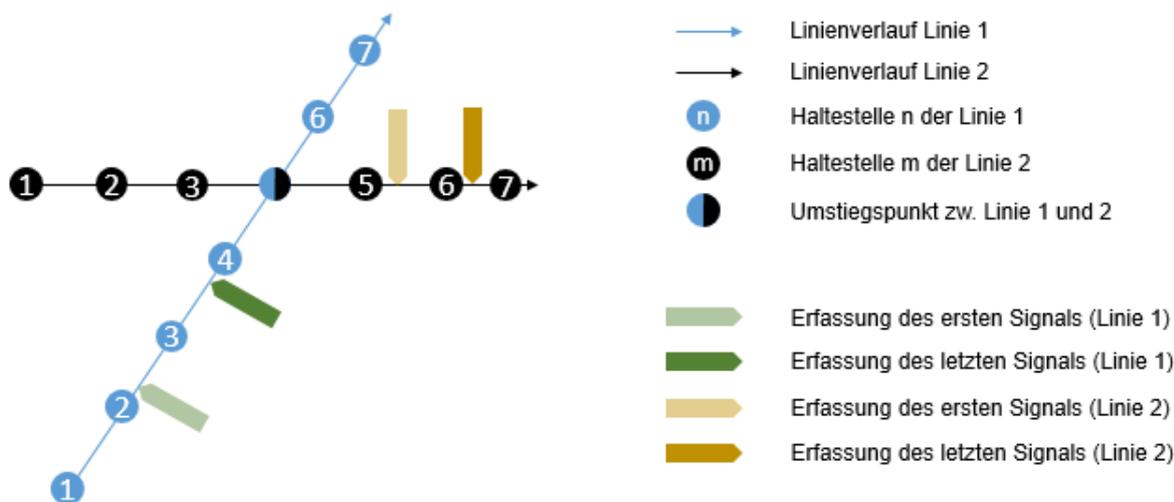


Abbildung 3-9: Skizze zur Präzisierung von Ein- und Ausstiegshaltestelle auf Basis von weiteren Fahrzeugdaten

Dem Beispiel liegt folgende Linienfahrt zugrunde:

- Einstieg an Haltestelle 2 der Linie 1, Ausstieg (Umstieg) an Haltestelle 5 und
- Einstieg (Umstieg) an Haltestelle 4 der Linie 2, Ausstieg an Haltestelle 7.

Aus den Signalen würden entsprechend dem beschriebenen, linienfahrtspezifischen Vorgehen zwei Linienfahrten abgeleitet werden:

- für Linie 1 eine Fahrgastfahrt von Haltestelle 2 nach Haltestelle 4 und
- für Linie 2 eine Fahrgastfahrt von Haltestelle 5 nach Haltestelle 7.

Die Kombination der Informationen der beiden Linienfahrten wird dabei bislang nicht genutzt, kann aber zur Erhöhung der Genauigkeit beitragen.

Die MAC-Adresse wird zunächst auf Linie 1 erfasst und in späteren Verlauf auf einer weiteren Linienfahrt (Linie 2). Sofern die Bedingungen erfüllt sind, dass

- die beiden betrachteten Linienfahrten eine gemeinsame (Umstiegs-)Haltestelle besitzen und
- die Zeitpunkte der Erfassung in Abgleich mit dem IST-Fahrplandaten¹⁰ von Linie 1 und Linie 2 auf einen direkten Umstieg von Linie 1 zu Linie 2 hinweisen,

ist mit hoher Wahrscheinlichkeit davon auszugehen, dass ein Umstieg erfolgt ist. Unter direktem Umstieg wird hier die Nutzung des nächsten Fahrzeugs an der Umstiegshaltestelle verstanden, das zu der (angenommenen) Zielhaltestelle führt.

Daraus können folgende Erkenntnisse für die beiden Linienfahrten abgeleitet werden:

- Fahrgast fährt auf Linie 1: Keine zusätzlichen Informationen zur Quelle der Fahrgastfahrt. Das Ziel der Fahrgastfahrt kann jedoch präzisiert werden zu Haltestelle 5.
- Fahrgast fährt auf Linie 2: Die Quelle der Fahrgastfahrt kann präzisiert werden zu Haltestelle 4. Zu dem Ziel der Fahrgastfahrt liegen keine weiteren Informationen vor.

In dem Beispiel wurde eine Fahrgastfahrt mit zwei Linien bzw. einem Umstieg erläutert. Das Vorgehen ist analog für weitere Linien bzw. Umstiege erweiterbar.

3.3.2.3 Kritische Würdigung

Bei der Anwendung des beschriebenen Vorgehens sind folgende Punkte zu beachten:

¹⁰ Unter Berücksichtigung von Unschärfen sind auch SOLL-Fahrplandaten nutzbar.

Unsicherheit bei der Festlegung der Ein- und Ausstiegshaltestelle: Für die Verfahrensentwicklung ist festzuhalten, dass der Ein- und Austrittszeitpunkt eines Fahrgastes auf Grundlage des ersten bzw. letzten erfassten Signal eines Endgeräts nicht bzw. nur näherungsweise durch das vorgestellte Vorgehen bzw. durch die vorliegenden Daten bestimmt werden kann. Somit liegt eine Unsicherheit hinsichtlich der Ein- und Ausstiegshaltestelle vor, die im Sendeverhalten der Endgeräte begründet ist.

Dabei kann auch die Lage der Haltestelle von Bedeutung sein: In urbanen Räumen sind die Abstände zwischen Haltestellen räumlich und zeitlich geringer als in ländlichen Räumen (Haltestellenabstand, Fahrzeit zwischen Haltestellen).

Fehler in der Datengrundlage zur Erstellung der Quelle-Ziel-Matrizen: Bei der Interpretation der Quelle-Ziel-Matrizen ist zu berücksichtigen, dass

- die Datengrundlage nur eine Teilmenge der Fahrgastfahrten abbildet,
- trotz Einsatz verschiedener Filter weiterhin Stördaten in der Datengrundlage vermutet werden können (false positives) und
- trotz Einsatz verschiedener Filter Daten, die grundsätzlich als Fahrgastdaten einzuordnen wären, bei der Erstellung der Quelle-Ziel-Matrizen unberücksichtigt bleiben (false negatives).

Teilmenge der Fahrgastfahrten: Die Anzahl mitgeführter Endgeräte eines Fahrgasts kann zwischen 0 und theoretisch einer Vielzahl von WLAN- und Bluetooth-fähigen Endgeräten schwanken, lässt sich nicht vorhersagen und führt zu unbekanntem Verzerrungen der Datengrundlage. Führt ein Fahrgast kein Endgerät mit sich, können entsprechend keine WLAN- bzw. Bluetooth-Signale zu dieser Fahrgastfahrt erfasst werden. Ebenso besteht die Möglichkeit, dass insbesondere Kurzstreckenfahrten bzw. Fahrgastfahrten mit geringer Reisedauer nicht erfasst werden, da einige Endgeräte einen mittleren Abstand zwischen Signalen von mehreren Minuten aufweisen.

Unvollständige Erfassung der Signale: Grundsätzlich ist es möglich, eine größere Anzahl an Signalen eines Endgeräts zu empfangen. Dies hat Waltari und Kangasharju (2016) bereits mit dem „Wireless Shark“ gezeigt, indem 13 WLAN-Adapter genutzt wurden, um alle Kanäle des 2,4 GHz-Netzes abzu hören. Dabei wurde u. a. festgestellt, dass in den meisten Fällen alle Kanäle bei einem Probe Requests Burst (WLAN) genutzt werden. Dennoch gibt es Unterschiede in der Häufigkeit der empfangenen Probe Requests auf den einzelnen Kanälen – d. h. einzelne Kanäle werden häufiger genutzt. Es bleibt

jedoch unklar, ob durch diese Mehrausstattung – bzw. ergänzend auch für alle Kanäle des 5-GHz-Netzes sowie der Bluetooth-Kanäle – und die damit verbundene größere Anzahl an erfassten Signalen ein signifikanter Mehrwert für die Erfassungsgenauigkeit hervorgeht. Ein Mehrwert wird nur dann erzielt, wenn zu zusätzlichen Zeitpunkten, die nicht durch den bereits genutzten Adapter erfasst wurden, Signale eines Endgeräts erfasst werden können (Kapitel 2.4.1).

Betrachtung auf Ebene der Linienfahrt: Das Verfahren zielt darauf ab, Quelle-Ziel-Matrizen für einzelne Linienfahrten abzuleiten. Die linienfahrtspezifische Betrachtung für die ausgewählten Linien im Rahmen dieser Dissertation führt dazu, dass Verbundbeförderungsfälle, die ggf. aus mehreren Linienbeförderungsfällen mit Umstiegen bestehen, nicht abgebildet werden. Dies kann insbesondere bei der Verwendung der Matrizen für die Angebotsplanung zu Interpretationsschwierigkeiten führen, da die angegebene Quelle und das Ziel auf Linienfahrebene nicht in allen Fällen (Umsteiger) die Quelle und das Ziel des Weges des Fahrgasts widerspiegeln. Bei Erfassung der Input-Daten in allen Fahrzeugen in einem Untersuchungsraum kann für den Binnenverkehr des Untersuchungsraums eine Abbildung von Umstiegen erfolgen. Der ein- und ausgehende Verkehr (Bezug Untersuchungsraum, z. B. NVV-Gebiet) kann nicht bis zu seiner Quelle bzw. dem Ziel zurückverfolgt werden.

3.3.3 Bildung von Quelle-Ziel-Matrizen unter Berücksichtigung von Unschärfen

Anhand der Zeitpunkte der ersten und letzten Erfassung eines Signals kann (theoretisch) eine eindeutige Quelle-Ziel-Relation abgeleitet werden. Vor dem Hintergrund der bereits beschriebenen Unschärfen bei der Erfassung der Ein- und Ausstiegshaltestelle (Kapitel 2.5.3) ist diese eindeutige Zuordnung jedoch nicht zweifelsfrei dazu geeignet, die tatsächlich durchgeführten Fahrgastfahrten exakt abzubilden. Die aus dem Sendeverhalten resultierenden Unschärfen des Ein- und Ausstiegsorts werden durch die Fuzzy-Set-Theorie („Fuzzifizierung“¹¹) berücksichtigt.

¹¹ Mit „Fuzzifizierung“ wird die Anwendung der Fuzzy-Set-Theorie auf einen oder mehrere „scharfe“ Messwerte bezeichnet.

3.3.3.1 Grundlagen der Fuzzy-Set-Theorie

Dieser Abschnitt zielt darauf ab, ausgewählte Elemente der Fuzzy-Theorie vor dem Hintergrund der angestrebten Anwendung darzustellen. Für ausführliche Informationen sei hier auf (Zadeh 1965; Borchers 2013; Bandemer und Näther 2013) verwiesen.

Die von Zadeh (1965) entwickelte Fuzzy-Set-Theorie dient als Grundlage zur Modellierung von Unschärfen. In der Fuzzy-Set-Theorie wird von der klassischen Mengenlehre abgewichen, nach der ein Element x aus der Grundmenge X ($x \in X$) entweder eindeutig einer Menge A zuzuordnen, oder eindeutig nicht zur Menge A zuzuordnen ist. Stattdessen erlaubt die Fuzzy-Set-Theorie die Darstellung gradueller Zugehörigkeiten (Abbildung 3-10) (Nissen 2007). Der Unterschied zwischen „scharfen“ und „unscharfen“ Mengen ergibt sich somit dadurch, dass unscharfe Mengen eine Zugehörigkeit in einem Intervall, z. B. zwischen 0 und 1 annehmen können, während eine scharfe Abgrenzung auf binäre Werte, d. h. 0 oder 1, beschränkt ist (Borchers 2013).

Fuzzy Sets bieten die Möglichkeit, die Realitätsnähe einer Modellschätzung zu erhöhen. Klassische Modellschätzung basieren meist auf diskreten Eingangsparametern, die bekannte oder unbekannte Unschärfen beinhalten können. Diese scharfe Trennung kann in Abhängigkeit des betrachteten Sachverhalts keine angemessene Repräsentation der Realität darstellen und die Ergebnisse ebendieser Modellrechnungen beeinträchtigen. Die Fuzzy-Set-Theorie bezieht Unschärfen bewusst und systematisch ein, sodass hier bessere Ergebnisse als mit „scharfen“ Modellen erzielt werden können (Nissen 2007).

Als praktisches Beispiel dient häufig die Beschreibung von Temperaturen. Während die Zuordnung von „Frost“ auf Basis der Temperaturskala für Wertebereiche unter 0°C definiert ist, kann die Temperaturen auch durch Begriffe wie kalt, warm, heiß oder sehr heiß beschrieben werden (Abbildung 3-12). Die Beschreibung der Temperatur ist jedoch von der subjektiven Wahrnehmung und Einschätzung des Befragten abhängig (linguistische Unschärfe). Aufgrund der damit verbundenen Unschärfe existiert kein klarer Schwellenwert zwischen den genannten Klassen. Daher wird hier die Fuzzy-Logik genutzt, um die Unschärfen abzubilden. Dabei werden zunächst Temperaturen, die eindeutig zur Gruppe „kalt“ gehören (kleiner als -20°C) von der Gruppe der Temperaturen, die eindeutig nicht zur Gruppe „kalt“ gehören (größer als 40°C) abgegrenzt. Der Wertebereich zwischen diesen beiden Grenzen (hier:

-20°C bis 40°C) ist unscharf und wird durch eine Zugehörigkeitsfunktion beschrieben (beispielsweise durch eine lineare oder exponentielle Funktion).

Klassische Menge:



Unschärfe Menge („Graustufen“ zulässig):

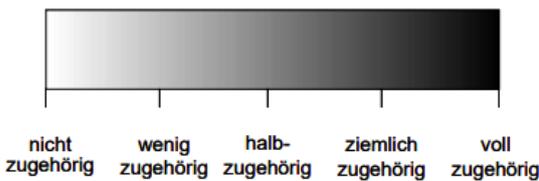


Abbildung 3-10: Scharfe Abgrenzung in der klassischen Mengentheorie und graduelle Zugehörigkeiten in der Fuzzy-Set-Theorie (Nissen 2007)

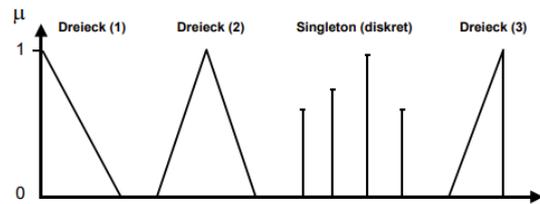
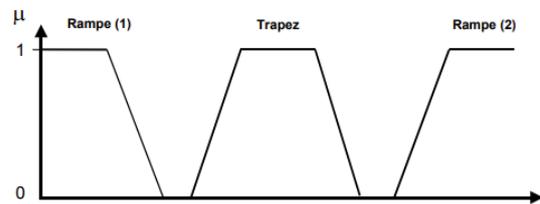


Abbildung 3-11: Beispiele geläufiger Formen von Zugehörigkeitsfunktionen (Nissen 2007)

Die Zugehörigkeitsfunktion (m_A) stellt die Zugehörigkeit eines Elements x zu einer unscharfen Menge A dar. Der Wert wird in der Regel innerhalb eines Wertebereichs zwischen 0 und 1 angegeben. 0 bedeutet dabei, dass das Element nicht zu der unscharfen Menge gehört, während der Wert 1 eine volle Zugehörigkeit angibt. Der Bereich zwischen 0 und 1 gibt die partielle Zugehörigkeit zu einer unscharfen Menge wieder. Eine Übersicht über geläufige Grundformen der Zugehörigkeitsfunktionen gibt Abbildung 3-11 (Nissen 2007). Diese Grundformen sind nicht abschließend und können entsprechend der Problemstellung angepasst und in ihrer Komplexität erhöht werden (Borchers 2013).

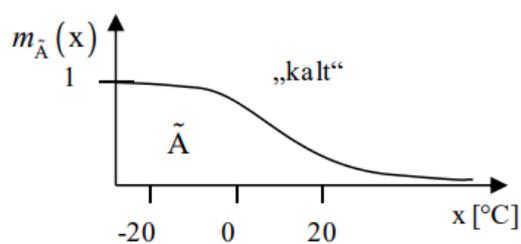
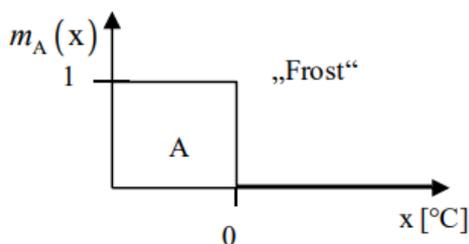


Abbildung 3-12: Beispiel für eine klassische Menge (links) und eine unscharfe Menge (rechts) (Kahlert 1995)

Für das Aufstellen der Zugehörigkeitsfunktionen sind detaillierte Kenntnisse des betrachteten Merkmals essentiell, da die Zugehörigkeitsfunktion so zu

wählen ist, dass diese das Problem adäquat abbildet. Für die Zugehörigkeitsfunktion sind nach Bandemer und Näther (2013) folgende grundlegenden Fragen zu klären:

- Welcher Wertebereich kann eindeutig ausgeschlossen werden?
- Welche möglichen Werte müssen berücksichtigt werden, um den gesuchten, exakten Wert abzubilden?

Um die Unschärfe möglichst exakt abzubilden, eignen sich für die Zugehörigkeitsfunktionen hochparametrisierte Funktionen (Nissen 2007). Mit der Erhöhung der Modellkomplexität steigt jedoch auch die Fehleranfälligkeit der Modellrechnung. Dieser Zusammenhang wird in Kapitel 3.4.3.1 im Rahmen der Beschreibung von Matrixschätzverfahren erläutert. Für weitergehende Informationen zur Fuzzifizierung und Zugehörigkeitsfunktionen sei an dieser Stelle an Borchers (2013) verwiesen.

3.3.3.2 Anwendung bei der Erstellung von Quelle-Ziel-Matrizen

Datengrundlage für die folgenden Ausführungen sind die von den Fahrzeugen erfassten Signale. Sollten darüber hinaus auch Daten berücksichtigt werden können, die von ausgestatteten Haltestellen erfasst werden (Haltestellendaten), sind geringfügige Anpassungen an dem Verfahren zielführend. Diese Anpassungen werden in Kapitel 3.3.3.3 beschrieben.

Zur Ermittlung einer adäquaten Berücksichtigung der Unschärfe ist zunächst das Problem klar abzugrenzen:

- Der **wahre Einstiegsort** kann unter Berücksichtigung der Unschärfe vor, jedoch nicht nach der ersten Erfassung liegen.
- Der **wahre Ausstiegsort** kann unter Berücksichtigung der Unschärfe nach, jedoch nicht vor der letzten Erfassung liegen.

Zur Verdeutlichung der Ermittlung möglicher Quelle-Ziel-Relationen auf Basis der Zuordnung der Quelle-Ziel-Relation anhand des ersten und letzten Signals dient Abbildung 3-13. Das rote „X“ kennzeichnet dabei die Zuordnung auf Basis der erfassten Signale ohne Berücksichtigung der Unschärfe.

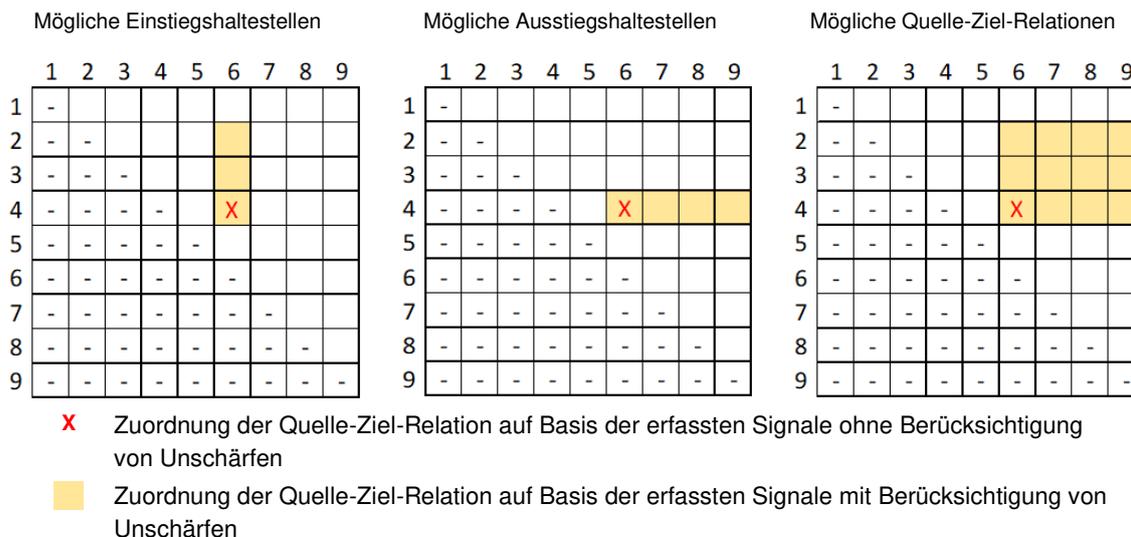


Abbildung 3-13: Ermittlung möglicher Quelle-Ziel-Relationen mit Fuzzifizierung

Für Ermittlung möglicher Einstiegshaltestellen werden alle Haltestellen, die das Fahrzeug entsprechend der IST-Fahrplandaten in einem festgelegten Zeitintervall vor Erfassung des ersten Signals des entsprechenden Endgeräts angefahren hat, als potentielle Einstiegshaltestellen berücksichtigt. Das Zeitintervall kann in Abhängigkeit des Input-Datums (WLAN, Bluetooth) variiert werden. Für das Zeitintervall wird im Rahmen dieser Dissertation ein Zeitraum von 20 Minuten sowohl für WLAN- als auch für Bluetooth-Daten gewählt. Entsprechend der Erkenntnisse aus der Analyse der WLAN- und Bluetooth-Daten ist bekannt, dass

- Smartphones im Standby teilweise über 15 Minuten kein Signal aussenden (Kapitel 2.5.1.1) und
- über 90% der MAC-Adressen* einen mittleren Abstand zwischen zwei Signalen von weniger als 20 Minuten besitzen (Kapitel 2.5.1.2 und 2.5.2).

Dieser vergleichsweise große Zeitraum ist dazu vorgesehen, dass die Betrachtung der Einstiegshaltestellen „auf der sicheren Seite“ erfolgt und mit hoher Wahrscheinlichkeit die reale Einstiegshaltestelle umfasst. Grundsätzlich sind auch andere Zeitintervalle denkbar.

Die Ermittlung der möglichen Ausstiegshaltestellen wird analog durchgeführt. Hier werden alle Haltestellen, die das Fahrzeug in dem Zeitintervall von 20 Minuten nach der Erfassung des letzten Signals angefahren hat (IST-Fahrplandaten), als potentielle Ausstiegshaltestellen berücksichtigt.

Die möglichen Quelle-Ziel-Relationen ergeben sich anschließend aus dem kartesischen Produkt der möglichen Ein- und Ausstiegshaltestellen. In dem Beispiel in Abbildung 3-13 kommen drei Haltestellen für den Einstieg in Betracht und vier Ausstiegshaltestellen. Daraus folgen zwölf mögliche Quelle-Ziel-Relationen ($3 \times 4 = 12$).

Der nächste Schritt hat das Ziel, den möglichen Relationen über Zugehörigkeitsfunktionen Wahrscheinlichkeiten zuzuordnen. Die Summe der Wahrscheinlichkeiten aller betrachteten Relationen für einen Fahrgast beträgt dabei immer 1. Die Zugehörigkeitsfunktion verteilt daher den Wert 1 (= eine Fahrgastfahrt) über alle möglichen Quelle-Ziel-Relationen. Der Wert der Zugehörigkeit einer Relation hängt dabei von der zeitlichen Differenz zwischen den erfassten Signalen ab. Die genaue Zuordnung der Zugehörigkeitswerte kann auf dieser Grundlage über zwei Vorgehensweisen ermittelt werden:

- **statische Berechnung** anhand fester Vorgaben aus einem Referenzdatensatz
- **dynamische Berechnung** anhand gerätespezifischer Kennwerte

Statische Berechnung

Bei der statischen Berechnung basiert die Zugehörigkeitsfunktion zur Abbildung der Unschärfe auf der Häufigkeitsverteilung des mittleren zeitlichen Abstands zwischen Bluetooth- bzw. WLAN-Signalen. Aufbauend auf den Untersuchungen zum Sendeverhalten (Kapitel 2.5.1 und Kapitel 2.5.2) kann eine Häufigkeitsverteilung des zeitlichen Abstands zwischen Bluetooth- bzw. WLAN-Signalen abgeleitet werden (Abbildung 3-14). Die dargestellte Häufigkeitsverteilung wird dazu genutzt, die Wahrscheinlichkeiten eines weiteren bzw. vorherigen Signals eines Endgeräts in minutenscharfen Zeitintervallen bis zum Maximalwert von 20 Minuten abzuschätzen und bildet die Grundlage des statischen Ansatzes¹².

Die Häufigkeitsverteilung berücksichtigt lediglich die MAC-Adressen*, deren mittlerer zeitlicher Abstand zwischen Signalen innerhalb des Betrachtungszeitraums von 20 Minuten liegt. Zeitliche Abstände von weniger als einer Sekunde wurden nicht berücksichtigt. Bei der Betrachtung der WLAN-Signale werden nur MAC-Adressen* betrachtet, deren lokaler Bit auf eine nicht-

¹² Hinweis: Die hier dargestellte Verteilung weicht von der in Kapitel 2.5 dargestellten Häufigkeitsverteilung der Anzahl erfasster WLAN- bzw. Bluetooth-Signale (Tabelle 2-12, Tabelle 2-14) ab, da hier lediglich Signale in einem Zeitraum von maximal 20 Minuten nach dem ersten Signal berücksichtigt wurden.

randomisierte MAC-Adresse hinweist, da randomisierte MAC-Adressen in der Verfahrensentwicklung nicht berücksichtigt werden¹³.

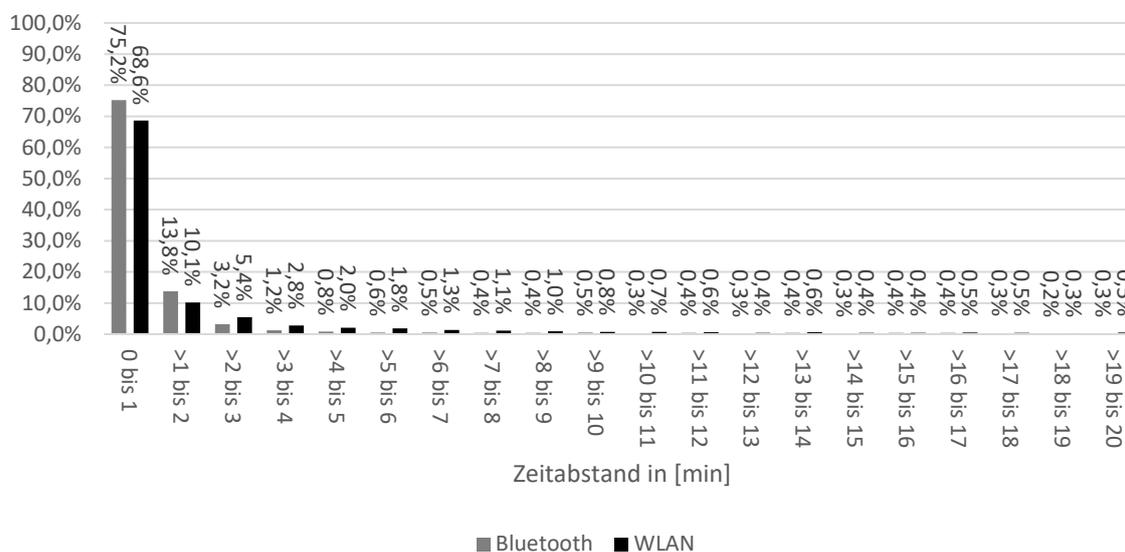


Abbildung 3-14: Häufigkeitsverteilung des zeitlichen Abstands zwischen Bluetooth- (grau, n = 13.862) und WLAN-Signalen (schwarz, n = 14.761) aggregiert in Minuten-Klassen (Dataset Bieland/Mobile Data Fusion 2022)

Zur Erläuterung dient folgendes Beispiel: Auf einer Linienfahrt wird ein unbekanntes Endgerät das erste Mal um 13:30:00 Uhr und das letzte Mal um 13:54:00 Uhr erfasst. Das nächste zu erwartende Signal des Endgeräts wird – entsprechend des statischen Ansatzes – mit einer Wahrscheinlichkeit (basierend auf der Häufigkeitsverteilung von Bluetooth-Signalen) von 75,2% innerhalb von einer Minute (13:54:00 Uhr bis 13:55:00 Uhr) erfasst werden, mit einer Wahrscheinlichkeit von 13,8% in einem Zeitraum von ein bis zwei Minuten (13:55:00 Uhr bis 13:56:00 Uhr), mit einer Wahrscheinlichkeit von rund 3% in einem Zeitraum von zwei bis drei Minuten (13:56:00 Uhr bis 13:57:00 Uhr) usw. Dies gilt analog für eine mögliche Erfassung vor dem ersten Signal, auch hier betragen die Wahrscheinlichkeiten für eine Erfassung vor einer Minute rund 75% (13:29:00 Uhr bis 13:30:00 Uhr) usw. Innerhalb des betrachteten Zeitraums sind dabei alle durch das Fahrzeug im IST-Fahrplan angefahrenen Haltestellen als mögliche Ein- und Ausstiegshaltestellen zu berücksichtigen.

¹³ Werden auch randomisierte MAC-Adressen verwendet – z.B. durch Entwicklung eines auf Information Elements und weiteren übertragenen Daten basierenden Derandomisierungs-Tools – ist die Häufigkeitsverteilung auf Basis der neuen Datengrundlage zu aktualisieren.

Analog zum Vorgehen zur Erstellung der Quelle-Ziel-Matrizen (Kapitel 3.3.2.1) ist auch bei der Zuordnung der zu berücksichtigenden Haltestellen zu den einzelnen Zeitintervallen zu beachten, dass der Halt eines Fahrzeugs an einer Haltestelle in der Regel aus der Ankunft des Fahrzeugs, der Haltezeit und der Abfahrt besteht. Um einen scharfen Bezugszeitpunkt für die Zuordnung der zu berücksichtigenden Haltestellen zu den einzelnen, minuten-scharfen Zeitintervallen vornehmen zu können, werden folgende Annahmen getroffen:

- Für die **Berechnung der Einstiegshaltestelle** wird angenommen, dass der Einstieg unmittelbar zum Zeitpunkt der Abfahrt des Fahrzeugs (IST-Fahrplandaten) stattfindet (Abbildung 3-15).
- Für die **Berechnung der Ausstiegshaltestelle** wird angenommen, dass der Ausstieg unmittelbar bei der Ankunft des Fahrzeugs an der Haltestelle (IST-Fahrplandaten) stattfindet (Abbildung 3-15).

Diese Annahmen tendieren auf die sichere Seite, sodass die Zuordnung von Wahrscheinlichkeiten zu den Zeitintervallen so erfolgt, dass die gesamte Haltezeit des Fahrzeugs an einer Haltestelle als Ein- bzw. Ausstiegszeit eines Fahrgasts berücksichtigt wird. Dadurch werden bei der Fuzzifizierung tendenziell mehr Haltestellen berücksichtigt, sodass – entsprechend des Fuzzy-Set-Ansatzes – eher alle möglichen Ein- und Ausstiegsorte in die Bewertung der Wahrscheinlichkeiten eingehen.

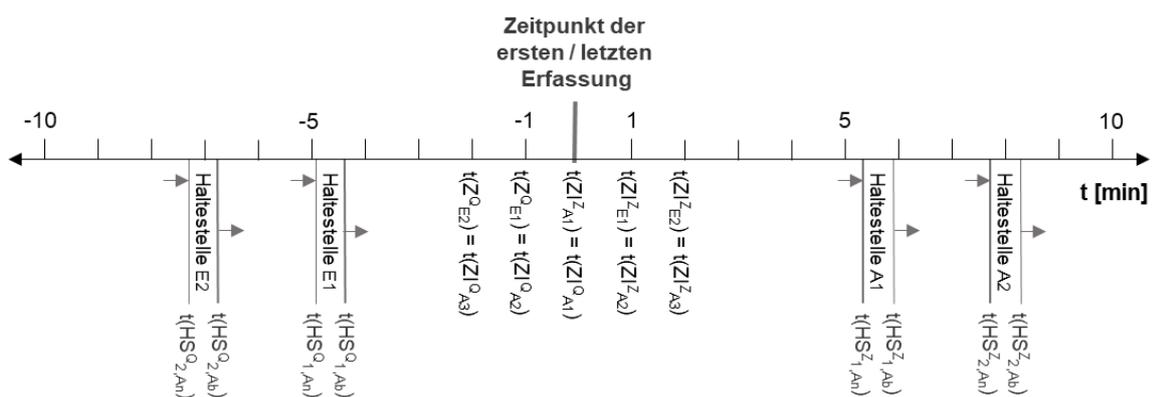


Abbildung 3-15: Skizze zur Verdeutlichung des Bezugszeitpunkts zur Zuordnung von relevanten Haltestellen zu den korrekten Zeitintervallen

Dabei ist zu beachten, dass der wahre Zeitpunkt des Zu- oder Ausstiegs nicht bekannt ist. Durch diese Zuordnung ergeben sich Unschärfen, da ein Ausstieg auch zu einem späteren Zeitpunkt, bzw. ein Zustieg zu einem früheren Zeitpunkt stattfinden kann, sodass ggf. zu viele Haltestellen berücksichtigt

werden. Die Auswirkungen dieser Unschärfen werden jedoch als marginal betrachtet, da sich diese Zuordnung nur im geringen Umfang auf das Ergebnis der Fuzzifizierung (Wahrscheinlichkeiten der Quelle-Ziel-Relationen) auswirkt.

Die Zuordnung von Wahrscheinlichkeiten für den Ein- oder Ausstieg zu den einzelnen Haltestellen im Fahrtverlauf (IST-Fahrplandaten) im Betrachtungszeitfenster von 20 Minuten vor der Erfassung des ersten Signals bzw. nach der Erfassung des letzten Signals **eines Endgeräts** erfolgt in fünf Schritten (Zur Veranschaulichung des Vorgehens dient eine Beispielrechnung für die Fuzzifizierung (statische Berechnung) auf Basis von Bluetooth-Daten in Anhang 2):

(1) Der erste Schritt besteht darin, die **Wahrscheinlichkeiten der Häufigkeitsverteilung in Abhängigkeit der zusätzlichen zur Verfügung stehenden Informationen für einzelne Endgeräte anzupassen**: Dazu werden nur die Zeitintervalle ab dem Erreichen der ersten Haltestelle bis zur maximal möglichen Betrachtungszeit (20 Minuten oder Start- bzw. Endzeitpunkt der Linienfahrt) berücksichtigt (relevante Zeitintervalle). Da ein Ein- oder Ausstieg zu 100% erfolgt, werden die Wahrscheinlichkeiten der folgenden Zeitintervalle gewichtet nach der Wahrscheinlichkeit im Status Quo auf 100% hochgerechnet. Die übrigen Wahrscheinlichkeiten in den Zeitintervallen werden auf 0% gesetzt. Dies gilt entsprechend

- Formel 3-3 für Zeitintervalle, in denen die erste Haltestelle noch nicht erreicht wurde (für die Berechnung der Einstiegshaltestelle $t(ZI_{E_x}^Q) > t(HS_{1,Ab}^Q)$; für die Berechnung der Ausstiegshaltestelle: $t(ZI_{E_x}^Z) < t(HS_{1,An}^Z)$) und entsprechend
- Formel 3-4 für Zeitintervalle, die im Bezug zum Start- bzw. Endzeitpunkt der Linienfahrt nach IST-Fahrplandaten bzw. durch den maximalen Betrachtungszeitraum von 20 Minuten außerhalb der Betrachtungszeit liegen (für die Berechnung der Einstiegshaltestelle für $t(ZI_{A_x}^Q) < t_{max}^Q(FP)$; für die Berechnung der Ausstiegshaltestelle: $t(ZI_{A_x}^Z) > t_{max}^Z(FP)$).

Relevante Zeitintervalle werden für die Berechnung der Einstiegshaltestelle durch die Bedingungen $t(ZI_{E_x}^Q) \leq t(HS_{1,Ab}^Q)$ und $t(ZI_{A_x}^Q) \geq t_{max}^Q(FP)$ ermittelt, für die Ausstiegshaltestelle gelten die Bedingungen $t(ZI_{E_x}^Z) \geq t(HS_{1,An}^Z)$ und $t(ZI_{A_x}^Z) \leq t_{max}^Z(FP)$.

Zum besseren Verständnis von Formel 3-4 (und analog auch Formel 3-3) dient Abbildung 3-16, in der die Anwendung der Formeln zur Identifikation der relevanten Zeitintervalle bei Berechnung der Ausstiegshaltestelle graphisch dargestellt ist.

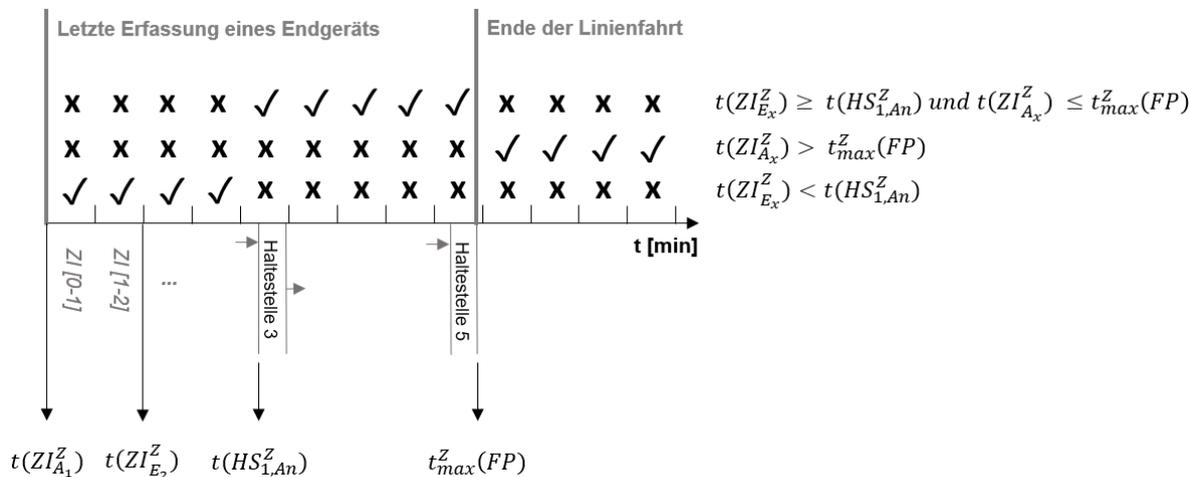


Abbildung 3-16: Skizze zur Verdeutlichung der Auswahl relevanter Zeitintervalle

Es ist zu berücksichtigen, dass sich die in den Formeln genannten Zeitpunkte (Anfangszeit, Endzeit, Ankunftszeit, Abfahrtszeit) und Zeitintervalle immer auf Uhrzeiten (Format: [HH:MM:SS]) beziehen. Für die Anwendung der Formeln kann die Uhrzeit grundsätzlich in Sekunden und mit einem festen Bezugszeitpunkt bestimmt werden. Hierfür eignet sich u. a. die Unixzeit, die die vergangenen Sekunden seit dem 01.01.1970, 00:00 Uhr UTC zählt.

Beispiel: Die letzte Erfassung von einem Signal eines Endgeräts erfolgt um 13:09:15 Uhr. Die nächste Haltestelle nach 13:09:15 Uhr wird um 13:11:00 Uhr erreicht. In dem ersten Zeitintervall $p(aZI_1^Z)$ zwischen 13:09:15 und 13:10:14 Uhr sollte mit einer Wahrscheinlichkeit von 75% das nächste Signal empfangen werden. Da hier jedoch keine Haltestelle angefahren wurde, konnte auch kein Ausstieg erfolgen. Somit sind die verbleibenden Wahrscheinlichkeiten der Zeitintervalle ab 13:10:15 Uhr ($t(ZI_{A_2}^Z)$) für den Ausstieg auf 100% hochzurechnen.

$$p(aZI_x^Q) = \begin{cases} \frac{100\%}{\sum_x p(ZI_x^Q)} * p(ZI_x^Q) & \text{für } t(ZI_{E_x}^Q) \leq t(HS_{1,Ab}^Q) \wedge t(ZI_{A_x}^Q) \geq t_{max}^Q(FP) \\ 0 & \text{für } t(ZI_{E_x}^Q) > t(HS_{1,Ab}^Q) \\ 0 & \text{für } t(ZI_{A_x}^Q) < t_{max}^Q(FP) \end{cases} \quad \text{Formel 3-3}$$

mit

$p(aZI_x^Q)$	Angepasste Wahrscheinlichkeit für ein Signal im Zeitintervall x für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
$p(ZI_x^Q)$	Wahrscheinlichkeit des Zeitintervalls x für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrgastfahrt gemäß Häufigkeitsverteilung
x	Index der Zeitintervalle von x=1 bis x=20 Minuten
$t(ZI_{E_x}^Q)$	Endzeit (E) des x-ten Zeitintervalls ab dem Signal der ersten Erfassung für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
$t(HS_{1,Ab}^Q)$	Abfahrtszeit (Ab) an der ersten möglichen Haltestelle vor Erfassung des ersten Signals für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
$t(ZI_{A_x}^Q)$	Anfangszeit (A) des x-ten Zeitintervalls ab dem Signal der ersten Erfassung für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
$t_{max}^Q(FP)$	Uhrzeit des maximal möglichen Betrachtungszeitraums bei Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrgastfahrt (oder Startzeit der Linienfahrt) nach IST-Fahrplan (FP)

$$p(aZI_x^Z) = \begin{cases} \frac{100\%}{\sum_x p(ZI_x^Z)} * p(ZI_x^Z) & \text{für } t(ZI_{E_x}^Z) \geq t(HS_{1,An}^Z) \wedge t(ZI_{A_x}^Z) \leq t_{max}^Z(FP) \\ 0 & \text{für } t(ZI_{E_x}^Z) < t(HS_{1,An}^Z) \\ 0 & \text{für } t(ZI_{A_x}^Z) > t_{max}^Z(FP) \end{cases} \quad \text{Formel 3-4}$$

mit

$p(aZI_x^Z)$	Angepasste Wahrscheinlichkeit für ein Signal im Zeitintervall x für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
$p(ZI_x^Z)$	Wahrscheinlichkeit des Zeitintervalls x für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrgastfahrt gemäß Häufigkeitsverteilung
x	Index der Zeitintervalle von x=1 bis x=20
$t(ZI_{E_x}^Z)$	Endzeit (E) des x-ten Zeitintervalls ab dem Signal der letzten Erfassung für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
$t(HS_{1,An}^Z)$	Ankunftszeit (An) an der ersten möglichen Haltestelle nach Erfassung des letzten Signals für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
$t(ZI_{A_x}^Z)$	Anfangszeit (A) des x-ten Zeitintervalls ab dem Signal der ersten Erfassung für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
$t_{max}^Z(FP)$	Maximal möglicher Betrachtungszeitraum bei Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrgastfahrt (oder Endzeit der Linienfahrt) nach IST-Fahrplan (FP)

(2) Im zweiten Schritt wird die **Anzahl an Haltestellen je Zeitintervall bestimmt, die als potentielle Ein- bzw. Ausstiegshaltestelle zu berücksichtigen sind (nHS)**. Zur Verdeutlichung dient hier Abbildung 3-17, in der die relevanten Zeitintervalle (ZI) für die Einstiegshaltestelle (ZI[0-1] bis ZI[5-6]) eingezeichnet sind. Durch die Darstellung kann leicht die Anzahl der relevanten Haltestellen (nHS) der jeweiligen Zeitintervalle abgelesen werden.

Es ist zu berücksichtigen, dass bei Betrachtung der Einstiegshaltestellen alle Haltestellen als relevant gelten, die bis zur Endzeit des Zeitintervalls

angefahren wurden. Im ersten Zeitintervall (ZI[0-1]), das die Zeiten von 13:06:00 bis 13:05:01 Uhr umfasst, gibt es eine relevante Haltestelle (Haltestelle 4). Im ZI[4-5], d. h. wenn ein Signal zwischen 13:02:00 bis 13:01:01 Uhr ausgesendet, aber nicht durch das Fahrzeug erfasst wurde, kann der Einstieg an insgesamt drei Haltestellen (Haltestellen 2 bis 4) stattgefunden haben.

Das Vorgehen bei der Ermittlung der relevanten Ausstiegshaltestellen verläuft analog zum Vorgehen der Einstiegshaltestellen. Die Zeitintervalle für die Betrachtung der Ausstiegshaltestelle wurden aus Gründen der Lesbarkeit nicht eingezeichnet.

Haltestelle 5 kommt hier weder als Ein- noch als Ausstiegshaltestelle in Frage, da

- das erste Signal des Endgeräts bereits vor Ankunft an Haltestelle 5 und
- das letzte Signal des Endgeräts nach der Abfahrt von Haltestelle 5 erfasst wurde.

Die Anzahl der relevanten Haltestellen bestimmt sich somit aus Formel 3-5 und Formel 3-6.

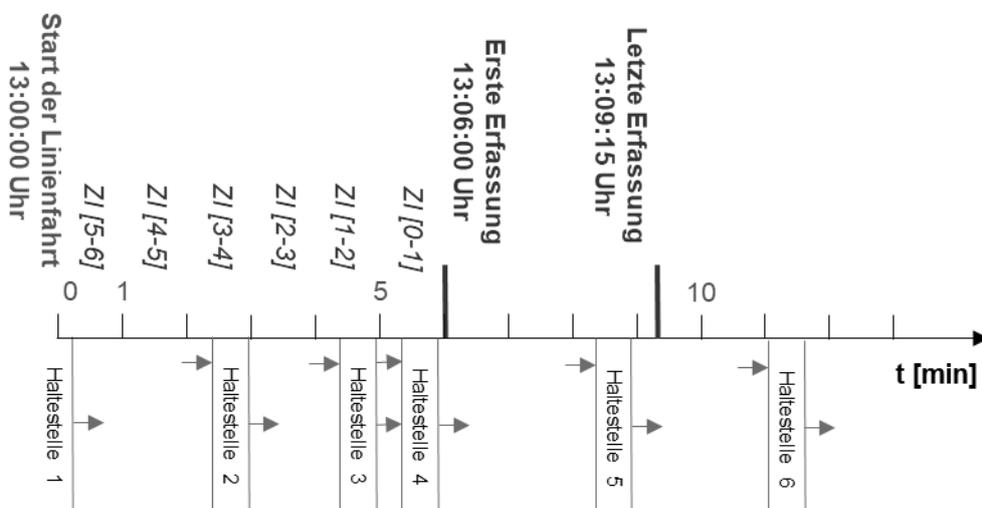


Abbildung 3-17: Bestimmung der Anzahl relevanter Haltestellen (Beispiel: Einstiegshaltestellen)

$$nHS^Q(ZI_x^Q) = \{nHS^Q \in HS \mid t(ZI_{E_x}^Q) \leq t(HS_{o,Ab}^Q)\} \quad \text{Formel 3-5}$$

mit

$nHS^Q(ZI_x^Q)$	Anzahl aller relevanten Einstiegshaltestellen (Menge aller nHS^Q aus der Menge der Haltestellen) des x-ten Zeitintervalls für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrgastfahrt
$t(ZI_{E_x}^Q)$	Endzeit (E) des x-ten Zeitintervalls ab dem Signal der ersten Erfassung für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
$t(HS_{o,Ab}^Q)$	Abfahrtszeit (Ab) an der o-ten Haltestelle vor Erfassung des ersten Signals für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt

$$nHS^Z(ZI_x^Z) = \{nHS^Z \in HS \mid t(ZI_{E_x}^Z) \geq t(HS_{d,An}^Z)\} \quad \text{Formel 3-6}$$

mit

$nHS^Z(ZI_x^Z)$	Anzahl aller relevanten Ausstiegshaltestellen (Menge aller nHS^Z aus der Menge der Haltestellen) des x-ten Zeitintervalls für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrgastfahrt
$t(ZI_{E_x}^Z)$	Endzeit (E) des x-ten Zeitintervalls ab dem Signal der ersten Erfassung für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
$t(HS_{d,An}^Z)$	Ankunftszeit (An) an der d-ten Haltestelle nach Erfassung des letzten Signals für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt

(3) Bei der Betrachtung von Zeitintervallen ist es möglich, dass mehrere **relevante Haltestellen** zu berücksichtigen sind. Im Beispiel von Abbildung 3-17 liegen im Zeitintervall „ZI[4-5]“ drei relevante Haltestellen (Haltestellen 2 bis 4). Da ein Ein- oder Ausstieg nur an den relevanten Haltestellen stattfinden kann und keine weiteren Informationen zum Ein- oder Ausstiegszeitpunkt vorliegen, wird die Wahrscheinlichkeit eines Ein- oder Ausstiegs für jede relevante Haltestelle des Zeitintervalls als gleichverteilt angenommen. Die Berechnung des Gewichtungsfaktor ist in Formel 3-7 und Formel 3-8 dargestellt.

$$g(ZI_x^Q, HS_o^Q) = \frac{100\%}{nHS^Q(ZI_x^Q)} \quad \text{Formel 3-7}$$

mit

$g(ZI_x^Q, HS_o^Q)$	Gewichtungsfaktor für den Einstieg an der o-ten Haltestellen des x-ten Zeitintervall für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
$nHS^Q(ZI_x^Q)$	Anzahl aller relevanten Einstiegshaltestellen (Menge aller nHS^Q aus der Menge der Haltestellen) des x-ten Zeitintervalls für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrgastfahrt

$$g(ZI_x^Z, HS_d^Z) = \frac{100\%}{nHS^Z(ZI_x^Z)} \quad \text{Formel 3-8}$$

mit

- $g(ZI_x^Z, HS_d^Z)$ Gewichtungsfaktor für den Ausstieg an der d-ten Haltestellen des x-ten Zeitintervall für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
- $nHS^Z(ZI_x^Z)$ Anzahl aller relevanten Ausstiegshaltestellen (Menge aller nHS^Z aus der Menge der Haltestellen) des x-ten Zeitintervalls für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrgastfahrt

(4) Die **Ermittlung der gewichteten Wahrscheinlichkeit für den Ein- oder Ausstieg** an einer Haltestelle innerhalb eines Zeitintervalls wird in Formel 3-9 und Formel 3-10 gezeigt.

$$p_g(ZI_x^Q, HS_o^Q) = g(ZI_x^Q, HS_o^Q) * p(aZI_x^Q) \quad \text{Formel 3-9}$$

mit

- $p_g(ZI_x^Q, HS_o^Q)$ gewichtete Wahrscheinlichkeit für den Einstieg an der o-ten Haltestellen des x-ten Zeitintervall für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
- $g(ZI_x^Q, HS_o^Q)$ Gewichtungsfaktor für den Einstieg an der o-ten Haltestellen des x-ten Zeitintervall für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
- $p(aZI_x^Q)$ Angepasste Wahrscheinlichkeit für ein Signal im Zeitintervall x für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt

$$p_g(ZI_x^Z, HS_d^Z) = g(ZI_x^Z, HS_d^Z) * p(aZI_x^Z) \quad \text{Formel 3-10}$$

mit

- $p_g(ZI_x^Z, HS_d^Z)$ gewichtete Wahrscheinlichkeit für den Ausstieg an der d-ten Haltestellen des x-ten Zeitintervall für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
- $g(ZI_x^Z, HS_d^Z)$ Gewichtungsfaktor für den Ausstieg an der d-ten Haltestellen des x-ten Zeitintervall für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
- $p(aZI_x^Z)$ Angepasste Wahrscheinlichkeit für ein Signal im Zeitintervall x für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt

(5) Durch **Aufsummieren der Einzelwahrscheinlichkeiten** für den Ein- bzw. Ausstieg der Haltestellen über alle Zeitintervalle ergibt sich die Wahrscheinlichkeit für den Ein- oder Ausstieg an der jeweiligen Haltestelle (Formel 3-11, Formel 3-12). Entsprechend der Annahme, dass ein Signal mit hoher Wahrscheinlichkeit (75%) innerhalb von einer Minute wieder erfasst werden sollte, erhält hier die nächste Haltestelle erwartungsgemäß die höchste Wahrscheinlichkeit für den Ausstieg.

$$p_g(HS_o^Q) = \sum_i p_g(ZI_x^Q, HS_o^Q) \quad \text{Formel 3-11}$$

mit

$p_g(HS_o^Q)$	gewichtete Wahrscheinlichkeit für den Einstieg an der o-ten Haltestelle für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
$p_g(ZI_x^Q, HS_o^Q)$	gewichtete Wahrscheinlichkeit für den Einstieg an der o-ten Haltestellen des x-ten Zeitintervall für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt

$$p_g(HS_d^Z) = \sum_i p_g(ZI_x^Z, HS_d^Z) \quad \text{Formel 3-12}$$

mit

$p_g(HS_d^Z)$	gewichtete Wahrscheinlichkeit für den Ausstieg an der d-ten Haltestelle für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt
$p_g(ZI_x^Z, HS_d^Z)$	gewichtete Wahrscheinlichkeit für den Ausstieg an der d-ten Haltestellen des x-ten Zeitintervall für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt

Abbildung 3-18 zeigt das Ergebnis der Fuzzifizierung einer theoretischen Quelle-Ziel-Relation, die auf den Daten (erstes und letztes Signal) eines Endgeräts basiert. Dabei wurden die Werte für die Ein- und Ausstiegshaltestellen entsprechend der gegebenen Beispielrechnung genutzt. Die Haltezeit wurde auf 15 Sekunden festgelegt. Werte, die 0,01 unterschreiten, wurden mit „<0,01“ abgekürzt. Die Berechnung der Werte für die möglichen Quelle-Ziel-Relationen basieren auf Formel 3-13.

$$p_g(HS_{od}) = p_g(HS_o^Q) * p_g(HS_d^Z) \quad \text{Formel 3-13}$$

mit

$p_g(HS_{od})$	gewichtete Wahrscheinlichkeit für die Quelle-Ziel-Relation von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$p_g(HS_o^Q)$	gewichtete Wahrscheinlichkeit für den Ausstieg an der o-ten Haltestelle für die Berechnung der Einstiegshaltestelle (Quelle (Q)) der Fahrt
$p_g(HS_d^Z)$	gewichtete Wahrscheinlichkeit für den Ausstieg an der d-ten Haltestelle für die Berechnung der Ausstiegshaltestelle (Ziel (Z)) der Fahrt

Das Vorgehen wird anschließend für alle Endgeräte wiederholt. Die Ergebnisse – spezifische Quelle-Ziel-Matrizen für jedes Endgerät – werden dabei

- in einzelnen Matrizen oder
- in Vektoren für die einzelnen Matrixelemente (d_{od})

gespeichert. Die Fuzzifizierung wird durch das Aufsummieren der Matrizen aller Endgeräte abgeschlossen. Dieses Vorgehen bietet den Vorteil, dass weniger Nullwerte zu erwarten sind und somit das Problem nicht struktureller Nullwerte (Problem of non-structural zeros) (Kapitel 3.4.3.2) reduziert wird.

Mögliche Einstiegshaltestellen										Mögliche Ausstiegshaltestellen									
	1	2	3	4	5	6	7	8	9		1	2	3	4	5	6	7	8	9
1	-				<0,01					1	-								
2	-	-			0,01					2	-	-							
3	-	-	-		0,08					3	-	-	-						
4	-	-	-	-	0,91					4	-	-	-	-	0,98	0,02	<0,01	<0,01	
5	-	-	-	-	-					5	-	-	-	-	-				
6	-	-	-	-	-	-				6	-	-	-	-	-	-			
7	-	-	-	-	-	-	-			7	-	-	-	-	-	-	-		
8	-	-	-	-	-	-	-	-		8	-	-	-	-	-	-	-	-	
9	-	-	-	-	-	-	-	-	-	9	-	-	-	-	-	-	-	-	-

Mögliche Quelle-Ziel-Relation									
	1	2	3	4	5	6	7	8	9
1	-				<0,01	<0,01	<0,01	<0,01	
2	-	-			0,01	<0,01	<0,01	<0,01	
3	-	-	-		0,08	<0,01	<0,01	<0,01	
4	-	-	-	-	0,89	0,02	<0,01	<0,01	
5	-	-	-	-	-				
6	-	-	-	-	-	-			
7	-	-	-	-	-	-	-		
8	-	-	-	-	-	-	-	-	
9	-	-	-	-	-	-	-	-	-

Abbildung 3-18: Ermittlung der Zugehörigkeitswerte aller möglichen Quelle-Ziel-Relationen

Unsicherheiten ergeben sich im statischen Ansatz aufgrund der festgelegten Wahrscheinlichkeiten für die Erfassung eines weiteren Signals. Gerätespezifische Eigenheiten im Sendeverhalten werden hier nicht berücksichtigt.

Dynamische Berechnung

Das Vorgehen im dynamischen Ansatz baut auf den Berechnungsschritten des statischen Verfahrens auf. Im statischen Ansatz werden die Wahrscheinlichkeiten für eine Signalerfassung vor dem ersten erfassten Signal bzw. nach dem letzten erfassten Signal auf Basis von Referenzdaten in minutenscharfen Zeitintervallen festgelegt. Diese Wahrscheinlichkeiten haben einen direkten Einfluss auf die Zugehörigkeitswerte und damit auf das Ergebnis der Fuzzifizierung.

Schwäche der statischen Vorgehensweise ist, dass jedes Endgerät die gleichen Wahrscheinlichkeiten zugewiesen bekommt. Durch die Häufigkeitsverteilung sowie die Untersuchungen zum Sendeverhalten ist jedoch ersichtlich, dass einzelne Endgeräte durchaus große Zeitabstände zwischen Signalen vorweisen. Der statische Ansatz berücksichtigt spezifische Sendeeigenschaften einzelner Geräte jedoch nicht. Somit erhalten bislang Geräte, die (theoretisch) alle 10 Sekunden ein Signal aussenden, die gleichen Wahrscheinlich-

keiten innerhalb der vorgegebenen Zeitintervalle, wie Geräte, die (theoretisch) alle 300 Sekunden ein Signal aussenden.

Um diese Unterschiede bei der Fuzzifizierung zu berücksichtigen, werden bei der dynamischen Berechnung keine Wahrscheinlichkeiten der Zeitintervalle vorgegeben, sondern Wahrscheinlichkeiten für jedes Endgerät neu ermittelt (d. h. in Abhängigkeit des individuellen Sendeverhaltens bestimmt). Dazu werden alle Signale einer MAC-Adresse während der Dauer einer Linienfahrt in die Betrachtung aufgenommen. Die Wahrscheinlichkeit für einen Ein- oder Ausstieg werden je Zeitintervall auf Basis der erfassten Daten festgelegt (Formel 3-14).

$$p_E(ZI_x) = \frac{nS_E(ZI_x)}{\sum_x nS_E(ZI_x)} \quad \text{Formel 3-14}$$

mit

$p_E(ZI_x)$	Wahrscheinlichkeit für den Ein- oder Ausstieg des Endgeräts E im Zeitintervall x
$nS_E(ZI_x)$	Anzahl der erfassten Signale des Endgeräts E während des Zeitintervalls x

Sollten von einzelnen Endgeräten nur wenige Signale erfasst worden sein, kann folglich nur eine geringe Anzahl an Zeitabständen bestimmt werden. Wahrscheinlichkeit des nächsten Signals für einzelne Endgeräte kann bei nur zwei erfassten Signalen (einem Zeitabstand) entsprechend zu 100% in einem Zeitintervall liegen. Daher ist hier ein Grenzwert für die Anzahl erfasster Signale bzw. Anzahl ermittelter Zeitabstände denkbar, um diesem Problem zu begegnen. Bei zu wenig vorliegenden Zeitabständen besteht die Möglichkeit, auf die bestehende statische Berechnung zurückzugreifen.

Die weiteren Berechnungsschritte werden wie im Fall der statischen Berechnung durchgeführt und werden an dieser Stelle nicht erneut beschrieben.

3.3.3.3 Theoretische Überlegungen zur Präzisierung der Ein- und Ausstiegshaltestelle

Analog zum Vorgehen bei der Erstellung von Quelle-Ziel-Matrizen ohne Fuzzifizierung können auch hier theoretische Überlegungen abgeleitet werden, die zu einer Erhöhung der Genauigkeit bei der Bestimmung der Ein- und Ausstiegshaltestelle führen können.

Einbezug von Haltestellendaten

Auch das Vorgehen zur Fuzzifizierung der Quelle-Ziel-Relationen auf Basis der Input-Daten kann analog zu Kapitel 3.3.2.2 durch Daten, die durch die Ausstattung der Haltestellen erfasst werden, präzisiert werden:

Zusätzliche Informationen zur Einstiegshaltestelle: Kann aus den zusätzlichen Informationen die Schlussfolgerung gezogen werden, dass der Fahrgast an einer konkreten Haltestelle zugestiegen ist, sind für die weiteren Schritte der Fuzzifizierung lediglich die möglichen Ausstiegshaltestellen zu berücksichtigen (Abbildung 3-19, links) und die Ausstiegswahrscheinlichkeiten entsprechend Formel 3-13 zu berechnen.

Zusätzliche Informationen zur Ausstiegshaltestelle: Kann aus den zusätzlichen Informationen die Schlussfolgerung gezogen werden, dass der Fahrgast mindestens bis zu einer Haltestelle gefahren ist, die nach der zugeordneten Ausstiegshaltestelle liegt, ist dies ebenfalls bei dem Vorgehen der Fuzzifizierung zu berücksichtigen. In Abbildung 3-19 (Mitte und rechts) sind mögliche Quelle-Ziel-Relationen auf Basis zusätzlicher Informationen zur Ausstiegshaltestelle dargestellt. In dem Beispiel wurde der Fahrgast das letzte Mal an (oder nach) Haltestelle 5 durch das Fahrzeug erfasst (markiert durch das „X“ in rot). Nach der letzten Erfassung im Fahrzeug wurde die MAC-Adresse durch die Haltestelle 8 wiedererfasst (zum Zeitpunkt des Halts). Ein Ausstieg an Haltestelle 6 oder 7 kann somit ausgeschlossen und die möglichen Quelle-Ziel-Relationen effektiv eingegrenzt werden. Trotz Erfassung an Haltestelle 8 ist zu berücksichtigen, dass der Ausstieg auch zu einem späteren Zeitpunkt stattgefunden haben kann.

In diesem Fall ist analog zur Beschreibung der Einstiegshaltestelle vorzugehen: Für die weiteren Schritte der Fuzzifizierung sind lediglich die möglichen Einstiegshaltestellen zu berücksichtigen und die Einstiegswahrscheinlichkeiten entsprechend Formel 3-13 zu berechnen. Für die Berechnung der Werte für die einzelnen Quelle-Ziel-Relationen wird entsprechend davon ausgegangen, dass die Erfassung an der Haltestelle die letzte Erfassung des Geräts darstellt.

Quelle-Ziel-Relationen bei Zusatzinformationen zur Einstiegshaltestelle (ohne Zusatzinformationen zur Ausstiegshaltestelle)

	1	2	3	4	5	6	7	8	9
1	-								
2	-	-							
3	-	-	-						
4	-	-	-	-		X			
5	-	-	-	-	-				
6	-	-	-	-	-	-			
7	-	-	-	-	-	-	-		
8	-	-	-	-	-	-	-	-	
9	-	-	-	-	-	-	-	-	-

Quelle-Ziel-Relationen bei Zusatzinformationen zur Ausstiegshaltestelle (ohne Zusatzinformationen zur Einstiegshaltestelle)

	1	2	3	4	5	6	7	8	9
1	-								
2	-	-							
3	-	-	-						
4	-	-	-	-		X			
5	-	-	-	-	-				
6	-	-	-	-	-	-			
7	-	-	-	-	-	-	-		
8	-	-	-	-	-	-	-	-	
9	-	-	-	-	-	-	-	-	-

Quelle-Ziel-Relationen bei Zusatzinformationen zur Ausstiegshaltestelle (mit Zusatzinformationen zur Einstiegshaltestelle)

	1	2	3	4	5	6	7	8	9
1	-								
2	-	-							
3	-	-	-						
4	-	-	-	-		X			
5	-	-	-	-	-				
6	-	-	-	-	-	-			
7	-	-	-	-	-	-	-		
8	-	-	-	-	-	-	-	-	
9	-	-	-	-	-	-	-	-	-

- X Zuordnung der Quelle-Ziel-Relation auf Basis der erfassten Signale ohne Berücksichtigung von Unschärfen
- Zuordnung der Quelle-Ziel-Relation auf Basis der erfassten Signale mit Berücksichtigung von Unschärfen

Abbildung 3-19: Ermittlung möglicher Quelle-Ziel-Relationen unter Berücksichtigung von zusätzlichen Daten

Die Wahrscheinlichkeit für das Auftreten zusätzlicher Informationen zur Bestimmung der Ausstiegshaltestelle wird im Vergleich zur Anwendung zusätzlicher Informationen zur Bestimmung der Einstiegshaltestelle als gering angesehen: Während an Einstiegshaltestellen in der Regel Wartezeiten der Fahrgäste vorliegen, fehlen ebendiese Aufenthaltszeiten an den Ausstiegshaltestellen.

Einbezug von weiteren Fahrzeugdaten

Der Einbezug weiterer Fahrzeugdaten kann in Abhängigkeit der vorhandenen Daten entsprechend dem in Kapitel 3.3.2.2 beschriebenem Vorgehen zu einer genauen Angabe der Ein- bzw. Ausstiegshaltestelle führen.

In dem Fall ist die Fuzzifizierung auf Basis der zusätzlichen Informationen zur Einstiegshaltestelle (Abbringer) bzw. Ausstiegshaltestelle (Zubringer) analog zu den Beschreibungen bei Einbezug von Haltestellendaten anzupassen (Abbildung 3-19).

3.3.3.4 Kritische Würdigung

Bei der Anwendung des beschriebenen Verfahrens zur Fuzzifizierung ist neben der grundsätzlichen Kritik bei der Erstellung von Quelle-Ziel-Matrizen auf Basis der WLAN- und Bluetooth-Daten (Kapitel 3.3.2.3) zu beachten, dass es zu einer Überschätzung der Wahrscheinlichkeiten „kleiner“ Zeitklassen (WLAN) kommt.

Die derzeitige Wahrscheinlichkeitsverteilung des statischen Ansatzes basiert auf allen erfassten Signalen eines Fahrzeugs und somit auf Signalen von Fahrgästen und Nicht-Fahrgästen¹⁴ (Dataset Bieland/Mobile Data Fusion 2022). Vor dem Hintergrund, dass bei aktiver Nutzung von Smartphones i. d. R. eine Vielzahl an Signalen ausgesendet werden, kann aus Sicht des Autors eine Verzerrung der genutzten Wahrscheinlichkeitsverteilung aufgrund von Störsignalen vorliegen, durch die insbesondere „kleine“ Zeitklassen überschätzt werden:

Aufgrund zufälliger Situationen, bei der sich eine oder mehrere Personen in Reichweite des Erfassungsfahrzeugs befinden und ihr Smartphone aktiv nutzen, werden diese Endgeräte ggf. mehrfach während einer (Vorbei-)Fahrt erfasst. Sofern mehr als ein Signal erfasst wurde, kann für diese Geräte eine zeitliche Differenz gebildet werden, die in die Wahrscheinlichkeitsverteilung eingeht und zu einer Erhöhung der Wahrscheinlichkeiten von Zeitintervallen mit geringem zeitlichen Abstand führen. Dagegen gehen tendenziell Signale von Endgeräten, die sich nur kurz im Empfangsbereich des Erfassungsfahrzeugs befanden, jedoch nicht aktiv genutzt werden, aufgrund der – wenn überhaupt Signale empfangen wurden – einmaligen Erfassung nicht in die Erstellung der Wahrscheinlichkeitsverteilung ein, da bei einmaliger Erfassung kein zeitlicher Abstand zwischen Signalen ermittelt werden kann.

Darüber hinaus liegen insbesondere bei WLAN-Signalen Unsicherheiten aufgrund des Nutzungsverhaltens der Smartphones vor: Wird das Smartphone aktiv genutzt, wird es mit hoher Wahrscheinlichkeit in kurzer Zeit mehrfach erfasst. Wird das Smartphone jedoch nicht genutzt (Inaktivität) und / oder ist mit einem Access Point verbunden (Fahrzeug-WLAN), ist die Wahrscheinlichkeit hoch, das Gerät nur selten (ein bis keinmal) während einer Linienfahrt zu erfassen.

3.4 Beschreibung von Verfahren zum Matrixausgleich

Ziel des Verfahrens ist die Darstellung von Fahrgastwegen einer Linienfahrt (Einzelwege von Fahrgästen) auf Grundlage der gefilterten WLAN- und Bluetooth-Daten (Ausgangsmatrix). Neben den Matrixelementen der Ausgangsmatrix liegen die wahren Ein- und Ausstiegszahlen je Haltestelle aus den AFZS-Daten als Randsummen vor. Da sich die Randsummen aus der

¹⁴ Wie in Kapitel 3.2.3 sowie Kapitel 4.7 beschrieben, kann die Filterung nicht dem Anspruch gerecht werden, ausschließlich Daten von Nicht-Fahrgästen zu identifizieren.

Ausgangsmatrix und den AFZS-Daten mit einer hohen Wahrscheinlichkeit unterscheiden, werden mathematische Verfahren genutzt, um die Ausgangsmatrix so anzupassen, dass die Randsummen der AFZS-Daten eingehalten werden. Diese mathematischen Verfahren werden als Matrixausgleichsverfahren bezeichnet. Matrixausgleichsverfahren haben das Ziel, eine Quelle-Ziel-Matrix (Bewertungsmatrix) zu finden, die die Verkehrsnachfrage unter den gegebenen Rahmenbedingungen (Randsummen) und Input-Daten (Ausgangsmatrix) wiedergibt.

Im Folgenden werden zunächst mathematische Verfahren zum Ausgleich der abgeleiteten Quelle-Ziel-Matrix (Ausgangsmatrix) auf Basis der Randsummen vorgestellt. Darüber hinaus werden Verfahren vorgestellt, die rein auf den Randsummen basieren und keine Ausgangsmatrix benötigen. Die im Rahmen des entwickelten Verfahrens ausgewählten Hochrechnungsverfahren werden in Kapitel 3.4.4 beschrieben.

3.4.1 Verfahren auf Basis von Ausgangsmatrizen

Steigerungsfaktorenmodelle stellen laut Lohse und Schnabel (2011) eine Methode zur groben und schnellen Hochrechnung

- einer repräsentativ (Stichprobe, z. B. Erhebungsdaten aus Befragungsdaten oder weiterer Datenerfassungen) erfassten Verkehrsstrommatrix (=Quelle-Ziel-Matrix, Ausgangsmatrix) auf die Grundgesamtheit bzw.
- von einem Analysezustand – vollständig oder repräsentativ erfasste Verkehrsstrommatrix (Ausgangsmatrix) – auf einen Prognosezustand (Bewertungsmatrix) dar.

Grundsätzlich lassen sich einseitig und zweiseitig gekoppelte Modelle unterscheiden. Einseitig gekoppelte Modelle kommen dann zum Einsatz, wenn Informationen zum Quellverkehrs- oder Zielverkehrsaufkommen vorliegen. Im vorliegenden Anwendungsfall liegen Informationen sowohl zum Quell- als auch zum Zielverkehrsaufkommen vor. Daher beziehen sich die folgenden Beschreibungen auf zweiseitig gekoppelte Modelle.

Steigerungsfaktorenmodelle folgen dem Grundsatz, dass das Verkehrsaufkommen im Prognosefall ($V(P)$, $Q_o(P)$, $Z_d(P)$) durch Anwendung von Steigerungsfaktoren (f , f_{q_o} , f_{z_d}) auf Basis des Verkehrsaufkommens im Analysefall ($V(A)$, $Q_o(A)$, $Z_d(A)$) hochgerechnet werden kann (Formel 3-15 bis Formel 3-17) (Lohse und Schnabel 2011; Ortúzar und Willumsen 2004). Das Verkehrsaufkommen im Analysefall ist dabei als Ausgangsmatrix zu sehen, die

die Bewertungen der einzelnen Quelle-Ziel-Beziehungen, beispielsweise auf Basis von Befragungsdaten, Attraktivitäten oder Widerständen wiedergibt (Bregman 1967; Lamond und Stewart 1981; Ben-Akiva et al. 1985). Die Steigerungsfaktoren ergeben sich aus der Differenz zwischen Prognose- und Analysewert der Quell- und Zielverkehrsaufkommen.

$$D(P) = f * D(A) \quad \text{Formel 3-15}$$

$$Q_o(P) = f q_o * Q_o(A) \quad \text{Formel 3-16}$$

$$Z_d(P) = f z_d * Z_d(A) \quad \text{Formel 3-17}$$

mit

o	Nummer der Quellverkehrszelle
d	Nummer der Zielverkehrszelle
$D(A), D(P)$	Verkehrsaufkommen D in der Analyse (A) und Prognose (P)
$Q_o(A), Q_o(P)$	Quellverkehrsaufkommen der Verkehrszelle o in der Analyse (A) und Prognose (P)
$Z_d(A), Z_d(P)$	Zielverkehrsaufkommen der Verkehrszelle d in der Analyse (A) und Prognose (P)
$f, f q_o, f z_d$	Steigerungsfaktoren für die Verkehrsaufkommen V, Q_o und Z_d

Die Steigerungsfaktoren ergeben sich aus der Umstellung der genannten Gleichungen (Formel 3-18):

$$f = \frac{D(P)}{D(A)} \quad f q_i = \frac{Q_o(P)}{Q_o(A)} \quad f z_j = \frac{Z_d(P)}{Z_d(A)} \quad \text{Formel 3-18}$$

Für die Berechnung der Verkehrsströme ist die Art der Randsummen und der daraus folgenden Bedingungen für das Rechenverfahren relevant. Es können unelastische (fixierte und harte) sowie elastische Randsummen unterschieden werden (Lohse und Schnabel 2011).

- Fixierte Randsummen: Fixierte Randsummen müssen durch die Berechnung des jeweiligen Verfahrens exakt eingehalten werden. In Abgrenzung dazu stehen
- harte Randsummen: Bei harten Randsummen muss der Randsummenwert lediglich näherungsweise eingehalten werden.

Für fixierte und harte Randsummen wird vorausgesetzt, dass diese ausreichen, um den Prognosewert in ausreichender Genauigkeit zu berechnen. Weitere beeinflussende Faktoren spielen hier eine untergeordnete Rolle. Bei elastischen Randsummen wirken weitere Faktoren, wie etwa die Erreichbarkeit auf die Verteilung des Gesamtverkehrsaufkommens ein (Lohse und Schnabel 2011).

Im Weiteren wird lediglich auf die Berechnungsverfahren für fixierte Randsummen eingegangen. Im vorliegenden Anwendungsfall (Ermittlung der Quelle-Ziel-Beziehungen von Fahrgästen im ÖPNV) liegen fixierte Randsummen auf Grundlage der AFZS in hoher Qualität und Auflösung vor. Die AFZS liefern die Ein- und Aussteigerzahlen (Q_o , Z_d) je Haltestelle auf dem Linienweg einer Linienfahrt. Die Matrixelemente des Verkehrsaufkommens im Analysefall ergeben sich aus den erfassten WLAN- und Bluetooth-Daten.

Die zu berechnenden Verkehrsströme im Prognosefall (d_{od}) müssen entsprechend des Grundmodells mit unelastischen Randsummenbedingungen (Lohse und Schnabel 2011) (Formel 3-19) den drei Randsummenbedingungen (Formel 3-20 bis Formel 3-22) genügen:

$$d_{od} = B_{od} * f_{q_o} * f_{z_d} \quad \text{Formel 3-19}$$

$$\hat{Q}_o = Q_o = \sum_d d_{od} \quad \text{Formel 3-20}$$

$$\hat{Z}_d = Z_d = \sum_o d_{od} \quad \text{Formel 3-21}$$

$$\hat{D} = \sum_o \hat{Q}_o = \sum_d \hat{Z}_d = \sum_o \sum_d d_{od} \quad \text{Formel 3-22}$$

mit

o	Haltestellennummer der Einstiegshaltestelle (Quelle)
d	Haltestellennummer der Ausstiegshaltestelle (Ziel)
d_{od}	Fahrgastfahrten zwischen den Haltestellen o und d
B_{od}	Bewertung der Verkehrsbeziehung zwischen o und d
f_{q_o}, f_{z_d}	Steigerungsfaktoren für die Verkehrsaufkommen Q_o und Z_d
\hat{Q}_o	Vorgegebene Einsteigerzahlen an Haltestelle o (Quellverkehrsaufkommen, Randsumme)
Q_o	Berechnete Einsteigerzahlen an Haltestelle o
\hat{Z}_d	Vorgegebene Aussteigerzahlen an Haltestelle d (Zielverkehrsaufkommen, Randsumme)
Z_d	Berechnete Aussteigerzahlen an Haltestelle d
\hat{D}	Vorgegebene Fahrgastzahlen (Randsumme)

Zur Lösung des Grundmodells eines bi- oder n-linearen Gleichungssystem, wie es mit unelastischen Randsummenbedingungen der Fall ist, werden häufig Näherungsverfahren eingesetzt, sogenannte Steigerungsfaktorenmodelle. Bei diesen Modellen „werden die Randsummenbedingungen eingehalten und die Minimierung des Informationsgewinns gewährleistet“ (Lohse und Schnabel 2011). Mit dem Prinzip der Informationsminimierung wird sichergestellt, dass die Verkehrsstrommatrix im Prognosefall die bekannten Informationen der Verkehrsstrommatrix im Analysezustand auf Grundlage der

Randsummenbedingungen erhält. Der Minimierungsbegriff nimmt Bezug auf die augenscheinlich zusätzlichen Informationen, die eine vollständig geschätzte (prognostizierte) Verkehrsstrommatrix gegenüber der Ausgangsmatrix bzw. den Randsummen enthält (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 1995). Näheres zur Ermittlung des Informationsgewinns als Differenz zwischen den Verkehrsstrommatrizen im Analyse- und Prognosefall zeigen (Lohse und Schnabel 2011).

Die Steigerungsfaktoren f_{q_o} (Einsteiger) und f_{z_d} (Aussteiger) sind so zu bestimmen, dass die vorgegebenen Randsummen (\hat{D} , \hat{Q}_o , \hat{Z}_d) bei der Berechnung der Verkehrsströme eingehalten werden. B_{od} stellt hier die Ausgangsmatrix als Grundlage der Lineartransformation (Hochrechnung) zur Lösungsfindung dar und kann auf Zähl- oder Befragungsdaten basieren bzw. im Anwendungsfall auf der aus WLAN- und Bluetooth-Daten abgeleiteten Quelle-Ziel-Matrix. Formel 3-22 impliziert, dass alle Wege, die begonnen werden, auch ein Ziel im Untersuchungsraum haben. Dies ist für den Anwendungsfall der ÖPNV-Linienfahrten gegeben.

Formel 3-23 bis Formel 3-25 verdeutlichen darauf aufbauend einen prinzipiellen Lösungsweg (Lohse und Schnabel 2011).

$$\hat{Q}_o = \sum_d d_{od} = \sum_d B_{od} * f_{q_o} * f_{z_d} = f_{q_o} * \sum_d B_{od} * f_{z_d}$$

$$\rightarrow f_{q_o} = \frac{\hat{Q}_o}{\sum_d B_{od} * f_{z_d}}$$

Formel 3-23

$$\hat{Z}_d = \sum_o d_{od} = \sum_o B_{od} * f_{q_o} * f_{z_d} = f_{z_d} * \sum_o B_{od} * f_{q_o}$$

$$\rightarrow f_{z_d} = \frac{\hat{Z}_d}{\sum_o B_{od} * f_{q_o}}$$

Formel 3-24

$$d_{od} = B_{od} * \frac{\hat{Q}_o}{\sum_d B_{od} * f_{z_d}} * \frac{\hat{Z}_d}{\sum_o B_{od} * f_{q_o}}$$

Formel 3-25

Werden diese Formeln als Rekursionsformeln definiert, kann die Gleichung iterativ gelöst werden (Lohse und Schnabel 2011). Eine *Rekursionsformel* (lat. recurrere bedeutet zurücklaufen) bezeichnet eine Gleichung, die sich selbst abbildet. Die Elemente eines Iterationsschritts (A_n) bestimmen sich aus dem Wert des Elements im vorherigen Schritt (A_{n-1}). Um den iterativen Lösungsprozess anzustoßen, wird

- die Ausgangsmatrix des Lösungsprozesses neu bezeichnet: $B_{od} = d_{od}(1)$ und
- die Faktoren f_{q_o} und f_{z_d} mit 1 angesetzt.

Der Iterationsprozess führt zu einer Annäherung an die Lösung entsprechend der gegebenen Rahmenbedingungen. Dabei nähern sich mit zunehmender Iteration die ermittelten Randsummen den vorgegebenen Randsummen an. Das heißt, dass sich der Quotient zwischen geschätzter und gemessener Anzahl an Ein- und Aussteigern je Haltestelle (Formel 3-18) zunehmend dem Wert 1 annähert. Gleichzeitig nähern sich die Faktoren f_{q_o} und f_{z_d} einem konstanten Wert an, sodass das Gleichungssystem gegen eine stabile Lösung konvergiert (Lohse und Schnabel 2011).

Notwendige Bedingung zur Lösbarkeit des Gleichungssystems ist die Existenz einer Matrix, die den Randsummenbedingungen genügt sowie allen Fahrgastfahrten eine Quelle und ein Ziel zuordnet (Lohse 1997). Eine Konvergenz kann also nicht für alle Ausgangsmatrizen gewährleistet werden. Gleichzeitig bestehen für lösbar bilineare Gleichungssysteme meist mehrere Lösungsmöglichkeiten. In dem Fall ist die (Näherungs-)Lösung (Verkehrstrommatrix im Prognosefall) zu wählen, die, wie oben beschrieben, den geringsten Informationsgewinn besitzt und die Randsummenbedingungen hinreichend genau erfüllt (Lohse und Schnabel 2011).

Die durch Steigerungsfaktorenmodelle ermittelte Lösungen sind stark von der Ausgangsmatrix abhängig, da dort Informationen zu den Bewertungen der einzelnen Quelle-Ziel-Relationen enthalten sind. Daher ist der Aufstellung der Ausgangsmatrix höchste Aufmerksamkeit zu widmen:

- In der Ausgangsmatrix vorhandene und ggf. unbekannte Verzerrungen (durch Erhebungsfehler) können durch die Steigerungsfaktorenmodelle nicht ausgeglichen werden und wirken sich daher auf die Genauigkeit der Schätzung aus. Die Ergebnisse der Schätzung sind daher vor dem Hintergrund systematischer Verzerrungen stets zu überprüfen (Keppeler und Kröpel 2014).
- Besitzen einzelne Verkehrsströme in der Ausgangsmatrix den Wert „0“, werden diese aufgrund der multiplikativen Form (Formel 3-25) auch in der berechneten Matrix den Wert „0“ behalten (Problem of non-structural zeros (Ben-Akiva et al. 1985)). Bei einem Fall, in dem eine ganze Zeile oder Spalte aus „0“-Werten besteht, während die Randsumme einen Wert größer als „0“ besitzt, kann das Iterationsverfahren entsprechend

nicht konvergieren, sodass keine gültige Lösung gefunden werden kann. Nach Ortúzar und Willumsen (2004) ist eine geringe Konvergenz nach wenigen Iterationsschritten ein mögliches Zeichen, dass falsche „0“-Werte bestehen und an den entscheidenden Stellen in der Ausgangsmatrix eingesetzt sind.

Ein Abbruch des Iterationsprozesses kann bei Erreichen einer hohen Übereinstimmung zwischen den berechneten und vorgegebenen Werten des Quell- und Zielverkehrsaufkommens erfolgen. Die Berechnung kann auf Basis der Steigerungsfaktoren sowie eines Schwellenwertes (E) erfolgen (Formel 3-26, Formel 3-27). Wird der Schwellenwert E – beispielsweise $E = 0,01$ – unterschritten, liegt eine ausreichende Genauigkeit (= geringe Abweichung) der Randsummen vor.

$$|1 - f_{q_o}| < E \quad \text{Formel 3-26}$$

$$|1 - f_{z_d}| < E \quad \text{Formel 3-27}$$

Bei großen Verkehrsaufkommen sollten die Genauigkeitsanforderungen erhöht werden, etwa durch einen Schwellenwert, der auf relativen Abweichungen basiert (Lohse und Schnabel 2011). Gleichzeitig ist der Fall zu berücksichtigen, in dem das Iterationsverfahren nicht konvergiert und das genannte Abbruchkriterium nie erreicht wird. Stattdessen finden zwar weitere Iterationen statt, die Genauigkeit der Abschätzung kann jedoch nicht weiter verbessert werden. Für diese Fälle ist eine maximale Zahl an Iterationen zu wählen, um den Algorithmus zu stoppen (beispielsweise $n = 100$ Iterationen).

Im Ergebnis gleichen Steigerungsfaktorenmodelle die Randsummen zwischen Analyse- und Prognosefall auf Basis der Bewertungsinformationen der Stichprobe (Ausgangsmatrix) aus. Steigerungsfaktorenmodelle dürfen entsprechend nur dann angewandt werden, wenn die Ausgangsmatrix den Prognosezustand repräsentiert (Lohse und Schnabel 2011). Die Anwendung der Steigerungsfaktorenmodelle sollten vorzugsweise je Linienfahrt angewandt werden (Keppeler und Kröpel 2014; Neumann 2017).

Bekannte Steigerungsfaktorenmodelle zur Bestimmung der Verkehrsströme d_{od} sind das *Furness-Verfahren*, das *Fratar-Verfahren*, das *Detroitmodell* und das *Multimodell* sowie weitere Varianten (Lohse und Schnabel 2011). Diese basieren „mehr oder weniger“ (Lohse 1997, S. 97) auf den hier dargestellten Beziehungen in Formel 3-23 und Formel 3-24 und werden als iterative

Randsummenverfahren-Verfahren (engl.: iterative proportional fitting (IPF)) bezeichnet. Die Wahl des IPF-Verfahrens beeinflusst das Ergebnis damit nicht maßgeblich. Unterschiede zwischen den Verfahren liegen im Rechenaufwand und der Konvergenzgeschwindigkeit aufgrund zusätzlicher Faktoren (Lohse und Schnabel 2011; Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2021).

3.4.1.1 Furness-Verfahren

Das Furness-Verfahren wurde bereits in den 1930er Jahren verwendet und dient der iterativen Lösung des bilinearen Problems (Furness 1965; Lohse 1997) (Formel 3-25). Bei der Anwendung des Furness-Verfahrens werden zunächst geeignete Startwerte für $f_{q_o}(1)$, $f_{z_d}(1)$ und $d_{od}(1) = B_{od}$ festgelegt. Anschließend werden anhand von Formel 3-23 und Formel 3-24 die Näherungswerte für $f_{q_o}(n)$ und $f_{z_d}(n)$ in den Iterationsschritten abwechselnd auf Basis der vorliegenden Randsummen berechnet und so die Matrixelemente $d_{od}(n)$ bestimmt (Formel 3-28). Mit der abwechselnden Herangehensweise werden auch in der Matrix abwechselnd die Zeilen oder die Spalten so korrigiert, dass die jeweiligen Randsummenbedingungen exakt eingehalten werden. Gleichzeitig werden die Abweichungen der jeweils anderen Summe von den vorgegebenen Randsummen schrittweise besser erfüllt. Das Furness-Verfahren wird als rasch konvergierend bezeichnet (Lohse 1997; Mekky 1983).

$$d_{od}(n+1) = \begin{cases} B_{od}(n) * f_{q_o}(n+1) * f_{z_d}(n) = d_{od}(n) * \frac{\hat{Q}_o}{\sum_r d_{or}} & \text{für } n \text{ gerade} \\ B_{od}(n) * f_{q_o}(n) * f_{z_d}(n+1) = d_{od}(n) * \frac{\hat{Z}_d}{\sum_s d_{sd}} & \text{für } n \text{ ungerade} \end{cases} \quad \text{Formel 3-28}$$

mit

$d_{od}(n)$	Verkehrsstrom zwischen den Haltestellen o und d in Iterationsschritt n (für n = 0: Ausgangsmatrix)
o, d, r, s	Index einer Haltestelle als Quelle oder Ziel einer Fahrt
B_{od}	Bewertung der Verkehrsbeziehung zwischen o und d in Iterationsschritt n (für n = 0: Ausgangsmatrix)
$f_{q_o}(n)$	Steigerungsfaktor der Quelle (Haltestelle) o im Iterationsschritt n
$f_{z_d}(n)$	Steigerungsfaktor des Ziels (Haltestelle) d im Iterationsschritt n
\hat{Q}_o	Vorgegebene Einsteigerzahlen an Haltestelle o (Quellverkehrsaufkommen, Randsumme)
\hat{Z}_d	Vorgegebene Aussteigerzahlen an Haltestelle d (Zielverkehrsaufkommen, Randsumme)

3.4.1.2 Fratar-Verfahren

Die zugrundeliegende Methode des Fratar-Verfahrens (Fratar 1954; Lohse 1997) führt zwei Schritte des Furness-Verfahrens simultan unter Bildung

zusätzlicher Steigerungsfaktoren durch, die die durchschnittlichen Hochrechnungsfaktoren der jeweils anderen Randsummen berücksichtigen (Formel 3-29).

$$d_{od}(n+1) = d_{od}(n) * f_{q_o}(n) * f_{z_d}(n) * \frac{1}{2} \left[\frac{1}{\overline{f_{q_o}(n)}} + \frac{1}{\overline{f_{z_d}(n)}} \right] \quad \text{Formel 3-29}$$

$$f_{q_o}(n) = \frac{\hat{Q}_o}{Q_o(n)} = \frac{\hat{Q}_o}{\sum_s d_{os}(n)} \quad ; \quad f_{z_d}(n) = \frac{\hat{Z}_d}{Z_d(n)} = \frac{\hat{Z}_d}{\sum_r d_{rd}(n)} \quad \text{Formel 3-30}$$

$$\overline{f_{q_o}(n)} = \frac{\sum_s d_{os}(n) * f_{z_s}(n)}{\sum_s d_{os}(n)} \quad ; \quad \overline{f_{z_d}(n)} = \frac{\sum_r d_{rd}(n) * f_{q_r}(n)}{\sum_r d_{rd}(n)} \quad \text{Formel 3-31}$$

mit:

$d_{od}(n)$	Verkehrstrom zwischen den Haltestellen o und d in Iterationsschritt n (für n = 0: Ausgangsmatrix)
o, d, r, s	Index einer Haltestelle als Quelle oder Ziel einer Fahrt
$f_{q_o}(n)$	Steigerungsfaktor der Quelle (Haltestelle) o in Iterationsschritt n (Formel 3-30)
$f_{z_d}(n)$	Steigerungsfaktor des Ziels (Haltestelle) d in Iterationsschritt n (Formel 3-30)
$\overline{f_{q_o}(n)}$	Durchschnittlicher Steigerungsfaktor der Quelle (Haltestelle) o in Iterationsschritt n unter Einbezug des Steigerungsfaktors des Ziels s (z_s) (Formel 3-31)
$\overline{f_{z_d}(n)}$	Durchschnittlicher Steigerungsfaktor des Ziels (Haltestelle) d in Iterationsschritt n unter Einbezug des Steigerungsfaktors der Quelle r (q_r) (Formel 3-31)
\hat{Q}_o	Vorgegebene Einsteigerzahlen an Haltestelle o (Quellverkehrsaufkommen, Randsumme)
$Q_o(n)$	Berechnete Einsteigerzahlen an Haltestelle o in Iterationsschritt n
\hat{Z}_d	Vorgegebene Aussteigerzahlen an Haltestelle d (Zielverkehrsaufkommen, Randsumme)
$Z_d(n)$	Berechnete Aussteigerzahlen an Haltestelle d in Iterationsschritt n

3.4.1.3 Detroit-Verfahren

Der Iterationsprozess des Detroit-Verfahrens (Formel 3-32) ähnelt grundsätzlich dem des Furness-Verfahrens. Im Detroit-Verfahren werden die Näherungswerte der Steigerungsfaktoren f_{q_o} und f_{z_d} gleichzeitig in jedem Iterationsschritt ermittelt. Darüber hinaus wird ein zusätzlicher von o und d unabhängiger Korrekturfaktor, der nach Lohse (1997) zur Sicherung der Konvergenz des Iterationsprozesses dient, eingeführt. Dieser Faktor berücksichtigt das Verhältnis zwischen gezählten (\hat{D}) und geschätzten (d_{sr}) Fahrgästen.

$$d_{od}(n+1) = d_{od}(n) * f_{q_o}(n) * f_{z_d}(n) * \frac{1}{\hat{D}} * \sum_s \sum_r d_{sr}(n) \quad \text{Formel 3-32}$$

mit:

$d_{od}(n)$	Verkehrsstrom zwischen den Haltestellen o und d in Iterationsschritt n (für n = 0: Ausgangsmatrix)
o, d, r, s	Index einer Haltestelle als Quelle oder Ziel einer Fahrt
$f_{q_o}(n)$	Steigerungsfaktor der Quelle (Haltestelle) o in Iterationsschritt n (Formel 3-30)
$f_{z_d}(n)$	Steigerungsfaktor des Ziels (Haltestelle) d in Iterationsschritt n (Formel 3-30)
\hat{D}	Vorgegebene Fahrgastzahlen (Randsumme)

3.4.1.4 MULTI-Verfahren

Das MULTI-Verfahren (Formel 3-33) wurde durch (Lohse und Schnabel 2011; Lohse 1977) auf Basis des Detroit- und Fratar-Verfahrens mit dem Ziel, eine Beschleunigung der Konvergenz zu erreichen, entwickelt. Zur Erhöhung der Konvergenzgeschwindigkeit werden daher zusätzliche Faktoren gegenüber dem Detroit-Verfahren eingeführt. Dazu zählen Faktoren aus dem Fratar-Verfahren (Formel 3-30, Formel 3-31), die die durchschnittlichen Hochrechnungsfaktoren der jeweils anderen Randsummen berücksichtigen sowie ein Faktor ($f(n)$, Formel 3-34), der das Verhältnis zwischen gezählten und geschätzten Fahrgästen berücksichtigt.

$$d_{od}(n+1) = d_{od}(n) * \frac{f_{q_o}(n)}{\overline{f_{q_o}}(n)} * \frac{f_{z_d}(n)}{\overline{f_{z_d}}(n)} * f(n) \quad \text{Formel 3-33}$$

$$f(n) = \frac{\hat{D}}{D(n)} = \frac{\hat{D}}{\sum_o \sum_d d_{od}(n)} \quad \text{Formel 3-34}$$

mit:

$d_{od}(n)$	Verkehrsstrom zwischen den Haltestellen o und d in Iterationsschritt n (für n = 0: Ausgangsmatrix)
$f_{q_o}(n)$	Steigerungsfaktor der Quelle (Haltestelle) o in Iterationsschritt n (Formel 3-30)
$f_{z_d}(n)$	Steigerungsfaktor des Ziels (Haltestelle) d in Iterationsschritt n (Formel 3-30)
$\overline{f_{q_o}}(n)$	Durchschnittlicher Steigerungsfaktor der Quelle (Haltestelle) o in Iterationsschritt n unter Einbezug des Steigerungsfaktors des Ziels d (z_d) (Formel 3-31)
$\overline{f_{z_d}}(n)$	Durchschnittlicher Steigerungsfaktor des Ziels (Haltestelle) d in Iterationsschritt n unter Einbezug des Steigerungsfaktors der Quelle o (q_o) (Formel 3-31)
$f(n)$	Korrekturfaktor für den Iterationsschritt n (Formel 3-34)
\hat{D}	Vorgegebene Fahrgastzahlen (Randsumme)
$D(n)$	Fahrgastzahlen im Iterationsschritt n

3.4.1.5 Durchschnittsfaktorenmethode

Die Durchschnittsfaktorenmethode, oder auch Mittelwertverfahren, beschreibt ein Verfahren, das ursprünglich für die Hochrechnung von

Befragungsdaten von Verkehrserhebungen im ÖPNV-Betrieb entwickelt wurde und explizit das häufig bei Fahrgastbefragungen auftretende Problem der Untererfassung von Kurzstreckenfahrten aufgreift (Keppeler und Kröpel 2014).

Das sogenannte „short trip bias“ (Navick und Furth 1994) beschreibt die Verzerrung in den Befragungsdaten, die dadurch auftreten, dass Kurzstreckenfahrer aufgrund der mit der kurzen Reisedauer verbundenen kürzeren Aufenthaltszeit im Fahrzeug eine geringere Auswahlwahrscheinlichkeit bei der Befragung haben. Dabei besitzen auch weitere Faktoren eine Bedeutung, wie die Reisedauer, die Anzahl an Personen im Fahrzeug oder die Anzahl der Interviewer (Sommer et al. 2011; Keppeler und Kröpel 2014).

In dem Verfahren erfolgt die Hochrechnung in nur einem Iterationsschritt unter Berücksichtigung eines mittleren Hochrechnungsfaktors für die Ein- und Aussteigerseite. Dabei werden die Differenzen der Ein- und Aussteiger zwischen Analyse- und Prognosefall für jeden Halt auf Basis einer Mittelwertbildung über den Iterationsprozess ausgeglichen (Formel 3-36). Die Mittelwertbildung (MW) folgt Formel 3-35 und berücksichtigt einen Faktor k . Erläuterungen zu dem Faktor, Quellen hinsichtlich des genutzten Verfahrens oder eine Herleitung des eingeführten Faktors k werden nicht gegeben (Keppeler und Kröpel 2014).

$$MW_{od} = \sqrt[k]{\frac{fq_o^k + fz_d^k}{2}} \quad \text{Formel 3-35}$$

$$d_{od}(n+1) = d_{od}(n) * \sqrt[k]{\frac{fq_o^k + fz_d^k}{2}} \quad \text{Formel 3-36}$$

mit:

MW_{od}	Mittelwert der Steigerungsfaktoren unter Berücksichtigung von k
o, d	Index einer Haltestelle als Quelle (o) oder Ziel (d) der Linienfahrt
k	Parameter k (z. B. $k = 1$ oder $k = 2$)
$d_{od}(n)$	Verkehrsstrom zwischen den Haltestellen o und d in Iterationsschritt n (für $n = 0$: Ausgangsmatrix)
$fq_o(n)$	Steigerungsfaktor der Quelle (Haltestelle) o in Iterationsschritt n (Formel 3-30)
$fz_d(n)$	Steigerungsfaktor des Ziels (Haltestelle) d in Iterationsschritt n (Formel 3-30)

3.4.2 Verfahren ohne Ausgangsmatrix

3.4.2.1 Verfahren nach Li & Cassidy

Li und Cassidy (2007) haben ein Verfahren (weiter-)entwickelt, das eine Quelle-Ziel-Matrix allein auf Basis von Randsummen berechnet. In diesem Verfahren wird entsprechend keine Ausgangsmatrix benötigt, sodass auch Verzerrungen bei der Erfassung der Daten keine Bedeutung haben. Die Matrix des Verfahrens soll als Vergleichsmatrix dienen.

Das entwickelte Verfahren basiert auf der rekursiven Methode zur Bestimmung der Fahrgastnachfrage, die erstmals durch Tsygalnitsky entwickelt wurde (Tsygalnitsky 1977). Da die Berechnungsvorschrift dabei unverändert bleibt (direkte Rekursion), können die Ergebnisse der Vorschleife in den neuen Aufruf eingehen. Dies wird solange durchgeführt, bis eine Abbruchbedingung erfüllt wird. Ziel dabei ist es, ein komplexes System in kleinere Probleme aufzuteilen und dadurch zu lösen. Bei der Iteration werden in Abgrenzung zur Rekursion dagegen bestimmte Anweisungen solange ausgeführt, bis eine Abbruchbedingung erfüllt wird (Vries und Weiß 2021).

Tsygalnitsky nutzt eine rekursive Methode, um aussteigende Fahrgäste zu identifizieren. Dazu gibt er jedem der im Fahrzeug befindlichen Fahrgäste die gleiche Wahrscheinlichkeit an einer Haltestelle auszusteigen. Zusteigende Fahrgäste werden dem Pool der im Fahrzeug befindlichen Fahrgäste zugefügt. Dabei kann auch eine Mindestreisedistanz der Fahrgäste eingeführt werden, sodass ein Fahrgast vor Erreichen der Mindestreisedistanz nicht als Aussteiger in Frage kommt (Tsygalnitsky 1977; Li und Cassidy 2007; Simon und Furth 1985).

Li und Cassidy (2007) weisen darauf hin, dass die Annahme einer gleichen Ausstiegswahrscheinlichkeit je Haltestelle unter allen Fahrgästen im Fahrzeug unabhängig von der Bedeutung von Haltestellen oder der zurückgelegten Distanz unrealistisch ist. Laut Ji et al. (2014) führt die Annahme von Tsygalnitsky zu einer Überrepräsentation von Fahrgastfahrten mit mittlerer Reiseweiten.

Fahrten von Fahrgästen beginnen oder enden in der Regel an Aktivitätszentren (z. B. Innenstadtbereiche, Schulen). Fahrgastfahrten zwischen Haltestellen ohne besondere Zielgelegenheiten sind dagegen unüblich. Vor diesem Hintergrund haben Li und Cassidy Haltestellen im Umfeld von Aktivitätenstandorten als „*Major-Haltestellen*“ bezeichnet, die übrigen Haltestellen – etwa in Wohngebieten – als „*Minor-Haltestellen*“. Die Bezeichnungen beziehen sich

dabei lediglich auf die Aktivitäten in der Umgebung, nicht auf die Zahl der Ein- und Aussteiger. Eine Erhebung in San Francisco stützt ihre Behauptung – hier wurden 85% der Wege an einer Major-Haltestelle begonnen oder beendet, während 15% der Fahrgastfahrten zwischen Minor-Haltestellen stattgefunden haben. Dabei ist zu berücksichtigen, dass sich gerade in urbanen Räumen eine Mischung von Aktivitätenstandorten (Major-Haltestellen) und Wohngelegenheiten (Minor-Haltestellen) vorliegen. Daher kann davon ausgegangen werden, dass die Ausstiegswahrscheinlichkeiten entgegen dem Ansatz von Tsygalnitsky nicht gleichverteilt sind.

Aufbauend auf der Annahme, dass Major-Haltestellen im Wesentlichen der Ausgangs- oder Zielpunkt von ÖV-Fahrten sind, haben Li und Cassidy ihr Verfahren entwickelt. Der Algorithmus wählt vor Erreichen einer Haltestelle unter

- den Fahrgästen, die an einer Major-Haltestelle zugestiegen sind (a) und
- den Fahrgästen, die an einer Minor-Haltestelle zugestiegen sind (b)

die aussteigenden Fahrgäste (Anzahl entsprechend der Randsumme) aus. Die Wahrscheinlichkeit für den Ausstieg eines Fahrgasts an der Haltestelle s wird mit p_{as} für Major-Haltestellen bzw. p_{bs} für Minor-Haltestellen bezeichnet. Durch die Variable α kann eine Unterscheidung der Ausstiegswahrscheinlichkeit zwischen Major- und Minor-Haltestellen berücksichtigt werden (Formel 3-37). α_a gilt dabei für Major-Haltestellen, α_b für Minor-Haltestellen. Die Methode von Tsygalnitsky sieht dabei keine Differenzierung zwischen Major- und Minor-Haltestellen vor, sodass hier $\alpha = \alpha_a = \alpha_b = 0,5$ gilt.

$$\frac{p_{as}}{p_{bs}} = \frac{1 - \alpha}{\alpha} \quad \text{Formel 3-37}$$

Die Wahrscheinlichkeit für den Ausstieg von Fahrgästen wird durch Formel 3-38 ermittelt. Diese Formel zielt darauf ab, den Erwartungswert der aussteigenden Fahrgäste an allen Haltestellen der Linienfahrt zu ermitteln, die an einer Major-Haltestelle eingestiegen sind (\widehat{n}_a). Die Anzahl der Fahrgäste im Fahrzeug mit Einstieg an eine Minor- oder Major-Haltestelle ($N_{a,i}, N_{b,i}$) sind aus Zähl- und Daten (z. B. AFZS) bekannt. Der Erwartungswert für diejenigen Fahrgäste, die an einer Minor-Haltestelle zugestiegen sind (\widehat{n}_b) ergibt sich aus der Differenz zwischen der Anzahl aussteigender Fahrgäste und dem Erwartungswert \widehat{n}_a (Formel 3-39).

$$\widehat{n}_{a,o} = \frac{(1 - \alpha_a) * N_{a,o}}{(1 - \alpha_a) * N_{a,o} + \alpha_a * N_{b,o}} * n_o \quad \text{Formel 3-38}$$

$$\widehat{n}_{b,o} = n_o - \widehat{n}_{a,o} \quad \text{Formel 3-39}$$

mit

$\widehat{n}_{a,o}$	Erwartungswert der Anzahl aussteigender Fahrgäste an Haltestelle o, mit Einstieg an einer Major-Haltestelle
$\widehat{n}_{b,o}$	Erwartungswert der Anzahl aussteigender Fahrgäste an Haltestelle o, mit Einstieg an einer Minor-Haltestelle
α_a	Faktor zur Bestimmung der Ausstiegswahrscheinlichkeit für Major-Haltestellen
$N_{a,o}$	Anzahl Fahrgäste an Haltestelle o im Fahrzeug mit Einstieg an Major-Haltestelle
$N_{b,o}$	Anzahl Fahrgäste an Haltestelle o im Fahrzeug mit Einstieg an Minor-Haltestelle
n_o	Anzahl aussteigender Fahrgäste (=Randsomme) an Haltestelle o

Auf Basis der aussteigenden Fahrgäste, differenziert nach Einstieg an einer Major- oder Minor-Haltestelle, soll im nächsten Schritt die Quelle-Ziel-Matrix erstellt werden. Da eine Linienfahrt in der Regel mehr als eine Major- und Minor-Haltestelle bedient, werden in diesem Schritt die Einstiegshaltestellen der aussteigenden Fahrgäste ermittelt. Dazu wird der Anteil der Einsteiger je Major- bzw. Minor-Haltestelle auf alle Einsteiger der Haltestellen, die im Fahrtverlauf vor der aktuell betrachteten Haltestelle liegen, bezogen.

Die Anzahl der Fahrgäste auf den jeweiligen Quelle-Ziel-Relationen ergibt sich durch Multiplikation des berechneten Anteils mit der Anzahl der aussteigenden Fahrgäste an der betrachteten Haltestelle. Dazu wird entsprechend Formel 3-40 die Ausstiegswahrscheinlichkeit für die zugestiegenen Fahrgäste an allen Haltestellen, differenziert nach Major- und Minor-Haltestellen, bestimmt. Das Vorgehen für aussteigende Fahrgäste mit Einstieg an einer Minor-Haltestelle erfolgt analog.

Beispiel: Betrachtet wird die Haltestelle #5 (Major-Haltestelle). An der Haltestelle steigen vier Fahrgäste aus, die vorher an Major-Haltestellen eingestiegen sind ($\widehat{n}_{a,5}$). Die Haltestellen #1 und #3 waren Major-Haltestellen, die Haltestellen #2 und #4 Minor-Haltestellen. An jeder der Haltestellen sind jeweils drei Fahrgäste zugestiegen. Vor Haltestelle #5 hat kein Ausstieg stattgefunden. Auf Basis von Formel 3-40 ergibt sich, dass ein Anteil von 50% (Einsteiger an Haltestelle #1 / Einsteiger an Haltestelle #1 und #3 = 3/6) der Aussteiger an Haltestelle #5 an Haltestelle #1 zugestiegen sind. Die verbleibenden 50% sind an Haltestelle 3 zugestiegen.

$$n_{a,od} = \frac{E_{ad}}{\sum_{d=1}^o E_{ad}} * \widehat{n}_{a,o} \quad \text{Formel 3-40}$$

mit

$n_{a,od}$	Anzahl aussteigender Fahrgäste an Haltestelle o mit Einstieg an Major-Haltestelle d
E_{ad}	Anzahl der einsteigenden Fahrgäste an Major-Haltestelle d
$\sum_{d=1}^o E_{ad}$	Summe der einsteigenden Fahrgäste an allen Major-Haltestellen d bis zur betrachteten Major-Haltestelle o
$\widehat{n}_{a,o}$	Erwartungswert der Anzahl aussteigender Fahrgäste an Haltestelle o, mit Einstieg an einer Major-Haltestelle

Darüber hinaus besteht die Möglichkeit, dass der Erwartungswert der Anzahl aussteigender Fahrgäste, mit Einstieg an einer Major-Haltestelle ($\widehat{n}_{a,o}$), die Anzahl an Fahrgästen überschreitet, die an einer Major-Haltestelle eingestiegen sind ($N_{a,o}$). In diesem Fall ist der „Überschuss“ aus den Fahrgästen zu ziehen, die an einer Minor-Haltestelle zugestiegen sind (analog zu $n_{a,od}$).

Wie eingangs erwähnt basiert das Verfahren von Li und Cassidy auf der Annahme, dass die Ausstiegswahrscheinlichkeiten in Abhängigkeit des Haltestellentyps variieren. Daher wurden die Werte für α_a und α_b näherungsweise durch ein iteratives Verfahren bestimmt. Das Verfahren hatte das Ziel, das Distanzmaß D auf Basis der mittleren berechneten und mittleren gemessenen Fahrzeugbelegung einer Auswahl von Linienfahrten auf Basis variabler Werte für α_a und α_b zu optimieren (Formel 3-41, $D \rightarrow \text{Min!}$). Die Auswahl der Linienfahrten erfolgte dabei unter Berücksichtigung möglichst homogener Einflussfaktoren wie Tageszeiten, Tagestyp oder Jahreszeit. Als Referenzwert kann dabei $\alpha_a = \alpha_b = 0,5$ (Tsygalnitsky) herangezogen werden.

In einem Fallbeispiel wurden alle Werte für α_a und α_b im Wertebereich von 0,1 bis 0,9 in Schritten von 0,1 getestet. Im Ergebnis der Untersuchung war das Distanzmaß für $\alpha_a = 0,1$ und $\alpha_b = 0,3$ am geringsten (Li und Cassidy 2007).

$$D = \sqrt{\frac{1}{J} * \sum_{j=1} (\hat{x}_{j,\alpha_a,\alpha_b} - x_j)^2} \rightarrow \text{MIN!} \quad \text{Formel 3-41}$$

mit

D	Distanzmaß D auf Basis der mittleren berechneten und mittleren gemessenen Fahrzeugbelegung
J	Anzahl der betrachteten Linienfahrten
j	Index der Linienfahrten
$\hat{x}_{j,\alpha_a,\alpha_b}$	Mittlere geschätzte Belegung der Linienfahrt j mit den Faktoren α_a und α_b
x_j	Mittlere gemessene Belegung der Linienfahrt j

3.4.2.2 Entropiemaximierung – Verfahren nach Willumsen

Nach Ortúzar und Willumsen (2004) wurde der mathematische Ansatz der Entropiemaximierung erstmals durch Wilson (1974) entwickelt und für die Nachfragemodellierung (Verkehrsverteilung) genutzt. Das Prinzip stammt aus der Physik und geht davon aus, dass sich ein System mit größter Wahrscheinlichkeit in dem Zustand einstellt, in dem es unter den gegebenen Zwängen die größte Ausgeglichenheit aufweist. Die Maximierung der Entropie (=Unsicherheit) bedeutet dabei, dass eine Vorhersage zum Zustand eines Einzelelements aufgrund der maximalen Freizügigkeit der Entscheidung am schwierigsten (unsichersten) ist. Diesem Prinzip zufolge tritt beim viermaligen Werfen einer Münze mit größter Wahrscheinlichkeit der ausgeglichene Zustand auf: zwei Mal Zahl und zwei Mal Kopf. Somit ist die Vorhersage eines Einzelwurfs am unsichersten (Köhler 2001).

Die Entropiemaximierung zielt darauf ab, die Verkehrsverteilung in einem Untersuchungsraum zu modellieren. Dabei ist zu berücksichtigen, dass die Verkehrsnachfrage aus den Entscheidungen und Eigenschaften aller einzelnen Individuen im Untersuchungsraum resultiert (mikroskopische Dimension). Da diese Eigenschaften in der Regel nicht bekannt sind, findet eine Aggregation ebendieser Informationen mit Hilfe einer Quelle-Ziel-Matrix statt, ohne auf die Eigenschaften der Individuen einzugehen (makroskopische Dimension). Für jede beliebige Quelle-Ziel-Matrix bestehen dabei eine Vielzahl an Kombinationen auf Mikro-Ebene, die zu dem gleichen Ergebnis auf der Makro-Ebene führen. Die Informationsbeschaffung auf der aggregierten Makro-Ebene ist durch Beobachtungen durchführbar und damit schneller und günstiger zu erheben.

Nach Willumsen (1978) ist unter allen möglichen Quelle-Ziel-Matrizen die Matrix die Wahrscheinlichste, die im Rahmen der vorgegebenen Rahmenbedingungen (etwa harte Randsummen) die größtmögliche Anzahl an Mikrozuständen abbildet (Formel 3-42).

$$E = \frac{D!}{\prod_{od} d_{od}!}$$

Formel 3-42

mit

E	Anzahl Mikrozustände einer Matrix
D	Anzahl aller Fahrgastfahrten im Untersuchungsgebiet
d_{od}	Anzahl Fahrgastfahrten zwischen der Quelle o und dem Ziel d

Lohse und Schnabel (2011) haben auf Grundlage dieser Formel das Optimierungsproblem in Formel 3-44 abgeleitet. Dieses Optimierungsproblem ergibt sich aus Formel 3-43: Der Term $D * \ln(D) - D$ ist konstant, da sich die Fahrgastfahrten und Randsummen im Untersuchungsraum nicht ändern. Lediglich die Gesamtzahl der Fahrgastfahrten im Untersuchungsraum muss bekannt sein. Um nun die Entropie zu maximieren, ist es also notwendig, die Matrixelemente so auszuwählen, dass der nicht-konstante Teil des Terms minimiert wird (Formel 3-44).

$$\ln(E) = D * \ln(D) - D - \sum_o \sum_d (d_{od} * \ln(d_{od}) - d_{od}) \quad \text{Formel 3-43}$$

$$\sum_o \sum_d (d_{od} * \ln(d_{od}) - d_{od}) \rightarrow \text{Min!} \quad \text{Formel 3-44}$$

Falls historische Informationen zu der Verkehrsverteilung bzw. Realisierungswahrscheinlichkeiten von Fahrgastfahrten der Beziehung o nach d vorliegen, kann das Optimierungsproblem entsprechend Formel 3-45 angepasst werden (Wang 2008). Je näher also die geschätzte Verkehrsverteilung an der realen Verteilung liegt, desto höher die Entropie.

$$\sum_o \sum_d \left(d_{od} * \ln\left(\frac{d_{od}}{B_{od}}\right) - 1 \right) \rightarrow \text{Min!} \quad \text{Formel 3-45}$$

mit

d_{od}	Anzahl Fahrgastfahrten zwischen der Quelle o und dem Ziel d
B_{od}	Bewertung der Verkehrsbeziehung zwischen o und d (hier auf Basis historischer Daten)

In der Realität kann grundlegend das Prinzip der maximalen Freizügigkeit für die Modellierung des Nutzerverhaltens berücksichtigt werden. Gleichzeitig muss jedoch auch die spezifische Empfindlichkeit und Nutzenorientierung der verschiedenen Verkehrsteilnehmer und Personen(-gruppen) beachtet werden. Daher stellt die Modellierung auf Basis der maximalen Entropie eine nicht ausreichende Annäherung an die Lösung einer Verkehrsverteilungsaufgabe dar (Lohse und Schnabel 2011).

3.4.2.3 Zufallsmodell

Das Zufallsmodell ist ein Spezialfall zur Lösung eines bilinearen Gleichungssystems, das auf der Entropiemaximierung basiert. Bei dem Zufallsmodell hat die Ausgangsmatrix (Bewertungen oder bereits erfasste Stichproben) keinen

Einfluss auf das Ergebnis. Das bedeutet, dass alle Verkehrsbeziehungen als gleichrangig angesehen werden, sodass eine Verkehrsstrommatrix mit maximaler Entropie bei Einhaltung der Randsummenbedingungen entsteht. Das Zufallsmodell folgt Formel 3-46 (Lohse und Schnabel 2011).

$$d_{od} = \frac{\widehat{Q}_o * \widehat{Z}_d}{\widehat{D}} \quad \text{Formel 3-46}$$

mit

d_{od}	Verkehrsstrom zwischen den Haltestellen o und d
\widehat{Q}_o	Vorgegebene Einsteigerzahlen an Haltestelle o (Quellverkehrsaufkommen, Randsumme)
\widehat{Z}_d	Vorgegebene Aussteigerzahlen an Haltestelle d (Zielverkehrsaufkommen, Randsumme)
\widehat{D}	Vorgegebene Fahrgastzahlen (Randsumme)

Mit dem Zufallsmodell können nach Lohse und Schnabel (2011) der KFZ-Verkehr und öffentliche Personenverkehr für Klein- und Mittelstädte mit einem Stadtdurchmesser von bis zu 6 km gute Abbildung der Realität erzielt werden. Hierbei ist der Begriff „gut“ nicht näher definiert.

3.4.3 Kritische Würdigung der Verfahren zur Hochrechnung

3.4.3.1 Grundlagen zu Modellfehlern

Die Güte einer Vorhersage durch Modellrechnungen wird nach Ortúzar und Willumsen aus der Komplexität des Modells, beschrieben durch die Anzahl notwendiger Variablen, und der Datenqualität der Erhebungsdaten bestimmt. In der Realität findet häufig ein Kompromiss zwischen dem notwendigen Aufwand für die Erhebung von Daten in ausreichender Qualität und den Zielen bzw. der Genauigkeit der Modellrechnung statt: Ein Modell auf mikroskopischer Ebene setzt eine Vielzahl von Informationen (Variablen) zu den Individuen voraus, deren Erhebung mit einem hohen Aufwand und einer hohen Anfälligkeit für Fehler einhergehen.

Fehler in der erfassten Datengrundlage können dazu führen, dass trotz eines korrekten Modellalgorithmus die Modellergebnisse unbrauchbar sind. Die Modellkomplexität sollte daher auf die Qualität der Daten abgestimmt werden, sodass Daten mit geringer Qualität in eher einfache, weniger komplexe, dafür aber robuste Modelle eingehen.

In Abbildung 3-20 wird der Zusammenhang zwischen Modellkomplexität und Modellfehler dargestellt. Der Modellfehler ergibt sich dabei aus dem Spezifizierungsfehler (e_s) und Messfehler (e_m) (Ortúzar und Willumsen 2004).

Spezifizierungsfehler (e_s) entstehen u. a. dadurch, dass die dem Modell zugrundeliegende Fragestellung nicht ausreichend untersucht ist. Zu Spezifizierungsfehlern gehören der Einbezug von irrelevanten Variablen, das Auslassen von relevanten Variablen, die Nutzung von falschen Funktionen (lineare Funktion zur Darstellung eines in der Realität nichtlinearen Zusammenhangs).

Demgegenüber stehen Messfehler (e_m). Je mehr Variablen in ein Modell einbezogen werden (Komplexität), desto mehr Variablen müssen repräsentativ ermittelt werden. Mit steigender Zahl an zu erhebenden Variablen steigt somit auch der Messfehler, bspw. durch nicht eindeutige Fragestellungen, falsche Interpretation der Antworten, Fehler bei der Übertragung der Antworten oder ungenaue Variablen, die die Wahrnehmung und Entscheidung des Befragten nicht vollständig wiedergeben.

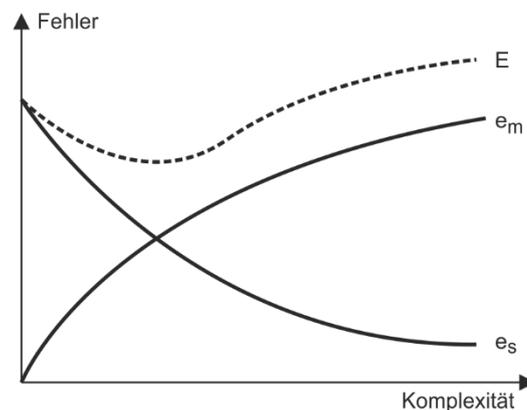


Abbildung 3-20: Abhängigkeit zwischen Fehler und Modellkomplexität (Ortúzar und Willumsen 2004, S. 72)

Entsprechend dieser Überlegungen ist in dem geplanten Anwendungsfall aufgrund der beschriebenen Ungenauigkeit der WLAN- und Bluetooth-Daten (Kapitel 2.5.3) ein Verfahren mit eher geringer Komplexität zu wählen.

3.4.3.2 Betrachtung der Verfahren

Im Folgenden kommt es zu einer kritischen Betrachtung der vorgestellten Verfahren. Dabei werden Verfahren mit und ohne Ausgangsmatrix unterschieden.

Bei beiden Verfahrenstypen – Verfahren mit und ohne Startmatrix – ist zu berücksichtigen, dass sie auf AFZS-Daten basieren. Ohne die Erfassung von AFZS-Daten stehen keine Informationen zu den Ein- und Aussteigern je Haltestelle zur Verfügung und somit keine Randsummen. Die Anwendung der vorgestellten Verfahren ist somit nicht möglich. Verkehrsunternehmen

verfügen in der Regel nicht über eine Vollausrüstung an AFZS-Geräten in allen Linienfahrzeugen. Stattdessen wird nur ein Teil der Fahrzeuge ausgestattet und alternierend auf verschiedenen Linien als Messfahrzeug eingesetzt. Die Einsatzplanung basiert auf einer Stichprobenplanung in Abhängigkeit des Ausrüstungsumfangs mit AFZS je Fahrzeugtyp (Gelenkbus, Standardlinienbus etc.) (Verband Deutscher Verkehrsunternehmen e. V. (VDV) 2018). Daher ist bei einer Anwendung bzw. potentiellen Umsetzung des entwickelten Verfahrens zu berücksichtigen, dass die Erfassungshardware nur in Fahrzeuge eingebaut wird, die bereits über ein AFZS-System verfügen.

Verfahren mit Startmatrix

Abhängigkeit von der Ausgangsmatrix: Steigerungsfaktorenmodelle besitzen eine starke Abhängigkeit von der Ausgangsmatrix (Prinzip der Informationsminimierung) und den Randsummen. Während die Randsummen mit hoher Genauigkeit vorliegen (AFZS-Daten), werden die Matrixelemente der Ausgangsmatrix auf Basis der gefilterten WLAN- und Bluetooth-Signale erstellt. In der Ausgangsmatrix vorhandene und ggf. unbekannte Verzerrungen können nicht bzw. nur bedingt ausgeglichen werden und wirken sich daher negativ auf die Modellergebnisse aus. Etwaige Unschärfen sind in Kapitel 2.5 und Kapitel 3.3.2.3 beschrieben.

Problem struktureller Nullen: Besitzen einzelne Verkehrsströme in der Ausgangsmatrix den Wert „0“, werden diese aufgrund der multiplikativen Form (Formel 3-25) auch in der Bewertungsmatrix den Wert „0“ behalten (Problem of non-structural zeros (Ben-Akiva et al. 1985). Bestehen ganze Zeilen oder Spalten aus „0“-Werten, während die Randsumme einen Wert größer als „0“ besitzt, kann das Iterationsverfahren entsprechend nicht konvergieren, so dass keine gültige Lösung gefunden werden kann – für die entsprechende Zeile oder Spalte kann nie die (Soll-)Randsumme erreicht werden. Nach Ortúzar und Willumsen (2004) ist eine geringe Konvergenz nach wenigen Iterationsschritten ein mögliches Zeichen, dass die Ausgangsmatrix an entscheidenden Stellen „0“-Werte besitzt. Mit der Fuzzifizierung werden einzelne, auf Basis des ersten und letzten Signals einer einzigartigen MAC-Adresse ermittelten Quelle-Ziel-Relationen, unter Berücksichtigung der Unschärfe auf eine Vielzahl von Quelle-Ziel-Relationen, aufgeteilt. Die relevanten Quelle-Ziel-Relationen erhalten Wahrscheinlichkeitswerte zwischen 0 und 1, die die Wahrscheinlichkeit der Relation auf Basis der Signale widerspiegelt. Damit wird ein Schritt dahingehend unternommen, den Anteil von

Nullen in der Quelle-Ziel-Matrix zu reduzieren. Selbst kleinste Werte über 0 führen dazu, dass eine Konvergenz erreicht werden kann.

Verfahren ohne Startmatrix

Auf Basis der Randsummen allein stehen in der Regel mehr als eine mögliche Quelle-Ziel-Matrix als mögliche Lösung zur Auswahl.

Im Entropiemodell werden die Wahrscheinlichkeiten für den Ein- und Ausstieg an einer beliebigen Haltestelle im Linienverlauf mit dem gleichen Gewicht bzw. der gleichen Präferenz bewertet. Das trifft bei ÖV-Fahrten in den wenigsten Fällen zu, da entweder Ein- oder Ausstieg meist an einem Knoten liegen, bei denen bestimmten Aktivitäten nachgegangen werden kann. Fahrgastfahrten zwischen Haltestellen, bei denen keine Aktivitäten in der Nähe durchgeführt werden können, sind unwahrscheinlich.

Diesen Umstand greifen Li und Cassidy (2007) in ihrem Verfahren auf und unterscheiden Major- und Minor-Haltestellen anhand der Möglichkeiten von Aktivitäten in der Umgebung der Zielhaltestelle. In Deutschland kommt es im Zuge räumlicher Konzepte wie der dezentralen Konzentration zusammen mit dem Konzept der kurzen Weg (Brake 1998) zu einer Mischung von Wohn- und Aktivitätenstandorten, sodass keine klare Abgrenzung zwischen Major- und Minor-Haltestellen besteht.

Eine Möglichkeit zur objektiven Klassifizierung von Major- und Minor-Haltestellen stellt die Einteilung der Haltestellen auf Basis der AFZS-Daten dar: Im Anwendungsfall (NVV) besteht eine breite Datengrundlage an AFZS-Daten für alle Linien. Die Daten können nach Schichten (Tagestyp, Zeitschicht) ausgewertet werden, um relevante Ausstiegshaltestellen (Major-Haltestellen), differenziert nach der gewählten Schichtung, zu bestimmen. Bei einer großen Anzahl austeigender Fahrgäste – basierend auf dem relativen Anteil an allen Fahrgästen der Linienfahrt – kommt es zur Einteilung zu der Gruppe Major-Haltestellen. Analog kommt es bei Unterschreiten eines Grenzwerts zur Einteilung in die Gruppe der Minor-Haltestellen. Die Betrachtung der Einsteigerzahlen ist für das Verfahren unerheblich.

3.4.4 Auswahl von Verfahren zur Hochrechnung

Für die Hochrechnung im Rahmen des entwickelten Verfahrens werden

- das Fratar-Verfahren sowie
- das Verfahren nach Li und Cassidy (2007)

genutzt.

Das Fratar-Verfahren, ein in der Verkehrsplanung bekanntes Verfahren, zählt zu den Steigerungsfaktorenmodellen bzw. zu den iterativen Randsummenverfahren und benötigt eine Ausgangsmatrix. Die Ausgangsmatrix stellt hier die aus den gefilterten WLAN- und Bluetooth-Daten abgeleitete Quelle-Ziel-Matrix für die betrachtete Linienfahrt dar. Entsprechend der Ausführungen basieren die vorgestellten Steigerungsfaktorenmodelle auf den in Formel 3-23 und Formel 3-24 dargestellten Beziehungen und liefern daher ähnliche Ergebnisse. Die Auswahl des Verfahrens hat somit wenig Wirkung auf die zu erwarteten Verfahrensergebnisse.

Um dem Problem struktureller Nullen zu begegnen und eine Hochrechnung zu ermöglichen, werden „0“-Werte in der Ausgangsmatrix überall dort mit geringen Hilfswerten über 0 (z. B. 0,01) ersetzt, wo ganze Zeilen oder Spalten aus Nullwerten bestehen, während die Randsumme einen Wert größer als 0 besitzt. Aufgrund der multiplikativen Form des Hochrechnungsverfahrens können so die Steigerungsfaktoren ihre Funktion erfüllen und die Matrix entsprechend den Randsummen ausgleichen, sodass es zu einer stabilen Lösung hin konvergieren kann. Durch die Hochrechnung nehmen einzelne Hilfswerte einen Wert von 1 oder mehr an, andere Hilfswerte werden reduziert ($<0,01$).

(Hilfs-)Werte der einzelnen Quelle-Ziel-Relationen können durch die Hochrechnung sehr kleine Werte annehmen. Zur Erhöhung der Lesbarkeit und Interpretierbarkeit der Bewertungsmatrix wurde daher ein zusätzliches Verfahren implementiert, das kleine Werte nach der Hochrechnung identifiziert und ggf. entfernt (Reduzierungsverfahren). Die Reduzierung umfasst dabei alle Werte der Bewertungsmatrix – sowohl Werte auf Basis von Hilfswerten als auch auf Basis der erfassten WLAN- und Bluetooth-Daten, da beide Werte mit Unsicherheiten verbunden sind (Kapitel 3.3.2.3).

Zur Identifikation eines „kleinen Werts“ wird ein Grenzwert – etwa 0,001 – eingesetzt. Werte, die kleiner als der Grenzwert sind, werden als „kleine Werte“ angesehen. Darauf aufbauend sucht das Verfahren nach Beendigung der Hochrechnung nach dem kleinsten Wert unterhalb des Grenzwerts und entfernt diesen. Anschließend wird die Hochrechnung (Fratar) erneut durchgeführt. Sofern die Hochrechnung ohne den entfernten Wert fehlschlägt, wird die Löschung rückgängig gemacht. Kommt das Verfahren zu einer Lösung, wird erneut der kleinste Wert unterhalb des Grenzwerts entfernt und die

Hochrechnung erneut durchgeführt. Das Verfahren wird solange durchgeführt, bis die Hochrechnung keine Werte unterhalb des Grenzwerts aufweist.

Um die Notwendigkeit der Ausgangsmatrix aufzuzeigen, werden zwei Verfahren genutzt:

- Fratar-Verfahren, das rein auf Hilfwerten basiert (Anwendung ohne Berücksichtigung der Ausgangsmatrix) und
- Li-Cassidy-Verfahren.

Die Anwendung dieser Verfahren benötigt keine Ausgangsmatrix, sodass die Matrixelemente allein auf Basis der Randsummen bestimmt werden. Das Verfahren nach Li-Cassidy hat gegenüber der Entropiemaximierung den Vorteil, dass die Ausstiegswahrscheinlichkeit eines Fahrgasts nicht gleichverteilt, sondern in Abhängigkeit der jeweiligen Haltestelle bestimmt wird. Die Einteilung in Major- und Minor-Haltestellen erfolgt basierend auf den Ein- und Ausstiegszahlen der einzelnen Haltestellen im Verhältnis zu dem Mittelwert über alle Haltestellen der Linienfahrt (AFZS-Daten). Major-Haltestellen zeichnen sich dadurch aus, dass sie an Standorten liegen, an denen der Fahrgast Aktivitäten wie Einkauf, Freizeit oder ähnliches durchführen kann und eine überdurchschnittlich hohe Zahl an Ein- und Ausstiegen erfolgt. Als Grenzwert für die Zuordnung in die Gruppe der Major-Haltestellen wird eine Überschreitung der mittleren Anzahl der Ein- und Aussteiger um 50% festgelegt. Somit wird gegenüber einer festen Zuordnung von Major- und Minor-Haltestellen die Dynamik von Zielgelegenheiten in Abhängigkeit von Tageszeiten oder Jahreszeiten berücksichtigt.

Folgendes Beispiel dient zur Verdeutlichung: Im Mittel werden pro Haltestelle einer fiktiven Linienfahrt zehn Ein- und Aussteiger gezählt. Somit werden nur die Haltestellen als Major-Haltestellen gekennzeichnet, an denen mindestens 15 Ein- und Aussteiger (10 Ein- und Aussteiger $\times 1,5$) gezählt wurden.

Ziel des entwickelten Ansatzes zur Ableitung von Quelle-Ziel-Matrizen auf Basis der WLAN- und Bluetooth-Datenerfassung ist es, bessere Verfahrensergebnisse zu liefern als ein Verfahren, welches die Matrixelemente ohne diese zusätzliche Datengrundlage berechnet.

4 Ergebnisse des Verfahrens in der praktischen Anwendung

4.1 Einführung

In diesem Kapitel werden die Ergebnisse der exemplarischen Anwendung des Verfahrens im Untersuchungsgebiet des Nordhessischen Verkehrsverbunds (NVV) dargestellt. Dazu werden zunächst die ausgewählten Linien und Haltestellen im NVV beschrieben, die mit der technischen Hardware zur Datenerfassung ausgestattet werden.

Neben Bussen werden ausgewählte Haltestellen, die bereits über einen Stromanschluss verfügen, mit der Erfassungsinfrastruktur ausgestattet. Haltestellen ohne vorhandenen Stromanschluss konnten aus Kostengründen nicht ausgestattet werden (Kapitel 4.2).

Ziel der Anwendung des Verfahrens ist es, die Verfahrensergebnisse zu bewerten und anhand der Ergebnisse die Qualität des Verfahrens unter Nutzung der WLAN- und Bluetooth-Daten sowie der untersuchten Parametereinstellungen zu bewerten. Darüber hinaus sollen die Parametereinstellungen ermittelt werden, die zu Ergebnissen mit der höchstmöglichen Güte führen. Abbildung 4-1 zeigt das Vorgehen zur Bewertung der Verfahrensergebnisse in Erweiterung zu Abbildung 3-1.

Die Bewertung der Verfahrensergebnisse erfolgt unter Zuhilfenahme verschiedener Gütemaße (Kapitel 4.4), die die Unterschiede der Modellschätzung (Bewertungsmatrix) im Bezug zu der realen Verteilung der Quelle-Ziel-Relationen der betrachteten Linienfahrten (Referenzmatrix) quantifizieren. Die Referenzmatrizen basieren auf Fahrgastbefragungen, die als Vollerhebung einzelner Linienfahrten der ausgestatteten Buslinien durchgeführt wurden (Kapitel 0). Aufgrund von Verweigerern sowie großer Fahrgastnachfrage auf einzelnen Linienabschnitten, können voraussichtlich nicht alle Fahrgäste erfasst werden, sodass die Ergebnisse der Fahrgastbefragung bei Bedarf anhand der AFZS-Daten hochgerechnet werden.

Die Ergebnisse der Verfahrensanwendung werden umfassend in Kapitel 4.6 zusammengefasst. Dabei werden die Gütemaße für alle berücksichtigten Parametereinstellungen (Kapitel 4.5) bestimmt. Die Auswertung der Ergebnisse der betrachteten Parameterkombinationen erfolgt sowohl differenziert für die einzelnen Gütemaße als auch in Form einer gütemaßübergreifenden Bewertung. Darüber hinaus werden die gütemaßübergreifende Bewertung für

verschiedene Klasseneinteilungen der Referenzfahrten – nach Linien und nach Fahrgastnachfrage – durchgeführt und die Erkenntnisse beschrieben.

In Kapitel 4.6.5 erfolgt eine qualitative Bewertung und Diskussion der Verfahrensergebnisse vor dem Hintergrund der in Kapitel 1 beschriebenen Einsatzbereiche. Das Kapitel schließt mit einer Untersuchung von sieben Thesen, die die Ursachen für Unschärfen betrachten (Kapitel 4.7).

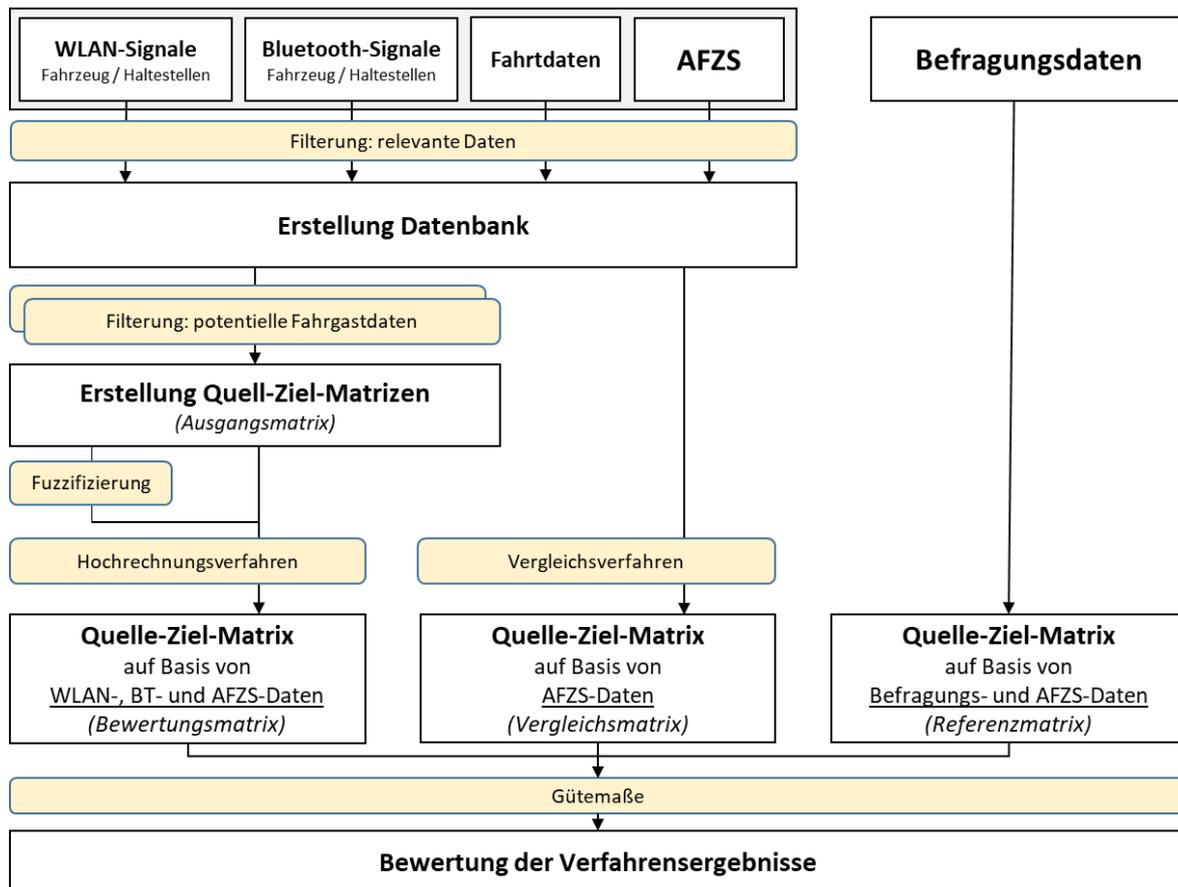


Abbildung 4-1: Überblick zum methodischen Vorgehen (Bewertung)

4.2 Auswahl Linien und Haltestellen im Untersuchungsgebiet

Das Untersuchungsgebiet liegt im Verbundgebiet des Nordhessischen Verkehrsverbunds (NVV). In Abstimmung mit den beteiligten Partnern des Projekts Mobile Data Fusion (BMVI) wurden drei Linien für die Erprobungsphase ausgewählt, auf denen alle Fahrzeuge sowie ausgewählte Haltestellen mit der für die Erfassung notwendigen Hardware ausgerüstet werden. Die Linienauswahl erfolgte vorrangig anhand der folgenden beiden Kriterien:

- Berücksichtigung städtischer und ländlicher Verkehrsräume,
- Umsteigebeziehungen zwischen den Linien.

Daher wurden für die Umsetzung die Buslinien 54, 100 und 500 ausgewählt, sodass sowohl städtische als auch ländliche Verkehrsräume berücksichtigt sind. Darüber hinaus bestehen Umsteigebeziehungen zwischen den Buslinien 54 und 500.

Linie 54: Baunatal – Niedenstein – Gudensberg

Die Linie 54 ist eine Zubringer-Linie der Linie 500 und verkehrt werktags zwischen 07:00 Uhr und 23:30 Uhr etwa im Stundentakt in der Region zwischen Baunatal, Niedenstein und Gudensberg. Zu Schulzeiten wird der Linienweg in Gudensberg bis zur Grundschule verlängert. An Samstagen, Sonn- und Feiertagen verkehrt die Linie zwischen 07:30 Uhr und 23:30 Uhr ebenfalls etwa im Stundentakt allerdings ausschließlich auf dem Abschnitt Baunatal – Niedenstein. Ergänzend zum Angebot der Busfahrten werden AST-Fahrten angeboten, die den restlichen Abschnitt zwischen Niedenstein und Gudensberg abdecken. Einige AST-Fahrten werden auch nachts angeboten (Nordhessischer Verkehrsverbund (NVV) 2022c). Auf der Linie 54 liegt die werktägliche Nachfrage bei etwa 350 Fahrgästen pro Tag (Stand 11/2020, Dataset Bieland/Mobile Data Fusion 2022).

Linie 500: Kassel – Gudensberg – Fritzlar – Edertal – Bad Wildungen

Bei der Schnellbuslinie 500 kommen als Alleinstellungsmerkmal Doppeldeckerfahrzeuge zum Einsatz. Darüber hinaus besteht in den Fahrzeugen der Buslinie 500 für die Fahrgäste bereits ein WLAN-Zugang. Die Linie 500 verkehrt werktags zwischen 05:00 Uhr und 00:00 Uhr etwa im Stundentakt, in der Hauptverkehrszeit (HVZ) im 30-Minuten-Takt. Sie verbindet Bad Wildungen, Fritzlar, Gudensberg und Edertal mit der Kasseler Innenstadt und dem Fernbahnhof Kassel-Wilhelmshöhe. An Samstagen reduziert sich das Angebot auf einen Zweistundentakt zwischen 07:00 Uhr und 23:00 Uhr. An Sonn- und Feiertagen besteht das Linienverkehrsangebot zwischen 09:00 Uhr und 23:00 Uhr im Zweistundentakt (Nordhessischer Verkehrsverbund (NVV) 2022b). Auf der Linie 500 liegt die werktägliche Nachfrage bei über 1.300 Fahrgästen (Stand 11/2020, Dataset Bieland/Mobile Data Fusion 2022).

Linie 100: Kassel Bahnhof Wilhelmshöhe – Schäferberg – Calden Flughafen

Die Linie 100 ist eine Regionalbuslinie zwischen Calden und Kassel. Sie verbindet den Flughafen Kassel-Calden, die Ortsmitte der Stadt Calden und den Ortsteil Schäferberg der Gemeinde Espenau mit der Kasseler Innenstadt, der Universität am Holländischen Platz und dem Fernbahnhof Kassel-

Wilhelmshöhe; auf dem (längsten) Abschnitt Calden bis Kassel Halitplatz verkehrt sie als Schnellbuslinie mit nur wenigen Zwischenhalten. Täglich zwischen ca. 04:50 Uhr und 23:50 Uhr wird ein Stundentakt angeboten (Nordhessischer Verkehrsverbund (NVV) 2022a). Auf der Linie 100 liegt die werktägliche Nachfrage bei etwa 650 Fahrgästen (Stand 11/2020, Dataset Bieleland/Mobile Data Fusion 2022).

Nähere Informationen zu den gewählten Buslinien können der Fahrplanauskunft des NVV entnommen werden.

Haltestellen

Bei der Ausstattung von Haltestellen ist zu beachten, dass unter einem Haltestellennamen – etwa „Katzensprung“ in Kassel – in der Regel mehrere Haltestellenstandorte zusammengefasst werden. Diese liegen in der Regel in enger räumlicher Distanz zueinander, so dass die Signale für beide Fahrrichtungen aufgezeichnet werden können.

Insgesamt wurden 29 Haltestellen ausgestattet. Es ist zu berücksichtigen, dass aufgrund von technischen Problemen – i. W. fehlendes Signal, Lücken in der Stromversorgung, unbekannte Ausfälle sowie fehlender Möglichkeiten zur Wartung – bei einem Großteil der Haltestellen zum Zeitpunkt der Fahrgastbefragung die Datenerfassung ausgefallen ist bzw. sehr große Zeitlücken von mehr als 20 Stunden pro Tag aufweist. Insgesamt waren lediglich vier Haltestellen vorhanden, von denen Daten empfangen werden (Tabelle 4-1).

In Tabelle 4-1 sind die Haltestellenbezeichnung sowie der Standort und die Fahrtrichtung der Haltestellen angegeben, die Daten durchgehend erfassen und weiterleiten. Die Angabe der Fahrtrichtung dient der Identifikation der Straßenseite der konkreten Haltestelle, in der die Erfassungsinfrastruktur verbaut wurde. Die Angabe „beide“ wurde bei Haltestellen genutzt, sofern keine Straßenseite angegeben werden kann - etwa bei Busbahnhöfen in Rondell- oder Parallel-Form).

Tabelle 4-1: Übersicht der ausgestatteten Haltestellen mit Datenerfassung (init innovation in traffic systems SE 2022)

Lfd-Nr	Haltestellenbezeichnung	Ortsteil	Richtung	Linien
1	Bahnhof Wilhelmshöhe	Kassel, Stadt	beide	500, 100
2	Bahnhof	Bad Wildungen	Kassel	500
3	Stadtmitte	Altenbauna	Gudensberg	54
4	Jugendzentrum	Altenbauna	Gudensberg	54

4.3 Erfassung von Referenzdaten

Die in der Erprobungsphase ausgestatteten Busse und Haltestellen erfassen WLAN- und Bluetooth-Signale, die im Rahmen des Verfahrens unter Zuhilfenahme der AFZS-Daten zu Quelle-Ziel-Matrizen hochgerechnet werden (Bewertungsmatrix). Um die Qualität der Hochrechnung zu bewerten (Kapitel 4.4), sind Daten zu der realen Nachfrage für einzelne Linienfahrten notwendig (Referenzmatrix).

Die Referenzdaten wurden im Rahmen einer Fahrgastbefragung durch das Fachgebiet Verkehrsplanung und Verkehrssysteme der Universität Kassel ermittelt und aufbereitet (Bieland und Briegel 2021). Die Fahrgastbefragung zielte darauf ab, für ausgewählte Linienfahrten der ausgestatteten Linien die realen Quelle-Ziel-Beziehungen aller Fahrgäste der Linienfahrt zu erfassen. Die Befragung fand zwischen September 2020 und November 2020 auf den Linien 54, 100 und 500 statt.

Bei der Auswahl der Linienfahrten für die Fahrgastbefragung wurden die Merkmale Linienweg, Tagestyp und Zeitschicht berücksichtigt. Dabei sollten stets alle Fahrgäste der Linienfahrten befragt werden (Vollerhebung).

Hinsichtlich des Tagestyps werden dabei Werktage und Wochenenden (inkl. Feiertage) differenziert. An Werktagen werden zusätzlich zwei Zeitschichten unterschieden: Hauptverkehrszeit sowie Neben- und Schwachverkehrszeit. Unter Hauptverkehrszeit (HVZ) werden die Zeiten zwischen 06:00 und 09:00 Uhr, 12:00 und 14:00 Uhr sowie 16:00 und 19:00 Uhr verstanden (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2012). Zu diesen Zeiten ist mit einem hohen Verkehrsaufkommen zu rechnen. Dabei finden hauptsächlich Fahrgastfahrten mit den Fahrtzwecken Arbeit und Bildung statt. Die Neben- und Schwachverkehrszeiten (NVZ, SVZ) umfassen die verbleibenden Zeitintervalle. Hier ist mit einer gegenüber der HVZ geringeren

Nachfrage zu rechnen. Wesentliche Fahrtzwecke sind Freizeit, Einkauf und Begleitung.

Die Befragungen wurden durch geschultes Personal als CAPI (Computer Assisted Personal Interview) unter Verwendung eines standardisierten Fragebogens mit Smartphones vorgenommen. Die Inhalte der Befragung wurden auf die Ziele der Befragung (Referenzmatrix, Vollerhebung) abgestimmt. Die Befragung bestand aus einer Erhebung der

- Einstiegshaltestelle des Fahrgasts auf der Linienfahrt,
- Ausstiegshaltestelle des Fahrgasts auf der Linienfahrt sowie
- einer Filterfrage zu weiteren Linienfahrten des Fahrgastweges (Umstiege), sofern weitere, ausgestattete Linien genutzt wurden.

Im Rahmen der Datenaufbereitung wurden die Befragungsdaten in Quelle-Ziel-Matrizen überführt. Aufgrund von Verweigerern sowie nicht befragten Fahrgästen (Kurzstreckenfahrer, hohe Nachfrage auf einzelnen Linienabschnitten) kam es zu Differenzen zwischen

- der Gesamtzahl an automatisch gezählten Fahrgästen (nach AFZS) und
- der Summe der Personenfahrten auf Basis der Befragung (Quelle-Ziel-Matrix).

Die Differenzen wurden durch das Fratar-Verfahren (Kapitel 3.4.1.2) auf Basis der AFZS-Daten der Linienfahrt ausgeglichen. Die daraus entstandenen Quelle-Ziel-Matrizen werden im Weiteren als Referenzmatrizen bezeichnet.

Während der Fahrgastbefragungen kam es zu verschiedenen Problemen, so dass die Referenzdaten hinsichtlich ihrer Verwertbarkeit in zwei Klassen unterteilt werden. Die Probleme umfassen:

- Differenzen zwischen den Fahrgastzahlen nach AFZS und auf Basis der Befragung,
- Ausfälle des AFZS (u. a. keine Datenerfassung, fehlende Zeitstempel) oder
- falsch hinterlegte Haltestellenfolgen in den AFZS aufgrund von Baustellenumleitungen.

Die Klasseneinteilung berücksichtigt dabei die Datenqualität der AFZS-Zähl-daten (APC_QUALITY), die Anzahl gezählter Fahrgäste nach AFZS sowie die Anzahl befragter Fahrgäste durch das Interviewpersonal (Tabelle 4-2). Die Qualität der AFZS-Daten (APC_QUALITY) wird softwareseitig durch das

AFZS bestimmt. Sie ergibt sich aus dem Betrag der Differenz der durch das AFZS gezählten Ein- und Aussteiger in Bezug zu der Summe der Ein- und Aussteiger (Formel 4-1). Wurden etwa fünf Einsteiger und sechs Aussteiger gezählt, ergibt sich eine APC-Quality von 90,9% (init innovation in traffic systems SE 2021).

$$APC_Quality = 100 - \frac{|Einsteiger - Aussteiger|}{Einsteiger + Aussteiger} * 100 \quad \text{Formel 4-1}$$

Tabelle 4-2: Klasseneinteilung der Referenzfahrten nach Verwertbarkeit

Klasse	Bedingung
uneingeschränkt verwertbare Referenzfahrten	(relative Abweichung der Fahrgastzahl <= 10 % oder absolute Abweichung <= 5 Fahrgästen) und APC_QUALITY >= 80
Unbrauchbare Referenzfahrten	Alle Referenzfahrten, die nicht uneingeschränkt verwertbar sind.

Befragungen wurden auf 71 Linienfahrten durchgeführt und dabei über 1.800 Fahrgäste interviewt. Von den durchgeführten Befragungen konnten lediglich 34 Linienfahrten (rund 48%) der Klasse „uneingeschränkt verwertbare Referenzfahrten“ zugeordnet werden.

Die Anzahl uneingeschränkt verwertbarer Referenzfahrten der Fahrgastbefragungen sind differenziert nach Linie, Tagestyp (Werktag, Wochenende) und Zeitschicht (Hauptverkehrszeit HVZ und Neben- bzw. Schwachverkehrszeit NVZ/SVZ) in Tabelle 4-3 dargestellt.

Tabelle 4-3: Übersicht zu den uneingeschränkt verwertbaren Fahrgastbefragungen

Linie	Tagtyp	Zeitschicht	Anzahl uneingeschränkt verwertbarer Linienfahrten
54	Werktag	HVZ	5
		NVZ/SVZ	1
	Wochenende	-	0
100	Werktag	HVZ	6
		NVZ/SVZ	5
	Wochenende	-	10
500	Werktag	HVZ	0
		NVZ/SVZ	1
	Wochenende	-	6
Summe			34

4.4 Ableitung von Gütemaßen zur Bewertung der Verfahrensqualität

4.4.1 Definition von Gütemaßen zur Bewertung der Verfahrensqualität

Das Verfahren liefert eine Schätzung der Quelle-Ziel-Relationen für jede eingesetzte Parameterkombination und jede Linienfahrt. Auf Grundlage der erfassten Referenzdaten (Kapitel 0) können Abweichungen zwischen Modellergebnissen und Erhebungsergebnissen quantifiziert werden. Vor dem Hintergrund einer möglichen Anwendung des Verfahrens (u. a. Angebotsplanung, Einnahmenaufteilung, Tarifentwicklung) ist eine hohe Übereinstimmung zwischen den Verfahrensergebnissen (Bewertungsmatrix) und der realen Verteilung der Nachfrage (Referenzmatrix) notwendig. Abweichungen werden durch geeignete Gütemaße ermittelt und zur Bewertung der Modellergebnisse genutzt.

Die Gründe für Abweichungen zwischen Modell- und Erhebungswerten sind vielfältig. Im Wesentlichen sind zu beachten:

- Abweichungen in den Referenzdaten: Bei den durchgeführten Fahrgastbefragungen wurden alle Fahrgäste einer Linienfahrt angesprochen (Vollerhebung). Durch Interview-Verweigerungen sind trotz Vollerhebung nicht alle Quelle-Ziel-Beziehungen bekannt. Für einen Teil der Verweigerer konnte der Ein- und Ausstiegsort auf Grundlage von Beobachtungen des Erhebungspersonals ergänzt werden. Zum Ausgleich weiterer etwaiger Differenzen wurden die Daten auf Grundlage der AFZS-Daten hochgerechnet (Fratat-Verfahren).
- Abweichungen durch Erstellung der Ausgangsmatrix: Sowohl die Filterung als auch die Anzahl mitgeführter Endgeräte durch die Fahrgäste kann dazu führen, dass Verzerrungen bereits in die Ausgangsmatrix eingehen (Kapitel 3.3.3.4).

Zur Bewertung der Modellergebnisse werden Gütemaße definiert, die Aussagen über die relative und/oder absolute Abweichung zwischen den Bewertungs- und Referenzmatrizen in aggregierter Form treffen. Dabei sollen die Gütemaße sowohl zur Bewertung der Modellergebnisse im Ganzen als auch in Entfernungsklassen (Kapitel 4.4.1.1) betrachtet werden. Zur Interpretation der Gütemaße wird darüber hinaus der Idealwert der spezifischen Gütemaße genannt, der bei einer perfekten Modellschätzung erreicht wird. Je größer die

Abweichung von dem Idealwert ist, desto geringer ist die Modellgüte. In Kapitel 4.4.2 werden weiterhin Grenzwerte für Gütemaße festgelegt. Mit den Grenzwerten werden akzeptable Ergebnisse von Ergebnissen, die eine zu große Abweichung zwischen Bewertungs- und Referenzmatrix aufweisen, abgegrenzt. Wird der Grenzwert überschritten, gilt die Modellschätzung mit den gewählten Parametern als nicht zweckmäßig.

Das Verfahren wird anhand von verschiedenen Eingangsparametern (Filter, Hochrechnungsverfahren) durchgeführt und deren Wirkung auf die Güte der Verfahrensergebnisse anhand der Gütemaße bestimmt. Die Ergebnisse werden genutzt, um die optimale Parametrisierung zu ermitteln.

4.4.1.1 Entfernungsklassen

Die Gütemaße werden unabhängig von einer Klasseneinteilung als Gesamtwert ausgegeben, wodurch ein erster Vergleich zwischen den gewählten Parameterkombinationen durchgeführt werden kann. Dabei besteht die Möglichkeit, dass ein oder mehrere Parametereinstellungen ähnliche bzw. gleiche Werte annehmen. Um dennoch Unterschiede in der Verfahrensqualität analysieren zu können, sollen die Gütemaße daher weiter nach Entfernungsklassen differenziert werden.

Es ist davon auszugehen, dass die im Rahmen des Verfahrens erzeugten Quelle-Ziel-Matrizen hinsichtlich der Ein- und Ausstiegshaltestelle verzerrt sind und Kurzstreckenfahrten (insbesondere Fahrgastfahrten unter vier Minuten) untererfasst werden (Kapitel 2.5.3). Eine Einordnung der Verfahrensergebnisse nach Entfernungsklassen trägt daher dazu bei, die Ergebnisse (Vergleich mit Referenzdaten) hinsichtlich möglicher Verzerrungen einzuordnen. Dazu werden die in Tabelle 4-4 genannten drei Entfernungsklassen gebildet.

Für die Bewertung der Verfahrensergebnisse in Entfernungsklassen ist neben der Bewertungs- und Referenzmatrix eine Entfernungsmatrix mit der räumlichen Distanz zwischen den Haltestellen für die betreffende Linienfahrt (oder das Gesamtgebiet) notwendig.

Tabelle 4-4: Einteilung und Beschreibung der Entfernungsklassen

Entfernungs- klasse	Reisedistanz	Beschreibung
Kurzstrecke	0 km bis 1,6 km	<p>Unter einer Kurzstrecke werden im NVV Fahrgastfahrten mit Bus oder Tram verstanden, die inklusive der Einstiegshaltestelle nicht mehr als fünf Haltestellen umfassen (Fahrtlänge von vier Haltestellen nach Einstieg). Im Eisenbahnverkehr (RegioTram) wird die Kurzstrecke über die Reisedistanz von 3 km begrenzt (Nordhessischer Verkehrsverbund (NVV) 2022d).</p> <p>Der mittlere Haltestellenabstand auf städtischen Hauptverkehrsstraßen in Ober- und Mittelzentren beträgt 400 m (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2010). Die Entfernungsklassen werden unter Berücksichtigung der Definition des NVV sowie der Annahme eines mittleren Haltestellenabstands von 400 m festgesetzt.</p>
Mittlere Strecke	>1,6 km bis 8 km	<p>Die mittlere Strecke wurde anhand der Ausdehnung der Siedlungs- und Verkehrsfläche des Stadtgebiets Kassel festgelegt. Das Gebiet ist im Westen durch den Bergpark Wilhelmshöhe begrenzt, im Süden durch Oberzwehren, im Osten durch Bettenhausen und im Norden durch Nordholland. Die Ausdehnung des Gebiets beträgt sowohl in Nord-Süd- als auch West-Ost-Richtung etwa 11 km.</p> <p>Als mittlere Strecke wird die mittlere Distanz von 5,5 km gewählt. Darüber hinaus wird ein Umwegefaktor von 1,5 berücksichtigt, sodass die Klasse für die mittlere Strecke auf 8 km (gerundet) festgesetzt wird.</p>
Langstrecke	>8 km	<p>Strecken, die länger als 8 km sind. Fahrgastfahrten mit einer Reisedistanz von mehr als 8 km haben mit hoher Wahrscheinlichkeit Quelle oder Ziel der Fahrgastfahrt im Umland Kassels.</p>

Die Entfernungsklassen wurden bei der softwaretechnischen Umsetzung der Gütemaße nur zum Teil berücksichtigt (Berechnung des GEH-Mittelwerts, Abweichung der Verkehrsleistung).

4.4.1.2 Fehlerquadratsumme und Wurzel der Fehlerquadratsumme

Die Ähnlichkeit metrisch skalierten Merkmale kann unter Verwendung des Konzepts der euklidischen Distanz gemessen werden (Schübler o.J.). Dieses Konzept wird mit der Betrachtung der Fehlerquadratsumme, auch bekannt als Residuenquadratsumme, genutzt, um die Gesamtabweichung zwischen den einzelnen Matrixelementen der geschätzten Verkehrsströme und den Verkehrsströmen der Referenzdaten zu bestimmen (Formel 4-2) (Wang 2008). Je näher die Fehlerquadratsumme (FQS) an dem Wert 0 liegt, desto geringer sind die Differenzen zwischen den Matrixelementen aus dem Verfahren sowie den Referenzdaten. Analog gilt: je höher die Fehlerquadratsumme, desto stärker sind die Abweichungen.

$$FQS = \sum_i \sum_j (d_{od,m} - d_{od,c})^2 \quad \text{Formel 4-2}$$

mit

FQS	Fehlerquadratsumme der Fahrgastfahrten
$d_{od,m}$	Berechnete Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$d_{od,c}$	Reale Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d

Insbesondere dann, wenn die Fehlerquadratsumme sehr große Werte annimmt, bietet sich die Wurzel der Fehlerquadratsumme an (Formel 4-3), um die Werte ohne Informationsverlust zu reduzieren und damit die Lesbarkeit und Interpretierbarkeit zu erhöhen.

$$WFQS = \sqrt{\sum_o \sum_d (d_{od,m} - d_{od,c})^2} \quad \text{Formel 4-3}$$

mit

$WFQS$	Wurzel der Fehlerquadratsumme der Fahrgastfahrten
$d_{od,m}$	Berechnete Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$d_{od,c}$	Reale Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d

Gewichtung der Wurzel der Fehlerquadratsumme

Abweichungen der berechneten Quelle-Ziel-Matrix von den real durchgeführten Fahrgastfahrten (Referenzmatrix) sind umso bedeutender für die Modellgüte, je größer die Abweichung, gemessen in Reisedistanz, ist. Folgendes Beispiel dient zur Verdeutlichung der Bedeutung der Gewichtung. Betrachtet wird eine reale Fahrgastfahrt von Haltestelle 1 nach Haltestelle 3 – dabei gibt die Zahl der Haltestelle die Reihenfolge der Haltestelle in Fahrtrichtung an – sowie zwei Modellergebnisse:

- Das Modellergebnis 1 zeigt eine Fahrgastfahrt von Haltestelle 1 nach Haltestelle 2
- Das Modellergebnis 2 zeigt eine Fahrgastfahrt von Haltestelle 1 nach Haltestelle 22

Eine Berechnung der Fehlerquadratsumme ohne Gewichtung ergibt hier jeweils den gleichen Wert. Dabei ist die Modellgüte im ersten Fall (Abweichung von einer Haltestelle) als wesentlich besser anzusehen als in Fall 2 (Abweichung von 19 Haltestellen). Um diese Bedeutung zu berücksichtigen, kann der berechnete Wert der Wurzel der Fehlerquadratsumme jeweils mit der Entfernung (räumliche Distanz) zwischen den Haltestellen gewichtet werden.

Für die Durchführung der Gewichtung ist neben den Matrizen auf Basis der Referenzdaten sowie der Verfahrensergebnisse eine Entfernungsmatrix für die betreffende Linie (oder das Gesamtgebiet) notwendig. Die Gewichtung erfolgt entsprechend Formel 4-4 durch Multiplikation der räumlichen Distanz zu dem bestehenden Formelteil der FQS.

$$WGFQS_{g(rD)} = \sqrt{\sum_o \sum_d rD_{od} * (d_{od,m} - d_{od,c})^2} \quad \text{Formel 4-4}$$

mit

$WGFQS_{g(rD)}$	Wurzel der Fehlerquadratsumme gewichtet nach räumlicher Distanz
rD_{od}	Räumliche Distanz von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$d_{od,m}$	Berechnete Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$d_{od,c}$	Reale Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d

Relativierung der Wurzel der Fehlerquadratsumme

Der Absolutwert der Fehlerquadratsumme besitzt für sich allein gestellt zunächst nur eine begrenzte Aussagekraft, da dieser aggregierte Wert bedeutende Merkmale der Modellrechnung nicht berücksichtigt. Bei der Relativierung wird der absolute Wert der Wurzel der Fehlerquadratsumme zu einer oder mehreren Größen ins Verhältnis gesetzt, um die Interpretierbarkeit zu vereinfachen. Für die Relativierung kommt u. a. die Summe der durchgeführten Fahrgastfahrten oder die Anzahl der möglichen Quelle-Ziel-Relationen in Betracht.

Folgendes Beispiel dient zur Verdeutlichung der Bedeutung der Relativierung. Betrachtet werden zwei Modellergebnisse mit einem Ergebnis der Wurzel der Fehlerquadratsumme von 500 (Tabelle 4-5). Die alleinige Betrachtung der WFQS würde die Interpretation nahelegen, dass die Modellrechnungen die gleiche Güte besitzen. Dabei ist jedoch zu berücksichtigen, dass im ersten Fall insgesamt 1.000 Fahrgastfahrten und im zweiten Fall insgesamt 500.000 Fahrgastfahrten modelliert wurden.

Tabelle 4-5: Beispiel zur Relativierung der WFQS

WFQS	Anzahl Fahrgastfahrten	Relative WFQS (Anzahl der Fahrgastfahrten)
500	1.000	0,5
500	500.000	0,001

Die WFQS von 500 ist vor diesem Hintergrund im ersten Fall als vergleichsweise hoch anzusehen, während die Modellgüte im zweiten Fall deutlich

besser ist. Um diesen Zusammenhang zu berücksichtigen, wird der absolute Wert der WFQS mit der Gesamtzahl der Fahrgastfahrten in Beziehung gesetzt (=relativiert, Formel 4-5). Je höher die betrachtete Anzahl der Fahrgastfahrten bei einem konstanten WFQS-Wert ist, desto geringer ist der relativierte Wert.

$$WRFQS_{r(nF)} = \frac{\sqrt{\sum_o \sum_d (d_{od,m} - d_{od,c})^2}}{\sum_o \sum_d d_{od,c}} \quad \text{Formel 4-5}$$

mit

$WRFQS_{r(nF)}$	Wurzel der relativen Fehlerquadratsumme (Relativierung über Gesamtzahl der Fahrgastfahrten)
$d_{od,m}$	Berechnete Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$d_{od,c}$	Reale Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d

Bei der softwaretechnischen Umsetzung der Gütemaße wurde lediglich die Fehlerquadratsumme umgesetzt. Eine Berücksichtigung der Gewichtung nach Entfernungsklassen oder Relativierung auf Basis der Fahrgastfahrten konnte dabei aus ökonomischen Gründen des Umsetzungspartners nicht implementiert werden.

4.4.1.3 GEH-Wert

Als Gütemaß für den Vergleich der hochgerechneten Ergebnisse (Bewertungsmatrix) mit denen der Erhebung (Referenzmatrix) wird der GEH-Wert eingesetzt, der die Übereinstimmung von Wertepaaren beschreibt. Der GEH-Wert wurde 1970 speziell für die Verkehrsplanung (Vergleich von Verkehrsstärkewerten) entwickelt und nach seinem Erfinder Geoffrey E. Havers benannt. Der GEH-Wert diente ursprünglich als Gütemaß zur Bewertung der Modellschätzungen der stündlichen Verkehrsstärke verschiedener Zählstellen. Der GEH-Wert zeichnet sich dadurch aus, dass dieser sowohl die relative als auch die absolute Abweichung zwischen Modell- und Zählergebnis berücksichtigt (Formel 4-6). Durch diese Eigenschaft besitzt er eine hohe Bedeutung bei der Bewertung von Modellergebnissen (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2015; Friedrich et al. 2019a).

Der GEH-Wert berücksichtigt gleichzeitig

- den absoluten Fehler: Differenz zwischen geschätztem und wahren Wert und

- den relativen Fehler: Quotient aus absolutem Fehler und dem wahren Wert

eines Wertepaarvergleichs. Eine separate Betrachtung ist nicht zielführend, wie das Beispiel in Tabelle 4-6 verdeutlicht. Hier werden drei Beispiele von Wertepaaren aus erdachten Modell- und Messdaten dargestellt: Während der absolute Fehler bei großen Wertepaaren (Beispiel 1) ebenfalls einen vergleichsweise großen Wert annimmt, beträgt der relative Fehler hier 5%. In Beispiel 2 beträgt der relative Fehler weiterhin 5%. Aufgrund der geänderten Größenordnung des Wertepaares besitzt der absolute Fehler hier jedoch „nur“ den Wert 5. Weiter zeigt Beispiel 3, dass bei konstantem absolutem Fehler von 5 der relative Fehler – aufgrund einer erneuten Anpassung des Wertepaares – lediglich 0,1% beträgt.

Tabelle 4-6: Beispiele zu absolutem und relativem Fehler

	Beispiel 1	Beispiel 2	Beispiel 3
Modellergebnis	95.000	95	4.995
Wahrer Wert	100.000	100	5.000
Absoluter Fehler	5.000	5	5
Relativer Fehler	5%	5%	0,1%
GEH-Wert	16	0,5	0,1

Der GEH-Wert nimmt umso kleinere Werte an, je geringer die Abweichung zwischen einem modellierten und realen Wert ist. Das bedeutet, je kleiner der GEH-Wert ist, desto besser ist die Modellgüte (The Highways Agency (UK) et al. 1996) – der Idealwert liegt also bei 0. Entsprechend der Beispiele in Tabelle 4-6 kann aus dem GEH-Wert das Beispiel mit der höchsten Güte abgelesen werden: Beispiel 3. Bei der alleinigen Betrachtung der absoluten Abweichung wären die Beispiele 2 und 3 gleichwertig, bei alleiniger Betrachtung der relativen Abweichung wären die Beispiele 1 und 2 gleichwertig.

Für die Bewertung einer Modellierung der stündlichen Verkehrsstärke von Zählstellen wurde bereits 1996 in The Highways Agency (UK) et al (1996) ein GEH-Wert von 5 oder geringer gefordert. Dieser Grenzwert wird auch in aktuellen Regelwerken genutzt (Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) 2015; Department for Transport 2020). Ein konstanter GEH-Grenzwert von 5 erlaubt beispielsweise bei Verkehrsstärken von 1.000 Fahrzeugen pro Stunde eine Abweichung der Modellergebnisse von 16%, während bei Verkehrsstärken von 10.000 Fahrzeugen pro Stunde die maximale Abweichung 5% zur Einhaltung des Grenzwerts beträgt.

Bei Festlegung eines Grenzwerts ist zu berücksichtigen, dass der GEH-Wert nicht selbstskalierend ist. Der Grenzwert ist somit immer in Abhängigkeit des Wertebereichs des Anwendungsfalls anzupassen (Pestel et al. 2016).

Bei einer Übertragung des Grenzwerts auf Modellergebnisse für die Abschätzung der Verkehrsstärke pro Tag – das heißt, die Werte der betrachteten Paare werden gegenüber den Stundenwerten (deutlich) größer – sind GEH-Werte für eine hohe Modellgüte zwischen 10 und 20 ausreichend (Schulze 2016).

Werden dagegen Fahrgäste auf Linienfahrten betrachtet, liegen im Vergleich zu stündlichen Verkehrsstärken (deutlich) geringere Werte vor, sodass auch ein geringerer Grenzwert gewählt werden sollte. Geringe Werte führen – wie Tabelle 4-6 zeigt – dazu, dass trotz geringen absoluten Fehlers der relative Fehler hohe Werte annimmt.

In Abbildung 4-2 werden GEH-Werte für Wertepaare im Bereich von 0 bis 30 dargestellt. Dabei sind geringe GEH-Werte grün eingefärbt; mit steigendem Wert wechselt die Farbe zu rot. Aus der Abbildung geht hervor, dass der GEH-Wert unter der Annahme einer konstanten absoluten Abweichung – z. B. von 1 – einen umso kleineren Wert annimmt, je größer die verglichenen Wertepaare sind.

Darüber hinaus kann der Toleranzbereich der Modellergebnisse in Abhängigkeit eines GEH-Grenzwerts abgelesen werden: Wird von einem GEH-Grenzwert von 2 ausgegangen, liegen für einen wahren Wert von 2 die Modellergebnisse zwischen 0 (absolute Abweichung von 2) als auch 6 (absolute Abweichung von 4) im Toleranzbereich für akzeptable Modellergebnisse.

Der GEH-Wert wird zur Bewertung der Ähnlichkeit zwischen Bewertungs- und Referenzmatrix auf Basis der Personenfahrten (Formel 4-6, Formel 4-7) genutzt.

		berechneter Wert																														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
tatsächlicher Wert	0	1,4	2,0	2,4	2,8	3,2	3,5	3,7	4,0	4,2	4,5	4,7	4,9	5,1	5,3	5,5	5,7	5,8	6,0	6,2	6,3	6,5	6,6	6,8	6,9	7,1	7,2	7,3	7,5	7,6	7,7	
	1	1,4	0,0	0,8	1,4	1,9	2,3	2,7	3,0	3,3	3,6	3,8	4,1	4,3	4,5	4,7	4,9	5,1	5,3	5,5	5,7	5,9	6,0	6,2	6,4	6,5	6,7	6,8	6,9	7,1	7,2	7,4
	2	2,0	0,8	0,0	0,6	1,2	1,6	2,0	2,4	2,7	3,0	3,3	3,5	3,8	4,0	4,2	4,5	4,7	4,9	5,1	5,2	5,4	5,6	5,8	5,9	6,1	6,3	6,4	6,6	6,7	6,9	7,0
	3	2,4	1,4	0,6	0,0	0,5	1,0	1,4	1,8	2,1	2,4	2,7	3,0	3,3	3,5	3,8	4,0	4,2	4,4	4,6	4,8	5,0	5,2	5,4	5,5	5,7	5,9	6,0	6,2	6,4	6,5	6,6
	4	2,8	1,9	1,2	0,5	0,0	0,5	0,9	1,3	1,6	2,0	2,3	2,6	2,8	3,1	3,3	3,6	3,8	4,0	4,2	4,4	4,6	4,8	5,0	5,2	5,3	5,5	5,7	5,8	6,0	6,2	6,3
	5	3,2	2,3	1,6	1,0	0,5	0,0	0,4	0,8	1,2	1,5	1,8	2,1	2,4	2,7	2,9	3,2	3,4	3,6	3,8	4,0	4,2	4,4	4,6	4,8	5,0	5,2	5,3	5,5	5,7	5,8	6,0
	6	3,5	2,7	2,0	1,4	0,9	0,4	0,0	0,4	0,8	1,1	1,4	1,7	2,0	2,3	2,5	2,8	3,0	3,2	3,5	3,7	3,9	4,1	4,3	4,5	4,6	4,8	5,0	5,2	5,3	5,5	5,7
	7	3,7	3,0	2,4	1,8	1,3	0,8	0,4	0,0	0,4	0,7	1,0	1,3	1,6	1,9	2,2	2,4	2,7	2,9	3,1	3,3	3,5	3,7	3,9	4,1	4,3	4,5	4,7	4,9	5,0	5,2	5,3
	8	4,0	3,3	2,7	2,1	1,6	1,2	0,8	0,4	0,0	0,3	0,7	1,0	1,3	1,5	1,8	2,1	2,3	2,5	2,8	3,0	3,2	3,4	3,6	3,8	4,0	4,2	4,4	4,5	4,7	4,9	5,0
	9	4,2	3,6	3,0	2,4	2,0	1,5	1,1	0,7	0,3	0,0	0,3	0,6	0,9	1,2	1,5	1,7	2,0	2,2	2,4	2,7	2,9	3,1	3,3	3,5	3,7	3,9	4,1	4,2	4,4	4,6	4,8
	10	4,5	3,8	3,3	2,7	2,3	1,8	1,4	1,0	0,7	0,3	0,0	0,3	0,6	0,9	1,2	1,4	1,7	1,9	2,1	2,4	2,6	2,8	3,0	3,2	3,4	3,6	3,8	4,0	4,1	4,3	4,5
	11	4,7	4,1	3,5	3,0	2,6	2,1	1,7	1,3	1,0	0,6	0,3	0,0	0,3	0,6	0,8	1,1	1,4	1,6	1,8	2,1	2,3	2,5	2,7	2,9	3,1	3,3	3,5	3,7	3,8	4,0	4,2
	12	4,9	4,3	3,8	3,3	2,8	2,4	2,0	1,6	1,3	0,9	0,6	0,3	0,0	0,3	0,6	0,8	1,1	1,3	1,5	1,8	2,0	2,2	2,4	2,6	2,8	3,0	3,2	3,4	3,6	3,8	3,9
	13	5,1	4,5	4,0	3,5	3,1	2,7	2,3	1,9	1,5	1,2	0,9	0,6	0,3	0,0	0,3	0,5	0,8	1,0	1,3	1,5	1,7	1,9	2,2	2,4	2,6	2,8	2,9	3,1	3,3	3,5	3,7
	14	5,3	4,7	4,2	3,8	3,3	2,9	2,5	2,2	1,8	1,5	1,2	0,8	0,6	0,3	0,0	0,3	0,5	0,8	1,0	1,2	1,5	1,7	1,9	2,1	2,3	2,5	2,7	2,9	3,1	3,2	3,4
	15	5,5	4,9	4,5	4,0	3,6	3,2	2,8	2,4	2,1	1,7	1,4	1,1	0,8	0,5	0,3	0,0	0,3	0,5	0,7	1,0	1,2	1,4	1,6	1,8	2,0	2,2	2,4	2,6	2,8	3,0	3,2
	16	5,7	5,1	4,7	4,2	3,8	3,4	3,0	2,7	2,3	2,0	1,7	1,4	1,1	0,8	0,5	0,3	0,0	0,2	0,5	0,7	0,9	1,2	1,4	1,6	1,8	2,0	2,2	2,4	2,6	2,7	2,9
	17	5,8	5,3	4,9	4,4	4,0	3,6	3,2	2,9	2,5	2,2	1,9	1,6	1,3	1,0	0,8	0,5	0,2	0,0	0,2	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3	2,5	2,7
	18	6,0	5,5	5,1	4,6	4,2	3,8	3,5	3,1	2,8	2,4	2,1	1,8	1,5	1,3	1,0	0,7	0,5	0,2	0,0	0,2	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3	2,4
	19	6,2	5,7	5,2	4,8	4,4	4,0	3,7	3,3	3,0	2,7	2,4	2,1	1,8	1,5	1,2	1,0	0,7	0,5	0,2	0,0	0,2	0,4	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,0	2,2
	20	6,3	5,9	5,4	5,0	4,6	4,2	3,9	3,5	3,2	2,9	2,6	2,3	2,0	1,7	1,5	1,2	0,9	0,7	0,5	0,2	0,0	0,2	0,4	0,6	0,9	1,1	1,3	1,4	1,6	1,8	2,0
	21	6,5	6,0	5,6	5,2	4,8	4,4	4,1	3,7	3,4	3,1	2,8	2,5	2,2	1,9	1,7	1,4	1,2	0,9	0,7	0,4	0,2	0,0	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6	1,8
	22	6,6	6,2	5,8	5,4	5,0	4,6	4,3	3,9	3,6	3,3	3,0	2,7	2,4	2,2	1,9	1,6	1,4	1,1	0,9	0,7	0,4	0,2	0,0	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6
	23	6,8	6,4	5,9	5,5	5,2	4,8	4,5	4,1	3,8	3,5	3,2	2,9	2,6	2,4	2,1	1,8	1,6	1,3	1,1	0,9	0,6	0,4	0,2	0,0	0,2	0,4	0,6	0,8	1,0	1,2	1,4
	24	6,9	6,5	6,1	5,7	5,3	5,0	4,6	4,3	4,0	3,7	3,4	3,1	2,8	2,6	2,3	2,0	1,8	1,5	1,3	1,1	0,9	0,6	0,4	0,2	0,0	0,2	0,4	0,6	0,8	1,0	1,2
	25	7,1	6,7	6,3	5,9	5,5	5,2	4,8	4,5	4,2	3,9	3,6	3,3	3,0	2,8	2,5	2,2	2,0	1,7	1,5	1,3	1,1	0,8	0,6	0,4	0,2	0,0	0,2	0,4	0,6	0,8	1,0
	26	7,2	6,8	6,4	6,0	5,7	5,3	5,0	4,7	4,4	4,1	3,8	3,5	3,2	2,9	2,7	2,4	2,2	1,9	1,7	1,5	1,3	1,0	0,8	0,6	0,4	0,2	0,0	0,2	0,4	0,6	0,8
	27	7,3	6,9	6,6	6,2	5,8	5,5	5,2	4,9	4,5	4,2	4,0	3,7	3,4	3,1	2,9	2,6	2,4	2,1	1,9	1,7	1,4	1,2	1,0	0,8	0,6	0,4	0,2	0,0	0,2	0,4	0,6
	28	7,5	7,1	6,7	6,4	6,0	5,7	5,3	5,0	4,7	4,4	4,1	3,8	3,6	3,3	3,1	2,8	2,6	2,3	2,1	1,9	1,6	1,4	1,2	1,0	0,8	0,6	0,4	0,2	0,0	0,2	0,4
	29	7,6	7,2	6,9	6,5	6,2	5,8	5,5	5,2	4,9	4,6	4,3	4,0	3,8	3,5	3,2	3,0	2,7	2,5	2,3	2,0	1,8	1,6	1,4	1,2	1,0	0,8	0,6	0,4	0,2	0,0	0,2
	30	7,7	7,4	7,0	6,6	6,3	6,0	5,7	5,3	5,0	4,8	4,5	4,2	3,9	3,7	3,4	3,2	2,9	2,7	2,4	2,2	2,0	1,8	1,6	1,4	1,2	1,0	0,8	0,6	0,4	0,2	0,0

Abbildung 4-2: GEH-Werte für Wertepaare im Bereich von 0 bis 30

GEH-Wert auf Basis von Personenfahrten

Die Bewertungs- und Referenzmatrix stellen die Personenfahrten zwischen allen Quelle-Ziel-Relationen der betreffenden Linienfahrt dar. Der GEH-Wert wird entsprechend Formel 4-6 für alle Matrixelemente bestimmt. Der GEH-Wert kann dabei auch für Teilmatrizen bestimmt werden, etwa durch Differenzierung auf Basis von Entfernungsklassen.

$$GEH_{P,od} = \sqrt{\frac{2 * (d_{od,m} - d_{od,c})^2}{d_{od,m} + d_{od,c}}} \quad \text{Formel 4-6}$$

mit

$GEH_{P,od}$	GEH-Wert für die Bewertung der Personenfahrten einer Linienfahrt (P) des Wertepaars o, d
$d_{od,m}$	Berechnete Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$d_{od,c}$	Reale Anzahl von Fahrgastfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d

Mittelwertbildung des GEH-Werts

Für den Vergleich mehrerer Wertepaare, etwa zum Vergleich von Quelle-Ziel-Matrizen auf Basis aller Wertepaare einer Linienfahrt kann der Mittelwert des GEH-Werts genutzt werden (Pillat 2014). Die Berechnung erfolgt entsprechend Formel 4-7.

$$\overline{GEH}_P = \frac{1}{nVB} * \sum_o \sum_d GEH_{P,od} \quad \text{Formel 4-7}$$

mit

\overline{GEH}_P	Mittelwert des GEH-Werts für die Bewertung aller Personenfahrten einer Linienfahrt (P) einer Quelle-Ziel-Matrix
$GEH_{P,od}$	GEH-Wert für die Bewertung der Personenfahrten einer Linienfahrt (P) des Wertepaars o, d
nVB	Anzahl der möglichen Verbindungen der Linienfahrt (Formel 4-8)

$$nVB = \frac{n * (n - 1)}{2} \quad \text{Formel 4-8}$$

mit

nVB	Anzahl der möglichen Verbindungen der Linienfahrt
n	Anzahl der Haltestellen der Linienfahrt

Basierend auf dem GEH-Wert wurden von Friedrich et al. (2019b) der SQV-Wert (Scalable Quality Value) entwickelt. Mit dem SQV-Wert sollen die wesentlichen Probleme des GEH-Werts (nicht selbstskalierend, Notwendigkeit zur Neubewertung der Grenzwerte bzw. zulässigen Abweichung, Interpretationsschwierigkeiten) behoben werden. Grundsätzlich benötigt auch der SQV-Wert einen Skalierungsfaktor f , um die Bewertung an der typischen Größenordnung der betrachteten Kenngröße auszurichten. Für den Anwendungsfall, in dem Matrix 1 aus einer Stichprobe (Erhebungsdaten) und Matrix 2 auf Basis einer Vollerhebung erstellt wurde, sehen Friedrich et al. den mittleren, gewichteten SQV-Wert über alle Matrixelemente vor. Beim mittleren, gewichteten SQV-Wert über alle Matrixelemente wird der SQV-Wert jedes Matrixelements berechnet (Formel 4-9). In den Vergleich der Wertepaare gehen jedoch nicht die originären Größen ein, sondern relative Häufigkeiten. Der Skalierungsfaktor ist konstant und ergibt sich auf Basis der Zellenanzahl mit $f = 1/Z^2$. Der SQV-Wert wurde nachträglich, d. h. nach Abschluss der Softwareentwicklung bekannt, sodass dieser nicht umgesetzt wurde. Der GEH-Wert ist für den Zweck der Verfahrensbewertung ebenso geeignet wie der SQV-Wert. Die Definition von Grenzwerten des GEH-Werts wird in Kapitel 4.4.2 vorgenommen.

$$\overline{SQV} = \sum_o \sum_d (SQV_{od} * d_{od,c}^{rel}) \quad \text{Formel 4-9}$$

mit

\overline{SQV}	Mittlerer, gewichteter SQV-Wert über alle Matrixelemente
SQV_{od}	SQV-Wert der relativen Häufigkeiten des Wertepaars von Einstiegshaltestelle o nach Ausstiegshaltestelle d (Formel 4-10)
$d_{od,c}^{rel}$	Relative Häufigkeit des realen Wertepaars von Einstiegshaltestelle o nach Ausstiegshaltestelle d (Formel 4-11)

$$SQV_{od} = \frac{1}{1 + \sqrt{\frac{(d_{od,m}^{rel} - d_{od,c}^{rel})^2}{f * d_{od,c}^{rel}}}} \quad \text{Formel 4-10}$$

mit

SQV_{od}	SQV-Wert der relativen Häufigkeiten des Wertepaars von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$d_{od,c}^{rel}$	Relative Häufigkeit des realen Wertepaars von Einstiegshaltestelle o nach Ausstiegshaltestelle d (Formel 4-11)
$d_{od,m}^{rel}$	Relative Häufigkeit des berechneten Wertepaars von Einstiegshaltestelle o nach Ausstiegshaltestelle d (Formel 4-11)
f	Skalierungsfaktor ($1/Z^2$)

$$d_{od,c}^{rel} = \frac{d_{od,c}}{\sum_o \sum_d d_{od,c}} ; d_{od,m}^{rel} = \frac{d_{od,m}}{\sum_o \sum_d d_{od,m}} \quad \text{Formel 4-11}$$

mit

$d_{od,c}^{rel}$	Relative Häufigkeit des realen Wertepaars von Einstiegshaltestelle o nach Ausstiegshaltestelle d (Formel 4-11)
$d_{od,m}^{rel}$	Relative Häufigkeit des berechneten Wertepaars von Einstiegshaltestelle o nach Ausstiegshaltestelle d (Formel 4-11)
$d_{od,m}$	Berechnete Anzahl von Personenfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$d_{od,c}$	Reale Anzahl von Personenfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d

4.4.1.4 Abweichung der Verkehrsleistung

Der beschriebene GEH-Wert gibt die Güte der Modellergebnisse in Form eines aggregierten Wertes wieder. Aufgrund der Aggregation gehen Informationen verloren, sodass die Richtung der Abweichung – wurde beispielsweise die Anzahl der Fahrgäste auf der betrachteten Relation über- oder unterschätzt – nicht abgelesen werden kann. Somit sind auch systematische Über- oder Unterschätzungen der Anzahl an Personenfahrten auf der Linienfahrt bzw. der Verkehrsleistung nicht ersichtlich. Zur Identifikation ebendieser

systematischen Fehler wird die Abweichung der Verkehrsleistung als weiteres Gütemaß genutzt.

Auf Ebene der Gesamtfahrt können aufgrund des Hochrechnungsverfahrens (Ausgleich auf Randsummen) keine Abweichungen zwischen der Bewertungs- und Referenzmatrix entstehen, solange die Randsummen genau eingehalten werden. Daher wird die Abweichung lediglich innerhalb der Entfernungsklassen bestimmt. Die Berechnung der Abweichung der Verkehrsleistung in Entfernungsklassen erfolgt gemäß Formel 4-12.

$$dVL(rD) = \left(\frac{VL(rD)_m}{VL(rD)_c} - 1 \right) * 100 \quad \text{differenziert für} \quad \begin{array}{l} rD_{od} \leq 2 \text{ km} \\ rD_{od} > 2 \text{ km bis } \leq 8 \text{ km} \\ rD_{od} > 8 \text{ km} \end{array} \quad \text{Formel 4-12}$$

mit

$dVL(rD)$	Abweichung der Verkehrsleistung in Abhängigkeit der räumlichen Distanzklasse (rD) in Prozent
$VL(rD)_m$	Verkehrsleistung in Abhängigkeit der räumlichen Distanzklasse (rD) für die hochgerechneten Daten (m) bzw. die realen Daten (c) (Formel 4-13, Formel 4-14)
$VL(rD)_c$	
rD_{od}	Entfernung zwischen den Haltestellen o und d in km

$$VL(rD)_m = \sum_o \sum_d VL_{od,m} = \sum_o \sum_d rD_{od} * d_{od,m} \quad \text{Formel 4-13}$$

$$VL(rD)_c = \sum_o \sum_d VL_{od,c} = \sum_o \sum_d rD_{od} * d_{od,c} \quad \text{Formel 4-14}$$

mit

$VL_{od,m}$	Hochgerechnete (m) Verkehrsleistung in Personenkilometern (Pkm) von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$VL_{od,c}$	Reale (c) Verkehrsleistung in Personenkilometern (Pkm) von Einstiegshaltestelle o nach Ausstiegshaltestelle d
rD_{od}	Entfernung zwischen den Haltestellen o und d in km
$d_{od,m}$	Hochgerechnete (m) Anzahl von Personenfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d
$d_{od,c}$	Reale (c) Anzahl von Personenfahrten von Einstiegshaltestelle o nach Ausstiegshaltestelle d

4.4.1.5 Treffermaß (Anteil der Treffer)

In Ergänzung zum GEH-Wert, der die Übereinstimmung zwischen Wertepaaren bewertet, dient dieses Gütemaß der Beschreibung des Anteils an Quelle-Ziel-Relationen in der Bewertungsmatrix, die sowohl in der Bewertungs-, als auch in der Referenzmatrix („Treffer“) belegt sind. Als belegt gilt eine Quelle-Ziel-Relation, wenn sie mit einem Wert größer 0 belegt ist. Je höher der Anteil an Treffern ist, desto höher ist die Modellgüte. Eine Auswertung dieses

Gütemaßes auf Basis der Ergebnisse der Filterungen (Ausgangsmatrizen) kann darüber hinaus Hinweise auf die Güte der Filterung geben.

Die Anzahl an Treffern kann maximal der Anzahl belegter Relationen in der Referenzmatrix entsprechen. Die Anzahl der Treffer wird durch den Vergleich zwischen Bewertungs- und Referenzmatrix bestimmt. Ein Treffer liegt dann vor, wenn dieselbe Quelle-Ziel-Relation der Bewertungs- und Referenzmatrix einen Wert größer 0 aufweist ($d_{od,c} > 0$ und $d_{od,m} > 0$). Der Anteil der Treffer gibt den Anteil an Quelle-Ziel-Relationen in der Bewertungsmatrix an, der zwischen Bewertungs- und Referenzmatrix übereinstimmt. Die Berechnung des Trefferanteils folgt Formel 4-15.

$$dTR_{BM/RM} = \frac{n\ddot{U}_{RM/BM}}{nR_{BM}} \quad \text{Formel 4-15}$$

mit

$dTR_{BM/RM}$	Anteil Treffer (übereinstimmende Werte in Bewertungsmatrix (BM))
$n\ddot{U}_{RM/BM}$	Anzahl übereinstimmend gefüllter Quelle-Ziel-Relationen ($n\ddot{U}$) zwischen Referenzmatrix (RM) und Bewertungsmatrix (BM)
nR_{BM}	Anzahl belegter Relationen (nR) in der Bewertungsmatrix (BM)

Abbildung 4-3 verdeutlicht das Vorgehen: In diesem Fall ergibt sich der Anteil der Treffer aus einer Übereinstimmung zwischen den Matrizen (eine Übereinstimmung) sowie der Gesamtzahl belegter Zellen in der Bewertungsmatrix (zwei belegte Zellen). Der Trefferanteil beträgt somit 50%.

Bewertungsmatrix				
Von/nach	0	1	2	3
0		1		
1				
2				1
3				

Referenzmatrix				
Von/nach	0	1	2	3
0			1	
1			1	
2				1
3				

Trefferanteil				
Von/nach	0	1	2	3
0		x		
1				
2				✓
3				

Abbildung 4-3: Beispiel zur Berechnung des Anteils der Treffer

4.4.1.6 Überfüllungsfaktor

Der Überfüllungsfaktor basiert auf der Auswertung der besetzten Relationen der Bewertungs- und Referenzmatrix (Anzahl Fahrgäste auf der Relation größer 0: $d_{od,c} > 0$ und $d_{od,m} > 0$). Dazu wird der Quotient zwischen besetzten Relationen der Bewertungsmatrix und der besetzten Relationen der Referenzmatrix (Formel 4-16) gebildet. Ein Überfüllungsfaktor von 1 bedeutet, dass die gleiche Anzahl an Relationen in den betrachteten Matrizen befüllt ist und somit keine Überfüllung vorliegt. Der Wert kann jedoch keine Aussage

über die Lage der besetzten Relationen sowie der Anzahl an Fahrgäste je Relation treffen.

Je höher der Überfüllungsfaktor ist, desto mehr Relationen sind in der Bewertungsmatrix besetzt. Bei einem Überfüllungsfaktor von 2 sind gegenüber der Referenzmatrix doppelt so viele Relationen in der Bewertungsmatrix besetzt. Höhere Werte des Überfüllungsfaktors werden insbesondere bei den Hochrechnungsverfahren ohne Reduktionsverfahren sowie beim Verfahren nach Li und Cassidy erwartet. Nimmt der Überfüllungsfaktor einen Wert kleiner 1 an, sind in der Bewertungsmatrix weniger Relationen als in der Referenzmatrix belegt.

$$\ddot{U}F_{BM/RM} = \frac{nR_{BM}}{nR_{RM}} \quad \text{Formel 4-16}$$

mit

$\ddot{U}F_{BM/RM}$	Überfüllungsfaktor der Bewertungsmatrix (BM) gegenüber der Referenzmatrix (RM)
nR_{BM}	Anzahl belegter Relationen (nR) in der Bewertungsmatrix (BM)
nR_{RM}	Anzahl belegter Relationen (nR) in der Referenzmatrix (RM)

Eine perfekte Übereinstimmung zwischen Referenz- und Bewertungsmatrix kann durch die kombinierte Betrachtung von Überfüllungsfaktor und Trefferanteil erreicht werden: Nimmt der Überfüllungsfaktor den Wert 1 an, während der Anteil Treffer 100% beträgt, stimmen die beiden Matrizen vollkommen überein.

4.4.2 Definition von Grenzwerten der Gütemaße

Es wird davon abgesehen Grenzwerte für die Fehlerquadratsumme und Abweichung der Verkehrsleistung zu ermitteln, da diese lediglich die absolute Abweichung bewerten. Anders verhält es sich bei dem GEH-Wert, bei dem die Ergebnisse je Wertepaare ermittelt und sowohl die absolute als auch die relative Abweichung berücksichtigt werden. Das Treffermaß sowie der Überfüllungsfaktor dienen als ergänzende, leicht interpretierbare Gütemaße.

Da der GEH-Wert nicht selbstskalierend ist, sind Grenzwerte in Abhängigkeit des jeweiligen Anwendungsfalls zu bestimmen. Dazu wurde eine Auswertung der Anzahl an Fahrgästen je Relation auf Basis der hochgerechneten Befragungsdaten durchgeführt. Aus der Auswertung geht hervor, dass der Anteil der Quelle-Ziel-Relationen mit Werten größer als fünf Fahrgästen 0,05% an allen betrachteten Relationen beträgt. Werden nur Relationen betrachtet, die von Fahrgästen genutzt wurden (Werte der Relationen >0), liegt der Anteil

der Relationen mit Werten größer als 5 bei rund 3% (Dataset Bieland/Mobile Data Fusion 2022).

Das bedeutet, dass bei den ausgewählten Referenzfahrten i. d. R. fünf oder weniger Fahrgäste gemeinsam an einer Haltestelle ein- oder aussteigen. Größere Zahlen sind i. W. an Umstiegspunkten zu erwarten und stellen Ausnahmen dar.

In Tabelle 4-7 sind die absoluten und relativen Abweichungen (gerundet) zwischen geschätzten und wahren Werten der Wertepaare für sechs GEH-Grenzwerte zwischen 0,5 und 5,0 dargestellt. Die Werte für die Betrachtungsgröße (Personenfahrten) liegen zwischen 1 und 20 und bilden den angenommenen Wertebereich für einzelne Quelle-Ziel-Relationen einer Linienfahrt ab (maximale Fahrgäste zwischen zwei Haltestellen). Dabei ist zu berücksichtigen, dass die Werte für Personenfahrten überwiegend kleiner als 5 sind.

Tabelle 4-7: Absolute und relative Abweichung in Abhängigkeit des GEH-Grenzwerts (gerundet, Fahrgastzahlen je Relation)

Wahrer Wert	GEH-Grenzwert											
	0,5	0,7	1,0	1,5	2,0	5,0	0,5	0,7	1,0	1,5	2,0	5,0
	tolerierte absolute Abweichung (Wertebereich)						tolerierte relative Abweichung (maximaler Wert)					
1	[0;1]	[-1;1]	[-1;1]	[-1;2]	[-1;3]	[-2;14]	57%	83%	>100%	>100%	>100%	>100%
3	[-1;1]	[-1;1]	[-2;2]	[-2;3]	[-3;5]	[-4;17]	31%	45%	67%	>100%	>100%	>100%
5	[-1;1]	[-1;2]	[-2;3]	[-3;4]	[-4;6]	[-7;19]	24%	34%	50%	79%	>100%	>100%
10	[-2;2]	[-2;2]	[-3;3]	[-4;5]	[-5;7]	[-11;23]	16%	23%	34%	53%	74%	>100%
15	[-2;2]	[-3;3]	[-4;4]	[-5;6]	[-7;9]	[-14;27]	13%	19%	28%	43%	59%	>100%
20	[-2;2]	[-3;3]	[-4;5]	[-6;7]	[-8;10]	[-17;29]	11%	16%	24%	36%	50%	>100%

Zur Erläuterung der Tabelle dient folgendes Ablesebeispiel (grau markiert):

- Der wahre Wert (Referenzdaten) liegt vor: es sind fünf Fahrgäste auf der betrachteten Quelle-Ziel-Relation zwischen Haltestelle A und Haltestelle B gefahren.
- Der GEH-Grenzwert beträgt 2,0.
- Daraus ergibt sich ein Toleranzbereich der absoluten Abweichung der Modellergebnisse gegenüber dem wahren Wert zwischen -4 und 6.
- Bezogen auf den Anwendungsfall bedeutet dies, dass ein Modellergebnis von 1 bis 11 Fahrgastfahrten auf dieser Relation als akzeptabel eingestuft würde. Der Toleranzbereich berechnet sich durch Addition des

wahren Werts und den Maximalwerten des Wertebereichs der absoluten Abweichung (-4 und 6).

Eine Toleranz (relative Abweichung) von 100% oder höher bedeutet, dass ein Modellergebnis von 0 bzw. dem mehr als Doppelten gegenüber dem wahren Wert als akzeptabel gilt. Der wahre Wert der Quelle-Ziel-Relationen auf den betrachteten Linienfahrten liegt der Annahme entsprechend überwiegend bei 5 oder geringer. Um in diesem Wertebereich eine mögliche 100%-Toleranz (relative Abweichung) zu vermeiden, ist ein GEH-Grenzwert von weniger als 0,90 notwendig. Dabei ist zu berücksichtigen, dass bei einem geringen wahren Wert (etwa 1) bereits bei einer geringen absoluten Abweichung (etwa 1) eine hohe relative Abweichung vorliegt.

Darüber hinaus ist zu beachten, dass insbesondere bei Linienfahrten mit wenig Fahrgästen ein Großteil der verglichenen Wertepaare den Wert 0 aufweisen. Dies ist insbesondere bei der Mittelwertbildung des GEH-Werts zu beachten.

Auf Grundlage dieser Überlegungen werden die in Tabelle 4-8 beschriebenen Grenzwerte für den GEH-Wert zur Bewertung akzeptabler Abweichungen der Bewertungsmatrix gegenüber der Referenzmatrix auf Basis der Personenfahrten gewählt.

Tabelle 4-8: Definition von Grenzwerten des GEH-Werts zur Bewertung auf Basis der Personenfahrten

Grenzwert	Anwendung	Kurzbeschreibung
0,7	Einzuhalten bei 95% der einzeln betrachteten Verbindungen.	Durch den Grenzwert von 0,7 ist bei einer wahren Anzahl von Fahrgästen zwischen zwei Haltestellen von 3 eine absolute Abweichung von 1 im Toleranzbereich; bei einem wahren Wert von 20 steigt die tolerierte Abweichung auf 3 an. Der Grenzwert dient dazu, vergleichsweise kleine Unterschiede zwischen Referenz- und Bewertungsmatrix zu identifizieren.
1,5	Einzuhalten bei allen der einzeln betrachteten Verbindungen.	Durch den Grenzwert von 1,5 ist bei einer wahren Anzahl von Fahrgästen zwischen zwei Haltestellen von 3 eine absolute Abweichung von 2 bis 3 im Toleranzbereich; bei einem wahren Wert von 20 steigt die tolerierte Abweichung auf 6 bis 7 an. Der Grenzwert dient dazu, vergleichsweise große Unterschiede zwischen Referenz- und Bewertungsmatrix zu identifizieren.

Die Grenzwerte sind sowohl für das Gesamtergebnis als auch innerhalb der Entfernungsklassen gültig. Da hier zwei Grenzwerte für den GEH-Wert

bestehen, ist bei der Auswertung darauf zu achten, dass die einzelnen Referenzfahrten beide Grenzwerte einhalten.

4.4.3 Auswahl der Gütemaße für die Bewertung der Verfahrensergebnisse

Für die Bewertung der Verfahrensergebnisse werden in Abhängigkeit folgende Gütemaße in Abhängigkeit der umgesetzten Software verwendet:

- **GEH-Wert (Anzahl Fahrgäste):** Entsprechend den Erläuterungen wird hier auf die Einhaltung der gesetzten Grenzwerte (1,5 und 0,7) hin geprüft. Ziel ist es, die Parametersets zu identifizieren, die bei den meisten – im Idealfall bei allen – betrachteten Referenzfahrten die Grenzwerte durch einzelne Parameterkombinationen einhalten.
- **Treffermaß (Anteil Treffer):** Für das Treffermaß wurden keine Grenzwerte zur Abgrenzung akzeptabler von nicht akzeptablen Modellschätzungen festgelegt. Die Berücksichtigung des Treffermaßes dient zur Ermittlung von Unterschieden zwischen Varianten, die bei einer ähnlichen Anzahl an Referenzfahrten die Grenzwerte einhalten. Der Anteil Treffer wird je Parameterkombination als Mittelwert über alle Referenzfahrten angegeben.
- **Überfüllungsfaktor:** Für den Überfüllungsfaktor wurden ebenfalls keine Grenzwerte zur Abgrenzung akzeptabler von nicht akzeptablen Modellschätzungen festgelegt. Aus der Kombination des Überfüllungsfaktors mit dem Treffermaß können Informationen zur Menge falsch belegter Relationen in der Bewertungsmatrix abgeleitet werden. Der Überfüllungsfaktor wird je Parameterkombination als Mittelwert über alle Referenzfahrten angegeben.
- **Fehlerquadratsumme (FQS):** Für die Fehlerquadratsumme wurde kein Grenzwert zur Abgrenzung akzeptabler von nicht akzeptablen Modellschätzungen festgelegt. Die Berücksichtigung der Fehlerquadratsumme erfolgt aufgrund der metrischen Skalierung des Gütemaßes zur Ermittlung von Unterschieden zwischen Varianten, die bei einer ähnlichen Anzahl an Referenzfahrten die Grenzwerte einhalten. Die Fehlerquadratsumme einer Parameterkombination zeigt in der Ergebnisanalyse den aufsummierten Wert der Fehlerquadratsummen über alle Referenzfahrten der betrachteten Parameterkombination.

- **Betrag der Abweichung der Verkehrsleistung (dVL) in den Entfernungsklassen:** Bei der Betrachtung der Abweichung der Verkehrsleistung ist die Wirkrichtung (Überschätzung, Unterschätzung) nicht relevant. Daher wird hier der Betrag der Abweichung (Absolutwert) berücksichtigt. Auch für die Abweichung wurde kein Grenzwert festgelegt. Die Abweichung der Verkehrsleistung einer Parameterkombination zeigt in der Ergebnisanalyse den aggregierten Wert der Abweichung der Verkehrsleistung auf Ebene der Gesamtfahrt über alle Referenzfahrten der betrachteten Parameterkombination.

4.4.4 Kritische Würdigung

Die Gütemaße sind das Kernelement zur Bewertung der Verfahrensgüte. Bei der Interpretation der Gütemaße sind folgende Punkte zu beachten:

Berücksichtigung von leeren Relationen: Der GEH-Wert wird für jedes Wertepaar einzeln bestimmt. Damit werden auch Wertepaare verglichen, die in der Bewertungs- und/oder Referenzmatrix den Wert 0 besitzen. Dabei ist zu berücksichtigen, dass die Anzahl an möglichen Relationen einer Linienfahrt bei den betrachteten Linien deutlich größer ist als die Anzahl der Fahrgäste und somit ein Großteil der Relationen mit 0 besetzt ist. Dies führt insbesondere bei der Mittelwertbildung des GEH-Werts für eine Linienfahrt dazu, dass dieser tendenziell zu einer guten bis sehr guten Güte führt. Zwar sind ein Großteil der Relationen, auf denen keine Fahrgäste gefahren sind, korrekt durch das Verfahren abgebildet, jedoch verzerrt die Mittelwertbildung die Interpretierbarkeit, da dadurch Informationen zu einzelnen (starken) Unterschieden zwischen Referenz- und Bewertungsmatrix versteckt werden.

Automatisches Einhalten des Grenzwerts in Abhängigkeit des Anteils an Relationen: Der GEH-Wert von 0,7 ist bei 95% der Relationen einzuhalten. Die Anzahl belegter Relationen ist unter den betrachteten Referenzfahrten häufig geringer als die Anzahl der Fahrgäste, da einzelne Fahrgäste die gleiche Ein- und Ausstiegshaltestelle nutzen. Bei einer Linienfahrt mit 20 Haltestellen ergeben sich 190 mögliche Relationen in einer Fahrtrichtung. Da dieser GEH-Grenzwert lediglich auf 95% der Relationen eingehalten werden muss, erfüllen alle Referenzfahrten, auf denen nur 9 Relationen besetzt sind, das Gütemaß automatisch. Alternativ ist auch eine absolute Anzahl an Relationen denkbar, für die der Grenzwert überschritten werden darf – ggf. in Abhängigkeit der Anzahl an Relationen in der Referenzmatrix mit einer Belegung größer 0.

Fehlende Toleranz bei großen Abweichungen: Der GEH-1,5-Grenzwert ist auf 100% der Relationen einzuhalten. Dieser dient dazu, dass große Abweichungen, die auf eine geringe Modellgüte hinweisen, identifiziert werden können.

Grenzwerte des GEH-Werts: Sollten sich die Bandbreite der Anzahl an Fahrgästen ändern – etwa durch Formen der Aggregation mehrerer Linienfahrten – sind auch die Grenzwerte neu zu bewerten.

Wertepaarvergleich: Die betrachteten Gütemaße Fehlerquadratsumme, GEH-Wert und Treffermaß führen einen strikten Vergleich von Wertepaaren bzw. Relationen zwischen Bewertungs- und Referenzmatrix durch. Eine Abweichung von einer Haltestelle wird entsprechend im gleichen Maße bewertet wie eine Abweichung von zehn Haltestellen. Das führt dazu, dass die Quelle-Ziel-Matrizen haltestellenscharf bewertet werden. Gleichzeitig können die Gütemaße in diesem Fall jedoch keine Aussagen darüber treffen, ob es sich bei einer schlechten Bewertung um Abweichungen von wenigen Haltestellen zwischen dem Ein- und Ausstieg handelt oder ob einzelne Relationen mit starker Abweichung belegt wurden. Diese Abweichungen können durch die Gewichtung der FQS anhand der Reisedistanz (Entfernungsmatrix) einbezogen werden.

Interpretierbarkeit aggregierter Werte: Die Fehlerquadratsumme sowie die Abweichung der Verkehrsleistung werden als aggregierter Wert ausgegeben. Somit stehen keine weiteren Informationen zur Verfügung, etwa darüber, ob es sich um einzelne starke Ereignisse oder eine Vielzahl kleiner Ereignisse handelt. Das Treffermaß trifft darüber hinaus keine Aussage über die Höhe der Abweichung.

Differenz der Verkehrsleistung auf Ebene der Gesamtfahrt: Das Verfahren nach Li-Cassidy bestimmt die Matrixelemente so, dass die Randsummen exakt eingehalten werden, somit kann auf Ebene der Gesamtfahrt keine Abweichung der Verkehrsleistung entstehen. Mögliche Abweichungen wären auf Rundungsfehler zurückzuführen. Für das Fratar-Verfahren gilt: Je höher die Genauigkeitsparametrisierung für das Beenden der Iterationen, desto besser werden auch hier die Randsummen getroffen und desto geringer sind die Abweichungen der Verkehrsleistung auf Ebene der Gesamtfahrt. Dementsprechend sind bei Betrachtung der Verkehrsleistung vor allem die Ergebnisse in den einzelnen Entfernungsklassen von Bedeutung.

Bezug zur Bewertungsmatrix im Treffermaß: Der Anteil der Treffer bezieht sich auf die Bewertungsmatrix. Es wird angegeben, welcher Anteil der belegten Relationen auch in der Referenzmatrix belegt ist. Unklar bleibt an der Stelle jedoch, welcher Anteil der in der Referenzmatrix belegten Relationen auch in der Bewertungsmatrix belegt ist.

4.5 Parametrisierung des Verfahrens für die exemplarische Anwendung

Entsprechend der Erkenntnisse aus Kapitel 3.2 und Kapitel 3.4 sind verschiedene Parametrisierungen zur Filterung von Fahrgastdaten sowie zur Hochrechnung – sowohl für das Fratar- als auch für das Verfahren nach Li und Cassidy – möglich. Die untersuchten Parametrisierungen sind in den folgenden Kapiteln dargestellt.

4.5.1 Berücksichtigte Parameter der Filterung

Aus der Literatur können für das Verfahren keine optimalen Einstellungen der Parameter zur Filterung abgeleitet werden. Auch aus den bisherigen Ausführungen aus den Kapiteln 2.5 und 3.2 gehen keine begründeten, festen Werte für einzelne zu berücksichtigende Filterindikatoren hervor. Entsprechend der Ausführungen in Kapitel 3.2.2 werden die folgenden Filterindikatoren verwendet:

- GPS-Luftlinienentfernung,
- Anzahl Linienfahrten mit der gleichen Linie pro Tag und
- mittlere Signalstärke (RSSI) einzelner Endgeräte auf der Linienfahrt (nur WLAN).

Die verwendeten Filterindikatoren sowie die untersuchten Parameterausprägungen sind in Tabelle 4-9 dargestellt. Dabei werden Hochrechnungen für verschiedene Parameterkombinationen entsprechend der aufgeführten Parameterausprägungen berechnet, für die Referenzdaten vorliegen. Anhand der Gütemaße werden anschließend die Parametereinstellungen ermittelt, bei denen die geringste Abweichung zwischen Verfahrensergebnissen und realen Beobachtungsdaten vorliegen.

Tabelle 4-9: Filterindikatoren und Parameterausprägungen für die Verfahrenskalibrierung

Parametergruppe	Parameter	Parameterausprägungen
Datenquelle	Genutzte Input-Daten	[WLAN], [Bluetooth], [WLAN+Bluetooth]
Filtereinstellungen	Anzahl Linienfahrten am Tag	[<5]
	GPS-Luftlinienentfernung	[0], [100], [200], [400], [600]
	RSSI-Wert (WLAN)	[-90], [-70], [-60]
Fuzzifizierung	Nutzung von Fuzzy-Sets	[Ja], [Nein]

Die im Rahmen des Verfahrens geprüften Parameterkombinationen (Filtersets) sind in Tabelle 4-10 dargestellt. Die Filtersets setzen sich aus verschiedenen Kombinationen der Filtereinstellungen zusammen – teils werden alle Filtereinstellungen verwendet (ID 01, 05, 09 und 13), teils werden nur einzelne Parameter der Filtereinstellungen genutzt (ID 50 – 62). Die Filtersets mit den ID 90 und ID 99 nehmen eine gesonderte Rolle ein: Mit ID 90 erfolgt faktisch keine Filterung, sodass alle erfassten Daten in die Erstellung der Startmatrizen eingehen¹⁵. ID 99 ist das Gegenteil: Hier werden restlos alle Daten herausgefiltert, sodass eine leere Ausgangsmatrix entsteht¹⁶. Dieses Filterset dient zur Erstellung der Vergleichsmatrizen für das Fratar-Verfahren, um einerseits den Nutzen der Input-Daten (WLAN- und Bluetooth-Daten) und andererseits die Wirkung der Filterung nachzuweisen.

¹⁵ Damit alle Daten in die Ausgangsmatrix eingehen können, werden hier die Filtereinstellungen so konfiguriert, dass die Filterung keine Wirkung hat (GPS < 1 Mio. km).

¹⁶ Damit keine Daten in die Ausgangsmatrix eingehen können, werden hier die Filtereinstellungen auf extreme (unerreichbare) Werte gesetzt (RSSI = 1.000 dBm).

Tabelle 4-10: Zuordnung zwischen Filterset und Parameterausprägungen

Daten- quelle	Filterset-ID	Anzahl Linien- fahrten (gleiche Linie) pro Tag	GPS-Luftlinien- entfernung (>=) [m]	mittlerer RSSI- Wert (>=) [dBm]	Fuzzifizierung	
[WLAN] ; [Bluetooth] ; [WLAN+Bluetooth]	01	<5	200	-90	-	
	02	-	200	-90	-	
	05	<5	200	-70	-	
	06	-	200	-70	-	
	09	<5	400	-90	-	
	10	-	400	-90	-	
	13	<5	400	-70	-	
	14	-	400	-70	-	
	20	-	100	-90	x	
	21	-	100	-70	-	
	30	-	600	-90	-	
	31	-	600	-70	-	
	32	-	600	-60	x	
	33	-	400	-60	-	
	34	-	200	-60	-	
	35	-	100	-60	-	
	40	<5	100	-90	x	
	41	<5	100	-70	-	
	42	<5	600	-90	-	
	43	<5	600	-70	-	
	44	<5	600	-60	x	
	45	<5	40	-60	-	
	46	<5	200	-60	-	
	47	<5	100	-60	-	
	50	-	100	-	x	
	51	-	200	-	x	
	52	-	400	-	x	
	53	-	600	-	x	
	60	-	-	-90	-	
	61	-	-	-70	-	
	62	-	-	-60	-	
	90	<i>keine Filterung</i>	-	Ungleich 1.000.000	-	-
	99	<i>vollständige Filterung</i>	-	-	1.000	-

Jedes der 33 betrachteten Filtersets wird dabei für jede der erhobenen Datenquellen (WLAN, Bluetooth sowie WLAN und Bluetooth) separat betrachtet.

Die Signale der jeweiligen Quelle werden je Filterset auf den einzelnen Quelle-Ziel-Relationen aufsummiert.

Die Fuzzifizierung wurde aufgrund des vergleichsweise hohen Aufwands zur Erzeugung und Kontrolle für acht ausgewählte Filtersets durchgeführt. Die Auswahl der Filtersets erfolgte dabei unter Berücksichtigung erster Ergebnisse der Verfahrensbewertungen (u. a. wurden die Filtersets mit der höchsten Güte für die Fuzzifizierung genutzt). Die ausgewählten Filtersets sind ebenfalls in Tabelle 4-10 dargestellt.

4.5.2 Berücksichtigte Parameter der Hochrechnung

Die verwendeten Parametrisierungen für die Hochrechnungsverfahren (Fratar-Verfahren und Verfahren nach Li und Cassidy) sind in Tabelle 4-11 dargestellt. Analog zu den Parameterkombinationen für die Filterung werden auch verschiedene Parameterkombinationen für die Hochrechnung betrachtet. Die Berechnung erfolgt auch hier für die betrachteten Linienfahrten, für die Referenzdaten vorliegen. Die Hilfswerte dienen dem Umgang mit dem Problem strukturelle Nullen und werden auf Matrixelementen ohne Nachfrage eingesetzt. Die Bewertung der Ergebnisse erfolgt auf Basis der Gütekriterien. Ziel ist es, die Parameterkombinationen zu ermitteln, die zu der geringsten Abweichung zwischen Bewertungs- und Referenzmatrix führen.

Tabelle 4-11: Differenzierungsmerkmale und Ausprägungen für die Kalibrierung des Verfahrens

Differenzierungsmerkmal	Beschreibung	Ausprägung
Hochrechnungsverfahren	Fratar	
	- Hilfswerte	[0,0001], [1,0]
	- Grenzwert Reduktionsverfahren ¹⁷	[0], [0,1], [0,5], [1,0]
Li-Cassidy	Li-Cassidy	
	- Einteilung Major (Bezug Mittelwert)	[50%], [100%]
	- Alpha Major	[0,1], [0,3], [0,5]
	- Alpha Minor	[0,1], [0,3], [0,5], [0,7], [0,9]
Fratar (ohne Ausgangsmatrix)	Fratar (ohne Ausgangsmatrix)	
	- Hilfswert	[1,0]
	- Grenzwert Reduktionsverfahren ¹⁷	[0], [0,1], [0,5], [1,0]

Im Rahmen des Verfahrens werden acht Parameterkombinationen für das Fratar-Verfahren (Tabelle 4-12) sowie 20 Parameterkombinationen für das Li-Cassidy-Verfahren (Tabelle 4-13) untersucht. Abbruchkriterium für das Fratar-Verfahren ist eine erreichte Genauigkeit von 0,01 (WVI Prof. Dr. Wermuth Verkehrsforschung und Infrastrukturplanung GmbH 2021) – d. h. eine

¹⁷ 0 bedeutet, dass keine Reduktion stattfindet.

Abweichung zwischen den Randsummen der Bewertungs- und Referenzmatrix von maximal 0,01 (Fahrgäste). Sollte die Genauigkeit nicht erreicht werden, bricht das Verfahren nach 13.000 Iterationsschritten ab. Das Li-Casidy-Verfahren betrachtet zwei Möglichkeiten zur Einteilung von Major- und Minor-Haltestellen sowie verschiedene Werte für Alpha Major und Alpha Minor. Alpha Major und Alpha Minor wirken sich auf den Anteil aussteigender Fahrgäste aus, die an Major- oder Minor-Haltestellen eingestiegen sind. Entsprechend Kapitel 3.4.2.1 wird die höchste Übereinstimmung mit den realen Quelle-Ziel-Matrizen bei Alpha Major = 0,1 und Alpha Minor = 0,3 erreicht.

Tabelle 4-12: Parametrisierung Fratar-Verfahren

Genauigkeit	Hilfswert	Wert Reduktionsverfahren
0,01	0,00001	0
0,01	0,00001	0,1
0,01	0,00001	0,5
0,01	0,00001	1
0,01	1	0
0,01	1	0,1
0,01	1	0,5
0,01	1	1

Tabelle 4-13: Parametrisierung Li-Casidy-Verfahren

Einteilung Major-Haltestellen	Alpha Major	Alpha Minor
Ein- und Aussteigerzahlen mehr als 50% bzw. 100% über Mittelwert	0,1	0,3
	0,1	0,5
	0,1	0,7
	0,1	0,9
	0,3	0,1
	0,3	0,3
	0,3	0,5
	0,3	0,7
	0,3	0,9
	0,5	0,5

Die Einteilung in Major- und Minor-Haltestellen erfolgt auf Basis der Summe von Ein- und Aussteigern an den einzelnen Haltestellen der Linienfahrt. Grundsätzlich sollte die Summe der Ein- und Aussteiger an jeder Major-Haltestellen so groß wie der Mittelwert der Ein- und Aussteiger über alle Haltestellen sein. Im Anwendungsfall werden Haltestellen erst dann als Major-Haltestelle klassifiziert, wenn die Summe der Ein- und Aussteiger 50% bzw. 100% über dem Mittelwert liegt. Das führt dazu, dass bei einem Mittelwert von 10 Ein- und Aussteigern, Major-Haltestellen mindestens 15 (50%) bzw. 20 (100%) Ein- und Aussteiger aufweisen müssen.

Die Darstellung der Hochrechnungssets (HR-Sets) in der Ergebnisauswertung erfolgt auf folgende Weise:

- für Fratar (FRAT): [Verfahren]_h[Hilfswert]_r[Grenzwert Reduktionsverfahren]_[----]

- für Li-Cassidy (LICA): [Verfahren]_HS[Wert zur Einteilung in Major-Haltestellen]_a[Wert Alpha Major]_b[Wert Alpha Minor].

Daraus ergibt sich für die Nutzung des Fratar-Verfahrens mit der Berücksichtigung von Hilfwerten in Höhe von 0,00001 und der Reduzierung mit einem Grenzwert von 1,0 folgende Beschreibung: „FRAT_h1,00000_r1,0_----“. Diese Beschreibung wird in der Darstellung der Ergebnisse wieder aufgegriffen. Ein Reduktionsgrenzwert von 0 (r0,0) bedeutet, dass keine Reduktion stattfindet.

4.6 Ermittlung der Verfahrensergebnisse

In diesem Kapitel werden die durch das Verfahren bewerteten Quelle-Ziel-Matrizen analysiert. Ziel der Ergebnisanalyse ist die Ermittlung der Parametereinstellungen, die die beobachtete Realität (entsprechend den Referenzdaten, Kapitel 0) am besten treffen.

Zu diesem Zeitpunkt liegen die Bewertungsmatrizen, die Referenzmatrizen und Vergleichsmatrizen vor. Darüber hinaus wurden zur Bewertung der Ähnlichkeit der Matrizen die beschriebenen Gütemaße berechnet. Das weitere Vorgehen gliedert sich in drei Schritte.

1. **Prüfung der Eingangsdaten auf Vollständigkeit:** Vollständige Eingangsdaten (WLAN- und Bluetooth-Daten, AFZS-Daten) sind Voraussetzung für eine korrekte Hochrechnung sowie die Berechnung der Gütemaße. Daher werden in diesem Schritt die Eingangsdaten auf Vollständigkeit geprüft, um Fehler bei der Interpretation der Ergebnisse zu vermeiden.
2. **Vorbereitung der Datenanalyse:** Auf Basis der Software (HR-Tool) werden die ausgewählten Parameterkombinationen genutzt, um für ausgewählte Linienfahrten Bewertungsmatrizen zu erzeugen. Diese Parameterkombinationen werden im ersten Schritt erzeugt und in das HR-Tool implementiert. Die berechneten Bewertungsmatrizen werden anschließend auf Basis der umgesetzten Gütemaße bewertet. Die Ergebnisse werden in einer Datenbank abgelegt, die aus mehreren untereinander verknüpften Tabellen besteht. Zur Vorbereitung der Datenanalyse werden daher im zweiten Schritt die vorliegenden Daten so strukturiert, dass in einer Tabelle die wesentlichen, zur Interpretation notwendigen Informationen abgelegt werden.

- 3. Darstellung der Ergebnisse aus der Erprobungsphase:** Anhand der Gütemaße wird die Übereinstimmung zwischen den Verfahrensergebnissen mit den realen Beobachtungsdaten gemessen. In diesem Schritt wird zunächst das Vorgehen zur Bewertung der Parameterkombinationen beschrieben. Anschließend werden die Ergebnisse der einzelnen Gütemaße übersichtlich aufbereitet und vorgestellt. Darüber hinaus werden zunächst die Parameterkombinationen zur Filterung und Hochrechnung identifiziert, die insgesamt die besten Ergebnisse erzielen. In einem weiteren Schritt wird geprüft, ob es Überschneidungen oder Abweichungen zwischen den optimalen Parametereinstellungen bei differenzierter Betrachtung der Linienfahrten nach Merkmalen (Nachfrage, Anzahl Haltestellen) gibt. Zur Bewertung werden die Gütemaße sowie die festgelegten Grenzwerte herangezogen.

4.6.1 Prüfung der Eingangsdaten auf Vollständigkeit

Das technische System zur Datenerfassung (Kapitel 2.4) erfasst und speichert eine Vielzahl von Bluetooth- und WLAN-Signalen, die durch Fahrgäste und Nicht-Fahrgäste ausgesendet werden. In Kapitel 2.5.3 wurden bereits Unsicherheiten bei der Vollständigkeit und Genauigkeit der erfassten Daten erläutert. Die erfassten Daten bilden die Grundlage für alle weiteren Verfahrensschritte. Fehlende oder fehlerhafte Daten haben somit einen unmittelbaren Einfluss auf die Verfahrensergebnisse. Weitere Verfahrensschritte wie

- die Filterung zur Identifikation von Daten, die mit hoher Wahrscheinlichkeit von Fahrgästen stammen, oder
- die Fuzzifizierung zum Umgang mit Unsicherheiten bei der Genauigkeit der erfassten Ein- und Ausstiegshaltestelle

erlauben zwar eine Aufbereitung des Datenbestands, bieten jedoch nicht die Möglichkeit – bzw. nur in einem begrenzten Maße – Fehler in den Grunddaten auszugleichen.

Die Hochrechnung der Daten erfolgt anhand des Fratar-Verfahrens. Steigerungsfaktorenmodellen folgen dem Prinzip der Informationsminimierung (Kapitel 3.4.1), mit dem sichergestellt werden soll, dass das Berechnungsergebnis (Bewertungsmatrix) die Informationen der Gewichtungsmatrix (Ausgangsmatrix) erhält. Die durch Steigerungsfaktorenmodelle ermittelten Lösungen sind damit stark von den Informationen der Ausgangsmatrix abhängig (Verteilung der Fahrgäste auf die Quelle-Ziel-Relationen der betrachteten

Linienfahrt auf Basis der gefilterten Inputdaten). Fehlende oder unvollständige Ausgangsdaten haben unmittelbare Auswirkungen auf die Ergebnisse.

Daher sollen an dieser Stelle die Eingangsdaten auf Vollständigkeit geprüft werden. Dazu gehören die durchgehende Erfassung von WLAN- und Bluetooth-Signalen sowie der AFZS-Daten.

Die Analyse der erfassten Rohdaten ist auf Linienfahrten begrenzt, für die Referenzmatrizen der Klasse „uneingeschränkt verwertbare Referenzfahrten“ (34 Linienfahrten) vorliegen. Die Analyse basiert auf einer Prüfung der Ausgangsmatrix für das Filterset #90 (keine Filterung). Somit wird hier das Maximum an belegten Zellen (Quelle-Ziel-Relationen) angezeigt. In Abbildung 4-4 ist die vollständige Datenerfassung einer Linienfahrt der Linie 100 dargestellt. Auffällig ist die starke Befüllung von Quelle-Ziel-Relationen, die entsprechend der Auswertung nur eine Reisedistanz von einer Haltestelle besitzen. Dies ist auf die Logik der Zuordnung von Ein- und Ausstiegshaltestelle zurückzuführen: Endgeräte, die nur einmalig erfasst wurden, wird die zuletzt vor dem erfassten Signal angefahrne Haltestelle als Einstiegshaltestelle zugeordnet, sowie die als nächstes angefahrne Haltestelle als Ausstiegshaltestelle. Der Großteil dieser Fälle kann durch die weitere Filterung bspw. mit dem Filter zur *GPS-Luftlinienentfernung* durch einen Grenzwert größer 0 m, identifiziert werden.

Fahrten mit einer Distanz von mehr als einer Haltestelle treten deutlich seltener auf, aber auch hier ist die Wahrscheinlichkeit von *false positives* hoch. Als Beispiel können hier Signale von Endgeräten wartender Fahrgäste gezählt werden, die bereits bei der Haltestellenanfahrt sowie kurz nach der Abfahrt an der Haltestelle erfasst werden. Dies gilt analog für Endgeräte in der näheren Umgebung von Haltestellen (z. B. in Wohnungen).

In Abbildung 4-5 wird dagegen eine durch die Prüfung als „unvollständige Datenerfassung“ klassifizierte Linienfahrt gezeigt. Die Sichtprüfung ist notwendig, da die Erfassungseinheit selbst nicht hinsichtlich von Stromausfällen oder fehlender Erfassung überwacht wird. Gründe für den Ausfall sind nicht bekannt.

Haltestellenindex von / nach	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Σ
0		91				1							1	1	2			1	97
1			12	2							1								15
2				20														3	23
3					27	2													29
4						37	1						1						39
5							19												19
6								27							1				28
7									22	2				1					25
8										48	5				1	1	1		56
9											85	1	2	1			1		90
10												43	6					1	50
11													77	5	4	2			88
12														100	1	2	1	2	106
13															165	6	2		173
14																93	4	4	101
15																	52	4	56
16																		37	37
17																			0
Σ	0	91	12	22	27	40	20	27	22	50	91	44	87	108	174	104	61	52	1032

Abbildung 4-4: Beispiel Quelle-Ziel-Matrix mit vollständiger Erfassung (Fahrt der Linie 100, 22.09.2020, 15:33 Uhr - WLAN- und Bluetooth-Daten, keine Filterung) (Dataset Bieland/Mobile Data Fusion 2022)

Haltestellenindex von / nach	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Σ
0																			0
1			12																12
2																			0
3																			0
4																			0
5							24	8				1							33
6								74	6	3	1	2							86
7									29			1							30
8										25									25
9											28	6							34
10												12							12
11																			0
12																			0
13																			0
14																			0
15																			0
16																			0
17																			0
Σ	0	0	12	0	0	0	24	82	35	28	29	22	0	0	0	0	0	0	232

Abbildung 4-5: Beispiel Quelle-Ziel-Matrix mit unvollständiger Erfassung (Fahrt der Linie 100, 11.10.2020, 15:35 Uhr - WLAN- und Bluetooth-Daten, keine Filterung) (Dataset Bieland/Mobile Data Fusion 2022)

Die fehlerhafte Datenerfassung zeigt sich darin, dass

- eine Erfassung zwischen Haltestelle 1 und 2 stattfand,
- keine Daten zwischen den Haltestellen 2 und 5 erfasst wurden,
- die Erfassung zwischen den Haltestellen 5 und 11 wieder einsetzte und
- danach keine weiteren Daten erfasst wurden.

Dieses unstete Bild mit völlig leeren Zeilen und Spalten zeigt sich nur bei wenigen Datenerfassungen, da die Fahrzeuge durchgehend Daten erfassen – zum Teil von Fahrgästen, zum Teil von Störquellen, sodass auch bei Leerfahrten i. d. R. die Diagonale mit einer Reisedistanz von einer Haltestelle besetzt ist (Erfahrungswissen).

Die Haltestellennummer leitet sich aus der Reihenfolge der angefahrenen Haltestellen entsprechend des Fahrplans ab. Zur Vermeidung von Fehlern wurden zusätzlich die Fahrtdaten anhand der Journey ID und Vehicle ID mit den Ergebnissen der AFZS-Datenerfassung sowie der Fahrgastbefragung verglichen, um u. a. Linienfahrten mit einem verkürztem Linienweg auszuschließen. Insgesamt konnten fehlerhafte Linienfahrten i. W. durch vermehrt auftretende leere Zellen entlang der Diagonalen mit einer Reisedistanz von einer Haltestelle identifiziert werden.

Insgesamt sind 5 der 34 Referenzfahrten von Datenausfällen betroffen (Tabelle 4-14). Da diese Referenzfahrten eine unvollständige Datengrundlage liefern, gehen sie nicht in die weitere Auswertung ein. Die Auswertungen beziehen sich somit auf die 29 verbleibenden Referenzfahrten.

Tabelle 4-14: Referenzfahrten mit Datenausfällen

Linie	Datum der Fahrt	Abfahrtszeit	Abfahrtshaltestelle
54	13.10.2021	17:28:00	Altenbauna Stadtmitte
54	13.10.2021	18:19:00	Gudensberg Rathaus
100	27.09.2020	15:35:00	Calden Flughafen Terminal
500	10.10.2020	15:33:00	Bad Wildungen Breiter Hagen
500	30.10.2020	09:06:00	Kassel Bahnhof Wilhelmshöhe

4.6.2 Vorbereitung der Datenanalyse

Zunächst werden die beschriebenen Parametrisierungen in die Software übertragen. Dazu werden die Klasseneinteilung der Referenzfahrten (uneingeschränkt verwertbar, unbrauchbar) eingetragen, Parametersets für die Filterung und Hochrechnung definiert sowie eindeutige Bezeichnungen

vergeben, um die verschiedenen Varianten systematisch voneinander unterscheiden zu können. Mit den vorgenommenen Einstellungen werden die Hochrechnungen im HR-Tool durchgeführt.

Die Ergebnisse der Hochrechnung und Bewertung anhand der Gütemaße werden durch das Hochrechnungs-Tool (HR-Tool) in der MySQL-Datenbank abgelegt. Anschließend werden die erzeugten Tabellen der MySQL-Datenbank in Excel eingelesen und über Microsoft PowerQuery eine neue Tabelle erzeugt, die eine klare Zuordnung der Ergebnisse zu einzelnen Linienfahrten (Tag der Fahrt, Uhrzeit der Abfahrt, Starthaltestelle der Fahrt, Liniennummer) ermöglicht. Darüber hinaus werden Informationen zu

- den gewählten Filtereinstellungen zur Erzeugung der Startmatrizen,
- dem gewählten Hochrechnungsverfahren und dessen Parametereinstellungen,
- der gewählten Datenquelle (WLAN, Bluetooth, WLAN+Bluetooth),
- der Erzeugung der Startmatrizen (Nutzung der Fuzzifizierung) sowie
- der Linie, dem Datum der betrachteten Fahrt, der Abfahrtszeit der betrachteten Linienfahrt und der ersten Haltestelle im Linienverlauf

ergänzt. Da die Berechnung der Gütemaße „Treffermaß“ und „Überfüllungsfaktor“ nicht in der Software vorgesehen sind, werden diese nachträglich durch Nutzung des Programms Microsoft PowerBI berechnet (Dataset Bieleland/Mobile Data Fusion 2022).

4.6.3 Vorgehen zur Bewertung der Parameterkombinationen

An dieser Stelle wird das Vorgehen zur Bewertung der Parameterkombinationen (Filterung, Hochrechnung) ermittelt, die anhand der ermittelten Gütemaße die höchste Übereinstimmung zwischen den Verfahrensergebnissen (Bewertungsmatrix) und den realen Beobachtungsdaten (Referenzmatrix) für die betrachteten Referenzfahrten aufzeigen.

Das Vorgehen zur Auswahl der geeignetsten Parameterkombinationen unter allen betrachteten Kombinationen zeigt Abbildung 4-6.

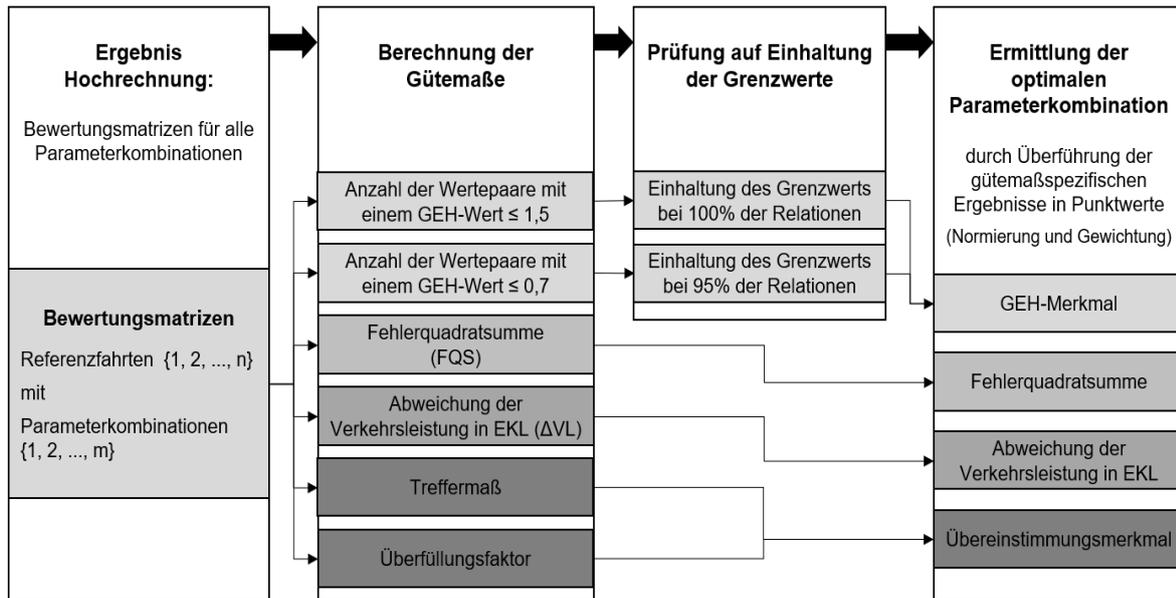


Abbildung 4-6: Vorgehen zur Ermittlung der optimalen Parameterkombination für das entwickelte Verfahren

Dazu werden die Ergebnisse der Hochrechnung aufbereitet und differenziert nach Parameterkombination, Referenzfahrt und Gütemaß gegliedert. Ziel ist es, die Vielzahl an Ergebnissen zu ordnen und die Parameterkombinationen zu identifizieren, die im Rahmen

- der spezifischen Gütemaße (Kapitel 4.6.4.1) und
- einer gütemaßübergreifenden Bewertung (Kapitel 4.6.4.2)

die besten Ergebnisse erzielen. Bei der gütemaßübergreifenden Betrachtung bezeichnen die „besten Ergebnissen“ die Parameterkombinationen,

- die bei einem möglichst hohen Anteil der Referenzfahrten die Grenzwerte der GEH-Werte einhalten sowie
- die anhand der weiteren Gütemaße Treffermaß, Überfüllungsfaktor, Fehlerquadratsumme und Betrag der Abweichung der Verkehrsleistung die geringsten Abweichungen zur Referenzmatrizen aufweisen.

Dazu werden die gütemaßspezifischen Bewertungen in vier Merkmalen zusammengeführt und anhand ihrer originären Bewertungsergebnisse normiert. Darüber hinaus findet eine Gewichtung der Merkmale nach der Bedeutung für die Verfahrensbewertung statt. Im Ergebnis werden Punktwerte je Parameterkombination in Abhängigkeit der erreichten Werte der Gütemaße über alle Referenzfahrten ermittelt. Das Vorgehen wird tiefergehend zusammen mit den Ergebnissen in Kapitel 4.6.4.2 erläutert.

4.6.4 Darstellung der Ergebnisse aus der Erprobungsphase

Für die Ermittlung der Ergebnisse wurde das HR-Tool Version 1.10.1 genutzt. Die Erstellung der Startmatrizen erfolgte mit StauFi-Tool Version vom 13.06.2022. Die Ergebnisse können der Datenbank entnommen werden (Dataset Bieland/Mobile Data Fusion 2022), die Auswertungen werden in den folgenden Kapiteln beschrieben.

4.6.4.1 Übersicht der Verfahrensergebnisse

Das Verfahren wurde mit

- 123 verschiedenen Parametereinstellung für die Filterung¹⁸ und
- 28 verschiedenen Parametereinstellung für die Hochrechnung

für die vorliegenden 29 Referenzfahrten durchgeführt. Da die Hochrechnung mit Li-Cassidy rein auf den Randsummen basiert, wurde diese nur einmalig durchgeführt (unterschiedliche Filtereinstellungen haben keine Wirkung auf das Ergebnis). Somit können insgesamt über 30.000 Bewertungsmatrizen für die Analyse genutzt werden¹⁹.

Aufgrund der Komplexität des Verfahrens sowie der Ergebnisanalysen wird in diesem Kapitel zunächst eine Übersicht der Ergebnisse dargestellt. In den folgenden Kapiteln werden die Ergebnisse detailliert beschrieben. Dazu werden die berücksichtigten Parameterkombinationen bewertet sowie die relativ beste Parameterkombination abgeleitet.

In Tabelle 4-15 werden statistische Lagemaße (Maximalwert, Minimalwert, Mittelwert und Median) für die Anzahl an Referenzfahrten gezeigt, die die GEH-Grenzwerte einhalten. Darüber hinaus wird die Anzahl an Parameterkombinationen, bestehend aus Filter- und HR-Set, aufgeführt, die beispielsweise den Maximalwert erreichen.

¹⁸ 33 Filtersets mit jeweils drei Datenquellen (99 Parameterkombinationen) sowie ergänzend 8 Filtersets für die Fuzzifizierung mit jeweils drei Datenquellen (24 Parameterkombinationen).

¹⁹ Insgesamt sind es 32.683 Matrizen. Diese setzen sich zusammen aus

- 1.107 Parameterkombinationen (123 Filtersets x 8 Kombinationen für das Fratar-Verfahren + 123 Filtersets x 1 Kombination (keine Hochrechnung)) für 29 Referenzfahrten (32.103 Matrizen) sowie
- 20 Parameterkombinationen für Li-Cassidy (da Li-Cassidy allein auf den realen Randsummen basiert, gehen hier keine weiteren Parameterkombinationen ein) für 29 Referenzfahrten (580 Matrizen).

Tabelle 4-15: Übersicht der Anzahl von Grenzwert-Einhaltungen bei allen untersuchten Parameterkombinationen

Einhaltung GEH-1,5-Grenzwert (100%)	Maximalwert	Minimalwert	Mittelwert	Median
Anzahl Referenzfahrten der betrachteten Parameterkombination mit Einhaltung des Grenzwerts	13	5	8,95	9
Anzahl Parameterkombinationen mit Einhaltung des Maximal- / Minimalwerts (Filter & Hochrechnung)	4	61	-	-
Einhaltung GEH-0,7-Grenzwert (95%)	Maximalwert	Minimalwert	Mittelwert	Median
Anzahl Referenzfahrten der betrachteten Parameterkombination mit Einhaltung des Grenzwerts	13	6	8,25	8
Anzahl Parameterkombinationen mit Einhaltung des Maximal- / Minimalwerts (Filter & Hochrechnung)	4	24	-	-

An dieser Stelle lässt sich bereits erkennen, dass die erforderliche Genauigkeit der einzelnen Quelle-Ziel-Matrizen bei einem Großteil der betrachteten Referenzfahrten durch keine der berücksichtigten Parameterkombination erreicht werden kann. Einzelne Parameterkombinationen halten die Grenzwerte lediglich bei maximal 45% der Referenzfahrten (13 von 29) ein. Diese Maximalwerte werden von je vier Parameterkombinationen für den GEH-1,5- und GEH-0,7-Grenzwert erreicht. Bei einer Anwendung des Verfahrens in der Praxis besteht daher eine hohe Wahrscheinlichkeit, dass die Zuordnung der Fahrgäste auf die Quelle-Ziel-Relationen in den ermittelten Quelle-Ziel-Matrizen (Bewertungsmatrizen) Fehler aufweist, die oberhalb der gesetzten Grenzwerte liegen.

Im Folgenden werden die besten Parameterkombinationen differenziert für jedes Gütemaß dargestellt. Wichtig für die Interpretation der Ergebnisse der besten Parameterkombinationen ist jeweils der Vergleich mit den Vergleichsverfahren. Die Vergleichsverfahren beschreiben Hochrechnungen, denen keine zusätzlichen Informationen zu Quelle-Ziel-Relationen auf Basis der WLAN- und Bluetooth-Daten zugrunde liegen.

Die Optima der Parameterkombinationen in den jeweiligen Gütekriterien werden im Hinblick auf die Ergebnisse der Vergleichsverfahren beschrieben. Eine umfassende Bewertung unter Berücksichtigung aller Gütemaße findet in Kapitel 4.6.4.2 statt.

Betrachtung zwischen gütemaßspezifischen Optima und Vergleichsverfahren

Als Vergleichsverfahren dienen die Hochrechnung mit dem Verfahren nach Li und Cassidy sowie das Fratar-Verfahren mit dem Filterset #99 (FullFilter). Durch das Filterset #99 gehen keine WLAN- und/oder Bluetooth-Daten in die Ausgangsmatrix ein, sodass alle Relationen mit dem gleichen Hilfswert besetzt werden. Somit besitzen alle Relationen das gleiche Gewicht (Einser-Matrix).

Die Hochrechnungsvarianten des Fratar-Verfahrens unterscheiden sich damit im Hilfswert und dem Reduktionsverfahren. Da jedoch auf allen Relationen der gleiche Hilfswert eingesetzt wird, hat auch die Höhe des Hilfswerts keinen Einfluss auf die Ergebnisse.

Mit der Berücksichtigung der WLAN- und Bluetooth-Daten entstehen dagegen Ausgangsmatrizen, in denen einzelne Relationen verschieden gewichtet werden. Diese Gewichtungen können bessere oder schlechtere Ergebnisse erzeugen als die Einser-Matrix, in der keine Informationen zur Belegung der Relationen vorliegen.

In Tabelle 4-16 sind die Ergebnisse der Vergleichsverfahren auf Basis des Fratar-Verfahrens vollständig, sowie exemplarisch die besten Ergebnisse der Verfahren nach Li und Cassidy angezeigt, die die besten Ergebnisse auf Basis der Einhaltung der GEH-Grenzwerte sowie den höchsten Trefferanteil erzielen. Die Anzahl an Referenzfahrten, die sowohl das GEH-1,5- als auch das GEH-0,7-Kriterium einhalten, beträgt bei allen Hochrechnungen nach Li und Cassidy vier Referenzfahrten.

Tabelle 4-16: Ergebnisse der Vergleichsverfahren (Li-Cassidy, Fratar mit Filterset 99 (FullFilter)) unter Hervorhebung (graue Markierung) der besten gütemaßspezifischen Ergebnisse

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung [m]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
	P99_FullFilter	FRAT_h0,00001_r0,0 ----	-	9	8	4	478	2642	38%	2,9
	P99_FullFilter	FRAT_h0,00001_r0,1 ----	-	9	8	4	489	2020	42%	2,5
	P99_FullFilter	FRAT_h0,00001_r0,5 ----	-	11	8	5	667	2560	53%	1,4
	P99_FullFilter	FRAT_h0,00001_r1,0 ----	-	6	9	4	1005	1783	51%	1,1
		LICA_HS100%-a0,3_b0,3	-	11	7	4	551	0	38%	2,8
		LICA_HS050%-a0,3_b0,9	-	8	7	4	635	0	41%	2,6
		LICA_HS050%-a0,1_b0,3	-	10	8	4	599	0	38%	2,8
		LICA_HS050%-a0,5_b0,5	-	10	7	4	461	0	38%	2,9

Ein Vergleich der Vergleichsverfahren zeigt, dass

- beide Vergleichsverfahren den GEH-1,5-Grenzwert bei maximal 11 von 29 Referenzfahrten einhalten,
- der GEH-0,7-Grenzwert bei maximal 9 (Fratar) bzw. 8 (Li-Cassidy) von 29 Referenzfahrten eingehalten wird,
- die Fehlerquadratsumme beim Fratar-Verfahren erwartungsgemäß mit dem Reduktionsgrenzwert steigt (siehe dazu Betrachtung Gütemaß: Fehlerquadratsumme), während die Werte durch Nutzung des Verfahrens nach Li-Cassidy weniger streuen,
- die Abweichung der Verkehrsleistung auf Ebene der Gesamtfahrt bei Li-Cassidy den Wert 0 annimmt, während das Fratar-Verfahren Differenzen in Höhe von mindestens 1.787 m auf Ebene der Gesamtfahrt zeigt,
- die Anzahl übereinstimmender Referenzmatrizen mit Einhaltung des GEH-1,5- und GEH-0,7-Grenzwerts liegt maximal bei vier (Li-Cassidy) bzw. fünf (Fratar) Referenzfahrten,
- der Anteil Treffer beträgt bei Nutzung des Fratar-Verfahrens maximal 53% bei einem Überfüllungsfaktor von 1,4, während Li-Cassidy aufgrund

der hohen Überfüllung von mindestens 2,6 lediglich einen Trefferanteil von 41% erreicht.

Darüber hinaus wurde eine Bewertung der Parameterkombinationen ohne Hochrechnung durchgeführt. Die Ergebnisse der Gütemaße zeigen, dass die Hochrechnung zu einer deutlichen Steigerung der Güte beiträgt. Ohne Hochrechnung kommt es zu folgenden mittleren Bewertungen:

- die Differenz der Verkehrsleistung liegt im Mittel bei über 5.000 km (dies ist auf zum Teil leere Startmatrizen aufgrund restriktiver Filter zurückzuführen),
- die Fehlerquadratsumme liegt im Mittel bei über 22.000,
- der Trefferanteil liegt im Mittel bei 16%,
- der Überfüllungsfaktor liegt im Mittel bei 2,5 (mit einer sehr hohen Streuung, da sowohl Startmatrizen aus restriktiven (geringer Überfüllungsfaktor) als auch aus wenig restriktiven Filtersets (sehr hoher Überfüllungsfaktor) eingehen).

Betrachtung Gütemaß: GEH-Wert (Grenzwert 1,5 und Grenzwert 0,7)

Die Ergebnisse in Tabelle 4-15 zeigen, dass der GEH-Grenzwert von 1,5, einzuhalten bei 100% der Quelle-Ziel-Relationen (Anzahl Fahrgäste) je Referenzfahrt, unter allen betrachteten Parameterkombinationen maximal bei 13 der 29 betrachteten Referenzfahrten eingehalten wird. Der Maximalwert von 13 Referenzfahrten wird durch die in Tabelle 4-17 gezeigten Parameterkombinationen eingehalten. Dabei sind auch Werte der weiteren Gütemaße angegeben. Eine vergleichende Betrachtung der Gütemaße erfolgt jedoch in Kapitel 4.6.4.2 bei der Ermittlung der optimalen Parametereinstellungen.

Tabelle 4-17: Parameterkombinationen mit maximaler Einhaltung des Gütemaßes GEH 1,5

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung [m]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN_BT	P50_GPS100	FRAT_h1,00000_r0,0 ----	-	13	8	4	474	2696	38%	2,9
WLAN_BT	P50_GPS100	FRAT_h1,00000_r0,1 ----	-	13	8	4	486	2087	42%	2,4
WLAN	P51_GPS200	FRAT_h1,00000_r0,0 ----	x	13	9	8	570	2109	38%	2,8
WLAN	P51_GPS200	FRAT_h1,00000_r0,1 ----	x	13	9	8	588	1417	46%	1,8

Der Grenzwert wird mit 13 Referenzfahrten gegenüber den Vergleichsverfahren bei zwei zusätzlichen Referenzfahrten eingehalten. Weitere 38 Parameterkombinationen halten den Grenzwert auf 12 Referenzfahrten ein.

Der Minimalwert liegt entsprechend Tabelle 4-15 bei fünf Referenzfahrten. Das heißt, dass unabhängig von der gewählten Parameterkombination der GEH-1,5-Grenzwert bei fünf Referenzfahrten eingehalten wird. Eine Betrachtung der Referenzmatrizen dieser Linienfahrten zeigt, dass die Verteilung der Randsummen bei drei Referenzfahrten nur eine mögliche Lösung der Hochrechnung zulässt. Für eine weitere Referenzfahrt besteht mehr als ein mögliches Hochrechnungsergebnis, jedoch ist auch hier die Anzahl der möglichen Matrizen aufgrund der Verteilung der Randsummen vergleichsweise gering (wenig mehrfach besetzte Zeilen und Spalten). Gleichzeitig bestehen fünf Referenzfahrten, auf denen keine der betrachteten Parameterkombinationen den GEH-1,5-Grenzwert einhält.

Die Betrachtung des GEH-Grenzwert von 0,7, einzuhalten bei 95% der Quelle-Ziel-Relationen (Anzahl Fahrgäste) je Referenzfahrt, zeigt ebenfalls, dass unter allen betrachteten Parameter-Kombinationen der Grenzwert bei maximal 13 der 29 betrachteten Referenzfahrten von vier Parameterkombination eingehalten wird (Tabelle 4-18). Der Minimalwert liegt hier bei sechs Referenzfahrten. Dabei ist zu berücksichtigen, dass der Grenzwert lediglich bei 95% aller betrachteten Relationen, inklusive der Relationen, auf denen keine Fahrgäste gefahren sind, einzuhalten ist. Referenzfahrten, auf denen 5% oder weniger der Relationen besetzt sind, erfüllen das Gütemaß automatisch wie in Kapitel 4.4.4 beschrieben. Bei einer Linienfahrt mit insgesamt 200

möglichen Relationen würde somit jede Bewertungsmatrix, in der nicht mehr als 10 Relationen belegt sind, das Gütekriterium einhalten – unabhängig von der Lage sowie dem Wert der belegten Relationen in der Bewertungsmatrix. Unter allen Referenzfahrten sind hier fünf Referenzfahrten, bei denen 100% der Parameterkombination den Grenzwert automatisch aufgrund der geringen Anzahl belegter Relationen einhalten.

Tabelle 4-18: Parameterkombinationen mit maximaler Einhaltung des Gütemaßes GEH 0,7

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung [m]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN_BT	P50_GPS100	FRAT_h0,00001_r0,5_----	x	10	13	8	689	1458	57%	1,1
WLAN_BT	P50_GPS100	FRAT_h1,00000_r0,5_----	x	10	13	8	701	1290	56%	1,1
WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5_----	x	10	13	8	737	2031	58%	1,1
WLAN_BT	P51_GPS200	FRAT_h1,00000_r1,0_----	x	9	13	7	855	1436	58%	1,0

Mit der Einhaltung des Grenzwerts bei 13 Referenzfahrten wird der Grenzwert gegenüber den Vergleichsverfahren bei vier zusätzlichen Referenzfahrten eingehalten. Weitere 20 Parameterkombinationen halten den Grenzwert auf 12 Referenzfahrten ein.

Bei dem Vergleich der Minimalwerte zwischen GEH-1,5- und GEH-0,7-Grenzwert ist das automatische Einhalten des GEH-0,7-Grenzwerts zu berücksichtigen, wodurch der GEH-0,7-Grenzwert zum Teil bei mehr Referenzfahrten eingehalten werden kann als der – in Bezug auf die betrachteten Relationen – vergleichsweise restriktivere Grenzwert von 1,5.

Tabelle 4-19 visualisiert die bisherigen Ergebnisse zu den GEH-Grenzwerten in Matrixform. Es wird die Anzahl an Parameterkombinationen gezeigt, die die GEH-Grenzwerte von 0,7 und 1,5 einhalten, differenziert nach der Anzahl an Referenzfahrten, auf denen die Grenzwerte eingehalten werden.

Tabelle 4-19: Gegenüberstellung der Ergebnisse je Parameterkombination zur Einhaltung der GEH-Grenzwerte (1,5 und 0,7)

GEH 0,7 \ GEH 1,5		Anzahl Referenzfahrten mit Einhaltung des GEH-0,7-Grenzwerts														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	>13
Anzahl Referenzfahrten mit Einhaltung des GEH-1,5-Grenzwerts	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	58	3	0	0	0	0
	6	0	0	0	0	0	0	0	2	1	12	31	1	0	0	0
	7	0	0	0	0	0	0	0	1	13	26	6	10	4	0	0
	8	0	0	0	0	0	0	0	13	38	57	12	20	8	6	0
	9	0	0	0	0	0	0	0	15	30	71	37	19	8	6	1
	10	0	0	0	0	0	0	0	12	58	99	25	24	18	5	3
	11	0	0	0	0	0	0	0	4	7	132	6	7	5	1	0
	12	0	0	0	0	0	0	0	2	3	27	3	1	1	1	0
	13	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0
	>13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Ablesebeispiel (grau markiert): Es gibt eine Parameterkombination, die den GEH-0,7-Grenzwert bei 12 Referenzfahrten einhält und den GEH-1,5-Grenzwert bei 11 Referenzfahrten.

Aus der Tabelle geht hervor, dass insgesamt acht Parameterkombinationen die beiden GEH-Grenzwerte bei mindestens 11 (der insgesamt 29) Referenzfahrten einhalten.

Aus der Tabelle geht jedoch nicht hervor, ob es sich bei den Referenzfahrten, bei denen die Grenzwerte durch die jeweils betrachtete Parameterkombination eingehalten werden, um die gleichen oder unterschiedliche Referenzfahrten handelt. Dies wurde durch eine separate Auswertung ergänzt (Übereinstimmung GEH1,5 und GEH0,7). Das Maximum an Referenzfahrten einer Parameterkombination, die die Grenzwerte für die beiden betrachteten GEH-Werte einhält, liegt bei acht Referenzfahrten.

Die maximale Übereinstimmung in den Vergleichsverfahren wird durch das Fratar-Verfahren mit einer Reduzierung von 0,5 erreicht – hier beträgt der Wert 5. Die Hochrechnung mit Li-Cassidy erzielt maximal 4 Übereinstimmungen der Einhaltung von GEH-1,5- und GEH-0,7-Grenzwert. Durch die

Nutzung der WLAN- und Bluetooth-Daten sowie der Filterung wird hier eine Steigerung von drei Referenzfahrten erreicht (Maximalwert 8).

Betrachtung Gütemaß: Fehlerquadratsumme

Der dargestellte Wert der Fehlerquadratsumme ergibt sich als Summe der Fehlerquadratsummen über alle Referenzfahrten der betrachteten Parameterkombination. Die besten Ergebnisse der Fehlerquadratsumme werden bei

- dem Verfahren nach Li-Cassidy sowie
- dem Fratar-Verfahren ohne Reduktionsverfahren bzw. Reduktionsverfahren mit einem geringen Grenzwert von bis zu 0,1

erreicht. Dies ist auf die Ermittlung der Fehlerquadratsumme zurückzuführen: Durch das Verfahren nach Li-Cassidy bzw. das Fratar-Verfahren unter Nutzung von Hilfswerten jedoch ohne Reduktion enthält die Bewertungsmatrix eine Vielzahl an Relationen, die mit sehr geringen Werten (Anzahl Fahrgäste) nahe 0 belegt sind.

Somit entstehen im Vergleich der Bewertungs- mit der Referenzmatrix eine Vielzahl an sehr geringen Abweichungen, die durch die quadratische Formel der Fehlerquadratsumme nochmals verringert werden ($0,1 \times 0,1 = 0,01$). Das führt dazu, dass Relationen, in denen die Abweichung weniger als 1 beträgt, die Fehlerquadratsumme einen geringeren Wert annimmt als die Abweichung selbst. Der Effekt ist umso größer, je näher die Abweichung an 0 liegt. Im Gegensatz dazu führt eine hoher Reduktionsgrenzwert zwar zu einer geringen Anzahl Abweichungen, die jedoch größere Werte annehmen.

Da das Verfahren nach Li-Cassidy allein auf AFZS-Daten basiert, ist hier das Filterset irrelevant. Auf eine tabellarische Darstellung wird aufgrund der großen Anzahl an Parameterkombinationen mit sehr geringer FQS verzichtet.

Betrachtung Gütemaß: Abweichung der Verkehrsleistung

Die Verkehrsleistung auf Ebene der Gesamtfahrt wird durch das Verfahren nach Li-Cassidy exakt eingehalten. Dies ist auf die präzise Einhaltung der Randsummen durch das Verfahren sowie die Verteilung der Fahrgäste auf die Quelle-Ziel-Relationen und die Verteilung der Entfernungen zwischen den Haltestellen zurückzuführen (Kapitel 4.4.4).

Werden nur die Hochrechnungen unter Nutzung des Fratar-Verfahrens betrachtet (Tabelle 4-20), wird deutlich, dass es durch Einbezug der Fuzzifizierung zu besseren Ergebnissen in der Abweichung der Verkehrsleistung auf Ebene der Gesamtfahrt kommt.

Neben der Abweichung auf Ebene der Gesamtfahrt wurden die Abweichungen der Verkehrsleistung in den einzelnen Entfernungsklassen betrachtet. Für die Auswertung wurde die Summe des Betrags der Abweichung in den einzelnen Entfernungsklassen gebildet. Der Mittelwert über alle Referenzfahrten dieser Summe liegt bei 320 km – dies entspricht einer mittleren Abweichung von rund 11 km je betrachteter Referenzfahrt. Wird die Abweichung auf die Anzahl der Fahrgäste auf den einzelnen Linienfahrten bezogen, beträgt die Abweichung rund 500 m je Fahrgast (realer Fahrgäste, AFZS-Daten). Die Abweichung je Fahrgast streut dabei stark zwischen dem Minimalwert von 0 m (wie bei der Beschreibung der Ergebnisse des Gütemaß GEH-Werts aufgeführt, besitzen einige Matrizen aufgrund der Verteilung der Randsummen nur eine mögliche Lösung) und einem Maximalwert von über 4 km.

Tabelle 4-20: Parameterkombinationen mit maximaler Einhaltung des Gütemaßes Abweichung der Verkehrsleistung (ohne Li-Cassidy)

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung [km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN	P50_GPS100	FRAT_h1,00000_r1,0_----	x	9	10	5	921	0,93	51%	1
WLAN	P53_GPS600	FRAT_h1,00000_r1,0_----	x	9	10	7	889	1,04	56%	1,0

Die Verfahren nach Li und Cassidy liegen im Mittel bei 270 km mit einer geringen Streuung von 26 km. Der geringste Wert beträgt 241 km. Das Fratar-Vergleichsverfahren erreicht einen Mittelwert von 320 km mit einer Streuung von 49 km. Der geringste Wert beträgt hier 233 km. Unter allen Parameterkombinationen, die die WLAN- und Bluetooth-Daten nutzen, erreichen zwei Kombinationen einen Wert von unter 240 km (Tabelle 4-21).

Dabei wird deutlich, dass die besten Ergebnisse auch hier durch die Verfahren mit einem geringen Reduktionsgrenzwert erzielt werden.

Tabelle 4-21: Parameterkombinationen mit maximaler Einhaltung des Gütemaßes Abweichung der Verkehrsleistung (Abweichungen der Verkehrsleistung in den Entfernungsklassen)

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung in den Entfernungsklassen in [100 km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN	P31_GPS600_ RSSI70	FRAT_h0,00001_ r0,0 ----	-	9	7	5	618	2,3	38%	2,9
WLAN	P44_L5_GPS6 00_RSSI70	FRAT_h0,00001_ r0,0 ----	-	9	7	5	618	2,3	38%	2,9

Betrachtung Gütemaß: Treffermaß (Anteil Treffer)

Der Anteil Treffer beschreibt den Anteil der belegten Relationen in der Bewertungsmatrix, der mit den Relationen in der Referenzmatrix übereinstimmt. Abweichungen der Anzahl an Fahrgästen werden hier nicht betrachtet. Der Anteil Treffer wird als Mittelwert über alle Referenzfahrten der betrachteten Parameterkombinationen ausgegeben. Je höher der Anteil Treffer ist, desto besser schneidet eine Parameterkombination bei diesem Gütemaß ab.

Die Ergebnisse der Parameterkombinationen mit den besten Ergebnissen für das Gütemaß Treffermaß sind in Tabelle 4-22 dargestellt. Der maximale mittlere Anteil an Treffern liegt bei 59%. Dabei beträgt die Streuung (Standardabweichung) unter den in der Tabelle betrachteten Parameterkombination zwischen 15% und 21%.

Tabelle 4-22: Parameterkombinationen mit maximaler Einhaltung des Gütemaßes Treffermaß (Anteil Treffer)

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung [km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN	P50_GPS100	FRAT_h0,00001_r1,0_----	-	10	11	6	860	1,93	59%	1,0
WLAN_BT	P20_L5_GPS100_RSSI90	FRAT_h0,00001_r1,0_----	-	8	11	5	867	1,36	59%	1,0
WLAN_BT	P50_GPS100	FRAT_h0,00001_r1,0_----	-	10	11	6	824	1,26	59%	1,0
WLAN_BT	P51_GPS200	FRAT_h0,00001_r1,0_----	x	8	10	7	733	1,19	59%	1,1
WLAN_BT	P40_L5_GPS100_RSSI90	FRAT_h0,00001_r1,0_----	-	8	11	7	845	1,10	59%	0,9

Erwartungsgemäß führen Hochrechnungsverfahren mit einem hohen Grenzwert zur Reduktion (1,0) zu besseren Ergebnissen beim Anteil der Treffer. Grund dafür ist, dass durch den hohen Reduktionsgrenzwert die Anzahl an belegten Relationen in der Bewertungsmatrix insgesamt abnimmt. Bei konstanter Anzahl von korrekten Relationen in der Referenzmatrix ist der Anteil an Treffern (mindestens ein Treffer notwendig) bei Bewertungsmatrizen umso höher, je weniger Relationen insgesamt belegt sind.

Mit einem mittleren Anteil von Treffern von 59% beträgt die Differenz zum besten Vergleichsverfahren (Fratar-Verfahren mit einem Reduktionsgrenzwert von 0,5) 6%-Punkte. Die Verfahren nach Li und Cassidy nehmen Werte zwischen 37% und 41% an.

Betrachtung Gütemaß: Überfüllungsfaktor

Die separate Betrachtung des Gütemaßes Überfüllungsfaktor besitzt nur eine geringe Aussagekraft. Der ideale Wert von 1 (Mittelwert über alle Referenzfahrten) wird durch 146 Parameterkombinationen eingehalten. Die Streuung (Standardabweichung) des Überfüllungsfaktors über die betrachteten Referenzfahrten liegt für diese 146 Parameterkombinationen zwischen 0,08 und maximal 0,16. Auch hier kommt es erwartungsgemäß bei Verfahren mit einer hohen Reduktion zu besseren Ergebnissen. Erwartungsgemäß sind bei dem Idealwert von 1,0 lediglich die HR-Sets „FRAT_h0,00001_r1,0_----“ und „FRAT_h1,00000_r1,0_----“ vertreten.

Das Filterset dagegen besitzt nur geringen Einfluss: 114 der 123 Filtersets erreichen (mit dem entsprechenden Hochrechnungsverfahren) einen Überfüllungsfaktor von 1,0. Auf eine tabellarische Darstellung wird aufgrund der großen Anzahl an Parameterkombinationen verzichtet.

Der geringste Überfüllungsfaktor der Vergleichsverfahren wird durch das Fratar-Verfahren mit einem Reduktionsgrenzwert von 1,0 erreicht (Überfüllungsfaktor 1,1). Für alle weiteren Vergleichsverfahren werden höhere Werte erreicht – Li-Cassidy-Verfahren sind mit einem Überfüllungsfaktor von mindestens 2,6 deutlich überfüllt.

Zwischenfazit

Die gütemaßspezifischen Auswertungen zeigen, dass keine Parameterkombination, bestehend aus Datenquelle, Filterset und HR-Set, mehrfach bei den besten Ergebnissen genannt wurde. Auffällig ist auch, dass sich unter allen Parameterkombinationen keine Kombination mit der alleinigen Datenquelle Bluetooth befindet. Die Bluetooth-Daten sind daher auf Basis dieser Ergebnisse lediglich als ergänzend zu betrachten. Regelmäßig tritt dagegen die Datenquelle WLAN und Bluetooth auf sowie die Filtersets GPS 100 und GPS 200. Bei der Betrachtung der Hochrechnungsverfahren zeigt sich, dass insgesamt kleine Hilfswerte (0,00001) bei Verfahren ohne Fuzzifizierung gegenüber Hilfswerten mit dem Wert 1 bessere Ergebnissen erzielen. Die Verwendung kleiner Hilfswerte führt dazu, dass durch WLAN- und Bluetooth-Daten besetzte Relationen ein (deutlich) höheres Gewicht besitzen. Bei Nutzung der Fuzzifizierung wird dieser Effekt abgeschwächt. Insgesamt nutzen etwa die Hälfte der aufgeführten Parameterkombinationen die Fuzzifizierung.

Die Vergleichsverfahren führen bei den allen betrachteten Gütemaßen, mit Ausnahme des Gütemaßes Abweichung der Verkehrsleistung, zu Ergebnissen mit geringerer Güte als andere Parameterkombinationen. Das Verfahren nach Li-Cassidy führt bei Betrachtung der Abweichung der Verkehrsleistung auf Ebene der Gesamtfahrt zwar definitionsgemäß zu perfekten Ergebnissen, jedoch werden die Unschärfen des Verfahrens durch die Betrachtung der Abweichung der Verkehrsleistung nach Entfernungsklassen deutlich: Auch hier schneiden Parameterkombinationen, die auf die WLAN- und Bluetooth-Daten zurückgreifen, besser ab als die Vergleichsverfahren.

Auf Basis dieser Erkenntnisse kann bereits festgehalten werden, dass die Nutzung der WLAN- und Bluetooth-Daten gegenüber den Vergleichsverfahren einen Mehrwert besitzen.

4.6.4.2 Ableitung der optimalen Parameterkombinationen

Grundsätzlich ist an dieser Stelle anzumerken, dass keine der Parameterkombinationen sich als geeignet erweisen, da bei einem Großteil der Referenzfahrten die Grenzwerte (Gütemaß GEH) nicht eingehalten werden. Mit dem Vorgehen soll dennoch die beste Parameterkombination unter den betrachteten Parameterkombinationen und unter Berücksichtigung der betrachteten Gütemaße identifiziert werden. Eine Analyse der Gründe für geringe Qualität der Verfahrensergebnisse folgt, aufbauend auf den Kapiteln zur kritischen Würdigung bei der Herleitung des Verfahrens, in Kapitel 4.7.

Eine geeignete Parameterkombination für die Filterung und Hochrechnung erzielt idealerweise bei allen Gütemaßen der betrachteten Referenzfahrten gute Ergebnisse. Für die Ermittlung der besten Parameterkombination werden daher die ermittelten Ergebnisse der Gütemaße wie in Tabelle 4-23 beschrieben zusammengefasst, normiert und gewichtet. Die Zusammenfassung zielt darauf ab, möglichst disjunkte Merkmale für die Ermittlung der optimalen Parameterkombination zu erzeugen. Für die Normierung wurden nur Parameterkombinationen genutzt, bei denen eine Hochrechnung mit dem Fratar-Verfahren oder dem Verfahren nach Li und Cassidy stattfand. Der Vergleichsfall „keineHR“, in dem keine Hochrechnung durchgeführt wurde, wird nicht berücksichtigt.

Dazu werden das GEH-Merkmal sowie das Übereinstimmungsmerkmal eingeführt. Das GEH-Merkmal fasst die Gütemaße GEH-1,5-Grenzwert, GEH-0,7-Grenzwert und die Anzahl übereinstimmender Referenzfahrten mit Einhaltung der beiden GEH-Grenzwerte zusammen, sodass die Ergebnisse des GEH-Werts nur einmalig in die Bewertung eingehen. Mit dem Übereinstimmungsmerkmal werden das Treffermaß (Anteil Treffer) und der Überfüllungsfaktor zusammengefasst, da diese nicht disjunkt sind: Je größer der Wert des Überfüllungsfaktors (>1), desto mehr Zellen sind in der Bewertungsmatrix gegenüber Referenzmatrix belegt und desto geringer wird entsprechend der Trefferanteil (Treffermaß).

Die Gewichtung bezieht die Bedeutung der einzelnen Merkmale für die Bewertung mit ein. Durch die GEH-Grenzwerte werden (zu) starke Abweichungen zwischen Bewertungs- und Referenzmatrix identifiziert, die den in Kapitel 4.4.2 definierten Schwellenwert überschreiten. Die GEH-Werte berücksichtigen gegenüber der Fehlerquadratsumme oder der Abweichung der Verkehrsleistung sowohl die relative als auch die absolute Abweichung zwischen

Bewertungs- und Referenzmatrix. Die Ergebnisse des **GEH-Merkmals** besitzen somit gegenüber den weiteren Gütemaßen die höchste Bedeutung.

Das **Übereinstimmungsmerkmal** bewertet ähnlich wie das GEH-Merkmal die Korrektheit der belegten Quelle-Ziel-Relationen zwischen Bewertungs- und Referenzmatrix. Hier wird durch das Treffermaß unabhängig von dem Wert der Zellen geprüft, wie häufig die korrekten Zellen belegt sind. In Kombination mit dem Überfüllungsfaktor können so Informationen zur Anzahl falsch belegter Relationen abgeleitet werden. Auch kann durch die kombinierte Betrachtung die vollständige Übereinstimmung zwischen Bewertungs- und Referenzmatrix identifiziert werden. Vor dem Hintergrund des Fratar-Verfahrens zur Hochrechnung und dem zugehörigen Prinzip der Informationsminimierung kommt der korrekten Belegung der Quelle-Ziel-Relationen eine hohe Bedeutung zu.

Die Summe des Betrags der **Abweichung der Verkehrsleistung** in den Entfernungsklassen fasst die Abweichungen der einzelnen Entfernungsklassen zusammen. Mit diesem Gütemaß werden insbesondere Ergebnisse identifiziert, deren Quelle-Ziel-Relationen zwischen Bewertungs- und Referenzmatrix große Abstände (Haltestellen) aufweisen. Die Bewertung kann ergänzend zu dem Übereinstimmungsmerkmal gewertet werden, um die Höhe der Abweichungen zu quantifizieren. Im Vergleich mit den vorhergehenden Merkmalen besitzt die Abweichung der Verkehrsleistung eine geringere Bedeutung.

Mit der **Fehlerquadratsumme** werden die Gesamtabweichungen zwischen den Matrizen bestimmt. Allerdings kann die reine Betrachtung der Fehlerquadratsumme zu Fehlinterpretationen führen, da - wie bei der Betrachtung der Ergebnisse der Gütemaße beschrieben - eine große Anzahl sehr kleiner Abweichungen zu geringen Werten der Fehlerquadratsumme führen, obwohl hier eine Vielzahl der Quelle-Ziel-Relationen falsch besetzt ist. Die Fehlerquadratsumme besitzt somit im Vergleich mit den weiteren Bewertungsmerkmalen die geringste Bedeutung.

Tabelle 4-23: Vorgehen zur Normierung und Gewichtung der parameterkombinationspezifischen Ergebnisse

Merkmal	Normierung	Gewicht
GEH-Merkmal <i>(GEH-1,5-Grenzwert, GEH-0,7-Grenzwert, Anz. Übereinstimmen der Referenzfahrten mit Einhaltung der beiden GEH-Grenzwerte)</i>	<p>Das GEH-Merkmal fasst die Bewertung des GEH-1,5-Grenzwerts, GEH-0,7-Grenzwerts sowie die Übereinstimmung GEH1,5 und GEH0,7 zusammen. Die drei einzelnen Gütemaße gehen jeweils mit dem gleichen Gewicht von 1/3 ein. Dabei wird für jedes der drei Gütemaße der erreichte Wert je Parameterkombination auf einer Skala von 0 bis zum maximal erreichbaren Wert normiert. Die Einhaltung des Gütemaßes auf allen 29 Referenzfahrten bildet den maximal erreichbaren Wert der drei Gütemaße (Formel 4-17).</p> <p>Beispiel:</p> <ul style="list-style-type: none"> - GEH-1,5-Grenzwert: 10 - GEH-0,7-Grenzwert: 10 - Anz. Übereinstimmung: 3 <p>führen zu einer Bewertung von 0,264 ($1/3 \times ((10 + 10 + 3)/29)$)</p> <p>Bei der Berücksichtigung der Anzahl übereinstimmender Referenzfahrten, die den GEH-1,5- und GEH-0,7-Grenzwert einhalten, ist zu berücksichtigen, dass bei sehr guten Prognoseergebnissen automatisch Übereinstimmungen entstehen – etwa wenn alle Referenzfahrten beide GEH-Grenzwerte vollständig einhalten. Somit erhält dieser perfekte Fall entsprechend der Gewichtung den Wert 1.</p>	50%
Fehlerquadratsumme	<p>Normierung des erreichten Werts je Parameterkombination auf Basis des Quotienten aus dem Betrag der Differenz des erreichten Werts zum Maximalwert (schlechteste Parameterkombination) als Dividend und dem Maximalwert als Divisor (Formel 4-18). Der Wert 0 führt zu einer Bewertung von 1, der Maximalwert zu einer Bewertung von 0.</p>	5%
Summe des Betrags der Abweichung der Verkehrsleistung in den Entf.-Klassen	Analog zu Fehlerquadratsumme	10%
Übereinstimmungsmerkmal <i>(Treffermaß; Anteil Treffer und Überfüllungsfaktor)</i>	<p>Das Übereinstimmungsmerkmal fasst die Ergebnisse zum Treffermaß (Anteil Treffer) und Überfüllungsfaktor zusammen. Die beiden Gütemaße gehen mit dem gleichen Gewicht von jeweils 1/2 ein. Zur Bestimmung werden zunächst beide Werte linear normiert:</p> <p>Überfüllungsfaktor: Normierung auf Basis des erreichten Werts je Parameterkombination auf einer Skala von 0 bis 1 (perfektes Ergebnis) (Formel 4-19). Eine starke Überfüllung (Überfüllungsfaktor >2) sowie eine leerer Matrix (Überfüllungsfaktor = 0) führen zu einer Bewertung von 0, ein Überfüllungsfaktor von 1 zu einer Bewertung von 1.</p> <p>Treffermaß: Normierung auf Basis des erreichten Werts je Parameterkombination auf einer Skala von 0 bis zum maximal erreichbaren Wert von 1. Der maximal erreichbare Wert ist die vollständige Übereinstimmung und somit 100% Anteil von Treffern.</p> <p>Die Zusammenführung der beiden Werte in das Übereinstimmungsmerkmal erfolgt durch Multiplikation der normierten Werte des Treffermaßes und des Überfüllungsfaktors. Somit wird das Treffermaß mit dem Überfüllungsfaktor und damit der Anzahl belegter Zellen in Beziehung gesetzt. Eine hohe Trefferquote (0,6) bei gleichzeitig hohem Überfüllungsfaktor (1,8) führt zu einer geringen Bewertung. Zur Verdeutlichung dient Abbildung 4-7.</p> <p>Beispiel:</p> <ul style="list-style-type: none"> - Treffermaß: 0,6 - Normierter Überfüllungsfaktor: 0,2 <p>führen zu einer Bewertung von 0,12 ($0,6 \cdot 0,2$).</p>	35%

$$f(GEH_i) = \frac{1}{3} * (n_{GEH1,5_i} + n_{GEH0,7_i} + n_{\ddot{U}_i}) / n_{Ref} \tag{Formel 4-17}$$

mit

- f(GEH_i) Ergebnis der Normierung des GEH-Merkmals von Parameterkombinationen i
- n_{GEH1,5_i} Anzahl Einhaltung des GEH-1,5-Grenzwerts der betrachteten Parameterkombinationen i über alle Referenzfahrten
- n_{GEH0,7_i} Anzahl Einhaltung des GEH-0,7-Grenzwerts der betrachteten Parameterkombinationen i über alle Referenzfahrten
- n_{Ü_i} Anzahl übereinstimmender Referenzfahrten mit Einhaltung der beiden GEH-Grenzwerte der betrachteten Parameterkombinationen i
- i Index der betrachteten Parameterkombination
- n_{Ref} Anzahl der Referenzfahrten (Maximalwert zur Einhaltung der Grenzwerte)

$$f(FQS_i) = \frac{|FQS_i - FQS_{max}|}{FQS_{max}} \tag{Formel 4-18}$$

mit

- f(FQS_i) Ergebnis der Normierung der Fehlerquadratsumme von Parameterkombinationen i
- FQS_i Fehlerquadratsumme der betrachteten Parameterkombinationen i
- i Index der betrachteten Parameterkombination
- FQS_{max} Maximalwert der Fehlerquadratsumme aller Parameterkombinationen

$$f(\ddot{U}F_i) = \begin{cases} \ddot{U}F_i & \text{für } \ddot{U}F_i \leq 1 \\ |2 - \ddot{U}F_i| & \text{für } \ddot{U}F_i > 1 \wedge \ddot{U}F_i \leq 2 \\ 0 & \text{für } \ddot{U}F_i > 2 \end{cases} \tag{Formel 4-19}$$

mit

- f(ÜF_i) Ergebnis der Normierung des Überfüllungsfaktors von Parameterkombinationen i
- ÜF_i Mittlerer Überfüllungsfaktor der betrachteten Parameterkombinationen i
- i Index der betrachteten Parameterkombination

		Überfüllungsfaktor																				
		0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2
Treffermaß (Anteil Treffer)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10%	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	9%	8%	7%	6%	5%	4%	3%	2%	1%	0%
	20%	0%	2%	4%	6%	8%	10%	12%	14%	16%	18%	20%	18%	16%	14%	12%	10%	8%	6%	4%	2%	0%
	30%	0%	3%	6%	9%	12%	15%	18%	21%	24%	27%	30%	27%	24%	21%	18%	15%	12%	9%	6%	3%	0%
	40%	0%	4%	8%	12%	16%	20%	24%	28%	32%	36%	40%	36%	32%	28%	24%	20%	16%	12%	8%	4%	0%
	50%	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	45%	40%	35%	30%	25%	20%	15%	10%	5%	0%
	60%	0%	6%	12%	18%	24%	30%	36%	42%	48%	54%	60%	54%	48%	42%	36%	30%	24%	18%	12%	6%	0%
	70%	0%	7%	14%	21%	28%	35%	42%	49%	56%	63%	70%	63%	56%	49%	42%	35%	28%	21%	14%	7%	0%
	80%	0%	8%	16%	24%	32%	40%	48%	56%	64%	72%	80%	72%	64%	56%	48%	40%	32%	24%	16%	8%	0%
	90%	0%	9%	18%	27%	36%	45%	54%	63%	72%	81%	90%	81%	72%	63%	54%	45%	36%	27%	18%	9%	0%
	100%	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%	0%

Abbildung 4-7: Bewertung des Übereinstimmungsmerkmals unter Berücksichtigung verschiedener Ergebnisse des Treffermaß und Überfüllungsfaktors in Prozent (Schrittweite: 0,1)

Die Normierung wird entsprechend den Ausführungen auf Basis linearer Funktionen durchgeführt. Dem liegt die Annahme zugrunde, dass mit stärkerer Abweichung des Bewertungsergebnisses des Gütemaßes vom Optimalzustand die Güte gleichmäßig bis zum Minimum (z. B. Einhaltung GEH-0,7-Grenzwert bei keiner Referenzfahrt) sinkt.

Die Ergebnisse der besten Parameterkombinationen entsprechend der Normierung und Gewichtung (Punktwert) zeigt Tabelle 4-24. Tabelle 4-25 stellt die originären Ergebnisse der Gütemaße dar.

Die identifizierten Parameterkombinationen zählen auch bei Änderung der Gewichtung stets zu den 10% der besten Parameterkombinationen. Dabei ist zu berücksichtigen, dass mit einer Änderung der Gewichtung die Bewertung der Parameterkombinationen unterschiedliche Werte annimmt. Die identifizierten Parameterkombinationen können somit als robust bezeichnet werden. Als weitere Gewichtungsoptionen wurden u. a. geprüft:

- gleiche Gewichtung der vier Merkmale (Gewicht jeweils 25%)
- gleiche Gewichtung der Gütemaße, sodass das GEH-Merkmal (umfasst drei von sieben Gütemaßen) mit Faktor 3 (43%), das Übereinstimmungsmerkmal (umfasst zwei von sieben Gütemaßen) mit Faktor 2 (29%) und beiden verbleibenden Merkmale jeweils mit Faktor 1 (jeweils 14%) eingehen.

Aus der Bewertung der Ergebnisse geht hervor, dass sieben Parameterkombinationen mit 0,40 Punkten die höchsten Punktwerte und damit die besten Ergebnisse erzielen. Die ermittelten Parameterkombinationen nutzen als Datenquelle sowohl WLAN- als auch Bluetooth-Daten. Für die Hochrechnung werden Reduktionsgrenzwerte von 0,5 oder höher verwendet. Als Filtersets eignen sich GPS-Filter. Das beste Ergebnis mit 0,4027 Punkten erreicht das Filterset P50_GPS100 zusammen mit einer Hochrechnung mit Hilfwert 0,00001 und Reduktionsgrenzwert 1,0,

Tabelle 4-24: Beste Parameterkombinationen auf Basis der Normierung und Gewichtung nach Tabelle 4-23

Datenquelle	Filterset	HR-Set	Fuzzy	GEH-Merkmal	FQS	Summe des Betrags der Abweichung der Verkehrsleistung in den EKL	Übereinstimmungsmerkmal	Summe (Normierung)
WLAN_BT	P50_GPS100	FRAT_h0,00001_r0,5 ----	x	0,36	0,33	0,31	0,50	0,40
WLAN_BT	P50_GPS100	FRAT_h0,00001_r1,0 ----		0,31	0,20	0,32	0,59	0,40
WLAN_BT	P51_GPS200	FRAT_h0,00001_r0,5 ----	x	0,33	0,29	0,36	0,52	0,40
WLAN_BT	P51_GPS200	FRAT_h0,00001_r1,0 ----	x	0,30	0,18	0,36	0,58	0,40
WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5 ----	x	0,36	0,28	0,28	0,51	0,40
WLAN_BT	P51_GPS200	FRAT_h1,00000_r1,0 ----	x	0,33	0,17	0,23	0,57	0,40
WLAN_BT	P52_GPS400	FRAT_h0,00001_r1,0 ----		0,33	0,16	0,27	0,57	0,40

Tabelle 4-25: Beste Parameterkombinationen mit originären Gütemaßen

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung in den Enfernungsklassen in [100 km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN_BT	P50_GPS100	FRAT_h0,00001_r0,5 ----	x	10	13	8	689	3,3	57%	1,1
WLAN_BT	P50_GPS100	FRAT_h0,00001_r1,0 ----		10	11	6	824	3,2	59%	1,0
WLAN_BT	P51_GPS200	FRAT_h0,00001_r0,5 ----	x	10	11	8	732	3,0	58%	1,1
WLAN_BT	P51_GPS200	FRAT_h0,00001_r1,0 ----	x	9	10	7	848	3,1	59%	1,0
WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5 ----	x	10	13	8	737	3,4	58%	1,1
WLAN_BT	P51_GPS200	FRAT_h1,00000_r1,0 ----	x	9	13	7	855	3,7	58%	1,0
WLAN_BT	P52_GPS400	FRAT_h0,00001_r1,0 ----		10	12	7	862	3,5	58%	1,0

Das beste unter den berücksichtigten Verfahren ohne Ausgangsdaten (WLAN- und Bluetooth-Daten) wird durch die Hochrechnung mit dem Fratar-Verfahren und einem Reduktionsgrenzwert von 1,0 erreicht. Hier wird ein

Punktwert von 0,29 erzielt. Die Datenquelle und die Höhe des Hilfswerts haben keinen Einfluss auf das Ergebnis des Vergleichsverfahrens.

Unter den besten 10% der Parameterkombinationen, ermittelt auf Basis des 90%-Quantils, verbleiben 116 Parameterkombinationen mit einer Punktbeurteilung von mindestens 0,35. Unter diesen 116 Parameterkombinationen sind 44 Parameterkombinationen, die Fuzzy-Sets nutzen. Bei der Interpretation ist zu beachten, dass für nur acht der Filtersets die Fuzzifizierung eingesetzt wurde (Kapitel 4.5.1). Entsprechend dem maximal erreichten Punktwert der Vergleichsverfahren (0,29 Punkte) ist keins der Vergleichsverfahren in den besten 10% der Parameterkombinationen (Punktwert von mindestens 0,35) enthalten.

4.6.4.3 Bewertung der Parameterkombinationen nach Linien und Nachfrageklassen

Als Ergänzung zur Auswertung der Bewertung der Parameterkombinationen über alle Linienfahrten sollen mit diesem Kapitel die Ergebnisse auf Unterschiede in den optimalen Parameterkombinationen anhand differenzierter Auswertungen

- nach Linie (indirekt: Anzahl Relationen, Raum) und
- nach Nachfrageklasse (Fahrgastnachfrage je Linienfahrt, Datenbasis: AFZS)

geprüft werden. Die Ausprägungen der Merkmale werden im Folgenden zusammen mit den Ergebnissen der Auswertungen beschrieben. Die Ergebnisse sind im Detail in Anhang 3 und Anhang 4 dargestellt. In diesem Kapitel werden die wesentlichen Erkenntnisse zusammengefasst.

Auswertung nach betrachteter Linie

Die Ergebnisse der Auswertungen in Abhängigkeit der Linie zeigt, dass mit den Linien erwartungsgemäß eine unterschiedliche Anzahl an bedienten Haltestellen und damit eine unterschiedliche Anzahl möglicher Quelle-Ziel-Relationen einhergeht. Dabei ist zu beachten, dass die Anzahl an Haltestellen je Zeitschicht und Tagestyp (leicht) variiert.

- Linie 100 (etwa 140 Relationen) ($n = 20$)
- Linie 54 (etwa 300 Relationen) ($n = 4$)
- Linie 500 (etwa 500 Relationen) ($n = 5$)

In Tabelle 4-26 ist ein Ausschnitt der Ergebnisse je betrachteter Linie dargestellt. Dabei ist zu berücksichtigen, dass je Linie jeweils zwei der Parameterkombinationen mit dem höchsten Punktwert dargestellt werden. Insgesamt gibt es – wie in Anhang 3 beschrieben – zum Teil weitere Parameterkombinationen, die ebenfalls den höchsten Punktwert erreichen.

Tabelle 4-26: Übersicht (Ausschnitt) der besten Parameterkombinationen auf Basis der Normierung und Gewichtung (Tabelle 4-23) nach betrachteter Linie

Linie	Datenquelle	Filterset	HR-Set	Fuzzy	GEH-Merkmal	FQS	Summe des Betrags der Abweichung der Verkehrsleistung in den EKL	Übereinstimmungsmerkmal	Summe (Normierung)
100	WLAN_BT	P32_GPS600_RSSI60	FRAT_h1,00000_r0,5_----	x	0,3	0,34	0,45	0,47	0,38
	WLAN_BT	P51_GPS200	FRAT_h1,00000_r1,0_----	x	0,25	0,21	0,32	0,57	0,37
	Vergleichsverfahren mit höchstem Punktwert (P99_Fullfilter FRAT_h0,00001_r1,0_----)								
54	WLAN	P51_GPS200	FRAT_h1,00000_r0,5_----	x	1	0,69	0,77	0,7	0,86
	WLAN	P51_GPS200	FRAT_h1,00000_r1,0_----	x	1	0,69	0,77	0,7	0,86
	Vergleichsverfahren mit höchstem Punktwert (P99_Fullfilter FRAT_h0,00001_r0,5_----)								
500	WLAN_BT	P09_L5_GPS400_RSSI90	FRAT_h0,00001_r1,0_----		0,4	0,25	0,55	0,49	0,44
	WLAN_BT	P52_GPS400	FRAT_h0,00001_r1,0_----		0,4	0,28	0,51	0,51	0,44
	Vergleichsverfahren mit höchstem Punktwert (P99_Fullfilter FRAT_h0,00001_r1,0_----)								

Die Ergebnisse zeigen, dass es in Abhängigkeit der betrachteten Linie zu unterschiedlichen optimalen Parameterkombinationen kommt. Aus dem Vergleich mit den in Kapitel 4.6.4.2 ermittelten Parameterkombinationen geht hervor, dass diese bei Betrachtung der verschiedenen Linien nicht durchgehend zu den höchsten, aber je betrachteter Linie mindestens zu den 20% der besten Ergebnisse auf Basis der Ermittlung der Punktwerte führen. Die in Kapitel 4.6.4.2 ermittelten Parameterkombinationen können somit auf Basis der durchgeführten Tests mit drei Gruppen und einem Stichprobenumfang von insgesamt 29 Referenzfahrten als robust gegenüber den unterschiedlichen Linien angesehen werden. Dabei führt insbesondere die Parameterkombination unter Nutzung von WLAN- und Bluetooth-Daten, dem Filterset P51_GPS200 unter Verwendung der Hochrechnung mit dem Fratar-Verfahren (Hilfswert 1,0000, Reduktionsgrenzwert 0,5) je betrachteter Klasse zu

Ergebnissen, die einen höheren Punktwert als 90% der weiteren betrachteten Parameterkombinationen erreichen. Die Vergleichsverfahren (ohne Ausgangsdaten) führen zu Ergebnissen mit geringerer Güte als die in Kapitel 4.6.4.2 ermittelten Parameterkombinationen. Unter den Vergleichsverfahren erreicht das Fratar-Verfahren mit einem Reduktionsgrenzwert von 1,0 oder 0,5 die besten Ergebnisse.

Aus Tabelle 4-26 ist weiterhin ersichtlich, dass der erreichte Punktwert für Linie 54 deutlich höher ist als bei den weiteren betrachteten Linien. Dies ist einerseits auf die geringere Fahrgastnachfrage bei den erfassten Referenzfahrten der Linie 54 zurückzuführen: Mit geringerer Nachfrage wird auch der mögliche Lösungsraum kleiner (vgl. Kapitel 4.6.5). Gleichzeitig ist dabei zu berücksichtigen, dass die Stichprobe lediglich vier Referenzfahrten umfasst, von denen zwei nur eine mögliche Lösung besitzen.

Auswertung nach Nachfrageklassen

Neben der Unterscheidung nach Linien findet eine weitere Unterscheidung nach Nachfrageklassen statt. Die Nachfrageklassen basieren auf den durch die AFZS ermittelten Fahrgastzahlen für die gesamte Linienfahrt. Die betrachteten Referenzfahrten werden in drei Klassen unterteilt. Die Klassengrenzen wurden die auf Basis der Terzilgrenzen (33%, 66%) der Fahrgastnachfrage festgelegt:

- Nachfrageklasse 1: 0 bis 13 Fahrgäste je Linienfahrt (n = 6)
- Nachfrageklasse 2: 14 bis 21 Fahrgäste je Linienfahrt (n = 14)
- Nachfrageklasse 3: über 21 Fahrgäste je Linienfahrt (n = 9)

In Tabelle 4-27 ist ein Ausschnitt der Ergebnisse je betrachteter Nachfrageklasse dargestellt. Dabei ist zu berücksichtigen, dass je Linie jeweils zwei der Parameterkombinationen mit dem höchsten Punktwert dargestellt werden. Insgesamt gibt es – wie in Anhang 4 beschrieben – zum Teil weitere Parameterkombinationen, die ebenfalls den höchsten Punktwert erreichen.

Die Ergebnisse der Auswertung in Abhängigkeit der Fahrgastnachfrage zeigen – wie bereits die Ergebnisse der Auswertung in Abhängigkeit der Linie – dass es in Abhängigkeit der betrachteten Nachfrageklasse zu unterschiedlichen optimalen Parameterkombinationen kommt. Aus dem Vergleich mit den in Kapitel 4.6.4.2 ermittelten Parameterkombinationen geht hervor, dass die über alle Referenzfahrten ermittelten optimalen Parameterkombinationen nicht durchgehend zu den höchsten, aber je Nachfrageklasse mindestens zu den besten 15% der besten Ergebnisse auf Basis der Ermittlung der

Punktwerte führen. Die in Kapitel 4.6.4.2 ermittelten Parameterkombinationen können somit auf Basis der durchgeführten Tests mit drei Gruppen und einem Stichprobenumfang von insgesamt 29 Referenzfahrten als robust gegenüber der Fahrgastnachfrage angesehen werden. Auch hier führt die Parameterkombination unter Nutzung von WLAN- und Bluetooth-Daten, dem Filterset P51_GPS200 unter Verwendung der Hochrechnung mit dem Fratar-Verfahren (Hilfswert 1,0000, Reduktionsgrenzwert 0,5) je betrachteter Klasse zu Ergebnissen, die einen höheren Punktwert als 90% der weiteren betrachteten Parameterkombinationen erreicht. Die Vergleichsverfahren (ohne Ausgangsdaten) führen zu Ergebnissen mit geringerer Güte als die in Kapitel 4.6.4.2 ermittelten Parameterkombinationen. Unter den Vergleichsverfahren erreicht das Fratar-Verfahren mit einem Reduktionsgrenzwert von 1,0 oder 0,5 die besten Ergebnisse.

Tabelle 4-27: Übersicht (Ausschnitt) der besten Parameterkombinationen auf Basis der Normierung und Gewichtung (Tabelle 4-23) nach Nachfrageklassen

Nachfrageklasse	Datenquelle	Filterset	HR-Set	Fuzzy	GEH-Merkmal	FQS	Summe des Betrags der Abweichung der Verkehrsleistung in den EKL	Übereinstimmungsmerkmal	Summe (Normierung)
1 0 bis 13 Fahrgäste	WLAN_BT	P51_GPS200	FRAT_h0,00001_r0,5_----	x	0,89	0,68	0,78	0,8	0,84
	WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5_----	x	0,89	0,68	0,73	0,8	0,83
	Vergleichsverfahren mit höchstem Punktwert (P99_Fullfilter FRAT_h0,00001_r0,5_----)								0,57
2 14 bis 21 Fahrgäste	WLAN_BT	P32_GPS600_RSSI60	FRAT_h1,00000_r1,0_----	x	0,26	0,27	0,57	0,53	0,39
	WLAN	P44_L5_GPS600_RSSI70	FRAT_h1,00000_r1,0_----	x	0,26	0,2	0,5	0,54	0,38
	Vergleichsverfahren mit höchstem Punktwert (P99_Fullfilter FRAT_h0,00001_r0,5_----)								0,26
3 über 21 Fahrgäste	WLAN_BT	P01_L5_GPS200_RSSI90	FRAT_h1,00000_r0,5_----		0,22	0,15	0,43	0,5	0,34
	WLAN_BT	P52_GPS400	FRAT_h1,00000_r1,0_----		0,22	0,19	0,44	0,49	0,34
	Vergleichsverfahren mit höchstem Punktwert (P99_Fullfilter FRAT_h0,00001_r1,0_----)								0,24

Aus Tabelle 4-27 ist weiterhin ersichtlich, dass der erreichte Punktwert in der ersten Nachfrageklasse deutlich höher liegt als in den Nachfrageklassen 2 und 3. Dies lässt den Schluss zu, dass die Ergebnisqualität umso höher ist, je geringer die Nachfrage. Dieser Zusammenhang lässt sich grundsätzlich

auf den mit der geringeren Nachfrage verbundenen kleineren Lösungsraum zurückführen (vgl. Kapitel 4.6.5).

4.6.5 Bewertung der Ergebnisse

Im Anwendungsfall liegen die Randsummen der Matrizen auf Basis von AFZS-Daten sowie Informationen mit unbekannter Genauigkeit zu den Ein- und Ausstiegshaltestellen der Fahrgäste vor (WLAN- und Bluetooth-Daten). Die Datengrundlage enthält jedoch nicht ausschließlich Informationen zu den Fahrgästen, sondern wird durch Stördaten beeinflusst. Auf diese Verzerrungen wird detailliert in Kapitel 4.7 eingegangen.

In Abhängigkeit der Fahrgastanzahl und der möglichen Relationen der betrachteten Linienfahrt besteht eine Vielzahl möglicher Matrizen als Lösung, die die vorgegebenen Randsummen einhalten (Friedrich et al. 2021). Von allen möglichen Matrizen bildet jedoch nur eine (bei ganzzahligen Werten) die Realität ab.

Ziel des Matrixschätzverfahrens ist es, die wahrscheinlichste unter den möglichen Matrizen auszuwählen. Im Fall der Einser-Matrix liegen keine Informationen zur realen Verteilung der Fahrgäste zwischen den Haltestellen vor. Im Fall der Nutzung von WLAN- und Bluetooth-Daten liegen Informationen zu der Verteilung vor, wenn auch verzerrt, die – wie die Ergebnisse zeigen – zu besseren Ergebnissen führen als Alternativen ohne Ausgangsdaten. Die Ergebnisse der Hochrechnung – mit oder auch ohne Verwendung von WLAN- und Bluetooth-Daten – können hier also nur zufällig die korrekte Matrix treffen, da die reale Matrix nicht zwingend die wahrscheinlichste Matrix darstellt.

Die hohe Bedeutung (vollständig) korrekter Ausgangsdaten wurde durch die WVI Prof. Dr. Wermuth Verkehrsforschung und Infrastrukturplanung GmbH (2022) nachgewiesen: In einer ergänzenden Analyse wurde die Wirkung von zufälligen Reduzierungen der belegten Relationen in den Referenzmatrizen auf die Ergebnisse des Fratar-Verfahrens geprüft. Dazu wurden die belegten Relationen der bestehenden Referenzmatrizen in Schritten von 10% reduziert, ohne zusätzlich WLAN- oder Bluetooth-Daten einzubeziehen. Eine Reduzierung von 10% gegenüber der vollständigen Referenzmatrix bedeutet, dass in der Ausgangsmatrix nur noch 90% der Relationen belegt sind, die in der Referenzmatrix belegt sind. Damit das Fratar-Verfahren konvergieren kann, wurde eine zunehmende Anzahl an Hilfswerten notwendig. Die Ergebnisse zeigen, dass mit zunehmender Reduzierung die Ergebnisgüte erwartungsgemäß abnimmt.

Hintergrund ist, dass durch die zunehmende Anzahl an Hilfwerten und damit belegten Relationen die Anzahl möglicher Matrizen zur Lösung der Hochrechnung unter Einhaltung der Randsummen ansteigt. Das Fratar-Verfahren wählt dabei nur eine Lösung – die wahrscheinlichste Lösung – aus, die in der Regel nicht mit der Realität (Referenzmatrix) übereinstimmt.

Der Mehrwert des Verfahrens kann auf Basis des Vergleichs der erzielten Ergebnisse unter Berücksichtigung der WLAN- und Bluetooth-Daten mit einer gleichverteilten Matrix abgeleitet werden, die entsprechend keine Informationen zur Verteilung der Quelle-Ziel-Relationen in der Realität enthält.

Bereits mit der Auswertung der Einhaltung der GEH-Grenzwerte wird deutlich, dass die erforderliche Genauigkeit der Verfahrensergebnisse nicht konsequent durch einzelne der untersuchten Parameterkombinationen über alle Referenzfahrten eingehalten werden kann. Der Maximalwert einer Parameterkombination liegt bei 13 Referenzfahrten – bei einzelnen Referenzfahrten ist es mit keiner Parameterkombination gelungen, die definierten Grenzwerte einzuhalten.

Die gütemaßspezifischen Auswertungen zeigen, dass je Gütemaß unterschiedliche Parameterkombinationen zu den besten Ergebnissen führen. Dabei ist auffällig, dass bei den besten Ergebnissen stets WLAN- und Bluetooth-Daten genutzt wurden. Auch die Vergleichsverfahren führen bei keinem der untersuchten Gütemaße zu den besten Ergebnissen.

Diese Erkenntnisse werden mit der Ableitung der optimalen Parameterkombination bestätigt: Die besten Ergebnisse werden bei Nutzung von WLAN- und Bluetooth-Daten als Datenquelle, einem auf GPS-Luftlinienentfernung basierenden Filterset (100m, 200m und 400m) und der Hochrechnung mit dem Fratar-Verfahren und einem Reduktionsgrenzwert von mindestens 0,5 erzielt. Die ermittelten Parameterkombinationen haben sich auch bei differenzierter Betrachtung der Referenzfahrten nach Linien und Nachfrageklassen als robust erwiesen. Insbesondere die Parameterkombination unter Verwendung von WLAN- und Bluetooth-Daten, dem Filterset P51_GPS200 mit Fuzzifizierung und Hochrechnung mit dem Fratar-Verfahren (Hilfswert 1,000, Reduktionsgrenzwert 0,5) liegt durchweg unter den besten 10% der betrachteten Parameterkombinationen. Darüber hinaus schneiden die ermittelten Parameterkombinationen besser ab als

- das beste Verfahren ohne Ausgangsdaten,
- das beste Verfahren ohne Filterung der Eingangsdaten sowie

- das beste Verfahren ohne Hochrechnung der Ausgangsmatrizen.

Dies bestätigt den Mehrwert des Verfahrens sowie der entwickelten Verfahrensschritte: Die Nutzung der WLAN- und Bluetooth-Daten als Datenquelle erzielt bessere Werte als Alternativen, denen keinen weiteren Informationen zur Verteilung der Fahrgäste zur Verfügung stehen. Mit der Filterung der Daten können Stördaten zwar nicht eindeutig identifiziert und entfernt werden, jedoch wird die Datengrundlage für die Erstellung der Ausgangsmatrix gegenüber Alternativen ohne Filterung verbessert. Darüber hinaus wirkt die Hochrechnung erwartungsgemäß durch den Ausgleich der Randsummen positiv auf die Güte der Ergebnisse. Unter den besten Parameterkombinationen befinden sich sowohl Alternativen mit als auch ohne Fuzzifizierung. Bei einem Vergleich der Teilmenge an Parameterkombinationen, für die die Fuzzifizierung durchgeführt wurde, zeigt sich, dass die Fuzzifizierung nicht eindeutig zu besseren oder schlechteren Ergebnissen gegenüber dem Vergleichsfall ohne Fuzzifizierung führt: Etwa die Hälfte der Parameterkombinationen erzielt mit Fuzzifizierung, die andere Hälfte führt ohne Fuzzifizierung zu besseren Ergebnissen. Werden weiterhin nur die Parameterkombinationen unter Nutzung des Fratar-Verfahrens und eines Reduktionsfaktors von 0,5 oder 1,0 betrachtet (diese Einstellung zur Hochrechnung hat tendenziell zu besseren Ergebnissen geführt), zeigt sich, dass die Alternativen mit Fuzzifizierung bei zwei Drittel der verbleibenden Parameterkombinationen zu besseren Ergebnissen führen. Daraus kann abgeleitet werden, dass die Fuzzifizierung bei dem größeren Teil der betrachteten Parameterkombinationen zu einem Mehrwert führt. Gleichzeitig bestehen auch Parameterkombinationen, die ohne Fuzzifizierung Ergebnisse mit einer höheren Güte liefern.

Für die Vergleichsverfahren hat sich gezeigt, dass das Hochrechnungsverfahren nach Li und Cassidy auf Basis der gewählten Alternativen geringere Punktwerte erreicht als das Fratar-Verfahren. Dabei ist anzumerken, dass das Verfahren nach Li-Cassidy einen Großteil der Zellen mit Werten kleiner 1 befüllt. Daher kann die Eignung des Verfahrens für die Anwendung auf einzelnen Linienfahrten mit einer geringen Anzahl an Fahrgästen und entsprechend belegten Relationen angezweifelt werden. Forschungsbedarf besteht aus Sicht des Autors für einen Vergleich auf Basis größerer Datenbestände, etwa über einen Monat oder über längere Zeiträume aggregierte Daten einer Linienfahrt. Somit würden die Belegungen in den Zellen

auf Basis des Li-Cassidy-Verfahrens ebenfalls Werte größer 1 annehmen können²⁰.

Unter den geeignetsten Parameterkombinationen befinden sich über alle Referenzfahrten, sowie nach Linien und Nachfrageklassen, Varianten mit und ohne Fuzzifizierung. Dabei ist zu berücksichtigen, dass die Fuzzifizierung nur für einen Teil der Filtersets getestet wurde. Die Ergebnisse zeigen, dass Varianten mit Fuzzifizierung nicht durchweg zu besseren Ergebnissen führen als Varianten ohne Fuzzifizierung. Grund dafür können aus Sicht der Autors Verzerrungen in der Datengrundlage sein: Mit dem Einbezug von false positives (Filterung) sowie dem Fehlen von Fahrgastdaten (Filterung, Datenerfassung) kommt es zu unbekanntem Verzerrungen in der Ausgangsmatrix. Das Vorliegen von Verzerrungen zeigt bereits die Auswertung zur Einhaltung der GEH-Grenzwerte. Auf etwaige Verzerrungen und Fehlerquellen wird detailliert in Kapitel 4.7 eingegangen. Es ist jedoch davon auszugehen, dass die Fuzzifizierung bei einem Datenset, das nur die Daten von Fahrgästen beinhaltet, zu besseren Ergebnissen als Varianten ohne Fuzzifizierung führen würde, da damit die Unschärfen bei der Zuordnung der Ein- und Ausstiegshaltestelle berücksichtigt werden.

Die Fuzzifizierung bietet den Vorteil, dass eine geringere Anzahl an Hilfwerten notwendig ist und die Ausgangsmatrix mit weiteren Werten befüllt wird, die jedoch einen Bezug zu den erfassten Daten besitzt. Hilfwerte dagegen werden ohne weitere Informationen überall dort eingesetzt, wo sie notwendig sind.

Das Verfahren überschreitet im derzeitigen Zustand die festgelegten GEH-Grenzwerte bei einem Großteil der Referenzfahrten. Die Verfahrensergebnisse besitzen damit eine hohe Wahrscheinlichkeit, dass die Zuordnung der Fahrgäste auf die Quelle-Ziel-Relationen Fehler aufweist, die oberhalb der gesetzten Grenzwerte liegen. Vor dem Hintergrund der in Kapitel 1.3 aufgezeigten Einsatzbereiche von Quelle-Ziel-Daten sowie der damit einhergehenden Anforderungen an die räumliche und zeitliche Genauigkeit ist jedoch zu berücksichtigen, dass bei bislang genutzten Quelle-Ziel-Daten unbekannte Fehler vorliegen. Grund für die Fehler ist, dass eine repräsentative Erhebung von Quelle-Ziel-Verflechtungen mit klassischen Verfahren praktisch nicht möglich ist bzw. mit einem erheblichen Aufwand verbunden wäre (Sommer 2002). Fehler in den bestehenden Daten sind nicht quantifiziert.

²⁰ Das Reduktionsverfahren ist nur für Hochrechnungen mit dem iterierenden Fratar-Verfahren geeignet.

Demgegenüber wurden im Rahmen der Verfahrensbewertung die Ungenauigkeiten der Verfahrensergebnisse bereits für 29 Referenzfahrten anhand verschiedener Gütemaße ermittelt.

Quelle-Ziel-Daten stammen – wie in Kapitel 1.3 beschrieben – bislang aus Verkehrsmodellen und basieren auf Haushalts- und/oder Fahrgastbefragungen, die die Quelle-Ziel-Verflechtungen in der Regel nicht repräsentativ erfassen. Auf Befragungen basierende Daten haben den Nachteil, dass die zeitliche Aussagekraft begrenzt ist, sofern die Befragungen nicht in engem Turnus wiederkehrend durchgeführt werden.

Die Anforderungen an die räumliche und zeitliche Genauigkeit sind abhängig vom Anwendungszweck innerhalb der Einsatzbereiche sowie den Planern vorliegenden Daten. Mit dem aufgezeigten Verfahren werden dauerhaft Daten erfasst, die eine bislang nicht vorhandene Datengrundlage bilden. Somit wird es den ÖPNV-Betreibern ermöglicht, Informationen über wiederkehrende Events (etwa Fußballspiele, Konzerte o. ä.) zu sammeln und das Angebot entsprechend anzupassen, um Kapazitätsüber- sowie -unterschreitungen zu vermeiden. Gleichzeitig können die Daten auf Tarifzonenebene aggregiert werden und in die Entwicklung von Tarifen eingehen.

Weiterhin liegen stets aktuelle Daten vor, auf deren Basis verschiedene Formen der Aggregation möglich sind:

- räumliche Aggregation (etwa Verkehrszellen, Tarifzonen)
- zeitliche Aggregation („Musterfahrten“ nach Tagestyp und/oder Zeitschicht)

Mit der Aggregation kommt es zu einer Reduzierung der Informationen, sodass davon auszugehen ist, dass die Qualität der Ergebnisse erhöht wird.

4.7 Prüfung von Ursachen für die Güte der Verfahrensergebnisse

Zur Ermittlung der Ursachen für die teilweise geringe Qualität der Verfahrensergebnisse sollen an dieser Stelle – in Ergänzung der Kapitel zur kritischen Würdigung der einzelnen Verfahrensschritte – sieben Thesen geprüft werden. Dabei werden insbesondere die Datenquelle, Filterung und Erzeugung von Quelle-Ziel-Matrizen betrachtet. Folgende Thesen wurden untersucht:

- These 1 (Datengrundlage): „Der Einfluss der Anzahl an Fahrgästen auf die Anzahl der empfangenen Signale ist gering. Das bedeutet, dass der Großteil der empfangenen Signale von Nicht-Fahrgästen stammt.“
- These 2 (Datengrundlage): „Es ist nur eine Teilmenge der Fahrgäste erfassbar: Nur ein Teil der Fahrgäste führt ein WLAN- und/oder Bluetooth-fähiges Endgerät mit sich.“
- These 3 (Filterung): „Die erfassten Informationen aus den WLAN- und Bluetooth-Daten lassen keine eindeutige Filterung von Fahrgastdaten zu.“
- These 4 (Erzeugung von Quelle-Ziel-Matrizen): „Die Zuordnung der erfassten (und gefilterten) Signale zu Ein- und Ausstiegshaltestellen realer Fahrgäste ist ungenau.“
- These 5 (Fuzzifizierung): „Bei der Ableitung der Häufigkeitsverteilung zwischen Bluetooth- und WLAN-Signalen kommt es zu einer Überschätzung der Wahrscheinlichkeiten „kleiner“ Zeitintervalle.“
- These 6 (Hochrechnung): „Fehlende oder verzerrte Informationen zu den Quelle-Ziel-Beziehungen der Fahrgäste in der Ausgangsmatrix können durch das Hochrechnungsverfahren nicht ausgeglichen werden und führen somit zu Verfahrensergebnissen mit geringer Güte.“

Darüber hinaus wurde im Hinblick auf die Entwicklung der Datengrundlage These 7 geprüft:

- These 7 (Datengrundlage): „Der Anteil von Smartphones mit aktiver Randomisierung nimmt zu.“

Sofern nicht anders dargestellt basieren die folgenden Auswertungen auf den bis 12/2021 erfassten Daten (Dataset Bieland/Mobile Data Fusion 2022).

These 1 (Datengrundlage): „Der Einfluss der Anzahl an Fahrgästen auf die Anzahl der empfangenen Signale ist gering. Das bedeutet, dass der Großteil der empfangenen Signale von Nicht-Fahrgästen stammt.“

Die erfassten Signale bilden die Grundlage für das Verfahren. Im Idealfall werden nur Daten von Fahrgästen empfangen. Dies ist jedoch aufgrund der Empfangsreichweite der Erfassungshardware nicht möglich. Somit ist die Erfassung von Stördaten nicht zu vermeiden. Mit dem Vorliegen von Stördaten wird der Verfahrensschritt der Filterung notwendig, um die Ausgangsmatrizen möglichst korrekt mit Daten von Fahrgästen zu besetzen.

Zur Prüfung fand eine linienspezifische Betrachtung der Signalrate (Anzahl aller erfasster WLAN- und Bluetooth-Signale ohne Filterung je Fahrtminute) von Fahrzeugen der Linien 54, 100 und 500 an einem Stichtag (04.11.2021) statt. Dazu wurden jeweils alle Linienfahrten von je zwei Fahrzeugen der Linien berücksichtigt. Bei der Fahrzeugauswahl wurde – wenn vorhanden – ein Fahrzeug gewählt, das am Stichtag bei mindestens einer Linienfahrt keine Fahrgäste hatte (nur möglich bei Linie 54 am 04.11.2021). Linienfahrten zum Betriebshof wurden bei der Fahrzeugauswahl sowie der Auswertung nicht berücksichtigt, da diese nicht der Personenbeförderung dienen. Auch wurden Daten von der Analyse ausgeschlossen, bei denen die Datenerfassung teilweise oder vollständig gestört war (erkennbar an fehlenden WLAN- und Bluetooth-Signalen auf einzelnen Linienabschnitten, Kapitel 4.6.1).

Im Ergebnis wurde die Signalrate zusammen mit der Fahrgastzahl differenziert für die einzelnen Linienfahrten der betrachteten Fahrzeuge ermittelt und in Abbildung 4-8 dargestellt. Zusätzlich sind Regressionsgeraden sowie das Bestimmtheitsmaß ergänzt. Der Betrachtungszeitraum der Linienfahrt beginnt bei der Abfahrtszeit an der ersten Haltestelle und endet bei der Ankunftszeit an der letzten Haltestelle der Linienfahrt.

Aus der Abbildung geht hervor, dass das Bestimmtheitsmaß einen Maximalwert von 0,44 (Linie 54) annimmt. Das Bestimmtheitsmaß R^2 entspricht dem Quadrat des Pearson-Korrelationskoeffizienten und ermittelt die Streuung der Daten gegenüber der vorliegenden linearen Regression. Das Bestimmtheitsmaß nimmt Werte zwischen 0 und 1 an, wobei 1 eine perfekte Übereinstimmung der ermittelten Regressionsgeraden mit den Datenpunkten bedeutet.

Entsprechend den Werten für das Bestimmtheitsmaß der einzelnen Linien kann festgehalten werden, dass mit steigender Fahrgastzahl grundsätzlich auch die Anzahl der Signale pro Fahrtminute steigt (positive Korrelation). Gleichzeitig kann daraus sowie aus der Darstellung in Abbildung 4-8 abgeleitet werden, dass große Schwankungen der Signalrate bei verschiedenen Fahrgastzahlen vorliegen. Für Linie 100 etwa liegen für Fahrgastzahlen zwischen 15 und 20 Fahrgästen mittlere Signalaraten zwischen 150 und 500 Signalen pro Fahrtminute vor. Dieser Umstand ist auch bei den weiteren Linien zu beobachten. Wird darüber hinaus die Signalrate auf die Anzahl der Fahrgäste bezogen, zeigt sich, dass die Signalrate pro Fahrgast mit steigender Fahrgastzahl sinkt (negative Korrelation). Unter der vereinfachenden Annahme, dass die Endgeräte der Fahrgäste im Mittel über die Linienfahrten und Linien ein ähnliches Sendeverhalten besitzen, ist insbesondere vor dem

Hintergrund der großen Unterschiede zwischen den Linien der Schluss möglich, dass der überwiegende Teil der empfangenen Signale von Nicht-Fahrgästen stammt (Stördaten).

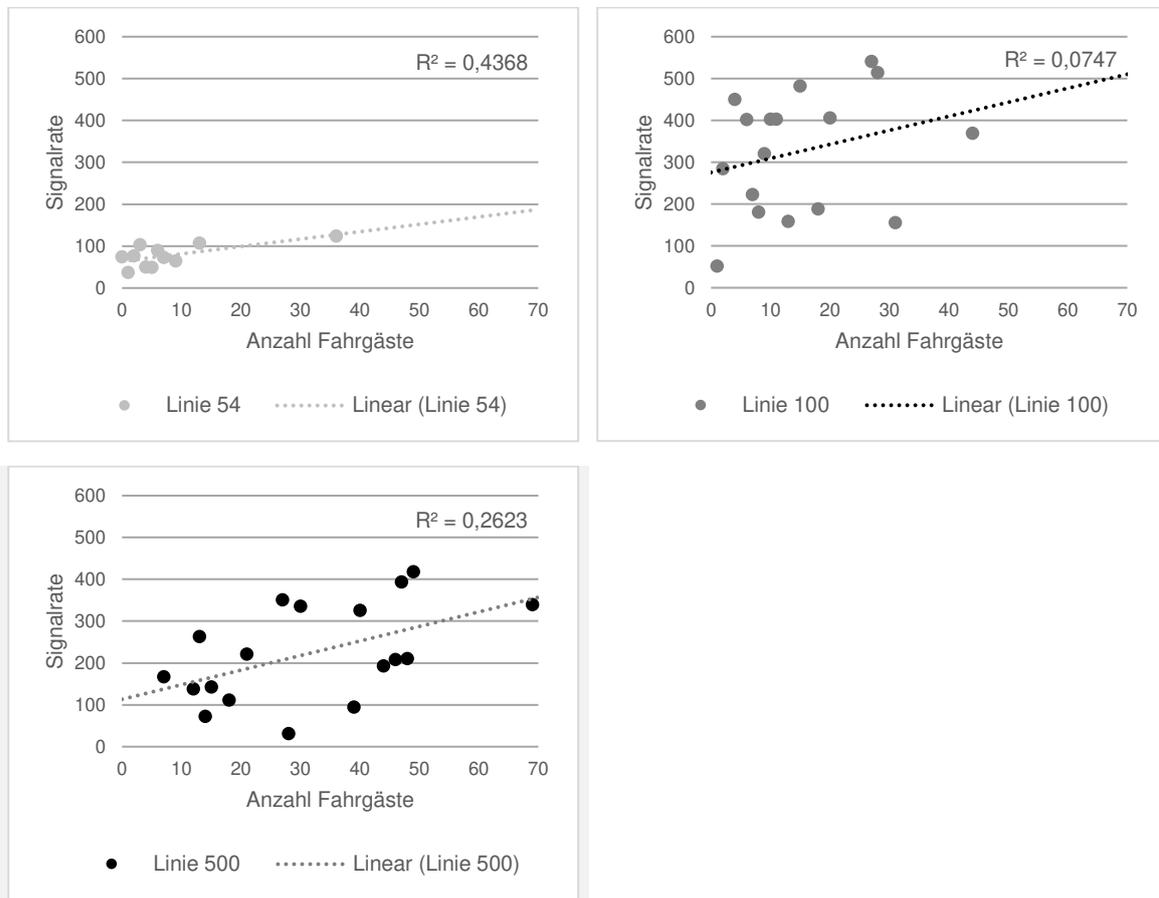


Abbildung 4-8: Gegenüberstellung von Fahrgastzahl und Signalrate (Anzahl WLAN- und Bluetooth-Signale pro Fahrtminute) auf ausgewählten Linienfahrten der Linien 54, 100 und 500

Dies verdeutlicht die Betrachtung der Signalraten der verschiedenen Linien: Die Signalrate der Linie 54 ist deutlich geringer als die Signalrate der Linien 100 oder 500. Grund dafür kann in dem Linienweg der Linie 54 gesehen werden, der hauptsächlich durch ländliche Gebiete außerhalb der Stadt Kassel führt. In ländlichen Gebieten ist im Gegensatz zu städtischen Gebieten mit weniger Störsignalen zu rechnen, da in ländlichen Gebieten die Bevölkerungs- und Arbeitsplatzdichte geringer sind. Die höhere Korrelation bei Linie 500 gegenüber Linie 100 kann ebenso auf den höheren Anteil des Linienwegs bzw. der Haltestellen in ländlichen Räumen hindeuten: Linie 100 besitzt nur drei Haltestellen (von 18) außerhalb des Stadtgebiets Kassels, während Linie 500 außerhalb Kassels 22 Haltestellen (von 40) bedient.

Bei der Interpretation muss berücksichtigt werden, dass die Signale der Fahrgäste nicht separat betrachtet werden können. Die Ergebnisse können aus Sicht des Autors auf Störsignale zurückgeführt werden, die aufgrund zufälliger Situationen (eine oder mehrere Personen befinden sich zufällig in Reichweite des Busses und nutzen ihr Smartphone aktiv) während der Linienfahrt erfasst werden. Darüber hinaus zeigt sich bei Leerfahrten (Anzahl Fahrgäste = 0) eine große Streuung bei der Anzahl erfasster Daten. Daraus folgend ist die Menge an Stördaten grundsätzlich variabel und nicht prognostizierbar. Dabei ist jedoch die kleine Stichprobe von nur zwei Fahrten zu berücksichtigen.

Die These wird entsprechend der Ausführungen als wahr angenommen.

These 2 (Datengrundlage): „Es ist nur eine Teilmenge der Fahrgäste erfassbar: Nur ein Teil der Fahrgäste führt ein WLAN- und/oder Bluetooth-fähiges Endgerät mit sich.“

Die Datengrundlage auf Basis der erfassten WLAN- und Bluetooth-Daten ist unvollständig, da jeder Fahrgast entweder kein, ein oder mehrere WLAN- oder Bluetooth-fähige Endgeräte mitführen kann. Insbesondere bei schwach besetzten Linienfahrten wirken sich Fahrgäste, die nicht erfassbar sind, stark auf den Informationsgehalt der Ausgangsmatrix aus. Der Anteil der fehlenden Daten ist dabei unbekannt. Fehlende Daten wirken sich negativ auf die Qualität des Inputs (Ausgangsmatrix) für das Hochrechnungsverfahren aus. Andererseits ist es bei schwach besetzten Linienfahrten möglich, dass aufgrund der Verteilung der Randsummen ggf. nur eine mögliche Lösung besteht. In dem Fall wären keine Daten in der Ausgangsmatrix notwendig.

Zur Prüfung dieser These wurde eine Frage zum Mitführen von Smartphones in einer zusätzlichen Fahrgastbefragung des Fachgebiets VPVS aufgenommen. Die Auswertung zeigt, dass der Anteil der Befragten, die mindestens ein (eingeschaltetes) Smartphone dabei hatten, im Mittel aller Befragungsfahrten rund 86% beträgt. Die durchschnittliche Anzahl an Smartphones unter den Befragten beträgt hingegen 0,97. Im Vergleich zu dem genannten Prozentsatz machen sich die „Mehrfachbesitzenden“ bemerkbar.

Die These wird entsprechend der Ausführungen als wahr angenommen.

These 3 (Filterung): „Die erfassten Informationen aus den WLAN- und Bluetooth-Daten lassen keine eindeutige Filterung von Fahrgastdaten zu.“

Für die Hochrechnung ist es von hoher Bedeutung, die Quelle-Ziel-Relationen in der Ausgangsmatrix möglichst korrekt zu besetzen, da das Fratar-Verfahren dem Prinzip der Informationsminimierung folgt. Daher ist es zum einen essentiell, Fahrgäste von Nicht-Fahrgästen zu unterscheiden und zum anderen, die Ein- und Ausstiegshaltestelle genau zu erfassen. Der These liegt die Vermutung zugrunde, dass der Großteil der empfangenen Daten sogenannte „Stördaten“ sind. Der Anteil der Daten von Fahrgästen ist somit vergleichsweise gering.

Die Anzahl an Fahrgästen mit WLAN- bzw. Bluetooth-fähigem Endgerät und auch die Anzahl an Endgeräten je Fahrgast sind auf Basis der erfassten WLAN- und Bluetooth-Daten nicht bekannt. Daher kann kein direkter Vergleich der Anzahl gefilterter Endgeräte mit der Anzahl an Endgeräten von Fahrgästen im Fahrzeug erfolgen.

Zur Prüfung dieser These wurde daher hilfsweise die absolute und relative Abweichung der Fahrgastzahlen zwischen dem SOLL-Wert (AFZS-Daten) und dem IST-Wert (Ausgangsmatrix) bestimmt. Da für Bluetooth-Daten keine Signalstärke bestimmt wird (RSSI), werden die entsprechenden Filtersets nicht berücksichtigt. Darüber hinaus werden die Filtersets P90 (keine Filterung) aufgrund der (beabsichtigten) fehlenden Filterwirkung und P60 (RSSI90) aufgrund der sehr geringen Filterwirkung nicht betrachtet. Die genannten Filter überschätzen die Werte der Fahrgastzahlen stark und haben großen Einfluss auf die Ergebnisse des Mittelwerts und der Standardabweichung.

Im Ergebnis wird deutlich, dass die mittlere Anzahl an Fahrgästen in den Ausgangsmatrizen geringer ist als die Anzahl an Fahrgästen nach AFZS. Dieses Ergebnis ist zunächst plausibel, da nicht jeder Fahrgast ein Smartphone bzw. WLAN-/Bluetooth-fähiges Endgerät mitführt. Daher ist davon auszugehen, dass bei einer korrekten Filterung die Anzahl an Fahrgästen auf Basis der erfassten Daten geringer oder gleich der durch das AFZS ermittelten Fahrgastzahl ist.

Eine detaillierte Betrachtung der Abweichung zwischen der Anzahl an Fahrgästen nach AFZS und der Anzahl an Fahrgästen in der Ausgangsmatrix, differenziert für alle betrachteten Referenzfahrten und Filtersets zeigt, dass

die relative Abweichung (gemessen als Quotient der absoluten Abweichung zwischen der Differenz von IST- und SOLL-Wert und dem SOLL-Wert) entsprechend im Mittel für die Datenquellen Bluetooth oder WLAN einen negativen Wert annimmt. Die Fahrgastzahlen werden durch die einzelnen Filtersets auf einigen der betrachteten Referenzfahrten (stark) unterschätzt, während die Fahrgastzahlen beim gleichen Filterset bei anderen Referenzfahrten überschätzt werden. Bei sehr restriktiven Filtern (P30_GPS600_RSSI60) werden die Fahrgastzahlen in der Ausgangsmatrix durchweg unterschätzt. Dies wird durch die Streuung der Werte um den Mittelwert deutlich:

- Bluetooth Mittelwert: -40%
mittl. Streuung (Standardabweichung): 31%
- WLAN Mittelwert: -45%
mittl. Streuung (Standardabweichung): 217%
- WLAN+Bluetooth Mittelwert: 57%
mittl. Streuung (Standardabweichung): 145%

Diese Schwankungen sind vor dem Hintergrund (theoretisch) stark unterschiedlicher Fahrgastgruppen (Schülerverkehr, Studenten, Pendler etc.) unter den betrachteten Linienfahrten durchaus möglich. Um diesen Einfluss zu prüfen, wurden homogene Gruppen nach Tagestyp (Montag bis Freitag sowie Samstag und Sonntag) und Zeitschicht (>00:00 – 06:00, >06:00 – 09:00, >09:00 – 16:00, >16:00 – 19:00, >19:00 – 24:00) gebildet und separat betrachtet. Eine Gruppe besteht aus mindestens drei Referenzfahrten des gleichen Tagestyps und der gleichen Zeitschicht. Aufgrund der geringen Anzahl an Referenzfahrten der Klasse „uneingeschränkt verwertbare Referenzfahrten“ liegen lediglich fünf Gruppen vor, die jeweils mit drei Referenzfahrten besetzt sind.

Die Analyse der homogenen Gruppen bestätigt die Ergebnisse der Auswertung über alle Referenzfahrten: Auch hier zeigen sich innerhalb der Gruppen sowie innerhalb der Filtersets Unterschiede zwischen den Fahrgastzahlen der Ausgangs- und Referenzmatrix.

Die These wird entsprechend der Ausführungen als wahr angenommen: Unter Nutzung der verwendeten Filtersets kann nicht gewährleistet werden, dass die Anzahl der Fahrgäste nach Filterung durchweg kleiner oder gleich der gezählten Anzahl der Fahrgäste ist. Die hohe Streuung der Differenz der berechneten und realen Fahrgastzahl innerhalb der Parameterkombinationen über die betrachteten Referenzfahrten zeigt, dass es nicht möglich ist,

gleichzeitig false negatives als auch false positives zu reduzieren. Im gefilterten Datensatz verbleiben somit mit hoher Wahrscheinlichkeit Stördaten.

Darüber hinaus wurde an den Auswertungen zum Treffermaß deutlich, dass, unabhängig von der Korrektheit der reinen Anzahl an Fahrgästen in Ausgangs- und Referenzmatrix, keine korrekte Zuordnung der Quelle-Ziel-Relationen auf Basis der erfassten Daten möglich ist.

These 4 (Erzeugung von Quelle-Ziel-Matrizen): „Die Zuordnung der erfassten (und gefilterten) Signale zu Ein- und Ausstiegshaltestellen realer Fahrgäste ist ungenau.“

An dieser Stelle soll der Nachweis erfolgen, dass Unschärfen bei der Zuordnung der Ein- und Ausstiegshaltestelle auf Basis der erfassten Daten dazu führen, dass die Informationsgrundlage der Ausgangsmatrix verzerrt ist. Da das Fratar-Verfahren zur Hochrechnung – wie bereits beschrieben – dem Prinzip der Informationsminimierung folgt, kommt das Verfahren erwartungsgemäß zu Ergebnissen mit (zu) geringer Qualität.

Zur Prüfung wurde ein Vergleich der belegten Quelle-Ziel-Relationen der Ausgangsmatrizen auf Basis der erfassten WLAN- und Bluetooth-Daten ohne Filterung (Filterset P90) mit den belegten Quelle-Ziel-Relationen der Referenzmatrizen verglichen. In die betrachteten Ausgangsmatrizen gehen somit ungefiltert alle erfassten Signale ein – sie stellen somit das Maximum der möglichen Übereinstimmung zwischen den betrachteten Matrizen dar. Eine Filterung (etwa durch GPS-Filter) kann nicht zu einer höheren Anzahl übereinstimmender Zellen führen, sondern die Anzahl belegter Relationen lediglich reduzieren.

In Tabelle 4-28 sind die Ergebnisse der Analyse dargestellt. Zunächst wird deutlich, dass bei keiner der 29 betrachteten Linienfahrten die Ausgangsmatrix auf Basis der ungefilterten Eingangsdaten alle Zellen (Quelle-Ziel-Relationen), die in der Referenzmatrix belegt sind, trifft. Der höchste Anteil liegt bei 75% und stellt eine Ausnahme dar. Der Mittelwert liegt mit 29% auf demselben Niveau wie der Median mit 29%. Das 75%-Quantil liegt bei 35%.

Tabelle 4-28: Ergebnisse der Anzahl und des Anteils übereinstimmender Zellen zwischen Ausgangs- und Referenzmatrix (Dataset Bieland/Mobile Data Fusion 2022)

Linie	Datum der Linienfahrt	Startzeit der Linienfahrt	Starthaltestelle	Anzahl übereinstimmender (belegter) Zellen	Anzahl belegter Zellen (Referenzmatrix)	Anteil übereinstimmender (belegter) Zellen
054	28.09.2020	11:58:00	Altenbauna Stadtmitte	2	10	20%
054	28.09.2020	12:49:00	Gudensberg Rathaus	2	8	25%
054	28.09.2020	17:28:00	Altenbauna Stadtmitte	2	9	22%
054	28.09.2020	18:19:00	Gudensberg Rathaus	1	6	17%
100	17.09.2020	09:33:00	Kassel Bhf. Wilhelmshöhe	2	12	17%
100	17.09.2020	10:35:00	Calden Flughafen Terminal	4	13	31%
100	17.09.2020	12:33:00	Kassel Bhf. Wilhelmshöhe	7	21	33%
100	17.09.2020	13:35:00	Calden Flughafen Terminal	7	17	41%
100	19.09.2020	09:33:00	Kassel Bhf. Wilhelmshöhe	0	7	0%
100	19.09.2020	10:35:00	Calden Flughafen Terminal	1	4	25%
100	20.09.2020	19:33:00	Kassel Bhf. Wilhelmshöhe	2	13	15%
100	20.09.2020	20:35:00	Calden Flughafen Terminal	1	3	33%
100	22.09.2020	15:33:00	Kassel Bhf. Wilhelmshöhe	8	19	42%
100	22.09.2020	16:35:00	Calden Flughafen Terminal	9	12	75%
100	27.09.2020	14:33:00	Kassel Bhf. Wilhelmshöhe	1	10	10%
100	09.10.2020	16:35:00	Calden Flughafen Terminal	6	9	67%
100	11.10.2020	14:33:00	Kassel Bhf. Wilhelmshöhe	5	13	38%
100	11.10.2020	15:35:00	Calden Flughafen Terminal	2	15	13%
100	13.10.2020	12:33:00	Kassel Bhf. Wilhelmshöhe	5	15	33%
100	13.10.2020	13:35:00	Calden Flughafen Terminal	5	10	50%
100	13.10.2020	19:33:00	Kassel Bhf. Wilhelmshöhe	4	12	33%
100	13.10.2020	20:35:00	Calden Flughafen Terminal	2	7	29%
100	31.10.2020	19:33:00	Kassel Bhf. Wilhelmshöhe	5	12	42%
100	31.10.2020	20:35:00	Calden Flughafen Terminal	0	12	0%
500	20.09.2020	14:06:00	Kassel Bhf. Wilhelmshöhe	7	22	32%
500	20.09.2020	15:33:00	Bad Wildungen Breiter Hagen	6	24	25%
500	26.09.2020	09:06:00	Kassel Bhf. Wilhelmshöhe	2	26	8%
500	01.11.2020	14:06:00	Kassel Bhf. Wilhelmshöhe	4	21	19%
500	01.11.2020	18:06:00	Kassel Bhf. Wilhelmshöhe	9	26	35%

Insgesamt lässt sich anhand dieser Auswertung festhalten, dass die Qualität der Eingangsdaten gering ist. Bei der Interpretation ist zu berücksichtigen, dass bei einer Übereinstimmung zwischen Ausgangs- und Referenzmatrix weiterhin unklar ist, ob die Daten der betrachteten MAC-Adresse tatsächlich von einem Fahrgast stammen, der auf dieser Relation unterwegs war.

Aufgrund der Erkenntnisse zum Sendeverhalten besteht die Möglichkeit, dass die korrekt zugeordnete Relation auf Daten eines anderen Fahrgast-Endgeräts basieren.

Darüber hinaus sind eine Vielzahl an Zellen in der Ausgangsmatrix falsch befüllt, d. h. dass diese Zellen nicht in der Referenzmatrix befüllt sind. Dieser Umstand geht nicht in die Analyse ein und wird nicht in der Tabelle dargestellt. Die große Anzahl falsch besetzter Relationen wirkt sich dabei negativ auf die Verfahrensergebnisse aus, ähnlich wie bei der Verwendung einer großen Anzahl an Hilfswerten.

Ein geringer Anteil an übereinstimmenden Zellen kann im weiteren Verfahrensablauf zu einer vergleichsweise hohen Anzahl an benötigten Hilfswerten führen (Kapitel 3.4.4) damit das Verfahren unter Zuhilfenahme der Steigerungsfaktoren zu einer stabilen Lösung hin konvergieren kann. Mit dem Einsatz von Hilfswerten geht vor dem Hintergrund des Prinzips der Informationsminimierung das Problem einher, dass sie keine Information zu der realen Verteilung der Fahrgäste auf die Quelle-Ziel-Relationen enthalten und daher nicht zu einer Steigerung der Datenqualität beitragen können. Aufgrund der geringen Güte der Daten in der Ausgangsmatrix gegenüber der Referenzmatrix ist es daher unwahrscheinlich, dass die ermittelte Lösung des Steigerungsfaktorenmodells korrekt ist (sofern mehrere Lösungen möglich sind).

Die These wird entsprechend der Ausführungen als wahr angenommen.

These 5 (Fuzzifizierung): „Bei der Ableitung der Häufigkeitsverteilung zwischen erfassten Bluetooth- und WLAN-Signalen kommt es zu einer Überschätzung der Wahrscheinlichkeiten „kleiner“ Zeitintervalle.“

Die Häufigkeitsverteilung der Erfassung von WLAN- und Bluetooth-Signalen basiert auf erfassten Daten in einem Fahrzeug über einem Zeitraum von etwa einem Monat. In die Häufigkeitsverteilung sind dabei alle Geräte eingegangen, von denen mehr als ein Signal erfasst wurde. Dabei ist zu berücksichtigen, dass hier Auswahleffekte bei der Erfassung der Daten vorliegen: selten sendende Geräte, die sich außerhalb des Busses befinden, gehen nicht in die Auswertung ein, da von diesen Geräten maximal ein Signal bei der Vorbeifahrt eines Busses erfasst werden kann. Anders verhält es sich bei Geräten außerhalb des Busses, die in kurzen Zeitabständen Signale aussenden. Diese können auch bei einer Vorbeifahrt mehrfach erfasst werden und gehen somit in die Häufigkeitsverteilung – die mindestens zwei Signale zur Ermittlung des zeitlichen Abstands zwischen Signalen benötigt – ein. Durch diese

Geräte kann insbesondere der Anteil an geringen Zeitabständen zwischen zwei Signalen erhöht werden. Dieser Effekt könnte theoretisch durch eine gezielte Filterung auf die reinen Fahrgastdaten entfernt werden. Diese Filterung ist jedoch nicht möglich (s. These 3).

Die These kann aufgrund des Fehlens belastbarer Daten zur Prüfung weder widerlegt noch angenommen werden. Aufgrund der Ausführungen liegt jedoch eine Tendenz vor, die These als wahr anzunehmen.

These 6 (Hochrechnung): „Fehlende oder verzerrte Informationen zu den Quelle-Ziel-Beziehungen der Fahrgäste in der Ausgangsmatrix können durch das Hochrechnungsverfahren nicht ausgeglichen werden und führen somit zu Verfahrensergebnissen mit geringer Güte.“

Wie die vorangegangenen Thesen gezeigt haben, bestehen Probleme bei der Erfassung der Datengrundlage, der Filterung auf Fahrgastdaten sowie der korrekten Zuordnung von erfassten Daten der Endgeräte zu den Quelle-Ziel-Relationen der Fahrgäste. An dieser Stelle soll die Wirkung des Hochrechnungsverfahrens (Fratar-Verfahren) geprüft werden:

Bei Eingang korrekter und unverzerrter Daten in der Ausgangsmatrix kommt das Hochrechnungsverfahren zu Ergebnissen mit einer hohen Güte. Mit zunehmendem Anteil nicht oder falsch besetzter Quelle-Ziel-Relationen in der Ausgangsmatrix gegenüber der Referenzmatrix wird die Informationsgrundlage für das Hochrechnungsverfahren jedoch verzerrt – die Güte der Verfahrensergebnisse sinkt (WVI Prof. Dr. Wermuth Verkehrsforschung und Infrastrukturplanung GmbH 2022). In diesen Fällen ist die vom Fratar-Verfahren bestimmte Lösung zwar die wahrscheinlichste Lösung, diese wird in der Regel jedoch nicht mit der Realität (Referenzmatrix) übereinstimmen (bei Vorliegen mehrerer möglicher Lösungen).

Die These wird entsprechend der Ausführungen als wahr angenommen.

These 7 (Datengrundlage): „Der Anteil von Smartphones mit aktiver Randomisierung nimmt zu.“

Die Einführung der Randomisierung führt dazu, dass ein steigender Anteil an Endgeräten nicht (ohne weiteres) räumlich und zeitlich nachverfolgbar ist. Dieser Prozess wird durch die softwareseitige Standardisierung der Randomisierung durch die OS-Hersteller verstärkt. Folglich wird die Datengrundlage nicht-randomisierter Daten für das Verfahren reduziert.

Die Entwicklung der Verbreitung von Android-Betriebssystemen zeigt, dass aktuelle Versionen innerhalb von etwa einem Jahr nach dem Release den höchsten Anteil an allen Android Geräten (weltweit) besitzen (Statista 2022b). Im Juli 2021 hatte Android 11 die höchste Verbreitung (Statista 2022a). Die MAC-Randomisierung ist seit Android Version 6 möglich, ab Android Version 10 ist sie standardmäßig aktiviert.

Die Verbreitung neuer iOS-Versionen erfolgt zügiger – innerhalb von drei Monaten wurde Version 14 auf über 70% der aktiv genutzten iPhones genutzt (Floemer 2020).

Diese Diskrepanz in der Verbreitungsgeschwindigkeit ist darauf zurückzuführen, dass bei Android Geräten die Updates über die verschiedenen Hersteller der Smartphones (Samsung, Xiaomi, Huawei etc.) bereitgestellt werden müssen, wodurch sich teils erhebliche Verzögerungen nach einem Release ergeben. Bei iOS steht nur Apple als Hersteller, sodass es hier einfacher und schneller geht.

Abbildung 4-9 zeigt die Entwicklung des Anteils randomisierter MAC-Adressen an allen erfassten WLAN-Signalen über einen Zeitraum von rund zwei Jahren. Die Auswertung basiert auf den erfassten Daten eines Fahrzeugs der Linie 500 an dem jeweiligen Stichtag. Zur Darstellung der Entwicklung wurden Stichtage in einem Intervall von etwa zwölf Wochen ausgewertet. Die Stichtage liegen jeweils an einem Werktag (Montag bis Donnerstag).

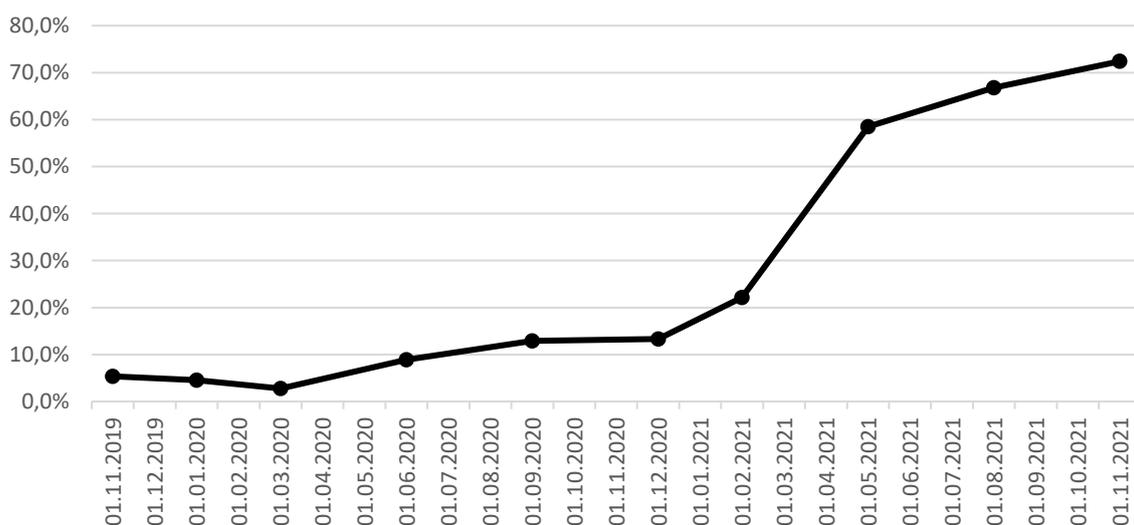


Abbildung 4-9: Entwicklung des Anteils randomisierter MAC-Adressen an allen erfassten WLAN-Signalen im Zeitraum vom 21.11.2019 bis 04.11.2021

Aus der Abbildung geht hervor, dass der Anteil randomisierter MAC-Adressen an allen erfassten Signalen seit Beginn der Datenerfassung zunimmt. Die

Entwicklung ähnelt einer Diffusionskurve, die die Ausbreitung einer Innovation von der Entwicklung bis zur Marktdurchdringung wiedergibt. Insbesondere seit Dezember 2020 ist der Anteil an randomisierten MAC-Adressen unter allen erfassten WLAN-Signalen durch den Hochlauf von Android 10 und 11 erwartungsgemäß stark gestiegen (MAC-Randomisierung ist seit Android 10 standardmäßig aktiviert). Vor dem Hintergrund der Rechercheergebnisse ist davon auszugehen, dass diese Entwicklung in ähnlicher Form bei allen Linien und Fahrzeugen zu beobachten ist.

Während der Fahrgasterhebung (11/2020) lag der Anteil an randomisierten MAC-Adressen unter allen erfassten WLAN-Signalen des Fahrzeugs bei unter 20%. Es ist jedoch anzunehmen, dass der Anteil der MAC-Adressen mit Randomisierung von Fahrgästen höher ist, da Störsignale aus den umliegenden Wohnungen tendenziell aufgrund der Verbindung zu einem Access Point feste MAC-Adressen besitzen. Mit der Einführung von Android 11 wird es den Nutzern ermöglicht, die MAC-Adresse auch bei bestehender Verbindung zu einem Access Point zu randomisieren, sodass hier nicht mehr die wahre MAC-Adresse angezeigt wird (Wallen 2021). Somit ist davon auszugehen, dass sich der Anteil an randomisierten MAC-Adressen unter den Störsignalen zunehmend erhöht. Ob diese Einstellung standardmäßig durch die OS-Betreiber eingeschaltet wird, ist unklar.

Die These wird entsprechend der Ausführungen als wahr angenommen.

5 Weiterentwicklung des Verfahrens (räumliche Aggregation)

5.1 Einführung

In diesem Kapitel wird exemplarisch eine Form der Weiterentwicklung des Verfahrens zur Ermittlung von Quelle-Ziel-Relationen auf Basis der bestehenden Input-Daten untersucht. In Kapitel 3.3.2.2 wurden bereits Möglichkeiten zur Präzisierung des Ein- und Ausstiegsorts von Fahrgästen auf Basis von Haltestellendaten und weiteren Fahrzeugdaten diskutiert.

Aus den bisherigen Ausführungen geht hervor, dass es insbesondere bei der Zuordnung der Ein- und Ausstiegshaltestelle der Fahrgäste auf Basis der erfassten Signale zu Unschärfen kommt. Um diese Unschärfen zu reduzieren, soll in diesem Kapitel eine räumliche Aggregation von Haltestellen der bestehenden Linien untersucht werden. Dazu wird eine exemplarische Prüfung durchgeführt, bei der die räumliche Aggregation auf Basis von Stadtteilgrenzen in Kassel sowie Ortsgrenzen für Gebiete außerhalb der Stadt Kassel erfolgt. Mit der räumlichen Aggregation kommt es zu einer Reduzierung der Informationen in den Quelle-Ziel-Relationen. Die Anzahl möglicher Relationen nimmt ab, die Ein- und Ausstiegshaltestellen werden entsprechend auf Stadtteil bzw. Ortsebene ermittelt. Aufgrund dieser Anpassung bei der Genauigkeit des Ein- und Ausstiegsorts soll geprüft werden, ob die Verfahrensqualität gesteigert werden kann.

5.2 Vorgehen räumliche Aggregation

Die exemplarische Prüfung der räumlichen Aggregation wird für zehn Referenzfahrten (Tabelle 5-1) durchgeführt:

- fünf Referenzfahrten der Linie 100 und
- fünf Referenzfahrten der Linie 500.

Bei der Auswahl der Referenzfahrten wurden für Linie 500 alle vorliegenden Referenzfahrten berücksichtigt. Für Linie 100 wurden sowohl Fahrten mit sehr guten als auch Fahrten mit sehr schlechten Ergebnissen entsprechend der GEH-Grenzwerte berücksichtigt.

Tabelle 5-1: Ausgewählte Linienfahrten für die räumliche Aggregation (Linie 100, Linie 500)

Linie	Datum der Fahrt	Abfahrtszeit	Abfahrtshaltestelle
100	11.10.2020	14:33:00	Kassel, Bahnhof Wilhelmshöhe
100	13.10.2020	12:33:00	Kassel, Bahnhof Wilhelmshöhe
100	13.10.2020	19:33:00	Kassel, Bahnhof Wilhelmshöhe
100	17.09.2020	12:33:00	Kassel, Bahnhof Wilhelmshöhe
100	20.09.2020	20:35:00	Calden, Flughafen KS Terminal
500	01.11.2020	14:06:00	Kassel, Bahnhof Wilhelmshöhe
500	01.11.2020	18:06:00	Kassel, Bahnhof Wilhelmshöhe
500	20.09.2020	14:06:00	Kassel, Bahnhof Wilhelmshöhe
500	20.09.2020	15:33:00	Bad Wildungen, Breiter Hagen
500	26.09.2020	09:06:00	Kassel, Bahnhof Wilhelmshöhe

Eine Aggregation der Haltestellen auf Basis der Verkehrszellen des Verkehrsmodells der Stadt Kassel war nicht möglich, da jeder Verkehrszelle im Stadtgebiet maximal eine Haltestelle der betrachteten Linien zugeordnet werden konnte (Prüfung mit Geometrie-Werkzeugen in QGIS 3.16). Daher wurde alternativ eine Aggregation auf Basis der Stadtteile Kassels vorgenommen. Dazu wurden die Haltestellen aggregiert, die innerhalb eines Stadtteils Kassels (Bad Wilhelmshöhe, Vorderer Westen, Mitte etc.) liegen. Außerhalb des Stadtgebiets von Kassel wurden die Haltestellen auf Gemeindeebene, zum Beispiel Bad Wildungen, Fritzlar, Gudensberg oder Calden zusammengefasst.

Durch die räumliche Aggregation reduzieren sich die Anzahl an Haltestellen

- auf Linie 500 von 40 auf 7 - diese starke Aggregation ist auf die große Anzahl an Haltestellen außerhalb Kassels zurückzuführen (22 Haltestellen außerhalb Kassels wurden zu 3 Haltestelle aggregiert)
- auf Linie 100 von 18 auf 6 Haltestellen.

Eine Kartendarstellung der aggregierten Haltestellen ist in Anhang 5 beigefügt.

Die exemplarische Prüfung der räumlichen Aggregation wurde dabei für die ausgewählten Linienfahrten für je drei Filtersets durchgeführt:

- P50_GPS100
- P51_GPS200 mit Fuzzifizierung
- P99_FullFilter als Vergleichsfall

Die Filtersets wurden so gewählt, dass je ein Filterset mit und ohne Fuzzifizierung betrachtet wird sowie ein Vergleichsfall. Für die Filtersets wurde dabei auf die relativ besten Filtersets zurückgegriffen (Kapitel 4.6.4). Als Datenquelle werden WLAN- und Bluetooth-Daten genutzt. Die Hochrechnung wird entsprechend der Ergebnisse und der geringen Güte bei Anwendung des Verfahrens nach Li und Cassidy ausschließlich mit dem Fratar-Verfahren durchgeführt.

Mit der räumlichen Aggregation geht eine weitreichende Datenaufbereitung (u. a. Erstellung neuer Zählfahrten, Zielmatrizen, Startmatrizen) einher, damit die Daten in das bestehende Datenmodell der MySQL-Datenbank integriert werden können. Für die räumliche Aggregation wurden neue Kodierungen der Haltestellen (Stop_Codes) sowie eine neue Indexierung der Haltestellen in Abhängigkeit der Fahrtrichtung und des Linienwegs (Ausgangsmatrix, Referenzmatrix) berücksichtigt. Haltestellenabstände wurden nicht ermittelt, sodass keine Auswertungen nach Entfernungsklassen oder hinsichtlich der Verkehrsleistung möglich sind.

5.3 Ergebnisse räumliche Aggregation

Für die Ermittlung der Ergebnisse wurde das HR-Tools (Version 1.10.1²¹) genutzt. Die Erstellung der Startmatrizen erfolgte mit StauFi-Tool Version vom 13.06.2022 (Software Mobile Data Fusion 2022). Die Ergebnisse sind der Datenbank (Dataset Bieland/Mobile Data Fusion 2022) zu entnehmen.

Da sich im Zuge der Aggregation die Anzahl an Relationen bei gleichbleibender Zahl an Fahrgästen verringert, nehmen sowohl die Fahrgäste je Relation als auch die Randsummen höhere Werte an. Um eine Vergleichbarkeit der Ergebnisse auf Basis des nicht selbstskalierenden GEH-Werts zu erhalten, ist daher eine Anpassung der GEH-Grenzwerte aufgrund der generell höheren Werte für die einzelnen Quelle-Ziel-Relationen notwendig.

In Kapitel 4.4.2 wurde das Vorgehen zur Ermittlung der Grenzwerte des GEH-Werts beschrieben. Dabei wurde der Grenzwert auf Basis der Anzahl an realen Fahrgästen je Relation (Befragungsdaten) ermittelt, der bei 3% der von den Fahrgästen genutzten Relationen (Werte der betrachteten Relationen >0) überschritten wird. Während der Wert bei den nicht aggregierten Fahrten fünf Fahrgäste über alle Referenzfahrten beträgt, liegt für die in diesem

²¹ Das HR-Tool berücksichtigt bei der Ermittlung der Anzahl an Relationen für die Speicherung in der Datenbank nicht die Diagonale (Binnenverkehr, der bei der Aggregation erzeugt wird).

Kapitel betrachteten zehn Referenzfahrten der Wert bei acht Fahrgästen. Das heißt, dass durch die Aggregation der Anteil an Quelle-Ziel-Relationen mit Werten größer als acht Fahrgäste nur bei 3% der Relationen mit Fahrgästen überschritten wird.

Analog zum Vorgehen zur Ermittlung der GEH-Grenzwerte werden die GEH-Grenzwerte entsprechend der gewährten tolerierten relativen Abweichung angepasst. Die bestehenden GEH-Grenzwerte besitzen eine Toleranz von 34% (GEH-Grenzwert 0,7) bzw. 79% (GEH-Grenzwert 1,5). Wird die tolerierte relative Abweichung auf die erhöhten Fahrgastzahlen auf Basis der Aggregation und Auswahl der Referenzfahrten übertragen, ergeben sich folgende GEH-Grenzwerte:

- 0,9 einzuhalten bei 95% (vorher 0,7)
- 1,9 einzuhalten bei 100% der Relationen (vorher 1,5)

Tabelle 5-2 und Tabelle 5-3 zeigen jeweils eine Ergebnisübersicht für die zehn ausgewählten Referenzfahrten sowohl im originären Zustand der WLAN- und Bluetooth-Daten (ohne Aggregation) als auch die Ergebnisse nach der Aggregation. Dazu werden folgende Gütemaße differenziert für die verwendeten Filtersets (P50_GPS100, P51_GPS200_Fuzzy und P99_Full-Filter) betrachtet:

- GEH-1,5-Grenzwert (ohne Aggregation) / GEH-1,9-Grenzwert (mit Aggregation): markiert werden die Referenzfahrten und Filtersets, die bei einem der acht verwendeten HR-Sets (Fratar-Verfahren), den jeweiligen GEH-Grenzwert vollständig einhalten.
- GEH-0,7-Grenzwert (ohne Aggregation) / GEH-0,9-Grenzwert (mit Aggregation): markiert werden die Referenzfahrten und Filtersets, die bei einem der acht verwendeten HR-Sets (Fratar-Verfahren), den jeweiligen GEH-Grenzwert bei 95% der betrachteten Relationen einhalten.

Die weiteren Gütemaße (Treffermaß, Überfüllungsfaktor, Fehlerquadratsumme) werden qualitativ im Rahmen der Bewertung der Ergebnisse (Kapitel 5.4) betrachtet.

Tabelle 5-2: Ergebnisübersicht ausgewählter Referenzfahrten zur räumlichen Aggregation (ohne Aggregation)

Li- nie	Datum, Abfahrtszeit, Abfahrtshaltestelle	GEH 1,5			GE0,7		
		P50	P52	P99	P50	P52	P99
100	11.10.2020_14:33:00_Bahnhof Wilhelmshöhe			x			
100	13.10.2020_12:33:00_Bahnhof Wilhelmshöhe	x	x	x			
100	13.10.2020_19:33:00_Bahnhof Wilhelmshöhe	x	x	x			
100	17.09.2020_12:33:00_Bahnhof Wilhelmshöhe	x		x			
100	20.09.2020_20:35:00_Flughafen KS Terminal	x	x	x	x	x	x
500	01.11.2020_14:06:00_Bahnhof Wilhelmshöhe	x			x	x	
500	01.11.2020_18:06:00_Bahnhof Wilhelmshöhe				x	x	x
500	20.09.2020_14:06:00_Bahnhof Wilhelmshöhe	x			x	x	x
500	20.09.2020_15:33:00_Breiter Hagen					x	
500	26.09.2020_09:06:00_Bahnhof Wilhelmshöhe						

Aus der Tabelle geht hervor, dass der GEH-1,5-Grenzwert ohne räumliche Aggregation in insgesamt 14 Fällen (Kombination aus Referenzfahrt und Filterset) eingehalten wird. Bei drei Referenzfahrten der Linie 500 wird der Grenzwert durch keins der betrachteten Filtersets eingehalten. Rückt der GEH-0,7-Grenzwert in den Fokus zeigt sich, dass dieser bei fünf Referenzfahrten durch kein Filterset eingehalten wird – vier der Linie 100 und eine der Linie 500.

Tabelle 5-3: Ergebnisübersicht ausgewählter Referenzfahrten zur räumlichen Aggregation (mit Aggregation)

Li- nie	Datum, Abfahrtszeit, Abfahrtshaltestelle	GEH 1,9			GE0,9		
		P50	P52	P99	P50	P52	P99
100	11.10.2020_14:33:00_Bahnhof Wilhelmshöhe		x	x			x
100	13.10.2020_12:33:00_Bahnhof Wilhelmshöhe	x	x				
100	13.10.2020_19:33:00_Bahnhof Wilhelmshöhe	x	x	x			
100	17.09.2020_12:33:00_Bahnhof Wilhelmshöhe			x			
100	20.09.2020_20:35:00_Flughafen KS Terminal	x	x	x	x	x	x
500	01.11.2020_14:06:00_Bahnhof Wilhelmshöhe			x			
500	01.11.2020_18:06:00_Bahnhof Wilhelmshöhe						
500	20.09.2020_14:06:00_Bahnhof Wilhelmshöhe						
500	20.09.2020_15:33:00_Breiter Hagen		x	x			
500	26.09.2020_09:06:00_Bahnhof Wilhelmshöhe			x			

Bei der Alternative mit räumlicher Aggregation wird der angepasste GEH-1,9-Grenzwert in insgesamt 15 Fällen (Kombination aus Referenzfahrt und Filterset) eingehalten. Bei zwei Referenzfahrten der Linie 500 wird der Grenzwert durch kein Filterset eingehalten. Die Betrachtung des GEH-0,9-Grenzwerts

zeigt, dass dieser nur bei vier Referenzfahrten – ausschließlich der Linie 100 – eingehalten wird. Eine nach Filterset differenzierte Betrachtung zeigt, dass die Aggregation beim

- Filterset P50 ohne Fuzzifizierung zu schlechteren Ergebnissen führt,
- Filterset P52 mit Fuzzifizierung je nach Grenzwert zu besseren (GEH-1,9 bzw. 1,5) oder schlechteren (GEH-0,9 bzw. 0,7) Ergebnissen führt,
- Filterset P99 ebenfalls je nach Grenzwert zu besseren (GEH-1,9 bzw. 1,5) oder schlechteren (GEH-0,9 bzw. 0,7) Ergebnissen führt.

Die Ergebnisse zum Trefferanteil, Überfüllungsfaktor und Fehlerquadratsumme sind nicht dargestellt. Die Auswertungen zeigen jedoch, dass die Gütemaße erwartungsgemäß aufgrund der geringen Anzahl an Wertepaaren zu besseren Ergebnissen führen als im Fall ohne Aggregation.

5.4 Bewertung der Ergebnisse

Ein Vergleich der Ergebnisse zeigt, dass mit der räumlichen Aggregation die Anzahl Referenzfahrten und Filtersets, die den GEH-1,5- bzw. GEH-1,9-Grenzwert einhalten, um eine Referenzfahrt zunimmt, während die Anzahl an Fahrten, die den GEH-0,7- bzw. GEH-0,9-Grenzwert einhalten, um acht Fahrten abnimmt. Darüber hinaus zeigt sich, dass mit der räumlichen Aggregation

- der Trefferanteil höher ist,
- der Überfüllungsfaktor geringere Werte annimmt und
- auch die Fehlerquadratsumme geringere Werte annimmt.

Diese Erkenntnisse sind plausibel. Mit der räumlichen Aggregation kommt es dazu, dass die Anzahl möglicher Relationen der betrachteten Referenzfahrten stark reduziert wird, während die Belegung der Quelle-Ziel-Relationen höhere Werte annimmt. Dies hat direkte Auswirkungen auf die Gütemaße:

- GEH-1,5-Grenzwert / GEH-1,9-Grenzwert (einzuhalten bei 100% der Relationen): Mit der Aggregation kommt es dazu, dass sowohl die einzelnen Relationen mit höheren Werten belegt werden und entsprechend auch die Randsummen höhere Werte annehmen. Gleichzeitig wird die Anzahl an möglichen Relationen zum Ausgleich auf die Randsummen reduziert, die bei der Iteration berücksichtigt werden. Somit können auch eingesetzte Hilfswerte häufiger Werte größer 1 annehmen. Dies wird insbesondere dann problematisch, wenn auf der realen Relation der Wert

0 vorliegt, da hier die GEH-Grenzwerte (1,5 sowie 1,9) bereits bei einem Wert von 2 überschritten werden (Tabelle 4-7 in Kapitel 4.4.2).

- GEH-0,7-Grenzwert / GEH-0,9-Grenzwert (einzuhalten bei 95% der Relationen): In Kapitel 4.4.4 wurde erläutert, dass aufgrund der großen Anzahl an Relationen ohne Aggregation und der Vorgabe zur Einhaltung des Grenzwerts bei 95% der Relationen bei einem Teil der betrachteten Fahrten automatisch eingehalten wird. Dieser Fall tritt bei den aggregierten Daten nicht mehr auf, da die Anzahl möglicher Relationen stark reduziert wird: Die Anzahl möglicher Relationen beträgt 15 bzw. 21 Relationen. Eine Abweichung oberhalb des Grenzwerts ist bei 5% der Relationen erlaubt (entspricht bei 21 Relationen etwa 1,05 Relationen). Somit darf der Grenzwert beim Vergleich aller Wertepaare lediglich einmal abweichen. Es ist abzuwägen, ob der Grenzwert entsprechend angepasst werden sollte, da die 95% unter Berücksichtigung der Anzahl Relationen in den betrachteten Referenzfahrten gewählt wurden. Eine Anpassung ist nicht erfolgt.
- Treffermaß: Der Anteil an Treffern hat sich mit der Aggregation im Mittel erhöht. Dies ist auf folgenden Zusammenhang zurückzuführen: Je geringer die Anzahl möglicher Relationen, desto höher ist bei zufälliger Belegung einer der möglichen Relationen die Wahrscheinlichkeit für eine korrekte Belegung. Dies kann durch folgendes Beispiel veranschaulicht werden. Es liegen 20 mögliche Relationen vor, von denen nur eine in der Referenzmatrix befüllt ist. Die Wahrscheinlichkeit einer zufälligen korrekten Befüllung liegt somit bei $1/20$ (5%). Liegen dagegen 200 mögliche Relationen vor, reduziert sich die Wahrscheinlichkeit auf $1/200$ (0,5%). Die Erhöhung des mittleren Trefferanteils infolge der räumlichen Aggregation ist somit plausibel.
- Überfüllungsfaktor: Der Überfüllungsfaktor nimmt ab, weil zwei Effekte zusammenkommen. Zum einen stehen weniger mögliche Relationen zur Verfügung. Gleichzeitig ist der Anteil belegter Relationen für die ausgewählten Referenzfahrten und Filtersets gestiegen. Zur Verdeutlichung dient folgendes Beispiel: Die Anzahl möglicher Relationen beträgt 15 bzw. 21 Relationen. Sind in der Referenzmatrix 11 der Relationen belegt, kann der Überfüllungsfaktor keinen Wert größer 2 annehmen.
- Fehlerquadratsumme: Die geringere Fehlerquadratsumme im Fall mit Aggregation kann auf die geringere Anzahl an belegten Relationen

zurückgeführt werden sowie auf geringere Differenzen bei den Wertepaarvergleichen. Die höchsten Differenzen zwischen dem Fall mit und ohne Aggregation ergeben sich erwartungsgemäß (aufgrund der quadratischen Form) bei Anwendung des Reduktionsgrenzwerts von 1,0. Auch bei Anwendung der Fuzzifizierung steigt die Differenz, da insbesondere im Fall ohne Aggregation so eine Vielzahl an Relationen belegt werden.

Auf Basis der Ergebnisse ist an dieser Stelle festzuhalten, dass die räumliche Aggregation in der beschriebenen Form (Aggregation auf Stadtteil- bzw. Ortsebene) nicht zu einer generellen Verbesserung der Verfahrensgüte über alle Referenzfahrten und (betrachteten) Filtersets führt. Gleichzeitig ist anzumerken, dass durch die Aggregation die Einhaltung des GEH-1,9-Grenzwerts für zwei Fahrten der Linie 500 ermöglicht wurde, die im Fall ohne Aggregation durch keine Parameterkombination eingehalten werden konnte (GEH-1,5-Grenzwert).

Bei der Interpretation ist zu berücksichtigen, dass es sich an dieser Stelle um eine exemplarische Prüfung zur räumlichen Aggregation mit geringem Stichprobenumfang handelt. Die Ergebnisse gelten somit lediglich für die ausgewählten Linienfahrten und Parameterkombinationen. Darüber hinaus wurde lediglich eine Form der Aggregation geprüft. Weiterhin sind die bereits diskutierten Schwächen der Datengrundlage und des Verfahrens auch bei der räumlichen Aggregation vorhanden – für die Erzeugung der Quelle-Ziel-Matrizen aufgrund der Aggregation jedoch in abgeschwächter Form.

Die Treffermaß und der Überfüllungsfaktor behalten ihre Gültigkeit. Für die Fehlerquadratsumme bietet sich eine Normierung über die Anzahl an möglichen bzw. belegten Relationen an (nicht betrachtet). Die Differenz der Verkehrsleistung wurde, wie oben beschrieben, nicht bestimmt (Fehlen von Haltestellenabständen).

6 Zusammenfassung, Fazit und weiterer Forschungsbedarf

Mit diesem abschließenden Kapitel werden die Problemstellung, der methodische Ansatz des Verfahrens und die zentralen Ergebnisse der Verfahrensanwendung zusammengefasst. Dazu werden die Erkenntnisse in Form von Schlussfolgerungen dargestellt. Die vorliegende Dissertation wird durch die Darstellung eines Ausblicks sowie des weiteren Forschungsbedarfs abgeschlossen.

6.1 Zusammenfassung

Einführung und Problemstellung

Demografische und sozioökonomische Entwicklungen sowie Veränderungen im Mobilitätsverhalten führen zu weniger Zwangskunden im ÖPNV. Gleichzeitig steigt das Kundenpotential, das sich situationsabhängig für ein bestimmtes Verkehrsmittel entscheidet (multimodales Verhalten). Darüber hinaus werden im Zuge der Verkehrswende tarifliche Maßnahmen wie Nulltarif oder Bürgertickets diskutiert, die zu starken Fahrpreissenkungen für die Kunden des öffentlichen Verkehrs (ÖV) führen und somit das Potential zu einer hohen und kurzfristigen Steigerung der Fahrgastnachfrage besitzen.

Aktuelle und zuverlässige Informationen zur Nachfrage erhalten damit einen Bedeutungsgewinn: Um die bestehenden Potentiale zu nutzen und auf ggf. kurzfristige Nachfragesteigerungen reagieren zu können, sind genaue und aktuelle Daten zum Mobilitätsverhalten und zur Fahrgastnachfrage notwendig. Insbesondere die Angebotsplanung braucht eine verbesserte Datengrundlage, um der rasch verändernden Entwicklung der Nachfrage künftig gerecht werden zu können. Diese Daten stammen derzeit aus Nachfragemodellen, deren Datengrundlage durch personalintensive und damit aufwändige Erhebungen erfasst wird.

Gleichzeitig steigt der Smartphone-Besitz in Deutschland seit Jahren an. Vor dem Hintergrund, dass die meisten Smartphones regelmäßig Signale ausstrahlen, um WLAN-Netze (Wireless Local Area Network) oder Bluetooth-Geräte zu finden, fallen bei eingeschaltetem WLAN und/oder Bluetooth in den in Reichweite gelegenen WLAN- und Bluetooth-Routern entlang des realisierten Weges Registrierungs- und ggf. Nutzungsdaten an. Diese Daten entstehen ohne das Zutun des Nutzers und können in Verknüpfung mit anderen Datenquellen (AFZS) ein Gesamtbild der Verkehrsnachfrage im ÖPNV-Netz

aufbauen, das für ÖPNV-Planung und Marketing verwendet werden kann. Diese stets aktuelle Datengrundlage kann insbesondere in den Bereichen der (Angebots-)Planung, Einnahmenaufteilung sowie Tarifkonzeption einen Mehrwert bieten. An dieser Stelle setzt die vorliegende Dissertation an, mit dem Ziel, diese Daten zunächst auf Linienfahrbene für die Planungs- und Marketingprozesse der Verkehrsunternehmen und -verbände bereitstellen zu können.

Ziel der Arbeit

Das zentrale Ziel dieser Dissertation war daher, ein Verfahren zu entwickeln, das basierend auf

- AFZS-Daten, die in vielen Verkehrsunternehmen und –verbänden vorliegen, sowie
- automatisch erfassbaren Daten unterschiedlicher Quellen (WLAN- und Bluetooth-Daten)

die Fahrgastnachfrage in Form von linienspezifischen Quelle-Ziel-Matrizen automatisch, mit ausreichender Genauigkeit und für lange Erhebungszeiträume ermittelt. Dabei wurden die datenschutzrechtlichen Belange bereits bei der Erfassung und Aufbereitung der Daten berücksichtigt.

Linien- und tagesscharfe Quelle-Ziel-Matrizen ermöglichen ÖPNV-Betreibern nach Wochentagen und Tageszeiten differenzierte Auswertungen, deren Ergebnisse in den Planungs- und Betriebsprozess eingebracht werden können. Gleichzeitig ist die benötigte Infrastrukturausrüstung vergleichsweise kostengünstig.

Status Quo der Erfassung von Nachfragedaten

Die derzeitige Erfassung von Nachfragedaten im ÖPNV baut insbesondere auf Fahrgastzählungen, Befragungen sowie der Auswertung von Vertriebsdaten auf. Während Fahrgastzählungen in der Regel automatisiert über AFZS-Systeme durchgeführt werden, können Informationen zu den Fahrtwegen oder genutzten Tickets der Fahrgäste nur über aufwändige Fahrgastbefragungen erhoben werden, die in der Regel nur eine geringe zeitliche Gültigkeit besitzen.

Darüber hinaus können aus dem Verkauf von Fahrausweisen je nach Ticketart unterschiedliche Informationen gewonnen werden: Während im Bartarif zum Teil Informationen zur Start- und Zielhaltestelle enthalten sind, wird bei Zeitkarten in der Regel lediglich die Tarifzone erfasst – Informationen zur

Anzahl der Fahrten fehlen jedoch. Gleichzeitig besitzen Daten, die im Zuge des elektronischen Fahrgeldmanagements (EFM) erfasst werden, ein großes Potential: Mit der dritten Stufe des EFM soll eine automatische Fahrpreisberechnung eingeführt werden. Für diese Preisbestimmung ist die genaue Erfassung des Ein- und Ausstiegsorts mittels In-Out-Systemen notwendig. Diese Genauigkeit kann derzeit jedoch noch nicht automatisiert in Form eines BIBO-Systems erreicht werden. Grundsätzlich bieten die elektronischen Vertriebsdaten jedoch bei flächendeckendem Einsatz die Möglichkeit einer umfassenden Darstellung der Nachfrage. Wenn für die Nutzung von E-Tickets eine Registrierung mit Namen, Adresse und Geburtsdatum notwendig ist, können neben fahrtenbezogenen Daten auch Informationen über die demographische Struktur der E-Ticket Kunden erfasst werden.

Neben diesen klassischen Nachfragedatenquellen werden bereits weitere, innovative Datenquellen diskutiert:

- **Verbindungsankunftsdaten:** Aktuell werden Versuche unternommen, Daten, die bei der Nutzung von Verbindungsankünften entstehen, für Planungs- und Informationszwecken im ÖPNV zu operationalisieren. Dazu werden u. a. Verbindungsankunftsdaten mit weiteren Datensätzen (etwa Daten aus dem elektronischen Ticketing) kombiniert, um Belegungsprognosen für einzelne Linienfahrten zu erstellen. Die Nutzbarkeit von Daten aus Verbindungsankünften ist jedoch stark von dem Nutzungsverhalten der elektronischen Fahrplanauskunft durch die Fahrgäste abhängig – wie etwa der Häufigkeit der Anfragen für eine Verbindung oder der Genauigkeit der Anfrage (Ein- und Ausstiegsort). Aufgrund der (unbekannten) Verzerrungen sind die Verbindungsankunftsdaten daher nur schwer zu interpretieren.
- **Mobilfunkdaten:** Mobilfunkdaten basieren auf der Kommunikation von Mobiltelefonen und Mobilfunkmasten und werden durch Telekommunikationsdienstleister erfasst. Die Zuordnung von Signalen einzelner Endgeräte zu Funkzellen ermöglicht die zeitliche und räumliche Nachverfolgung einzelner Endgeräte. Die Genauigkeit der Standortbestimmung ist dabei von verschiedenen Einflussfaktoren, wie etwa der Ausstattung des Raums mit Basisstationen bzw. Funkmasten, gerätespezifischen Eigenschaften des Mobilfunktelefons sowie dem genutzten Verfahren zur Positionsbestimmung abhängig. Darüber hinaus ist eine valide Zuordnung

der Signale zu verschiedenen Verkehrsmodi aufgrund der teils ähnlichen Bewegungsmuster derzeit nur eingeschränkt möglich.

- WLAN- und Bluetooth-Daten: Die Nutzung von WLAN- und Bluetooth-Daten ist bereits in verschiedenen Forschungsprojekten geprüft worden. Fokus dieser Projekte war die Abschätzung von Fahrgastzahlen, die Ermittlung von Reise- und Wartezeiten sowie die Nachverfolgung von Wegen. Dabei ist zu berücksichtigen, dass die Validierung der entwickelten Methoden anhand von quantitativen Messungen im realen Testumfeld in der Regel fehlt und die Forschungsergebnisse teils stark aggregiert wurden. Darüber hinaus wird die Nachverfolgbarkeit einzelner Endgeräte durch den zunehmenden Schutz der Privatsphäre seitens der Smartphone-Hersteller erschwert. Diese verändern softwareseitig die bislang „festen“ Identifikationsmerkmale einzelner Endgeräte durch die sogenannte MAC-Randomisierung.

Die diversen Ansätze, um neue Datenquellen zur Nachfrageermittlung zu nutzen, zeigen das große Interesse der Forschung und ÖPNV-Betreiber, ihre Datengrundlage für Planungs- und Betriebsprozesse sowie das Marketing zu verbessern.

Betrachtete Datenquellen im Verfahren

Das im Rahmen dieser Dissertation entwickelte Verfahren basiert auf der Nutzung von WLAN- und Bluetooth-Daten, mit dem Ziel, die Signale einzelner Endgeräte räumlich und zeitlich nachverfolgen zu können und so Quelle-Ziel-Matrizen der Fahrgastnachfrage einer Linienfahrt ableiten zu können.

Unter WLAN-Daten werden Probe Requests verstanden. Beim Verbindungsaufbau zwischen einem Access Point und einem Endgerät muss das Endgerät zunächst über das Vorhandensein des Access Points informiert sein. Dazu kann das Endgerät selbst sogenannte „Probe Requests Frames“ (kurz Probe Requests) auf allen Kanälen des jeweiligen Frequenzbereichs an die Broadcast-MAC-Adresse versenden. Dies wird von allen Access Points in Reichweite empfangen, die die in den Probe Requests übermittelten Informationen hinsichtlich der Möglichkeit eines Datenaustauschs überprüfen. Mit den Probe Requests werden unter anderem die MAC-Adresse, Sequence Number und Information Elements (Informationen zu den Hardwareeigenschaften und Merkmale zur Datenübertragung) übertragen. Ist eine Datenübertragung möglich, sendet der Access Point eine Probe Response.

Bluetooth-Daten bezeichnen im Rahmen dieser Dissertation Inquiry-Response-Nachrichten. Bluetooth befindet sich auf Endgeräten standardmäßig im Standby-Modus, um Energie zu sparen. Erst wenn eine Verbindung aufgebaut werden soll, wird der Inquiry-Prozess (Verbindung zwischen zwei unbekanntem Geräten) gestartet. Das Gerät, das den Inquiry-Prozess anstößt und Inquiry-Nachrichten aussendet, wird als Master bezeichnet. Das Gerät, das nach Inquiry-Nachrichten scannt, wird als Slave bezeichnet. Der Inquiry-Prozess dient dem Master dazu, eine Liste von Bluetooth-Geräten zu erstellen, die sich in räumlicher Nähe befinden. Dazu versendet der Master mehrfach auf festgelegten Kanälen (Frequenzen) Inquiry Requests aus. Slaves, die gefunden werden wollen, scannen ebendiese Kanäle und antworten dem Master mit einer Inquiry Response Nachricht, die u. a. Geräte-Adresse, Clockwerte, lokalen Namen, unterstützte Funktionen (Telefonie, Audioausgabe, Navigation etc.) beinhaltet.

Darüber hinaus werden Daten von AFZS genutzt. AFZS zählen mit hoher Genauigkeit die ein- und aussteigenden Fahrgäste an Haltestellen einzelner Linienfahrten. Auf Grundlage der erfassten Daten können die Fahrzeugbelegung, die Verkehrsmenge (Linienbeförderungsfälle) und Verkehrsleistung (Personenkilometer, Pkm) abgeleitet werden. Da die einzelnen Fahrtweiten der Fahrgäste nicht bekannt sind, können keine Aussagen zur Fahrtweitenverteilung oder zu Quelle-Ziel-Relationen getroffen werden. Die AFZS-Daten dienen als Randsummen für die Hochrechnung der auf Basis von WLAN- und Bluetooth-Daten abgeleiteten Quelle-Ziel-Matrizen.

Verfahrensschritte

Die Datenaufzeichnung der WLAN- und Bluetooth-Daten beginnt in Fahrzeugen ab dem Zeitpunkt, zu dem das Fahrzeug gestartet wird. Grundsätzlich sollten auch Daten an Haltestellen aufgezeichnet werden, um die Zuordnungsgenauigkeit von Ein- und Ausstiegshaltestelle zu erhöhen, jedoch konnten die Sensoren in den Haltestellen nicht zuverlässig betrieben werden. Bei der Datenerfassung werden alle Signale in Empfangsreichweite der Erfassungseinheiten gespeichert, von denen jedoch nur eine Teilmenge für das Verfahren Relevanz besitzt. Daher werden die erfassten Signale in zwei Stufen gefiltert.

Stufe 1 filtert unplausible und nicht nutzbare Daten, wie etwa Daten mit fehlerhafter Zuordnung von Fahrtinformationen (Tag der Fahrt, Linie) oder Daten zu Zeitpunkten, in denen sich kein Fahrgast im Fahrzeug befand. Die verbleibenden Daten werden in einer Datenbank gespeichert.

Von den erfassten Daten sind nur die Daten der Fahrgäste für die Erstellung der Quelle-Ziel-Relationen relevant. Da alle Signale in Empfangsreichweite aufgezeichnet werden, ist es nicht zu vermeiden, dass neben den Signalen von Fahrgästen zu großen Teilen auch Signale von Nicht-Fahrgästen – etwa Fußgänger, Radfahrer, Pkw-Fahrer und -Mitfahrer oder auch Signale von Bewohnern oder Beschäftigten in der umliegenden Bebauung – miterfasst werden.

Aufbauend auf veröffentlichten Forschungsergebnissen und Erkenntnissen aus selbst durchgeführten Untersuchungen erfolgt daher in Stufe 2 eine weitere Datenfilterung. Diese Filterung hat das Ziel, mit hoher Wahrscheinlichkeit Daten von Nicht-Fahrgästen herauszufiltern, etwa auf Basis der Luftlinienentfernung zwischen der ersten und letzten Erfassung eines Endgeräts. Dabei können false positives (Daten von Nicht-Fahrgästen, die als Fahrgäste interpretiert werden) und false negatives (Daten von Fahrgästen, die als Nicht-Fahrgäste interpretiert werden) nicht ausgeschlossen werden. Die zweite Stufe der Filterung ist entsprechend mit Unsicherheiten verbunden, sodass im Rahmen der Verfahrensentwicklung verschiedene Filtereinstellungen in Szenarien untersucht und hinsichtlich ihrer Güte verglichen wurden.

Die verbliebenen Daten werden in einer Quelle-Ziel-Matrix aufbereitet. Die Zeilen und Spalten bestehen aus den Haltestellen der betrachteten Linienfahrt, die Matrixelemente ergeben sich aus der Zuordnung des ersten und letzten erfassten Signals eines Endgeräts zu Haltestellen, die den potentiellen Ein- und Ausstiegsort markieren. Das Ergebnis dieses Modellschritts ist die Ausgangsmatrix und bildet die Basis für die Hochrechnung. Es ist zu berücksichtigen, dass Unsicherheiten bei der Festlegung der Ein- und Ausstiegshaltestelle auf Grund des unbekanntem Sendeverhaltens der Endgeräte vorliegen.

Das Sendeverhalten von Probe Requests wurde bereits durch verschiedene Wissenschaftler untersucht. Das Versenden von Probe Requests hängt demnach von verschiedenen Faktoren ab. Dazu zählen allen voran die Nutzung des Endgeräts (aktive Nutzung / Standby) sowie der WLAN-Verbindungsstatus des Endgeräts. Weitere Faktoren, wie Hersteller und Betriebssystem(-version) des Endgeräts, die Anzahl bekannter Netzwerke, WLAN-spezifische Einstellungen des Endgeräts, Energiemodus bzw. Batteriezustand des Endgeräts haben ebenfalls Einfluss auf die Sendetätigkeit. Unter Berücksichtigung der Erkenntnisse der Wissenschaft sowie aus eigenen Untersuchungen kann festgehalten werden, dass keine Sendemuster einzelner Endgeräte

ableitbar sind. Die erste Erfassung eines Endgeräts kann somit sowohl unmittelbar mit dem Einstieg in das Fahrzeug als auch mehrere Minuten nach dem Einstieg erfolgen.

Um diese Unsicherheiten bei der Festlegung der Ein- und Ausstiegshaltestelle zu berücksichtigen, wurde die Fuzzy-Set-Theorie angewandt, die eine scharfe Zuteilung von Ein- und Ausstiegsort vermeidet und stattdessen unscharfe Mengen für einzelne Quelle-Ziel-Relationen ermittelt. Dazu wurde eine Zugehörigkeitsfunktion auf Basis der Häufigkeitsverteilung des mittleren zeitlichen Abstands zwischen erfassten Bluetooth- bzw. WLAN-Signalen gebildet. Für den Einstiegsort werden dabei alle Haltestellen, die bereits zwanzig Minuten vor Erfassung des ersten Signals eines Endgeräts angefahren wurden, als potentielle Einstiegsorte berücksichtigt. Dies gilt analog für alle kommenden Haltestellen nach Erfassung des letzten Signals eines Endgeräts für den Ausstiegsort. Die Fuzzy-Sets führen dazu, dass mehrere Quelle-Ziel-Relationen mit Wahrscheinlichkeitswerten für jeden betrachteten Fahrgast befüllt werden. Die Summe der Wahrscheinlichkeitswerte je Fahrgast beträgt dabei stets 100%.

Nach der Ermittlung der Ausgangsmatrizen erfolgt die Hochrechnung. Mit der Hochrechnung wird die Ausgangsmatrix mittels eines angepassten Steigerungsfaktorenmodells, das die mathematische Schwäche von Steigerungsfaktorenmodellen bei strukturellen Nullwerten berücksichtigt, schrittweise so korrigiert, dass die Randsummen eingehalten werden. Randsummen bilden hier die Ein- und Aussteigerzahlen je Haltestelle, die mit hoher Genauigkeit durch das AFZS vorliegen. Im Ergebnis liegt die auf Basis der Ausgangsmatrix und der AFZS-Daten hochgerechnete Bewertungsmatrix vor. Steigerungsfaktorenmodelle besitzen eine starke Abhängigkeit von der Ausgangsmatrix (Prinzip der Informationsminimierung) und den Randsummen. Während die Randsummen mit hoher Genauigkeit vorliegen (AFZS-Daten), werden die Matrixelemente der Ausgangsmatrix auf Basis der gefilterten WLAN- und Bluetooth-Signale – mit allen vorliegenden Unschärfen – erstellt. In der Ausgangsmatrix vorhandene und unbekanntere Verzerrungen können daher nicht bzw. nur bedingt ausgeglichen werden und wirken sich negativ auf die Modellergebnisse aus.

Die Hochrechnung hat zum Ziel, die wahrscheinlichste Matrix unter allen möglichen Matrizen, die die Randsummen einhalten, zu bestimmen. Dabei ist zu beachten, dass unter allen möglichen Lösungen nur eine die Realität abbildet – dies muss jedoch nicht die wahrscheinlichste Lösung sein.

Neben der Hochrechnung mit einem Steigerungsfaktorenmodell, das die WLAN- und Bluetooth-Daten verwendet, wurden zur Überprüfung des Mehrwerts der Eingangsdaten zwei weitere Hochrechnungsverfahren verwendet, die keine Eingangsdaten benötigen:

- Verfahren nach Li und Cassidy, das lediglich auf den Randsummen (AFZS-Daten) basiert und
- Fratar-Verfahren, in dem alle notwendigen Quelle-Ziel-Relationen auf Basis der Randsummen im ersten Iterationsschritt den gleichen Wert erhalten.

Aufgrund der vorliegenden Datenmenge wurden für die Schritte

- Erstellung der Datenbank,
- Erzeugung der Ausgangsmatrizen sowie
- Hochrechnung der Daten (Erzeugung Bewertungsmatrizen)

softwaretechnische Lösungen verwendet.

Zentrale Ergebnisse und Bewertung der Anwendung des Verfahrens

Das Verfahren wurde exemplarisch im Untersuchungsgebiet des Nordhessischen Verkehrsverbunds (NVV) angewandt. Dazu wurden zunächst drei ausgewählte Linien (Linie 100, 500 und 54) mit der technischen Hardware zur Datenerfassung (Raspberry Pi 3 Modell B+ mit WLAN Adapter, GSM-Modul und Speicherkarte) ausgestattet.

Die Bewertung der Verfahrensergebnisse erfolgt unter Zuhilfenahme verschiedener Gütemaße, die die Unterschiede der Modellschätzung (Bewertungsmatrix) im Bezug zu der realen Verteilung der Quelle-Ziel-Relationen der betrachteten Linienfahrten (Referenzmatrix) quantifizieren. Die Referenzmatrizen basieren auf Fahrgastbefragungen, die als Vollerhebung einzelner Linienfahrten der ausgestatteten Buslinien durchgeführt wurden. Insgesamt stehen 29 Referenzfahrten für die Auswertung zur Verfügung.

Zur Bewertung der Ergebnisse wurden folgende Gütemaße verwendet:

- GEH-Grenzwerte: Durch die GEH-Grenzwerte werden (zu) starke Abweichungen zwischen Bewertungs- und Referenzmatrix identifiziert. Die Bewertung durch GEH-Werte berücksichtigt sowohl die relative als auch die absolute Abweichung zwischen Bewertungs- und Referenzmatrix. Der GEH-Wert nimmt im Idealfall den Wert 0 an – in diesem Fall sind die Werte des Wertepaarvergleichs zwischen Bewertungs- und

Referenzmatrix identisch. Je höher der GEH-Wert, desto größer sind die Abweichungen.

- **Treffermaß (Anteil Treffer):** Mit dem Treffermaß wird unabhängig von dem Wert der Zellen (Quelle-Ziel-Relationen) geprüft, welcher Anteil der Zellen in der Bewertungsmatrix auch in der Referenzmatrix belegt ist. Je höher der Anteil der Treffer ist, desto mehr Übereinstimmungen gibt es mit der Referenzmatrix. Dabei wird jedoch keine Aussage zu zusätzlich falsch belegten Zellen getroffen. Informationen dazu können erst in Kombination mit dem Überfüllungsfaktor abgeleitet werden.
- **Überfüllungsfaktor:** Der Überfüllungsfaktor gibt an, ob in der Bewertungsmatrix mehr, weniger oder gleich viele Zellen gegenüber der Referenzmatrix belegt sind. Der Überfüllungsfaktor beträgt im Idealfall 1,0. Je höher die Abweichung von diesem Wert, desto größer ist die Differenz in der Anzahl der belegten Relationen zwischen Bewertungs- und Referenzmatrix. In Kombination mit dem Trefferanteil können so Informationen zur Anzahl falsch belegter Relationen abgeleitet werden.
- **Summe des Betrags der Abweichung der Verkehrsleistung in den Entfernungsklassen:** Mit diesem Gütemaß werden insbesondere Ergebnisse identifiziert, deren Quelle-Ziel-Relationen zwischen Bewertungs- und Referenzmatrix große Abstände (Haltestellen) aufweisen. Je näher die Summe der Abweichungen der Verkehrsleistung in den einzelnen Entfernungsklassen an dem Wert 0 liegt, desto geringer sind die Differenzen zwischen den Matrixelementen der Teilmatrizen.
- **Fehlerquadratsumme:** Die euklidische Distanz misst die Gesamtabweichung zwischen den einzelnen Matrixelementen der Bewertungs- und Referenzmatrix. Je näher die Fehlerquadratsumme (FQS) an dem Wert 0 liegt, desto geringer sind die Differenzen zwischen den Matrixelementen aus dem Verfahren sowie den Referenzdaten. Allerdings kann die reine Betrachtung der Fehlerquadratsumme zu Fehlinterpretationen führen, da eine große Anzahl sehr kleiner Abweichungen zu geringen Werten der Fehlerquadratsumme führt, obwohl hier eine Vielzahl der Quelle-Ziel-Relationen falsch besetzt sind.

Zur Erstellung der Bewertungsmatrizen für die 29 Referenzfahrten wurden Parameterkombinationen unter Berücksichtigung von

- 123 Filtersets aus drei verschiedenen Datenquellen, verschiedenen Filtereinstellungen und der Nutzung der Fuzzifizierung bei acht ausgewählten Filtereinstellungen sowie
- 28 Hochrechnungssets mit verschiedenen Einstellungen für das Fratar- und Li-Cassidy-Verfahren

gebildet. Insgesamt wurden über 30.000 Bewertungsmatrizen für die Auswertung ermittelt – mehr als 1.000 je Referenzfahrt. Zum Umgang mit strukturellen Nullen werden Nullwerte in der Ausgangsmatrix überall dort mit geringen Hilfswerten über 0 (z. B. 0,001) ersetzt, wo aufgrund der Randsummen der Quell- und Zielseite Werte größer 0 notwendig sind, damit das Verfahren zu einer stabilen Lösung hin konvergieren kann. Aufgrund der Datenerfassung sowie zum Teil restriktiver Filter kann eine große Anzahl von Hilfswerten notwendig werden. Diese Hilfswerte führen dazu, dass einige der Quelle-Ziel-Relationen auch nach Beendigung des Fratar-Verfahrens (Erreichen eines parametrisierten Genauigkeitsgrenzwerts) sehr kleine Werte besitzen. Zur Erhöhung der Lesbarkeit und Interpretierbarkeit der Bewertungsmatrix wurde daher ein zusätzliches Verfahren implementiert, das kleine Werte nach der Hochrechnung identifiziert und ggf. entfernt (Reduzierungsverfahren). Zur Identifikation eines „kleinen Werts“ wird ein Grenzwert – etwa 0,001 – eingesetzt. Werte, die kleiner als der Grenzwert sind, werden als „kleine Werte“ angesehen. Darauf aufbauend sucht das Verfahren nach Beendigung der Hochrechnung nach dem kleinsten Wert unterhalb des Grenzwerts und entfernt diesen. Anschließend wird die Hochrechnung (Fratar) erneut durchgeführt. Sofern die Hochrechnung ohne den entfernten Wert fehlschlägt, wird die Löschung rückgängig gemacht. Kommt das Verfahren zu einer Lösung, wird erneut der kleinste Wert unterhalb des Grenzwerts entfernt und die Hochrechnung erneut durchgeführt. Das Verfahren wird solange durchgeführt, bis die Hochrechnung keine Werte unterhalb des Grenzwerts aufweist. Abbruchkriterium für das Fratar-Verfahren ist eine erreichte Genauigkeit von 0,01 – d. h. eine Abweichung der Randsummen zwischen Bewertungs- und Referenzmatrix von maximal 0,01 (Fahrgäste). Sollte die Genauigkeit nicht erreicht werden, bricht das Verfahren nach 13.000 Iterationsschritten ab.

Der Mehrwert des Verfahrens wird auf Basis des Vergleichs der erzielten Ergebnisse unter Berücksichtigung der WLAN- und Bluetooth-Daten mit den Vergleichsverfahren abgeleitet, die keine Informationen zur Verteilung der Quelle-Ziel-Relationen in der Realität enthalten (Li-Cassidy, Fratar mit Einser-Matrix).

Die Auswertung der Ergebnisse der betrachteten Parameterkombinationen erfolgte sowohl differenziert für die einzelnen Gütemaße als auch in Form einer gütemaßübergreifenden Bewertung. Darüber hinaus wird die gütemaßübergreifende Bewertung für verschiedene Klasseneinteilungen der Referenzfahrten – nach Linien und nach Fahrgastnachfrage – durchgeführt. Für die Bewertung der Ergebnisse wurde auf eine Normierung der einzelnen Parameterkombinationen auf Basis der Ergebnisse über alle Referenzfahrten zurückgegriffen. Dazu wurden die erzielten Ergebnisse in den betrachteten Gütemaßen in vier Merkmalen zu Punktwerten zusammenfasst. Die betrachteten Merkmale

- GEH-Merkmal,
- Übereinstimmungsmerkmal,
- Summe der Beträge der Abweichung der Verkehrsleistung in den betrachteten EKL und
- Fehlerquadratsumme

wurden darüber hinaus entsprechend ihrer Bedeutung für die Verfahrensergebnisse gewichtet. Der ermittelte Punktwert ergibt sich aus dem Summenprodukt der Gewichtung mit dem erreichten Punktwert der Merkmale.

Auf Basis dieses Vorgehens wurden sieben Parameterkombinationen ermittelt, die die höchsten Punktwerte erzielen. Diese Parameterkombinationen greifen sowohl auf WLAN- als auch auf Bluetooth-Daten als Quelle zurück. Die Eingangsdaten zur Erstellung der Ausgangsmatrix werden hier durch eine Filterung auf Basis der GPS-Luftlinienentfernung reduziert. Bei differenzierter Betrachtung der Referenzfahrten nach Linien und Nachfrageklassen hat sich insbesondere die Parameterkombination aus WLAN- und Bluetooth-Daten mit Filterset P51_GPS200 mit Fuzzifizierung und Hochrechnung mittels Fratar-Verfahren (Hilfswert 1,0000, Reduktionsgrenzwert 0,5) als robust erwiesen: Diese Parameterkombination liegt in jeder betrachteten Klasse mindestens unter den 10% der besten Parameterkombinationen.

Darüber hinaus schneiden die ermittelten Parameterkombinationen stets besser ab als

- das beste Verfahren ohne Ausgangsdaten,
- das beste Verfahren ohne Filterung der Eingangsdaten sowie
- das beste Verfahren ohne Hochrechnung der Ausgangsmatrizen.

Dies bestätigt den Mehrwert des Verfahrens sowie der entwickelten Verfahrensschritte: Die Nutzung der WLAN- und Bluetooth-Daten als Datenquelle erzielt bessere Werte als Alternativen, denen keine weiteren Informationen zur Verteilung der Fahrgäste zur Verfügung stehen. Mit der Filterung der Daten können Stördaten zwar nicht eindeutig identifiziert und entfernt werden, jedoch wird die Datengrundlage für die Erstellung der Ausgangsmatrix gegenüber Alternativen ohne Filterung verbessert. Darüber hinaus wirkt die Hochrechnung erwartungsgemäß durch den Ausgleich der Randsummen positiv auf die Güte der Ergebnisse. Unter den besten Parameterkombinationen befinden sich sowohl Alternativen mit, als auch ohne Fuzzifizierung. Die Fuzzifizierung führt bei dem größeren Teil der betrachteten Parameterkombinationen zu einem Mehrwert. Gleichzeitig bestehen auch Parameterkombinationen, die ohne Fuzzifizierung Ergebnisse mit einer höheren Güte bestimmen.

Gleichzeitig ist zu berücksichtigen, dass einzelne betrachtete Parameterkombinationen insgesamt jedoch nur bei maximal 13 der 29 betrachteten Referenzfahrten (etwa 45%) die GEH-Grenzwerte einhalten. Auf einzelnen Referenzfahrten konnten die Grenzwerte durch keine der Parameterkombinationen eingehalten werden. Die erforderliche Genauigkeit wird somit bei einem Großteil der betrachteten Referenzfahrten nicht eingehalten. Daraus ergibt sich die Folgerung, dass das Verfahren in der praktischen Anwendung eine hohe Wahrscheinlichkeit besitzt, dass die Zuordnung der Fahrgäste auf die Quelle-Ziel-Relationen Fehler oberhalb der gesetzten Grenzwerte aufweist.

Die Gründe für diese Ergebnisse wurden ausführlich untersucht: Zusammenfassend bestehen Probleme bei der Datenerfassung, sodass neben den erforderlichen Daten von Fahrgastendgeräten auch eine Vielzahl von Stördaten erfasst und gespeichert werden. Mit der Erfassung der Stördaten wird eine Filterung notwendig, die jedoch auf Basis der erfassbaren Informationen der Probe Request und Inquiry-Response-Nachrichten nicht dazu geeignet ist, alle Stördaten zielgerichtet zu entfernen. Auf Basis der gefilterten Daten werden anschließend die Ausgangsmatrizen (Quelle-Ziel-Matrizen auf Basis der WLAN- und Bluetooth-Daten) erzeugt. Diese beinhalten bereits die beschriebenen Verzerrungen aus den vorangegangenen Verfahrensschritten. Darüber hinaus kommt es durch das heterogene Senderverhalten von Endgeräten zu Unschärfen bei der Zuordnung der Ein- und Ausstiegshaltestelle. Das genutzte Hochrechnungsverfahren (Fratar-Verfahren) basiert auf dem Prinzip der Informationsminimierung, sodass es die Verteilung und Informationen der

Ausgangsmatrix nach Möglichkeit und unter Berücksichtigung der Randsummen beibehält. In den erstellten Ausgangsmatrizen fehlen zum Teil an für das Fratar-Verfahren notwendigen Stellen der Quelle-Ziel-Matrix Einträge größer 0, damit das Verfahren konvergieren kann. Dies kann einerseits auf eine unvollständige Datengrundlage (nicht alle Fahrgäste führen ein Endgerät mit sich) und andererseits auf restriktive Filtersets zurückgeführt werden. Daher werden Hilfswerte eingesetzt, die jedoch keine Informationen zu der realen und gesuchten Quelle-Ziel-Matrix besitzen und somit die Anzahl der möglichen Lösungen erhöhen. Gleichzeitig wird dadurch die Wahrscheinlichkeit, die korrekte Quelle-Ziel-Matrix zu bestimmen, verringert.

Bei der Bewertung der Verfahrensergebnisse vor dem Hintergrund der angestrebten Einsatzbereiche (Angebotsplanung, Tarifplanung, Einnahmenaufteilung, Kalibrierung von Verkehrsmodellen) ist zu beachten, dass die Anforderungen an die räumliche und zeitliche Genauigkeit abhängig vom Anwendungszweck sind. Darüber hinaus ist die Qualität bisher genutzter Quelle-Ziel-Daten zu berücksichtigen. Quelle-Ziel-Daten stammen bislang aus Verkehrsmodellen und basieren auf Haushalts- und/oder Fahrgastbefragungen. Da eine repräsentative Erhebung von Quelle-Ziel-Verflechtungen mit einem erheblichen Aufwand verbunden wäre, sind Hochrechnungen notwendig, die zu bislang nicht quantifizierten Verzerrungen und Ungenauigkeiten in den bestehenden Daten führen. Demgegenüber wurden im Rahmen der Verfahrensbewertung die Ungenauigkeiten der Verfahrensergebnisse bereits für 29 Referenzfahrten anhand verschiedener Gütemaße ermittelt.

Mit dem aufgezeigten Verfahren besteht die Möglichkeit, dauerhaft Daten zu erfassen und somit eine bislang nicht vorhandene Datengrundlage zu bilden. Den ÖPNV-Betreibern wird es damit u. a. ermöglicht, Informationen zur Nachfrage wiederkehrender Events (etwa Fußballspiele, Konzerte o. ä.) zu sammeln und das Angebot entsprechend anzupassen, um etwa Kapazitätsüber- sowie -unterschreitungen zu vermeiden. Gleichzeitig können die Daten auf Tarifzonenebene aggregiert werden und in die Entwicklung von Tarifen eingehen.

Weiterentwicklung

Aufgrund der Güte der ermittelten Ergebnisse wurde exemplarisch eine Weiterentwicklung des Verfahrens in Form einer räumlichen Aggregation von Haltestellen geprüft: Zur Reduktion von Unschärfen bei der Zuordnung der Ein- und Ausstiegshaltestellen wurden dazu Haltestellen auf Stadtteilebene (in Kassel) bzw. Ortsteilebene (außerhalb von Kassel) aggregiert. Mit der

räumlichen Aggregation kommt es zu einer Reduzierung der Informationen, indem die Anzahl möglicher Relationen abnimmt – die Ein- und Ausstiegshaltestellen müssen entsprechend nicht mehr haltestellenscharf vorliegen, sondern auf Stadtteilebene. Da im Zuge der Aggregation die Anzahl möglicher Relationen der betrachteten Referenzfahrten stark reduziert wird, während die Belegung der Quelle-Ziel-Relationen höhere Werte annimmt, wurden weiterhin die GEH-Grenzwerte für die Bewertung der Ergebnisse der Aggregation angepasst.

Die räumliche Aggregation in der beschriebenen Form führte zu keiner generellen Verbesserung der Verfahrensgüte über alle Referenzfahrten und (betrachteten) Filtersets. Gleichzeitig ist anzumerken, dass durch die Aggregation die Einhaltung des GEH-1,9-Grenzwerts für zwei Fahrten der Linie 500 ermöglicht wurde, die im Fall ohne Aggregation durch keine Parameterkombination eingehalten werden konnte (GEH-1,5-Grenzwert).

6.2 Fazit und Einsatzbereiche des Verfahrens

Mit der vorliegenden Dissertation wurde ein Verfahren entwickelt, das die Fahrgastnachfrage für jede erfasste Linienfahrt anhand tagesaktueller, haltestellenscharfer Quelle-Ziel-Matrizen bereitstellt. Dazu werden Daten aus verschiedenen Quellen (AFZS-Daten, Fahrzeugdaten, WLAN- und Bluetooth-Daten) zusammenführt und strukturiert in einer Datenbank ablegt. Die Daten werden hinsichtlich ihrer Eignung bewertet und mehrstufig gefiltert. Sie gehen anschließend als Ausgangsdaten in die Erstellung von Quelle-Ziel-Matrizen ein und bilden die Eingangsdaten für die Hochrechnung. Bei der Hochrechnung werden Steigerungsfaktorenmodelle genutzt, um die erstellten Quelle-Ziel-Matrizen auf die realen Randsummen, die mittels AFZS-Daten vorliegen, auszugleichen. Für das gewählte Hochrechnungsverfahren (Fratar-Verfahren) war aufgrund der teils stark reduzierten Ausgangsmatrizen eine Anpassung notwendig: der Umgang mit strukturellen Nullwerten, sodass das Verfahren gegen eine stabile Lösung konvergieren kann.

Die Erfassung der Daten mit kostengünstiger Infrastruktur funktioniert in den Fahrzeugen und bietet Verkehrsunternehmen eine Möglichkeit, selbstständig Daten zu erfassen. Dabei ist zu berücksichtigen, dass die Erfassungsinfrastruktur im Idealfall nur in Fahrzeugen verbaut wird, die bereits über AFZS verfügen. Für Fahrzeuge ist nicht zu erwarten, dass höherpreisige Infrastruktur zu besseren Daten führt. Allerdings sind bei dauerhafter Anwendung die speziellen Vorschriften für Einbauten in Fahrzeuge des öffentlichen Verkehrs

zu beachten. Anders verhält es sich für die Datenerfassung an den Haltestellen. Diese hat nur in Einzelfällen funktioniert, sodass die entsprechenden Daten nicht in die Auswertungen eingehen konnten. Das Potential der Haltestellendaten liegt insbesondere darin, die Genauigkeit der Einstiegshaltestellen zu erhöhen und ggf. die Filterung zu verbessern.

Insgesamt hat sich gezeigt, dass eine Vielzahl an Datenpunkten erfasst werden, die zu großen Teilen nicht von Fahrgästen, sondern von Störquellen stammen. Die Informationen der erfassten Datenpakete sind darüber hinaus nicht geeignet, die Daten von Fahrgästen eindeutig von Stördaten trennen zu können. Die damit bestehenden Verzerrungen können nicht allein auf Basis der Hochrechnung und damit den Ausgleich der Matrixelemente auf die Randsummen ausgeglichen werden. Trotz dieser Verzerrungen konnte ein Mehrwert der Nutzung von WLAN- und Bluetooth-Daten gegenüber Vergleichsverfahren erreicht werden, die nur auf den Randsummen basieren. Ein Vergleich mit vorliegenden Quelle-Ziel-Daten als Ergebnis von bestehenden Nachfragemodellen konnte aufgrund des Fehlens ebendieser Daten im Untersuchungsgebiet nicht durchgeführt werden. Bei der Interpretation der Verfahrensergebnisse ist zu berücksichtigen, dass die Bewertung des Verfahrens bislang drei Linien im ÖPNV mit einer Stichprobe von insgesamt 29 Referenzfahrten umfasst.

Für die aufgeführten Einsatzbereiche von Quelle-Ziel-Daten bestehen unterschiedliche Anforderungen an die räumliche und zeitliche Genauigkeit. Diese Anforderungen sind abhängig vom Anwendungszweck innerhalb der Einsatzbereiche. Mit dem aufgezeigten Verfahren wird es den ÖPNV-Betreibern ermöglicht, kontinuierlich Daten zu erfassen und diese z. B. im Rahmen der Angebots- und Tarifplanung zu nutzen.

Das in dieser Dissertation entwickelte Verfahren kann als Grundbaustein genutzt werden, um die Datenqualität und -quantität von Quelle-Ziel-Daten zu erhöhen. Zur Optimierung der Ergebnisse können weitere oder zusätzliche Datenquellen integriert werden, die eine höhere Qualität als die in dieser Arbeit genutzten WLAN- und Bluetooth-Daten besitzen. Daher besteht an dieser Stelle weiterhin Forschungsbedarf, auf den im nächsten Kapitel eingegangen wird.

6.3 Ausblick und Forschungsbedarf

Das entwickelte Verfahren zur Abbildung der Verkehrsnachfrage in Form von linienspezifischen Quelle-Ziel-Matrizen auf Basis von AFZS-, Fahrzeug-

sowie WLAN- und Bluetooth-Daten besitzt Unschärfen, die nicht durch die bestehenden Verfahrensschritte ausgeglichen werden konnten. Zur Verfeinerung und Weiterentwicklung des vorliegenden Verfahrens besteht daher weiterer methodischer und datenseitiger Forschungsbedarf. Die grundsätzliche Eignung der Hochrechnung konnte für Daten, die eine hohe Genauigkeit besitzen, nachgewiesen werden. Die datenbedingte Unschärfe entsteht bereits mit der Datenerfassung (Stördaten) und kann durch ggf. weitere oder alternative Datenquellen minimiert werden. Das vorliegende Verfahren kann dabei als Grundbaustein für die Integration genutzt werden.

Zukünftig ist davon auszugehen, dass der Anteil an randomisierten WLAN- und Bluetooth-Daten unter allen erfassten Daten ansteigt, sodass die nutzbaren Daten in Zukunft tendenziell reduziert werden (These 7, Kapitel 4.7). Die zunehmende Standardisierung der MAC-Randomisierung bei neuen Endgeräten lässt darauf schließen, dass Maßnahmen zum Schutz der Privatsphäre der Smartphone-Nutzer seitens der OEM (Original Equipment Manufacturer) zunehmen werden. Zwar bestehen Möglichkeiten, einen Teil der randomisierten Daten auf einzelne Endgeräte zurückzuführen, jedoch sind diese mit zusätzlichem Aufwand verbunden, ohne dass ein Mehrwert zugesichert werden kann.

Im Folgenden wird der weitere Forschungsbedarf für das bestehende Verfahren sowie des Verfahrens unter Berücksichtigung weiterer möglicher Datenquellen beschrieben.

Integration weiterer Datenquellen: Das Verfahren ist offen für weitere bzw. andere Datenquellen, die dazu dienen, die Genauigkeit der Ein- und Ausstiegshaltestelle zu erhöhen. Gleichzeitig fallen im ÖPNV zunehmend Verkehrs- und Smartphone-Daten an. Diese Daten besitzen ein großes Potential, um einerseits die Angebotsplanung für die Fahrgäste und Kunden zu optimieren, aber auch um die Effizienz in den Verkehrsverbänden und -unternehmen zu steigern. Bei Nutzung mehrerer Datenquellen ist das Vorgehen zur Integration in das bestehende Verfahren zu prüfen (u.a. Ergänzung der Filtersets entsprechend der zusätzlichen Informationen einer neuen Datenquelle), insbesondere sofern es sich bei den Datenquellen nicht um disjunkte Mengen an Daten einzelner Fahrgäste bzw. Endgeräte handelt.

- **Verbindungsanfragen:** Auskunftsanfragen, die an die Verbindungsauskunft der Verkehrsunternehmen bzw. -verbände gerichtet werden,

enthalten u. a. Angaben²² zur Starthaltestelle und Zielhaltestelle in Klarschrift und als ID, Datum und Zeit des Eingangs der Anfrage im System sowie Datum und Zeit der angefragten Verbindung. Diese Daten bieten damit grundsätzlich hochaufgelöste Nachfragedaten, besitzen jedoch auch Schwächen: Es bleibt unklar, welche Routenvorschläge der Kunde angezeigt bekommt, welche Verbindung der Fahrgast tatsächlich nutzt, ob der Fahrgast einmalig oder mehrfach die gleiche Anfrage stellt, ob die Eingaben des Fahrgasts korrekt sind sowie ob die angefragte Uhrzeit der Verbindung sich auf die Abfahrts- oder Ankunftszeit bezieht. Daneben ist zu ermitteln, welche Nutzergruppen Verbindungsanfragen in welcher Form nutzen: Werden etwa Verbindungsanfragen für routinisierte Wege in urbanen Zentren überhaupt genutzt? Die Ankunftsanfragen können perspektivisch als Referenzgröße für die Ermittlung linienübergreifender Quelle-Ziel-Beziehungen herangezogen werden. Da über die Verbindungsauskünfte auch Fahrausweise erworben werden können (NVV), bieten sich diese Daten ebenfalls für die Auswertung an. Verkaufsdaten gehen mit sehr hoher Wahrscheinlichkeit auf real durchgeführte Fahrgastfahrten zurück, werden jedoch lediglich auf Ebene der Tarifzonen gespeichert.

- **Bus-WLAN:** Verkehrsunternehmen bieten zunehmend kostenfreies Internet über einen WLAN-Zugang in den ÖV-Fahrzeugen an. Somit stehen den Unternehmen bzw. den Dienstleistern in Abhängigkeit der Nutzungsbedingungen des Angebots umfangreiche Daten zu der Nutzung (u. a. Zeitpunkt und Dauer der Verbindung) zur Verfügung. Der kostenfreie Zugang kann dabei so eingerichtet werden, dass der Fahrgast vor der Verbindung einige Informationen hinterlegen muss. Um die Repräsentativität der Antworten sowie deren Korrektheit zu überprüfen, sind jedoch weitere Fahrgastbefragungen notwendig.
- **App-Daten:** Mit der App des Verkehrsverbunds sowie den entsprechenden Berechtigungen durch die Nutzer ist es möglich, die Wege sowie weitere Informationen (Ticketart, soziodemographische Merkmale etc.) der App-Nutzer zu speichern. Mit Nutzung der App liegt somit ein eindeutiges Identifizierungsmerkmal vor, das nicht von der MAC-Randomisierung beeinflusst werden kann. Hier sind jedoch Auswahleffekte zu

²² Geprüft für Auskunftsanfragedaten des NVV, geliefert durch Firma HaCon.

berücksichtigen, da nur die Fahrgäste erfasst werden können, die die App aktiv nutzen (gegenüber dem beschriebenen passiven Verfahrensansatz). Dies kann durch Incentives in Form von Vergünstigungen von Fahrausweisen oder der Berechtigung zur Nutzung des Bus-WLAN unterstützt werden.

- **Stadtweites WLAN:** Verschiedene Städte in Deutschland bieten bereits stadtweit freies WLAN an (Hamburg, Berlin). Zur Steigerung der Erfassungsgenauigkeit des Starts und Ziels einer Fahrgastfahrt besteht grundsätzlich auch die Möglichkeit, Daten dieser stadtweiten WLAN-Netze einzubeziehen.
- **E-Ticket-Daten:** Sofern die dritte Stufe des EFM eingesetzt wird, liegt die Fahrgastnachfrage für Nutzer der automatisierten Fahrpreisberechnung vor.

Bei Nutzung zusätzlicher Datenquellen sollte darüber hinaus

- die Priorisierung von Datenquellen zur Unterscheidung von Datenquellen mit hoher Güte (etwa App-Daten, E-Ticket-Daten) und Daten mit geringerer Güte (etwa Daten aus einem Stadtweiten WLAN) sowie
- Verfahren zur Fusionierung der Daten (Additiv, Multiplikativ, Mittelwerte etc.)

methodisch und empirisch analysiert werden.

Prüfung des Verfahrens für Anwendungsfälle mit langen Haltestellenabständen: Das Verfahren wurde bislang für ÖPNV-Anwendungsfälle getestet, in denen überwiegend kurze Haltestellenabstände vorliegen. Mit steigendem zeitlichen Abstand zwischen Haltestellen ist jedoch davon auszugehen, dass die Erfassungsquote der Fahrgäste zwischen den Haltestellen zunimmt und damit die Genauigkeit der Zuordnung von Ein- und Ausstiegshaltestelle steigt. Somit besteht aus Sicht des Autors großes Potential für die Anwendung des Verfahrens für Anwendungsfälle, die lange Fahrtabschnitte besitzen. IC-Fahrten der Deutsche Bahn beispielsweise besitzen Haltestellenabständen von teilweise mehr als einer Stunde auf. Darüber hinaus kann aufgrund der hohen Geschwindigkeit von deutlich weniger Störsignalen (während der Fahrt, nicht bei Halt bzw. Einfahrt in Bahnhof) ausgegangen werden.

Prüfung weiterer Formen zur Aggregation: Mit der Aggregation (zeitlich / räumlich) gehen einerseits Informationen verloren, andererseits besteht die Möglichkeit, Unschärfen auszugleichen. Bislang wurde eine Form der

räumlichen Aggregation exemplarisch geprüft. Daneben bieten sich weitere Möglichkeiten zur Aggregation an, insbesondere unter Verwendung ggf. weiterer Datenquellen.

Prüfung weiterer Zugehörigkeitsfunktionen (Fuzzifizierung): Neben der Aggregation bietet auch die Fuzzifizierung einen Umgang mit Unschärfen aus den erfassten Daten. Im Rahmen der Fuzzifizierung sind weitere Verteilungen der Zugehörigkeitsfunktionen zu prüfen, wie beispielsweise eine Verteilung, die auf dem Mittelwert der relativen Häufigkeiten je Endgerät und Zeitintervall basiert.

Erhöhung der Stichprobe (Referenzfahrten): Bislange wurde das Verfahren für 29 Referenzfahrten betrachtet, die Stichprobe für die Bewertung der Verfahrensergebnisse ist damit gering. Je geringer die Stichprobe, desto größer wird der Standardfehler. Zufällige Ereignisse haben damit einen sehr großen Einfluss auf das Ergebnis. Das Verfahren sollte daher auf weiteren Linienfahrten und Linien geprüft werden.

Differenzierung ländliche Räume und urbane Räume: Die Analysen in Kapitel 4.7 deuten darauf hin, dass die Güte der Verfahrensergebnisse bei Linien mit einem hohen Anteil an Haltestellen außerhalb des Stadtgebiets von Kassel höher ist. Für eine detaillierte Auswertung dieser Unterschiede sind Teilmatrizen für ländliche sowie für urbane Quelle-Ziel-Relationen zu bilden und die Ergebnisse auf Basis der bestehenden Gütemaße gegenüberzustellen.

Betrachtung von Verbundbeförderungsfällen: Derzeit zielt das Verfahren darauf ab, Quelle-Ziel-Relationen auf Linienfahrtebene zu betrachten. Für die Angebotsplanung und Einnahmenaufteilung ist die Betrachtung von Verbundfahrten von Relevanz. Daher ist auf Basis zusätzlicher Fahrzeugdaten (Kapitel 3.3.2.2) oder zusätzlicher Datenquellen die Güte des Verfahrens zur Ermittlung von Umstiegen bzw. Verbundfahrten zu prüfen.

Vergleich der Verfahrensergebnisse mit bestehenden Quelle-Ziel-Daten: Die Bewertung des Verfahrens im Rahmen dieser Dissertation erfolgte auf Basis vollständig vorliegender Referenzmatrizen. Vorteile des Verfahrens, etwa durch die dauerhafte Datenerfassung und die Möglichkeiten der Aggregation sind weitergehend zu prüfen. Dabei besteht insbesondere Forschungsbedarf bei dem Vergleich der Verfahrensergebnisse mit derzeit vorliegenden Daten in den Verkehrsverbänden. Diese basieren u. a. auf hochgerechneten Nachfragedaten (Fahrgastbefragung) und können daher

ebenso Verzerrungen beinhalten. Das Ausmaß der Verzerrungen in den bisher genutzten Daten ist dabei – aufgrund des erheblichen Aufwands zur Ermittlung realer Quelle-Ziel-Verflechtungen – unbekannt.

Die zunehmende Verfügbarkeit von Daten bietet nicht nur im ÖPNV-Sektor vielfältige Potentiale zur Operationalisierung. Einer Datennutzung muss jedoch stets eine intensive Datenaufbereitung und kritische Bewertung vorangestellt werden. Erst nach umfangreichen methodischen und empirischen Analysen, die die Brauchbarkeit und Übertragbarkeit eines Verfahrens auf Basis neuer Datenquellen reproduzierbar nachweisen und die erforderlichen Anforderungen einhalten, kann dieses in die Praxisanwendung übergehen.

Literaturverzeichnis

- Abrahamsson, Torgil (1998): Estimation of Origin-Destination Matrices Using Traffic Counts - A Literature Survey. International Institute for Applied Systems Analysis. Online verfügbar unter [https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.248&rep=rep1&type=pdf#:~:text=The%20Origin%2DDestination%20\(OD\),describing%20transportation%20in%20a%20region.&text=The%20OD%20matrix%20can%20be,in%20a%20%5Ctarget%20OD%20matrix.](https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.248&rep=rep1&type=pdf#:~:text=The%20Origin%2DDestination%20(OD),describing%20transportation%20in%20a%20region.&text=The%20OD%20matrix%20can%20be,in%20a%20%5Ctarget%20OD%20matrix.), zuletzt geprüft am 14.04.2021.
- AIRNET (2015): Aufgaben der einzelnen Schichten. Online verfügbar unter https://www.airnet.de/cr1-gfe/de/html/InfoAustKomp_learningObject7.xml, zuletzt aktualisiert am 17.03.2015, zuletzt geprüft am 18.11.2019.
- ALMEX GmbH (2021): Kreisverkehr Schwäbisch Hall. Online verfügbar unter <https://www.almex.de/portfolio-item/kreisverkehr-schwabisch-hall/>, zuletzt geprüft am 13.04.2021.
- Android Developers (2021): Android 6.0 Changes. Online verfügbar unter <https://developer.android.com/about/versions/marshmallow/android-6.0-changes.html>, zuletzt aktualisiert am 11.03.2021, zuletzt geprüft am 14.04.2021.
- Android Source (2020): Datenschutz: MAC-Randomisierung. Online verfügbar unter <https://source.android.com/devices/tech/connect/wifi-mac-randomization>, zuletzt aktualisiert am 09.01.2020, zuletzt geprüft am 13.04.2021.
- Apple Inc. (2020): Use Bluetooth and Wi-Fi in Control Center. Online verfügbar unter <https://support.apple.com/en-us/HT208086>, zuletzt geprüft am 19.03.2020.
- Apple Inc. (2021a): Informationen zu iOS 12-Updates. Online verfügbar unter <https://support.apple.com/de-de/HT209084>, zuletzt geprüft am 14.04.2021.
- Apple Inc. (2021b): Informationen zu iOS 13-Updates. Online verfügbar unter <https://support.apple.com/de-de/HT210393>, zuletzt geprüft am 14.04.2021.

- Apple Inc. (2022): Datenschutz und Privatsphäre bei WLAN-Verbindungen. Online verfügbar unter <https://support.apple.com/de-de/guide/security/secb9cb3140c/web>, zuletzt aktualisiert am 28.03.2022, zuletzt geprüft am 28.03.2022.
- Axel Springer SE (2021): Geschwindigkeit und Reichweite. Online verfügbar unter <https://www.computerbild.de/artikel/cb-Ratgeber-Handy-Allesueber-Bluetooth-3177119.html>, zuletzt geprüft am 14.04.2021.
- Badertscher, Kurt; Gubelmann, Josef; Scheuring, Johannes (2006): Wirtschaftsinformatik Grundlagen - Informations- und Kommunikationssysteme gestalten. Grundlagen mit zahlreichen Illustrationen, Beispielen, Repititionsfragen und Antworten. 1. Aufl. Zürich: Compendio Bildungsmedien AG (compendio Bildungsmedien).
- Baeta, Nuno; Fernandes, Agnelo; Ferreira, Joao (2017): Mining Users Mobility at Public Transportation. In: *Inteligencia Artificial* 20 (59), S. 32. DOI: 10.4114/intartif.vol20iss59pp32-41.
- Bai, Lu; Ireson, Neil; Mazumdar, Suvodeep; Ciravegna, Fabio (2017): Lessons learned using wi-fi and Bluetooth as means to monitor public service usage. In: Seungyon "Claire" Lee, Leila Takayama und Khai Truong (Hg.): Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. UbiComp '17: The 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Maui Hawaii, 11 09 2017 15 09 2017. New York, NY, USA: ACM, S. 432–440.
- Bandemer, Hand; Näther, Wolfgang (2013): Fuzzy Data Analysis: Springer Verlag.
- Basalamah, Anas (2016): Crowd Mobility Analysis using WiFi Sniffers. In: *International Journal of Advanced Computer Science and Applications* 7 (12). DOI: 10.1006/cviu.1998.0716.
- Baun, Christian (2018): Computernetze kompakt. 4. Auflage. Berlin: Springer Vieweg (IT kompakt). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-662-57469-0>.
- Becker, Johannes K.; Li, David; Starobinski, David (2019): Tracking Anonymized Bluetooth Devices. In: *Proceedings on Privacy Enhancing Technologies* 2019 (3), S. 50–65. DOI: 10.2478/popets-2019-0036.

- Ben-Akiva, Moshe; Macke, Peter; Hsu, Poh Ser (1985): Alternative methods to estimate route level trip tables and expand on-board surveys. Washington: Transportation Research Board (Transportation research record, 1037).
- Bieland, Dominik (2020): Internes Arbeitspapier zum technischen System der Datenerfassung in MobileDataFusion. nicht veröffentlicht.
- Bieland, Dominik; Briegel, Ramón (2021): Internes Arbeitspapier zur Fahrgastbefragung MobiDat Phase 1 (unveröffentlicht). Kassel.
- Bieland, Dominik; Sommer, Carsten (2021): Wirkungsabschätzung alternativer Finanzierungsinstrumente im ÖPNV am Beispiel des Rhein-Main-Verkehrsverbundes. Heureka 2021 - Optimierung in Verkehr und Transport. Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV); Verband Deutscher Verkehrsunternehmen e. V. (VDV). Online-Veranstaltung, 15.04.2021.
- Bitkom (2018): Anzahl der Smartphone-Nutzer in Deutschland in den Jahren 2009 bis 2018 (in Millionen). Online verfügbar unter <https://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenuutzer-in-deutschland-seit-2010/>, zuletzt geprüft am 22.11.2019.
- BLIC GmbH; KCW GmbH (2014): CICO-, CIBO und BIBO-basierte ÖPNV-Vertriebssysteme in Ballungs-räumen weltweit Markterkundung. Online verfügbar unter <https://docplayer.org/893383-Cico-cibo-und-bibo-basierte-oepnv-vertriebssysteme-in-ballungsraeumen-weltweit-markterkundung.html>, zuletzt aktualisiert am 18.05.2021.
- Bluetooth SIG (2009): Bluetooth Baseband. Online verfügbar unter https://www.ampedrftech.com/datasheets/cod_definition.pdf, zuletzt geprüft am 14.04.2021.
- Bluetooth SIG (2015): Bluetooth Technology Protecting Your Privacy. Online verfügbar unter <https://www.bluetooth.com/blog/bluetooth-technology-protecting-your-privacy/>, zuletzt aktualisiert am 02.04.2015, zuletzt geprüft am 14.04.2021.
- Bluetooth SIG (2019): Bluetooth Core Specification v5.2. Online verfügbar unter https://www.bluetooth.org/docman/handlers/download-doc.ashx?doc_id=478726, zuletzt geprüft am 15.04.2021.
- Borchers, Ralf (2013): Anwendung von Unsicherheitsmodellen am Beispiel der Verkehrserfassung unter Nutzung von Mobilfunkdaten. Institut für

- Ingenieurgeodäsie der Universität Stuttgart. Stuttgart. Online verfügbar unter https://elib.uni-stuttgart.de/bitstream/11682/3945/1/Dissertation_Ralf_Borchers.pdf, zuletzt geprüft am 14.04.2021.
- Brake, Klaus (1998): Dezentrale Konzentration in Großstadtreionen. In: *Raumforschung und Raumordnung Spatial Research and Planning* 56 (5-6), S. 343–351. DOI: 10.1007/BF03183757.
- Bregman, L. M. (1967): Eine Relaxationsmethode zur Bestimmung des gemeinsamen Punktes konvexer Mengen und ihre Anwendung zur Lösung konvexer Optimierungsaufgaben (in Russisch). In: *Shurnal vycisl. mat. i mat fiz.* 7 1967, S. 620–631.
- Brik, Vladimir; Banerjee, Suman; Gruteser, Marco; Oh, Sangho (2008): Wireless Device Identification with Radiometric Signatures. Online verfügbar unter http://www.winlab.rutgers.edu/~gruteser/papers/brik_paradis.pdf, zuletzt geprüft am 14.04.2021.
- Bucsky, Péter (2020): Modal share changes due to COVID-19: The case of Budapest. In: *Transportation research interdisciplinary perspectives* 8, S. 100141. DOI: 10.1016/j.trip.2020.100141.
- Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) (2021): Extended Mobile Network Data - Erschließung von Mobilfunkdaten zur nachfrageorientierten Planung im öffentlichen Verkehr – xMND. Online verfügbar unter <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mfundprojekte/xmnd.html>, zuletzt geprüft am 15.04.2021.
- Bundesnetzagentur (2015): Funkanwendungen auf den ISM-Bändern. Online verfügbar unter <https://emf3.bundesnetzagentur.de/pdf/ISM-BNetzA.pdf>, zuletzt geprüft am 28.11.2019.
- Bundesnetzagentur (2018): Vfg. 151/2018. Allgemeinzuteilung von Frequenzen in den Bereichen 5150 MHz - 5350 MHz und 5470 MHz - 5725 MHz für Funkanwendungen zur breitbandigen Datenübertragung, WAS/WLAN. Online verfügbar unter https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Telekommunikation/Unternehmen_Institutionen/Frequenzen/Allgemeinzuteilungen/2010_07_WLAN_5GHz_pdf.pdf?__blob=publicationFile&v=7.
- Bundesnetzagentur (2020): Allgemeinzuteilungen. Mobilfunk, Kurzstreckenfunk, DECT, WLAN, CB-Funk, etc. Online verfügbar unter

https://www.bundesnetzagentur.de/DE/Sachgebiete/Telekommunikation/Unternehmen_Institutionen/Frequenzen/Allgemeinzuteilungen/allgemeinzuteilungen-node.html, zuletzt geprüft am 03.02.2020.

c.c.com Moser GmbH (2020): BLIDS Projekte. Online verfügbar unter <https://www.blids.cc/blids-projekte/>, zuletzt geprüft am 13.04.2021.

Cisco Systems, Inc (2014): Cisco Connected Mobile Experiences (CMX) CVD - Location Fundamentals. Chapter 13. Online verfügbar unter https://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Borderless_Networks/Unified_Access/CMX/CMX_LocFund.pdf, zuletzt geprüft am 15.04.2021.

Cisco Systems, Inc (2018): 802.11 Association process explained. Online verfügbar unter https://documentation.meraki.com/MR/WiFi_Basics_and_Best_Practices/802.11_Association_process_explained, zuletzt geprüft am 28.11.2019.

Cunche, Mathieu; Kaafar, Mohamed-Ali; Boreli, Roksana (2014): Linking wireless devices using information contained in Wi-Fi probe requests. In: *Pervasive and Mobile Computing* 11, S. 56–69. DOI: 10.1016/j.pmcj.2013.04.001.

Cunche, Mathieu; Matte, Celestin (2016): On Wi-Fi Tracking and the Pitfalls of MAC Address Randomization. Online verfügbar unter <https://ido2016.sciencesconf.org/122873/document>, zuletzt geprüft am 04.03.2020.

DATACOM Buchverlag GmbH (2021): Bluetooth-Sicherheit. Online verfügbar unter <https://www.itwissen.info/Bluetooth-Sicherheit-Bluetooth-security.html#:~:text=Bluetooth%20kennt%20drei%20Sicherheitsstufen%3A%20die,flexible%20Zugriffe%20f%C3%BCr%20unterschiedliche%20Sicherheitsanforderungen.&text=Reklame%3A%20Alles%20rund%20um%20Bluetooth.>, zuletzt geprüft am 14.04.2021.

Dataset Bieland/Mobile Data Fusion (2020): Auswertung zur Konstanz von HT Capabilities in WLAN Probe Requests von Fahrzeug MDF 15 im Zeitraum 07.08.2020 bis 09.08.2020. Vorliegend an der Universität Kassel, Fachgebiet Verkehrsplanung und Verkehrssysteme.

Dataset Bieland/Mobile Data Fusion (2022): Rohdaten sowie Einstellungen für Filter- und HR-Sets mit Ergebnissen der Hochrechnung im Rahmen im Rahmen der Dissertation von Dominik Bieland (Ersteller, MySQL-

Datenbank). Vorliegend an der Universität Kassel, Fachgebiet Verkehrsplanung und Verkehrssysteme.

Dataset Mobile Data Fusion (2020): Rohdaten des AFZS (NVV) vom 03.04.2020. Vorliegend an der Universität Kassel, Fachgebiet Verkehrsplanung und Verkehrssysteme.

Demir, Levent (2013): Wi-Fi tracking: what about privacy. In: *HAL-inria*. Online verfügbar unter <https://hal.inria.fr/hal-00859013>, zuletzt geprüft am 13.04.2021.

Department for Transport (2020): TAG UNIT M3.1. Highway Assignment Modeling. Online verfügbar unter https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/938864/tag-m3-1-highway-assignment-modelling.pdf, zuletzt geprüft am 29.03.2022.

der-wirtschaftsingenieur.de (2010): OSI - Referenzmodell | Der Wirtschaftsingenieur.de. Online verfügbar unter <https://www.der-wirtschaftsingenieur.de/index.php/open-system-interconnection-referenzmodell/>, zuletzt geprüft am 18.11.2019.

Deutscher Bundestag (2017): Datenschutzrechtliche Zulässigkeit des WLAN-Trackings. Unter Mitarbeit von Wissenschaftliche Dienste. Online verfügbar unter <https://www.bundestag.de/resource/blob/538890/3dfae197d2c930693aa16d1619204f58/wd-3-206-17-pdf-data.pdf>, zuletzt geprüft am 14.04.2021.

Digi-Key (Hg.) (2018): 802.11x-Module und -Entwicklungskits können die Entwicklung drahtloser IoT-Designs vereinfachen. Vereinfachte Entwicklung mit Wi-Fi Modulen. Online verfügbar unter <https://www.digikey.de/de/articles/80211x-modules-dev-kits-iot-wireless-design-efforts>, zuletzt aktualisiert am 11.01.2018, zuletzt geprüft am 13.04.2021.

Dionicio, Rowell (2016): Reducing WiFi Channel Interference. Proper design helps minimize co-channel and adjacent channel interference, which degrade WLAN performance. Wireless infrastructure. Online verfügbar unter <https://www.networkcomputing.com/wireless-infrastructure/reducing-wifi-channel-interference>, zuletzt aktualisiert am 01.06.2016, zuletzt geprüft am 13.04.2021.

- Dontsova, Inga; Zeiser, Holger; Thiesing, Gustav (2020): Datenschutzkonzept. AP400. internes Arbeitspapier im Projekt Mobile Data Fusion (BMVI).
- Dunlap, Matthew; Li, Zhibin; Henrickson, Kristian; Wang, Yinhai (2016): Estimation of Origin and Destination Information from Bluetooth and Wi-Fi Sensing for Transit. In: *Transportation Research Record* 2595 (1), S. 11–17. DOI: 10.3141/2595-02.
- Elektronik-Kompendium.de (2018): Ortung und Positionsbestimmung mit Mobilfunk. Online verfügbar unter <https://www.elektronik-kompendium.de/sites/kom/1201061.htm>, zuletzt aktualisiert am 13.04.2021, zuletzt geprüft am 13.04.2021.
- Elektronik-Kompendium.de (2019a): MAC-Adresse. Online verfügbar unter <https://www.elektronik-kompendium.de/sites/net/1406201.htm>, zuletzt geprüft am 20.11.2019.
- Elektronik-Kompendium.de (2019b): WLAN-Beacons. Online verfügbar unter <https://www.elektronik-kompendium.de/sites/net/2010231.htm>, zuletzt geprüft am 02.12.2019.
- Elektronik-Kompendium.de (2019c): WLAN-Frequenzen und -Kanäle. Online verfügbar unter <https://www.elektronik-kompendium.de/sites/net/1712061.htm>, zuletzt geprüft am 10.12.2019.
- Elektronik-Kompendium.de (2021a): Bluetooth 1.0/1.1/1.2 (IEEE 802.15). Online verfügbar unter <https://www.elektronik-kompendium.de/sites/kom/0803301.htm><https://www.elektronik-kompendium.de/sites/kom/0803301.htm>, zuletzt geprüft am 13.04.2021.
- Elektronik-Kompendium.de (2021b): Bluetooth Low Energy (4.0 / 4.1 / 4.2). Online verfügbar unter <https://www.elektronik-kompendium.de/sites/kom/1805171.htm>, zuletzt geprüft am 13.04.2021.
- Elektronik-Kompendium.de (2021c): WLAN-Roaming. Online verfügbar unter <https://www.elektronik-kompendium.de/sites/net/1407081.htm>, zuletzt geprüft am 14.04.2021.
- Europäische Union (2021): Datenschutz-Grundverordnung. DSGVO, vom 04.03.2021.
- flinc GmbH (2016): Studie zeigt: Shuttle-System sorgt für 97% weniger Autos in Hamburg. Online verfügbar unter <https://flinc32.rssing.com/channel-52403089/latest-article8.php>, zuletzt geprüft am 15.04.2021.

Floemer, Andreas (2020): iOS 14 verbreitet sich schneller als iOS 13 – dank neuem Homescreen und Widgets. In: *t3n Magazin*, 2020. Online verfügbar unter <https://t3n.de/news/ios-14-verbreitet-schneller-ios-13-1346157/>, zuletzt geprüft am 02.04.2022.

Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) (1995): Hinweise zur Schätzung von Verkehrsbeziehungen mit Hilfe von Querschnittszählungen. W1. Köln: FGSV (Technische Regelwerke).

Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) (2010): Empfehlungen für Planung und Betrieb des öffentlichen Personennahverkehrs. Forschungsprojekt des Forschungsprogramms Stadtverkehr (FoPS). Ausg. 2010. Köln (FGSV 050).

Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) (2012): Empfehlungen für Verkehrserhebungen (EVE). Ausgabe 2012. Köln: FGSV Verlag GmbH (FGSV R2 - Merkblätter und Empfehlungen, 125).

Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) (2015): Handbuch für die Bemessung von Straßenverkehrsanlagen. HBS 2015. Ausg. 2015, Stand: 18.9.2015. Forschungsgesellschaft für Straßen- und Verkehrswesen. Köln (FGSV W1 - Wissensdokumente, FGSV 299 B).

Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) (2021): Verfahren zur Ermittlung der maßgebenden Verkehrsnachfrage für die Planung und Bemessung von Straßen – Zwischenbericht für die Bearbeitergruppe „Verkehrsnachfrage“ des AA 3.13 „Qualität des Verkehrsablaufs“.

Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV) (2022): Empfehlungen zum Einsatz von Verkehrsnachfragemodellen für den Personenverkehr. Ausg. 2022. Köln.

Forschungs-Informationen-System (FIS) (2019): Elektronische Dienste für den Personenverkehr. Hg. v. Bundesministerium für Verkehr und digitale Infrastruktur (BMVI). Online verfügbar unter <https://www.forschungsinformationssystem.de/servlet/is/451343/?clslid0=0&clslid1=0&clslid2=0&clslid3=0>, zuletzt aktualisiert am 12.03.2019, zuletzt geprüft am 15.04.2021.

Forschungs-Informationen-System (FIS) (2020): Elektronisches Fahrgeldmanagement. Mobilität und Verkehr. Unter Mitarbeit von Plank-

Wiedenbeck. Hg. v. Bundesministerium für Verkehr und digitale Infrastruktur (BMVI). Online verfügbar unter <https://www.forschungsinformationssystem.de/servlet/is/339460/>, zuletzt aktualisiert am 27.03.2020, zuletzt geprüft am 13.04.2021.

Fratar, T. J. (1954): Vehicular trip distribution by successive approximations. In: *Traffic Quarterly* 1, S. 53–65.

Freudiger, Julien (2015): How talkative is your mobile device? In: Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks. the 8th ACM Conference. New York, New York, 6/22/2015 - 6/26/2015. New York, NY, USA: ACM, S. 1–6.

Friedrich, Markus; Pestel, Eric; Pillat, Juliane; Heidl, Udo; Schiller, Christian; Simon, Robert (2019a): Anforderungen an städtische Verkehrsnachfragemodelle. FE 70.919/2015. Online verfügbar unter <https://fops.de/wp-content/uploads/2021/02/FE-70.0919-2015-Anf-an-staedt-Verkehrsnachfragemodelle-Schlussbericht.pdf>, zuletzt geprüft am 28.05.2022.

Friedrich, Markus; Pestel, Eric; Schiller, Christian; Simon, Robert (2019b): Scalable GEH: A Quality Measure for Comparing Observed and Modeled Single Values in a Travel Demand Model Validation. In: *Transportation Research Record* 2673 (4), S. 722–732. DOI: 10.1177/0361198119838849.

Friedrich, Markus; Schilling, Magdalena; Sonnleitner, Jörg; Vortisch, Peter; Buck, Sebastian; Baumann, Marvin et al. (2021): Verfahren zur Ermittlung der maßgebenden Verkehrsnachfrage für die Planung und Bemessung von Straßen. Zwischenbericht 3. FE 01.0203/2019/AGB.

Furness, K. P. (1965): Time function iteration. In: *Traffic Engin. Control* (7), S. 458–460.

Gast, Matthew S. (2021): 802.11 Wireless Networks: The Definitive Guide. Chapter 4, 2nd Edition. Online verfügbar unter <https://www.oreilly.com/library/view/80211-wireless-networks/0596100523/ch04.html>, zuletzt aktualisiert am 13.04.2021, zuletzt geprüft am 13.04.2021.

Gessler, Ralf; Krause, Thomas (2015): Wireless-Netzwerke für den Nahbereich. Wiesbaden: Springer Fachmedien Wiesbaden.

Gruteser, Marco; Grunwald, Dirk (2005): Enhancing Location Privacy in Wireless LAN Through Disposable Interface Identifiers: A Quantitative

Analysis. In: *Mobile Netw Appl* 10 (3), S. 315–325. DOI: 10.1007/s11036-005-6425-1.

Haase, Marc; Handy, Matthias (2004): BlueTrack - Imperceptible Tracking of Bluetooth Devices. Universität Rostock. Online verfügbar unter https://www.amd.e-technik.uni-rostock.de/veroeff/Haase_Handy_Ubi-COMP2004_BlueTrack-Imperceptible_Tracking_of_Bluetooth_Devices-Paper.pdf, zuletzt geprüft am 15.04.2021.

HaCon Ingenieurgesellschaft mbH (2020): RMV präsentiert Auslastungsprognose. Smarte Algorithmen für sicheren Abstand im ÖV. Online verfügbar unter <https://www.hacon.de/news/presse/rmv-praesentiert-auslastungsprognose/>, zuletzt aktualisiert am 13.04.2021, zuletzt geprüft am 13.04.2021.

Heinemann Verlag GmbH: OSI-Modell | it-administrator.de. Online verfügbar unter <https://www.it-administrator.de/lexikon/osi-modell.html>, zuletzt geprüft am 18.11.2019.

Hidayat, Arief; Terabe, Shintaro; Yaginuma, Hideki (2018): WiFi Scanner Technologies for Obtaining Travel Data about Circulator Bus Passengers: Case Study in Obuse, Nagano Prefecture, Japan. In: *Transportation Research Record* 2672 (45), S. 45–54. DOI: 10.1177/0361198118776153.

Hidayat, Arief; Terabe, Shintaro; Yaginuma, Hideki (2020): Bus Passenger Volume and Origin-Destination Based on Field Surveys Using a Wi-Fi Scanner. In: *Transportation Research Procedia* 20, S. 1376–1389. Online verfügbar unter <https://www.sciencedirect.com/science/article/pii/S235214652030586X?via%3Dihub>, zuletzt geprüft am 28.05.2022.

Hofmann, Stefan (2020): Das sollten Sie über Bluetooth wissen – Versionen, Profile und mehr. Alle Versionen & Unterschiede in der Übersicht. Online verfügbar unter <https://www.mobilefidelity-magazin.de/das-sollten-sie-ueber-bluetooth-wissen-versionen-profile-und-mehr>, zuletzt aktualisiert am 21.10.2020, zuletzt geprüft am 14.04.2021.

Huang, Haosheng; Cheng, Yi; Weibel, Robert (2019): Transport mode detection based on mobile phone network data: A systematic review. In: *Transportation Research Part C: Emerging Technologies* 101, S. 297–312. DOI: 10.1016/j.trc.2019.02.008.

- Hutter, Dominik (2016): MVG - Kostenloses Wlan für Bus, Tram und U-Bahn. In: *Süddeutsche Zeitung*, 09.03.2016. Online verfügbar unter <https://www.sueddeutsche.de/muenchen/nahverkehr-stadtrat-zwingt-mvg-zu-kostenlosem-wlan-1.2897913>, zuletzt geprüft am 26.11.2021.
- IEEE 802 LAN/MAN Standards Committee (2004): OVERVIEW AND GUIDE TO THE IEEE 802 LMSC, zuletzt geprüft am 02.12.2019.
- IEEE Computer Society (2016): IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012). IEEE Standard for Information technology--Telecommunications and information exchange between systems Local and metropolitan area networks--Specific requirements - Part 11: Wireless LAN Medium Access. New York: IEEE. Online verfügbar unter <http://ieeexplore.ieee.org/servlet/opac?punumber=7786993>.
- ifmo - Institut für Mobilitätsforschung (Hg.) (2011): Mobilität junger Menschen im Wandel – multimodaler und weiblicher. München.
- insoft GmbH (2019): Indoor Analytics in Echtzeit. Online verfügbar unter https://www.insoft.com/de/loesungen/anwendungen/indoor-analytics/gclid/cj0kcqia89zvbrdoarisaoiepbpp7firocazbxagoownjpfq-victwd1hvinfqomia9ivpu4efp-r8oaavfgealw_wcb, zuletzt geprüft am 16.12.2019.
- init innovation in traffic systems SE (2021): Benutzerhandbuch. MOBILEstatistics Web Client.
- init innovation in traffic systems SE (2022): Sensorliste - MobileDataFusion. Stand 01/2022.
- INRIX Inc. (2020): INRIX Verkehrsstudie: Stau verursacht Kosten in Milliardenhöhe. Online verfügbar unter <https://inrix.com/press-releases/2019-traffic-scorecard-german/>, zuletzt geprüft am 15.04.2021.
- Institute of Electrical and Electronics Engineers (IEEE) (2019): IEEE Public Listing (OUI). Online verfügbar unter <http://standards-oui.ieee.org/oui.txt>, zuletzt geprüft am 20.11.2019.
- Ipsos Retail Performance (2016): Wifi Analytics - How Can It Help With Customer Counting? » Ipsos » Market leader in people counting, management and behaviour insights. Online verfügbar unter <https://www.ipsos-retailperformance.com/en/insights/instore-wifi-analytics/>, zuletzt aktualisiert am 31.05.2022, zuletzt geprüft am 22.06.2022.

- Issoufaly, Taher; Tournoux, Pierre Ugo (2017): BLEB: Bluetooth Low Energy Botnet for large scale individual tracking. In: 2017 1st International Conference on Next Generation Computing Applications (NextComp). 2017 1st International Conference on Next Generation Computing Applications (NextComp). Mauritius, 19.07.2017 - 21.07.2017: IEEE, S. 115–120.
- Jarvis, Nicholas; Hata, John; Wayne, Nicholas; Raychoudhury, Vaskar; Gani, Md Osman (2019): MiamiMapper: Crowd Analysis using Active and Passive Indoor Localization through Wi-Fi Probe Monitoring. In: Geyong Min und Ahmed Mostefaoui (Hg.): Proceedings of the 15th ACM International Symposium on QoS and Security for Wireless and Mobile Networks - Q2SWinet'19. the 15th ACM International Symposium. Miami Beach, FL, USA, 25.11.2019 - 29.11.2019. New York, New York, USA: ACM Press, S. 1–10.
- Ji, Y.; Mishalani, R. G.; McCord, M. R. (2014): Iteratively Improving the Base Matrix of the IPF Method for Estimating Transit Route-level OD Flows from APC Data. In: *Journal of Transportation Engineering* (5).
- Ji, Yuxiong; Zhao, Jizhou; Zhang, Zhiming; Du, Yuchuan (2017): Estimating Bus Loads and OD Flows Using Location-Stamped Farebox and Wi-Fi Signal Data. In: *Journal of Advanced Transportation* 2017, S. 1–10. DOI: 10.1155/2017/6374858.
- Kahlert, J. (1995): Fuzzy control für Ingenieure. Analyse, Synthese und Optimierung von Fuzzy-Regelungssystemen. Unter Mitarbeit von Jörg Kahlert. Braunschweig, Wiesbaden: Vieweg.
- Karlsson, Jimi; Persson, Albin (2002): Device and Service Discovery in Bluetooth Networks. Master Thesis. Blekinge Institut of Technology, zuletzt geprüft am 15.04.2021.
- Keppeler, Bernhard; Kröpel, Stefan (2014): Verkehrsleistung und Reiseweiten sachgerecht abbilden: Hochrechnung von Stichprobenerhebungen in Fahrzeugen. In: *Der Nahverkehr* 32 (6), S. 30–39.
- Köhler, Uwe (2001): Analyse der Anwendung von Verkehrsnachfragemodellen. [Bericht zum Forschungs- und Entwicklungsvorhaben 01.144/1996/HGB des Bundesministeriums für Verkehr, Bau- und Wohnungswesen]. Bonn: Bundesministerium für Verkehr, Bau- und Wohnungswesen, Abt. Straßenbau, Straßenverkehr (Forschung Straßenbau und Straßenverkehrstechnik, 804).

- Kuhnimhof, Tobias (2012): Mobilitätstrends junger Erwachsener. In: *Internationales Verkehrswesen* 64 (2).
- KVG Kieler Verkehrsgesellschaft mbH: KVG rüstet Busflotte weiter mit WLAN aus. Online verfügbar unter https://www.kvg-kiel.de/presse/pressemittelungen/pm8_KVGruestetweitereBussemitWLANaus.pdf, zuletzt geprüft am 26.11.2021.
- Kyne, Oisín (2018): MAC Address De-Anonymisation. Online verfügbar unter <http://arxiv.org/pdf/1805.07613v1>, zuletzt geprüft am 15.04.2021.
- Lamond, B.; Stewart, N. F. (1981): Bregman's balancing method. In: *Transportation Research Part B: Methodological* 15 (4), S. 239–248. DOI: 10.1016/0191-2615(81)90010-2.
- Li, Yuwei; Cassidy, Michael J. (2007): A generalized and efficient algorithm for estimating transit route ODs from passenger counts. In: *Transportation Research Part B: Methodological* 41 (1), S. 114–125. DOI: 10.1016/j.trb.2006.04.001.
- Liebig, T.; Wagoum, A. U. K. (2012): Modelling microscopic pedestrian mobility using bluetooth. In: Joaquim Filipe (Hg.): Proceedings of the 4th International Conference on Agents and Artificial Intelligence. Vilamoura, Algarve, Portugal, 6 - 8 February, 2012. S.l.: SciTePress, S. 270–275.
- Loh, Desmond C. C.; Yuan Cho, Chia; Pheng Tan, Chung; Seng Lee, Ri (2008): Identifying Unique Devices through Wireless Fingerprinting. National University of Singapore; DSO National Laboratories. Online verfügbar unter <https://www2.seas.gwu.edu/~cheng/6547/Readings/p46-desmond.pdf>, zuletzt geprüft am 14.04.2021.
- Lohse, Dieter (1977): Berechnung von Personenverkehrsströmen. In: *Wissenschaft und Technik im Straßenwesen* (17).
- Lohse, Dieter (1997): Ermittlung von Verkehrsströmen mit n-linearen Gleichungssystemen unter Beachtung von Nebenbedingungen einschließlich Parameterschätzung. (Verkehrsnachfragemodellierung: Erzeugung, Verteilung, Aufteilung). Abschlußbericht zum DFG-Forschungsthema. Unter Mitarbeit von Gerald Bachner, Birgit Dugge und Heidrun Teichert. Dresden.

- Lohse, Dieter; Schnabel, Werner (2011): Grundlagen der Strassenverkehrstechnik und der Verkehrsplanung. 3., vollst. überarb. Aufl. 2 Bände. Berlin, Wien, Zürich: Beuth (2).
- Lukaß, Tom (2016): Offline-Tracking: Kundenfrequenzmessung in Ladengeschäften - datenschutz notizen | News-Blog der datenschutz nord Gruppe. Online verfügbar unter <https://www.datenschutz-notizen.de/offline-tracking-kundenfrequenzmessung-in-ladengeschaeften-4515135/>, zuletzt aktualisiert am 28.01.2021, zuletzt geprüft am 22.06.2022.
- Martin, Jeremy; Alpuche, Douglas; Bodeman, Kristina; Brown, Lamont; Fenske, Ellis; Foppe, Lucas et al. (2019): Handoff All Your Privacy - A Review of Apple's Bluetooth Low Energy Continuity Protocol. Online verfügbar unter <http://arxiv.org/pdf/1904.10600v2>.
- Martin, Jeremy; Mayberry, Travis; Donahue, Collin; Foppe, Lucas; Brown, Lamont; Riggins, Chadwick et al. (2017): A Study of MAC Address Randomization in Mobile Devices and When it Fails. In: *Proceedings on Privacy Enhancing Technologies* 2017 (4), S. 365–383. DOI: 10.1515/popets-2017-0054.
- Matte, Celestin (2017): Wi-Fi Tracking: Fingerprinting Attacks and Counter-Measures. Online verfügbar unter <https://hal.archives-ouvertes.fr/tel-01659783/document>, zuletzt geprüft am 23.01.2020.
- Matte, Célestin; Cunche, Mathieu; Rousseau, Franck; Vanhoef, Mathy (2016): Defeating MAC Address Randomization Through Timing Attacks. In: Matthias Hollick, Panos Papadimitratos und William Enck (Hg.): WiSec '16. Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks : July 18-20, 2016, Darmstadt, Germany. the 9th ACM Conference. Darmstadt, Germany, 7/18/2016 - 7/20/2016. WiSec; Association for Computing Machinery; ACM Conference on Security & Privacy in Wireless and Mobile Networks; Security and Privacy Week; SPW. New York, New York: The Association for Computing Machinery, S. 15–20.
- Matte, Célestin; Cunche, Mathieu; Toubiana, Vincent (2018): Does disabling Wi-Fi prevent my smartphone from sending Wi-Fi frames? Online verfügbar unter <https://hal.inria.fr/hal-01575519v2/document>, zuletzt geprüft am 19.03.2020.
- Medicus, Manuel (2020): Android-Versionen: Verteilung und Übersicht 2008-2020. Online verfügbar unter

<https://www.connect.de/ratgeber/android-versionen-verteilung-ueberblick-liste-3200452-8635.html>, zuletzt aktualisiert am 17.12.2020, zuletzt geprüft am 14.04.2021.

Mekky, A. (1983): A direct method for speeding up the convergence of the Furness biproportional method. : *Transp. Res.* 17B(1983)1, S. 1-11. In: *Transportation Research* 17B, S. 1–11.

Messe Berlin (2020): BUS2BUS - Datamining im ÖPNV. Online verfügbar unter <https://www.bus2bus.berlin/de/blog/datamining-im-%C3%B6pnv/>, zuletzt geprüft am 13.04.2021.

Mietke, Detlef (2019): Elektromagnetische Wellen. Ausbreitung elektromagnetischer Wellen. Online verfügbar unter <https://elektroniktutor.de/geraetetechnik/elmagwel.html>, zuletzt geprüft am 25.11.2019.

Mikkelsen, Lars; Buchakchiev, Radoslav; Madsen, Tatiana; Schwefel, Hans Peter (2016): Public transport occupancy estimation using WLAN probing. In: Magnús Jónsson, Jacek Rak, Arun Somani, Dimitri Papadimitriou und Alexey Vinel (Hg.): *Proceedings of 2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*. September 13-15, 2016, Halmstad, Sweden. 2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM). Halmstad, Sweden, 9/13/2016 - 9/15/2016. *International Workshop on Resilient Networks Design and Modeling; RNDM*. Piscataway, NJ: IEEE, S. 302–308.

Mishalani, Rabi G.; McCord, Mark R.; Reinhold, Ted (2016): Use of Mobile Device Wireless Signals to Determine Transit Route-Level Passenger Origin–Destination Flows: Methodology and Empirical Evaluation. In: *Transportation Research Record* 2544 (1), S. 123–130. DOI: 10.3141/2544-14.

Myrvoll, Tor A.; Håkegård, Jan E.; Matsui, Tomoko; Septier, François (2017): Counting Public Transport Passenger Using WiFi Signatures of Mobile Devices. Piscataway, NJ: IEEE. Online verfügbar unter <http://ieeexplore.ieee.org/servlet/opac?punumber=8307147>, zuletzt geprüft am 20.02.2020.

Navick, David S.; Furth, Peter G. (1994): Distance-Based Model for Estimating a Bus Route Origin-Destination Matrix. In: *Transportation Research Record* (1433), S. 16–22.

- Neumann, Marcus (2017): Hochrechnung von Fahrgastbefragungen im Öffentlichen Verkehr – Ansätze zur Vermeidung von Stichprobenverzerrungen. Masterarbeit. Dresden.
- Niedersächsisches Ministerium für Wirtschaft, Arbeit, Verkehr und Digitalisierung (2020): Grünes Licht für kostenloses WLAN in Hannovers Stadtbahnen | Nds. Ministerium für Wirtschaft, Arbeit, Verkehr und Digitalisierung. Online verfügbar unter <https://www.mw.niedersachsen.de/startseite/aktuelles/presseinformationen/grunes-licht-fur-kostenloses-wlan-in-hannovers-stadtbahnen-185186.html>, zuletzt aktualisiert am 26.11.2021, zuletzt geprüft am 26.11.2021.
- Nissen, Volker (2007): Ausgewählte Grundlagen der Fuzzy Set Theorie. Technische Universität Ilmenau (Ilmenauer Beiträge zur Wirtschaftsinformatik). Online verfügbar unter https://www.db-thueringen.de/servlets/MCRFileNodeServlet/dbt_derivate_00011797/IBzWI_2007-03.pdf, zuletzt geprüft am 14.04.2021.
- Nitti, Michele; Pinna, Francesca; Pintor, Lucia; Pilloni, Virginia; Barabino, Benedetto (2020): iABACUS: A Wi-Fi-Based Automatic Bus Passenger Counting System. In: *Energies* 13 (6), S. 1446. DOI: 10.3390/en13061446.
- NL Times (2013): Train Wifi Continues. Online verfügbar unter <https://nltimes.nl/2013/04/22/train-wifi-continues>, zuletzt geprüft am 15.04.2021.
- Nobis, Claudia (2013): Multimodale Vielfalt - Quantitative Analyse multimodalen Verkehrshandelns. Dissertation an der Humboldt-Universität zu Berlin.
- Nobis, Claudia.; Eisenmann, Christine (2020): Die Folgen der Corona-Pandemie für den ÖPNV. In: *Der Nahverkehr* (39), S. 10–15.
- Nordhessischer Verkehrsverbund (NVV) (2020): Fotos Mobile Data Fusion: Einbau Hardware in Haltestelle. nicht veröffentlicht.
- Nordhessischer Verkehrsverbund (NVV) (2022a): Fahrplanauskunft Linie 100. Online verfügbar unter <https://www.nvv.de/fileadmin/nvv/download/fahrplaene/100.pdf>, zuletzt geprüft am 10.01.2022.
- Nordhessischer Verkehrsverbund (NVV) (2022b): Fahrplanauskunft Linie 500. Online verfügbar unter <https://www.nvv.de/fileadmin/nvv/download/fahrplaene/500.pdf>, zuletzt geprüft am 10.01.2022.

- Nordhessischer Verkehrsverbund (NVV) (2022c): Fahrplanauskunft Linie 54. Online verfügbar unter <https://www.nvv.de/fileadmin/nvv/download/fahrplaene/054.pdf>, zuletzt geprüft am 10.01.2022.
- Nordhessischer Verkehrsverbund (NVV) (2022d): NVV: Einzelfahrkarte. Online verfügbar unter <https://www.nvv.de/tickets-preise/tickets/ticketsimueberblick/einzelfahrkarte>, zuletzt aktualisiert am 29.03.2022, zuletzt geprüft am 29.03.2022.
- NS Groep N.V. (2020a): NS privacy statement. Online verfügbar unter https://www.ns.nl/binaries/_ht_1586180771563/content/assets/ns-nl/privacy/ns-privacy-statement, zuletzt geprüft am 21.07.2020.
- NS Groep N.V. (2020b): Zitplaatszoeker | Reisinformatie. Online verfügbar unter <https://www.ns.nl/reisinformatie/service-verbeteren/zitplaatszoeker.html>, zuletzt geprüft am 23.07.2020.
- NS Groep NV (2021): Privacy | NS. In and around the station. Online verfügbar unter <https://www.ns.nl/en/privacy/in-and-around-the-station.html>, zuletzt aktualisiert am 26.11.2021, zuletzt geprüft am 26.11.2021.
- O'Reilly Online Learning (2022): 802.11 Wireless Networks: The Definitive Guide, 2nd Edition. Online verfügbar unter <https://www.oreilly.com/library/view/80211-wireless-networks/0596100523/ch04.html>, zuletzt aktualisiert am 28.03.2022, zuletzt geprüft am 28.03.2022.
- Oransirikul, Thongtat; Nishide, Ryo; Piumarta, Ian; Takada, Hideyuki (2014): Measuring Bus Passenger Load by Monitoring Wi-Fi Transmissions from Mobile Devices. In: *Procedia Technology* 18, S. 120–125. DOI: 10.1016/j.protcy.2014.11.023.
- Oransirikul, Thongtat; Piumarta, Ian; Takada, Hideyuki (2019): Classifying Passenger and Non-passenger Signals in Public Transportation by Analysing Mobile Device Wi-Fi Activity. In: *Journal of Information Processing* 27 (0), S. 25–32. DOI: 10.2197/ipsjip.27.25.
- Ortúzar, Juan; Willumsen, Luis (2004): Modelling transport. 3. ed., reprint. Chichester: Wiley.
- Ortúzar, Juan de Dios; Willumsen, Luis G. (2011): Modelling Transport. 4th ed. Chichester, UK: John Wiley & Sons, Ltd. Online verfügbar unter <https://onlinelibrary.wiley.com/doi/book/10.1002/9781119993308>.

- Paradedada, Diego Benites; Kraus Junior, Werner; Carlson, Rodrigo Castelan (2019): Bus passenger counts using Wi-Fi signals: some cautionary findings. In: *TRANSPORTES* 27 (3), S. 115–130. DOI: 10.14295/transportes.v27i3.2039.
- Pattanusorn, Woramate; Nilkhamhang, Itthisek; Kittipiyakul, Somsak; Ek-kachai, Kittipong; Takahashi, Atsushi (2016): Passenger estimation system using Wi-Fi probe request. In: 2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES). 2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES). Bangkok, Thailand, 20.03.2016 - 22.03.2016: IEEE, S. 67–72. Online verfügbar unter <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7467124>, zuletzt geprüft am 15.04.2021.
- Pestel, Eric; Friedrich, Markus; Heidl, Udo; Pillat, Juliane; Schiller, Christian; Schimpf, Martin (2016): Qualitätssicherung von Verkehrsnachfragemodellen. Quality Control of Travel Demand Models. Online verfügbar unter https://www.isv.uni-stuttgart.de/vuv/publikationen/downloads/20161000_SVT_Pestel-Friedrich-Heidl-Pillat-Schiller-Schimpf_Qualitaetssicherung-von-Verkehrsnachfragemodellen.pdf, zuletzt geprüft am 29.03.2022.
- Pillat, Juliane (2014): Methoden zur Analyse und Prognose des Verkehrsaufkommens unter Berücksichtigung des Wetters auf Autobahnen. Dissertation der Universität Stuttgart. In: *Veröffentlichungen aus dem Institut für Straßen- und Verkehrswesen* (59). Online verfügbar unter https://elib.uni-stuttgart.de/bitstream/11682/606/1/Dissertation_Pillat_A5.pdf, zuletzt geprüft am 02.04.2022.
- Pinelli, Fabio; Nair, Rahul; Calabrese, Francesco; Berlingerio, Michele; Di Lorenzo, Giusy; Sbodio, Marco Luca (2016): Data-Driven Transit Network Design From Mobile Phone Trajectories. In: *IEEE Trans. Intell. Transport. Syst.* 17 (6), S. 1724–1733. DOI: 10.1109/TITS.2015.2496783.
- PTV Planung Transport Verkehr AG (2018): smartStations: Die Haltestelle als Einstieg in die multimodale Mobilität. Unter Mitarbeit von Annette Kindl, Oliver Wolf, Markus Gläser und Christian Reuter. Hg. v. Bundesministerium für Verkehr und digitale Infrastruktur (BMVI). Online

verfügbar unter https://fops.de/wp-content/uploads/2020/07/700918_smartStation_Studie.pdf, zuletzt geprüft am 18.06.2021.

Pu, Ziyuan; Zhu, Meixin; Cui, Zhiyong; Wang, Yin Hai (2019): Mining Public Transit Ridership Flow and Origin-Destination Information from Wi-Fi and Bluetooth Sensing Data. Online verfügbar unter <https://arxiv.org/ftp/arxiv/papers/1911/1911.01282.pdf>, zuletzt geprüft am 14.04.2021.

Rakebrandt, Andreas (2013): Automatic Passenger counting with expandable features. APC Webinar. Online verfügbar unter https://www.itscanada.ca/files/MOBILE-APC_ITSCanada_webinar.pdf, zuletzt geprüft am 12.04.2021.

ISO 7498:1984: Reference Model of Open Systems Interconnection for CCITT Applications. Online verfügbar unter <https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=2820&lang=en>, zuletzt geprüft am 18.11.2019.

Rohde, Olaf (2021): MobiDat | Genauigkeit der GPS-Daten. Kassel, 15.02.2021. E-Mail an Dominik Bieland.

Rothengatter, Werner; Zhang, Junyi; Hayashi, Yoshitsugu; Nosach, Anastasiia; Wang, Kun; Oum, Tae Hoon (2021): Pandemic waves and the time after Covid-19 - Consequences for the transport sector. In: *Transport policy* 110, S. 225–237. DOI: 10.1016/j.tranpol.2021.06.003.

Rouse, Margaret (2005): Metropolitan Area Network (MAN). Online verfügbar unter <https://www.computerweekly.com/de/definition/Metropolitan-Area-Network-MAN>, zuletzt geprüft am 02.12.2019.

Rütermann, Tim; Benabbas, Aboubakr; Nicklas, Daniela (2019): Know Thy Quality: Assessment of Device Detection by WiFi Signals. In: 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). Kyoto, Japan, 3/11/2019 - 3/15/2019. Piscataway, NJ: IEEE, S. 639–644.

Ryu, Seunghan; Park, Byungkyu Brian; El-Tawab, Samy (2020): WiFi Sensing System for Monitoring Public Transportation Ridership: A Case

- Study. In: *KSCE J Civ Eng* 24 (10), S. 3092–3104. DOI: 10.1007/s12205-020-0316-7.
- Sapiezynski, Piotr; Stopczynski, Arkadiusz; Gatej, Radu; Lehmann, Sune (2015): Tracking Human Mobility using WiFi signals. In: *PLoS ONE* 10 (7), e0130824. DOI: 10.1371/journal.pone.0130824.
- Sauter, Martin (2013): Grundkurs Mobile Kommunikationssysteme. UMTS, HSPA und LTE, GSM, GPRS, Wireless LAN und Bluetooth. 5., überarb. und erw. Aufl. 2013. Wiesbaden, s.l.: Springer Fachmedien Wiesbaden. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-658-01461-2>.
- Schelewsky, Marc (2014): Tracking mit Smartphones: Einführung in die Technik und Stand der Forschung. Hg. v. Marc Schelewsky, Helga Jonuschat, Benno Bock und Korinna Stephan. Wiesbaden (Smartphones unterstützen die Mobilitätsforschung – Neue Einblicke in das Mobilitätsverhalten durch Wege-Tracking).
- Schmidt, Alexander; Männel, Tobias (2017): Potenzialanalyse zur Mobilfunkdatennutzung in der Verkehrsplanung. Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO. Online verfügbar unter <https://www2.iao.fraunhofer.de/images/iao-news/telefonica-studie.pdf>, zuletzt geprüft am 15.04.2021.
- Schmietendorf, Gaby (2011): Verkehrsdatenerfassung mit Bluetooth-Detektion: Möglichkeiten und Grenzen. Diplomarbeit. Technische Universität Dresden, Dresden.
- Schnieder, Lars (2018): Betriebsplanung im öffentlichen Personennahverkehr. Ziele, Methoden, Konzepte. 2. Auflage. Berlin, Heidelberg: Springer Vieweg.
- Schrader, Samuel (2020): Die-SARS-CoV-2-Pandemie-und-Strategien-für-den-ÖPNV_mib_V1.01. Online verfügbar unter https://mobilityinstitute.com/wp-content/uploads/2020/05/Die-SARS-CoV-2-Pandemie-und-Strategien-f%C3%BCr-den-%C3%96PNV_mib_V1.01.pdf, zuletzt geprüft am 28.03.2022.
- Schreiner, Rüdiger (2019): Computernetzwerke. Von den Grundlagen zur Funktion und Anwendung. 7., aktualisierte Auflage.
- Schulze, Christoph (2016): PTV AG Dokumentvorlage PTVAGSingle.dotm. Online verfügbar unter <https://zukunft-elbinsel.de/wp->

content/uploads/2020/03/U21.1-Bericht_A26-Ost-August-16_final_A.pdf, zuletzt geprüft am 29.03.2022.

Schüßler, Martina (o.J.): Clusteranalyse. Universität Kassel - FB07.

Sebayang, Andreas (2019): VVS: S-Bahn-Netz der Region Stuttgart bietet vollständig WLAN - Golem.de. In: *Golem.de*, 31.07.2019. Online verfügbar unter <https://www.golem.de/news/vvs-s-bahn-netz-der-region-stuttgart-bietet-vollstaendig-wlan-1907-142888.html>, zuletzt geprüft am 26.11.2021.

Simon, J.; Furth, Peter G. (1985): Generating a bus route o–d matrix from on–off data. In: *ASCE Journal of Transportation Engineering* (6), 583–593.

Skinner, Katie; Novak, Jason (2015): Privacy and your app. System Framework. #WWDC15. Online verfügbar unter https://devstreaming-cdn.apple.com/videos/wwdc/2015/703kp2dwbwkr/703/703_privacy_and_your_app.pdf?dl=1, zuletzt geprüft am 14.04.2021.

Software Mobile Data Fusion (2022): Softwaretools (StauFi, HR-Tool). Vorliegend an der Universität Kassel, Fachgebiet Verkehrsplanung und Verkehrssysteme.

Sommer, Carsten (2002): Erfassung des Verkehrsverhaltens mittels Mobilfunktechnik. Konzept, Validität und Akzeptanz eines neuen Erhebungsverfahrens. Dissertation. Braunschweig (Schriftenreihe des Instituts für Verkehr und Stadtbauwesen der TU Braunschweig).

Sommer, Carsten (2022): LVA: Planung des ÖPNV. 08 - Angebotsplanung.

Sommer, Carsten; Bieland, Dominik (2018): Finanzierung des ÖPNV auf Basis von alternativen Finanzierungsinstrumenten. Kassel.

Sommer, Carsten; Bieland, Dominik; Leonhäuser, Daniel; Saake, Stefan (2019): U-hoch-3: Unbeschwert urban unterwegs (zweite Projektphase). Verkehrsplanung und Verkehrssysteme. Online verfügbar unter <https://www.uni-kassel.de/fb14bau/institute/ifv/verkehrsplanung-und-verkehrssysteme/forschung-und-dienstleistungen/forschungsprojekte/u-hoch-3-unbeschwert-urban-unterwegs-zweite-projektphase.html>, zuletzt aktualisiert am 13.04.2021, zuletzt geprüft am 13.04.2021.

- Sommer, Carsten; Sauer, Jochen (2016): Tarifgestaltung und Einnahmenaufteilung. Hg. v. N. Ostermann und W. Rollinger. Hamburg (Handbuch ÖPNV - Schwerpunkt Österreich).
- Sommer, Carsten; Sauer, Jochen; Bartels, S.; Spichal, M. (2011): Verkehrsnachfrage genau erfassen - Qualitätsstandards und -management bei Fahrgasterhebungen. In: *Der Nahverkehr* (5), S. 30–36.
- Somolinos, Batriz Aznar (2009): Characteristics and Modeling of Wireless LAN Channels for Multimedia Home Networks. Dissertation. Technischen Universität Dortmund, Dresden. Elektrotechnik und Informationstechnik.
- Song, Baoyang; Wynter, Laura (2017): Real-time public transport service-level monitoring using passive WiFi: a spectral clustering approach for train timetable estimation. Online verfügbar unter <http://arxiv.org/pdf/1703.00759v1>, zuletzt geprüft am 21.02.2020.
- StatCounter (2020): Mobile Operating System Market Share Europe. Online verfügbar unter <https://gs.statcounter.com/os-market-share/mobile/europe/#monthly-200901-202001>, zuletzt geprüft am 20.03.2020.
- Statista (2022a): Android - Anteile der Versionen Januar 2022 | Statista. Online verfügbar unter <https://de.statista.com/statistik/daten/studie/180113/umfrage/anteil-der-verschiedenen-android-versionen-auf-geraeten-mit-android-os/>, zuletzt aktualisiert am 02.04.2022, zuletzt geprüft am 02.04.2022.
- Statista (2022b): Verbreitung von Android-Versionen bis 2019 | Statista. Online verfügbar unter <https://de.statista.com/statistik/daten/studie/227253/umfrage/verbreitung-von-android-versionen/>, zuletzt aktualisiert am 02.04.2022, zuletzt geprüft am 02.04.2022.
- Stöber, Tim; Schmitt, Jens; Frank, Mario; Martinovic, Ivan (2013): Who do you sync you are? Smartphone Fingerprinting via Application Behaviour. Online verfügbar unter <http://dl.acm.org/citation.cfm?id=2462096>, zuletzt geprüft am 14.04.2021.
- Tektronix (2013): Wi-Fi: Overview of the 802.11 Physical Layer and Transmitter Measurements. Online verfügbar unter https://www.cnrood.com/en/media/solutions/Wi-Fi_Overview_of_the_802.11_Physical_Layer.pdf, zuletzt geprüft am 14.04.2021.

- Telefónica Germany GmbH & Co. OHG (2021): Telefonica Deutschland auf einen Blick. Online verfügbar unter <file:///F:/Domi/Downloads/Factsheet-Telefonica-Deutschland-20210225-.pdf>, zuletzt geprüft am 04.05.2021.
- Telefónica Germany NEXT GmbH (2018): So bewegt sich Deutschland. Online verfügbar unter <https://www.telefonica.de/analytics/anonymisierte-daten/studie-so-bewegt-sich-deutschland.html>, zuletzt aktualisiert am 13.04.2021, zuletzt geprüft am 13.04.2021.
- The Highways Agency (UK); The Scottish Office development Department; The Welsh Office; The Department of the Environment for Northern Island; Department of Transport (UK) (1996): Design Manual for Roads and Bridges. Online verfügbar unter <https://assets.highwaysengland.co.uk/roads/road-projects/a2-bean-ebbsfleet-junction-improvements/Orders/l.8+DMRB+Part+1+Traffic+Appraisal.pdf>, zuletzt geprüft am 29.03.2022.
- Thorun, Christian; Rauber, Julius; Bock, Benno; Schelewsky, Marc; Schönduwe, Robert (2018): Digitale Transformation. Big Data im öffentlichen Personennahverkehr.
- Tkachenko, Sergey (2016): Enable random MAC address in Windows 10 for Wi-Fi adapter. Online verfügbar unter <https://winaero.com/enable-random-mac-address-in-windows-10-for-wi-fi-adapter/>, zuletzt geprüft am 14.04.2021.
- TomTom International BV (2021): UNGLAUBLICH GENAUE VERKEHRSINFORMATIONEN. Online verfügbar unter https://www.tomtom.com/de_de/drive/tomtom-traffic/, zuletzt geprüft am 15.04.2021.
- Transport for London (TfL) (2016): How TfL uses 'big data' to plan transport services. Online verfügbar unter <https://www.intelligenttransport.com/transport-articles/19635/tfl-big-data-transport-services/>, zuletzt geprüft am 22.07.2020.
- Transport for London (TfL) (2017): Review of the TfL WiFi pilot. Online verfügbar unter <http://content.tfl.gov.uk/review-tfl-wifi-pilot.pdf>, zuletzt geprüft am 22.07.2020.
- Transport for London (TfL) (2019): Wi-Fi data collection - Transport for London. Online verfügbar unter <https://tfl.gov.uk/corporate/privacy-and-cookies/wi-fi-data-collection?cid=wifi-data-collection>, zuletzt geprüft am 22.07.2020.

- Transport for London (TfL) (2021): Wi-Fi data collection. Online verfügbar unter <https://tfl.gov.uk/corporate/privacy-and-cookies/wi-fi-data-collection>, zuletzt aktualisiert am September 2021, zuletzt geprüft am 26.11.2021.
- Trialog Publishers Verlagsgesellschaft (2017): Mobilfunkdaten mit gutem Potenzial für Verkehrsplanung. *Mobilität: Wissenschaft*. Online verfügbar unter <https://www.internationales-verkehrswesen.de/mobilfunkdaten-zur-verkehrsplanung/>, zuletzt geprüft am 13.04.2021.
- Tsygalnitsky, S. (1977): *Simplified Methods for Transportation Planning*. Cambridge.
- Universität Kassel (2020): Seatfinder. Online verfügbar unter <https://www.uni-kassel.de/ub/benutzung-service/lernort-bibliothek/seatfinder.html>, zuletzt aktualisiert am 13.04.2021, zuletzt geprüft am 13.04.2021.
- Universität Rostock (o.J.): Bluetooth Baseband. Universität Rostock. Online verfügbar unter [https://www.amd.e-technik.uni-rostock.de/ma/gol/lectures/wirlec/bluetooth_info/baseband.html#:~:text=Each%20packet%20consists%20of%203,\(0-2745%20bits\).&text=Access%20Code:%20Access%20code%20are,offset%20compensation%2C%20paging%20and%20inquiry.](https://www.amd.e-technik.uni-rostock.de/ma/gol/lectures/wirlec/bluetooth_info/baseband.html#:~:text=Each%20packet%20consists%20of%203,(0-2745%20bits).&text=Access%20Code:%20Access%20code%20are,offset%20compensation%2C%20paging%20and%20inquiry.), zuletzt geprüft am 13.04.2021.
- van Genabith, Roman (2019): iOS 13 / iPadOS 13 verbreiten sich gut: 20% Marktanteil nach einer Woche. Online verfügbar unter <https://www.apfelpage.de/news/ios-13-ipados-13-verbreiten-sich-gut-20-marktanteil-nach-einer-woche/>, zuletzt aktualisiert am 23.09.2019, zuletzt geprüft am 14.04.2021.
- Vanhoef, Mathy; Matte, Célestin; Cunche, Mathieu; Cardoso, Leonardo S.; Piessens, Frank (2016): Why MAC Address Randomization is not Enough. In: Xiaofeng Chen, XiaoFeng Wang und Xinyi Huang (Hg.): *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*. the 11th ACM. Xi'an, China, 5/30/2016 - 6/3/2016. New York, NY: ACM, S. 413–424.
- VCD Verkehrsclub Deutschland e.V. (2021): Multimodale Navigations-Apps: Mithilfe von Mobilitäts-Apps auch ohne Pkw flexibel unterwegs. Online verfügbar unter <https://www.vcd.org/themen/multimodalitaet/schwerpunktthemen/multimodale-navigations-apps/>, zuletzt geprüft am 18.06.2021.

VDV eTicket Service GmbH & Co. KG (o.J.): ((eTicket Deutschland. NFC im ÖPNV. Online verfügbar unter https://oepnv.eticket-deutschland.de/fileadmin/user_upload/NFC-Basispapier.pdf, zuletzt geprüft am 26.11.2021.

VDV-KA KG (2012): Handbuch ((eTicket Deutschland. Online verfügbar unter https://oepnv.eticket-deutschland.de/fileadmin/Daten/Fachpublikationen/KoMi_ErgebnisHandbuch_KA_V_1.0.pdf, zuletzt geprüft am 26.11.2021.

Verband Deutscher Verkehrsunternehmen e. V. (VDV) (Hg.) (2018): Automatische Fahrgastzählsysteme. Handlungsempfehlungen zur Anwendung von AFZS im öffentlichen Personennahverkehr, Version 2.1. 04/2018. Unter Mitarbeit von Silvia Köhler, Stephan Bobinger, Ronny Branick, Bernd Cerfontaine, Sven Krogull, Andreas Luther et al. Köln (VDV-Schrift 457).

Verband Deutscher Verkehrsunternehmen e. V. (VDV) (2020): In die Spur zurückfinden. Online verfügbar unter https://www.vdv-dasmagazin.de/story_01_titelstory_032020.aspx, zuletzt geprüft am 04.05.2021.

Versichele, Mathias; Neutens, Tijds; Delafontaine, Matthias; van de Weghe, Nico (2012): The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. In: *Applied Geography* 32 (2), S. 208–220. DOI: 10.1016/j.apgeog.2011.05.011.

Vries, Andreas de; Weiß, Volker (2021): Grundlagen der Programmierung. Vorlesungsskript für das erste Semester Wirtschaftsinformatik. Fachhochschule Südwestfalen, Campus Hagen. Online verfügbar unter <https://www4.fh-swf.de/media/java.pdf>, zuletzt geprüft am 14.04.2021.

Wallen, Jack (2021): How to enable enhanced randomize MAC addresses on Android. In: *TechRepublic*, 19.01.2021. Online verfügbar unter <https://www.techrepublic.com/article/how-to-enable-enhanced-randomize-mac-addresses-on-android/>, zuletzt geprüft am 28.03.2022.

Waltari, Otto; Kangasharju, Jussi (2016): The Wireless Shark. Identifying WiFi Devices Based on Probe Fingerprints. In: David Choffnes, Hamed Haddadi und Narseo Vallina-Rodriguez (Hg.): Proceedings of the First Workshop on Mobile Data - MobiData '16. the First Workshop. Singapore, Singapore, 30.06.2016 - 30.06.2016. New York, New York, USA: ACM Press, S. 1–6.

- Wang, Yun-Pang (2008): OPTIMIERUNG DER QUELLE-ZIEL-MATRIXSCHÄTZUNG HINSICHTLICH REDUNDANZSTÖRUNG SICH VERÄNDERNDER VERKEHRZUSTÄNDE. Dissertation. Online verfügbar unter <https://core.ac.uk/download/pdf/237446273.pdf>, zuletzt geprüft am 14.04.2021.
- Weinstein, Lauren (2016): How TfL usues 'big data' to plan transport services. Hg. v. intelligenttransport.com. Transport of London (TfL). Online verfügbar unter <https://www.intelligenttransport.com/transport-articles/19635/tfl-big-data-transport-services/>, zuletzt aktualisiert am 13.04.2021, zuletzt geprüft am 13.04.2021.
- Welotec GmbH (2017): IEEE 802.11 WLAN Standards im Vergleich. Online verfügbar unter <https://www.welotec.com/de/wlan-standards-vergleich>, zuletzt geprüft am 22.11.2019.
- Wermuth, Manfred (2005): Modellvorstellung zur Prognose. In: G. Steierwald, H. D. Künne und W. Vogt (Hg.): Stadtverkehrsplanung – Grundlagen, Methoden, Ziele. Berlin, Heidelberg: Springer Verlag.
- WiGLE.net (2019): All the networks. Found by Everyone. Online verfügbar unter <https://wigo.net/>, zuletzt geprüft am 11.12.2019.
- Willumsen, Luis G. (1978): Estimation of an O-D Matrix from Traffic Counts – A Review. Working Paper. Institute of Transport Studies. Leeds.
- Wilson, A. G. (1974): Urban and regional models ingeography and plamnning. London.
- WVI Prof. Dr. Wermuth Verkehrsforschung und Infrastrukturplanung GmbH (2021): Lenkungskreis 2. September 2021. digital. unveröffentlicht.
- WVI Prof. Dr. Wermuth Verkehrsforschung und Infrastrukturplanung GmbH (2022): MobiDat - Auswertungen für die Verfahrensentwicklung. Dokumentation AP 800. unveröffentlicht. Unter Mitarbeit von Sophie Laufer.
- Zadeh, L. A. (1965): Fuzzy sets. In: *Information and Control* 8 (3), S. 338–353. DOI: 10.1016/S0019-9958(65)90241-X.
- Zehl, Franziska; Follmer, Robert; Weber, Patrick; Schelewsky, Marc (2020): Die Vermessung der Mobilität in der Pandemie. Gedämpfte Hoffnung auf die Verkehrswende. Mobilitätsreport 03, Ergebnisse aus Beobachtungen per repräsentativer Befragung und ergänzendem Mobilitätstracking bis Ende Oktober. Berlin, Bonn. Online verfügbar unter <https://www.infas.de/fileadmin/pdf->

geschuetzt/infas_Mobilit%C3%A4tsreport_WZB_7331_20201217.pdf,
zuletzt geprüft am 04.05.2021.

Anhang

Anhang 1: Datensatzbeschreibung WLAN-, Bluetooth- und AFZS-Daten

Tab. 1: Übertragene Daten für die Verfahrensentwicklung aus Probe Request Frames

Datenquelle	Information / Merkmal	Kurzbeschreibung
Smartphone	EVENT_ID	Einzigartige ID des aufgezeichneten Events.
	TIME_STAMP	Zeitangabe des gesendeten Probe Requests. Die Zeitangaben werden zwischen Access point und den verbundenen mobilen Endgeräten durch die Timing Synchronization Function (TSF) synchronisiert.
	MAC_TYPE	Unterscheidung der Art der MAC-Adresse nach „WIFI“ (Probe Request) oder „BLUETOOTH“ (Inquiry Message). Die WLAN- und Bluetooth-MAC-Adresse sind innerhalb eines Endgeräts unterschiedlich, da sie auf dem genutzten Chip basieren.
	HASHED_MAC_ADDRESS	Die MAC-Adresse dient zur eindeutigen Identifikation eines Endgeräts. Auf Grundlage des Datenschutzkonzepts wird diese durch einen Hash-Wert anonymisiert. Der Hash-Algorithmus wird zu Beginn eines Kalendermonats angepasst.
	LOCALLY_ADMINISTERED	Gibt den Wert des "least significant Bits" zurück. Dieser gibt Aufschluss darüber, ob die MAC-Adresse lokal erzeugt wurde (MAC-Randomisierung). (Vergleich Kapitel 2 MAC-Mac Random)
	HT_CAPABILITIES	HT Capabilities (High Throughput) beschreiben die unterstützten Hardware-Eigenschaften des Geräts zum Datenaustausch. Sie umfassen MCS-Werte (Modulation and Coding Scheme), d. h. unterstützte Modulationsverfahren, Frequenzen und Übertragungsraten, die für die Kommunikation zwischen Access Point und Endgerät möglich sind. Die HT-Capabilities und Supported Rates werden als Hexwerte ausgegeben.
	SUPPORTED_RATES	
	EXTENDED_CAPABILITIES	Extended Capabilities umfassen Informationen zu den Eigenschaften eines Geräts, die die HT Capabilities ergänzen. Innerhalb der der HT Capabilities können lediglich acht Übertragungsraten unterschieden werden. Mit neuen Standards (ab 802.11g) stehen mehr als acht Übertragungsraten zur Verfügung. Diese können mit den Extended Capabilities kommuniziert werden. Diese Daten werden analog zu den HT Capabilities als Hexwerte ausgegeben.
	EXTENDED_SUPPORTED_RATES	
	SEQUENCE_NUMBER	Sequence Number (SN) des übertragenen (Probe Request) Frames. Der Zähler der Sequence Number wird mit jedem Frame (der eine SN nutzt) um 1 erhöht.
	WIFI_CHANNEL	Angabe des Kanals im 2,4 GHz Spektrums, auf dem der Probe Request empfangen wurde (statisch = 1).
	SIGNAL_STRENGTH	Bei der Erfassung eines Probe Request Frames misst der Access Point die Signalstärke des empfangenen Signals. Diese wird durch den received signal strength indicator (RSSI) angezeigt.
	VENDOR	Hersteller des getrackten Fahrgastgerätes auf Basis der MAC-Adresse. Zu unterscheiden von dem "Vendor-Datenpaket," in dem nicht durch den Standard 802.11 vorgegebene, herstellereigene Informationen übertragen werden können.
ÖV-Fahrzeug	SENSOR_ID	Einzigartige ID der Datenerfassungseinheit
	VEHICLE_ID	Einzigartige ID des Fahrzeugs
	JOURNEY_ID	ID der Linienfahrt des Fahrzeugs, in dem sich der Sensor (Raspberry Pi) befindet. Es findet ein Abgleich mit den Planungsdaten des NVV statt (VDV452-Export).
	OPERATING_DAY	Einsatztag der Fahrzeugs

Datenquelle	Information / Merkmal	Kurzbeschreibung
	PREVIOUS_CALLED_STOP_PLACE_ID	ID der Haltestelle, an der das Fahrzeug zuletzt gehalten hat oder gerade hält.
	PREVIOUS_CALL_INDEX	Index des Halts innerhalb der Fahrt, beginnend bei 0.
	LONGITUDE	Längengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs
	LATITUDE	Breitengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs
Haltestelle	STOP_PLACE_ID	Einzigartige ID der Haltestelle, an der sich der Raspberry Pi/Copilot-PC befindet (nur bei Haltestellendaten)

Tab. 2: Übertragene Daten für die Verfahrensentwicklung aus Inquiry Nachrichten

Datenquelle	Information / Merkmal	Kurzbeschreibung
Smartphone	EVENT_ID	Einzigartige ID des aufgezeichneten Events.
	TIME_STAMP	Zeitangabe des gesendeten Probe Requests. Die Zeitangaben werden zwischen Access Point und den verbundenen mobilen Endgeräten durch die Timing Synchronization Function (TSF) synchronisiert.
	MAC_TYPE	Unterscheidung der Art der MAC-Adresse nach „WIFI“ (Probe Request) oder „BLUETOOTH“ (Inquiry Message). Die WLAN- und Bluetooth-MAC-Adresse sind innerhalb eines Endgeräts unterschiedlich, da sie auf dem genutzten Chip basieren.
	HASHED_MAC_ADDRESS	Die MAC-Adresse dient zur eindeutigen Identifikation eines Endgeräts. Auf Grundlage des Datenschutzkonzepts wird diese durch einen Hash-Wert anonymisiert. Der Hash-Algorithmus wird zu Beginn eines Kalendermonats angepasst.
ÖV-Fahrzeug	SENSOR_ID	Einzigartige ID der Datenerfassungseinheit
	VEHICLE_ID	Einzigartige ID des Fahrzeugs
	JOURNEY_ID	ID der Linienfahrt des Fahrzeugs, in dem sich der Sensor (Raspberry Pi) befindet. Es findet ein Abgleich mit den Planungsdaten des NVV statt (VDV452-Export).
	OPERATING_DAY	Einsatztag des Fahrzeugs
	PREVIOUS_CALLED_STOP_PLACE_ID	ID der Haltestelle, an der das Fahrzeug zuletzt gehalten hat oder gerade hält.
	PREVIOUS_CALL_INDEX	Index des Halts innerhalb der Fahrt, beginnend bei 0.
	LONGITUDE	Längengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs
Haltestelle	LATITUDE	Breitengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs
	STOP_PLACE_ID	Einzigartige ID der Haltestelle, an der sich der Raspberry Pi/Copilot-PC befindet (nur bei Haltestellendaten)

Tab. 3: Übertragene Daten für die Verfahrensentwicklung aus dem AFZS-System

Information / Merkmal	Kurzbeschreibung
OPERATING_DAY	Einsatztag der Fahrzeugs
VEHICLE_ID	Einzigartige ID des Fahrzeugs
METERS	Gefahrene Distanz in Meter zwischen Einträgen (Haltestellen)
STOP_LONGNAME	Vollständiger Haltestellenname
STOP_SHORTNAME	vierstellige Kennung der Haltestelle (Buchstaben)
STOP_CODE	vierstellige Kennung der Haltestelle (Zahlen)
STOP_POS	Codierung der Halteposition bei Haltestellen mit mehr als zwei Positionen für die Fahrtrichtung (z.B. bei Busbahnhöfen)
PLANNED_ARRIVAL	Ankunftszeit an der Haltestelle gemäß Fahrplan
PLANNED_DEPARTURE	Abfahrtszeit an der Haltestelle gemäß Fahrplan
ACTUAL_ARRIVAL	Reale Ankunftszeit an der Haltestelle
ACTUAL_DEPARTURE	Reale Abfahrtszeit an der Haltestelle
DISTANCE_TO_NEXT	Distanz zur nächsten Haltestelle in Meter
DOORS_OPENING	Kennwert zum Status der Türöffnung an der Haltestelle (0 = geöffnet, 1 = nicht geöffnet)
GPS_LATITUDE	Breitengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs
GPS_LONGITUDE	Längengrad (WGS-84) auf Basis der GPS-Position des Fahrzeugs
STOP_TYPE	Grund für den Halt an der Haltestelle (u. a. geplanter Halt, Halt auf Fahrgastwunsch, Durchfahrt, Störfall)
PASSENGER_IN	Zählwert der Einsteiger nach Datenkorrektur durch internen Algorithmus
PASSENGER_OUT	Zählwert der Aussteiger nach Datenkorrektur durch internen Algorithmus
PASSENGER_LOAD	Belegung des Fahrzeugs als Differenz zwischen korrigierten Zählwerten der Ein- und Aussteiger
PASSENGER_IN_EXTRASTOP	Hilfswert für Einsteiger, sofern das System einen Einsteiger nicht einer Haltestelle zuordnen kann.
PASSENGER_OUT_EXTRASTOP	Hilfswert für Aussteiger, sofern das System einen Einsteiger nicht einer Haltestelle zuordnen kann.
PASSENGER_LOAD_EXTRASTOP	Belegung des Fahrzeugs unter Berücksichtigung der Extrastop-Belegung.
PASSENGER_IN_ORG	Original-Zählwert der Einsteiger durch das AFZS
PASSENGER_OUT_ORG	Original-Zählwert der Aussteiger durch das AFZS
PASSENGER_LOAD_ORG	Belegung des Fahrzeugs als Differenz zwischen Original-Zählwerten der Ein- und Aussteiger
WHEELCHAIR_COUNT	Anzahl an Events, die durch Rollstühle verursacht wurden.
BLOCK_ID	Umlauf-ID der Linienfahrt des Fahrzeugs.
LINE_ID	Linienkennung der Linienfahrt des Fahrzeugs.
JOURNEY_ID	ID der Linienfahrt des Fahrzeugs.
APC_QUALITY	Kennzahl zur Genauigkeit der AFZS-Daten (Ein- und Aussteigerzahlen) auf Basis des Umfangs der notwendigen Korrekturen in Prozent.

Anhang 2: Beispielrechnung für die Fuzzifizierung auf Basis von Bluetooth-Daten

Diese Beispielrechnung dient der Verdeutlichung des statischen Verfahrensansatzes zur Fuzzifizierung auf Basis von Bluetooth-Signalen. Dazu wird ein theoretisches Beispiel konstruiert, bestehend aus

- IST-Fahrplan einer erdachten Linienfahrt und
- dem Zeitpunkt des ersten und letzten Signals eines Endgeräts.

In Tab. 4 sind die IST-Fahrplandaten der An- und Abfahrtszeiten dargestellt. Die Linienfahrt beginnt bei Haltestelle 1 um 13:00:15 Uhr.

Tab. 4: IST-Fahrplandaten der theoretischen Linienfahrt (Grundlage der Beispielrechnung)

Haltestelle	Ankunftszeit (IST)	Abfahrtszeit (IST)
1		13:00:15
2	13:02:30	13:02:45
3	13:04:30	13:04:45
4	13:05:30	13:05:45
5	13:08:30	13:08:45
6	13:11:00	13:11:15
7	13:21:00	13:21:15
8	13:24:00	13:24:15
9	13:26:30	13:26:45

In diesem Beispiel liegt der Zeitpunkt

- der ersten Erfassung um 13:06:00 Uhr und
- der letzten Erfassung um 13:09:15 Uhr.

Dabei handelt es sich um Bluetooth-Signale, sodass die Wahrscheinlichkeitsverteilung für Bluetooth-Signale (Abbildung 3-14) der weiteren Berechnung zugrunde liegt. Die Berechnungen erfolgen auf Basis der in Kapitel 3.3.3.2 vorgestellten Formeln. Eine skizzenhafte Darstellung der Linienfahrt ist in Abb. 1 gegeben.

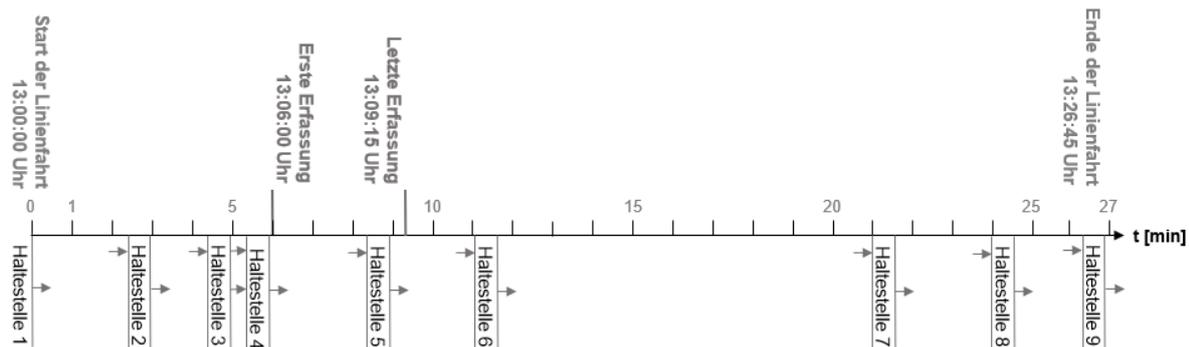


Abb. 1: Skizze der theoretischen Linienfahrt (Grundlage für die Beispielrechnung)

Schritt 1: Anpassung der Wahrscheinlichkeiten je Zeitintervall in Abhängigkeit der zusätzlichen Informationen ($p(aZI_x^Q)$, $p(aZI_x^Z)$)

Tab. 5 zeigt die minutenscharfen Zeitintervalle bis zur maximalen Betrachtungszeit von 20 Minuten sowie die unangepasste Wahrscheinlichkeit für den Ein- oder Ausstieg in dem betreffenden Zeitintervall. Daneben sind die aus den Zeitintervallen sowie den Zeitpunkten der ersten bzw. letzten Erfassung des Endgeräts abgeleiteten Uhrzeiten dargestellt, die sich aus den Zeitintervallen ergeben. Für die Berechnung der Einstiegshaltestelle wird ab dem Zeitpunkt der ersten Erfassung untersucht, welche Haltestellen zu welchen Zeitpunkten in der Vergangenheit angefahren wurden. Für die Berechnung der Ausstiegshaltestelle ist das Verfahren analog in die Zukunft gerichtet.

Die angepassten Wahrscheinlichkeiten ergeben sich auf Grundlage der zusätzlichen Informationen. Im Fall der Ausstiegshaltestelle

- wird im ersten Zeitintervall keine Haltestelle angefahren (d. h. im Zeitintervall 13:09:15 Uhr bis 13:10:14 Uhr kann kein Ausstieg stattgefunden haben, sodass die Ausstiegswahrscheinlichkeit sicher bei 0% liegt) und
- beträgt die restliche Fahrzeit der Linienfahrt weniger als 20 Minuten (die Linienfahrt endet um 13:26:45, sodass das letzte zu berücksichtigende Zeitintervall zwischen 13:26:15 Uhr bis 13:27:14 Uhr liegt. Danach kann kein Ausstieg mehr stattgefunden haben, sodass die Ausstiegswahrscheinlichkeit in den folgenden Zeitintervallen sicher 0% beträgt).

Daher wird die Wahrscheinlichkeitsverteilung ($p(ZI)$) auf die relevanten Zeitintervalle bezogen. Zur Verdeutlichung dient Abb. 1. Hier kann leicht abgelesen werden, dass bei der Berechnung der Einstiegshaltestelle gleich im ersten Zeitintervall eine Haltestelle erreicht wird und die insgesamt zu berücksichtigende Zeit bis zum Beginn der Fahrgastfahrt 6 Minuten beträgt. Die Wahrscheinlichkeiten der Zeitintervalle, die Zeiten von mehr als 6 Minuten

vor dem ersten Signal abdecken (vor 13:00:15 Uhr), sind entsprechend auf 0% zu setzen. Die Ergebnisse sind in Tab. 5 dargestellt.

Tab. 5: Ermittlung der angepassten Wahrscheinlichkeiten für die Berechnung der Ein- und Ausstiegshaltestelle

Zeitintervall [min]	Wahrscheinlichkeit für Ein-/Ausstieg in diesem Zeitintervall (ZI) $p(ZI)$	Zeitintervall für Berechnung der Einstiegshaltestelle [Uhrzeit]	angepasste Wahrscheinlichkeit für Einstieg in diesem Zeitintervall (ZI) $p(aZI_x^Q)$	Zeitintervall für Berechnung der Ausstiegshaltestelle [Uhrzeit]	angepasste Wahrscheinlichkeit für Ausstieg in diesem Zeitintervall (ZI) $p(aZI_x^Z)$
0 bis 1	75,2%	13:05:01 - 13:06:00	79,36%	13:09:15 - 13:10:14	0,00%
>1 bis 2	13,8%	13:04:01 - 13:05:00	14,53%	13:10:15 - 13:11:14	56,80%
>2 bis 3	3,2%	13:03:01 - 13:04:00	3,39%	13:11:15 - 13:12:14	13,24%
>3 bis 4	1,2%	13:02:01 - 13:03:00	1,32%	13:12:15 - 13:13:14	5,15%
>4 bis 5	0,8%	13:01:01 - 13:02:00	0,81%	13:13:15 - 13:14:14	3,18%
>5 bis 6	0,6%	13:00:01 - 13:01:00	0,59%	13:14:15 - 13:15:14	2,29%
>6 bis 7	0,5%	12:59:01 - 13:00:00	0,00%	13:15:15 - 13:16:14	2,26%
>7 bis 8	0,4%	12:58:01 - 12:59:00	0,00%	13:16:15 - 13:17:14	1,58%
>8 bis 9	0,4%	12:57:01 - 12:58:00	0,00%	13:17:15 - 13:18:14	1,67%
>9 bis 10	0,5%	12:56:01 - 12:57:00	0,00%	13:18:15 - 13:19:14	1,90%
>10 bis 11	0,3%	12:55:01 - 12:56:00	0,00%	13:19:15 - 13:20:14	1,40%
>11 bis 12	0,4%	12:54:01 - 12:55:00	0,00%	13:20:15 - 13:21:14	1,55%
>12 bis 13	0,3%	12:53:01 - 12:54:00	0,00%	13:21:15 - 13:22:14	1,25%
>13 bis 14	0,4%	12:52:01 - 12:53:00	0,00%	13:22:15 - 13:23:14	1,67%
>14 bis 15	0,3%	12:51:01 - 12:52:00	0,00%	13:23:15 - 13:24:14	1,37%
>15 bis 16	0,4%	12:50:01 - 12:51:00	0,00%	13:24:15 - 13:25:14	1,55%
>16 bis 17	0,4%	12:49:01 - 12:50:00	0,00%	13:25:15 - 13:26:14	1,76%
>17 bis 18	0,3%	12:48:01 - 12:49:00	0,00%	13:26:15 - 13:27:14	1,40%
>18 bis 19	0,2%	12:47:01 - 12:48:00	0,00%	13:27:15 - 13:28:14	0,00%
>19 bis 20	0,3%	12:46:01 - 12:47:00	0,00%	13:28:15 - 13:29:14	0,00%

Schritt 2: Ermittlung der relevanten Haltestellen je Zeitintervalle ($nHS^Q(ZI_x^Q), nHS^Z(ZI_x^Z)$)

Die Ermittlung der Anzahl relevanter Haltestellen je Zeitintervall erfolgt auf Basis von Formel 3-5 und Formel 3-6. Alternativ kann hier die entsprechende Anzahl leicht anhand von Abb. 1 sowie der festgelegten Bezugszeitpunkte abgelesen werden. Bezugszeitpunkt bei der Berechnung der Einstiegshaltestelle ist der Abfahrtszeitpunkt an einer Haltestelle, für die Berechnung der Ausstiegshaltestelle der Ankunftszeitpunkt des Fahrzeugs. Für die nicht relevanten Zeitintervalle gibt es keine relevanten Haltestellen (=0). Die Ergebnisse sind in Tab. 6 dargestellt.

Tab. 6: Ermittlung der relevanten Haltestellen für die Berechnung der Ein- und Ausstiegshaltestelle

Zeitintervall [min]	Zeitintervall für Berechnung der Einstiegshaltestelle [Uhrzeit]	Anzahl relevanter Haltestellen für Berechnung der Einstiegshaltestelle im Zeitintervall $nHS^Q(ZI_i^Q)$	Zeitintervall für Berechnung der Ausstiegshaltestelle [Uhrzeit]	Anzahl relevanter Haltestellen für Berechnung der Ausstiegshaltestelle $nHS^Z(ZI_j^Z)$
0 bis 1	13:05:01 - 13:06:00	1	13:09:15 - 13:10:14	0
>1 bis 2	13:04:01 - 13:05:00	2	13:10:15 - 13:11:14	1
>2 bis 3	13:03:01 - 13:04:00	2	13:11:15 - 13:12:14	1
>3 bis 4	13:02:01 - 13:03:00	3	13:12:15 - 13:13:14	1
>4 bis 5	13:01:01 - 13:02:00	3	13:13:15 - 13:14:14	1
>5 bis 6	13:00:01 - 13:01:00	4	13:14:15 - 13:15:14	1
>6 bis 7	12:59:01 - 13:00:00	0	13:15:15 - 13:16:14	1
>7 bis 8	12:58:01 - 12:59:00	0	13:16:15 - 13:17:14	1
>8 bis 9	12:57:01 - 12:58:00	0	13:17:15 - 13:18:14	1
>9 bis 10	12:56:01 - 12:57:00	0	13:18:15 - 13:19:14	1
>10 bis 11	12:55:01 - 12:56:00	0	13:19:15 - 13:20:14	1
>11 bis 12	12:54:01 - 12:55:00	0	13:20:15 - 13:21:14	2
>12 bis 13	12:53:01 - 12:54:00	0	13:21:15 - 13:22:14	2
>13 bis 14	12:52:01 - 12:53:00	0	13:22:15 - 13:23:14	2
>14 bis 15	12:51:01 - 12:52:00	0	13:23:15 - 13:24:14	3
>15 bis 16	12:50:01 - 12:51:00	0	13:24:15 - 13:25:14	3
>16 bis 17	12:49:01 - 12:50:00	0	13:25:15 - 13:26:14	3
>17 bis 18	12:48:01 - 12:49:00	0	13:26:15 - 13:27:14	4
>18 bis 19	12:47:01 - 12:48:00	0	13:27:15 - 13:28:14	0
>19 bis 20	12:46:01 - 12:47:00	0	13:28:15 - 13:29:14	0

Schritt 3: Berechnung der Gewichtungsfaktoren in Abhängigkeit der zu berücksichtigenden Haltestellen ($g(ZI_x^Q, HS_o^Q), g(ZI_x^Z, HS_d^Z)$)

Hier wird der Gewichtungsfaktor für die relevanten Haltestellen im betrachteten Zeitintervall bestimmt. Die Gewichtung erfolgt anhand der ermittelten Anzahl relevanter Haltestellen im betroffenen Zeitintervall (Tab. 7, Tab. 8). Der Gewichtungsfaktor ist für alle relevanten Haltestellen gleich.

Tab. 7: Ermittlung der Gewichtungsfaktoren für die Berechnung der Einstiegshaltestelle (erstes Signal: 13:06:00 Uhr)

Haltestelle gemäß Fahrplan	Abfahrtszeit (IST)	Anzahl relevanter Haltestellen	Gewichtungsfaktor			
			Haltestelle 4	Haltestelle 3	Haltestelle 2	Haltestelle 1
1	13:00:15	4	0,25	0,25	0,25	0,25
2	13:02:45	3	0,33	0,33	0,33	
3	13:04:45	2	0,50	0,50		
4	13:05:45	1	1,00			
5	13:08:45	0				
6	13:11:15	0				
7	13:21:15	0				
8	13:24:15	0				
9	13:26:45	0				

Tab. 8: Ermittlung der Gewichtungsfaktoren für die Berechnung der Ausstiegshaltestelle (letztes Signal: 13:09:15 Uhr)

Haltestelle gemäß Fahrplan	Anfahrtszeit (IST)	Anzahl relevanter Haltestellen	Gewichtungsfaktor			
			Haltestelle 6	Haltestelle 7	Haltestelle 8	Haltestelle 9
1	13:00:15	0				
2	13:02:30	0				
3	13:04:30	0				
4	13:05:30	0				
5	13:08:30	0				
6	13:11:00	1	1,00			
7	13:21:00	2	0,50	0,50		
8	13:24:00	3	0,33	0,33	0,33	
9	13:26:30	4	0,25	0,25	0,25	0,25

Schritt 4: Ermittlung der gewichteten Wahrscheinlichkeit für den Ein- oder Ausstieg an einer Haltestelle innerhalb eines Zeitintervalls ($p_g(ZI_x^Q, HS_o^Q), p_g(ZI_x^Z, HS_d^Z)$)

Auf Basis der ermittelten Gewichtungsfaktoren wird anschließend die gewichtete Wahrscheinlichkeit für den Ein- oder Ausstieg an den relevanten Haltestellen je Zeitintervall bestimmt, indem die angepasste Wahrscheinlichkeit auf Grundlage der zusätzlichen Informationen des Zeitintervalls ($p(aZI_x^Q)$, $p(aZI_x^Z)$) mit den Gewichtungsfaktoren multipliziert wird (Tab. 9, Tab. 10).

Tab. 9: Ermittlung der gewichteten Wahrscheinlichkeiten für die Berechnung der Einstiegshaltestelle

Zeitintervall für Berechnung der Einstiegshaltestelle [Uhrzeit]	angepasste Wahrscheinlichkeit für Einstieg in diesem Zeitintervall (ZI) $p(aZI_x^Q)$	Anzahl relevanter Haltestellen für Berechnung der Einstiegshaltestelle $nHS^Q(ZI_x^Q)$	gewichtete Wahrscheinlichkeit für den Einstieg an der Haltestelle im Zeitintervall			
			$p_g(ZI_x^Q, HS_4^Q)$	$p_g(ZI_x^Q, HS_3^Q)$	$p_g(ZI_x^Q, HS_2^Q)$	$p_g(ZI_x^Q, HS_1^Q)$
13:06:00 - 13:05:00	79,36%	1	79,36%	0,00%	0,00%	0,00%
13:04:59 - 13:04:00	14,53%	2	7,27%	7,27%	0,00%	0,00%
13:03:59 - 13:03:00	3,39%	2	1,69%	1,69%	0,00%	0,00%
13:02:59 - 13:02:00	1,32%	3	0,44%	0,44%	0,44%	0,00%
13:01:59 - 13:01:00	0,81%	3	0,27%	0,27%	0,27%	0,00%
13:00:59 - 13:00:00	0,59%	4	0,15%	0,15%	0,15%	0,15%
12:59:59 - 12:59:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:58:59 - 12:58:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:57:59 - 12:57:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:56:59 - 12:56:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:55:59 - 12:55:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:54:59 - 12:54:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:53:59 - 12:53:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:52:59 - 12:52:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:51:59 - 12:51:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:50:59 - 12:50:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:49:59 - 12:49:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:48:59 - 12:48:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:47:59 - 12:47:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
12:46:59 - 12:46:00	0,00%	0	0,00%	0,00%	0,00%	0,00%
Summe			89,18%	9,82%	0,86%	0,15%

Tab. 10: Ermittlung der gewichteten Wahrscheinlichkeiten für die Berechnung der Ausstiegshaltestelle

Zeitintervall für Berechnung der Ausstiegshaltestelle [Uhrzeit]	angepasste Wahrscheinlichkeit für Einstieg in diesem Zeitintervall (ZI) $p(aZI_x^Z)$	Anzahl relevanter Haltestellen für Berechnung der Einstiegshaltestelle $nHS^Z(ZI_x^Z)$	gewichtete Wahrscheinlichkeit für den Ausstieg an der Haltestelle im Zeitintervall			
			$p_g(ZI_x^Z, HS_6^Z)$	$p_g(ZI_x^Z, HS_7^Z)$	$p_g(ZI_x^Z, HS_8^Z)$	$p_g(ZI_x^Z, HS_9^Z)$
13:09:15 - 13:10:15	0,00%	0	0,00%	0,00%	0,00%	0,00%
13:10:16 - 13:11:15	56,80%	1	56,80%	0,00%	0,00%	0,00%
13:11:16 - 13:12:15	13,24%	1	13,24%	0,00%	0,00%	0,00%
13:12:16 - 13:13:15	5,15%	1	5,15%	0,00%	0,00%	0,00%
13:13:16 - 13:14:15	3,18%	1	3,18%	0,00%	0,00%	0,00%

Zeitintervall für Berechnung der Ausstiegshaltestelle [Uhrzeit]	angepasste Wahrscheinlichkeit für Einstieg in diesem Zeitintervall (ZI)	Anzahl relevanter Haltestellen für Berechnung der Einstiegshaltestelle	gewichtete Wahrscheinlichkeit für den Ausstieg an der Haltestelle im Zeitintervall			
13:14:16 - 13:15:15	2,29%	1	2,29%	0,00%	0,00%	0,00%
13:15:16 - 13:16:15	2,26%	1	2,26%	0,00%	0,00%	0,00%
13:16:16 - 13:17:15	1,58%	1	1,58%	0,00%	0,00%	0,00%
13:17:16 - 13:18:15	1,67%	1	1,67%	0,00%	0,00%	0,00%
13:18:16 - 13:19:15	1,90%	1	1,90%	0,00%	0,00%	0,00%
13:19:16 - 13:20:15	1,40%	1	1,40%	0,00%	0,00%	0,00%
13:20:16 - 13:21:15	1,55%	2	0,77%	0,77%	0,00%	0,00%
13:21:16 - 13:22:15	1,25%	2	0,62%	0,62%	0,00%	0,00%
13:22:16 - 13:23:15	1,67%	2	0,83%	0,83%	0,00%	0,00%
13:23:16 - 13:24:15	1,37%	3	0,46%	0,46%	0,46%	0,00%
13:24:16 - 13:25:15	1,55%	3	0,52%	0,52%	0,52%	0,00%
13:25:16 - 13:26:15	1,76%	3	0,59%	0,59%	0,59%	0,00%
13:26:16 - 13:27:15	1,40%	4	0,35%	0,35%	0,35%	0,35%
13:27:16 - 13:28:15	0,00%	0	0,00%	0,00%	0,00%	0,00%
13:28:16 - 13:29:15	0,00%	0	0,00%	0,00%	0,00%	0,00%
Summe			93,61%	4,14%	1,91%	0,35%

Schritt 5: Aufsummieren der Einzelwahrscheinlichkeiten ($p_g(HS_o^Q)$, $p_g(HS_d^Z)$)

Die Summe der Wahrscheinlichkeiten je Haltestelle ist bereits in den vorangegangenen Tabellen (Tab. 9, Tab. 10) dargestellt.

Schritt 6: Berechnung der Werte für die einzelnen Quelle-Ziel-Relationen ($p_g(HS_{od})$)

Aus den durchgeführten Schritten sind nun die Wahrscheinlichkeiten für den Ein- oder Ausstieg an den relevanten Haltestellen bekannt. Eine Darstellung in einer Quelle-Ziel-Matrix erfolgt in Tab. 11 und Tab. 12. Das Ergebnis der Fuzzifizierung für das betrachtete Endgerät auf der theoretischen Linienfahrt in diesem Beispiel ist in Tab. 13 dargestellt.

Sofern auf der Linienfahrt mehrere Endgeräte erfasst wurden, ist das Vorgehen zu wiederholen, die Ergebnismatrizen werden anschließend aufsummiert.

Tab. 11: Ergebnis der Fuzzifizierung der Einstiegshaltestelle

	1	2	3	4	5	6	7	8	9
1	-					<0,01			
2	-	-				0,01			
3	-	-	-			0,1			
4	-	-	-	-		0,89			
5	-	-	-	-	-				
6	-	-	-	-	-	-			
7	-	-	-	-	-	-	-		
8	-	-	-	-	-	-	-	-	
9	-	-	-	-	-	-	-	-	-

Tab. 12: Ergebnis der Fuzzifizierung der Ausstiegshaltestelle

	1	2	3	4	5	6	7	8	9
1	-								
2	-	-							
3	-	-	-						
4	-	-	-	-		0,94	0,04	0,02	<0,01
5	-	-	-	-	-				
6	-	-	-	-	-	-			
7	-	-	-	-	-	-	-		
8	-	-	-	-	-	-	-	-	
9	-	-	-	-	-	-	-	-	-

Tab. 13: Ergebnis der Fuzzifizierung der Quelle-Ziel-Relationen

	1	2	3	4	5	6	7	8	9
1	-					<0,01	<0,01	<0,01	<0,01
2	-	-				0,01	<0,01	<0,01	<0,01
3	-	-	-			0,09	<0,01	<0,01	<0,01
4	-	-	-	-		0,83	0,04	0,02	<0,01
5	-	-	-	-	-				
6	-	-	-	-	-	-			
7	-	-	-	-	-	-	-		
8	-	-	-	-	-	-	-	-	
9	-	-	-	-	-	-	-	-	-

Das Vorgehen kann analog für WLAN-Signale unter Berücksichtigung der Verteilung für den statischen Ansatz von WLAN durchgeführt werden. Auch bei Anwendung des dynamischen Ansatzes ist lediglich die Häufigkeitsverteilung zu verändern.

Anhang 3: Ergebnisdarstellung optimaler Parametereinstellungen nach Linie

In Tab. 14 bis Tab. 19 sind die Ergebnisse der Auswertungen auf Basis der Linien aufgeführt. Dabei wird auf das in Kapitel 4.6.4.2 beschriebene Vorgehen zur Normierung und Gewichtung zurückgegriffen. Die Linie steht dabei in direktem Zusammenhang mit der Anzahl der möglichen Relationen:

- Linie 100 (etwa 140 Relationen) ($n = 20$)
- Linie 54 (etwa 300 Relationen) ($n = 4$)
- Linie 500 (etwa 500 Relationen) ($n = 5$)

Linie 100

Zur Auswertung der Linie 100 stehen 20 der 29 Referenzfahrten zur Verfügung und damit die größte Stichprobe unter den betrachteten Linien. Anbei sind die besten Ergebnisse, die durch einzelne Parameterkombinationen erreicht werden, differenziert nach Gütemaßen, stichpunktartig aufgeführt:

- GEH-1,5-Grenzwert wird bei maximal 10 von 20 Referenzfahrten eingehalten
- GEH-0,7-Grenzwert wird bei maximal 6 von 20 Referenzfahrten eingehalten
- Fehlerquadratsumme besitzt den Mindestwert von 286
- Summe des Betrags der Abweichung der Verkehrsleistung über alle Entfernungsklassen liegt mindestens bei 110 km über alle betrachteten Referenzfahrten ($n = 20$) der Linie, dies entspricht rund 5.500 m je Referenzfahrt
- Anzahl übereinstimmender Referenzmatrizen mit Einhaltung des GEH-1,5- und GEH-0,7-Grenzwerts liegt maximal bei vier Referenzfahrten
- Anteil Treffer beträgt maximal 58%

Die Parameterkombinationen mit den besten Ergebnissen als Summe der normierten und gewichteten Gütemaßbewertungen zeigt Tab. 14. Tab. 15 zeigen die Ergebnisse der originären Gütemaße zu den Parameterkombinationen.

Tab. 14: Beste Parameterkombinationen auf Basis der Normierung und Gewichtung nach Tabelle 4-23

Datenquelle	Filterset	HR-Set	Fuzzy	GEH-Merkmal	FQS	Abweichung der Verkehrsleistung in den Entfernungsklassen in [100 km]	Übereinstimmungsmerkmal	Summe (Normierung)
WLAN_BT	P32_GPS600_RSSI60	FRAT_h1,00000_r0,5_----	x	0,30	0,34	0,45	0,47	0,38
WLAN_BT	P32_GPS600_RSSI60	FRAT_h1,00000_r1,0_----	x	0,25	0,22	0,49	0,54	0,37
WLAN_BT	P50_GPS100	FRAT_h0,00001_r0,5_----	x	0,30	0,36	0,42	0,47	0,37
WLAN_BT	P50_GPS100	FRAT_h1,00000_r0,5_----	x	0,30	0,36	0,40	0,46	0,37
WLAN_BT	P51_GPS200	FRAT_h0,00001_r0,5_----	x	0,30	0,34	0,37	0,49	0,37
WLAN_BT	P51_GPS200	FRAT_h0,00001_r1,0_----	x	0,25	0,20	0,33	0,58	0,37
WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5_----	x	0,30	0,37	0,39	0,47	0,37
WLAN_BT	P51_GPS200	FRAT_h1,00000_r1,0_----	x	0,25	0,21	0,32	0,57	0,37
WLAN_BT	P53_GPS600	FRAT_h1,00000_r0,5_----	x	0,28	0,37	0,46	0,47	0,37

Die besten Parameterkombinationen für Linie 100 umfassen mit Ausnahme der Parameterkombinationen ohne Fuzzifizierung alle in Kapitel 4.6.4.2 ermittelten Parameterkombinationen mit einem erreichten Punktwert von 0,37 Punkten. Die hohe Übereinstimmung kann vor allem dadurch erklärt werden, dass 20 der 29 Referenzfahrten auf Linie 100 entfallen und die Ergebnisse der Linie 100 somit ein großes Gewicht für die Gesamtbetrachtung besitzen. Darüber hinaus kommen zwei Parameterkombinationen auf Basis des restriktivsten Filters P32 zu Ergebnissen der gleichen oder höherer Güte innerhalb der Linie. Als HR-Set wird das Fratar-Verfahren mit Reduktionsgrenzwerten von 0,5 oder höher genutzt. Zum Vergleich: Das beste Ergebnis ohne Ausgangsdaten erhält 0,26 Punkte unter Nutzung des Fratar-Verfahrens mit Reduktionsgrenzwert 0,5.

Tab. 15: Beste Parameterkombinationen mit originären Gütemaßen

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleis- tung in den Entfernungsklas- sen in [100 km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN_BT	P32_GPS600_ RSSI60	FRAT_h1,00000 _r0,5_----	x	8	5	2	419	1,7	54%	1,1
WLAN_BT	P32_GPS600_ RSSI60	FRAT_h1,00000 _r1,0_----	x	7	4	2	501	1,5	56%	1,0
WLAN_BT	P50_GPS100	FRAT_h0,00001 _r0,5_----	x	7	6	2	412	1,8	56%	1,2
WLAN_BT	P50_GPS100	FRAT_h1,00000 _r0,5_----	x	7	6	2	408	1,8	56%	1,2
WLAN_BT	P51_GPS200	FRAT_h0,00001 _r0,5_----	x	7	6	2	421	1,9	57%	1,2
WLAN_BT	P51_GPS200	FRAT_h0,00001 _r1,0_----	x	6	5	2	511	2,0	58%	1,0
WLAN_BT	P51_GPS200	FRAT_h1,00000 _r0,5_----	x	7	6	2	404	1,8	57%	1,2
WLAN_BT	P51_GPS200	FRAT_h1,00000 _r1,0_----	x	6	5	2	502	2,1	58%	1,0
WLAN_BT	P53_GPS600	FRAT_h1,00000 _r0,5_----	x	7	5	3	406	1,6	55%	1,1

Die beste Parameterkombination ohne die Nutzung von Ausgangsdaten erhält eine Bewertung von 0,26 Punkten.

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
	P99_Fullfilter	FRAT_h0,00001_r0,5_----		0,26

Hinweis: Datenquelle und Höhe des Hilfswertes sind hier von keiner Bedeutung für die Ergebnisse. Eine Fuzzifizierung kann entsprechend nicht angewandt werden.

Linie 54

Zur Auswertung von Linie 54 stehen 4 der 29 Referenzfahrten zur Verfügung. Anbei sind die besten Ergebnisse, die durch einzelne Parameterkombinationen erreicht werden, differenziert nach Gütemaßen, stichpunktartig aufgeführt:

- GEH-1,5-Grenzwert wird bei maximal 4 von 4 Referenzfahrten eingehalten
- GEH-0,7-Grenzwert wird bei maximal 4 von 4 Referenzfahrten eingehalten
- Fehlerquadratsumme besitzt den Mindestwert von 17

- Summe des Betrags der Abweichung der Verkehrsleistung über alle Entfernungsklassen liegt mindestens bei 10 km über alle betrachteten Referenzfahrten ($n = 4$) der Linie, dies entspricht rund 2,5 km je Referenzfahrt
- Anzahl übereinstimmender Referenzmatrizen mit Einhaltung des GEH-1,5- und GEH-0,7-Grenzwerts liegt maximal bei vier Referenzfahrten
- Anteil Treffer beträgt maximal 85%

Bei der Interpretation ist zu beachten, dass – wie in Kapitel 4.6.4.1 (Gütemaß GEH-Wert) beschrieben – zwei der Referenzfahrten nur eine mögliche Lösungsmatrix besitzen. Die Parameterkombinationen mit den besten Ergebnissen als Summe der normierten und gewichteten Gütemaßbewertungen zeigt Tab. 16. Tab. 17 zeigt die Ergebnisse der originären Gütemaße zu den Parameterkombinationen.

Die besten Parameterkombinationen für Linie 54 erreichen einen Punktwert von 0,86 Punkten, der nächsthöchste Punktwert liegt bei 0,79 und ist nicht in der Tabelle aufgeführt. Diese besten Parameterkombinationen basieren auf WLAN-Daten und nutzen das Filterset P51_GPS200 zusammen mit der Fuzzifizierung. Für die Hochrechnung wird das Fratar-Verfahren mit einem Hilfswert von 1,0000 sowie einem Reduktionsgrenzwert von 0,5 bzw. 1,0 verwendet. Die in Kapitel 4.6.4.2 ermittelten Parameterkombinationen liegen mit mindestens 0,69 Punkten innerhalb der besten 213 aller 1127 Parameterkombinationen dieser Linie – der Maximalwert der drei Parameterkombinationen liegt bei 0,72. Dabei kommen die Verfahren mit Filterset P51 zu besseren Ergebnissen als die mit Filterset P50 oder Filterset P52. Zum Vergleich: Das beste Ergebnis ohne Ausgangsdaten erhält 0,69 Punkte unter Nutzung des Fratar-Verfahrens mit Reduktionsgrenzwert 0,5. Bei der Interpretation ist zu beachten, dass zwei der vier Referenzfahrten nur eine mögliche Lösung besitzen.

Tab. 16: Beste Parameterkombinationen auf Basis der Normierung und Gewichtung nach Tabelle 4-23

Datenquelle	Filterset	HR-Set	Fuzzy	GEH-Merkmal	FQS	Summe des Betrags der Abweichung der Verkehrsleistung in den EKL	Übereinstimmungsmerkmal	Summe (Normierung)
WLAN	P51_GPS200	FRAT_h1,00000_r0,5_----	x	1,00	0,69	0,77	0,70	0,86
WLAN	P51_GPS200	FRAT_h1,00000_r1,0_----	x	1,00	0,69	0,77	0,70	0,86

Tab. 17: Beste Parameterkombinationen mit originären Gütemaßen

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung in den Entfernungsklassen in [100 km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN	P51_GPS200	FRAT_h1,00000_r0,5_----	x	4	4	4	24	0,1	71%	1,0
WLAN	P51_GPS200	FRAT_h1,00000_r1,0_----	x	4	4	4	24	0,1	71%	1,0

Die besten Parameterkombinationen entsprechend Kapitel 4.6.4.2 erhalten hier Bewertungen von:

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
WLAN_BT	P50_GPS100	FRAT_h0,00001_r0,5_----	x	0,69
WLAN_BT	P50_GPS100	FRAT_h0,00001_r1,0_----		0,69
WLAN_BT	P51_GPS200	FRAT_h0,00001_r0,5_----	x	0,72
WLAN_BT	P51_GPS200	FRAT_h0,00001_r1,0_----	x	0,72
WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5_----	x	0,72
WLAN_BT	P51_GPS200	FRAT_h1,00000_r1,0_----	x	0,72
WLAN_BT	P52_GPS400	FRAT_h0,00001_r1,0_----		0,67

Die beste Parameterkombination ohne die Nutzung von Ausgangsdaten erhält eine Bewertung von 0,69 Punkten.

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
	P99_Fullfilter	FRAT_h0,00001_r0,5_----		0,69

Hinweis: Datenquelle und Höhe des Hilfswertes sind hier von keiner Bedeutung für die Ergebnisse. Eine Fuzzifizierung kann entsprechend nicht angewandt werden.

Linie 500

Zur Auswertung von Linie 500 stehen 5 der 29 Referenzfahrten zur Verfügung. Anbei sind die besten Ergebnisse, die durch einzelne Parameterkombinationen erreicht werden, differenziert nach Gütemaßen, stichpunktartig aufgeführt:

- GEH-1,5-Grenzwert wird bei maximal 1 von 5 Referenzfahrten eingehalten
- GEH-0,7-Grenzwert wird bei maximal 4 von 5 Referenzfahrten eingehalten
- Fehlerquadratsumme besitzt den Mindestwert von 142
- Summe des Betrags der Abweichung der Verkehrsleistung über alle Entfernungsklassen liegt mindestens bei 30 km über alle betrachteten Referenzfahrten ($n = 5$) der Linie, dies entspricht rund 6 km je Referenzfahrt
- Anzahl übereinstimmender Referenzmatrizen mit Einhaltung des GEH-1,5- und GEH-0,7-Grenzwerts liegt maximal bei einer Referenzfahrt
- Anteil Treffer beträgt maximal 57%

Die Parameterkombinationen mit den besten Ergebnissen als Summe der normierten und gewichteten Gütemaßbewertungen zeigt Tab. 18. Tab. 19 zeigt die Ergebnisse der originären Gütemaße zu den Parameterkombinationen.

Die beste Parameterkombination für Linie 500 stellt eins der Filtersets, die in Kapitel 4.6.5.2 ermittelt wurden: Filterset P52_GPS400 mit der Datenquelle WLAN- und Bluetooth-Daten. Als HR-Sets wird das Fratar-Verfahren mit Hilfswert 0,00001 und Reduktionsgrenzwert 1,0 genutzt. Neben dieser Parameterkombination kommen auch zwei weitere Parameterkombinationen ebenfalls zu 0,44 Punkten. Die in Kapitel 4.6.4.2 ermittelten Parameterkombinationen liegen mit mindestens 0,27 Punkten innerhalb der besten 143 aller 1127 Parameterkombinationen dieser Linie – der Maximalwert der drei Parameterkombinationen liegt bei 0,44. Zum Vergleich: Das beste Ergebnis ohne

Ausgangsdaten erhält 0,24 Punkte unter Nutzung des Fratar-Verfahrens mit Reduktionsgrenzwert 1,0.

Tab. 18: Beste Parameterkombinationen auf Basis der Normierung und Gewichtung nach Tabelle 4-23

Datenquelle	Filterset	HR-Set	Fuzzy	GEH-Merkmal	FQS	Summe des Betrags der Abweichung der Verkehrsleistung in den EKL	Übereinstimmungsmerkmal	Summe (Normierung)
WLAN_BT	P09_L5_GPS400_RSSI90	FRAT_h0,00001_r1,0_----		0,40	0,25	0,55	0,49	0,44
WLAN_BT	P52_GPS400	FRAT_h0,00001_r1,0_----		0,40	0,28	0,51	0,51	0,44
WLAN_BT	P14_GPS400_RSSI90	FRAT_h0,00001_r1,0_----		0,40	0,26	0,51	0,49	0,44

Tab. 19: Beste Parameterkombinationen mit originären Gütemaßen

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung in den Entfernungsklassen in [100 km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN_BT	P01_L5_GPS200_RSSI90	FRAT_h0,00001_r1,0_----	1	4	1	289	0,8	53%	0,9	
WLAN_BT	P09_L5_GPS400_RSSI90	FRAT_h0,00001_r1,0_----	1	4	1	277	0,8	55%	0,9	
WLAN_BT	P52_GPS400	FRAT_h0,00001_r1,0_----	1	4	1	283	0,8	53%	0,9	

Die besten Parameterkombinationen entsprechend Kapitel 4.6.4.2 erhalten hier Bewertungen von:

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
WLAN_BT	P50_GPS100	FRAT_h0,00001_r0,5_----	x	0,34
WLAN_BT	P50_GPS100	FRAT_h0,00001_r1,0_----		0,36
WLAN_BT	P51_GPS200	FRAT_h0,00001_r0,5_----	x	0,27
WLAN_BT	P51_GPS200	FRAT_h0,00001_r1,0_----	x	0,27
WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5_----	x	0,31
WLAN_BT	P51_GPS200	FRAT_h1,00000_r1,0_----	x	0,34
WLAN_BT	P52_GPS400	FRAT_h0,00001_r1,0_----		0,44

Die beste Parameterkombination ohne die Nutzung von Ausgangsdaten erhält eine Bewertung von 0,24 Punkten.

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
	P99_Fullfilter	FRAT_h0,00001_r1,0_----		0,24

Hinweis: Datenquelle und Höhe des Hilfwertes sind hier von keiner Bedeutung für die Ergebnisse. Eine Fuzzifizierung kann entsprechend nicht angewandt werden.

Anhang 4: Ergebnisdarstellung optimaler Parametereinstellungen nach Nachfrage

Tab. 20 bis Tab. 25 sind die Ergebnisse der Auswertungen auf Basis der realen Fahrgastnachfrage der jeweiligen Referenzfahrt (AFZS-Daten) aufgeführt. Die Nachfrage ist dabei in drei Klassen unterteilt. Die Klassengrenzen für die Nachfrage wurde auf Basis der Quantilsgrenzen (33%, 66%) ermittelt:

- Nachfrageklasse 1: 0 bis 14 Fahrgäste (n = 6)
- Nachfrageklasse 2: 14 bis 21 Fahrgäste (n = 14)
- Nachfrageklasse 3: über 21 Fahrgäste (n = 9)

Auch hier wird auf das in Tabelle 4-23 beschriebene Vorgehen zur Normierung und Gewichtung zurückgegriffen.

Nachfrageklasse 1: 0 bis 14 Fahrgäste

Zur Auswertung der Nachfrageklasse 1 stehen 6 der 29 Referenzfahrten zur Verfügung. Anbei sind die besten Ergebnisse, die durch einzelne Parameterkombinationen erreicht werden, differenziert nach Gütemaßen, stichpunktartig aufgeführt:

- GEH-1,5-Grenzwert wird bei maximal 6 von 6 Referenzfahrten eingehalten
- GEH-0,7-Grenzwert wird bei maximal 5 von 6 Referenzfahrten eingehalten
- Fehlerquadratsumme besitzt den Mindestwert von 21
- Summe des Betrags der Abweichung der Verkehrsleistung über alle Entfernungsklassen liegt mindestens bei 10 km über alle betrachteten Referenzfahrten (n = 6) der Nachfrageklasse, dies entspricht rund 1,7 km je Referenzfahrt
- Anzahl übereinstimmender Referenzmatrizen mit Einhaltung des GEH-1,5- und GEH-0,7-Grenzwerts liegt maximal bei fünf Referenzfahrten
- Anteil Treffer beträgt maximal 83%

Die Parameterkombinationen mit den besten Ergebnissen als Summe der normierten und gewichteten Gütemaßbewertungen zeigt Tab. 20. Tab. 21 zeigt die Ergebnisse der originären Gütemaße zu den Parameterkombinationen.

Die beste Parameterkombination in Nachfrageklasse 1 stellt mit 0,84 Punkten eins der Filtersets, die in Kapitel 4.6.5.2 ermittelt wurden: Filterset P51_GPS200 mit Fuzzifizierung und unter Verwendung von WLAN- und Bluetooth-Daten. Für die Hochrechnung wird das Fratar-Verfahren mit einem Hilfswert von 0,0001 sowie einem Reduktionsgrenzwert von 0,5 verwendet. Zwei weitere Parameterkombinationen kommen zu einer Bewertung von 0,83 Punkten. Sie verwenden ebenfalls die Datenquellen WLAN- und Bluetooth sowie das Filtersets P51_GPS200 mit Fuzzifizierung. Als HR-Sets wird hier das Fratar-Verfahren mit Hilfswert 0,00001 bzw. 1,0000 und Reduktionsgrenzwert 0,5 bzw. 1,0 verwendet. Die in Kapitel 4.6.4.2 ermittelten Parameterkombinationen liegen mit mindestens 0,70 Punkten innerhalb der besten 32 aller 1127 Parameterkombinationen in dieser Nachfrageklasse – der Maximalwert der Parameterkombinationen liegt bei 0,84. Zum Vergleich: Das beste Ergebnis ohne Ausgangsdaten erhält 0,57 Punkte unter Nutzung des Fratar-Verfahrens mit Reduktionsgrenzwert 0,5.

Tab. 20: Beste Parameterkombinationen auf Basis der Normierung und Gewichtung nach Tabelle 4-23

Datenquelle	Filterset	HR-Set	Fuzzy	GEH-Merkmal	FQS	Summe des Betrags der Abweichung der Verkehrsleistung in den EKL	Übereinstimmungsmerkmal	Summe (Normierung)
WLAN_BT	P51_GPS200	FRAT_h0,00001_r0,5_----	x	0,89	0,68	0,78	0,80	0,84
WLAN_BT	P51_GPS200	FRAT_h0,00001_r1,0_----	x	0,89	0,68	0,82	0,77	0,83
WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5_----	x	0,89	0,68	0,73	0,80	0,83

Tab. 21: Beste Parameterkombinationen mit originären Gütemaßen

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleis- tung in den Entfernungsklas- sen in [100 km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN_BT	P51_GPS200	FRAT_h0,00001_ r0,5_----	x	6	5	5	24	0,1	81%	1,0
WLAN_BT	P51_GPS200	FRAT_h0,00001_ r1,0_----	x	6	5	5	24	0,1	81%	0,9
WLAN_BT	P51_GPS200	FRAT_h1,00000_ r0,5_----	x	6	5	5	23	0,2	81%	1,0

Die besten Parameterkombinationen entsprechend Kapitel 4.6.4.2 erhalten hier Bewertungen von:

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
WLAN_BT	P50_GPS100	FRAT_h0,00001_ r0,5_----	x	0,82
WLAN_BT	P50_GPS100	FRAT_h0,00001_ r1,0_----		0,70
WLAN_BT	P51_GPS200	FRAT_h0,00001_ r0,5_----	x	0,84
WLAN_BT	P51_GPS200	FRAT_h0,00001_ r1,0_----	x	0,83
WLAN_BT	P51_GPS200	FRAT_h1,00000_ r0,5_----	x	0,83
WLAN_BT	P51_GPS200	FRAT_h1,00000_ r1,0_----	x	0,82
WLAN_BT	P52_GPS400	FRAT_h0,00001_ r1,0_----		0,70

Die beste Parameterkombination ohne die Nutzung von Ausgangsdaten erhält eine Bewertung von 0,57 Punkten.

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
	P99_Fullfilter	FRAT_h0,00001_ r0,5_----		0,57

Hinweis: Datenquelle und Höhe des Hilfwertes sind hier von keiner Bedeutung für die Ergebnisse. Eine Fuzzifizierung kann entsprechend nicht angewandt werden.

Nachfrageklasse 2: 14 bis 21 Fahrgäste

Zur Auswertung der Nachfrageklasse 2 stehen 14 der 29 Referenzfahrten zur Verfügung. Anbei sind die besten Ergebnisse, die durch einzelne Parameterkombinationen erreicht werden, differenziert nach Gütemaßen, stichpunktartig aufgeführt:

- GEH-1,5-Grenzwert wird bei maximal 7 von 14 Referenzfahrten eingehalten
- GEH-0,7-Grenzwert wird bei maximal 5 von 14 Referenzfahrten eingehalten
- Fehlerquadratsumme besitzt den Mindestwert von 195
- Summe des Betrags der Abweichung der Verkehrsleistung über alle Entfernungsklassen liegt mindestens bei 90 km über alle betrachteten Referenzfahrten ($n = 14$) der Nachfrageklasse, dies entspricht rund 6,5 km je Referenzfahrt
- Anzahl übereinstimmender Referenzmatrizen mit Einhaltung des GEH-1,5- und GEH-0,7-Grenzwerts liegt maximal bei vier Referenzfahrten
- Anteil Treffer beträgt maximal 58%

Die Parameterkombinationen mit den besten Ergebnissen als Summe der normierten und gewichteten Gütemaßbewertungen zeigt Tab. 22. Tab. 23 zeigt die Ergebnisse der originären Gütemaße zu den Parameterkombinationen.

In Nachfrageklasse 2 erreicht eine Parameterkombination den höchsten Punktwert von 0,39 Punkten. Diese Parameterkombination basiert auf WLAN- und Bluetooth-Daten und nutzt das Filterset P32_GPS600_RSSI60 mit Fuzzifizierung. Als Hochrechnungsverfahren wird das Fratar-Verfahren mit einem Hilfswert von 1,0000 und einem Reduktionsgrenzwert von 1,0 gewählt. Zwei weitere Parameterkombinationen erreichen 0,38 Punkte. Die in Kapitel 4.6.4.2 ermittelten Parameterkombinationen liegen mit mindestens 0,31 Punkten innerhalb der besten 157 aller 1127 Parameterkombinationen in dieser Nachfrageklasse – der Maximalwert der Parameterkombinationen liegt bei 0,36. Zum Vergleich: Das beste Ergebnis ohne Ausgangsdaten erhält 0,26 Punkte unter Nutzung des Fratar-Verfahrens mit Reduktionsgrenzwert 0,5.

Tab. 22: Beste Parameterkombinationen auf Basis der Normierung und Gewichtung nach Tabelle 4-23

Datenquelle	Filterset	HR-Set	Fuzzy	GEH-Merkmal	FQS	Summe des Betrags der Abweichung der Verkehrsleistung in den EKL	Übereinstimmungsmerkmal	Summe (Normierung)
WLAN_BT	P32_GPS600_RSSI60	FRAT_h1,00000_r1,0_----	x	0,26	0,27	0,57	0,53	0,39
WLAN_BT	P32_GPS600_RSSI60	FRAT_h1,00000_r0,5_----	x	0,29	0,35	0,56	0,45	0,38
WLAN	P44_L5_GPS600_RSSI70	FRAT_h1,00000_r1,0_----	x	0,26	0,20	0,50	0,54	0,38

Tab. 23: Beste Parameterkombinationen mit originären Gütemaßen

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung in den Entfernungsklassen in [100 km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN_BT	P32_GPS600_RSSI60	FRAT_h1,00000_r1,0_----	x	4	4	3	323	0,9	55%	1,0
WLAN_BT	P32_GPS600_RSSI60	FRAT_h1,00000_r0,5_----	x	5	4	3	291	0,9	53%	1,1
WLAN	P44_L5_GPS600_RSSI70	FRAT_h1,00000_r1,0_----	x	5	3	3	357	1,1	56%	1,0

Die besten Parameterkombinationen entsprechend Kapitel 4.6.4.2 erhalten hier Bewertungen von:

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
WLAN_BT	P50_GPS100	FRAT_h0,00001_r0,5_----	x	0,34
WLAN_BT	P50_GPS100	FRAT_h0,00001_r1,0_----		0,36
WLAN_BT	P51_GPS200	FRAT_h0,00001_r0,5_----	x	0,34
WLAN_BT	P51_GPS200	FRAT_h0,00001_r1,0_----	x	0,31
WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5_----	x	0,34
WLAN_BT	P51_GPS200	FRAT_h1,00000_r1,0_----	x	0,31
WLAN_BT	P52_GPS400	FRAT_h0,00001_r1,0_----		0,33

Die beste Parameterkombination ohne die Nutzung von Ausgangsdaten erhält eine Bewertung von 0,26 Punkten.

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
	P99_Fullfilter	FRAT_h0,00001_r0,5_----		0,26

Hinweis: Datenquelle und Höhe des Hilfswertes sind hier von keiner Bedeutung für die Ergebnisse. Eine Fuzzifizierung kann entsprechend nicht angewandt werden.

Nachfrageklasse 3: über 21 Fahrgäste

Zur Auswertung der Nachfrageklasse 3 stehen 9 der 29 Referenzfahrten zur Verfügung. Anbei sind die besten Ergebnisse, die durch einzelne Parameterkombinationen erreicht werden, differenziert nach Gütemaßen, stichpunktartig aufgeführt:

- GEH-1,5-Grenzwert wird bei maximal 2 von 9 Referenzfahrten eingehalten
- GEH-0,7-Grenzwert wird bei maximal 4 von 9 Referenzfahrten eingehalten
- Fehlerquadratsumme besitzt den Mindestwert von 221
- Summe des Betrags der Abweichung der Verkehrsleistung über alle Entfernungsklassen liegt mindestens bei 90 km über alle betrachteten Referenzfahrten (n = 9) der Nachfrageklasse, dies entspricht rund 10 km je Referenzfahrt
- Anzahl übereinstimmender Referenzmatrizen mit Einhaltung des GEH-1,5- und GEH-0,7-Grenzwerts liegt maximal bei einer Referenzfahrt
- Anteil Treffer beträgt maximal 57%

Die Parameterkombinationen mit den besten Ergebnissen als Summe der normierten und gewichteten Gütemaßbewertungen zeigt Tab. 24. Tab. 25 zeigt die Ergebnisse der originären Gütemaße zu den Parameterkombinationen.

In Nachfrageklasse 3 kommen drei Parameterkombinationen mit einem Punktwert von 0,34 zu den besten Ergebnissen. Diese Parameterkombinationen basieren durchweg auf WLAN- und Bluetooth-Daten und nutzen die Filtersets P52_GPS400, P01_L5_GPS200_RSSI90 und P09_L5_GPS400_RSSI90. Als HR-Sets wird hier das Fratar-Verfahren mit Hilfswert 1,0000 und Reduktionsgrenzwert 1,0 bzw. 0,5 genutzt. Die in Kapitel 4.6.4.2 ermittelten Parameterkombinationen liegen mit mindestens 0,25 Punkten innerhalb der besten 165 aller 1127 Parameterkombinationen in dieser Nachfrageklasse – der Maximalwert der drei Parameterkombinationen liegt bei 0,34. Zum

Vergleich: Das beste Ergebnis ohne Ausgangsdaten erhält 0,24 Punkte unter Nutzung des Fratar-Verfahrens mit Reduktionsgrenzwert 1,0.

Tab. 24: Beste Parameterkombinationen auf Basis der Normierung und Gewichtung nach Tabelle 4-23

Datenquelle	Filterset	HR-Set	Fuzzy	GEH-Merkmal	FQS	Summe des Betrags der Abweichung der Verkehrsleistung in den EKL	Übereinstimmungsmerkmal	Summe (Normierung)
WLAN_BT	P01_L5_GPS200_RSSI90	FRAT_h1,00000_r0,5_----		0,22	0,15	0,43	0,50	0,34
WLAN_BT	P09_L5_GPS400_RSSI90	FRAT_h1,00000_r1,0_----		0,22	0,15	0,49	0,48	0,34
WLAN_BT	P52_GPS400	FRAT_h1,00000_r1,0_----		0,22	0,19	0,44	0,49	0,34

Tab. 25: Beste Parameterkombinationen mit originären Gütemaßen

Datenquelle	Filterset	HR-Set	Fuzzy	GEH1,5	GEH0,7	Übereinstimmung (GEH1,5 und GEH0,7)	FQS	Abweichung der Verkehrsleistung in den Entfernungsklassen in [100 km]	mittlerer Anteil Treffer	mittlerer Überfüllungsfaktor
WLAN_BT	P01_L5_GPS200_RSSI90	FRAT_h1,00000_r0,5_----	1	4	1	481	1,6	55%	0,9	
WLAN_BT	P09_L5_GPS400_RSSI90	FRAT_h1,00000_r1,0_----	1	4	1	485	1,4	53%	0,9	
WLAN_BT	P52_GPS400	FRAT_h1,00000_r1,0_----	1	4	1	457	1,6	53%	0,9	

Die besten Parameterkombinationen entsprechend Kapitel 4.6.4.2 erhalten hier Bewertungen von:

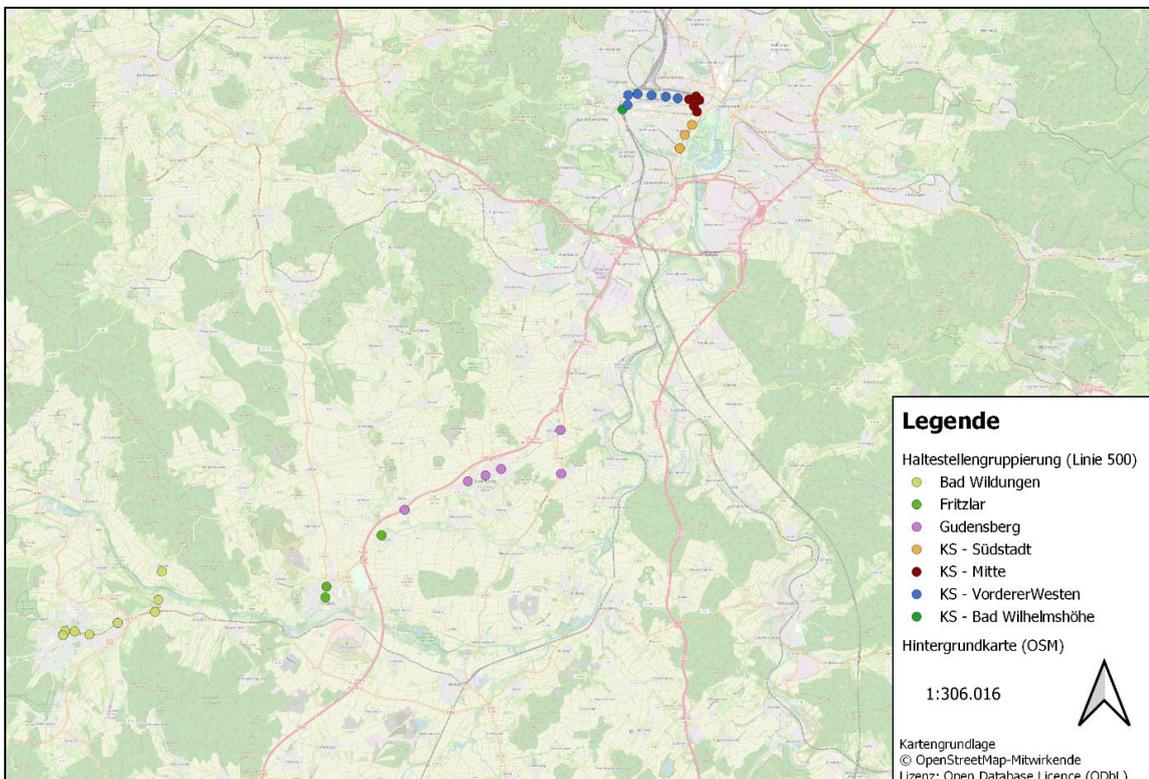
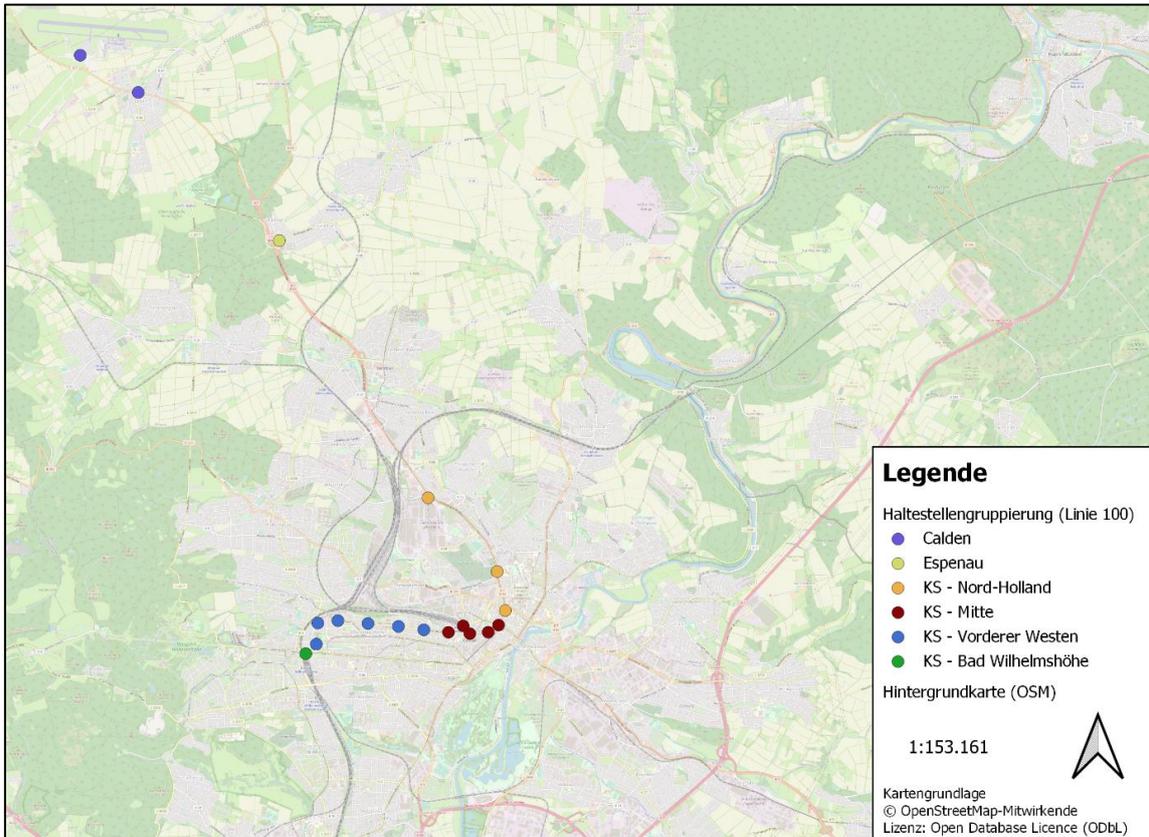
Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
WLAN_BT	P50_GPS100	FRAT_h0,00001_r0,5_----	x	0,28
WLAN_BT	P50_GPS100	FRAT_h0,00001_r1,0_----		0,29
WLAN_BT	P51_GPS200	FRAT_h0,00001_r0,5_----	x	0,25
WLAN_BT	P51_GPS200	FRAT_h0,00001_r1,0_----	x	0,26
WLAN_BT	P51_GPS200	FRAT_h1,00000_r0,5_----	x	0,27
WLAN_BT	P51_GPS200	FRAT_h1,00000_r1,0_----	x	0,30
WLAN_BT	P52_GPS400	FRAT_h0,00001_r1,0_----		0,34

Die beste Parameterkombination ohne Nutzung von Ausgangsdaten erhält die Bewertung von 0,24 Punkten.

Datenquelle	Filterset	HR-Set	Fuzzy	Summe (Normierung)
	P99_Fullfilter	FRAT_h0,00001_r1,0_----		0,24

Hinweis: Datenquelle und Höhe des Hilfwertes sind hier von keiner Bedeutung für die Ergebnisse. Eine Fuzzifizierung kann entsprechend nicht angewandt werden.

Anhang 5: Karten räumliche Aggregation (oben: Linie 100, unten: Linie 500)



Schriftenreihe des Institutes für Verkehrswesen

In der Schriftenreihe Verkehr des Institutes (bis zum Jahr 2002 des Fachgebietes) erschienen bisher folgende Hefte:

Heft Nr.:	Inhalt	Preis [€]
1	Köhler, Uwe: Verkehrsplanung im Widerstreit der Meinungen, 1989	(vergriffen)
2	1. Kasseler Nahverkehrs-Tage: Der ÖPNV zwischen Stadt und Umland, 1990	6,00
3	2. Kasseler Nahverkehrstage: Verknüpfungen von Fern- und Nahverkehr, 1992	6,00
4	Arndt, Karin: Entwicklung eines Verfahrens zur Abschätzung einer potentiellen P+R-Nachfrage im Berufsverkehr, 1993	9,00
5	3. Kasseler Nahverkehrstage: Stößt der öffentliche Verkehr an seine Grenzen ?, 1994	9,00
6	4. Kasseler Nahverkehrstage: Beiträge zum Stadt- und Regionalverkehr, 1996	12,00
7	Strauß, Susanne: City-Logistik - Ein Instrument zur Verringerung des städtischen Güterverkehrs, 1997	(vergriffen)
8	Klein, Angelika: Die ÖPNV Grundgebühr – Ein Instrument zur zukunftsorientierten Gestaltung und Finanzierung des öffentlichen Personennahverkehrs, 1998	12,00
9	5. Kasseler Nahverkehrstage: Nahverkehrspläne - eine erste Bilanz, 1998	12,00
10	6. Kasseler Nahverkehrstage: 10 Jahre Fachgebiet Verkehrssysteme und Verkehrsplanung, 2000	15,00
11	Klotz, Günther: Bahnbaubau und Bürgerbeteiligung - Eine Untersuchung über die Beteiligung der Bürger bei der Planung von Bahnanlagen am Beispiel der Neubaustrecke der deutschen Bundesbahn im Raum Kassel, 2002	15,00
12	7. Kasseler Nahverkehrstage: Wettbewerb im öffentlichen Verkehr, 2002	15,00
13	Zöllner, Ralf: Einsatzbereiche von Schienenregionalbahnen, 2002	15,00
14	Listl, Gerhard: Anwendung neuer Technologien zur Erfassung des Verkehrsablaufs, 2003	22,00
15	Ober-Sundermeier, Anja: Entwicklung eines Verfahrens zur Stauprognose an Engpässen auf Autobahnen unter besonderer Berücksichtigung von Arbeitsstellen, 2003	18,00

Heft Nr.:	Inhalt	Preis [€]
16	Winter, Olaf Markus: Analyse und Evaluation von Nahverkehrsplänen und die Aufstellung von Kriterien zur Bewertung von Standards im ÖPNV, 2005	(vergriffen)
17	Institut für Verkehrswesen (Hrsg.): Von den Anfängen bis zur Gegenwart - Verkehrstechnik an der Uni Kassel, Kolloquium anlässlich der Verabschiedung von Prof. Dr.-Ing. Heinz Zackor am 14. Juli 2005	18,00
18	Kobbeloer, Detlef: Dezentrale Steuerung von Lichtsignalanlagen in urbanen Verkehrsnetzen, 2007	29,00
19	Bertocchi, Timo: Einsatzbereiche von ÖPNV-Bedienungsformen im ländlichen Raum, 2009	18,00
20	Herzog, Torsten: Strategien und Potenziale zur Verbrauchsreduzierung bei Verkehrsstaus	29,00
21	Otto, Thomas: Kooperative Verkehrsbeeinflussung und Verkehrssteuerung an signalisierten Knotenpunkten	18,00
22	Leitzke, Christian: Reisezeitermittlung im motorisierten Individualverkehr mit Hilfe drahtloser Kommunikationstechnologien	29,00
23	Kühnel, Carsten: Verkehrsdatenerfassung mittels Floating Car Observer auf zweistreifigen Landstraßen	29,00
24	Nahverkehrs-Tage 2013: Neue Konzepte für Stadt und Land	24,00
25	Nahverkehrs-Tage 2015: Nahverkehr und Freizeit - Neue Kunden durch innovative Konzepte gewinnen	(vergriffen)
26	Rudolph, Felix: Qualitätsanalyse koordinierter Lichtsignalsteuerungen unter Verwendung kooperativ gewonnener Messgrößen, 2016	34,00
27	Weisheit, Toni: Ein Verfahren zur Prognose verkehrsabhängiger Schaltzeiten von Lichtsignalanlagen, 2017	34,00
28	Nahverkehrs-Tage 2017: Digital und Disruptiv - Neue Daten und Methoden für einen kundengerechten ÖPNV	29,00
29	Knese, Dennis: Integration der Elektromobilität in die Stadtplanung und Straßenraum- gestaltung – Lösungsansätze für Strategien, Konzepte und Maßnahmen,	39,00
30	Nahverkehrs-Tage 2019: Finanzierung des öffentlichen Verkehrs: zwischen leistungsorientierten E-Tarifen, preisgünstigen Flats und Drittnutzerfinanzierung	29,00

Schriftenreihe des Institutes für Verkehrswesen

Heft Nr.:	Inhalt	Preis [€]
31	Schmitt, Volker: In den ÖPNV integrierte Mitnahmesysteme im ländlichen Raum - Wirkungsanalyse und Bewertung	34,00
32	Weidemann, Tanja: Prozessdatenanalyse verkehrsabhängiger Lichtsignalsteuerungen zur Modellierung von Ersatzsteuerungen für 2-Phasen-Systeme	34,00
33	Saighani, Assadollah: Bewertungsverfahren für einen ökonomischen Vergleich städtischer Verkehrssysteme	39,00
34	Lambrecht, Franz: Entwicklung eines Verkehrssicherheitsprogramms für Radfahrende zwischen 11 und 14 Jahren mit dem Fokus auf metakognitiven Fähigkeiten	39,00
35	Bieland, Dominik: Ableitung von Quelle-Ziel-Matrizen im ÖPNV auf Basis von WLAN- und Bluetooth-Daten	39,00

Hefte 1 bis 13 und 16 sind zu bestellen über das:
Fachgebiet Verkehrsplanung und Verkehrssysteme
Universität Kassel
34109 Kassel
Fax-Nr. (0561) 804-3382
E-Mail: k.hager@uni-kassel.de

Hefte 14 und 15, sowie ab Heft 17 sind zu bestellen über die:
kassel university press
E-Mail: order.kup@uni-kassel.de

ISBN 978-3-7376-1120-6



9 783737 611206 >