Anwendung von Unsicherheitsmodellen am Beispiel der Verkehrserfassung unter Nutzung von Mobilfunkdaten

Von der Fakultät Luft- und Raumfahrttechnik und Geodäsie
der Universität Stuttgart zur Erlangung der Würde eines
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
genehmigte Abhandlung

Vorgelegt von

Dipl.-Ing. Ralf Borchers

aus Magdeburg

Hauptberichter: Prof. Dr.-Ing. habil. Volker Schwieger

Mitberichter: Prof. Dipl.-Ing. Dr. techn. Andreas Wieser

Prof. Dr.-Ing. Markus Friedrich

Tag der mündlichen Prüfung: 28.06.2012

Institut für Ingenieurgeodäsie der Universität Stuttgart

2013

FÜR MAREN SOPHIE, BEEKE ANNA UND SYLVIA

UND MEINEN KURZ VOR FERTIGSTELLUNG DIESER ARBEIT VERSTORBENEN VATER

Inhaltsverzeichnis

Kurzfassung **7**

Abstract **8**

1 Einführung **9**

1.1 Motivation und Zielsetzung 9

1.2 Vorgehensweise 10

2 Grundlagen ausgewählter Verkehrserfassungssysteme **12**

2.1 Einteilung des Verkehrs 12

2.2 Verfahren zur Verkehrserfassung und Verkehrsmittelidentifizierung 13

2.2.1 Verkehrserfassungssysteme mit infrastrukturgebundenen Sensoren 13

2.2.2 Verkehrserfassungssysteme mit mobilen Sensoren 14

2.3 Verkehrsdatenerfassung im Projekt DO-IT 16

3 Modellierung von Unsicherheiten **18**

3.1 Motivation 18

3.2 Grundbegriffe der Unsicherheiten 18

3.3 Mathematische Behandlung von Unsicherheiten 21

3.3.1 Wahrscheinlichkeitstheorie und zufällige Variabilität 21

3.3.2 Fuzzy-Theorie 27

3.3.3 Fuzzy-Randomness 43

4 Modellierung unsicherer Verkehrsdaten **50**

4.1 Einleitung 50

4.2 Modellierung unsicherer Ortungsinformationen 51

4.2.1 Rohdatengewinnung 51

4.2.2 Mobilfunkortung nach dem Signalpegel-Matching-Verfahren 51

4.2.3 Herleitung der Unsicherheiten für das Signalpegel-Matching mit Pegeldifferenzen 57

4.2.4 Datenpunktfusionierung mit FCM-Clustering 63

4.2.5 Verkehrsnetze und Kacheln digitaler Verkehrsnetze 70

4.2.6 Unsicherheitsgrade für den Aufenthalt in Kacheln 72

4.3 Bewegungsidentifikation 76

4.3.1 Szenarien der Bewegung 77

4.3.2 Bewegungsidentifikation mittels zufälliger Variabilität 78

4.3.3 Bewegungsidentifikation mittels Unschärfe 80

4.3.4 Bewegungsidentifikation mittels Fuzzy-Randomness 81

4.4 Identifikation des Verkehrsmittels anhand der Geschwindigkeit 83

4.4.1 Zufällige Variabilität des Verkehrsmittel-Ausschlusses 84

4.4.2 Unschärfe des Verkehrsmittel-Ausschlusses 84

4.4.3 Fuzzy-Randomness des Verkehrsmittel-Ausschlusses 85

4.5 Korrelation zu ÖV-Fahrzeugen 86

4.5.1 Georeferenzierung der Fahrplandaten 86

4.5.2 Prognose der Anwesenheitszeiten von ÖV-Fahrzeugen 89

4.5.3 Modellierung von Fahrplanabweichungen 90

4.6 Kartengestützte Plausibilitätsüberprüfung 97

4.6.1 Ein Verfahren zur Trajektoriensuche mit Raum- und Zeitrestriktionen 97

4.6.2 Auswahl der Teilgraphen und Bestimmung der Zeitfenster 99

4.6.3 Anpassung der Maximalgeschwindigkeiten der Verkehrsnetzgraphen 103

4.6.4 Optimierungsverfahren zur Auswahl der Verkehrsnetzteilgraphen 103

4.7 Verknüpfung der Identifizierungsergebnisse 104

5 Softwaretechnische Umsetzung 106
5.1 Erfassung und Anonymisierung der Mobilfunkdaten 106
5.2 Datenvorprozessierung und Datenhaltung 107
5.3 Umsetzung der Mobilfunkortung und der Identifizierungsverfahren 108

6 Validierung der Identifizierungsverfahren 109
6.1 Vorgehensweise 109
6.2 Versuchsanordnung 110
6.2.1 Versuchsgebiet 110
6.2.2 Geometrische Unsicherheiten des Signalpegel-Matchings 112
6.3 Kalibrierung und Evaluation der Identifizierungsverfahren 116
6.3.1 Bewegungsidentifikation 116
6.3.2 Identifikation des Verkehrsmittels anhand der Geschwindigkeit 117
6.3.3 Korrelation zu ÖV-Fahrzeugen 119
6.3.4 Kartengestützte Plausibilitätsprüfung 121
6.3.5 Zusammenfassung der Korrektheitsgrade und Effektivitätswerte 124
6.3.6 Zusammenfassung zur Kalibrierung und Evaluation der Identifikationsverfahren 125
6.4 Evaluation anhand anonymisierter Mobilfunkdaten 127
6.4.1 Vorgehensweise 127
6.4.2 Identifizierung anonymisierter Mobilfunkdaten 127
6.4.3 Verknüpfung der Identifikationsergebnisse 133

7 Zusammenfassung und Ausblick 136
7.1 Zusammenfassung 136
7.2 Ausblick 138

Literatur 139

Abkürzungsverzeichnis 144

Anhang 145
A Datentypen der Mobilfunkdaten für das Signalpegel-Matching 145
B Kennzahlen der Mobilfunkortung anonymisierter Mobilfunkdaten 146
C Integral oder numerische Integration für die card-Funktion 147
D Erhebung und Analyse von Fahrplanabweichungen 148
D.1 Linienbusse 148
D.2 Straßenbahnen 149
D.3 Züge der DB AG 150

Lebenslauf 153

Dank 154

Kurzfassung

Das hohe und stetig anwachsende Verkehrsaufkommen ist häufig Ursache für Überlastungen der Verkehrswege und deren Folgeerscheinungen, wie z. B. Verkehrsstaus und Unfälle. Mit Einführung von Verkehrsmanagement-systemen wird angestrebt, Effizienzverbesserungen des Verkehrsflusses auf den bestehenden Verkehrswegen zu erzielen. Als wesentliche Datengrundlage setzen Verkehrsmanagementsysteme Verkehrsdaten voraus, wie sie mittels Induktionsschleifen oder sogenannter Floating Car Data gewonnen werden können. Vor dem Hintergrund, dass Mobilfunkgeräte geortet werden können, eignen sich die als Floating Phone Data bezeichneten und aus Mobilfunkdaten hergeleiteten Verkehrsdaten ebenfalls zur Verkehrserfassung. Dieser auch wirtschaftlich attraktiven Verkehrserfassungsart stehen hohe Ortungsunsicherheiten von mehreren einhundert Metern sowie die fehlende Information, ob und in welchem Verkehrsmittel das Mobilfunkgerät mitgeführt wurde, gegenüber.

In der vorliegenden Arbeit wird eine Mobilfunkortung basierend auf dem Signalpegel-Matching eingesetzt, die die gemessenen Signalpegel der Mobilfunkgeräte des GSM-Mobilfunknetzes mit Referenzsignalpegelkarten vergleicht. Die zufälligen, systematischen aber auch unbekannt wirkenden Unsicherheiten werden mit Hilfe der zufälligen Variabilität, der Fuzzy-Theorie und der Fuzzy-Randomness Methodik modelliert. Im Anschluss werden Identifikationsverfahren vorgestellt, mit denen in Verkehrsmitteln des Motorisierten Individualverkehrs (z. B. in PKW oder LKW) generierte Mobilfunkdaten aus anonymisierten Mobilfunkdaten identifiziert werden können.

Zu Beginn wird geprüft, ob sich das Mobilfunkgerät in Bewegung befindet. Bewegt es sich, wird nachfolgend dessen Geschwindigkeit als Entscheidung bezüglich eines Verkehrsmittels herangezogen. Hintergrund ist, dass bauartbedingte oder administrative Gründe die Höchstgeschwindigkeit von Verkehrsmitteln begrenzen. Ist die Geschwindigkeit des Mobilfunkgerätes signifikant höher als die Höchstgeschwindigkeit des untersuchten Verkehrsmittels, kann dieses Verkehrsmittel ausgeschlossen werden. Da in öffentlichen Verkehrsmitteln generierte Mobilfunkdaten für die Erfassung des Motorisierten Individualverkehrs ungeeignet sind, werden sie im nächsten Schritt eliminiert. Aus Fahrplänen werden hierfür die Positionen der Fahrzeuge des Öffentlichen Verkehrs (z. B. Linienbusse, Straßenbahnen) prognostiziert und mit den Positionen des Mobilfunkgerätes sowohl zeitlich als auch räumlich verglichen. Abschließend wird geprüft, ob für die Positionsfolge des Mobilfunkgerätes eine Trajektorie auf dem Verkehrsnetzgraph des motorisierten Individualverkehrs (Straßennetz) generiert werden kann. Kann die Positionsfolge in den Verkehrsnetzgraph räumlich, topologisch und zeitlich eingepasst werden, ist sie grundsätzlich für die Verkehrslageerfassung des motorisierten Individualverkehrs geeignet.

Für stehende Mobilfunkgeräte ist in der Regel keine eindeutige Identifikation des Verkehrsmittels möglich, da jedes Verkehrsmittel stehen kann. Eine Unterscheidung zwischen Mobilfunkgeräten in Verkehrsstaus und beispielsweise nicht am Verkehr beteiligten Mobilfunkgeräten (z. B. stehende Fußgänger) ist infolgedessen zunächst nicht möglich. Diese Problemstellung wurde durch die Verknüpfung aktueller und vorgehender Identifizierungsergebnisse gelöst.

Um ihre Eignung und ihr Potential zu vergleichen, wurden die Identifikationsverfahren mit konsequenter Anwendung der mathematischen Unsicherheitsmodelle der zufälligen Variabilität, der Fuzzy-Theorie und des Fuzzy-Randomness entwickelt und softwaretechnisch umgesetzt. Die entwickelten Identifikationsverfahren wurden unter Verwendung realer Mobilfunkdaten validiert und evaluiert. Das auf der Fuzzy-Randomness Methodik basierende Identifikationsverfahren, ergab sowohl qualitativ als auch quantitativ die besten Identifikationsergebnisse.

Abstract

The high and steadily growing volume of traffic is frequently the cause for excessive strains on traffic routes and resulting consequences, such as traffic jams and accidents. The use of traffic management systems aims to achieve improvements in the efficiency of traffic flow on the existing traffic routes. Traffic management systems require traffic information, which can be obtained from induction loops or from so-called floating car data. Since cellular phones can be located these so-called floating phone data is suitable for traffic detection as well. This method of traffic detection, although economically attractive, is accompanied by a high level of imprecision in the localisation of up to several hundred metres as well as by missing information about whether and in which vehicle the cellular phones were transported.

In this paper, mobile positioning is based on a signal level matching method, wherein the position of the cellular phone is determined by use of the received signal strength at the mobile phone and matched to a signal strength map of the network provider. The random and systematic uncertainties are modelled alternatively using random variability, fuzzy theory or fuzzy randomness. In the following identification methods are presented, that indentify mobile phone data generated in motorized individual vehicles (e.g. in cars or trucks).

At the beginning, it is checked whether the cellular phone is in motion. If it moves, its velocity is taken to identity the means of transport. Background for this decision is that design-related or administrative reasons limit the speed of some means of transport. Some means of transport can be excluded if significantly higher speeds exist for the cellular phone. As mobile data generated in public transport are unsuitable for the detection of individual vehicles, these are filtered out in the next step. From timetables the positions of the vehicles of the public transport (e.g., busses, trams) are predicted and compared with the positions of the cellular phone both in time and space. Finally, it is checked whether for the positions of the cellular phone a trajectory on the graph of motorized individual transport (road network) can be generated. If the sequence of positions in the transport network graph can be fitted spatially, temporally and topologically, it is basically suitable for the detection of motorized individual traffic.

For static mobile phones no unique identification of the means of transport is generally possible, because each means of transport can temporally immobile. Mobile phones in traffic jams and mobile phones, which are not participating at traffic (e.g. immobile pedestrians), can initially not be distinguished. This problem was solved by linking current and previous identification results.

In order to compare their suitability and their potential, the identification methods were developed and implemented in software with the mathematical uncertainty models of random variability, the fuzzy theory, and the theory of fuzzy randomness. The developed identification methods were validated and evaluated using real anonymous mobile radio data. The best identification results in terms of both quality and quantity were achieved with using the uncertainty models based on fuzzy randomness.

1 Einführung

1.1 Motivation und Zielsetzung

Das hohe und stetig anwachsende Verkehrsaufkommen ist häufig Ursache für Überlastungen der Verkehrswege und deren Folgeerscheinungen, wie z. B. Verkehrsstaus und Unfälle. Neben einem weiteren Ausbau bestehender oder dem Neubau neuer Verkehrswege dienen insbesondere Verkehrsmanagementsysteme der Effizienzverbesserung des Verkehrsflusses. Als Datengrundlage setzen sie aktuelle Verkehrslagedaten (z. B. aus Mobilitätsdiensten) in Verbindung mit historischen Daten (z. B. zeitlich wiederkehrende Verkehrsphänomene, wie stockender Verkehr oder Stau zu Schulferienbeginn) voraus.

Zur Erfassung der aktuellen Verkehrslage existieren im Wesentlichen infrastrukturgebundene Sensoren auf Autobahnen und in städtischen Verkehrsnetzen (z. B. Induktionsschleifen, CCD-Kameras[1]) oder im Verkehr „mitschwimmende Sensoren“, die zumeist auf dem Ortungssystem des NAVSTAR **G**lobal **P**ositioning **S**ystem[2] (GPS) basierend als **F**loating **C**ar **D**ata (FCD) bezeichnet werden. Mit dem Hintergrund, dass Mobilfunkgeräte geortet werden können, sind sie zur Verkehrslageerfassung ebenfalls als mitschwimmende Sensoren einsetzbar. Mit ihnen erfasste Verkehrslageinformationen nennt man in Analogie zu den FCD auch **F**loating **P**hone **D**ata (FPD) (SCHWIEGER 2008; WILTSCHKO ET AL. 2007).

Im Jahr 2006 waren in Deutschland ca. 85,7 Millionen (EU.DESTATIS 2011) Mobilfunkgeräte angemeldet. Bezogen auf die Bevölkerungsanzahl von ca. 82,3 Millionen (DESTATIS 2011b) im selben Jahr, besaß damit statistisch gesehen, jeder Deutsche mindestens ein Mobilfunkgerät. Für die Verkehrslageerfassung wird mit Annahme, dass die Mobilfunkgeräte bei Reisen auf den Verkehrsnetzen mittransportiert werden, den Mobilfunkdaten ein hoher Stellenwert eingeräumt. Mit Mobilfunkdaten wird dadurch eine flächendeckende Verkehrsdatenerfassung auf wirtschaftlich attraktive Weise erreicht. Aus datenschutzrechtlichen Gründen ist jedoch ohne ausdrückliche Zustimmung des Mobilfunkteilnehmers die Erfassung, Verarbeitung und Speicherung von personenbezogenen Daten, wozu auch die Ortsinformationen aus der Mobilfunkortung zählen, nicht zulässig. Um dennoch eine große Anzahl von Mobilfunkdaten nutzen zu können, werden die Mobilfunkdaten anonymisiert.

Für die Verkehrslageerfassung mittels Mobilfunkdaten spricht somit die große Anzahl von Sensoren und die kostengünstige Datengewinnung, da die bestehende Infrastruktur der Mobilfunksysteme sowie die in der Bevölkerung vorhandenen Mobilfunkgeräte genutzt werden können. Aufgrund der anonymisierten Mobilfunkdaten ist jedoch die Fragestellung zu lösen, ob das Mobilfunkgerät in einem für die Verkehrslageerfassung relevanten Verkehrsmittel transportiert wurde. Beispielsweise sind für die Verkehrslage des **M**otorisierten **I**ndividual**v**erkehrs (MIV) die in den Fahrzeugen des **Ö**ffentlichen **V**erkehrs (ÖV) erzeugten Mobilfunkdaten zu identifizieren, da sie anderenfalls zu falschen Schlussfolgerungen führen würden. Zur Lösung dieses Problems sind die hohen metrischen Ortungsunsicherheiten aus der Mobilfunkortung von bis zu mehreren einhundert Metern zu berücksichtigen (SCHOLLMEYER & WILTSCHKO 2007).

Vor diesem Hintergrund ist es das Ziel der vorliegenden Arbeit, Verfahren zu entwickeln, mit denen das vom Mobilfunkteilnehmer verwendete Verkehrsmittel identifizieren werden kann. Hierfür sollen ausschließlich anonymisierte, nicht personenbezogene Ortungsdaten eingesetzt werden. Des Weiteren sind die

[1] Charge-coupled Devices (CCD) sind lichtempfindliche elektronische Bauelemente, die auf dem Prinzip des inneren Photoeffektes beruhen

[2] NAVSTAR GPS (Navigational Satellite Timing and Ranging - Global Positioning System)

Ortungsunsicherheiten mit geeigneten Unsicherheitsmodellen zu quantifizieren und entsprechend fortzupflanzen. Die entwickelten Verfahren sind softwaretechnisch umzusetzen und die Eignung anhand empirisch erfasster Mobilfunkdaten zu validieren und zu evaluieren.

1.2 Vorgehensweise

Im Abschnitt 2 wird zunächst ein Überblick gegeben, welche Methoden zur Verkehrsdatenerfassung und der damit verbundenen Verkehrsmittelidentifizierung in wissenschaftlichen Arbeiten existieren. Aufgrund des hohen Nutzungsgrades nimmt die Verkehrserfassung des MIV einen besonderen Stellenwert in Verkehrsmanagementsystemen ein. Dem aktuellen Trend entsprechend, diese Datengewinnung über die Ortung von Mobilfunkgeräten voranzutreiben, werden in der Literatur veröffentlichte Methoden zur **Identifizierung der im MIV befindlichen Mobilfunkgeräte** vorstellt.

Den Mobilfunkdaten anhaftende Unsicherheiten pflanzen sich unweigerlich auf die Qualität der Identifizierungsergebnisse fort. Kenntnis, Modellierung und Fortpflanzung der wirkenden Unsicherheiten sind somit für eine sinnvolle Beurteilung der Schlussfolgerungen unumgänglich. Insbesondere die Kenntnis der Unsicherheiten der Mobilfunkdaten stellt jedoch ein Problem dar, da die Unsicherheitseinflüsse auf die Mobilfunkdaten vielfältig und zum Teil nur näherungsweise bekannt sind. In Abschnitt 3 werden neben der einleitenden Klärung wesentlicher, für das Verständnis notwendiger Grundbegriffe, die mathematischen Grundbausteine für die Behandlung von Unsicherheiten zusammengefasst. Beschrieben werden die **zufällige Variabilität** (Stochastik) und die **Fuzzy-Theorie**, die beide im messtechnischen Umfeld häufig Anwendung finden, sowie das **Fuzzy-Randomness**, das eine Kombination der beiden erstgenannten Modellierungsarten darstellt. Mit dem Ziel die Eignung der jeweiligen Unsicherheitsmodellierung vergleichen und beurteilen zu können, werden sie jeweils auf die Identifizierungsverfahren angewendet.

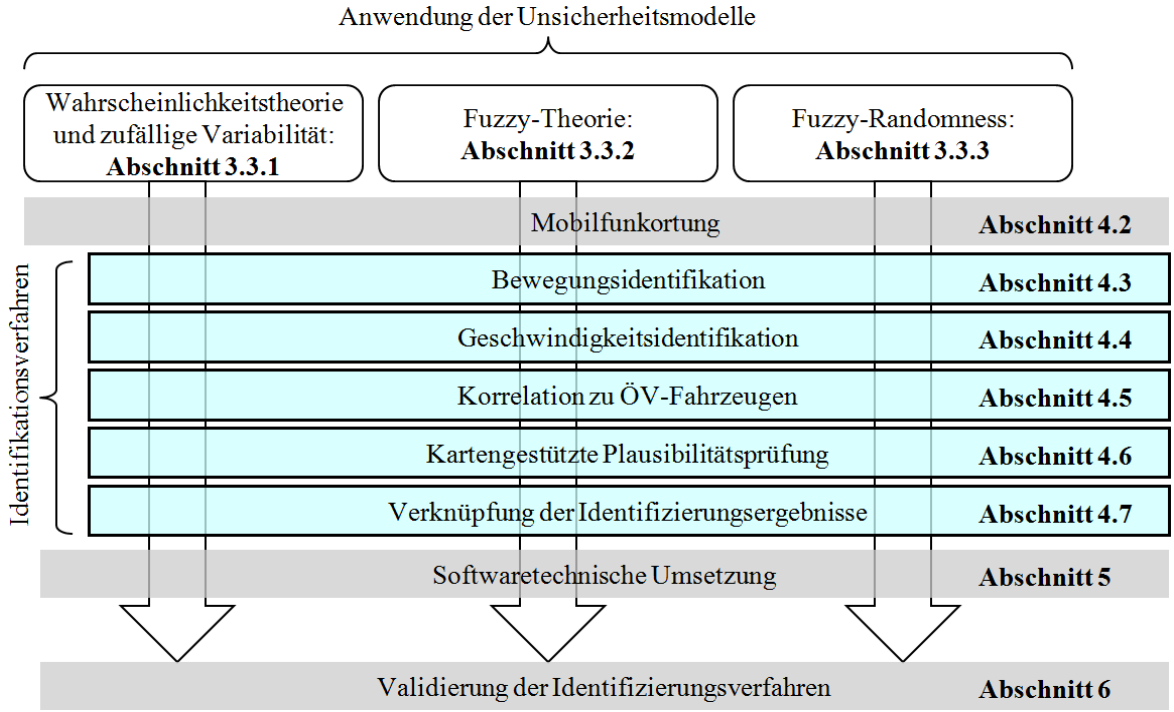


Abbildung 1-1: Abfolge der Verfahrensschritte

Abschnitt 4 behandelt die Kernaufgaben zur Entwicklung der Identifizierungsverfahren. Für jedes Verfahren werden die Grundidee erläutert und die zufällige Variabilität, die Fuzzy-Theorie und das Fuzzy-Randomness zur Unsicherheitsmodellierung angewendet. Eine Übersicht zu deren Abfolge zeigt Abbildung 1-1. Alle vorgestellten Identifizierungsverfahren bauen auf die Positionen der Mobilfunkgeräte auf, die aus den Mobilfunkdaten berechnet werden. Beginnend wird daher das eingesetzte Verfahren zur Mobilfunkortung erläutert. Aufgrund der hohen räumlichen Positionsunsicherheit im Vergleich zum zeitlichen Abstand der Positionen, wird mittels eines Verfahrens der Clusterbildung eine Reduzierung der Punktanzahl vorgestellt. Zur effizienten Anwendung digitaler Verkehrsnetze werden die Verwendung von Kacheln und die Übertragung der Positionsunsicherheiten auf Kacheln besprochen. Anschließend wird in der **Bewegungsidentifikation** der Frage nachgegangen, ob sich das Mobilfunkgerät in Bewegung befand. Bewegte es sich, werden in der **Geschwindigkeitsidentifikation** die Bewegungsgeschwindigkeiten abgeleitet und mit typischen Maximalgeschwindigkeiten der im Verkehr anzutreffenden Verkehrsmittel verglichen. Verkehrsmittel, die nicht mit den Geschwindigkeiten des Mobilfunkgerätes vereinbar sind, können daraufhin ausgeschlossen werden. Anhand der **Korrelation mit dem Öffentlichen Verkehr** wird anschließend überprüft, ob Ähnlichkeiten der Positionen des Mobilfunkgerätes mit den aus den Fahrplänen der ÖV-Fahrzeuge prognostizierten Positionen vorhanden sind. Zeigt die Korrelation keine Übereinstimmung mit dem ÖV, wird abschließend eine **kartengestützte Plausibilitätsprüfung** gegenüber den Charakteristika der Verkehrsnetze vorgenommen. Darin wird auf den Verkehrsnetzen eine Route gesucht, die sowohl räumlich und topologisch als auch zeitlich mit der Positionsfolge des Mobilfunkgerätes vereinbar ist. Verkehrsmittel auf deren Verkehrsnetzen keine Route gefunden werden kann, können daraufhin ebenfalls verworfen werden.

Die softwaretechnische Implementierung der Verfahren wird in Abschnitt 5 erläutert. Außerdem werden Hinweise für die Verknüpfung der Identifikationsergebnisse gegeben. In Abschnitt 6 erfolgt die Validierung der Identifikationsverfahren anhand realer Mobilfunkdaten. Die erzielten Identifikationsergebnisse werden darin mit Referenzdaten verglichen und bewertet.

2 Grundlagen ausgewählter Verkehrserfassungssysteme

2.1 Einteilung des Verkehrs

Zur langfristigen Planung von Verkehrsströmen, der Ermittlung des Verkehrsverhaltens bei bestimmten Situationen oder zur zeitnahen Erstellung von Verkehrsinformationen (z. B. Staumeldungen, Routenwahlempfehlungen) werden Verkehrslageinformationen der für die Verkehrslageerfassung relevanten Verkehrsmittel benötigt. Das betrifft die nachträgliche Bestimmung früherer Verkehrslagen durch Auswertung älterer Daten ebenso wie die Erfassung der aktuellen Verkehrslage. Da in der vorliegenden Arbeit Mobilfunkdaten für die Verkehrserfassung Verwendung finden, ist eine Aufstellung derjenigen Verkehrsmittel erforderlich, die von Mobilfunkteilnehmern genutzt werden können. Im Folgenden wird der Fokus auf die Nutzung von Landfahrzeugen gelegt. Fußgänger werden vereinfacht als ein Verkehrsmittel betrachtet.

Das Statistische Bundesamt (DESTATIS 2011a) unterteilt in einer ersten Ebene den Verkehr nach Personenverkehr und Güterverkehr. In nachfolgenden Ebenen wird speziell nach **Motorisiertem Individualverkehr** (MIV), **Nicht-Motorisiertem Individualverkehr** (NMIV) und **Öffentlichem Personenverkehr** (ÖV) klassifiziert. Mit dem Hintergrund der Verkehrslageerfassung sind auf Straßen fahrende Fahrzeuge des Güterverkehrs ebenfalls relevant, gehören jedoch nach DESTATIS (2011a) einer anderen Klasse als dem MIV an.

Das Kraftfahrt-Bundesamt (KBA 2009) klassifiziert nach der **Fahrzeugklasse** und unterscheidet **Krafträder** (KRAD), **Personenkraftwagen** (PKW), **Kraftomnibusse** (Reise-, ÖV- Bus), **Lastkraftwagen** (LKW), Zugmaschinen und sonstige Kfz. Bis auf die Fahrzeuge des Öffentlichen Verkehrs sind für die vorliegende Problemstellung alle dem MIV angehörig.

Für die Erfassung des Straßenverkehrs werden Geräte eingesetzt, deren Eignung nach den Vorgaben der Bundesanstalt für Straßenwesen (BAST) zu überprüfen ist. Die Prüfvorschrift (BAST 2011) sieht hierfür eine Fahrzeugklassifizierung von 8+1 Klassen vor. Die acht Grundklassen sind: KRAD, PKW, Lieferwagen (Lfw), PKW mit Anhänger (PKWA), LKW, LKW mit Anhänger (LKWA), SattelKfz, Bus. Der zusätzlichen neunten Klasse sind die Zählungen von „nicht realen Fahrzeugen“ zuzuordnen, die durch die Messeinrichtung beispielsweise bei Mehrfachzählungen desselben Fahrzeugs während des Prüfverfahrens entstehen. Anhand dieser Definitionen können bis auf die Fahrzeugklasse Bus, alle Grundklassen dem MIV zugeordnet werden. Die Aufgliederung „mit Anhänger“ bei PKW und LKW wird aus Vereinfachungsgründen im Folgenden nicht übernommen. Ebenso werden SattelKfz und LfW zur Vereinfachung als Untermenge der LKW verstanden. Wie im späteren Verlauf der Arbeit noch gezeigt wird, ist für die Identifizierung hingegen die Unterscheidung zwischen ÖV und **Individual Verkehr** (IV) wesentlich. Zwischen diesen beiden Verkehrsarten unterscheidet BAST 2011 jedoch nicht, so dass die Fahrzeuge der Grundklasse Bus anhand dieser Definition nicht genügend klassifiziert werden würden.

Der ÖV kann unterteilt werden in den klassischen, nach vordefinierten Fahrplänen und Fahrwegen verkehrenden ÖV und in den **flexiblen ÖV** (Flex-ÖV). Beim Flex-ÖV werden die Haltepunkte, die Abfahr- und Ankunftszeiten sowie der Fahrweg erst unmittelbar bei Fahrtbeginn festgelegt und können gegebenenfalls auch während der Fahrt in gewissen Grenzen noch geändert werden. Außerdem verwendet der Flex-ÖV im Wesentlichen das gleiche Straßennetz wie der MIV. Beispiele sind **Taxi**, **Anrufsammeltaxi** (AST) und **routenflexible Angebote** (Flexi-Bus) wie z. B. RudyBus (HENNINGER ET AL. 2003; WILTSCHKO & SCHOLLMEYER 2006). Der Flex-ÖV besitzt einerseits öffentlichen Charakter, da er den Bestimmungen des Personenbeförderungsgesetzes unterliegt, andererseits kann er aber individuell genutzt werden (GABLER 2013). Aufgrund dieser IV typischen

Eigenschaften eignen sich die Fahrzeuge des Flex-ÖV grundsätzlich für die Verkehrslageerfassung des MIV. Die Besonderheit, dass diese Fahrzeuge auch ÖV-eigene Fahrstreifen (z. B. Busspuren) nutzen dürfen, sei in diesem Zusammenhang aufgrund des geringen Vorkommens vernachlässigt.

Mit diesem Hintergrund wurde für die vorliegende Problemstellung eine angepasste Klassifizierung der Mobilfunkstationen bezüglich der verwendeten Verkehrsmittel vorgenommen, die in Abbildung 2-1 grafisch dargestellt ist. In der obersten Ebene wird zwischen „**ohne Verkehrsbeteiligung**“ (oVB) und „**mit Verkehrsbeteiligung**“ (VB) und im Anschluss unter VB weiter nach Öffentlicher Personenverkehr (ÖV) und Individualverkehr (IV) unterschieden. Die weiteren Untergruppen wurden im Wesentlichen nach Vorbild des Kraftfahrt-Bundesamtes gebildet, wobei die Verknüpfung des Flex-ÖV mit dem MIV eine Besonderheit aufgrund der spezifischen IV Eigenschaften ist. Die interessierende Zielgruppe des MIV ist im rechten Zweig mit grüner Hintergrundfarbe darstellt.

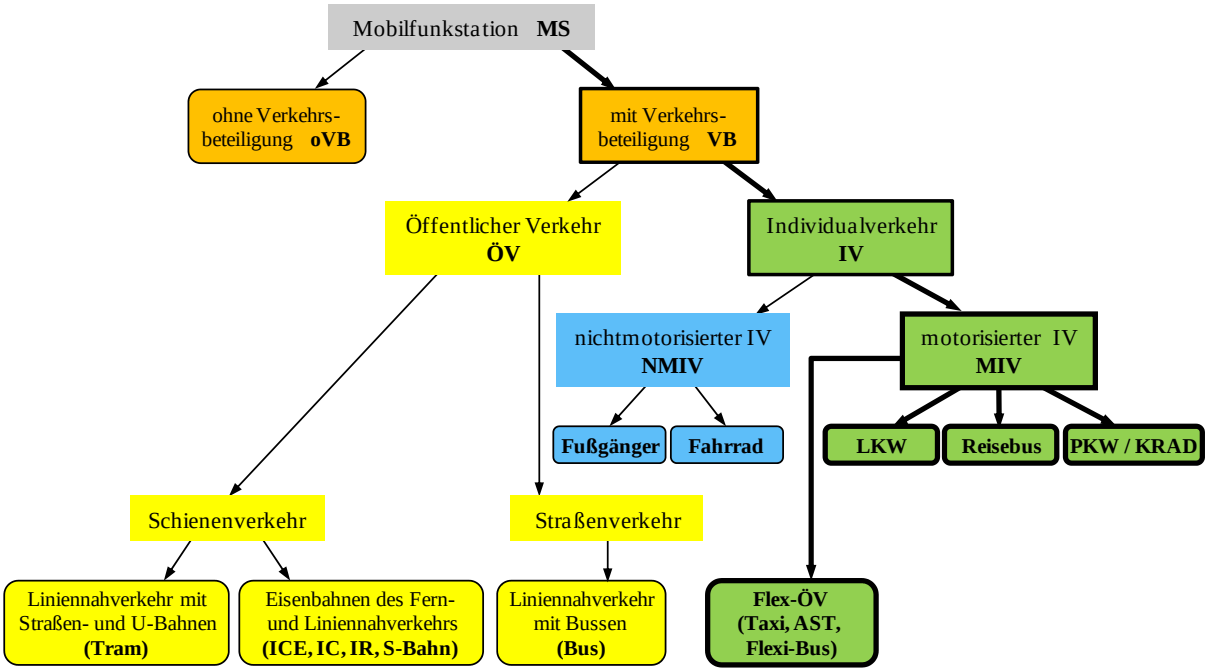


Abbildung 2-1: Klassifizierung der Verkehrsmittel für die Verkehrserfassung mit FPD

2.2 Verfahren zur Verkehrserfassung und Verkehrsmittelidentifizierung

2.2.1 Verkehrserfassungssysteme mit infrastrukturgebundenen Sensoren

Da sie seit mehreren Jahrzehnten im Verkehrswesen etabliert sind, können Erfassungssysteme mit infrastrukturgebundenen Sensoren zu den klassischen Systemen der Verkehrserfassung gezählt werden. Typische technische Realisierungsformen sind Induktionsschleifen oder Infrarotsensoren. Ihre Erfassungsregion ist bis auf einzelne Fahrstreifen eingrenzbar, weshalb sie ein wichtiger Bestandteil zur Steuerung von Lichtsignalanlagen oder auch zur fahrstreifenscharfen Verkehrszählung sind.

2.2.1.1 Ortsfeste Verkehrserfassung

Sind Fahrstreifen ausschließlich für die Nutzung spezifischer Verkehrsmittel vorgesehen, können die erfassten Fahrzeuge auf wenige mögliche Verkehrsmittel eingegrenzt werden oder gegebenenfalls sogar unmittelbar auf ein spezifisches Verkehrsmittel geschlossen werden. Diese Art der Identifizierung des Verkehrsmittels gelingt durch Definition zulässiger Verkehrsmittel oder Ausschluss von Verkehrsmitteln (z. B. Autobahnen – kein

Schienenverkehr, kein NMIV –, Busspuren – nur ÖV-Busse und außer Taxi kein MIV – oder Straßenbahnschienen – nur Tram –). Die Identifizierung der Verkehrsmittel gelingt somit durch die räumliche Eingrenzung der Erfassungsregion und Kenntnis über das dortige Vorkommen spezifischer Verkehrsmittel. Typische Geräte zur ortsfesten Verkehrserfassung sind in die Fahrbahn eingelassene Induktionsschleifen sowie an Brücken oder Masten montierte Infrarotdetektionssysteme.

Trotz der ortsfesten Installation derartiger Erfassungsanlagen zeigt LEHNHOFF (2005) in empirischen Analysen, dass das tatsächliche Verkehrsaufkommen nur bedingt mit den Messergebnissen übereinstimmt. Die zum Teil erheblichen Abweichungen der Detektordaten sind ursächlich auf die Registrierung von Fahrzeugen benachbarter Fahrstreifen, Mehrfachregistrierung von Fahrzeugen aufgrund vorbeifahrender Straßenbahnen, systematisch geringere Registrierung von Fahrzeuganzahlen und auch die fehlerhafte Zählung von real nicht vorhandenen Fahrzeugen zurückzuführen. Die Eignung derartiger Systeme zur Verkehrslagebestimmung und -prognose zeigt sich dennoch beim Einsatz in städtischen Verkehrsleitsystemen oder dem kommerziellen Einsatz, wie z. B. dem TMCpro der DDG (DDG 2011) als erweiterten Dienst zu den per Radiofunk ausgestrahlten digitalen Verkehrsinformationen Traffic Message Channel (TMC).

2.2.1.2 Ortsfeste Verkehrserfassung mit Verkehrsmittel erkennenden Systemen

Durch Auswertung des Signalverlaufes der Sensoren (z. B. Doppel-Induktionsschleifen) oder durch bildgebende Sensoren (z. B. CCD-Kameras) sind neben der Fahrzeugmenge und der Geschwindigkeit des Verkehrs auch Fahrzeugmaße ableitbar. Zumeist kann damit eine geeignetere Klassifizierung der erfassten Verkehrsmittel (z. B. LKW, LKW mit Anhänger) ermöglicht werden. Beispielsweise sind geringere Durchschnittsgeschwindigkeiten auf mehrspurigen Steigungsabschnitten gegebenenfalls auf eine höhere LKW-Verkehrsdichte zurückführbar, da Steigungen für PKW in der Regel keine Geschwindigkeitsreduzierung bedeuten. Eine fahrzeugscharfe Routenwahl gelingt z. B. durch eine Kennzeichenerfassung an mehreren Ortungen im Verkehrsnetz (FRIEDRICH ET AL. 2008a). Diese datenschutzrechtlich nicht unbedenklichen Erfassungsmethoden dienen wissenschaftlichen Untersuchungen zum Verhalten von Verkehrsteilnehmern bei Verkehrsinformationen (z. B. der Nutzung von Umleitungsrouten bei Staumeldungen).

2.2.2 Verkehrserfassungssysteme mit mobilen Sensoren

2.2.2.1 Erfassung von Bewegungsprofilen von Fahrzeugen

Die Kosten für Infrastruktur und Wartung sowie die daraus resultierende begrenzte Anzahl an Standorten zur Verkehrserfassung, führten zur Entwicklung und Einführung von im Verkehr „mitschwimmenden“ (engl.: floating) Sensoren. Die im Verkehr „mitschwimmenden“ Sensoren werden von den Betreibern der Verkehrserfassung explizit in die interessierenden Fahrzeuge integriert. Daraufhin erfolgt die Erfassung von Bewegungsprofilen, die an eine Zentrale gesendet werden.

Derartig gewonnene Verkehrslagedaten werden als FCD (vgl. Abschnitt 1.1, S. 9) bezeichnet und verwenden zur Positionsbestimmung im Wesentlichen GPS-Empfänger. BUSCH ET AL. (2004) schlagen beispielsweise den Einsatz von Fahrzeugflotten (Taxi) vor, die wegen vorhandener Dispositionssysteme ohnehin entsprechende Bewegungsprofile erfassen und im Rückschluss unmittelbar Nutzen aus der Verkehrslageprognose ziehen können. Für die Verkehrserfassung des MIV ist die Repräsentativität von FCD insbesondere von Fahrzeugflotten jedoch infrage zu stellen. Taxis dürfen beispielsweise gesonderte Fahrstreifen nutzen, wie z. B. Busspuren, und deren Fahrer/-innen sind in hohem Maße ortskundig. Engpässe werden daher frühzeitig umfahren und so Staus gegebenenfalls nicht oder nicht in ihrem tatsächlichen Ausmaß erfasst. Neben den Kosten für Fahrzeugausstattung und Einbau begrenzen aus wirtschaftlichen Gesichtspunkten vor allem die Kommunikationskosten

eine durchgreifende Ausweitung dieser Technik auf Fahrzeuge die keiner spezifischen Fahrzeugflotte angehören. Von Vorteil ist, dass die Bewegungsprofile mit der Information aus welchem Fahrzeug sie stammen, verknüpft sind. Die Verkehrsmittelklassifizierung steht damit unmittelbar zur Verfügung.

2.2.2.2 Erfassung von Bewegungsprofilen von Fahrzeugen innerhalb von Sensornetzen

Ein für die flächendeckende Verkehrserfassung vielversprechender Ansatz basiert auf der Auswertung von Mobilfunkdaten, die im Wesentlichen aus den Kommunikationssignalen zwischen den **mobilen Stationen** (MS Abk. für die englische Bezeichnung „**M**obile **S**tation") – in Deutschland als *Handys* (GB *mobile phone*, USA *cellular phone*) bezeichnet – und der Mobilfunkinfrastruktur gewonnen werden. Gemeinsam bilden sie das **Mobilfunknetz**. Als Techniken werden in Deutschland vorrangig die Mobilfunknetze des **Global System for Mobile Communications (**GSM, auch: **G**roupe **S**pécial **M**obile) oder **Universal Mobile Telecommunications System** (UMTS) eingesetzt.

Aus den Mobilfunkdaten werden Ortungsinformationen abgeleitet und daraus Bewegungsprofile für die Verkehrserfassung erstellt. Da die Ortungsgeräte (MS in Form von Handys) in der Bevölkerung und somit auch in den Fahrzeugen weit verbreitet sind sowie die Infrastruktur bereits vorhanden ist, entstehen im Vergleich zu den zuvor genannten Techniken nur geringe Kosten für den Abgriff der Kommunikationsdaten, was diesen Ansatz wirtschaftlich interessant macht. Trotz unterschiedlicher Ortungsansätze sind deutliche Parallelen zu den Eigenschaften von FCD erkennbar, weshalb man die gewonnenen Daten auch als **Net-FCD** (FASTENRATH 2008, S. 33) oder mit ausdrücklicher Kennung der Verkehrserfassung mittels Mobilfunkdaten als FPD (**Floating-Phone-Data**) (SCHWIEGER 2008, S. 6ff) bezeichnet.

Zur Lokalisierung beschreibt JUNIUS (1995) Verfahren, die nach dem Prinzip der Pegelmustererkennung einen Vergleich von gemessenen Signalen mit denen in Referenzdatenbanken arbeiten. In KENNEMANN (1997) wird dieses Verfahren aufgegriffen und sowohl im funktionalen Ansatz verfeinert als auch empirisch untersucht. In den Arbeiten werden Hinweise auf die erforderliche Beachtung der unterschiedlichen Verkehrsmittel gegeben. Möglichkeiten oder Verfahren, um die Mobilfunkdaten nach den Verkehrsmitteln, in denen sie generiert wurden, zu unterscheiden, werden jedoch nicht aufgeführt.

SOMMER (2002) erarbeitet und diskutiert die Datenerhebung verkehrstechnisch relevanter Merkmale mittels Mobiltelefon im Rahmen des Forschungsprojektes „TeleTravel System" (TTS) (WERMUTH ET AL. 2003). Per Tastatur wird das Verkehrsverhalten von den Nutzern in das von ihnen mitgeführte Mobiltelefon eingegeben. Durch technische Änderungen am Mobiltelefon wurde zusätzlich eine Ortungskomponente in der MS integriert (SOMMER 2002, S. 77f). Aus den Positionsdaten und den manuell eingegebenen Identifikationsmerkmalen ist das Verkehrsverhalten und in gewissem Rahmen auch die Verkehrslage ableitbar. Die Verfahren stützen sich somit auf ausgewählte und geschulte Probanden. Das Ziel dieser Arbeit visiert jedoch eher die Untersuchung der Nutzerakzeptanz für Mobiltelefon gestützte Erhebungsverfahren (elektronische Fragebögen) an und weniger die Verkehrserfassung. Eine automatisierte Identifizierung des vom Nutzer gewählten Verkehrsmittels wird angesprochen, jedoch nicht im Rahmen umsetzbarer Verfahren behandelt.

Im Projekt COX (Communication & Orientation eXpert, cox „Steuermann") wurden mobile Geräte als Lotsendienst für den MIV, NMIV und ÖV entwickelt, die sowohl Planungsfunktionen unterstützen als auch während der Reise aktuelle Verkehrs- und Zusatzinformationen in Form von Reiseinformationsdiensten bereitstellen. Der Reisende dient darin selbst als Sensor zur Verkehrslageerfassung, indem Bewegungszustände und Bewegungsprofile an eine Zentrale weiter geleitet werden (KRAMPE & KIESLICH 2005). Die anvisierte Positionsbestimmung basiert auf der Auswertung von Signalen der Telekommunikationsnetze (GSM, UMTS), lokaler Netze (z. B. WLAN, Bluetooth) und Koppelortung mit Beschleunigungs- und Drehratensensoren.

Vorrangig werden jedoch GPS/Galileo-Empfänger eingesetzt. Eine automatisierte Erkennung der Verkehrsmittelwahl diskutiert KRAMPE (2007, S. 50f). Trotz Einsatz von GPS-Empfängern, Drehratensensoren und Beschleunigungsmessern führten die Verfahren zu keinem befriedigenden Ergebnis, weshalb letztlich auf eine manuelle Eingabe am Gerät zurückgegriffen wurde.

In KOHLEN (2004) wird ein Verfahren zur Routenidentifikation vorgeschlagen, das im bereits oben genannten Projekt TTS erprobt werden konnte. Unsicherheiten der Mobilfunkdaten werden explizit mit Hilfe von unscharfen Mengen berücksichtigt und mit Fuzzy-Logiken behandelt. Um resultierende Mehrdeutigkeiten für mögliche gewählte Routen aufzulösen, wird mittels Einführung von Wahrscheinlichkeiten die wahrscheinlichste Route berechnet. Höhere Ortungsgenauigkeiten bewirken eine bessere Eingrenzung möglicher Routen. Das Hauptanwendungsgebiet wird im MIV gesehen, jedoch wird auch auf die mögliche Übertragbarkeit auf den ÖV hingewiesen. Verfahren zur Unterscheidung von Mobilfunkdaten aus dem MIV, ÖV oder einem anderen Verkehrsmittel, werden nicht explizit dargestellt.

2.3 Verkehrsdatenerfassung im Projekt DO-IT

Der Einbau von Sensoren in oder an die Fahrbahn oder in das Fahrzeug bedingt hohe Kosten dieser Systeme für Infrastruktur, Wartung und Betrieb. Bei Infrastruktur gebundenen Systemen sind insbesondere die geringe Anzahl der Erfassungsstandorte und bei FCD, deren geringe Verbreitung zu nennen. Diesen Nachteilen könnte mit dem Einsatz von FPD begegnet werden. Den Vorteilen der FPD in Datenmenge, Flächendeckung und den geringen Erfassungskosten, stehen jedoch insbesondere Schwierigkeiten bei der automatisierten Zuordnung der Mobilfunkdaten zum gewählten Verkehrsmittel gegenüber.

Das **Projekt DO-IT** (**D**aten**o**ptimierung für **i**ntegrierte **T**elematik) ist ein vom Bundesministerium für Wirtschaft und Technologie (BMWi) gefördertes Forschungsprojekt (Förderkennzeichen: 19 B 5007, Laufzeit 2005-2009) für die Erfassung der aktuellen Verkehrslage mit Hilfe von netzwerkbasierten Mobilfunkdaten anonymisierter Verkehrsteilnehmer (VM2010 PROJEKT DO-IT 2011). Die Mobilfunkdaten wurden für ein Demonstratorgebiet im Nord-Westen von Baden-Württemberg flächendeckend erfasst. Um eine spätere Markteinführung der entwickelten Systeme durch Vermeidung von zusätzlichen Investitionen seitens der Mobilfunkbetreiber zu erleichtern, wurden bewusst keine infrastrukturelle Änderungen an der Signalstruktur und dem Signalinhalt des bestehenden GSM-Mobilfunknetzes oder des Endgerätes vorgenommen. Auf Seiten der Infrastruktur wurden die für die Mobilfunkortung relevanten Mobilfunkdaten an der sogenannten **A-bis Schnittstelle** (A-bis Interface) zwischen der **Base Transceiver Station** (BTS) und dem **Base Station Controller** (BSC) verwendet. Mit Verwendung von Daten aus der A-bis Schnittstelle einer BTS wird die Erfassung von Mobilfunksignalen im Umfeld von wenigen km^2 ermöglicht. Um eine signaltechnische Verfolgung von MS über mehrere einhundert km^2 zu ermöglichen und somit vollständigere Bewegungsprofile der MS erstellen zu können, wurden zusätzlich Daten der A-Schnittstelle (A-Interface) zwischen BSC und **Mobile Switching Center** (MSC) erfasst.

Für die Verkehrserfassung ist neben der Mobilfunkortung die Identifizierung relevanter Mobilfunkdaten ein zentraler Punkt. Darin sollen diejenigen Mobilfunkdaten identifiziert werden, die den Verkehrsmitteln des MIV entstammen. Die nachfolgende Abbildung 2-2 veranschaulicht die Einbettung der Identifizierung in der Systemarchitektur des Projektes DO-IT (WILTSCHKO ET AL. 2006, 2007; WILTSCHKO & SCHWIEGER 2008; FRIEDRICH ET AL. 2008b; SCHWIEGER 2008; SCHLAICH ET AL. 2010).

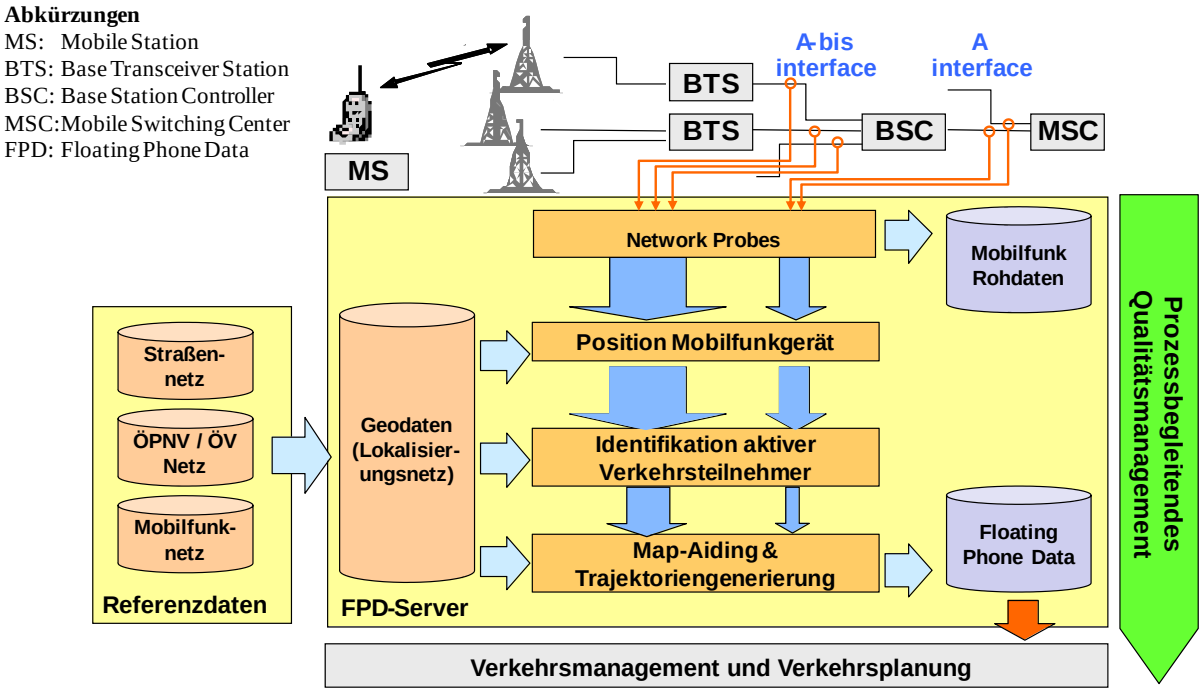


Abbildung 2-2: Systemarchitektur der Mobilfunkdatenerfassung zur FPD-Erzeugung im Projekt DO-IT (SCHWIEGER 2008, S. 7)

Um die Mobilfunkdaten für die Verkehrserfassung nutzen zu können, müssen die MS geortet werden. Bei der Ortung von MS ist insbesondere der Schutz personenbezogener Daten zu beachten. Nach der Charta der Grundrechte der Europäischen Union (CHARTA EU 2007/C 303/01, 2007) Artikel 8, zählt der Aufenthaltsort einer Person ausdrücklich dazu. Existiert keine ausdrückliche Einwilligung des Betroffenen oder speziell für Deutschland die Erlaubnis durch das Bundesdatenschutzgesetz (BDSG) (vgl. DÄUBLER ET AL. (2007) oder anderer entsprechender Rechtsvorschriften) ist die Ortung von MS der Mobilfunkkunden nicht gestattet. Eine wirksame Anonymisierung § 3 (6) BDSG oder Pseudonymisierung § 3 (6a) BDSG der Mobilfunkdaten sowie die Gewährleistung für die Datenübertragung, dass nur der benannte Empfänger Einsicht in die Daten erhält, sind sicherzustellen (KOHLEN 2004, S. 27f; SOMMER 2002, S. 102f). Die Zulässigkeit der Datenerhebung, -verarbeitung und -nutzung von personenbezogenen Daten regelt § 4 BDSG und die Bedingungen der Einwilligung § 4a BDSG. Aus diesem Grund verwenden die auf FPD basierten Verfahren anonymisierte Mobilfunkdaten oder streben deren Verwendung an.

3 Modellierung von Unsicherheiten

3.1 Motivation

Um die Schlussfolgerungen der Verkehrsmittelidentifizierung geeignet beurteilen zu können, müssen die Unsicherheiten der Mobilfunkdaten (Beobachtungen) adäquat modelliert werden. Das gilt im gleichen Maße für die weiteren, im Verarbeitungsprozess eingebundenen Informationen. Wie im späteren Abschnitt 4.2 noch gezeigt wird, sind insbesondere die Unsicherheitseinflüsse auf die Mobilfunkdaten vielfältig und zum Teil nur näherungsweise bekannt. Hauptaugenmerk ist daher auf die Wahl geeigneter mathematischer Formulierungsmöglichkeiten zur Behandlung von Beobachtungsunsicherheiten zu richten. Unsicherheiten, die aufgrund der Modellwahl entstehen, stellen einen eigenen Themenkomplex dar und werden in der vorliegenden Arbeit nicht behandelt.

3.2 Grundbegriffe der Unsicherheiten

Die Abweichung eines erfassten Elementes von der Wirklichkeit wird als **Unsicherheit** bezeichnet. Die Ursachen dieser Abweichungen können unterschiedlicher Natur sein (GLEMSER 2000). Bei technischen Anwendungen wird zwischen zwei Komponenten von Unsicherheiten unterschieden (HELTON & OBERKAMPF 2004; MÖLLER & BEER 2004; TRUCANO ET AL. 2006):

- **Aleatorische Unsicherheiten** sind nicht reduzierbare Unsicherheiten, die ausschließlich zufälligen (stochastischen) Charakter besitzen. Die Hinzunahme weiterer Informationen verringert die Unsicherheiten im Allgemeinen nicht, weshalb aleatorische Unsicherheiten als *objektiv* angesehen werden.
- **Epistemische Unsicherheiten** beinhalten neben einem zufälligen Charakter auch Unsicherheiten aus unvollständigem oder fehlendem Wissen der behandelten Phänomene. Aufgrund der Abhängigkeit vom Wissen über die Phänomene oder auch dem Wissensstand des Auswertenden gelten epistemische Unsicherheiten als *subjektiv*.

Während innerhalb der aleatorischen Unsicherheit Wahrscheinlichkeiten und Statistiken (z. B. mittels Wahrscheinlichkeitstheorie und Stochastik) geeignet quantifiziert und behandelt werden können, ist die Quantifizierung und Behandlung epistemischer Unsicherheiten im Allgemeinen schwieriger. Aus diesem Grund wird in der vorliegenden Arbeit der *objektive* zufällige Anteil der epistemischen Unsicherheiten mit Hilfe von Ansätzen aus der Wahrscheinlichkeitstheorie stochastisch als zufällige Variabilität behandelt. Die Behandlung der *subjektiven* nicht-zufälligen Anteile erfolgt durch Unschärfe. Hierbei wird davon ausgegangen, dass die Realität einen zufälligen Charakter hat, der jedoch nur unscharf beobachtet werden kann (NEUMANN 2009). Die für die vorliegende Arbeit grundlegen Begrifflichkeiten der zufälligen Variabilität und der Unschärfe werden wie folgt definiert:

- **Zufällige Variabilität**: Die Annäherung eines Messergebnisses x_i, im Folgenden auch als **Beobachtung** bezeichnet, an den wahren Wert $\hat{x}$ oder den Erwartungswert $E(x_i) = \overline{x}$ nennt man **Genauigkeit**. Der wahre Wert bleibt in der Regel verborgen. Unter dem Erwartungswert versteht man den Wert, gegen den das arithmetische Mittel bei unbegrenzter Wiederholung der Beobachtung konvergiert (WELSCH ET AL. 2000, S. 101f) basierend auf DIN 1319-1, DIN 18710-1, DIN 55350-13. Zur Angabe eines Genauigkeitsmaßes ist zu berücksichtigen, dass Beobachtungen Ergebnisse eines physikalischen Experiments sind, dessen Ausgang in einem gewissen Rahmen unsicher ist. Fehler sind

als Abweichungen der Beobachtungen von einem Modell zu verstehen. Man unterscheidet in der Wahrscheinlichkeitsrechnung zwischen **groben Fehlern** (Irrtum), **systematischen Abweichungen** δ und **zufälligen Abweichungen** ε. „*…Grobe Fehler sind durch Kontrollen zu vermeiden oder durch geeignete Maßnahmen aufzudecken und zu eliminieren. … Systematische Abweichungen … sind … durch geeignete Beobachtungsverfahren … zu erfassen … und durch Korrektionen zu eliminieren…*“ (WELSCH ET AL. 2000, S. 101). Als Differenz zwischen Erwartungswert und wahrem Wert werden systematische Abweichungen definiert mit $\delta = \overline{x} - \tilde{x}$. Als zufällige Abweichungen versteht man die Abweichung zwischen Erwartungswert und Beobachtung $\varepsilon_i = \overline{x} - x_i$. Eine von den zufälligen Abweichungen losgelöste Behandlung der groben Fehler und systematischen Abweichungen ist nicht unmittelbar möglich. Beispielsweise setzt die Aufdeckung eines groben Fehlers in einer Beobachtung eine hinreichend gute Beschreibung der systematischen und der zufälligen Abweichungen voraus. Systematische Abweichungen liegen erst dann vor, wenn sie mit Annahme einer gewissen Größenordnung für die zufällige Abweichung von dieser separierbar ist. Erst mit einer hinreichend guten, realistischen Beschreibung der Unsicherheiten für die Beobachtungen einschließlich ihrer Einflussfaktoren gegenüber einer theoretisch erwartbaren Häufigkeit, kann eine Aufdeckung von groben Fehlern ermöglicht werden.

- **Unschärfe**: Bei der mathematischen Beschreibung von Phänomenen oder auch Systemen, sind aufgrund von komplexen Zusammenhängen der Natur oder komplizierten, nur unvollständig beschreibbaren Strukturen, fokussierende Vereinfachungen und Reduzierungen in den meisten Anwendungsfällen unumgänglich. In Folge dessen entstehen Abweichungen eines Modells gegenüber der Wirklichkeit; es entstehen Unschärfen. Ein Wesen dieser Unschärfen ist, dass sie nicht-stochastisch sind (STEINMETZ 1998, S. 3). Die mathematische, auf der Mengentheorie nach GEORG CANTOR beruhende Behandlung von Unschärfen erfolgt z. B. mittels Intervallen (MOORE 1966; ALEFELD 1968; ALEFELD & HERZBERGER 1974; MAYER 1989). Im geodätischen Umfeld werden deren Anwendung von KUTTERER (1994) und SCHÖN (2003) mit Intervallen gezeigt. Als Erweiterung zu den Intervallen dienen die von ZADEH (1965) eingeführten unscharfen Mengen. Im Unterschied zu einer cantorschen Menge, in der die Bildmenge nur die Werte Null und Eins mittels der Indikatorfunktion $i(x) \in \{0;1\}$ annehmen kann, lassen sich mit unscharfen Mengen auch alle reellen Zahlen als Intervall zwischen Null und Eins $m(x) \in [0,1]$ als unscharfe Zuordnung definieren (BIEWER 1997, S. 54f). Mit $m(x)$ wird die entsprechende Zugehörigkeitsfunktion formuliert. KUTTERER (1994) und SCHÖN (2003) verwenden den Begriff „**Impräzision**“ als Oberbegriff für die in der vorliegenden Arbeit ausschließlich mit „Unschärfe“ bezeichneten Phänomene. Dem Begriff Unschärfe wird wegen der Anwendung der Fuzzy-Theorie und der Analogie zur Behandlung „unscharfer Mengen“ Vorzug gegeben.

Die **Wahrscheinlichkeitstheorie** beschäftigt sich mit der Behandlung von Phänomenen, deren Auftreten mit Hilfe von Wahrscheinlichkeiten, also der Häufigkeit ihres Auftretens beschrieben wird. Mit der eng zur Wahrscheinlichkeitstheorie verzahnten **Stochastik** werden speziell Phänomene mit zufälliger Variabilität behandelt. Die Stochastik erlangte in der Messtechnik große Bedeutung, da Messgrößen mit ihr sehr geeignet beschreibbar und beurteilbar sind, weshalb sich zur Behandlung aleatorischer Unsicherheiten die Wahrscheinlichkeitstheorie mit der Stochastik als Teilgebiet hervorragend eignet (Neumann 2009).

Sind Messungen, Daten oder allgemein Ausgangsinformationen in der Häufigkeit ihres Auftretens nicht beschreibbar oder nicht bekannt, oder werden systematische Zusammenhänge zwischen Größen vermutet und können abgeschätzt werden, dann eignen sich die Wahrscheinlichkeitstheorie und speziell die Stochastik nur bedingt. Weitere, aus der praktischen Anwendung der Wahrscheinlichkeitstheorie resultierende Probleme treten

auf, wenn z. B. Phänomene physikalisch leicht beschreibbar sind, jedoch mathematisch nur aufwendig oder nicht formuliert werden können. Zur Auswertung derartiger Daten oder Informationen eignet sich die von ZADEH (1965) mathematisch eingeführt **Fuzzy-Theorie** (fuzzy engl.: unscharf). Auf Basis der Mengentheorie erfolgt darin eine mathematische Behandlung von unscharfen Größen. Die Fuzzy-Theorie ist keine „unscharfe Theorie", sondern eine Theorie zur mathematischen Behandlung von Unschärfe.

Im Rahmen der gemeinsamen, epistemisch motivierten Modellierung von zufälliger Variabilität und Unschärfe unterscheidet KUTTERER (2002, S. 64) zwischen *unscharfen zufälligen Variablen* (engl.: fuzzy random variables, **FRV**), physikalisch motivierten *zufälligen unscharfen Variablen* (engl.: Random fuzzy variables, RFV) und aus Überlagerungen von Zufallszahlen und unscharfen Zahlen resultierenden *hybriden Zahlen* (engl.: hybrid numbers). Ausgehend von der Annahme, dass die originäre (abgelesene) Beobachtung alleinig Träger der zufälligen Variabilität ist und die Unschärfe durch deren Aufbereitung im Rahmen von Korrektionen und Reduktionen sowie dem Auswerteprozess selbst induziert werden, sind in der vorliegenden Arbeit ausschließlich die FRV von Interesse. Die FRV wurden mit den beiden Artikeln von KWAKERNAAK (1978, 1979) mathematisch eingeführt. Im geodätischen Umfeld beschäftigen sich KUTTERER (2002), SCHÖN (2003) und NEUMANN (2009) mit diesem Themengebiet sowie insbesondere mit der Behandlung und Beurteilung von FRV. Auf die dort verwendeten Formulierungen und Definitionen baut die vorliegende Arbeit maßgeblich auf.

Eine Übersicht und Einordnung der Wahrscheinlichkeitstheorie, Fuzzy-Theorie und des Fuzzy-Randomness gibt die nachfolgende Abbildung 3-1. In ihr sind außerdem die Abschnittsnummern vermerkt, in denen die jeweiligen mathematischen Definitionen erfolgen.

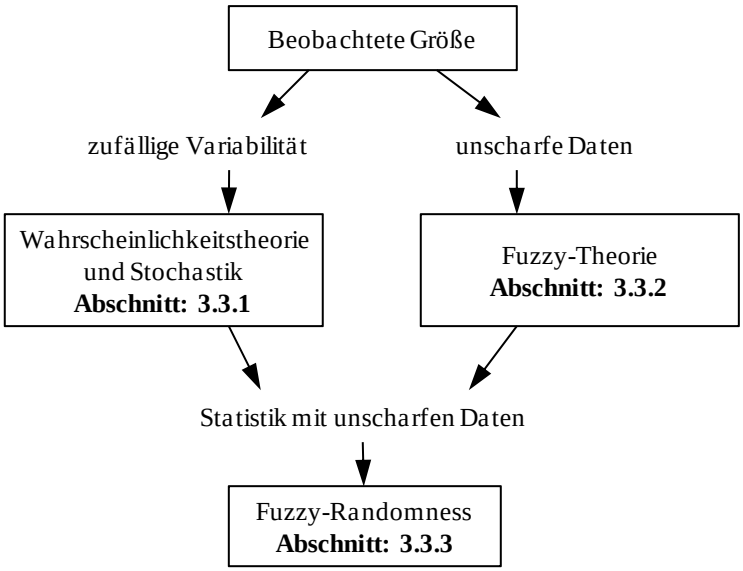


Abbildung 3-1: Zusammenhang zwischen Stochastik und Unschärfe nach VIERTL & HARETER (2006, S. 5)

3.3 Mathematische Behandlung von Unsicherheiten

3.3.1 Wahrscheinlichkeitstheorie und zufällige Variabilität

Werden Beobachtungen aus empirisch durchgeführten (physikalischen) Messungen gewonnen, können sie als Ergebnisse von Zufallsexperimenten betrachtet werden. Sie sind damit auch als zufällige Ereignisse interpretierbar (KOCH 1997, S. 81). Für zufällige Ereignisse ist charakteristisch, dass sie weder in Vorzeichen noch in Größe vorhergesagt werden können. Im Rahmen der Wahrscheinlichkeitstheorie werden geeignete mathematische Werkzeuge zur Modellierung, Behandlung und Beurteilung von Messgrößen als Zufallsereignisse sowie zur Ableitung von Aussagen aus diesen Größen bereitgestellt. Unter anderem ist es das Ziel, die **Wahrscheinlichkeit** $P(\mathrm{E})$ für das Auftreten eines **Ereignisses** E oder im Rahmen von Messungen die Wahrscheinlichkeit für das Auftreten von spezifischen Abweichungen gegenüber einem wahren Wert oder auch einem Erwartungswert zu bestimmen.

Die **Wahrscheinlichkeitstheorie** ist eine Teildisziplin der **Stochastik**. Die Stochastik umfasst in der Mathematik denjenigen Teilbereich, der sich mit zufälligen Ereignissen befasst. Die beschreibende Statistik ist darin ein wesentliches Element der Stochastik, um z. B. Variationen von Beobachtungsfolgen zu beschreiben.

In der klassischen Definition der Wahrscheinlichkeit $P(\mathrm{E})$ eines Ereignisses E ist es möglich, eine endliche Anzahl n sich gegenseitig ausschließender und gleichmöglicher Ereignisse in einem Experiment zu erzielen. Sind die Ergebnisse der Experimente vorhersagbar und die Anzahl $\mathrm{n_E}$ dieser Ergebnisse mit dem Ereignis E verbunden, folgt die **Wahrscheinlichkeit für das Auftreten** von E mit $P(\mathrm{E}) = \mathrm{n_E}/\mathrm{n}$ (KOCH 1997, S. 82 Formel (211.1)). Mit der Funktion *P* erfolgt die Abbildung der Ereignisse in das Intervall $[0,1]$. $P(\mathrm{E})$ liefert dann die Wahrscheinlichkeit oder auch das *Wahrscheinlichkeitsmaß* für das Auftreten eines bestimmten Ereignisses. Es gelten die folgenden drei Axiome der Wahrscheinlichkeit (KOCH 1997, S. 85; SACHS 2002, S. 39; VIERTL 2003, S. 21):

1. *Nichtnegativitätsaxiom*: Jedem Ereignis E aus B ist als Wahrscheinlichkeit eine reelle Zahl zwischen Null und Eins zugeordnet: $0 \le P(\mathrm{E}) \le 1$.
2. *Normierungsaxiom*: Ein sicheres Ereignis hat die Wahrscheinlichkeit Eins $P(\Omega) = 1$ und ein Unmögliches die Wahrscheinlichkeit Null $P(\varnothing) = 0$.
3. *Additivitätsaxiom*: Ist $\mathrm{E_1}, \mathrm{E_2}, \mathrm{E_3}, \ldots, \mathrm{E_n}$ von B eine Folge von endlich vielen oder abzählbar unendlich vielen sich gegenseitig ausschließenden Ereignissen $\mathrm{E_i} \cap \mathrm{E_j} = \varnothing, \mathrm{i} \neq \mathrm{j}$, dann ist die Wahrscheinlichkeit, dass von mehreren Ereignissen eines eintritt $P(\mathrm{E_1} \cup \mathrm{E_2} \cup \ldots)$, gleich die Summe der Wahrscheinlichkeiten der Ereignisse: $P\left(\bigcup_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{E_i}\right) = \sum_{\mathrm{i}=1}^{\mathrm{n}} P(\mathrm{E_i})$.

Auf Grundlage der Wahrscheinlichkeitsaxiome lassen sich wichtige Sätze zum Umgang mit Wahrscheinlichkeiten herleiten, die in den oben genannten Quellen gefunden werden können. Die für die vorliegende Arbeit relevanten Sätze sind in Tabelle 3-1 zusammengefasst.

Tabelle 3-1: Sätze zum Umgang mit Wahrscheinlichkeiten

Wahrscheinlichkeit des komplementären Ereignisses $\mathrm{E^c}$	$P(\mathrm{E^c}) = 1 - P(\mathrm{E})$
Additionssatz stochastisch unabhängiger Ereignisse	$P(\mathrm{E_1} \cup \mathrm{E_2}) = P(\mathrm{E_1}) + P(\mathrm{E_2})$
Multiplikationssatz stochastisch unabhängiger Ereignisse	$P\left(\bigcap_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{E_i}\right) = P(\mathrm{E_1}) \cdot \ldots P(\mathrm{E_n}) = \prod_{\mathrm{i}=1}^{\mathrm{n}} P(\mathrm{E_i})$

3.3.1.1 Zufallsvariablen und ihre Kenngrößen

Zählen alle beliebigen Ereignisse $x \in \Re$ zu den zufälligen Ereignissen, dann wird *X* als **Zufallsvariable** und im n-dimensionalen Fall als **Zufallsvektor** bezeichnet (KOCH 1997, S. 88f, S. 96f). Im Unterschied zu **diskreten Zufallsvariablen** mit endlich vielen oder abzählbar unendlich vielen Werten, wird diese Beschränkung bei **stetigen Zufallsvariablen** aufgehoben. Mit der Wahrscheinlichkeit $P(X \le x)$ eines zufälligen Ereignisses wird schließlich mit $\pi_0(x) = P(X \le x)$ die Wahrscheinlichkeitsverteilung der Zufallsvariablen *X*, als sogenannte **Verteilungsfunktion** $\pi_0(x)$ definiert. Sie gibt die Wahrscheinlichkeit an, mit der eine stetige Zufallsvariable *X* einen Wert kleiner oder gleich *x* annimmt, wobei $P(X \le x) = P(X < x)$ gilt (SACHS 2004, S. 93). Für diskrete Zufallsvariablen gilt $\pi_0(x) = P(X < x)$. Für die nachfolgenden Betrachtungen sind insbesondere stetige Zufallsvariablen von Interesse.

Als Grenzwerte der Verteilungsfunktion gelten $\pi_0(-\infty) = 0$ und $\pi_0(\infty) = 1$ (SACHS 2004, S. 90f). Für eine zumindest im Punkt x stetig differenzierbare Verteilungsfunktion ist:

$$\rho_0(\mathrm{x}) = \frac{\mathrm{d}\pi_0(\mathrm{x})}{\mathrm{dx}} \tag{3.1}$$

die **Wahrscheinlichkeitsdichtefunktion** von $\pi_0(\mathrm{x})$, die auch als Wahrscheinlichkeitsdichte oder Dichtefunktion bezeichnet wird. Beispielsweise ist mit den aus Messungen gewonnenen Werten einer Zufallsvariable über Integration der Wahrscheinlichkeitsdichte die Wahrscheinlichkeit berechenbar:

$$\pi_0(\mathrm{x}) = P(\mathrm{E} \le \mathrm{x}) = \int_{-\infty}^{x} \rho_0(\mathrm{t}) \cdot \mathrm{dt}. \tag{3.2}$$

Entsprechend des Normierungsaxiom ist die Wahrscheinlichkeit eines Ereignisses E definiert als eine reelle Zahl im Intervall $P(\mathrm{E}) \in [0,1]$, so dass $\int_{-\infty}^{+\infty} \rho_0(\mathrm{t}) \cdot \mathrm{dt} = 1$ gilt. Aufgrund dieser Normierungsbedingung folgt die Wahrscheinlichkeit für Realisierungen oberhalb der Grenze x mit:

$$P(\mathrm{E} > \mathrm{x}) = 1 - \pi_0(\mathrm{x}) = \int_{x}^{+\infty} \rho_0(\mathrm{t}) \cdot \mathrm{dt} \tag{3.3}$$

und für die Wahrscheinlichkeit, dass ein beliebiger Wert $\mathrm{x} \in \Re$ in ein Intervall $[\mathrm{a}, \mathrm{b}]$ fällt:

$$P(\mathrm{a} \le \mathrm{x} \le \mathrm{b}) = \pi_0(\mathrm{b}) - \pi_0(\mathrm{a}) = \int_{\mathrm{a}}^{\mathrm{b}} \rho_0(\mathrm{t}) \cdot \mathrm{dt}. \tag{3.4}$$

Eine bei der Behandlung von Zufallsvariablen wichtige Verteilungsfunktion ist die **Gauß'sche Normalverteilung** mit der Wahrscheinlichkeitsdichtefunktion $\rho_{\mathrm{Gauß}}$ und ihrer Verteilungsfunktion $\pi_{\mathrm{Gauß}}$

$$\begin{aligned} \rho_{\mathrm{Gauß}}(\mathrm{x}) &= \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2}\left(\frac{\mathrm{x}-\mu}{\sigma}\right)^2} \qquad \text{mit}: -\infty < \mathrm{x} < \infty, \sigma > 0 \\ \pi_{\mathrm{Gauß}}(\mathrm{x}) &= \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{\mathrm{x}} e^{-\frac{1}{2}\left(\frac{\mathrm{y}-\mu}{\sigma}\right)^2} \cdot \mathrm{dy} \quad \text{mit}: \sigma > 0. \end{aligned} \tag{3.5}$$

Mit dem Erwartungswert μ und der Varianz σ^2 werden die Funktionen vollständig festgelegt. Die Kennzeichnung, dass Zufallsvariablen dieser Verteilung folgen, erfolgt mit: $X \sim N(\mu, \sigma^2)$. *X* heißt dann **normalverteilt**. Eine geschlossene Lösung des Integrals für $\pi_{\mathrm{Gauß}}(\mathrm{x})$ existiert nicht. Zur automatisierten Auswertung wurden daher Näherungslösungen entwickelt (SACHS 2002 S. 145; KOCH 1997, S. 119f)[3].

[3] Gängige Auswerteprogramme, wie das im wissenschaftlichen Umfeld weit verbreitete Softwareprogramm MATLAB des Unternehmens The Mathworks®, stellen hierfür geeignete Berechnungsfunktionen zur Verfügung, vgl. THE MATHWORKS (2013)

Sich auf n-dimensionale Zufallsvariablen $X_1, X_2, \ldots, X_n$ stützende Funktionen $\rho_0(x_1, x_2, \ldots, x_n)$ werden als **multivariate Verteilungen** von $X_1, X_2, \ldots, X_n$ bezeichnet. Eine Erweiterung der zuvor genannten Beziehungen auf n Dimensionen ist im Allgemeinen möglich. Die vorliegende Arbeit beschränkt sich auf diejenigen der Dimension Zwei:

$$\rho_{\text{Gauß}}(x_1, x_2) = \frac{1}{2\pi \cdot \sqrt{\det \boldsymbol{\Sigma}}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\mathbf{x}-\boldsymbol{\mu})} = \frac{1}{2\pi \cdot \sigma_{x_1} \cdot \sigma_{x_2} \cdot \sqrt{1-\rho_{x_1 x_2}^2}} \cdot e^{-\frac{1}{2} \cdot \frac{\left(\frac{x_1-\mu_{x_1}}{\sigma_{x_1}}\right)^2 + \left(\frac{x_2-\mu_{x_2}}{\sigma_{x_2}}\right)^2 - 2 \cdot \rho_{x_1 x_2} \cdot \frac{x_1-\mu_{x_1}}{\sigma_{x_1}} \cdot \frac{x_2-\mu_{x_2}}{\sigma_{x_2}}}{1-\rho_{x_1 x_2}^2}} \tag{3.6}$$

$$\pi_{\text{Gauß}}(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \rho_{\text{Gauß}}(y_1, y_2) \cdot dy_1 \cdot dy_2 = P(X_1 \le x_1, X_2 \le x_2).$$

Darin ist $\rho_{\text{Gauß}}(x_1, x_2)$ die zweidimensionale Wahrscheinlichkeitsdichtefunktion und $\pi_{\text{Gauß}}(x_1, x_2)$ die zweidimensionale Verteilungsfunktion der Gauß' schen Normalverteilung. $\det \boldsymbol{\Sigma}$ kennzeichnet die Determinante der Varianz-Kovarianzmatrix $\boldsymbol{\Sigma}$ und $\boldsymbol{\mu}$ den Vektor der Erwartungswerte, siehe unten. y_1 sowie y_2 sind Integrationsvariablen. Beispielsweise folgt für das Ereignis $X_1 \le x_1, X_2 \le x_2$ bei angenommener Normalverteilung die Wahrscheinlichkeit mit $P(X_1 \le x_1, X_2 \le x_2) = \pi_{\text{Gauß}}(x_1, x_2)$. Zur Herleitung von Verteilungen ist die **Randverteilung** ein wichtiges Hilfsmittel. Mit ihr kann die Dichte mehrdimensionaler Zufallsvariablen auf die Dichte von Zufallsvariablen einer niedrigeren Dimension zurückgeführt werden (KOCH 1997, S. 96f, S. 98f).

Wichtige, aus der Wahrscheinlichkeitsdichtefunktion $\rho_0(x)$ ableitbare Kenngrößen sind der **Erwartungswert** $\mu = E(X)$ und die **Varianz** σ^2 als Maß der Streuung. Des Weiteren sind zur Funktionsbeschreibung die Schiefe und der Exzess (Wölbung) zu nennen, die im Folgenden jedoch nicht benötigt und daher nicht näher betrachtet werden. Mit Hilfe der **Momente** lassen sich die Kenngrößen über ihre Dichtefunktionen angeben, wobei für das k-te Moment einer multivariaten Verteilung gilt (KOCH 1997, S. 104; HARTUNG ET AL. 1995, S. 112f):

$$\mu_{x_i}^{(k)} = E\left(X_i^k\right) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} x_i^k \cdot \rho_0(x_1, x_2, \ldots, x_n) \cdot dx_1 \cdot dx_2 \cdot \ldots \cdot dx_n. \tag{3.7}$$

Das Moment $k = 1$ liefert den Erwartungswert μ_i. Die in Bezug auf den Erwartungswert mit $E\left(\left(X_i - \mu_i\right)^{k_i}\right)$ zentrierten Momente, werden **zentrale Momente** genannt. Das zweite zentrale Moment $k = 2$ liefert die Varianz σ_i^2. Die positive Quadratwurzel der Varianz wird **Standardabweichung** $\sigma_i = +\sqrt{\sigma_i^2}$ genannt.

Für die Erklärung des Zusammenhangs zweier Zufallsvariablen X_i und X_j dient die **Kovarianz**:

$$\begin{aligned} \sigma_{i,j} &= E\left(\left(X_i - \mu_i\right) \cdot \left(X_j - \mu_j\right)\right), \qquad |i \ne j \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(x_i - \mu_i\right) \cdot \left(x_j - \mu_j\right) \cdot \rho_0(x_i, x_j) \cdot dx_i \cdot dx_j. \end{aligned} \tag{3.8}$$

Zum Vergleich der beiden Zufallsvariablen dient die **Korrelation** in Form des normierten **Korrelationskoeffizienten** von X_i und X_j (die Ähnlichkeit der Symbole ρ_0 für die Wahrscheinlichkeitsdichtefunktion und $\rho_{i,j}$ für den Korrelationskoeffizienten ist ausschließlich redaktionell bedingt):

$$\rho_{i,j} = \frac{\sigma_{i,j}}{\sigma_i \cdot \sigma_j}, \quad |i \ne j, \ \sigma_i > 0, \sigma_j > 0. \tag{3.9}$$

Besteht eine **lineare Beziehung** zwischen den beiden Zufallsvariablen, so gilt $\rho_{i,j} = \pm 1$. Für $\rho_{i,j} = 0$ gelten die beiden Zufallsvariablen als **unkorreliert**. In den späteren Abschnitten werden häufig Varianzen und Kovarianzen auf kartesische Koordinaten von Punkten angewendet, weshalb dort die Bezeichnung σ_x^2, σ_y^2 und σ_{xy} verwendet wird.

Aufgrund der stets endlichen Anzahl n an Realisierungen einer Zufallsvariablen, sind die empirisch bestimmten Kenngrößen nur Schätzungen. Für den Erwartungswert gelingt die Schätzung des **empirischen Erwartungswertes** $\bar{x}$ über das arithmetische Mittel und darauf aufbauend die **empirische Varianz** s_x^2. Im Falle zwei- oder mehrdimensionaler Zufallsvariablen eignet sich die Matrizendarstellung. In der **Beobachtungsmatrix L** werden mit $l_{X,1}, l_{X,2}, \ldots, l_{X,n}$ und $l_{Y,1}, l_{Y,2}, \ldots, l_{Y,n}$ die Realisierungen der Zufallsvariablen *X* und *Y* zusammengefasst. Die Kenngrößen werden mit dem **Vektor der Mittelwerte** $\bar{\mathbf{l}}$, der geschätzten **Verbesserungsmatrix v** und der **empirischen Kovarianzmatrix $\mathbf{S_{LL}}$** bezeichnet (PELZER 1985, S. 41f). Sie werden berechnet mit:

eindimensional	zweidimensional
$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$	$\bar{\mathbf{l}}^T = \frac{1}{n} \cdot \mathbf{e} \cdot \mathbf{L} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \cdot \sum_{i=1}^{n} l_{X,i} & \frac{1}{n} \cdot \sum_{i=1}^{n} l_{Y,i} \end{bmatrix}^T$
$s_x^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$	$\mathbf{S}_{LL} = \begin{bmatrix} s_X^2 & s_{XY} \\ s_{YX} & s_Y^2 \end{bmatrix} = \frac{1}{n-1} \cdot \begin{bmatrix} \sum_{i=1}^{n} v_{X,i}^2 & \sum_{i=1}^{n} (v_{X,i} \cdot v_{Y,i}) \\ \sum_{i=1}^{n} (v_{Y,i} \cdot v_{X,i}) & \sum_{i=1}^{n} v_{Y,i}^2 \end{bmatrix} = \frac{1}{n-1} \cdot \mathbf{v} \cdot \mathbf{v}^T$ (3.10)
	mit: $\mathbf{e}^T = [1 \quad 1 \quad \ldots \quad 1]$, $\mathbf{L} = \begin{bmatrix} l_{X,1}, l_{X,2}, \ldots, l_{X,n} \\ l_{Y,1}, l_{Y,2}, \ldots, l_{Y,n} \end{bmatrix}$, $\mathbf{v} = \bar{\mathbf{l}} \cdot \mathbf{e}^T - \mathbf{L}$.

3.3.1.2 Varianz-Kovarianzfortpflanzung

Aus Zufallsvariablen abgeleitete Größen sind ebenfalls Zufallsvariablen. Die Fortpflanzung der Varianz-Kovarianzmatrizen erfolgt nach dem **Varianz-Kovarianzfortpflanzungsgesetz** (NIEMEIER 2002, S. 41f; WITTE & SCHMIDT 2000, S. 65f):

$$\mathbf{\Sigma}_{YY} = \mathbf{A} \cdot \mathbf{\Sigma}_{XX} \cdot \mathbf{A}^T \qquad \text{mit: } \mathbf{Y} = \mathbf{A} \cdot \boldsymbol{X} + \mathbf{b}. \tag{3.11}$$

Die Komponenten des Zufallsvektors **Y** sind darin lineare Funktionen des Zufallsvektors **X**. Diese im Allgemeinen nicht-linearen Funktionen werden durch Beschränkung auf ihre linearen Glieder der Taylor-Reihe approximiert. Dieser Vorgang wird **Linearisierung** genannt. Die Matrix **A** und der Vektor **b** enthalten nicht-stochastische Größen (NIEMEIER 2002, S. 47). Sie sind Konstanten innerhalb linearer Funktionen zwischen den Zufallsvariablen. **A** wird als Jacobische Matrix oder **Funktionsmatrix** bezeichnet (WITTE & SCHMIDT 2000):

$$\mathbf{A} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}. \tag{3.12}$$

In der **Varianz-Kovarianzmatrix** $\mathbf{\Sigma}_{XX}$ sind die a priori bekannten Varianzen und Kovarianzen enthalten. Das allgemeine Varianz-Kovarianzfortpflanzungsgesetz gilt auch bei empirisch ermittelten Varianz-Kovarianzmatrizen. Die Schätzung der Varianzen und Kovarianzen erfolgt in diesem Fall mit:

$$\mathbf{S}_{YY} = \mathbf{A} \cdot \mathbf{S}_{XX} \cdot \mathbf{A}^T. \tag{3.13}$$

Empirisch erfasste Messwerte sind in der Regel Ergebnis einer Summe von unterschiedlichen Einflüssen. Eine wesentliche Begründung für die Annahme, dass für diese Messwerte die Normalverteilung gilt, folgt aus dem **zentralen Grenzwertsatz** (CRAMÉR 1946). Er besagt, dass mit wachsender Anzahl unabhängiger zufälliger Einflussgrößen X_i mit beliebigen Verteilungen, deren Summe asymptotisch gegen die Normalverteilung strebt (NIEMEIER 2002, S. 36). Wenn nicht anders angegeben, wird für alle in der vorliegenden Arbeit verwendeten

Zufallsvariablen angenommen, dass sie normalverteilt sind oder aus normalverteilten Zufallsvariablen resultieren. Die entsprechenden Größen werden mit einem Unterstrich "_" beispielsweise mit $\underline{x}$ hervorgehoben.

Neben den zumeist aus dem geodätischen Umfeld genannten Quellen, finden sich Einführungen zur Wahrscheinlichkeitstheorie unter anderem in HARTUNG ET AL. (1995, S. 91f), HARTUNG & ELPELT (1999, S. 25f), RINNE (2008, S. 173f), ASSENMACHER (2003) und mit zahlreichen Beispielen in FAHRMEIR ET AL. (2010).

3.3.1.3 Hypothesentests

Zur objektiven Beurteilung von empirisch erhobenen Stichproben gegenüber Annahmen ihrer Merkmale oder auch gegenüber anderen Stichproben existieren in der beurteilenden Statistik **statistische Testverfahren**. Zu prüfende Merkmale können Annahmen der zugrundeliegenden Wahrscheinlichkeitsverteilungen oder die Gleichheit von Erwartungswerten gegenüber bekannten Werten sein. Vergleiche mit anderen Stichproben können auf Ebene empirisch erhobener Erwartungswerte oder Varianzen stattfinden. Ist die zugrunde liegende Verteilung bekannt oder definiert, wird von **Parametertests** gesprochen. Wird keine spezifische Verteilung vorausgesetzt, heißen sie nichtparametrische oder **parameterfreie Tests**. Insbesondere bei Prüfungen gegenüber einer hypothetischen Verteilung wird auch von **Anpassungstests** gesprochen (WITTE & SCHMIDT 2000, S. 135; SACHS 2002, S. 193f). Die zugrunde liegenden Annahmen werden als **Hypothesen** bezeichnet und beschreiben in Form der **Nullhypothese** H_0 deren statistische Übereinstimmung sowie mit der **Alternativhypothese** H_A die Nichtübereinstimmung.

In der vorliegenden Arbeit werden Testverfahren auf Basis der Normalverteilung $x_i \sim N$ sowie speziell Tests für die statistische Übereinstimmung $H_0 : E(\bar{x}) = \mu$ des Erwartungswertes μ gegenüber einem Schätzwert (einer Zufallsvariablen) aus der Stichprobe $\bar{x}$ angewendet. Für das Zutreffen der Testentscheidung wird eine Annahmebedingung für H_0 (z. B. $H_0 : \mu_1 = \mu_2$) formuliert, die mit der **Sicherheitswahrscheinlichkeit** P getestet wird. Die Möglichkeit, dass H_0 verworfen wird, obgleich sie richtig ist, wird mit der **Irrtumswahrscheinlichkeit** α_{Irrtum} angegeben, so dass gilt:

$$P = 1 - \alpha_{\text{Irrtum}} \tag{3.14}$$

Dazu wird das Quantil y_P der (angenommenen) Dichtefunktion ρ_0 gegenüber einer **Testgröße** $\underline{T}$ verglichen. Im Allgemeinen ist die Dichtefunktion $\rho_0(x) \in \Re$ aus den Dichtefunktionen $\rho_{\mathbf{x}}(\mathbf{x}) \in \Re^n$ der Beobachtungen $\mathbf{x}$ abzuleiten, wobei für diese in der vorliegenden Arbeit Normalverteilungen angenommen werden. Die Testgröße $\underline{T}$, auch Prüfgröße genannt (WITTE & SCHMIDT 2000, S. 130), folgt als Realisierung aus der Stichprobe $\mathbf{x} = x_1, x_2, \ldots, x_n$ der Beobachtungen. Die Nullhypothese ist anzunehmen, wenn sich der Wert der Testgröße $\underline{T}$ innerhalb des Bereiches befindet, der durch die Sicherheitswahrscheinlichkeit P für die Verteilungsfunktion festgelegt wird. Dieser Bereich wird als **Annahmebereich** A bezeichnet. Entsprechend bezeichnet man den von α_{Irrtum} abgedeckten Bereich der Verteilungsfunktion als **Verwerfungsbereich** V. Zum direkten Vergleich von $\underline{T}$ mit der angenommenen Verteilungsfunktion dient ein kritischer Wert zwischen A und V, der mittels des Quantils y_P ausgedrückt wird:

Fall I: $\underline{T}$ innerhalb A $\rightarrow H_0$ annehmen
Fall II: $\underline{T}$ innerhalb V $\rightarrow H_0$ verwerfen

für

	einseitige Fragestellungen	zweiseitige Fragestellungen
Fall I	$-\infty \le \underline{T} \le y_P$ oder $y_P \le \underline{T} \le \infty$	$y_{P=\alpha_{\text{Irrtum}}/2} \le \underline{T} \le y_{P=1-\alpha_{\text{Irrtum}}/2}$
Fall II	$\infty \ge \underline{T} > y_P$ oder $y_P > \underline{T} \ge -\infty$	$\underline{T} < y_{P=\alpha_{\text{Irrtum}}/2}$ und $y_{P=1-\alpha_{\text{Irrtum}}/2} < \underline{T}$

(3.15)

Das fehlerhafte Verwerfen von H_0 wird mit dem **Fehlerschluss 1. Art** und die fälschliche Annahme der H_0 mit dem **Fehlerschluss 2. Art** bezeichnet. In der Statistik wird in diesem Zusammenhang auch von **statistischen Hypothesentests** oder **Hypothesentests** gesprochen, die letztlich zur Beurteilung der statistischen Hypothesen dienen (VIERTL 1997, S. 125). Wird hierzu nur die Nullhypothese aufgestellt und dient der Test nur zur Prüfung der Annahme dieser Hypothese – in der Regel für $\alpha_{\text{Irrtum}} = 5\%$ –, spricht man auch von einem **Signifikanztest** (SACHS 2002, S. 193; WITTE & SCHMIDT 2000, S. 135). Die vorliegende Arbeit stützt sich im Wesentlichen auf diese Art von Test. Da, entsprechende Motivation vorausgesetzt, bei einigen Identifikationsverfahren prinzipiell auch Alternativhypothesen formulierbar wären, wird der Begriff *Hypothesentest* im Folgenden beibehalten.

Im Vorgriff auf die mengentheoretischen Darstellungen der nachfolgenden Abschnitte können die Hypothesen als klassische Mengen im $\mathbf{u} \in \Re^{\text{u}}$ für die zu prüfenden Parameter $\boldsymbol{\theta}$ im Falle der Nullhypothese $H_0 : \boldsymbol{\theta} \in \Theta_{H_0}$ sowie im Falle der Alternativhypothese $H_{\text{A}} : \boldsymbol{\theta} \in \Theta_{H_{\text{A}}}$ formuliert werden. Ihre eindeutige Testergebniszuordnung erfolgt mittels der Indikatorfunktionen $\text{i}_{H_0}(\mathbf{u})$ und $\text{i}_{H_{\text{A}}}(\mathbf{u})$ mengentheoretisch (NEUMANN 2009, S. 62):

$$\begin{aligned} &H_0 : \boldsymbol{\theta} \in \Theta_{H_0} && && H_{\text{A}} : \boldsymbol{\theta} \in \Theta_{H_{\text{A}}} \\ &\text{mit } \text{i}_{H_0}(\mathbf{u}) = \begin{cases} 1, & \mathbf{u} \in \Theta_{H_0} \\ 0, & \text{sonst} \end{cases} && \text{und} && \text{mit } \text{i}_{H_{\text{A}}}(\mathbf{u}) = \begin{cases} 1, & \mathbf{u} \in \Theta_{H_{\text{A}}} \\ 0, & \text{sonst.} \end{cases} \end{aligned} \tag{3.16}$$

Die Indikatorfunktion wird in Abschnitt 3.3.2 noch detaillierter eingeführt. Im Vektor **u** werden die zu prüfenden Parameter $\boldsymbol{\theta}$ zusammengefasst und mit Θ_{H_0} sowie $\Theta_{H_{\text{A}}}$ werden die sich nicht überlappenden Parameterräume bezeichnet. Der Annahmebereich A und dessen Indikatorfunktion können nun auf Basis der Hypothesen als Abbildung $\text{i}_{\text{A}} : \Theta_{H_0} \to \{0;1\}$ dargestellt werden. Für die nachfolgenden Verfahren ist aus Gründen der Vereinfachung insbesondere der eindimensionale Fall von Interesse. Versteht man den Verwerfungsbereich als mathematisches Komplement zum Annahmebereich $\text{V} = \text{A}^{\text{c}} = 1 - \text{A}$, kann schließlich folgende Definition aufgestellt werden:

$$\begin{aligned} &\text{A} \coloneqq \{x, \text{i}_{\text{A}}(x)\} \\ &\text{V} \coloneqq \{x, \text{i}_{\text{V}}(x) = 1 - \text{i}_{\text{A}}(x)\} \end{aligned} \qquad \text{mit: } \text{i}_{\text{A}} : \Re^{\text{u}} \to \{0,1\} \text{ und } x \in \Re. \tag{3.17}$$

Die Festlegung von Annahme- und Verwerfungsbereich erfolgt mit Hilfe der Irrtumswahrscheinlichkeit α_{Irrtum} des Fehlerschlusses 1. Art. Für die Signifikanzprüfung $H_0 : E(\underline{\text{T}}) = \mu = \mu_0 = 0$ und $H_{\text{A}} : E(\underline{\text{T}}) = \mu = \mu_0 + \delta \neq 0$, folgt der kritische Wert mit $\text{k} \coloneqq y_{1-\alpha_{\text{Irrtum}}}$ und im zweiseitigen Fall mit $\text{k} \coloneqq y_{1-\alpha_{\text{Irrtum}}/2}$, wobei aus Vereinfachungsgründen $\mu_0 = 0$ angenommen wird. Die Region des Annahmebereiches wird beschrieben mit dem Intervall $\text{A} = [0, \text{k}]$ für einseitige und mit $\text{A} = [-\text{k}, \text{k}]$ für zweiseitige Fragestellungen, woraus die Region des Verwerfungsbereiches mit $\text{V} = \text{A}^{\text{c}} = (\text{k}, +\infty)$ und mit $\text{V} = (-\infty, -\text{k}) \cup (\text{k}, +\infty)$ folgt. Die abschließende Testentscheidung gelingt mit der mengentheoretischen Prüfung der Testgröße $\underline{\text{T}}$ gegenüber A und V.

Die Testgröße liegt

Fall I:	ausschließlich im Annahmebereich:	$\underline{\text{T}} \cap \text{V} = \varnothing$	$\to H_0$ annehmen
Fall II:	ausschließlich im Verwerfungsbereich:	$\underline{\text{T}} \cap \text{A} = \varnothing$	$\to H_0$ verwerfen.

(3.18)

Für die Durchführung von Hypothesentests, insbesondere zur Durchführung von Parametertests, eignet sich der in Tabelle 3-2gezeigte Ablauf (JÄGER ET AL. 2005, S. 178f; WITTE & SCHMIDT 2000, S. 134f).

Tabelle 3-2: Ablauf statistischer Hypothesentests (Signifikanztests)

Schritt	Beschreibung
1.	Festlegung der Nullhypothese und gegebenenfalls der Alternativhypothese
2.	Berechnung der Testgröße $\underline{T}$ sowie Feststellung der zutreffenden Verteilung
3.	Definition und Konstruktion des Annahmebereiches A und des Verwerfungsbereiches V unter Berücksichtigung der Irrtumswahrscheinlichkeit α_{Irrtum} (Bestimmung des Quantils)
4.	Testentscheidung durch Prüfung, ob die Testgröße im Annahme- oder Verwerfungsbereich liegt

3.3.1.4 Eine Alternative zur binären Testentscheidung

Als Alternative zum Ablauf statistischer Hypothesentests kann ein Vergleich auch auf Ebene der Wahrscheinlichkeiten erfolgen. Unter Annahme einer gültigen Nullhypothese wird dazu die Wahrscheinlichkeit *p* für das Verwerfen dieser Nullhypothese unter folgenden Bedingungen berechnet:

$$p = \begin{cases} P\left(y_P \le \underline{T} \,|H_0\right) \quad \text{oder} \quad P\left(y_P \ge \underline{T} \,|H_0\right) & \text{einseitige Fragestellung} \\ 2 \cdot \min\left[P\left(y_{P=\alpha_{\text{Irrtum}}/2} \le \underline{T} \,|H_0\right), P\left(y_{P=1-\alpha_{\text{Irrtum}}/2} \ge \underline{T} \,|H_0\right)\right] & \text{zweiseitige Fragestellung} \end{cases} \tag{3.19}$$

Es wird somit die Wahrscheinlichkeit *p* berechnet, dass die Testgröße $\underline{T}$ in den Verwerfungsbereich fällt. Im Unterschied zur Testentscheidung in Abschnitt 3.3.1.3 geht y_P als Variable in die Berechnung ein. Insofern von Gültigkeit der Nullhypothese ausgegangen werden kann, gestattet *p* eine quantitative Beurteilung für das Zutreffen der mit H_0 formulierten Bedingung, für die $P = 1 - p$ der Wahrscheinlichkeitswert ist. Der Wertevergleich von *p* mit α_{Irrtum} liefert letztlich wieder die binäre Testentscheidung:

$$\begin{aligned} p \ge \alpha_{\text{Irrtum}} \quad &\Rightarrow H_0 \text{ annehmen} \\ p < \alpha_{\text{Irrtum}} \quad &\Rightarrow H_0 \text{ verwerfen.} \end{aligned} \tag{3.20}$$

Für alle theoretisch definierten Wahrscheinlichkeitsverteilungen lassen sich exakte Lösungswege formulieren. Lösungswege zur exakten Berechnung der Wahrscheinlichkeitswerte werden beispielsweise in LACKRITZ (1983) für χ^2-verteilte Größen sowie in LACKRITZ (1984) für *F*- und *t*-verteilte Größen gezeigt.

3.3.2 Fuzzy-Theorie

Ausgangspunkt bei der Entwicklung der Theorie der unscharfen Mengen (engl.: fuzzy set theory) war die Erweiterung einer klassischen Menge A der Grundmenge X um Elemente, die nicht eindeutig zu A zugeordnet werden können. Aufgrund der unscharfen Zuordnung entstehen *unscharfen Mengen* (**Fuzzy-Mengen**, engl.: fuzzy sets) (ZADEH 1965). Den Grad der Zugehörigkeit einer Fuzzy-Menge $\tilde{A}$ beschreibt die **Zugehörigkeitsfunktion** $m_{\tilde{A}}(x)$ (engl.: membership function), wodurch die mathematische Behandlung von Fuzzy-Mengen ermöglicht wird. Wesentlicher Unterschied zwischen einer „scharfen“ zu einer „unscharfen“ Mengenfunktion besteht in der Abbildung der Zugehörigkeit auf ein Intervall mit z. B. $[0,1]$ und nicht mehr mit Beschränkung auf die Binärmenge $\{0;1\}$. Eine Fuzzy-Menge kann nun formuliert werden mit:

$$\tilde{A} := \left\{\left(x, m_{\tilde{A}}(x)\right) \;\middle|\; x \in X\right\} \quad \text{mit}: \; m_{\tilde{A}}(x): X \to [0,1]. \tag{3.21}$$

Klassische „scharfe“ Mengen sind als Spezialfälle in der Definition von Fuzzy-Mengen eingebettet, wie die folgende Formulierung mit der Zugehörigkeitsfunktion $m_A(x)$ und der klassischen Menge M zeigt:

$$A := \left\{\left(x, m_A(x)\right) \;\middle|\; x \in X\right\} \quad \text{mit}: m_{\tilde{A}}(x) = \begin{cases} 1, & \text{falls } x \in M \\ 0, & \text{sonst.} \end{cases} \tag{3.22}$$

Das folgende Beispiel nach KAHLERT (1995) verdeutlicht den Unterschied zwischen der „scharfen“ und der „unscharfen“ Zuordnung. Die Zugehörigkeit von „Frost“ auf einer Temperaturskala wird durch den Gefrierpunkt des Wassers bei 0°C als Grenzwert scharf definiert. Die Abbildung 3-2a) zeigt die resultierende scharfe Menge. Da der Grenzwert für „kalt“ vom Gefühl von Personen abhängt, wird ihn jede Person unterschiedlich, auch zeitlich variierend, festlegen. Die in Abbildung 3-2b) mit Beispielwerten gezeigte Menge, ist unscharf.

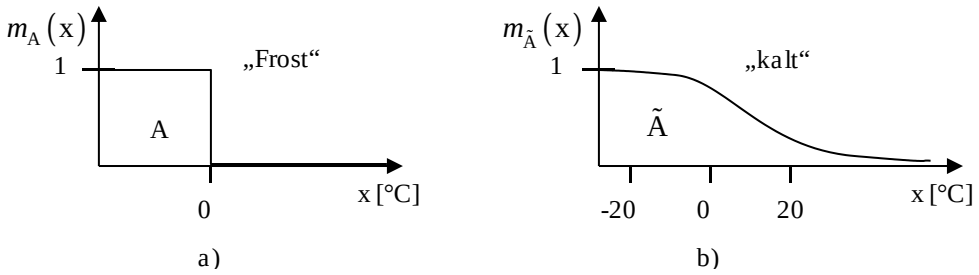


Abbildung 3-2: Beispiel für die Entstehung a) einer klassischen Menge und b) einer unscharfen Menge, nach KAHLERT (1995)

3.3.2.1 Grundlagen und Terminologie der Fuzzy-Theorie

Eine Fuzzy-Menge $\tilde{A}$ heißt **normale** (**normalisierte**) Fuzzy-Menge, wenn die Extrema der Zugehörigkeitsfunktion im Intervall $m_{\tilde{A}}(x) = [0,1] \,|\, \forall x \in \Re$ liegen und mindestens einmal der Zugehörigkeitswert Eins existiert $\exists x_0 \in \Re \,|\, m_{\tilde{A}}(x_0) = 1$, sonst **subnormale Fuzzy-Menge** (DUBOIS & PRADE 1980, S. 26; KAHLERT 1995, S. 16). Eine Fuzzy-Menge kann, abgesehen von einer leeren Fuzzy-Menge, immer normalisiert werden. Sie gilt zudem als konvex, wenn gilt (VIERTL 1996, S. 21; BIEWER 1997 S. 68):

$$\begin{aligned} &\forall\, x_1 \in X \text{ und } \forall\, x_2 \in X \text{ und } \forall\, \lambda \in [0,1]: \\ &\qquad m_{\tilde{A}}\left(\lambda \cdot x_1 + (1-\lambda) \cdot x_2\right) \geq \min\left(m_{\tilde{A}}(x_1), m_{\tilde{A}}(x_2)\right). \end{aligned} \tag{3.23}$$

Die Stellen für x_1 und x_2 dieser Bedingungen sind Abbildung 3-3a) mit den gepunkteten Linien symbolisiert. Die durch x_1 und x_2 begrenzte Menge in X, vgl. Abbildung 3-3b), entspricht der klassischen Menge, einem Intervall. Die vorliegende Arbeit verwendet, wenn nicht anders erwähnt, ausschließlich normalisierte konvexe Fuzzy-Mengen.

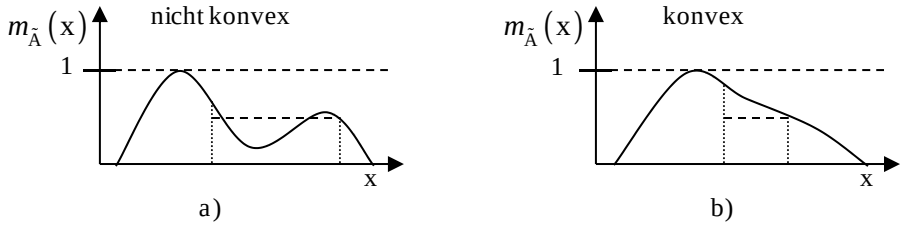


Abbildung 3-3: Beispiel einer a) nicht konvexen und b) konvexen Zugehörigkeitsfunktion (NEUMANN ET AL. 2006, S. 149)

Den nachfolgenden Anwendungen sei als weitere Forderung vorweg genommen, dass die Zugehörigkeitsfunktionen für deren mathematische Behandlung gegen Null streben müssen:

$$\lim_{x \to -\infty} m_{\tilde{A}}(x) = 0 \text{ und } \lim_{x \to \infty} m_{\tilde{A}}(x) = 0. \tag{3.24}$$

Wesentliche Eigenschaften von Fuzzy-Mengen sind Träger, Kern, Höhe, vgl. Abbildung 3-4, sowie die Mächtigkeit.

Der **Träger** einer Fuzzy-Menge (engl.: Support), auch als Einflussbereich oder stützende Menge bezeichnet, enthält alle diejenigen Elemente der Grundmenge X deren Zugehörigkeitsgrad größer Null ist:

$$\mathrm{supp}\left(\tilde{\mathrm{A}}\right)=\left\{\mathrm{x}\in \mathrm{X}\left|m_{\tilde{\mathrm{A}}}\left(\mathrm{x}\right)>0\right.\right\}. \tag{3.25}$$

Der **Kern** (engl.: core) einer Fuzzy-Menge, auch mit Toleranz bezeichnet, ist schließlich eine Untermenge von $\mathrm{supp}\left(\tilde{\mathrm{A}}\right)$ aller derjenigen Elemente mit dem Zugehörigkeitsgrad Eins (für normalisierte Fuzzy-Mengen):

$$\mathrm{core}\left(\tilde{\mathrm{A}}\right)=\left\{\mathrm{x}\in \mathrm{X}\left|m_{\tilde{\mathrm{A}}}\left(\mathrm{x}\right)=1\right.\right\}. \tag{3.26}$$

Die **Höhe** (engl.: height) einer Fuzzy-Menge ist die kleinste obere Grenze (lat.: Supremum) von $m_{\tilde{\mathrm{A}}}\left(\mathrm{x}\right)$ auf X:

$$\mathrm{height}\left(\tilde{\mathrm{A}}\right):=\sup m_{\tilde{\mathrm{A}}}\left(\mathrm{x}\right)\quad\left|\mathrm{x}\in \mathrm{X}\right.. \tag{3.27}$$

Die **Mächtigkeit** (engl.: cardinality) beschreibt die eingeschlossene Fläche der Fuzzy-Menge:

$$\mathrm{card}\left(\tilde{\mathrm{A}}\right)=\int_{\mathrm{X}} m_{\tilde{\mathrm{A}}}\left(\mathrm{x}\right)\cdot \mathrm{dx}\quad\left|\mathrm{x}\in \mathrm{X}\right.. \tag{3.28}$$

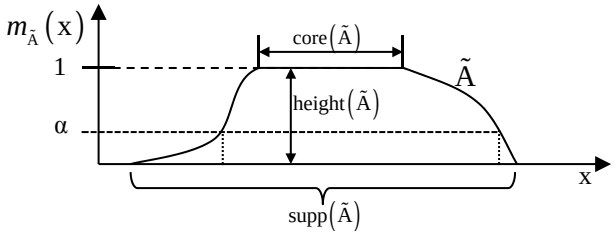


Abbildung 3-4: Träger, Kern und Höhe einer normalisierten konvexen Fuzzy-Menge

Aus dem **α-Schnitt** von $\tilde{\mathrm{A}}$ mit $\alpha\in\left[0,1\right]$ folgt eine scharfe (klassische) Menge mit den Elementen $\mathrm{x}\in \mathrm{X}$ (BOTHE 1995, S. 30). Es wird unterschieden zwischen:

$$\begin{aligned}&\alpha\text{-Schnitt:} && \tilde{\mathrm{A}}_{\alpha}:=\left\{\mathrm{x}\in \mathrm{X}\left|m_{\tilde{\mathrm{A}}}\left(\mathrm{x}\right)\geq\alpha\right.\right\}\\ &\text{strenger }\alpha\text{-Schnitt:} && \tilde{\mathrm{A}}_{\alpha}^{>}:=\left\{\mathrm{x}\in \mathrm{X}\left|m_{\tilde{\mathrm{A}}}\left(\mathrm{x}\right)>\alpha\right.\right\}.\end{aligned} \tag{3.29}$$

Für normierte Fuzzy-Mengen gilt für den Träger $\tilde{\mathrm{A}}_{0}^{>}\equiv\mathrm{supp}\left(\tilde{\mathrm{A}}\right)$ und für den Kern $\tilde{\mathrm{A}}_{1}^{=}\equiv\mathrm{core}\left(\tilde{\mathrm{A}}\right)$, wobei mit $\tilde{\mathrm{A}}_{0}^{>}$ die Menge der Elemente bei dem α-Schnitt $\alpha=0$ sowie deren Zugehörigkeit größer Null, und mit $\tilde{\mathrm{A}}_{1}^{=}$ die Menge der Elemente bei dem α-Schnitt $\alpha=1$ sowie deren Zugehörigkeit gleich Eins angegeben wird.

Mit Hilfe der α-Schnitte lässt sich eine weitere Notation für Fuzzy-Mengen einführen, die auch als **Auflösungsprinzip** bezeichnet wird. Die Fuzzy-Menge kann dadurch mit Hilfe ihre α-Schnitte angegeben werden, ohne auf die Zugehörigkeitsfunktion zugreifen zu müssen (TIZHOOSH 1998, S. 17). Mit Zusammenführung aller α-Schnitte $\tilde{\mathrm{A}}_{\alpha}$ kann die Fuzzy-Menge $\tilde{\mathrm{A}}$ wieder rekonstruiert werden, was sich auch auf die Zugehörigkeitsfunktion übertragen lässt:

$$\tilde{\mathrm{A}}=\int_{\alpha=0}^{\alpha=1}\alpha\cdot\tilde{\mathrm{A}}_{\alpha}\quad\Leftrightarrow\quad m_{\tilde{\mathrm{A}}}\left(\mathrm{x}\right)=\int_{0}^{1} m_{\tilde{\mathrm{A}}_{\alpha}}\left(\mathrm{x}\right)\cdot\mathrm{d}\alpha. \tag{3.30}$$

Jeder α-Schnitt von $\tilde{\mathrm{A}}$ entspricht einer klassischen Menge und damit einem klassischen Intervall mit Untergrenze $\tilde{\mathrm{A}}_{\alpha,\min}=\min\left(\tilde{\mathrm{A}}_{\alpha}\right)$ und Obergrenze $\tilde{\mathrm{A}}_{\alpha,\max}=\max\left(\tilde{\mathrm{A}}_{\alpha}\right)$ sowie in Intervallnotation:

$$\begin{aligned}\tilde{\mathrm{A}}_{\alpha} &= \left[\min\left(\tilde{\mathrm{A}}_{\alpha}\right),\max\left(\tilde{\mathrm{A}}_{\alpha}\right)\right]\\ m_{\tilde{\mathrm{A}},\alpha}\left(x\right) &= \left[\min\left(m_{\tilde{\mathrm{A}},\alpha}\left(x\right)\right),\max\left(m_{\tilde{\mathrm{A}},\alpha}\left(x\right)\right)\right].\end{aligned} \tag{3.31}$$

Zur Konstruktion oder auch Rekonstruktion von Fuzzy-Mengen sei auf die **Indikatorfunktion** i_A hingewiesen:

$$\forall x \in X:\ i_A(x) := m_{\tilde{A}}(x) = \begin{cases} 1, & \text{falls } x \in N \\ 0, & \text{sonst.} \end{cases} \tag{3.32}$$

Man kann sie als Spezialfall einer Zugehörigkeitsfunktion verstehen, da mit ihr eine Menge N von Fuzzy-Teilmengen aus der Grundmenge X selektiert wird. Aufgrund ihrer scharfen Zuordnungseigenschaft {0;1} kann sie auch als alternative Definitionsmöglichkeit für Intervalle und klassische Mengen dienen (VIERTL 2003, S. 9; BIEWER 1997, S. 50). Mit dem Ausdruck $\sup_{\alpha\in(0,1]}$ (supremum) sei als kleinste obere Grenze die Einhüllende der Zugehörigkeitsfunktion $m_{\tilde{A}}(x)$ aus den Teilmengen der α-Schnitte konstruiert:

$$m_{\tilde{A}}(x) = \sup_{\alpha\in(0,1]} \alpha \cdot i_{\tilde{A}_\alpha} \quad |\forall x \in X. \tag{3.33}$$

3.3.2.2 Bildung von Zugehörigkeitsfunktionen

Beim Aufstellen der Zugehörigkeitsfunktionen ist im Allgemeinen „Expertenwissen“ unumgänglich, dass durch Experimente gestützt werden kann. Dabei sind folgende Fragen zu klären (BANDEMER & NÄTHER 1992, S. 92):

- Welches sind die möglichen Beobachtungswerte, deren Möglichkeitsgrad dem des „exakten“ Wertes entspricht?
- Welche Werte können definitiv ausgeschlossen werden?

Aus den Antworten sind der Kern sowie der Träger der Fuzzy-Menge ableitbar. Beispielsweise kann für die Bestimmung einer Zugehörigkeitsfunktion für die Bewegungsgeschwindigkeit von Menschen $m_{\tilde{M}}(v)$ angenommen werden, dass alle Fußgänger Geschwindigkeiten von 0 $\frac{km}{h}$ bis zu 6 $\frac{km}{h}$ erreichen. Dieser Wertebereich eignet sich für die Definition des Kerns. Die Höchstgeschwindigkeit, die Usain Bolt während seines Weltrekords über 100 m im Jahr 2009 aufstellte, betrug 44,72 $\frac{km}{h}$. Geschwindigkeiten von über 45 $\frac{km}{h}$ können also selbst für Hochleistungssportler ausgeschlossen werden, woraus man auf einen Träger der Zugehörigkeitsfunktion in einem Bereich zwischen 0 $\frac{km}{h}$ und 45 $\frac{km}{h}$ schließen kann. Die nachfolgende Abbildung 3-5 veranschaulicht diese Definition.

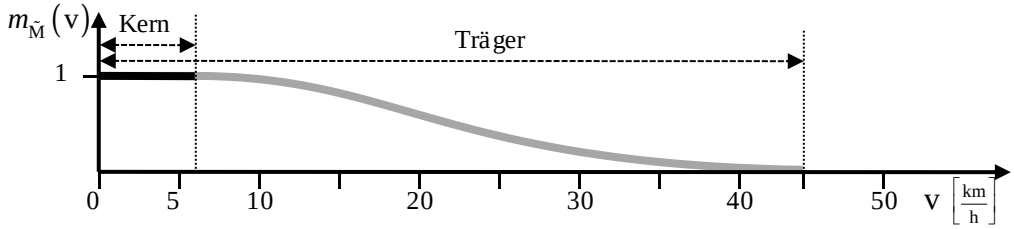


Abbildung 3-5: Beispiel einer Zugehörigkeitsfunktion für die Bewegungsgeschwindigkeit von Menschen

Zur Überbrückung des in der Abbildung 3-5 mit Grau gekennzeichneten Bereiches zwischen dem Kern und den Trägerenden ist eine geeignete monotone Funktion zu finden. Geeignete Funktionen lassen sich häufig dann definieren, wenn der Wert an der Übergangsstelle des Zugehörigkeitswertes 0,5 bekannt ist oder näherungsweise bestimmt werden konnte. DUBOIS & PRADE (1980, S. 255ff) geben weitere praktische Hinweise zur Gewinnung von Zugehörigkeitsfunktionen oder auch deren Ableitung aus empirisch erfassten Daten. Expertenwissen sowie die Forderung nach einer möglichst guten Anpassung an die real beobachtbaren Phänomene stehen auch dort im Vordergrund.

Zur Beschreibung der Funktionseigenschaften sollte die analytische Form der Funktion so einfach wie möglich sein. Während der Monotonienachweis wichtig ist, ist häufig die Abhängigkeit des Ergebnisses von der

analytischen Form der Zugehörigkeitsfunktion nur gering (BANDEMER & NÄTHER 1992, S. 92). Häufig bewirken einfachere analytische Formen auch einfachere Rechenoperationen und wirken sich dadurch günstig auf den rechentechnischen Aufwand aus. In Abhängigkeit von der angestrebten Funktionsbehandlung, kann sich eine darauf abgestimmte Funktionswahl vorteilhaft auswirken. So wählt beispielsweise KUTTERER (2002, S. 47f) für die Formulierung mehrdimensionaler Fuzzy-Mengen elliptische Funktionen, die sich anschließend analytisch elegant verknüpfen lassen.

3.3.2.3 Notationsformen unscharfer Mengen

In der Literatur finden sich je nach Anwendungsspektrum verschiedene Möglichkeiten zur mathematischen Formulierung von Fuzzy-Mengen. Unterschiede finden sich vorrangig in spezifischen Optimierungen für die Datenhaltung, Optimierungen zur rechentechnischen Auswertung oder auch in der Formeldarstellung.

Eine Fuzzy-Menge mit den Elementen als **geordnete Wertepaare**, vgl. auch Abbildung 3-6b), lässt sich formulieren mit (KAHLERT 1995; ZIMMERMANN 2001):

$$\tilde{A} := \left\{ \left(x_1, m_{\tilde{A}}(x_1)\right), \left(x_2, m_{\tilde{A}}(x_2)\right), \ldots, \left(x_n, m_{\tilde{A}}(x_n)\right) \quad | x \in X, n = 1, 2, \ldots \right\} \tag{3.34}$$

und heißt Fuzzy-Menge in X mit den Werten der Zugehörigkeitsfunktion $m_{\tilde{A}}(x)$. Ein einzelnes Wertepaar $\left(x, m_{\tilde{A}}(x)\right)$, wie es Abbildung 3-6a) zeigt, wird auch als **Singleton** bezeichnet. Die Darstellung mit Wertepaaren zeichnet sich durch sogenannte **look-up Tabellen** aus, in denen ausgewählte Funktionswerte der Zugehörigkeitsfunktion (Diskretisierung) abgelegt sind. Die Verarbeitungsgeschwindigkeit ist aufgrund der bereits vorliegenden Funktionswerte sehr hoch. Um eine genügend hohe Genauigkeit zu erzielen, ist jedoch sehr fein zu diskretisieren, wodurch hohe Datenmengen zu verwalten sind. Ebenso kann die begrenzte Menge der n auswertbaren Werte nachteilig sein.

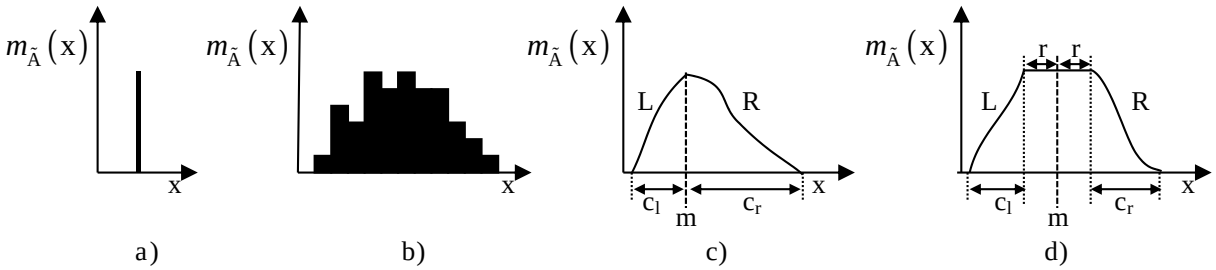


Abbildung 3-6: Darstellung der Zugehörigkeitsfunktion mit geordneten Wertepaaren a) als Singleton und b) mit mehreren Elementen sowie nach dem LR-Typ c) als Fuzzy-Zahl und d) Fuzzy-Intervall

Günstige Formulierungen für die analytische Arbeit mit Fuzzy-Mengen lassen sich mit Fuzzy-Zahlen und Fuzzy-Intervallen nach DUBOIS & PRADE (1980) angeben. Die Flanken der Zugehörigkeitsfunktionen werden darin mittels Referenzfunktionen beschrieben. Prinzipiell sind alle nicht-negativen, streng monoton fallenden Funktionen sowohl für die linke L als auch für die rechte R Referenzfunktion mit $L(0) = R(0) = 1$ gestattet (BANDEMER & GOTTWALD 1990, S. 49). Man spricht von **Fuzzy-Zahlen vom LR-Typ**, vgl. Abbildung 3-6c), und von **Fuzzy-Intervallen vom LR-Typ** (DUBOIS & PRADE 1980), vgl. Abbildung 3-6d).

Bezüglich des Mittelpunktes m und des Radius r für die Größe des Kerns erfolgt die Definition der Zugehörigkeitsfunktion eines LR-Fuzzy-Intervalls mit Hilfe der Spannweiten $c_l > 0$ für die linke und $c_r > 0$ die rechte Referenzfunktion mit Formel (3.35) (BOTHE 1995, S. 81f). Die Bezeichnung LR-Typ mit $\tilde{A} = (\bullet)_{LR}$ weist bereits auf die unterschiedlichen Referenzfunktionstypen für die linke und die rechte Flanke hin. Sind beide Referenzfunktionen vom identischen Typ, spricht man vom LL-Typ mit $\tilde{A} = (\bullet)_{LL}$. Existiert schließlich

Totalsymmetrie, so dass für die Spannweiten $s = c_l = c_r$ gilt, wird vom L-Typ mit $\tilde{A} = (\bullet)_L$ gesprochen. Im Allgemeinen kann man Fuzzy-Intervalle vom LR-Typ formulieren mit:

$$\tilde{A} = (m, r, c_l, c_r)_{LR} \Leftrightarrow m_{\tilde{A}}(x) = \begin{cases} 0 & ,(x \le m - r - c_l) \vee (m + r + c_r \le x) \\ L\left(\dfrac{m - x - r}{c_l}\right) & ,m - r - c_l < x < m - r \\ 1 & ,m - r \le x \le m + r \\ R\left(\dfrac{x - m - r}{c_r}\right) & ,m + r < x < m + r + c_r . \end{cases} \tag{3.35}$$

Für $r = 0$ vereinfachen sich die Formeln und es resultieren die Fuzzy-Zahlen vom LR-Typ $\tilde{A} = (m, c_l, c_r)_{LR}$. Im Falle totalsymmetrischer Referenzfunktionen folgt für Fuzzy-Zahlen $\tilde{A} = (m, s)_{LR}$. In der Praxis haben sich vor allem Zugehörigkeitsfunktionen mit linearen Referenzfunktionen wegen ihres geringen Bedarfs an Speichervolumen sowie Rechenzeit und dennoch günstigen Übertragungscharakteristiken für Steuerungen und Regelungen durchgesetzt (KAHLERT 1995, S. 15). Aufgrund ihrer charakteristischen Zugehörigkeitsfunktion heißen sie auch **dreieckige Fuzzy-Zahlen** bzw. **trapezförmige Fuzzy-Intervalle**.

3.3.2.4 Mengen- und Rechenoperationen mit Fuzzy-Mengen

Die Verknüpfung von Fuzzy-Mengen erfolgt mittels mengentheoretischer Operationen UND und ODER. Es gelten mit jeweils $\forall x \in X$:

$$\begin{array}{lll} \text{Vereinigungsmenge:} & \tilde{C} = \tilde{A} \cup \tilde{B} & \Leftrightarrow \quad m_{\tilde{C}}(x) = \max\left(m_{\tilde{A}}(x), m_{\tilde{B}}(x)\right) \\ \text{Schnittmenge:} & \tilde{C} = \tilde{A} \cap \tilde{B} & \Leftrightarrow \quad m_{\tilde{C}}(x) = \min\left(m_{\tilde{A}}(x), m_{\tilde{B}}(x)\right) \\ \text{Komplement:} & \tilde{C} = \tilde{A}^c & \Leftrightarrow \quad m_{\tilde{C}}(x) = 1 - m_{\tilde{A}}(x) \end{array} \tag{3.36}$$

Zur Realisierung der Operatoren sind bei unscharfen Mengen verschiedene Methoden denkbar. Obwohl alle Methoden bei Übergang auf klassische Mengen die gleichen Eigenschaften wie die Operatoren der klassischen Mengenlehre aufweisen (KAHLERT 1995, S. 20f), entstehen bei Anwendung auf unscharfe Mengen Vieldeutigkeiten. Um diese Vieldeutigkeiten zu umgehen, wurde das Konzept der sogenannten **t-Normen** und **s-Normen** eingeführt (BANDEMER & GOTTWALD 1990). Die mengentheoretisch mit dem Minimum-Operator $\min(\bullet)$ gebildete Schnittmenge, wird als t-Norm bezeichnet und entspricht dem „logischen UND“. Die Vereinigungsmenge entsteht mengentheoretisch mittels des Maximum-Operators $\max(\bullet)$. Sie wird als s-Norm bezeichnet und steht für das „logische inklusive ODER“ (GOODMAN & NGUYEN 1985; BOTHE 1995, S. 41f; ZIMMERMANN 2001, S. 29f; SPIES 1993, S. 228f; TIZHOOSH 1998, S. 228f). Es sei in diesem Zusammenhang angemerkt, dass es keine Vorschrift zur Wahl der „richtigen“ Methode gibt. Vielmehr ist anwendungsbezogen und dem Sachverhalt geeignet auszuwählen. Für alle gilt jedoch, dass die für t und s einsetzbaren Verknüpfungen den folgenden Bedingungen (T1) bis (T4) für die t-Normen sowie (S1) bis (S4) für die s-Normen mit $x, y, u, v \in [0,1]$ genügen müssen:

$$\begin{array}{llll} (T1) & t(x,y) = t(y,x) & (S1) & s(x,y) = s(y,x) \\ (T2) & t(x, t(y,z)) = t(t(x,y), z) & (S2) & s(x, s(y,z)) = s(s(x,y), z) \\ (T3) & \begin{array}{l} x \le u \wedge y \le v \quad \Rightarrow t(x,y) \le t(u,v) \\ u \le v, \forall x \in [0,1] \Rightarrow t(x,u) \le t(x,v) \end{array} & (S3) & \begin{array}{l} x \le u \wedge y \le v \quad \Rightarrow s(x,y) \le s(u,v) \\ u \le v, \forall x \in [0,1] \Rightarrow s(x,u) \le s(x,v) \end{array} \\ (T4) & t(x,1) = x, \quad t(x,0) = 0 & (S4) & s(x,1) = 1, \quad s(x,0) = x \end{array} \tag{3.37}$$

Die Anzahl der Methoden, die diese Bedingungen erfüllen, ist im Allgemeinen unbegrenzt. Die vorliegende Arbeit beschränkt sich daher auf die in Tabelle 3-3 gezeigten, von Zadeh eingeführten Varianten.

Tabelle 3-3: Beispiele für t-Normen und s-Normen

t-Norm		s-Norm	
Minimum	$t_0(u,v)=\min(u,v)$	Maximum	$s_0(u,v)=\max(u,v)$
algebraisches Produkt	$t_1(u,v)=u\cdot v$	algebraische Summe	$s_1(u,v)=u+v-u\cdot v$
beschränktes Produkt	$t_2(u,v)=\max(u+v-1,0)$	beschränkte Summe	$s_2(u,v)=\min(u+v,1)$

Für diese ausgewählten t- und s-Normen wird im Folgenden die Schreibweise mit Zugehörigkeitsfunktion und $\forall x \in X$ angegeben (BANDEMER & GOTTWALD 1990, S. 32f; BOTHE 1995, S. 42f):

Minimum: $\tilde{C}=\min(\tilde{A},\tilde{B}) \Leftrightarrow m_{\tilde{C}=\min(\tilde{A},\tilde{B})}(x)=\min(m_{\tilde{A}}(x),\, m_{\tilde{B}}(x))$ (3.38)

Maximum: $\tilde{C}=\max(\tilde{A},\tilde{B}) \Leftrightarrow m_{\tilde{C}=\max(\tilde{A},\tilde{B})}(x)=\max(m_{\tilde{A}}(x),\, m_{\tilde{B}}(x))$ (3.39)

algebraisches Produkt: $\tilde{C}=\tilde{A}\cdot\tilde{B} \Leftrightarrow m_{\tilde{C}=\tilde{A}\cdot\tilde{B}}(x)=m_{\tilde{A}}(x)\cdot m_{\tilde{B}}(x)$ (3.40)

algebraische Summe: $\tilde{C}=\tilde{A}+\tilde{B} \Leftrightarrow m_{\tilde{C}=\tilde{A}+\tilde{B}}(x)=m_{\tilde{A}}(x)+m_{\tilde{B}}(x)-m_{\tilde{A}}(x)\cdot m_{\tilde{B}}(x)$ (3.41)

beschränktes Produkt: $\tilde{C}=\tilde{A}\sqcap\tilde{B} \Leftrightarrow m_{\tilde{C}=\tilde{A}\sqcap\tilde{B}}(x)=\max(0,\, m_{\tilde{A}}(x)+m_{\tilde{B}}(x)-1)$ (3.42)

beschränkte Summe: $\tilde{C}=\tilde{A}\sqcup\tilde{B} \Leftrightarrow m_{\tilde{C}=\tilde{A}\sqcup\tilde{B}}(x)=\min(1,\, m_{\tilde{A}}(x)+m_{\tilde{B}}(x))$. (3.43)

Die Abbildung 3-7a) und Abbildung 3-7b) zeigen an einem Beispiel mit jeweils gleichen Fuzzy-Mengen $\tilde{A}$ und $\tilde{B}$ die Ergebnismengen der Formeln (3.40) bis (3.43).

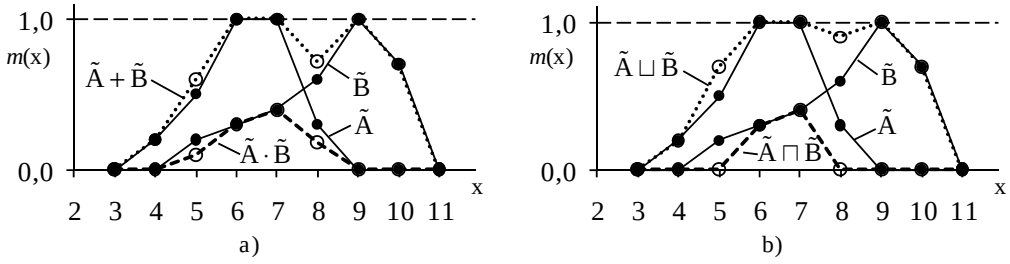


Abbildung 3-7: Verlauf der Zugehörigkeitsfunktionen für a) algebraisches Produkt und algebraische Summe sowie für b) beschränktes Produkt und beschränkte Summe

Die Verknüpfung $\times$ von Fuzzy-Mengen basiert auf dem sogenannten **Zadeh'schen Erweiterungsprinzip**, wonach eine Funktion $g: X^n \to Y$ zu einer Funktion auf Fuzzy-Mengen $\tilde{g}: \mathfrak{F}(X^n) \to \mathfrak{F}(Y)$ erweitert werden kann. Mit dem Supremum $\sup(\bullet)$ als asymptotischer Maximalwert sowie $\sup(\varnothing)=0$ folgt (KUTTERER 2002, S. 41; NEUMANN 2009, S. 42):

$$\tilde{B}=\tilde{g}(\tilde{A}_1,\ldots,\tilde{A}_n)=\tilde{g}(\tilde{\mathbf{A}}) \Leftrightarrow m_{\tilde{B}}(y)=\sup_{\substack{(x_1,\ldots,x_n)\in X_1\times\ldots\times X_n \\ g(x_1,\ldots,x_n)=y}} \min\left(m_{\tilde{A}_1}(x_1),\ldots,m_{\tilde{A}_n}(x_n)\right) \quad \forall y\in X. \tag{3.44}$$

Das kartesische Produkt ist darin eine wesentliche Verknüpfungsform, das auf Fuzzy-Mengen angewendet als **unscharfes kartesisches Produkt** bezeichnet wird (BANDEMER & GOTTWALD 1990, S. 37f) und Verknüpfungen zwischen unterschiedlichen Grundmengen erlaubt (z. B. $\tilde{A}\times\tilde{B}\in\mathfrak{F}(X\times Y)$ für $\tilde{A}\in\mathfrak{F}(X)$ und $\tilde{B}\in\mathfrak{F}(X)$).

Mit $\tilde{\mathbf{A}}$ wird in Formel (3.44) eine n-dimensionale Menge von Fuzzy-Mengen mit ihren Zugehörigkeitsfunktionen $m_{\tilde{A}}(\mathbf{x})$ für $\mathbf{x}\in X^n$ bezeichnet. Formel (3.21) kann dementsprechend auf n-Dimensionen erweitert

werden, wenn je Fuzzy-Menge mindestens einmal der Wert Eins existiert und die α-Schnitte kompakte klassische Mengen sind (VIERTL 1996, S. 33f; NEUMANN 2009, S. 42; KUTTERER 2002, S. 41):

$$\tilde{\mathbf{A}} := \left\{ \left(\mathbf{x}, m_{\tilde{A}}(\mathbf{x}) \right) \;\middle|\; \mathbf{x} \in X^n \right\} \quad \text{mit} : m_{\tilde{A}}(\mathbf{x}) : X^n \to [0,1] \tag{3.45}$$

$$\exists\, \mathbf{x_0} \in X^n : m_{\mathbf{A}}(\mathbf{x_0}) = 1 \quad \text{und} \quad \tilde{\mathbf{A}}_\alpha(x) := \left\{ \mathbf{x} \in X^n, m_{\tilde{A}}(\mathbf{x}) \geq \alpha \right\} \;\middle|\; \forall \alpha \in (0,1]. \tag{3.46}$$

Für den Fall, dass Abbildungen mehrerer Kombinationen $x_1, \ldots, x_n$ in Formel (3.44) denselben scharfen Funktionswert y liefern, ist das Minimum der beteiligten Zugehörigkeitsfunktionen zu nehmen. Die Konstruktion von $m_{\tilde{B}}(y)$ erfolgt anschließend als Supremum der Minimalwerte aller Kombinationsmöglichkeiten. Eine wesentliche Eigenschaft der Fuzzy-Theorie ist, dass die Funktionsauswertung auch α-schnittweise durchführbar ist (BANDEMER & GOTTWALD 1990, S. 41; BANDEMER & NÄTHER 1992, S. 26):

$$\tilde{B}_\alpha = g\left(\tilde{A}_{1,\alpha}, \ldots, \tilde{A}_{n,\alpha} \right) = g\left(\tilde{A}_\alpha \right) \;\middle|\; \forall\, \alpha \in (0,1]. \tag{3.47}$$

Mit den α-Schnitten konvexer Zugehörigkeitsfunktionen kann deren Funktionsauswertung auf die Intervallarithmetik zurückgeführt werden.

Angewendet auf Fuzzy-Zahlen sowie Fuzzy-Intervalle vom LR-Typ folgt mit den Umkehrfunktionen L^{-1} und R^{-1} der streng monotonen Referenzfunktionen L und R je α-Schnitt $\tilde{A}_\alpha$ das Intervall:

$$\begin{aligned} &\text{für LR-Fuzzy-Zahlen:} && \tilde{A}_\alpha = \left[\tilde{A}_{\alpha,\min}, \tilde{A}_{\alpha,\max} \right] = \left[m - L^{-1}(\alpha), m + R^{-1}(\alpha) \right]_{LR} \quad \text{und} \\ &\text{für LR-Fuzzy-Intervalle:} && \tilde{A}_\alpha = \left[\tilde{A}_{\alpha,\min}, \tilde{A}_{\alpha,\max} \right] = \left[m - r - L^{-1}(\alpha), m + r + R^{-1}(\alpha) \right]_{LR}. \end{aligned} \tag{3.48}$$

Die Funktionsauswertung in Formel (3.47) kann dann geschrieben werden mit:

$$\tilde{B}_\alpha = \left[\tilde{B}_{\alpha,\min}, \tilde{B}_{\alpha,\max} \right] = \left[\min\left(g(x_1, \ldots, x_n) \right), \max\left(g(x_1, \ldots, x_n) \right) \right] \quad \text{mit} : \; x_1 \in \tilde{A}_{1,\alpha}, \ldots, x_n \in \tilde{A}_{n,\alpha}. \tag{3.49}$$

Mit Hilfe des Zadeh'schen Erweiterungsprinzips wird dadurch die Funktionsauswertung α-schnittweise auf eine Wertebereichssuche der Eingangsintervalle zurückgeführt (DUBOIS & PRADE 1980, S. 37). Aufgrund der $\min(\bullet)$ und $\max(\bullet)$ Operatoren liefert diese Optimierungsaufgabe eine konvexe Einschließung der Zugehörigkeitsfunktion $m_{\tilde{B}}$ der Zielgröße. Das gilt ebenfalls bei Eingangsgrößen mit nicht-konvexen Zugehörigkeitsfunktionen (VIERTL 1996, S. 20).

Basierend auf dem Zadeh'schen Erweiterungsprinzip lassen sich auch die Grundrechenarten $(+, -, \cdot, /)$ für die Anwendung auf Fuzzy-Zahlen überführen. Sie werden als erweiterte Addition $\oplus$, erweiterte Subtraktion $\ominus$, erweiterte Multiplikation (auch erweitertes Produkt) $\odot$ und erweiterte Division $\oslash$ bezeichnet. Da in der vorliegenden Arbeit Fuzzy-Intervalle des LL-Typs Anwendung finden, wird nachfolgend deren Arithmetik beschrieben. Mit Hilfe der Konstruktion über die Indikatorfunktion aus Formel (3.33) wird zudem die α-schnittweise Auswertung mittels der Intervallarithmetik angegeben. Mit den Fuzzy-Intervallen:

$$\tilde{A} = \left(m_{\tilde{A}}, r_{\tilde{A}}, c_{\tilde{A},l}, c_{\tilde{A},r} \right)_{LL} \quad \text{und} \quad \tilde{B} = \left(m_{\tilde{B}}, r_{\tilde{B}}, c_{\tilde{B},l}, c_{\tilde{B},r} \right)_{LL} \tag{3.50}$$

gelingt die Überführung auf die entsprechenden Fuzzy-Zahlen leicht, indem der Radius des Kerns $r = 0$ gesetzt wird. Insbesondere die Arithmetikoperationen $\odot$ und $\oslash$ erfordern in der Regel Näherungslösungen. Effiziente und übersichtliche Lösungswege lassen sich mit Übergang auf die α-schnittweise Auswertung finden. Auf die resultierenden klassischen Mengen der α-Schnitte, können dann die auf die Intervallarithmetik zurückgehenden Operationen angewendet werden (BANDEMER & GOTTWALD 1990, S. 43f; BECTOR & CHANDRA 2005; BIEWER 1997, S. 127f; BOTHE 1995; DUBOIS & PRADE 1980, S. 53f; KAUFMANN & GUPTA 1991).

Für die **erweiterte Addition** gilt:

$$\begin{aligned}\tilde{A}\oplus\tilde{B} &= \left(m_{\tilde{A}}-r_{\tilde{A}}+m_{\tilde{B}}-r_{\tilde{B}},\ m_{\tilde{A}}+r_{\tilde{A}}+m_{\tilde{B}}+r_{\tilde{B}},\ c_{\tilde{A},l}+c_{\tilde{B},l},\ c_{\tilde{A},r}+c_{\tilde{B},r}\right)_{LL}\\ &\Leftrightarrow m_{\tilde{A}\oplus\tilde{B}}(x)=\sup_{\alpha\in(0,1]}\alpha\cdot i_{(\tilde{A}\oplus\tilde{B})_\alpha}(x)\quad\left|\forall x\in X,\ i_{(\tilde{A}\oplus\tilde{B})_\alpha}=\left[\tilde{A}_{\alpha,min}+\tilde{B}_{\alpha,min},\tilde{A}_{\alpha,max}+\tilde{B}_{\alpha,max}\right]\right.\end{aligned}. \tag{3.51}$$

Ein **negatives LL-Fuzzy-Intervall** $\tilde{B}^-$ ist definiert durch:

$$-\tilde{B}=\tilde{B}^-=-\left(m_{\tilde{B}},r_{\tilde{B}},c_{\tilde{B},l},c_{\tilde{B},r}\right)_{LL}=\left(-m_{\tilde{B}},r_{\tilde{B}},c_{\tilde{B},l},c_{\tilde{B},r}\right)_{LL}. \tag{3.52}$$

Diese Negation heißt auch monadisches Minus. Die Differenz zweier LL-Fuzzy-Intervalle folgt mit:

$$\begin{aligned}\tilde{A}\ominus\tilde{B} &= \tilde{A}\oplus\tilde{B}^-\\ &=\left(m_{\tilde{A}}-r_{\tilde{A}}-\left(m_{\tilde{B}}+r_{\tilde{B}}\right),\ m_{\tilde{A}}+r_{\tilde{A}}-\left(m_{\tilde{B}}-r_{\tilde{B}}\right),\ c_{\tilde{A},l}+c_{\tilde{B},r},\ c_{\tilde{A},r}+c_{\tilde{B},l}\right)_{LL}\\ &\Leftrightarrow m_{\tilde{A}\ominus\tilde{B}}(x)=\sup_{\alpha\in(0,1]}\alpha\cdot i_{(\tilde{A}\ominus\tilde{B})_\alpha}(x)\quad\left|\forall x\in X, i_{(\tilde{A}\ominus\tilde{B})_\alpha}=\left[\tilde{A}_{\alpha,min}-\tilde{B}_{\alpha,max},\tilde{A}_{\alpha,max}-\tilde{B}_{\alpha,min}\right]\right.\end{aligned}. \tag{3.53}$$

Die **Subtraktion der LL-Fuzzy-Intervalle** ist demnach nicht invers zur Addition. Die Multiplikation mit einer „scharfen" reellen Zahl $\lambda\in\Re$ erfolgt mit:

$$\begin{aligned}\lambda\odot\tilde{A} &= \left(\lambda\cdot m_{\tilde{A}},|\lambda|\cdot r_{\tilde{A}},|\lambda|\cdot c_{\tilde{A},l},|\lambda|\cdot c_{\tilde{A},r}\right)_{LL}\\ &\Leftrightarrow m_{\lambda\odot\tilde{A}}(x)=\sup_{\alpha\in(0,1]}\alpha\cdot i_{(\lambda\odot\tilde{A})_\alpha}(x)\quad\left|\forall x\in X, i_{(\lambda\odot\tilde{A})_\alpha}=\begin{cases}\left[\lambda\cdot\tilde{A}_{\alpha,min},\lambda\cdot\tilde{A}_{\alpha,max}\right], & \lambda>0\\ [0,0], & \lambda=0\\ \left[\lambda\cdot\tilde{A}_{\alpha,max},\lambda\cdot\tilde{A}_{\alpha,min}\right], & \lambda<0.\end{cases}\right.\end{aligned} \tag{3.54}$$

Im Falle der **Division durch eine „scharfe" reelle Zahl** $\lambda'\in\Re$ kann mit $\lambda=1/\lambda'$ für $\lambda'\neq 0$ ebenfalls auf die hier genannte Multiplikation zurückgegriffen werden. Insofern $\tilde{A}\oplus\tilde{B}$, $\tilde{A}\ominus\tilde{B}$ oder $\lambda\odot\tilde{A}$ für dreieckige Fuzzy-Zahlen oder trapezförmige Fuzzy-Intervalle α-schnittweise auszuwerten sind, genügt es die Schnitte bei $\alpha=0$ und $\alpha=1$ zu berechnen.

Für die **erweiterte Multiplikation** und **erweiterte Division** können derartigen Vereinfachungen im Allgemeinen weder für LR-, LL- noch für L-Fuzzy-Intervalle angewendet werden, da die Referenzfunktionen der Resultate ungleich derjenigen der Eingangsgrößen sind. Zudem sind die Referenzfunktionen meist nur näherungsweise lösbar. Daneben existieren eine Reihe von weiteren Vorschlägen die bessere Approximationen an die Zielgröße und/oder effizientere Rechenwege anvisieren (GIACHETTI & YOUNG 1997; OUSSALAH & SCHUTTER 2003).

Für die erweiterte Multiplikation folgt mit Beschränkung auf die α-schnittweise Auswertung:

$$\begin{aligned}\tilde{A}\odot\tilde{B} &:= m_{\tilde{A}\odot\tilde{B}}(x)\approx\sup_{\alpha\in(0,1]}\alpha\cdot i_{(\tilde{A}\odot\tilde{B})_\alpha}(x)\\ &\left|\forall x\in X, i_{(\tilde{A}\odot\tilde{B})_\alpha}=\begin{bmatrix}\min\left(\tilde{A}_{\alpha,min}\cdot\tilde{B}_{\alpha,min},\tilde{A}_{\alpha,min}\cdot\tilde{B}_{\alpha,max},\tilde{A}_{\alpha,max}\cdot\tilde{B}_{\alpha,min},\tilde{A}_{\alpha,max}\cdot\tilde{B}_{\alpha,max}\right),\\ \max\left(\tilde{A}_{\alpha,min}\cdot\tilde{B}_{\alpha,min},\tilde{A}_{\alpha,min}\cdot\tilde{B}_{\alpha,max},\tilde{A}_{\alpha,max}\cdot\tilde{B}_{\alpha,min},\tilde{A}_{\alpha,max}\cdot\tilde{B}_{\alpha,max}\right)\end{bmatrix}\right.\end{aligned}. \tag{3.55}$$

Die Zugehörigkeitsfunktion $m_{\tilde{A}\odot\tilde{B}}(x)$ für die erweiterte Multiplikation zeigt die Abbildung 3-8 an einem Beispiel. Mit Forderung das $0\notin\operatorname{supp}\left(\tilde{B}\right)$ folgt der **Kehrwert** $\tilde{B}^{-1}=1/\tilde{B}$ mit:

$$\begin{aligned}\tilde{B}^{-1} &:= m_{\tilde{B}^{-1}}(x) = \begin{cases}m_{\tilde{B}}(1/x), & \forall x\in\operatorname{supp}\left(\tilde{B}\right)\ \left|0\notin\operatorname{supp}\left(\tilde{B}\right)\right.\\ 0, & \text{sonst.}\end{cases}\\ &\approx\sup_{\alpha\in(0,1]}\alpha\cdot i_{\left(\tilde{B}^{-1}\right)_\alpha}(x)\quad\left|\forall x\in X,\ i_{\left(\tilde{B}^{-1}\right)_\alpha}=\left[\frac{1}{\tilde{B}_{\alpha,max}},\frac{1}{\tilde{B}_{\alpha,min}}\right]\right.\end{aligned}. \tag{3.56}$$

Für die erweiterte Division kann der Quotient auch als erweiterte Multiplikation $\tilde{A} \oslash \tilde{B} = \tilde{A} \odot \tilde{B}^{-1}$ für $0 \notin \text{supp}(\tilde{B})$ gebildet werden mit, vgl. das Beispiel in Abbildung 3-8:

$$\tilde{A} \oslash \tilde{B} := m_{\tilde{A} \oslash \tilde{B}}(x) \approx \sup_{\alpha \in (0,1]} \alpha \cdot i_{(\tilde{A} \oslash \tilde{B})_\alpha}(x)$$

$$\left| \forall x \in X, 0 \notin \text{supp}(\tilde{B}), \quad i_{(\tilde{A} \oslash \tilde{B})_\alpha} = \left[\min\left(\frac{\tilde{A}_{\alpha,\min}}{\tilde{B}_{\alpha,\min}}, \frac{\tilde{A}_{\alpha,\min}}{\tilde{B}_{\alpha,\max}}, \frac{\tilde{A}_{\alpha,\max}}{\tilde{B}_{\alpha,\min}}, \frac{\tilde{A}_{\alpha,\max}}{\tilde{B}_{\alpha,\max}} \right), \max\left(\frac{\tilde{A}_{\alpha,\min}}{\tilde{B}_{\alpha,\min}}, \frac{\tilde{A}_{\alpha,\min}}{\tilde{B}_{\alpha,\max}}, \frac{\tilde{A}_{\alpha,\max}}{\tilde{B}_{\alpha,\min}}, \frac{\tilde{A}_{\alpha,\max}}{\tilde{B}_{\alpha,\max}} \right) \right] \right. \quad (3.57)$$

$$= \left[\begin{array}{l} \min\left(\tilde{A}_{\alpha,\min} \cdot \tilde{B}^{-1}_{\alpha,\min}, \tilde{A}_{\alpha,\min} \cdot \tilde{B}^{-1}_{\alpha,\max}, \tilde{A}_{\alpha,\max} \cdot \tilde{B}^{-1}_{\alpha,\min}, \tilde{A}_{\alpha,\max} \cdot \tilde{B}^{-1}_{\alpha,\max} \right), \\ \max\left(\tilde{A}_{\alpha,\min} \cdot \tilde{B}^{-1}_{\alpha,\min}, \tilde{A}_{\alpha,\min} \cdot \tilde{B}^{-1}_{\alpha,\max}, \tilde{A}_{\alpha,\max} \cdot \tilde{B}^{-1}_{\alpha,\min}, \tilde{A}_{\alpha,\max} \cdot \tilde{B}^{-1}_{\alpha,\max} \right) \end{array} \right].$$

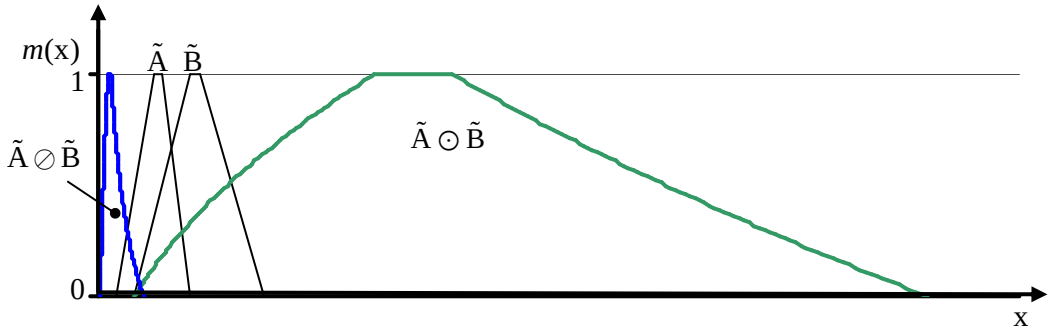


Abbildung 3-8: Zugehörigkeitsfunktionen der erweiterten Multiplikation und erweiterten Division

Fuzzy-Mengen der Dimension zwei oder mehr entstehen beispielsweise bei der Verknüpfung von Fuzzy-Mengen unterschiedlicher Grundgesamtheiten. Ein Beispiel für die Dimension Zwei ist die Verknüpfung $\tilde{C} = \tilde{A} \times \tilde{B}$ mit $\tilde{A} \subset X, \tilde{B} \subset Y$. Sind X und Y z. B. Grundgesamtheiten der reellen Zahlen, dann ist $\tilde{C}$ in $\Re^2$ und die Zugehörigkeitsfunktion $m_{\tilde{C}}(x,y)$ mit $x,y \in \Re^2$ definiert. $\tilde{C}$ ist damit Ergebnis der unscharfen Relation $\tilde{A} \times \tilde{B}$, die z. B. durch das unscharfe kartesische Produkt $\otimes$ festgelegt wird. Dieses Prinzip lässt sich auf n-dimensionale Relationen erweitern (DUBOIS & PRADE 1980, S. 68f). Die Resultate unscharfer Relationen können im Allgemeinen als Spezialfälle unscharfer Mengen angesehen werden (BANDEMER & GOTTWALD 1990, S. 53f), weshalb alle Voraussetzungen für die Arithmetik von Fuzzy-Mengen auch für die Arithmetik von n-dimensionalen Fuzzy-Mengen gelten. Da in der vorliegenden Arbeit trapezförmige Fuzzy-Intervalle der Dimension Zwei verwendet werden, werden nachfolgend einige wesentliche Berechnungsvorschriften erläutert. Als Kennzeichnung der Zweidimensionalität wird xy im Index angegeben (z. B. mit $\tilde{A}_{xy}$). Gleichzeitig weist xy auf den Koordinatensystemraum. Für die Arithmetik zweidimensionaler Fuzzy-Mengen werden folgenden Bedingungen gefordert (KAUFMANN & GUPTA 1991, S. 146ff): Eine zweidimensionale unscharfe Zahl $\tilde{A}_{xy} \subset \Re^2$ existiert dann, wenn die Zugehörigkeitsfunktionen mit (3.58) konvex (nach VIERTL 1996, S. 11) und jeder α-Schnitt mit (3.59) eine kompakte Fläche ist (nach VIERTL 1996, S. 33) sowie der Maximalwert Eins mit (3.60) existiert.

$$m_{\tilde{A}}(\lambda \cdot x + (1-\lambda) \cdot y) \geq \min\left(m_{\tilde{A}}(x), m_{\tilde{A}}(y)\right) \quad \left| x,y \in \Re \text{ und } \lambda \in [0,1] \right. \quad (3.58)$$

$$\tilde{A}_{xy,\alpha} := \left\{ x,y \in \Re : m_{\tilde{A}_{xy}}(x,y) \geq \alpha \right\} \quad \forall \alpha \in (0,1] \quad (3.59)$$

$$\exists (x,y) \in \Re^2 : m_{\tilde{A}_{xy}}(x,y) = 1. \quad (3.60)$$

Analog zur Formulierung von Fuzzy-Intervallen des LR-Typs im Eindimensionalen folgt in der Dimension Zwei:

$$\tilde{A}_{xy} := \left\{ \begin{bmatrix} m_{\tilde{A},x}, r_{\tilde{A},x}, c_{\tilde{A},l,x}, c_{\tilde{A},r,x} \\ m_{\tilde{A},y}, r_{\tilde{A},y}, c_{\tilde{A},l,y}, c_{\tilde{A},r,y} \end{bmatrix}_{LR}, m_{\tilde{A}_{xy}}(x,y) \;\middle|\; x,y \in \Re^2 \right\} \quad \text{mit:} \quad m_{\tilde{A}_{xy}}(x,y) \in [0,1]. \tag{3.61}$$

Im Falle linearer Referenzfunktionen ist das Abbild ein Pyramidenstumpf, wie ihn Abbildung 3-9a) zeigt. Gilt für beide Radien zudem $r_{\tilde{A},x} = r_{\tilde{A},y} = 0$, dann ist das Abbild pyramidenförmig, vgl. Abbildung 3-9b). Von beiden wird nachfolgend für die Beschreibung der MS-Positionen im xy- Raum häufig Gebrauch gemacht.

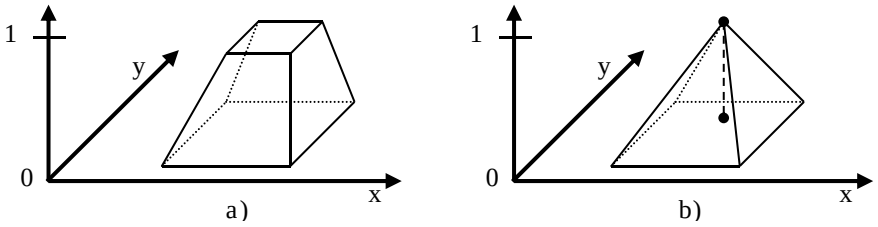


Abbildung 3-9: Zweidimensionale Fuzzy-Intervalle a) als Pyramidenstumpf und b) als Pyramide, nach KAUFMANN & GUPTA (1991)

Die oben im Eindimensionalen eingeführten erweiterten Operationen $\oplus$, $\ominus$, $\odot$ und $\oslash$ sind unmittelbar, unter Einhaltung von Restriktionen wie $0 \notin \tilde{B}_{xy}$, auf die mehrdimensionale Behandlung übertragbar.

3.3.2.5 Distanzen zwischen unscharfen Mengen

Allgemein kann eine Distanz zwischen zwei unscharfen Mengen $\tilde{A}$ und $\tilde{B}$ nach dem Erweiterungsprinzip gebildet werden mit (nach BANDEMER & GOTTWALD 1990, S. 26 Formel (2.60)):

$$m_{d(\tilde{A},\tilde{B})}(d) = \sup_{\substack{(x_1,x_2)\in\Re\times\Re \\ d(x_1,x_2)=d}} \min\left(m_{\tilde{A}}(x_1), m_{\tilde{B}}(x_2)\right) \tag{3.62}$$

sowie beschränkt auf den zweidimensionalen Raum $\Re^{(2)}$ mit dem euklidischen Abstand als kürzeste Distanz bestimmt werden mit:

$$m_{\tilde{d}(\tilde{A},\tilde{B})}(d) = \sup_{\substack{(x_1,y_1),(x_2,y_2)\in\Re^{(2)}\times\Re^{(2)} \\ d(x,y)=d=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}} \min\left(m_{\tilde{A}}(x_1,y_1), m_{\tilde{B}}(x_2,y_2)\right). \tag{3.63}$$

Einen geometrisch motivierten Lösungsweg zeigt GLEMSER (2000, S. 86ff), der wegen seines günstigeren Rechenaufwandes als (3.63) hier aufgegriffen werden soll. Alle möglichen Distanzen d zwischen beliebigen Objektpunkten $A \in \tilde{A}$ und $B \in \tilde{B}$ sowie deren Zugehörigkeitsgrade $m_{\tilde{d}}(d)$ bilden die unscharfe Menge $\tilde{d}$ der Distanz zwischen $\tilde{A}$ und $\tilde{B}$:

$$\tilde{d} := \left\{ \left(d_u \,\middle|\, m_{\tilde{d}}(d_u) \right) \right\} \quad \text{mit:} \quad m_{\tilde{d}}(d) = \min\left(m_{\tilde{A}}(A), m_{\tilde{B}}(B)\right), \; A \in \tilde{A}, \; B \in \tilde{B}. \tag{3.64}$$

Es wird gefordert, dass d jeweils die kürzeste Distanz zwischen den Objektpunkten in A und B ist. Mit $\tilde{d}$ soll zudem die kürzeste (unscharfe) Distanz zwischen den Objekten $\tilde{A}$ und $\tilde{B}$ beschrieben werden. Zur Behandlung der Objektunschärfen können die Distanzen α-schnittweise mit $\tilde{d}_\alpha$ zwischen $\tilde{A}_\alpha$ und $\tilde{B}_\alpha$ bestimmt werden. Von besonderem Interesse sind hierfür diejenigen Objektpunkte A_u und B_u auf beiden Objekten, die jeweils am dichtesten zu einander und diejenigen Objektpunkte A_o und B_o auf beiden Objekten, die jeweils am weitesten voneinander entfernt sind, vgl. dazu die Abbildung 3-10a). Überlappen sich $\tilde{A}_\alpha$ und $\tilde{B}_\alpha$ (und damit auch $\tilde{A}$ und $\tilde{B}$), ist die kürzeste Distanz zwischen ihnen Null. GLEMSER (2000) führt für diesen Fall negative Distanzen ein, die der kürzesten Distanz innerhalb des Überlappungsbereiches $\tilde{A}_\alpha \cap \tilde{B}_\alpha$ entsprechen, um das

jeweilige Objekt zu verlassen, vgl. Abbildung 3-10b). Negative Distanzen folgen sowohl für $\tilde{d}_{\alpha,\min}$ bei Existenz einer gemeinsamen Schnittmenge der beiden Objekte $\tilde{A}_\alpha \cap \tilde{B}_\alpha \neq \varnothing$, als auch für $\tilde{d}_{\alpha,\max}$, wenn eines der beiden Objekte vollständig im jeweils anderen liegt $\tilde{A}_\alpha \subset \tilde{B}_\alpha$ oder $\tilde{B}_\alpha \subset \tilde{A}_\alpha$.

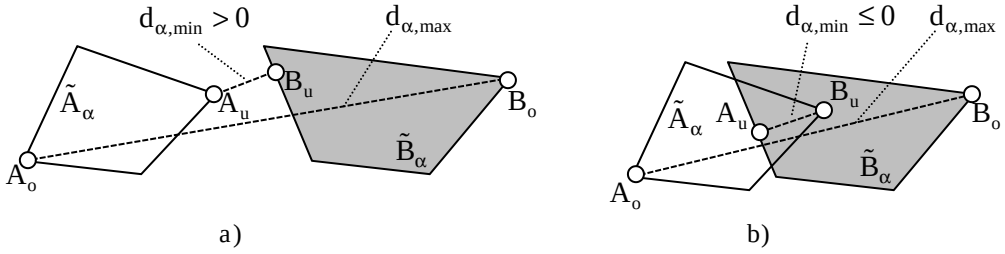


Abbildung 3-10: α-schnittweise Bestimmung der kürzesten Distanzen zwischen zwei a) disjunkten und b) sich überlappenden Objekten, nach GLEMSER (2000, S. 87)

Mit den kürzesten Distanzen $\tilde{d}_{\alpha,\min}$ zwischen den am dichtesten aneinander (innersten) liegenden Objektpunkten A_u und B_u sowie $\tilde{d}_{\alpha,\max}$ zwischen den entferntesten (äußersten) Objektpunkten A_o und B_o wird die unscharfe Distanz α-schnittweise konstruiert mit:

$$m_{\tilde{d}}(d) = \sup_{\alpha\in(0,1]} \alpha \cdot i_{\tilde{d}_\alpha}(d) \quad \text{mit:} \quad d \in \Re\,,\; i_{\tilde{d}_\alpha} = \left[\tilde{d}_{\alpha,\min}, \tilde{d}_{\alpha,\max}\right],$$

$$\tilde{d}_{\alpha,\min} = \begin{cases} +\sqrt{\left(x_{B_u} - x_{A_u}\right) + \left(y_{B_u} - y_{A_u}\right)} & ,\tilde{A}_\alpha \cap \tilde{B}_\alpha = \varnothing \\ -\sqrt{\left(x_{B_u} - x_{A_u}\right) + \left(y_{B_u} - y_{A_u}\right)} & ,\tilde{A}_\alpha \cap \tilde{B}_\alpha \neq \varnothing \end{cases} \tag{3.65}$$

$$\tilde{d}_{\alpha,\max} = \begin{cases} +\sqrt{\left(x_{B_o} - x_{A_o}\right) + \left(y_{B_o} - y_{A_o}\right)} & ,\tilde{A}_\alpha \not\subset \tilde{B}_\alpha \text{ und } \tilde{B}_\alpha \not\subset \tilde{A}_\alpha \\ -\sqrt{\left(x_{B_o} - x_{A_o}\right) + \left(y_{B_o} - y_{A_o}\right)} & ,\tilde{A}_\alpha \subset \tilde{B}_\alpha \text{ oder } \tilde{B}_\alpha \subset \tilde{A}_\alpha. \end{cases}$$

3.3.2.6 Maße für die Unschärfe

Um eine Fuzzy-Menge bezüglich ihrer Unschärfe bewerten oder Fuzzy-Mengen anhand ihrer Unschärfe miteinander vergleichen zu können, werden **Unschärfemaße** definiert. In der Literatur findet man als wesentliche Ansätze für Unschärfemaße die in der Abbildung 3-11 grafisch interpretierte Entropie und Energie von Fuzzy-Mengen. Entropiemaße eignen sich z. B. um Abweichungen und Vergleiche gegenüber klassischen „scharfen“ Mengen vorzunehmen. Energiemaße geben Abweichungen gegenüber der leeren Menge an. Grundlage beider Maße ist die Zugehörigkeitsfunktion der Fuzzy-Menge.

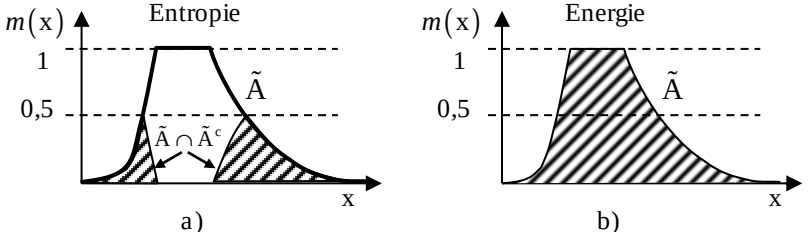


Abbildung 3-11: Entropie und Energie einer Fuzzy-Menge (Kutterer 2002, S. 52 Abb. 3.7)

Für das Unschärfemaß der **Entropie** wird eine nichtnegative Funktion h auf $\mathfrak{F}(X)$ definiert, für die die klassischen Mengen den Wert Null erhalten (BANDEMER & GOTTWALD 1990, S. 116f):

$$m_A(x) = m_{\tilde{A}_0}(x) \in \{0;1\} \quad \left|\forall x \in X \quad \Rightarrow h(A) = h\left(\tilde{A}_0\right) = 0.\right. \tag{3.66}$$

Als unschärfste Menge $\tilde{A}_{max}$ wird diejenige definiert, die sich selbst zum Komplement hat $\tilde{A}_{max}^{c} = \tilde{A}_{max}$. Dieser Fall stimmt für die Zugehörigkeitsfunktion:

$$m_{\tilde{A}}(x) \equiv 0{,}5 \quad \left| \forall x \in X \quad \Rightarrow h\left(\tilde{A}_{max}\right) = \max. \right. \tag{3.67}$$

Wesentlich für die Entropie unter anderem von Fuzzy-Intervallen ist, dass sie sich aus den Bereichen der Flanken der Zugehörigkeitsfunktion definiert, wohingegen der Kern keinen Beitrag liefert. Folglich haben klassische Intervalle, Punktmengen oder klassische Mengen verschwindende Entropiemaße. Die Berechnung der Entropie erfolgt beispielsweise mit (KUTTERER 2002, S. 52 Formel (3.48)):

$$h\left(\tilde{A}\right) = \int_{X} \min\left(m_{\tilde{A}}(x),\ 1 - m_{\tilde{A}}(x)\right) \cdot dx \ = \operatorname{card}\left(\tilde{A} \cap \tilde{A}^{c}\right). \tag{3.68}$$

Für den Vergleich zweier Fuzzy-Mengen ist $\tilde{B}$ eine **Verschärfung** von $\tilde{A}$, wenn gilt:

$$\begin{matrix} m_{\tilde{B}}(x) \le m_{\tilde{A}}(x) & \text{falls} & m_{\tilde{A}}(x) < 0{,}5 \\ m_{\tilde{B}}(x) \ge m_{\tilde{A}}(x) & \text{falls} & m_{\tilde{A}}(x) > 0{,}5 \end{matrix} \quad \Rightarrow \underbrace{h\left(\tilde{B}\right) \le h\left(\tilde{A}\right)}_{\text{„}\tilde{B}\text{ ist schärfer als }\tilde{A}\text{“}} \text{ oder auch: } \tilde{B} \prec_{h} \tilde{A}. \tag{3.69}$$

Für das Unschärfemaß der **Energie** wird eine nichtnegative Funktion e auf $\mathfrak{F}(X)$ definiert, für die die leere Menge $\varnothing$ den Wert Null erhält $e(\varnothing) = 0$. Der Maximalwert $e(X) = \max$ gilt für diejenige unschärfste Menge, die am stärksten von $\varnothing$ abweicht und somit im gesamten Grundbereich X vertreten ist. Außerdem soll für eine in der unscharfen Menge $\tilde{B}$ enthaltenen unscharfen Menge $\tilde{A}$ gelten: $\tilde{A} \subseteq \tilde{B} \Rightarrow e\left(\tilde{A}\right) \le e\left(\tilde{B}\right)$ (BANDEMER & GOTTWALD 1990, S. 120f; KUTTERER 2002, S. 53f). Die Höhe $\operatorname{height}\left(\tilde{A}\right)$ sowie die Mächtigkeit $\operatorname{card}\left(\tilde{A}\right)$ sind mögliche Energiemaße. Im Vergleich zur Höhe berücksichtigt die Mächtigkeit nicht ausschließlich die Stellen maximaler Zugehörigkeit sondern die Informationen des gesamten Trägers der Zugehörigkeitsfunktion, wodurch die mit der Fuzzy-Menge formulierten Unsicherheiten vollständig in der Mächtigkeit berücksichtigt werden. In der vorliegenden Arbeit wird daher für Vergleiche zwischen Fuzzy-Mengen auf Ebene der Energie die Mächtigkeit bevorzugt.

Für die Lösung der card-Funktion in der Praxis existieren in der Literatur verschiedene Vorschläge. KUTTERER (2002, S. 54f) zeigt beispielsweise eine Möglichkeiten zur Bestimmung der Energie von mehrdimensionalen unscharfen Vektoren sowie speziell unscharfen Vektoren vom parametrischen exponentiellen Typ, zweidimensionalen GAUß-schen Typ und elliptischen Typ (Bohnenform). Sind die Referenzfunktionen der Zugehörigkeitsfunktion stetig, eignet sich für gute Näherungen beispielsweise die SIMPSONSCHEN FORMEL. Auf ihr basierende Verfahren liefern in der Regel zudem rechentechnisch effiziente Lösungen (PAPULA 2009, S. 481ff).

In der vorliegenden Arbeit wird die Mächtigkeit auf dreieckige und trapezförmige Fuzzy-Intervalle sowie mit α-Schnittverfahren rekonstruierte Fuzzy-Mengen angewendet. Für die beiden Erstgenannten ergeben sich im Falle normaler (im Mittelpunkt ist der Zugehörigkeitswert Eins) Fuzzy-Mengen Vereinfachungen, deren Berechnungsformeln in Anhang C zusammengefasst sind. Für (re-)konstruierte normale und subnormale Fuzzy-Mengen aus α-Schnitten eignen sich Verfahren, die z. B. auf dem Trapezverfahren nach der Sehnentrapezformel (NEUMANN 2009) oder auch nach der Tangententrapezformel basieren. Beide Verfahren nähern den Verlauf einer Funktion $y = f(x)$ stückweise durch lineare Elemente an, um anschließend mittels Trapezen einfache Formulierungen zur Flächenberechnung zu erhalten. Die Summe aller trapezförmig an die Zugehörigkeitsfunktion angenäherten Flächen ergibt die folgende Näherung für $\operatorname{card}\left(\tilde{A}\right)$:

$$\operatorname{card}\left(\tilde{A}\right) \approx \sum_{i=1}^{n_\alpha - 1} \left(a_{i+1} - a_i\right) \cdot \frac{m_{\tilde{A}}\left(a_{i+1}\right) + m_{\tilde{A}}\left(a_i\right)}{2} \quad \text{mit: } n_\alpha \in 2 \cdot \mathbb{N}. \tag{3.70}$$

Mit Verwendung der Anzahl n äquidistanter Teilbereiche q auf der x-Achse (Abszisse) ergibt sich folgende Vereinfachung (nach NEUMANN 2009, S. 72ff), dessen Grundprinzip in Abbildung 3-12a) dargestellt ist:

$$\mathrm{card}\left(\tilde{A}\right) \approx q \cdot \left(\frac{m_{\tilde{A}}(a_0)}{2} + m_{\tilde{A}}(a_1) + \ldots + m_{\tilde{A}}(a_{n-1}) + \frac{m_{\tilde{A}}(a_n)}{2} \right) \quad \text{mit:} \quad q = \frac{\tilde{A}_{\alpha=0,\max} - \tilde{A}_{\alpha=0,\min}}{n}. \tag{3.71}$$

Nach Konstruktion normaler Fuzzy-Mengen mittels der α-Schnitte weisen die Wertepaare zudem folgende Eigenschaften auf: Die Anzahl der Wertepaare n_α ist geradzahlig mit $n_\alpha \in 2 \cdot \mathbb{N}$ und mit Ausnahme der Wertepaare $n_\alpha/2$ und $n_\alpha/2+1$ (deren Zugehörigkeitswertdifferenz Null beträgt) unterscheiden sich aufeinanderfolgende Zugehörigkeitswerte jeweils um $\Delta\alpha$. Die in Abbildung 3-12b) grafisch veranschaulichte Berechnung der card-Funktion vereinfacht sich damit zu:

$$\mathrm{card}\left(\tilde{A}\right) \approx \Delta\alpha \cdot \sum_{i=1}^{n_\alpha} \left(\frac{a_i - a_{i-1}}{2} \right) \tag{3.72}$$

Dass $\tilde{A}$ tatsächlich eine normale Fuzzy-Menge ist, ist mit Erfüllung der Bedingung $\mathrm{height}\left(\tilde{A}\right) = 1$ zu prüfen. Aufgrund der zwischen den Wertepaaren angenommenen linearen Verbindungselemente, resultieren sowohl für (3.70) und (3.71) als auch für (3.72) Approximationsfehler. An einer Beispielfunktion zeigen die Abbildung 3-12a) diese Approximationsfehler für (3.71) und die Abbildung 3-12b) für (3.72).

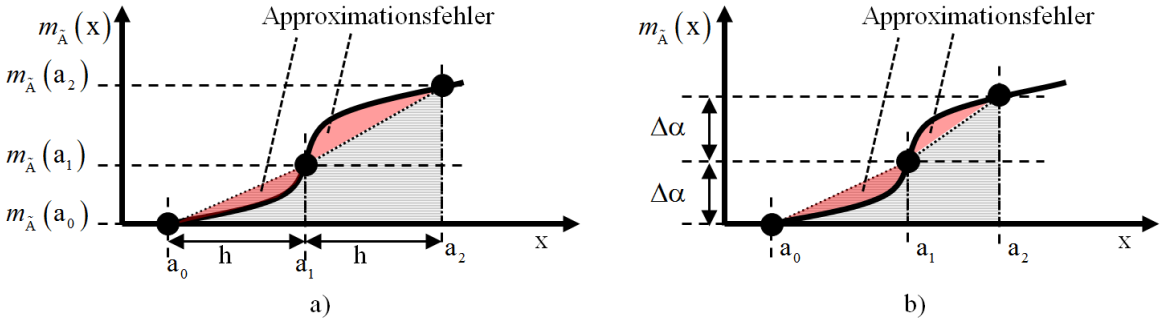


Abbildung 3-12: numerische Integration zur Lösung der card-Funktion a) mit äquidistanten Teilbereichen auf der Abszisse und b) mit äquidistanten Zugehörigkeitswerten

3.3.2.7 Hypothesentests mit unscharfen Daten

Zur Validierung auf Annahmen beruhender Aussagen mittels Daten oder allgemein zur Analyse von Daten sind Hypothesentests, wie sie bereits in Abschnitt 3.3.1 vorgestellt wurden, wesentliche Werkzeuge. Analog zu den statistischen Hypothesentests soll mit **Hypothesentests mit unscharfen Daten** eine aus unscharfen (Mess-) Daten abgeleitete Testgröße gegenüber Annahme- und Verwerfungsbereichen verglichen werden. Sowohl die Testgröße $T \Rightarrow \tilde{T}$ als auch Annahmebereich $A \Rightarrow \tilde{A}$ und Verwerfungsbereich $V \Rightarrow \tilde{V}$ können darin unscharf definiert sein. Die Abbildung 3-13 zeigt ein Beispiel.

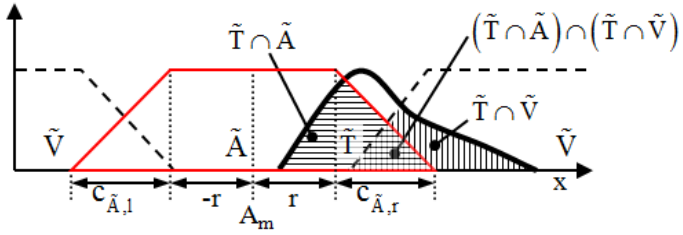


Abbildung 3-13: Hypothesentest mit unscharfen Daten nach KUTTERER (2002, S. 73)

Um unmittelbar Schlussfolgerungen ableiten zu können, soll ein „scharfes“ Maß als abschließendes Entscheidungskriterium geliefert werden. Während im Falle (klassischer) statistischer Hypothesentests binäre

Schlussfolgerungen mit $\rho \in \{0;1\}$ (z. B. mit $\rho = 0$ für H_0 annehmen oder $\rho = 1$ für H_0 verwerfen) resultieren, folgt im Hypothesentest mit unscharfen Daten ein kontinuierliches Maß $\rho_{\tilde{V}} \in [0,1]$ als Grad der Verwerfung der Nullhypothese H_0 gegenüber dem unscharfen Quantil $\tilde{y}$. Im Wesentlichen führt das zu einem Maß für die Übereinstimmung zweier Fuzzy-Mengen (KUTTERER 2002).

Ein **Vergleich von Fuzzy-Mengen** gelingt über den Grad der Übereinstimmung $\gamma \in [0,1]$ einer Fuzzy-Menge $\tilde{N}$ bezüglich einer Fuzzy-Menge $\tilde{M}$ (KAUFMANN & GUPTA 1991, S. 264)[4]:

$$\gamma_{\tilde{N}}\left(\tilde{M}\right) := \gamma\left(\tilde{M},\tilde{N}\right) = \frac{h\left(\tilde{M} \cap \tilde{N}\right)}{h\left(\tilde{M}\right)}. \tag{3.73}$$

Dazu wird die Fuzzy-Menge $\tilde{N} \in \mathfrak{F}(\Re)$ mit einer nicht leeren Fuzzy-Menge $\tilde{M} \in \mathfrak{F}(\Re), \tilde{M} \neq \varnothing$ verglichen. Die Funktion $h(\bullet)$ genügt dabei folgenden Bedingungen (KUTTERER 2004, S. 52f):

$$\begin{array}{lll} \text{minimale Übereinstimmung:} & \tilde{U} = \varnothing & \Leftrightarrow h\left(\tilde{U}\right) = 0 \\ \text{Monotonie:} & \tilde{U} \subseteq \tilde{V} & \Rightarrow h\left(\tilde{U}\right) \leq h\left(\tilde{V}\right) \end{array} \quad \left| \tilde{U},\tilde{V} \in \mathfrak{F}(\Re). \right. \tag{3.74}$$

Beispiele zur Realisierung von $h(\bullet)$ sind die Höhe $h\left(\tilde{U}\right) = \text{height}\left(\tilde{U}\right)$ aus Formel (3.27) oder auch die Mächtigkeit $h\left(\tilde{U}\right) = \text{card}\left(\tilde{U}\right)$ aus Formel (3.28). Ausführlich beschreibt NEUMANN (2009, S. 65ff) die Eigenschaften der beiden Methoden, um Testentscheidungen auf deren Basis zu treffen. In der vorliegenden Arbeit wird im Wesentlichen die Bildung mittels des card-Kriteriums verwendet. Das height-Kriterium wird weder Asymmetrien noch der Unschärfe selbst gerecht (KUTTERER 2002, S. 73). Ausnahme ist das Vorliegen von scharfen Daten, da dann $\text{card}\left(\tilde{U}\right) = 0$ gilt.

Neben dem Grad der Übereinstimmung kann unmittelbar der Grad der Nicht-Übereinstimmung $\delta \in [0,1]$ mit $\delta_{\tilde{N}}\left(\tilde{M}\right) := 1 - \gamma_{\tilde{N}}\left(\tilde{M}\right)$ formuliert werden. Mit den in (3.74) genannten Bedingungen lassen sich für beide Grade die nachfolgenden Eigenschaften ableiten:

		Übereinstimmung	Nicht-Übereinstimmung
$\tilde{M} \cap \tilde{N} = \varnothing$	$\Leftrightarrow$	$\gamma_{\tilde{N}}\left(\tilde{M} \cap \tilde{N}\right) = 0$	$\delta_{\tilde{N}}\left(\tilde{M} \cap \tilde{N}\right) = 1$
$\tilde{M} \cap \tilde{N} = \tilde{N}$	$\Leftrightarrow$	$\gamma_{\tilde{N}}\left(\tilde{M} \cap \tilde{N}\right) = 1$	$\delta_{\tilde{N}}\left(\tilde{M} \cap \tilde{N}\right) = 0$
$\tilde{N} = \varnothing$	$\Rightarrow$	$\gamma_{\tilde{N}}\left(\tilde{N}\right) = 1$	$\delta_{\tilde{N}}\left(\tilde{N}\right) = 0$
$\tilde{M} \subseteq \tilde{P} \wedge \tilde{N} \subseteq \tilde{Q}$	$\Rightarrow$	$\gamma_{\tilde{N}}\left(\tilde{M}\right) \leq \gamma_{\tilde{Q}}\left(\tilde{P}\right)$	$\gamma_{\tilde{N}}\left(\tilde{M}\right) \geq \gamma_{\tilde{Q}}\left(\tilde{P}\right)$

(3.75)

Darin entsprechen $\gamma_{\tilde{N}}(\bullet) = 0$ einer minimalen und $\gamma_{\tilde{N}}(\bullet) = 1$ einer maximalen Übereinstimmung sowie $\delta_{\tilde{N}}(\bullet) = 0$ einer minimalen und $\delta_{\tilde{N}}(\bullet) = 1$ einer maximalen Nicht-Übereinstimmung der Fuzzy-Mengen.

Mit dem Grad $\rho_{\tilde{V}}$ für die Verwerfung der Nullhypothese aus $\gamma_{\tilde{V}}\left(\tilde{T}\right)$ und $\delta_{\tilde{A}}\left(\tilde{T}\right)$ kann nun eine Testentscheidung abgeleitet werden mit:

$$\rho_{\tilde{V}}\left(\tilde{T}\right) := \overline{\rho}\left(\gamma_{\tilde{V}}\left(\tilde{T}\right),\ \delta_{\tilde{A}}\left(\tilde{T}\right) = 1 - \gamma_{\tilde{A}}\left(\tilde{T}\right)\right). \tag{3.76}$$

Die Funktion $\overline{\rho}(\bullet) \rightarrow [0,1]$ entspricht einer t-Norm oder einer s-Norm. In der vorliegenden Arbeit wird der $\min(\bullet)$-Operator verwendet, so dass folgt:

$$\rho_{\tilde{V}}\left(\tilde{T}\right) := \min\left(\gamma_{\tilde{V}}\left(\tilde{T}\right), \delta_{\tilde{A}}\left(\tilde{T}\right)\right) \tag{3.77}$$

[4] KAUFMANN & GUPTA (1991) bezeichnen den Grad der Übereinstimmung mit agreement index (dt.: Akzeptanzindex)

Ist der Annahmebereich $\tilde{A}$, wie in Abbildung 3-13 gezeigt, scharf definiert, gilt als Spezialfall $\rho_{\tilde{V}}\left(\tilde{T}\right)=\gamma_{\tilde{V}}\left(\tilde{T}\right)$. Sind das Quantil und/oder der Annahmebereich unscharf, ist eine eindeutige Testentscheidung nur für $\rho_{\tilde{V}}\left(\tilde{T}\right)=0\rightarrow$ "H_0 definitiv annehmen" und für $\rho_{\tilde{V}}\left(\tilde{T}\right)=1\rightarrow$ "H_0 definitiv verwerfen" möglich. Für alle $\rho_{\tilde{V}}\left(\tilde{T}\right)\in(0,1)$ gilt, dass das Testergebnis unscharf und damit streng formuliert „**uneindeutig**“, also unentscheidbar ist. Daraus lassen sich die folgenden drei Fälle für die Testentscheidung ableiten.

Die Testgröße liegt

Fall I:	ausschließlich im Annahmebereich	$\tilde{T}\cap\tilde{V}=\varnothing$	$\rightarrow H_0$ annehmen
Fall II:	ausschließlich im Verwerfungsbereich	$\tilde{T}\cap\tilde{A}=\varnothing$	$\rightarrow H_0$ verwerfen
Fall III:	sowohl im Annahme- als auch im Verwerfungsbereich	$\tilde{T}\cap\tilde{V}\neq\varnothing\wedge$ $\tilde{T}\cap\tilde{A}\neq\varnothing$	$\rightarrow$ uneindeutig.

(3.78)

Eine linguistische Testentscheidung kann durch die Kategorisierung des Wertebereichs von $\rho_{\tilde{V}}\left(\tilde{T}\right)$ mit einer oder mehrerer kritischer Grenzen $\rho_{\text{Krit}}\in[0,1]$ vorgenommen werden (KUTTERER 2002, S. 73). Abbildung 3-14 zeigt dafür eine grafische Umsetzung mit drei Kategorien.

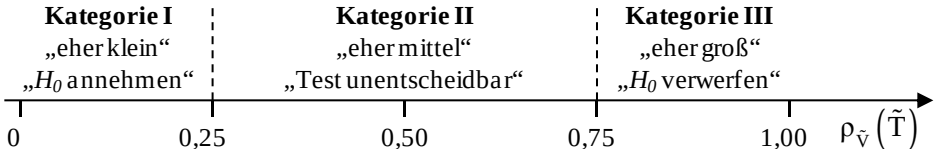


Abbildung 3-14: Kategorisierung des Verwerfungsindexes für Hypothesentests mit unscharfen Daten

Ein geringer Unterschied zwischen $\gamma_{\tilde{V}}\left(\tilde{T}\right)$ und $\delta_{\tilde{A}}\left(\tilde{T}\right)$ weist infolge der Unschärfen auf eine geringe Trennbarkeit der Hypothesen hin. Ist das der Fall, sollte keine oder zumindest keine scharfe Testentscheidung getroffen werden (KUTTERER 2002, S. 73).

Alternativ und für die Integration in einem automatisierten Prozess geeigneter, kann mittels eines einzelnen kritischen Wertes eine Testentscheidung erzwungen werden, wie sie Abbildung 3-15 zeigt:

$$\rho_{\tilde{V}}\left(\tilde{T}\right)\begin{Bmatrix}\leq\\>\end{Bmatrix}\rho_{\text{Krit}}\in[0,1]\Rightarrow\begin{cases}H_0 \text{ nicht verwerfen}\\ H_0 \text{ verwerfen.}\end{cases} \tag{3.79}$$

Ein Nachteil der binären Entscheidung ist die Aussparung desjenigen Bereiches, für den aufgrund der Unschärfe der „Test unentscheidbar“ ist. In Abbildung 3-14 entspricht das dem Fehlen der Kategorie II.

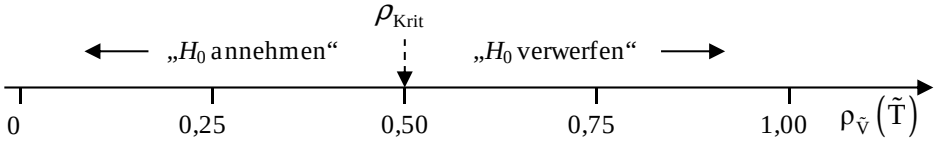


Abbildung 3-15: Binäre Testentscheidung mittels eines kritischen Wertes

Die numerische Wahl von ρ_{Krit} obliegt dem Anwender. Die Festlegung eines angemessenen Wertes ist mit Expertenwissen sowie unterstützend mit der sachlogischen Analyse von simulierten Daten und Re-Analyse realer Daten zu verknüpfen (NEUMANN 2009, S. 67). Ein geeigneter Näherungswert für die Analyse noch unbekannter Datenbestände ist $\rho_{\text{Krit}}=0,5$. Bei sicherheitsrelevanten Fragestellungen sollte ρ_{Krit} gegen Null gehen und für Tests im Sinne einer Ausreißersuche kann ρ_{Krit} nahe Eins gelten (NEUMANN & KUTTERER 2008, Abschnitt 2.2.4). Für sicherheitsrelevante Fragestellungen wird $\rho_{\text{Krit}}=0,1$ und für die Ausreißersuche $\rho_{\text{Krit}}=0,9$ empfohlen (NEUMANN 2009, S. 122, S. 126).

3.3.3 Fuzzy-Randomness

In den vorangegangenen Abschnitten 3.3.1 und 3.3.2 wurde für die zu bearbeitenden Daten angenommen, dass für sie entweder eine zufällige Variabilität oder Unschärfe vorliegt. Treffen jedoch beide Arten dieser Unsicherheiten zu, ist deren eine gemeinsame Modellierung und Behandlung erforderlich. Hierfür eignet sich das sogenannte **Fuzzy-Randomness** mit Hilfe von **Fuzzy-Random Variablen** (FRV).

3.3.3.1 Definition einer Fuzzy-Random-Variablen (FRV)

Nach NEUMANN (2009, S. 47) wird eine Fuzzy-Random-Variable wie folgt definiert: „*Sei* $(\Omega, \mathrm{B}, \mathrm{P})$ *ein Wahrscheinlichkeitsraum* […] (mit der Grundgesamtheit Ω, dem Ereignisraum B und dem Wahrscheinlichkeitsmaß P), *dann wird die unscharfe Abbildung* $\underline{\tilde{\boldsymbol{A}}}: \Omega \tilde{\rightarrow} \mathfrak{F}(\Re^n)$ *mit:*

$$\underline{\tilde{\boldsymbol{A}}}: \Omega \tilde{\rightarrow} \mathfrak{F}(\Re^n), \quad \underline{\tilde{\boldsymbol{A}}}(\boldsymbol{\omega}) = \tilde{\mathbf{A}}, \quad \boldsymbol{\omega} \in \Omega, \quad \tilde{\mathbf{A}} \in \mathfrak{F}(\Re^n) \tag{3.80}$$

als unscharfe Zufallsvariable oder (unscharfer Zufalls)-vektor bezeichnet, wenn für alle $\omega_i \in \Omega$ *und für alle* $\alpha \in (0,1]$ *die Menge* $\left(\underline{\tilde{\boldsymbol{A}}}(\boldsymbol{\omega})\right)_\alpha$ *beschränkt ist* […]„. Zudem sei $\underline{\tilde{\boldsymbol{A}}}(\boldsymbol{\omega})$ nach NEUMANN (2009, S. 47) mit:

$$\begin{aligned} \inf\left(\underline{\tilde{\boldsymbol{A}}}(\boldsymbol{\omega})\right)_\alpha &:= \inf\left\{\mathbf{x} \in \Re^n : m_{\underline{\tilde{\boldsymbol{A}}}(\boldsymbol{\omega})}(\mathbf{x}) \geq \alpha\right\} \\ \sup\left(\underline{\tilde{\boldsymbol{A}}}(\boldsymbol{\omega})\right)_\alpha &:= \sup\left\{\mathbf{x} \in \Re^n : m_{\underline{\tilde{\boldsymbol{A}}}(\boldsymbol{\omega})}(\mathbf{x}) \geq \alpha\right\} \end{aligned} \tag{3.81}$$

für jedes $\omega_i \in \Omega$ eine normale Fuzzy-Menge mit einer mindestens stückweise stetigen Zugehörigkeitsfunktion. Die Werte $\inf\left(\underline{\tilde{\boldsymbol{A}}}(\boldsymbol{\omega})\right)_\alpha$ und $\sup\left(\underline{\tilde{\boldsymbol{A}}}(\boldsymbol{\omega})\right)_\alpha$ sind endliche Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathrm{B}, \mathrm{P})$ (KUTTERER 2002, S. 64; KWAKERNAAK 1978, 1979). Die Definition der FRV basiert damit auf dem von ZADEH (1965) vorgeschlagenem Erweiterungsprinzip, wodurch z. B. bei Anwendung von Regeln und Methoden der reellen Stochastik auf FRV eine unscharfe Erweiterung und somit unscharfe Größen resultieren. Während in der klassischen Wahrscheinlichkeitstheorie jedem $\omega_i \in \Omega$ genau eine „scharfe“ reelle Zahl zugeordnet wird, wird im Falle von FRV durch die unscharfe Abbildung $\omega_i \tilde{\rightarrow} m_{\omega_i}$ eine Zugehörigkeitsfunktion zugeordnet (NEUMANN 2009, S. 49f).

Eine grafische Interpretation einer FRV als Fuzzy-Intervall des LR-Typs zeigt Abbildung 3-16. Der Mittelpunkt $\underline{x}_m$ der FRV wird als Träger der zufälligen Variabilität gedeutet und repräsentiert die originäre Beobachtung. $\underline{x}_m$ ist mit den wahrscheinlichkeitstheoretischen Methoden aus Abschnitt 3.3.1 zu behandeln. Die Unschärfe wird mittels Zugehörigkeitsfunktionen formuliert. Die Behandlung der Zugehörigkeitsfunktionen erfolgt mit Methoden der Fuzzy-Theorie aus Abschnitt 3.3.2. Eine als FRV formulierte Variable A wird mit $\underline{\tilde{\mathrm{A}}}$ und ihre Zugehörigkeitsfunktion mit $m_{\underline{\tilde{\mathrm{A}}}}(\mathrm{x})$ notiert. Der Mittelpunkt, der Träger der zufälligen Variabilität ist, wird mit $\underline{\mathrm{x}}_m$ gekennzeichnet. Vektoren werden entsprechend der Vektor- und Matrixschreibweise **fett** geschrieben.

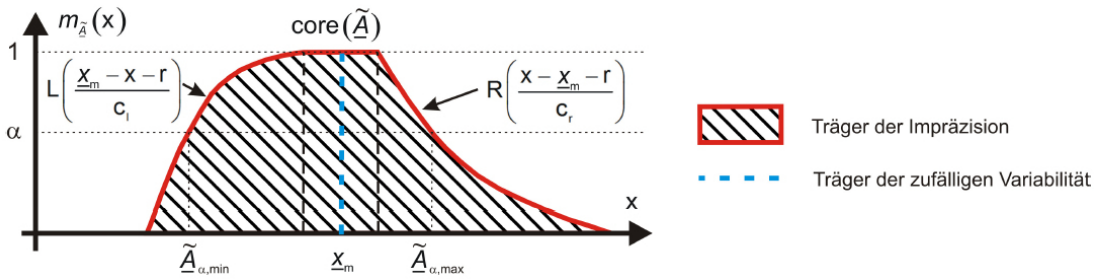


Abbildung 3-16: Grafische Interpretation einer FRV als Fuzzy-Intervall des LR-Typs
(NEUMANN 2009, S. 48 Abbildung 3.4)

3.3.3.2 Hypothesentests für FRV

Die Entwicklung von **Hypothesentests mit FRV** wurde im Fachbereich der Geodäsie wesentlich durch die grundlegenden Arbeiten von KUTTERER (1994) und KUTTERER (2002) sowie den daraus weiterentwickelten Methoden für Globaltests und Ausreißertests in NEUMANN (2009), NEUMANN & KUTTERER (2005 bis 2008) und NEUMANN ET AL. (2006) vorangetrieben. Die folgende Zusammenstellung wurde aus den genannten Quellen entnommen und in Teilen an die Anwendungen der vorliegenden Arbeit angepasst. Die Hypothesentests mit FRV sind als Erweiterung sowohl der Testverfahren mit scharfen, stochastischen Daten aus Abschnitt 3.3.1 als auch mit Fuzzy-Mengen aus Abschnitt 3.3.2 zu verstehen. Die Spezialfälle der scharfen stochastischen Daten und derjenigen ausschließlich Fuzzy-Mengen sind darin eingebettet.

Wesentlicher Unterschied der Hypothesentests mit FRV im Vergleich zu den klassischen statistischen Hypothesentests liegt darin, dass die Testgrößen nicht ausschließlich im Annahme- oder im Verwerfungsbereich, sondern auch teilweise in beiden vorhanden sein können. Sowohl die Hypothesen mit Annahme- und Verwerfungsbereich, als auch das Quantil der Verteilung und die Testgröße können als Fuzzy-Menge definiert werden. Mit Hilfe der Fuzzy-Theorie ist es möglich, diese Problemstellung mathematisch zu behandeln. Wenn im Folgenden von unscharfem Quantil der Verteilung $\underline{\tilde{y}}$ oder der unscharfer Testgröße $\underline{\tilde{T}}$ als dessen Realisierung, dem unscharfen Annahmebereich $\tilde{A}$ und dem unscharfen Verwerfungsbereich $\tilde{V}$ gesprochen wird, kennzeichnet das hochgestellte Tildesymbol „ ˜ “ die Unschärfe und der Unterstrich „ _ “ die Komponente der zufälligen Variabilität. Der Übergang auf scharfe Werte und Intervalle ist dabei jederzeit möglich.

Nachfolgend werden die grundlegenden Schritte für Hypothesentests mit unscharfen Daten zusammengefasst. Sie basieren auf KUTTERER (2002), KUTTERER (2004), NEUMANN & KUTTERER (2005, S. 147ff), NEUMANN (2009, S. 61ff), SCHÖN (2003, S. 86ff) und STEINMETZ (1998). Für die hier verwendeten Hypothesentests mit unscharfen Daten sei gemäß der Definition der Fuzzy-Zahlen und Fuzzy-Intervalle die Erfüllung der folgenden drei Kriterien vorausgesetzt (NEUMANN & KUTTERER 2005, S. 149; VIERTL 1996, S. 51):

1. die Zugehörigkeitsfunktion muss mindestens einen Funktionswert mit 1 als Maximalwert aufweisen

$$\exists\, x \in X : m_{\tilde{A}}(x) = 1 \tag{3.82}$$

2. die Form der Zugehörigkeitsfunktion muss konvex sein

$$\forall\, x_1 \in X \text{ und } \forall\, x_2 \in X \text{ und } \forall\, \lambda \in [0,1] \quad \text{gilt: } m_{\tilde{A}}(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \geq \min\left(m_{\tilde{A}}(x_1), m_{\tilde{A}}(x_2)\right) \tag{3.83}$$

3. und die Zugehörigkeitsfunktion muss gegen 0 streben

$$\lim_{x \to -\infty} m_{\tilde{A}}(x) = 0 \text{ und } \lim_{x \to \infty} m_{\tilde{A}}(x) = 0. \tag{3.84}$$

Hypothesentests werden zumeist nach einem Auswerteprozess angewendet, bei dem beispielsweise interessierende Parametern ermittelt oder geschätzt wurden. Die Bestimmung oder Schätzung der Parameter $\underline{\tilde{\theta}}_i \in \underline{\tilde{\boldsymbol{\theta}}}$ ist unter Berücksichtigung des Zadeh'schen Erweiterungsprinzips durchzuführen. Die theoretische Varianz $\sigma^2_{\theta_i}$ sei dazu als bekannt vorausgesetzt oder mittels Varianz-Kovarianzfortpflanzung bestimmt. Das unscharfe Quantil der Verteilung im eindimensionalen Fall, mit Zentrierung auf den unter der Nullhypothese gültigen Erwartungswert der scharfen Zufallsvariablen und Normierung mit deren theoretischer Standardabweichung, wird berechnet mit (KUTTERER 2002, S. 72 Formel 3.115):

$$\underline{\tilde{y}} = \frac{\underline{\tilde{\theta}}_i - \left(E\left(\underline{\theta}_i \,|\, H_0\right)\right)}{\sigma_{\theta_i}}. \tag{3.85}$$

Bei Verwendung von LR-Fuzzy-Intervallen $\tilde{\underline{T}} = \left[x_{\tilde{\underline{T}}_m}, r_{\tilde{\underline{T}}}, c_{\tilde{\underline{T}}_l}, c_{\tilde{\underline{T}}_r} \right]_{LR}$ mit $x \in \Re$ zur Konstruktion der unscharfen Testgröße folgt die Zugehörigkeitsfunktion:

$$m_{\tilde{\underline{T}}}(x) = \begin{cases} L_{\tilde{\underline{T}}}\left(\dfrac{\tilde{\underline{T}}_m - x - r_{\tilde{\underline{T}}}}{c_{\tilde{\underline{T}}_l}} \right), & x < \tilde{\underline{T}}_m - r_{\tilde{\underline{T}}}\ , c_{\tilde{\underline{T}}_l} > 0 \\ 1, & \tilde{\underline{T}}_m - r_{\tilde{\underline{T}}} \le x \le \tilde{\underline{T}}_m + r_{\tilde{\underline{T}}} \\ R_{\tilde{\underline{T}}}\left(\dfrac{x - \tilde{\underline{T}}_m - r_{\tilde{\underline{T}}}}{c_{\tilde{\underline{T}}_r}} \right), & x > \tilde{\underline{T}}_m + r_{\tilde{\underline{T}}}\ , c_{\tilde{\underline{T}}_r} > 0 \\ 0, & \text{sonst.} \end{cases} \tag{3.86}$$

Numerisch ist $\tilde{\underline{T}}_m$ für den Fall gemeinsamer Behandlung von zufälliger Variabilität und Unschärfe identisch mit der Testgröße $\underline{T}$ für den scharfen stochastischen Fall. Mit der Festlegung der Nullhypothese und Alternativhypothese, die letztlich aus sachlogischen Überlegungen abzuleiten sind, werden der Annahmebereich und als mengentheoretisches Komplement dazu der Verwerfungsbereich $\tilde{V} = \tilde{A}^c$ konstruiert. Beim Übergang auf Hypothesentests mit unscharfen Daten stellen die Zugehörigkeitsfunktion für den Annahmebereich $m_{\tilde{A}} : \Theta_0 \tilde{\rightarrow} [0,1]$ und die Zugehörigkeitsfunktion für den Verwerfungsbereich $m_{\tilde{V}} : \Theta_0 \tilde{\rightarrow} [0,1]$ unscharfe Abbildungen dar. Annahme- und Verwerfungsbereich sind daher unscharfe Größen:

$$\begin{aligned} &\tilde{V} = \tilde{A}^c = 1 - \tilde{A} \quad \Leftrightarrow \quad m_{\tilde{V}}(x) = 1 - m_{\tilde{A}}(x) \quad \text{mit:} \quad m_{\tilde{A}} : \Re^u \rightarrow [0,1] \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \forall\, x \in \Re \\ &\tilde{V} := \left\{ x, m_{\tilde{V}}(x) = m_{\tilde{A}^c}(x) = 1 - m_{\tilde{A}}(x) \right\} \qquad\qquad \tilde{A} := \left\{ x, m_{\tilde{A}}(x) \;\middle|\; x \in \Re \right\}. \end{aligned} \tag{3.87}$$

Dies stellt eine Verallgemeinerung zu Formel (3.17) aus Abschnitt 3.3.1 dar, weshalb der klassische Annahmebereich mit eindeutiger Testentscheidung als Spezialfall in (3.87) enthalten ist. Die Unschärfe beträgt dann Null.

Mit $\tilde{A}_m$ als den Mittelpunkt, $r_{\tilde{A}}$ dem Radius sowie $c_{\tilde{A},l}$ der linken und $c_{\tilde{A},r}$ der rechten Spannweite kann der Annahmebereich auch als LR-Fuzzy-Intervall konstruiert werden. Die Wahl der Referenzfunktionen unterliegt grundsätzlich keinen Einschränkungen, solange sie den oben genannten Kriterien genügen und sachlogisch genügend motivierbar sind. Zur Festlegung des Annahmebereichs wird mit Vorgabe der Irrtumswahrscheinlichkeit α_{Irrtum} das Sicherheitsniveau definiert, für das bei scharfem Quantil der Verteilung $\underline{T}$ und Standardnormalverteilung die Nullhypothese H_0 gilt. Der Annahmebereich ist als klassisches Intervall $\left[-r_{\tilde{A}}, r_{\tilde{A}} \right]$ mit dem Radius $r_{\tilde{A}}$ als das Quantil bei der Sicherheitswahrscheinlichkeit P definiert. Wird er beispielsweise mittels zweier verschiedener Sicherheitswahrscheinlichkeiten P_1 und P_2 fuzzyfiziert, bestimmen die aus den Quantilen gewonnenen Stützpunkte die Zugehörigkeitsfunktion. In Abbildung 3-17 ist hieraus ein trapezförmiger (unscharfer) Annahmebereich konstruiert worden und weist, bis auf die zufällige Komponente der Testgröße, hohe Ähnlichkeit zur Konstruktion in Abbildung 3-13 auf. Bei Wahl von z. B. $P_1 = 95\%$ und $P_2 = 99{,}5\%$ wird ein unscharfer Entscheidungsbereich zwischen den beiden Sicherheitswahrscheinlichkeiten aufgespannt, in dem sich H_0 und H_A überlappen. NEUMANN (2009, S. 62) motiviert eine derartige **Fuzzyfizierung** durch „…eine unzureichende Kenntnis des konkret zu testenden Wertes…“ beispielsweise der Frage ob die Fahrzeuggeschwindigkeit „etwa“ 50 km/h beträgt. Oder ob „…eine unzureichende Kenntnis über die zu verwendende Irrtumswahrscheinlichkeit für die Konstruktion des Annahmebereiches…“ z. B. der Frage ob „…die Beobachtung i mit einer Irrtumswahrscheinlichkeit von „etwa“ 5% ein Ausreißer…“ ist.

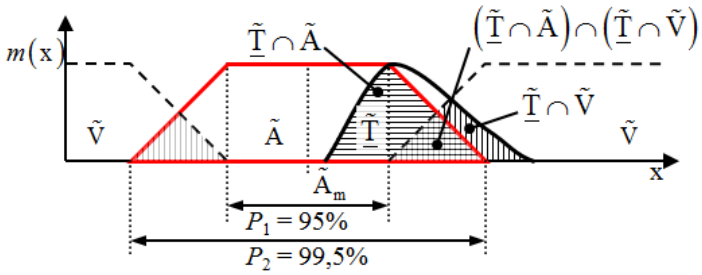


Abbildung 3-17: Mit zwei Sicherheitswahrscheinlichkeiten formulierte Annahme- und Verwerfungsbereiche nach KUTTERER (2002, S. 73)

Aus der Lage der Testgröße bezüglich zu Annahme- und Verwerfungsbereich lassen sich die folgenden drei Fälle für die Testentscheidung unterscheiden:

Die Testgröße liegt

Fall I :	ausschließlich im Annahmebereich	$\tilde{T} \cap \tilde{V} = \varnothing$	$\rightarrow H_0$ annehmen	(3.88)
Fall II :	ausschließlich im Verwerfungsbereich	$\tilde{T} \cap \tilde{A} = \varnothing$	$\rightarrow H_0$ verwerfen	
Fall III :	im Annahme- und im Verwerfungsbereich	$\tilde{T} \cap \tilde{V} \neq \varnothing \wedge \tilde{T} \cap \tilde{A} \neq \varnothing$	$\rightarrow$ uneindeutig.	

Eindeutige Testentscheidungen folgen ausschließlich dann, wenn die Testgröße entweder vollständig im Annahmebereich (Fall I) oder im Verwerfungsbereich (Fall II) liegt. Existieren Schnittmengen in beiden Bereichen (Fall III), kann aufgrund der Unschärfe keine eindeutige Aussage mehr getroffen werden. Eine mögliche Lösung mittels der Bestimmung des Grades der Übereinstimmung und des Grades der Nicht-Übereinstimmung wurde bereits im Abschnitt 3.3.2.7 auf Ebene der Fuzzy-Mengen angesprochen. Mit den dort genannten Formeln, gelten die entsprechenden Eigenschaften und Spezialfälle auch für den vorliegenden Fall. Mit dem Grad der Übereinstimmung des Verwerfungsbereiches mit der Testgröße $\gamma_{\tilde{V}}(\tilde{T})$ und dem Grad der Nicht-Übereinstimmung des Annahmebereiches mit der Testgröße $\delta_{\tilde{A}}(\tilde{T})$ bei Verwendung des height-Kriteriums oder des card-Kriteriums, folgt der Grad der Ablehnung der Nullhypothese bezüglich der Testgröße:

$$\gamma_{\tilde{V}}(\tilde{T}) = \text{height}(\tilde{T} \cap \tilde{V}) \quad \text{und} \quad \delta_{\tilde{A}}(\tilde{T}) = 1 - \text{height}(\tilde{T} \cap \tilde{A}) \quad \text{oder} \tag{3.89}$$

$$\gamma_{\tilde{V}}(\tilde{T}) = \frac{\text{card}(\tilde{T} \cap \tilde{V})}{\text{card}(\tilde{T})} \quad \text{und} \quad \delta_{\tilde{A}}(\tilde{T}) = 1 - \frac{\text{card}(\tilde{T} \cap \tilde{A})}{\text{card}(\tilde{T})}, \tag{3.90}$$

$$\rho_{\tilde{V}}(\tilde{T}) = \min\left(\gamma_{\tilde{V}}(\tilde{T}), \delta_{\tilde{A}}(\tilde{T})\right). \tag{3.91}$$

Eine binäre Testentscheidung kann durch Wahl eines geeigneten kritischen Wertes $\rho_{\text{krit}} \in [0,1]$, analog zu Abschnitt 3.3.2.7 Abbildung 3-15, erfolgen. Der Ablauf von Hypothesentests mit FRV ist in Tabelle 3-4 zusammengefasst dargestellt.

Abschließend können die Wahrscheinlichkeiten für den Fehlschluss 1. Art sowie den Fehlschluss 2. Art berechnet werden (KUTTERER 2002, S. 78f; NEUMANN 2009, S. 77f; NEUMANN & KUTTERER 2008). Beide Berechnungen werden im Folgenden nicht angewendet und sind daher in Tabelle 3-4 ausgeklammert gekennzeichnet.

Tabelle 3-4: Ablauf von Hypothesentests mit FRV

Schritt	Beschreibung
1.	Schätzung der unscharfen Parameter $\tilde{\underline{\boldsymbol{\theta}}}$
2.	Festlegung der Nullhypothese H_0 und gegebenenfalls der Alternativhypothese H_A
3.	Definition und Konstruktion des Annahmebereiches $\tilde{A}$ und des Verwerfungsbereiches $\tilde{V}$ in Form von Fuzzy-Mengen unter Berücksichtigung der Entscheidungskriterien, wie z. B. der Irrtumswahrscheinlichkeit α_{Irrtum} und dem unscharfen Quantil der Verteilung $\tilde{\underline{y}}$.
4.	Berechnung der unscharfen Testgröße $\tilde{\underline{T}}$ mit Hilfe des Erweiterungsprinzips.
5.	Berechnung des Grades der Verwerfung $\rho_{\tilde{V}}$ der Nullhypothese als Maß der Übereinstimmung der Testgröße mit dem Annahme- und Verwerfungsbereich.
6.	Definition des Entscheidungskriteriums ρ_{Krit} für die Annahme oder das Verwerfen der Nullhypothese oder gegebenenfalls einer uneindeutigen Testentscheidung.
(7.	Berechnung der Wahrscheinlichkeiten der Fehlschlüsse 1. und 2. Art im unscharfen Fall.)

3.3.3.3 Eine Alternative zur Testentscheidung mit unscharfen Wahrscheinlichkeiten

Alternativ zum vorherigen Vorgehen kann eine Testentscheidung für Hypothesentests mit FRV auch auf Basis unscharfer Wahrscheinlichkeiten stattfinden (FILZMOSER & VIERTL 2004; FILZMOSER & VIERTL 2006, S. 83f). Ausgangspunkt sind die klassischen Hypothesentests (mit präzisen Daten). Ebenfalls wird die klassische Unterteilung des Raumes der möglichen Werte in Annahmebereich A für H_0 sowie in Verwerfungsbereich V für H_A vorgenommen, so dass für H_0 mit der Irrtumswahrscheinlichkeit α_{Irrtum} der Verwerfungsbereich aufgespannt wird[5]:

$$\begin{aligned}
&\text{a) } P\left(\underline{T} \le \underline{y}_u \,|H_0\right) = \alpha_{\text{Irrtum}} && \left(\text{einseitig,V} = \left[-\infty, \underline{y}_u\right)\right)\\
&\text{b) } P\left(\underline{T} \ge \underline{y}_o \,|H_0\right) = \alpha_{\text{Irrtum}} && \left(\text{einseitig, V} = \left(\underline{y}_o, +\infty\right]\right)\\
&\text{c) } P\left(\underline{T} \le \underline{y}_a \,|H_0\right) = P\left(\underline{T} \ge \underline{y}_b \,|H_0\right) = \frac{\alpha_{\text{Irrtum}}}{2} && \left(\text{zweiseitig, V} = \left[-\infty, \underline{y}_a\right) \cup \left(\underline{y}_b, +\infty\right]\right)
\end{aligned} \tag{3.92}$$

Darin sind $\underline{y}_u$, $\underline{y}_o$ und $\left[\underline{y}_a, \underline{y}_b\right]$ die entsprechenden Quantile der zugrunde liegenden Verteilungsfunktion. Der Vergleich des Quantils der Verteilung mit der Testgröße erfolgt auf Ebene der Wahrscheinlichkeiten für das Eintreten der Bedingungen von Formel (3.92):

$$\begin{aligned}
&\text{a) } p = P\left(\underline{T} \le \underline{y}\right)\\
&\text{b) } p = P\left(\underline{T} \ge \underline{y}\right)\\
&\text{c) } p = 2 \cdot \min\left[P\left(\underline{T} \le \underline{y}\right), P\left(\underline{T} \ge \underline{y}\right)\right].
\end{aligned} \tag{3.93}$$

Für $p < \alpha_{\text{Irrtum}}$ ist die Nullhypothese zu verwerfen und entsprechend für $p \ge \alpha_{\text{Irrtum}}$ anzunehmen.

Infolge der Unschärfen der Beobachtungen ist deren Wahrscheinlichkeitsdichtefunktion ebenfalls als unscharfe Funktion $\underline{y} \to \tilde{\underline{y}}$ zu definieren, wodurch der Übergang von der scharfen Wahrscheinlichkeiten p auf die unscharfe Wahrscheinlichkeit $\tilde{p}$ mit $m_{\tilde{p}}(p) \in [0,1]$ und $p \in [0,1]$ motiviert wird. Die Auswertung kann nun α-schnittweise mit $\tilde{\underline{y}}_\alpha := m_{\tilde{\underline{y}},\alpha}(x) = \left[\tilde{\underline{y}}_{\alpha,\min}, \tilde{\underline{y}}_{\alpha,\max}\right]$, $\forall \alpha \in [0,1]$ als endlich abgeschlossene Intervalle erfolgen:

[5] die in FILZMOSER & VIERTL (2006) verwendeten Symbole wurden denen der vorliegenden Arbeit angepasst

$$\begin{aligned}
&\text{a)}\ \tilde{p}_\alpha := m_{\tilde{p},\alpha}(p) = \left[P\left(\underline{\mathrm{T}} \le \tilde{y}_{\underline{\alpha},\min}\right), P\left(\underline{\mathrm{T}} \le \tilde{y}_{\underline{\alpha},\max}\right)\right] \\
&\text{b)}\ \tilde{p}_\alpha := m_{\tilde{p},\alpha}(p) = \left[P\left(\underline{\mathrm{T}} \ge \tilde{y}_{\underline{\alpha},\min}\right), P\left(\underline{\mathrm{T}} \ge \tilde{y}_{\underline{\alpha},\max}\right)\right] \\
&\text{c)}\ \tilde{p}_\alpha := m_{\tilde{p},\alpha}(p) = \begin{cases} \left[2 \cdot P\left(\underline{\mathrm{T}} \le \tilde{y}_{\underline{\alpha},\min}\right),\ \min\left[1, 2 \cdot P\left(\underline{\mathrm{T}} \le \tilde{y}_{\underline{\alpha},\max}\right)\right]\right] & \text{für}: \mathrm{A_l} > \mathrm{A_r} \\ \left[2 \cdot P\left(\underline{\mathrm{T}} \ge \tilde{y}_{\underline{\alpha},\max}\right),\ \min\left[1, 2 \cdot P\left(\underline{\mathrm{T}} \ge \tilde{y}_{\underline{\alpha},\min}\right)\right]\right] & \text{für}: \mathrm{A_l} \le \mathrm{A_r} \end{cases}.
\end{aligned} \tag{3.94}$$

Die Fallunterscheidung bei c) untersucht, auf welcher Seite des Medians der Wahrscheinlichkeitsdichtefunktion sich der größere Unschärfeanteil von $\tilde{y}_{\underline{\alpha}}$ befindet. Dazu wird die untere Fläche $\mathrm{A_l}$ von $\tilde{y}_{\underline{\alpha},\min}$ bis zum Median und die obere Fläche $\mathrm{A_r}$ vom Median bis $\tilde{y}_{\underline{\alpha},\max}$ der Wahrscheinlichkeitsdichtefunktion berechnet. Zur Verdeutlichung dient das in FILZMOSER & VIERTL (2004 S. 28) gezeigte Beispiel, in dem anhand einer angenommen *F*-Verteilung mit $\mathrm{f_1} = 10$ und $\mathrm{f_2} = 12$ Freiheitsgraden versuchsweise die Signifikanz frei gewählter unscharfer Testgrößen $\underline{\mathrm{T}}_1$ bis $\underline{\mathrm{T}}_4$ für deren Zugehörigkeitsfunktionen $m_1(\mathrm{x})$ bis $m_4(\mathrm{x})$ im zweiseitigen Fall geprüft wird. Die $F_{10,12}$-Verteilung, deren Median sowie die Zugehörigkeitsfunktionen sind in Abbildung 3-18a) dargestellt. Zu erkennen ist, dass für $m_1(\mathrm{x})$ und $m_2(\mathrm{x})$ die Fläche $\mathrm{A_l}$ auf der linken Seite des Medians Null ist und für $m_4(\mathrm{x})$ die Fläche $\mathrm{A_r}$ rechts des Median Null beträgt. Die zu verwendende Fallunterscheidung in (3.94)-c) ist somit eindeutig. Im Falle von $m_3(\mathrm{x})$ sind hingegen Flächenbeiträge sowohl links als auch rechts des Medians vorhanden. Da augenscheinlich $\mathrm{A_r} > \mathrm{A_l}$ gilt, gilt entsprechend der untere Fall von (3.94)-c) für die Berechnung der unscharfen Wahrscheinlichkeit. Die α-schnittweise berechneten Zugehörigkeitsfunktionen der unscharfen Wahrscheinlichkeiten $m_{\tilde{p}_1}(p)$ bis $m_{\tilde{p}_4}(p)$ zeigt Abbildung 3-18b). Für den gewählten Schwellwert von $\alpha_{\mathrm{Irrtum}} = 0{,}10$ ist die Nullhypothese H_0 der mittels $m_1(\mathrm{x})$ und $m_4(\mathrm{x})$ charakterisierten Testgrößen $\underline{\mathrm{T}}_1$ und $\underline{\mathrm{T}}_4$ zu verwerfen. Für $m_2(\mathrm{x})$ und $m_3(\mathrm{x})$ und damit für $\underline{\mathrm{T}}_2$ und $\underline{\mathrm{T}}_3$ ist H_0 hingegen anzunehmen.

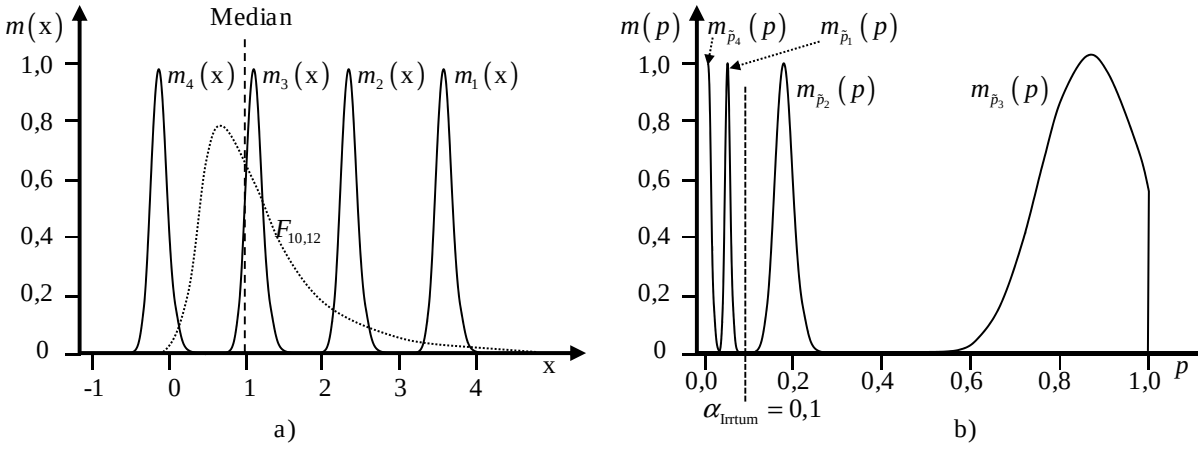


Abbildung 3-18: Beispiel a) einer Wahrscheinlichkeitsdichtefunktion und den Zugehörigkeitsfunktionen ausgewählter Testgrößen sowie b) die daraus folgenden unscharfen Wahrscheinlichkeiten
nach FILZMOSER & VIERTL (2004 S. 28, Fig. 4)

Die Zugehörigkeitsfunktion $m_{\tilde{p}}(p)$ wird mittels Zusammenführung der α-Schnitte über das Supremum $\sup(\bullet)$ konstruiert mit:

$$\tilde{p} := m_{\tilde{p}}(p) = \sup_{\alpha \in (0,1]} \alpha \cdot m_{\tilde{p},\alpha}(p). \tag{3.95}$$

Vergleicht man den Wertebereich des Trägers von $\tilde{p}$ mit $\mathrm{supp}(\tilde{p}) = \tilde{p}_{\alpha=0} = \left[\tilde{p}_{\alpha=0,\min}, \tilde{p}_{\alpha=0,\max}\right]$ gegenüber dem kritischen Signifikanzwert α_{Irrtum}, lassen sich bezüglich der Nullhypothese folgende Fälle unterscheiden:

$$\begin{aligned}
&\text{Fall I:} && \tilde{p}_{\alpha=0,\min} > \alpha_{\text{Irrtum}} && \Rightarrow H_0 \text{ definitiv annehmen} \\
&\text{Fall II:} && \tilde{p}_{\alpha=0,\max} < \alpha_{\text{Irrtum}} && \Rightarrow H_0 \text{ definitiv verwerfen} \\
&\text{Fall III:} && \alpha_{\text{Irrtum}} \in \tilde{p}_{\alpha=0} && \Rightarrow \text{Testentscheidung nicht eindeutig.}
\end{aligned} \tag{3.96}$$

Man erkennt sofort die Analogien der Fälle I und II zu denen des vorherigen Abschnitts 3.3.3.2. Ebenso folgt aus Fall III, dass keine eindeutige Entscheidung vorliegt. Mit Defuzzyfizierung von $\tilde{p}$ kann schließlich die binäre Testentscheidung (Annehmen oder Verwerfen von H_0) erzwungen werden. Als Verfahren der Defuzzyfizierung kann der in Abschnitt 3.3.2.7 vorgestellte Mengenvergleich dienen. Der Annahmebereich ist für diesen Fall zu definieren mit $\tilde{\mathrm{A}} = [\alpha_{\text{Irrtum}}, 1]$ und der Verwerfungsbereich mit $\tilde{\mathrm{V}} = [0, \alpha_{\text{Irrtum}})$. Der Wert der kritischen Grenze ρ_{Krit} ist, wie in den vorherigen Abschnitten auch, anwendungsbezogen geeignet zu wählen.

Es bleibt anzumerken, dass bei gleichen Verteilungsfunktionen und gleicher Wertewahl von α_{Irrtum} und ρ_{Krit} die Testverfahren des Hypothesentests mit FRV und mittels unscharfer Wahrscheinlichkeiten zu gleichen Ergebnissen bei der Testentscheidung führen.

4 Modellierung unsicherer Verkehrsdaten

4.1 Einleitung

Nach der allgemeinen Formulierung von Unsicherheiten in den vorangegangen Abschnitten werden nun die Identifizierungsverfahren für den Nachweis oder auch den Ausschluss von möglichen Verkehrsmitteln erarbeitet. Das Hauptaugenmerk liegt auf dem Vergleich der MS-Positionen mit denjenigen Merkmalen, die den Verkehrsmitteln des MIV entsprechen oder auch widersprechen. Die dazu infrage kommenden Verkehrsmittel und Verkehrsmittelgruppen sind die in Abschnitt 2.1 definierten oVB, NMIV, MIV und ÖV. Aufgrund des im Vergleich zu den anderen Verkehrsmitteln seltenen Einsatzes von Mobilfunkgeräten auf Fahrrädern des NMIV (die in der vorliegenden Arbeit verwendeten Ortungsverfahren erfordern, dass sich die MS im Gesprächsmodus befindet), werden sie im Rahmen der hier gezeigten Identifizierungsverfahren ausgeklammert[6].

Im nachfolgenden Abschnitt 4.2 **Modellierung unsicherer Ortungsinformationen** werden einleitend die Grundlagen und Methoden zur Ortung der MS erläutert, auf denen die Identifizierungsverfahren aufbauen. Wesentlich sind darin die Erarbeitung und insbesondere die Fortpflanzung der Unsicherheiten. Im Anschluss wird in Abschnitt 4.3 **Bewegungsidentifikation** untersucht, ob sich die MS in Bewegung befindet oder ob sie steht. Befindet sich die MS in Bewegung, wird in Abschnitt 4.4 die Geschwindigkeit der MS aus dessen Positionsfolge berechnet und mittels der **Geschwindigkeitsidentifikation** mit den Maximalgeschwindigkeiten von möglichen Verkehrsmitteln verglichen. Kommen nach dem Vergleich neben Verkehrsmitteln des MIV auch noch ÖV-Verkehrsmittel infrage, wird eine mögliche **Korrelation zu ÖV-Fahrzeugen** in Abschnitt 4.5 untersucht. Ist die MS keinem ÖV-Verkehrsmittel zuzuordnen, ist in einer abschließenden **kartengestützten Plausibilitätsprüfung** in Abschnitt 4.6 zu zeigen, dass entlang der MS-Positionsfolge sowohl räumlich als auch zeitlich auf dem digitalen Straßennetz eine möglich gewählte Route existiert. Die zusammengefasste Abfolge der Identifizierungsverfahren sowie die jeweiligen Abschnittsnummern zeigt Abbildung 4-1.

Zur Reduzierung fehlerhafter Verkehrslageerfassungen sollen Mobilfunkdaten nur dann an die Verkehrslageerfassung des MIV weitergeleitet werden, wenn der eindeutige Nachweis erbracht wurde, dass diese Mobilfunkdaten in einem Verkehrsmittel des MIV generiert wurden. Das ist dann der Fall, wenn die Verkehrsmittel aller anderen Verkehrsmittelgruppen ausgeschlossen werden können. Eine Abweichung von diesem Prinzip zum Zwecke der Erkennung von MS in Verkehrsstaus oder auch temporären Mehrdeutigkeiten bei der Verkehrsmittelnutzung, wird in Abschnitt 4.7 mit der **Verknüpfung der Identifizierungsergebnisse** besprochen.

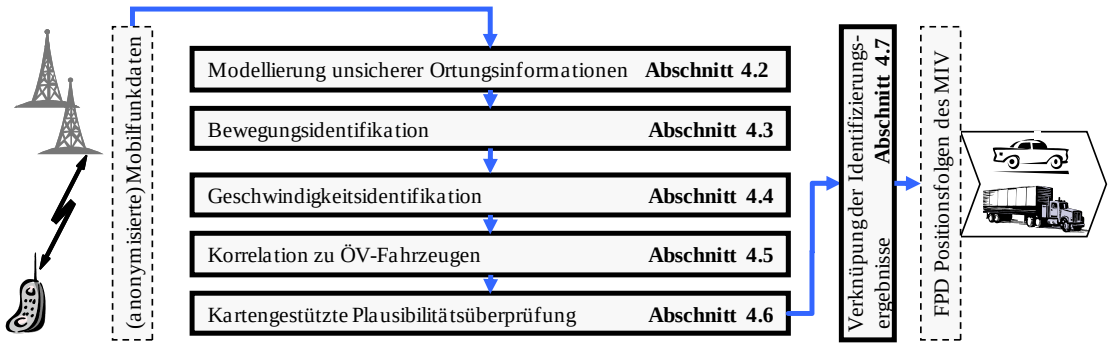


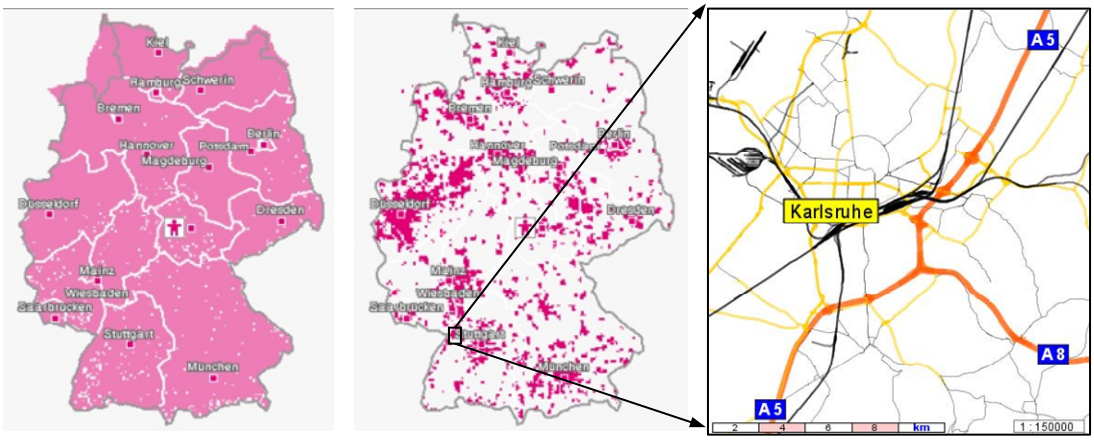
Abbildung 4-1: Ablauf der Identifikationsverfahren

[6] Mit steigender Verbreitung der Mobilfunkgeräte in der Bevölkerung sowie dem zukünftig steigenden Datentransfer zwischen MS und der Infrastruktur aufgrund von Web-, Mail- und Multimedia-Funktionen (SERVICEREPORT 2013), so dass die Ortung gegebenenfalls auch ohne aktive Gesprächsverbindung und bereits durch den bloßen Transport per Fahrrad möglich ist, muss diese Annahme neu bewertet werden.

4.2 Modellierung unsicherer Ortungsinformationen

4.2.1 Rohdatengewinnung

Datengrundlage der in der vorliegenden Arbeit verwendeten Mobilfunkortung sind infrastrukturseitig erfasste Mobilfunkdaten. Für eine flächendeckende Verkehrserfassung ist somit eine flächendeckende Erfassung von Mobilfunkdaten notwendig. In Deutschland kommen dafür vorrangig die Mobilfunknetze des GSM und des UMTS infrage. Während GSM mittlerweile eine fast vollständige Abdeckung bietet, befindet sich das UMTS-Netz derzeit noch im Ausbaustadium und wird aus Kostengründen auch in naher Zukunft nur in Ballungsräumen und entlang großer Verkehrswege installiert werden. Die jeweiligen Netzabdeckungen für GSM und UMTS in Deutschland sind in Abbildung 4-2a) und Abbildung 4-2b) dargestellt.



a) b) c)

Abbildung 4-2: Netzabdeckung für Deutschland in rot für a) GSM und b) UMTS des Mobilfunknetzes jeweils der Telekom Deutschland GmbH (Bildquelle: NETZABDECKUNG T-MOBILE (2010))
und c) Versuchsgebiet zur Mobilfunkerfassung der vorliegenden Arbeit

Da moderne Mobilfunkgeräte zwischen GSM und UMTS Mobilfunknetzen wechseln können, ist letztlich für die Mobilfunkortung eine Verknüpfung der Daten aus beiden Mobilfunknetzen anzustreben. Die vorliegende Arbeit baut im Wesentlichen auf die im Projekt DO-IT gewonnenen Mobilfunkdaten auf, worin ausschließlich Mobilfunkdaten des GSM erfasst und verarbeitet wurden. Das Versuchsgebiet befindet sich im Südwesten Deutschlands, im Großraum Karlsruhe, und umfasst eine Fläche von ca. 25 km^2 mit ca. 32 km Autobahnen und ca. 40 km Bundesstraßen, vgl. Abbildung 4-2c).

4.2.2 Mobilfunkortung nach dem Signalpegel-Matching-Verfahren

Wie die Verfahren der Verkehrserfassung, erfordert auch das funktechnische Grundprinzip für die Kommunikation im GSM-Netz die Kenntnis der MS-Position. Die Anforderungen an deren Darstellung sowie die Positionsgenauigkeit differieren jedoch. Im GSM-Netz erfolgt die Positionsdarstellung im Wesentlichen auf logischer Ebene in Form der Location-Area, der BTS, der Funkfrequenz und dem durch den **Timing Advance** (TA) angegebenen Zeitschlitz des Kommunikationskanals. Für FPD sind aber geografische Positionen von Interesse. Die Anwendbarkeit der GSM-Mobilfunkdaten setzt somit die Überführung der GSM-Positionsbeschreibung in geografische Koordinaten voraus. Für diese als **Mobilfunkortung** bezeichnete Problemstellung, finden sich in der Literatur mehrere Verfahren. Sie unterscheiden sich vorrangig durch den Einsatz unterschiedlicher, aus den Datenschnittstellen der Mobilfunknetze oder der Mobilfunkgeräte abgegriffener Mobilfunkdaten (KENNEMANN 1997; HELLEBRANDT ET AL. 1997; HELLEBRANDT & MATHAR 1999; KRIEGL

2000; WERMUTH ET AL. 2001; HOFMAN-WELLENHOF ET AL. 2003; CATREIN ET AL. 2004; KOHLEN 2004; RAMM ET AL. 2006; SCHWIEGER 2007; RAMM & SCHWIEGER 2007; KHALAF-ALLAH 2008).

In der vorliegenden Arbeit wird das im Projekt DO-IT eingesetzten und in RAMM & SCHWIEGER (2007) beschriebenen Ortungsprinzips des **Signalpegel-Matchings** angewendet. Für die Erfassung der Mobilfunkdaten muss bedingt dieses Verfahren eine aktive Verbindung zwischen MS und Infrastruktur des Mobilfunknetzes. Um sowohl den Unsicherheitshaushalt der eingesetzten Mobilfunkdaten, als auch die empirischen Analysen am Ende der vorliegenden Arbeit geeignet bewerten und einordnen zu können, wird dieses Ortungsprinzip in den folgenden Abschnitten im Detail betrachtet.

4.2.2.1 Das Ortungsprinzip des Signalpegel-Matchings

Das Grundprinzip des Signalpegel-Matchings basiert auf der bestmöglichen Anpassung der **gemessenen Signalpegel** (RX) an die Pegel der Referenzpegeldaten[7]. Die von einem MS gemessenen Signalpegel werden dazu mit denen einer georeferenzierten Signalpegelkarte verglichen (RAMM & SCHWIEGER 2007). An denjenigen Stellen mit bester Übereinstimmung zwischen gemessenen Pegel und Kartenpegel wird schließlich die Position des MS angenommen. Diese MS-Position wird nachfolgend mit $\mathbf{MS}_{xy}$ bezeichnet. Eine mögliche Lösung dieses Optimierungsproblems ist die analytische Auswertung beispielsweise mittels der Überprüfung der Definitheit der Hesse-Matrix (HELLEBRANDT ET AL. 1997). Dieser Weg wird im Folgenden nicht näher betrachtet. Stattdessen werden die Referenzpegel in Form georeferenzierter Rasterdaten, den sogenannten **Signalpegelkarten** eingesetzt.

Die Signalpegel der Signalpegelkarten werden in einer Zwei-dimensionalen Matrix **K** angeordnet. Ausgehend von einer georeferenzierten Ursprungskoordinate x_0, y_0 für den Mittelpunkt des Rasterelementes der linken oberen Ecke und einem Rasterpunktabstand a_K wird damit ein lokales kartesisches Koordinatensystem aufgespannt. Das Matrix-Koordinatensystem ist zudem achsparallel zum übergeordneten Koordinatensystem, das in der vorliegenden Arbeit das globale Gauß-Krüger-Koordinatensystem ist. Die Werte der Signalpegelkarten $k_{i,j}$ werden in Matrix-Darstellung formuliert mit:

$$\mathbf{K}^{BTS} = \begin{bmatrix} k_{1,1} & \cdots & & k_{1,m} \\ \vdots & \ddots & & \vdots \\ & & k_{i,j} & \\ & & & \ddots \\ k_{n,1} & & \cdots & k_{n,m} \end{bmatrix} \quad \text{mit: } i = 1,2...n,\ j = 1,2...m. \tag{4.1}$$

Das Kürzel BTS (Base Transceiver Station) steht für einen eindeutigen Mobilfunkmasten im Mobilfunknetz und kennzeichnet die eindeutige Zuordnung der Signalpegelkarte zu einer spezifischen Mobilfunkantenne der Infrastruktur. Die vom MS gemessenen Signalpegel werden mit RX_d bezeichnet. Mit dem Index d werden die vom MS gemessenen Signalpegel den Signalpegelkarten zugeordnet. Der Index $d = 0$ entspricht der ersten BTS (RAMM 2007). Sie wird als **bedienende BTS** bezeichnet, da zwischen ihr und dem MS die eigentliche Datenfunkübertragung stattfindet. Alle weiteren BTS werden als **Nachbarzellen** mit $d \in 1,\ldots,d_{max}$ bezeichnet, wobei im GSM-Mobilfunknetz deren Anzahl auf maximal $d_{max} = 6$ begrenzt ist (WALKE 2001). Je nach Empfangssituation variiert die Anzahl der Nachbarzellen, wodurch auch Fälle ohne Nachbarzelleninformation auftreten. Da die korrekte Zuordnung der gemessenen Pegel des MS zu den BTS und damit zu den „richtigen" Signalpegelkarten vorausgesetzt wird, wird nachfolgend das Kürzel BTS weggelassen. Die Angabe $\mathbf{K}_d$ und RX_d in einer Formel bedeuten dann, dass die BTS der Signalpegelkarte und die BTS des gemessenen Pegels gleich sind.

[7] Die Bezeichnung *Referenz* steht in diesem Zusammenhang für die *Referenzierung der Signalpegel auf ein Koordinatensystem*, also im Sinne wie *Georeferenzierung*, nicht aber im Sinne einer höheren Genauigkeit.

Jedem Kartenelement $k_{i,j}$ ist eine georeferenzierte Koordinate $k_{i,j} \rightarrow x, y$ zugeordnet. Die Zuordnung ist ebenfalls für die Rückrichtung $x, y \rightarrow k_{i,j}$ gültig:

$$\begin{aligned} x &= x_0 + (i-1)\cdot a_K \\ y &= y_0 + (j-1)\cdot a_K \end{aligned} \quad \text{und} \quad \begin{aligned} i &= 1 + (x - x_0)/a_K \\ j &= 1 + (y - y_0)/a_K . \end{aligned} \tag{4.2}$$

Aus diesem Zusammenhang resultiert die bereits oben erwähnte Bezeichnung *Referenzpegeldaten*. Aufgrund der in beide Richtungen eindeutigen Abbildung $k_{i,j} \leftrightarrow x, y$ ist die Angabe von i, j und x, y gleichbedeutend. Der Index des ersten Kartenelementes bei (x_0, y_0) ist $(i = 1, j = 1)$ und der des letzten ist $(i = n, j = m)$.

Im Verfahren des Signalpegel-Matchings wird diejenige Stelle in den Signalpegelkarten gesucht, an der sich die gemessenen Signalpegel optimal einpassen. Als Maß des Einpassungsgrades dient die Verbesserungsquadratsumme $v^2_{\Sigma,i,j}$, wobei an der Stelle mit dem wertemäßig kleinsten $v^2_{\Sigma,i,j}$ die MS-Position vermutet wird (CATREIN ET AL. 2004; HELLEBRANDT ET AL. 1997; HELLEBRANDT & MATHAR 1999; RAMM ET AL. 2006; RAMM & SCHWIEGER 2007; SCHWIEGER 2007):

Elementweise Darstellung	Matrizendarstellung
$\min\limits_{\substack{i=1\ldots n,\\ j=1\ldots m}} v^2_{\Sigma,i,j} \rightarrow \mathbf{MS}_{xy}$	$\min\limits_{\substack{i=1\ldots n,\\ j=1\ldots m}} \mathbf{v}^2_{\Sigma} \rightarrow \mathbf{MS}_{xy}$
$\text{mit}: v^2_{\Sigma,i,j} = \sum\limits_{d=0}^{d_{max}} \left(k_{i,j,d} - RX_d\right)^2$	$\text{mit}: \mathbf{v}^2_{\Sigma} = \sum\limits_{d=0}^{d_{max}} \left(\mathbf{K}_d - \mathbf{E}\cdot RX_d\right)^2 , \ \mathbf{E} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$

(4.3)

Zur Erhöhung der Recheneffizienz wird das Suchgebiet auf sogenannte **Sektoren** eingegrenzt. Die Flächen der Sektoren sind im Vorfeld so zu wählen, dass die MS-Positionen einerseits definitiv innerhalb der Sektoren liegen, andererseits die Flächengrößen der Sektoren aber möglichst klein sind. Für die Bestimmung der Sektorlage und -größe wird der folgende Zusammenhang genutzt: In GSM-Mobilfunknetzen findet die Informationsübertragung auf dem Luftweg über eine funktechnische Verbindung zwischen dem MS und der bedienenden BTS statt. Die BTS besitzen hierzu eindeutig zugeordnete Antennen. Im Rahmen der Installation und Inbetriebnahme der BTS-Antennen werden die geografische Koordinate, die **Hauptstrahlrichtung der Antenne** gegenüber der Nordrichtung und der **Öffnungswinkel** (typische Werte sind 120°, 180° oder 360°) festgelegt. Die Übertragung zwischen MS und BTS ist zudem in sogenannten Zeitschlitzen organisiert, wodurch die Anzahl der von einer BTS versorgbaren MS erhöht werden kann. Über welchen Zeitschlitz MS und BTS kommunizieren, gibt der TA-Wert an (ROTH 2002). Letztlich wird mit ihm die Synchronisierung zwischen BTS und MS realisiert sowie Laufzeit aufgrund der Ausbreitungsgeschwindigkeit der Signale zwischen BTS und MS berücksichtigt. Aus dem TA-Wert kann die Entfernung zwischen BTS und MS abgeleitet werden, die sich aus den ausgesendeten Datenpaketen relativ zum vorgegebenen Takt der Zeitschlitze berechnet (WALKE 2001, S. 225 Formel (3.5)):

$$d_{TA} = \frac{TA \cdot c \cdot t_{bit}}{2} = \frac{TA \cdot 3\cdot 10^8 \frac{m}{s} \cdot 3{,}69\cdot 10^{-6} s}{2} = TA \cdot 554\ m. \tag{4.4}$$

Der Wertebereich des TA-Wertes beträgt 0-63, wobei die 64 möglichen Schritte ganzzahlig sind. Signalausbreitungsgeschwindigkeit c und Dauer eines Bits t_{bit} sind Konstanten im GSM-Mobilfunknetz. Die ganzzahlige Auflösung des TA-Wertes bewirkt Distanzabweichungen von bis $½ \cdot 554$ m wegen Rundungseinflüssen.

Mit der BTS-Koordinate, der aus dem TA berechneten Distanz d_{TA}, der Hauptstrahlrichtung der Antenne und dem Öffnungswinkel wird die Lage und Ausdehnung des Sektor definiert (RAMM & SCHWIEGER 2007). Die Abbildung 4-3 zeigt schematisch den Sektor einer BTS im zweidimensional aufgespannten Raum. Zu berücksichtigen ist, dass sich Unsicherheiten der sektordefinierenden Größen (BTS-Koordinate, d_{TA},

Hauptstrahlrichtung und Öffnungswinkel) prinzipiell auf den Sektor fortpflanzen. Da mit Einführung der Sektoren ausschließlich Recheneffizienzsteigerungen durch die Eingrenzung des Suchgebietes für die MS-Position erzielt werden sollen, werden die Unsicherheiten des Sektors im Rahmen der vorliegenden Arbeit vereinfacht, in Form von konstanten radialen Zuschlägen von $\pm ½ \cdot 554\,\text{m}$ (zusätzlich zu den Rundungseinflüssen) sowie Zuschlägen beim Öffnungswinkel von ± 20 Grad behandelt.

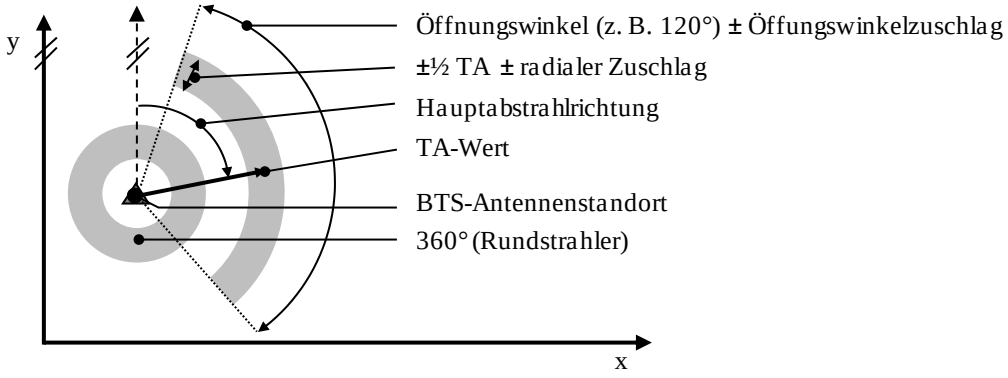


Abbildung 4-3: Zweidimensionale Abbildung von Sektoren für GSM-Verbindungen nach RAMM & SCHWIEGER (2007)

4.2.2.2 Gemessene Signalpegel und deren Eigenschaften

Die in der vorliegenden Arbeit eingesetzte Mobilfunkortung verwendet die am MS durchgeführten **Signalpegelmessungen** des GSM-Mobilfunknetzes. Die hierfür gemessenen Funksignale werden von jeder BTS kontinuierlich und mit gleicher Sendeleistung abgestrahlt. Davon abweichend wird die Sendeleistung der bedienenden BTS bei einer bestehenden Verbindung geregelt, um Funkstörungen zu minimieren. Die Signalpegelmessungen zur leistungsgesteuerten bedienenden BTS können je 480 Millisekunden 100 mal erfolgen und werden anschließend MS-intern gemittelt. Aufgrund der Leistungssteuerung muss ein Korrekturwert angebracht werden, um einen mit den Nachbarzellen und insbesondere mit den Signalpegelkarten vergleichbaren Signalpegelwert zu erhalten. Je 480 Millisekunden werden aufgeteilt auf die benachbarten BTS ebenfalls 100 Signalpegel erfasst und MS-intern gemittelt. Die Mittelwerte der bedienenden BTS $\overline{RX}_0$, der sechs vom Betreiber vorgegebenen sowie pegelstärksten Nachbar-BTS $\overline{RX}_d \,|\, d \in 1,2\ldots$, der TA-Wert und weitere Systemparameter werden in der Regel im 480 ms Zeitraster zur bedienenden BTS im Rahmen des **Messwertreports** übertragen. Die Auflösung der übermittelten Signalpegel beträgt 1 dBm (WALKE 2001, S. 219ff). Die Einheit dBm gibt einen Leistungspegel auf der Skala des dekadischen Logarithmus, bezogen auf 1 Milliwatt an (DETLEFSEN & SIART 2006).

Maßgebliche Charakteristik der Signalpegel entsteht durch den **Pegelverlauf aufgrund von Ausbreitungseffekten**. Im Wesentlich können die Einflüsse des Pegelverlaufs in die drei Bestandteile Kurzzeitschwund, Abschattung und dem, im nachfolgenden Abschnitt 4.2.2.3 näher erläuterten, Langzeitmittelwert aufgeteilt werden (WALKE 2001, S. 222). Alle drei Bestandteile sind ortsabhängige Größen und maßgeblich als Dämpfung des Signals zu verstehen. Die Summe dieser drei Bestandteile beschreibt in guter Näherung die Gesamtdämpfung des Signals entlang des Ausbreitungspfades und damit die empirisch beobachtbaren Pegelwerte.

Der **Kurzzeitschwund** wird durch Reflexionen und Beugungen einzelner Teilwellen verursacht. Für ein sich bewegendes MS werden derartige Einbrüche zeitlich nur kurz wahrgenommen, woher die Bezeichnung des Kurzzeitschwundes abgeleitet wurde. Empirische Untersuchungen zeigen, dass die durch den Kurzzeitschwund verursachte Standardabweichung der Pegelmessung im GSM-Netz ca. $\sigma_{RX} = 3{,}2\,\text{dB}$ beträgt (KENNEMANN 1997 S. 32).

Die **Abschattung** ist als Abschwächung des Signals aufgrund von Hindernissen zu verstehen. Sie kann in topografische Einflüsse (Geländeformen wie ebenes Gebiet, Hügel, Berge), morphologische Einflüsse (z. B. offenes Gebiet, Wald, Dorf, Stadt, Großstadt) und auch Einflüsse innerhalb von Gebäuden (z. B. Leichtbau, Betonwände, Wanddicken) unterschieden werden. Die Abschattungen können sich auf Gebiete von bis zu einigen 10 m und mehr auswirken. Für ein sich bewegendes MS ist die Signaldämpfung bei Abschattungen von längerer Dauer als beim Kurzzeitschwund, weshalb dieser Effekt auch als **Langzeitschwund** bezeichnet wird (KENNEMANN 1997, S. 29).

4.2.2.3 Prognostizierte Signalpegelkarten und deren Eigenschaften

Im Rahmen der Planung und Optimierung ihrer Mobilfunknetze erstellen Mobilfunkbetreiber in der Regel Signalpegelkarten für jeden BTS-Standort. Ein typisches Ergebnis einer derartigen Signalpegelkarte zeigt Abbildung 4-4. Man erkennt den Standort der BTS-Antenne in unmittelbarer Nähe zu den höchsten Signalpegeln.

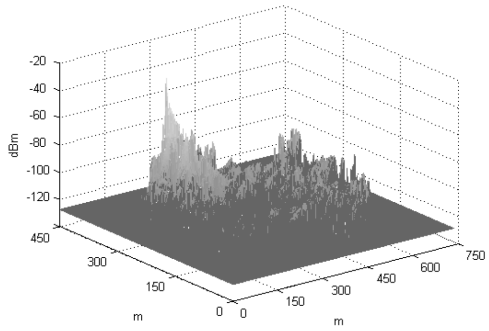


Abbildung 4-4: Signalpegel im örtlichen Koordinatensystem (RAMM & Schwieger 2007)

Ausdehnung und geometrische Auflösung der Signalpegelkarten sind unter anderem abhängig vom Anwendungszweck und den verfügbaren topografischen und morphologischen Datenbanken. Trotz bekannter Reflexionsgesetze und zahlreicher Dämpfungsmodelle ist die Prognose von Signalpegelkarten schwierig. Grund sind die zwingend erforderlichen detaillierten Informationen z. B. der Geländestruktur, Bebauung, Dachformen, Oberflächengestalt, verwendeten Baumaterialien oder Vegetation. Insbesondere für weit ausgedehnte Gebiete stehen derartige Daten nicht in vollem Umfang zur Verfügung oder können wirtschaftlich nicht erfasst werden. Außerdem sind Auswirkungen zeitlicher Veränderungen in der Bebauung, den Jahreszeiten sowie kurzfristige Wettereinflüsse ebenfalls zu berücksichtigen. Als weitere Einschränkung ist der für große Gebiete erforderliche hohe Rechenaufwand zu nennen (SCHMEINK 2005).

Für die Prognose von Pegelwerten werden daher in den Dämpfungsberechnungen neben Informationen zur Antenne, wie z. B. Höhe, Neigung, Sende- und Empfangscharakteristika sowie die Sendeleistung, häufig nur die entfernungsabhängigen Anteile und der Langzeitmittelwert berücksichtigt (LÜDERS 2001, S. 264ff). Der **Langzeitmittelwert** beschreibt einen gemittelten Signalpegel für ein großes Gebiet z. B. im Umkreis von 1-1,5 km. Charakteristisch für diesen Mittelwert ist, dass die Auswirkungen des Kurzzeitschwundes und kleinräumiger Abschattungen verschwinden. In OKUMURA ET AL. (1968) wurden mittels umfangreicher Messungen Langzeitmittelwerte für Gebiete um Tokio empirisch erfasst und kategorisiert. Im Anschluss wurden diese Messungsdaten in HATA (1980) mathematisch formuliert. In der Planung und Optimierung von Mobilfunknetzen sind die nach ihnen benannten *Okumura/Hata-Ausbreitungsmodelle* weit verbreitet und sind Ausgangspunkt zahlreicher Verbesserungen (BLAUSTEIN & ANDERSEN 2002; GENG & WIESBECK 1998; JUNIUS 1995; LEE 1997; LEE & MILLER 1998; LÜDERS 2001; PARSONS 2000; SIWIAK 1998). Die Prognose des Langzeitmittelwertes erfolgt für großräumige Gebiete (1 km-20 km) mit Simulationsprogrammen, die neben technischen Daten

zur Antenne, Daten der Topographie und Morphologie sowie die oben angesprochenen Ausbreitungsmodelle mit einbeziehen (WALKE 2001, S. 29f, S. 47f). Da die Einflüsse aus Kurzzeitschwund und Abschattung für diese (großräumigen) Prognosen in der Regel nicht prognostiziert werden, sind deren Auswirkungen beim Vergleich real gemessener Signalpegel zu prognostizierten Signalpegeln prinzipiell enthalten.

Die der vorliegenden Arbeit zugrunde liegenden Signalpegelkarten wurden durch die T-Mobile Deutschland[8] erstellt. Hierzu wurde die kommerzielle Software „PegaPlan“ der T-Systems International GmbH eingesetzt, deren Berechnungsroutinen nicht veröffentlicht sind.

4.2.2.4 Unsicherheiten der Signalpegelwerte

Ziel des Signalpegel-Matchings ist es, Positionen der MS zu bestimmen, ohne die spezifischen Eigenschaften des einzelnen Mobilfunkgerätes berücksichtigen zu müssen. Hersteller, Baureihe, Typ etc. der Mobilfunkgeräte sollen keinen unmittelbaren Einfluss auf die Positionsbestimmung haben, um ein möglichst geräteunabhängiges Ortungsverfahren anwenden zu können. Speziell werden hierzu die Vergleichbarkeit der gemessenen Signalpegel zwischen allen MS sowie deren Vergleichbarkeit zu den prognostizierten Signalpegelkarten gefordert.

Die Vergleichbarkeit der gemessenen Signalpegel zwischen den Mobilfunkendgeräten ist ebenfalls eine wesentliche Voraussetzung für die Funktion der GSM-Mobilfunknetze. Zu diesem Zweck wurden Grenzwerte für maximal zulässige Pegelabweichungen für die durch die GSM-Mobilfunkgeräte gemessenen Signalpegel festgelegt. Der **Root Means Square** (RMS) der Pegelabweichungen muss innerhalb ±4 dB von -110 dBm bis -70 dBm und ±6 dB im gesamten Empfangsbereich liegen (3GPP 2009, S. 39). Vor Inbetriebnahme müssen die Geräte mit entsprechenden Kalibrier- und Justierverfahren signaltechnisch so abgeglichen werden, dass der RMS innerhalb der geforderten Grenzen liegt. Während die originären Signalpegel selbst einem Zufallsprozess unterliegen, sind aufgrund der Justierverfahren systematische Abweichungen zu erwarten. Zusätzlich, und zum Teil zeitlich variierend, sind weitere systematische Abweichungen beispielsweise durch Temperatur, Luftdruck, Feuchte oder durch die Antennen, Kabel, elektronischen und anderen Bauelementen unumgänglich. Die gemessenen Signalpegel setzen sich somit aus einem zufälligen und einem systematischen Anteil zusammen:

$$\underline{RX}'_d = \overline{\underline{RX}}_d + \delta_{RX,d}. \tag{4.5}$$

Neben systematischen Abweichungen der Pegelmessung wirken sich auch die Abweichungen der prognostizierten Signalpegelkarten auf die MS-Ortung aus. Obwohl für jede BTS-Antenne eine eigene Signalpegelkarte prognostiziert wird, resultieren die darin enthaltenen Pegelwerte in der Regel maßgeblich aus den gleichen Dämpfungsmodellen (mit ihren Annahmen und Vereinfachungen in den Signalausbreitungsmodellen) sowie den gleichen topographischen und morphologischen Datenbeständen, vgl. Abschnitt 4.2.2.3. Es ist daher anzunehmen, dass Signalpegelkarten unterschiedlicher BTS-Antennen Korrelationen ausweisen. Gegenüber der Realität ist somit von systematisch wirkenden Pegelabweichungen der Signalpegelkarten $\delta_{k,i,j,d}$ auszugehen. Folgerichtig müssen die prognostizierten Signalpegel um diesen systematischen Anteil $k'_{i,j,d} = k_{i,j,d} + \delta_{k,i,j,d}$ ergänzt werden. Formel (4.3) wird erweitert zu:

$$\begin{aligned} \min_{\substack{i=1\ldots n,\\ j=1\ldots m}} v^2_{\Sigma,i,j} \rightarrow \mathbf{MS}_{xy} \qquad \text{mit}: \; v^2_{\Sigma,i,j} &= \sum_{d=0}^{d_{max}} \left(\left(k_{i,j,d} + \delta_{k,i,j,d} \right) - \left(\overline{\underline{RX}}_d + \delta_{RX,d} \right) \right)^2 \\ &= \sum_{d=0}^{d_{max}} \left(k_{i,j,d} - \overline{\underline{RX}}_d + \delta_{k,i,j,d} - \delta_{RX,d} \right)^2. \end{aligned} \tag{4.6}$$

[8] Die Geschäftseinheiten T-Mobile und T-Home wurden in Deutschland am 01.04.2010 in die Telekom Deutschland GmbH zusammengeführt (TELEKOM 2013)

4.2.2.5 Signalpegel-Matching mit Pegeldifferenzen

Da die systematischen Abweichungen auf mehrere Größen gleichartig wirken, kann der Einfluss dieser Abweichungen mittels Differenzbildung reduziert werden. Eine mögliche Anwendung dieses Prinzips auf das Signalpegel-Matching ist die Verwendung von Pegeldifferenzen. Mit Bildung der Pegeldifferenzen zwischen den Pegeln eines Zeitpunktes sowie jeweils dem Bezug auf die bedienende BTS konnten in BEETZ ET AL. (2009) und BEETZ ET AL. (2011) verbesserte Ortungsergebnisse gezeigt werden. Es kann schlussgefolgert werden, dass mit diesem Mobilfunkortungsverfahren systematische Abweichungen der Signalpegelmessung $\underline{\mathrm{RX}}_{\mathrm{d}}^{'}$ - zumindest diejenigen, die Offseteigenschaften aufweisen - eliminiert werden können. Systematische Abweichungen der Signalpegelkarten sind jedoch nur teilweise reduzierbar, da sie unterschiedliche Größe aufweisen können. In jeder Verbesserung der Pegeldifferenzen $v_{\mathrm{D,i,j,d}}$ verbleibt folglich ein Anteil an **systematischen Abweichungen mit unbekannter Größe und unbekanntem Vorzeichen** $\Delta_{\mathrm{i,j,d}}$. Im Folgenden wird aus Vereinfachungsgründen angenommen, dass die Bedingungen bei jedem Kartenelement des betreffenden Sektors näherungsweise gleich sind. Folglich kann für die systematischen Abweichungen vereinfacht von $\Delta_{\mathrm{d}} = \Delta_{\mathrm{i,j,d}}$ ausgegangen werden, so dass das für funktionale Modell der Mobilfunkortung mittels Pegeldifferenzen gilt:

$$\min_{\substack{\mathrm{i=1\ldots n,}\\ \mathrm{j=1\ldots m}}} v_{\Sigma\mathrm{D,i,j}}^2 \rightarrow \mathbf{MS}_{\mathrm{xy}} \quad \text{mit}: v_{\Sigma\mathrm{D,i,j}}^2 = \sum_{\mathrm{d=1}}^{\mathrm{d_{max}}} \left(v_{\mathrm{D,i,j,d}} \right)^2 = \sum_{\mathrm{d=1}}^{\mathrm{d_{max}}} \left(\mathrm{k_{i,j,0}} - \mathrm{k_{i,j,d}} + \overline{\mathrm{RX}}_{\mathrm{d}} - \overline{\mathrm{RX}}_{0} + \Delta_{\mathrm{d}} \right)^2. \tag{4.7}$$

Die Bildung von Pegeldifferenzen bedingt, dass neben der Pegelmessung der bedienenden BTS mindestens der Pegel einer Nachbar-BTS $\mathrm{d_{max}} \geq 1$ vorliegen muss.

4.2.3 Herleitung der Unsicherheiten für das Signalpegel-Matching mit Pegeldifferenzen

Zur Übertragung der Unsicherheiten aus dem Signalpegelraum in den Koordinatenraum werden in den folgenden Abschnitten die drei Arten der Unsicherheitsmodellierung (zufällige Variabilität, Fuzzy-Theorie, Fuzzy-Randomness) auf Formel (4.7) angewendet. Es wird davon ausgegangen, dass sich die tatsächliche MS-Position im Sektor befindet, der aus Signalpegelkarten ausgeschnitten wurde. Der Sektor selbst, wird durch die Kartenelemente in regelmäßige Flächenelemente unterteilt. Mit der Verbesserungsquadratsumme $v_{\Sigma\mathrm{D,i,j}}^2$ wird für jedes Kartenelement ein Übereinstimmungsmaß der gemessenen Pegel zu den prognostizierten Kartenpegeln berechnet. Fehlertheoretisch betrachtet, spiegeln die $v_{\mathrm{D,i,j,d}}$ in Formel (4.7) die Differenz $l_{\mathrm{d}} - \hat{l}_{\mathrm{d}}$ zwischen den Beobachtungen l_{d} gegenüber den wahren Werten $\hat{l}_{\mathrm{d}}$ [9] wider, obgleich die tatsächlichen wahren Werte, also die wahren Pegel in der Örtlichkeit, unbekannt bleiben. Die Differenz zwischen Beobachtung und wahrem Wert entspricht der wahren Abweichung $\eta_{\mathrm{d}} = \varepsilon_{\mathrm{d}} + \Delta_{\mathrm{d}} = l_{\mathrm{d}} - \hat{l}_{\mathrm{d}}$. Größe und Vorzeichen von ε_{d} und Δ_{d} bleiben jedoch verborgen. Für die in der Formel (4.7) ermittelten Verbesserungen $v_{\mathrm{D,i,j,d}}$ bedeutet das, dass sie sämtliche stochastischen und systematischen Abweichungen enthalten. Daraus folgt, dass auch die Verbesserungsquadratsummen $v_{\Sigma\mathrm{D,i,j}}^2$ sämtliche stochastischen und systematischen Abweichungen widerspiegeln.

Mit diesem Hintergrund weist der Wert der Verbesserungsquadratsumme eines Kartenelementes mit dem Gewicht $g_{\mathrm{i,j}}$ auf die Möglichkeit hin, als Lösung für die MS-Position infrage zu kommen. Mit der Normierung, so dass die Summe aller Gewichte der Kartenelemente eines Sektors dem Wert Eins entspricht, folgt:

$$g_{\mathrm{i,j}} = \frac{1 \big/ v_{\Sigma\mathrm{D,i,j}}^2}{\sum_{\mathrm{i=1}}^{\mathrm{n}} \sum_{\mathrm{j=1}}^{\mathrm{m}} \frac{1}{v_{\Sigma\mathrm{D,i,j}}^2}}, \quad \text{mit:} \quad g_{\mathrm{i,j}} \in [0,1], \quad \sum_{\mathrm{i=1}}^{\mathrm{n}} \sum_{\mathrm{j=1}}^{\mathrm{m}} g_{\mathrm{i,j}} = 1, \quad v_{\Sigma\mathrm{D,i,j}}^2 > 0. \tag{4.8}$$

[9] Abweichend von der im Vermessungswesen üblichen Schreibweise des wahren Wertes mit ~ (Tilde) wird hier ^ (Bogen) verwendet, um Verwechslungen mit den in der Literatur zur Fuzzy-Theorie häufig mit ~ gekennzeichneten unscharfen Mengen zu vermeiden.

Sowohl die Kartenelemente als auch die gemessenen Signalpegel stehen nur als ganzzahlige Werte mit 1 dB Schrittweite zur Verfügung. Aufgrund dessen nimmt $v^2_{\Sigma D,i,j}$ nicht selten den Wert Null an. Aus rechentechnischen Gründen wird für diese Fälle dem Kartenelement das maximale Gewicht zugeordnet. Formel (4.8) ändert sich dadurch in:

$$g_{i,j} = \frac{g^*_{i,j}}{\sum_{i=1}^{n}\sum_{j=1}^{m} g^*_{i,j}}, \quad \text{mit}: g^*_{i,j} = \begin{cases} g^*_{i,j} = \dfrac{1}{v^2_{\Sigma D,i,j}} & , \text{falls}: v^2_{\Sigma D,i,j} > 0 \\ g^*_{i,j} = 1 & , \text{falls}: v^2_{\Sigma D,i,j} = 0. \end{cases} \tag{4.9}$$

4.2.3.1 Zufällige Variabilität

Im Falle der zufälligen Variabilität wird angenommen, dass die räumliche Verteilung der Gewichte $g_{i,j}$ daher resultiert, dass sie im xy-Raum einer gemeinsamen Grundgesamtheit entstammen und zufälliger Natur sind. Darauf aufbauend können der Mittelpunkt der MS-Position $\underline{\mathbf{MS}}_{xy}$ und die Varianz-Kovarianzmatrix $\mathbf{S}_{\underline{\mathbf{MS}}_{xy}}$ als Schätzungen für die Parameter $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ der zweidimensionalen Normalverteilung, vgl. Formel (3.6), mit Hilfe der Kartenelementkoordinaten $\mathbf{k}_{xy,i,j}$ berechnet werden:

$$\begin{aligned} \underline{\mathbf{MS}}_{xy} &= \sum_{i=1}^{n}\sum_{j=1}^{m}\left(g_{i,j}\cdot\mathbf{k}_{xy,i,j}\right) \quad \text{mit: } \mathbf{k}_{xy,i,j} = \begin{pmatrix} x_{\mathbf{k}_{i,j}} \\ y_{\mathbf{k}_{i,j}} \end{pmatrix} \quad \left(\text{Koordinate des Kartenelementes}\right), \\ \mathbf{S}_{\underline{\mathbf{MS}}_{xy}} &= \sum_{i=1}^{n}\sum_{j=1}^{m} g_{i,j}\cdot\left(\underline{\mathbf{MS}}_{xy}-\mathbf{k}_{xy,i,j}\right)\cdot\left(\underline{\mathbf{MS}}_{xy}-\mathbf{k}_{xy,i,j}\right)^{T} \quad \text{mit: } \underline{\mathbf{MS}}_{xy} = \begin{pmatrix} x_{\underline{\mathbf{MS}}_{xy}} \\ y_{\underline{\mathbf{MS}}_{xy}} \end{pmatrix}. \end{aligned} \tag{4.10}$$

Die Anzahl der Kartenelemente je Sektor und damit der Wert für den Freiheitsgrad $f_{\underline{\mathbf{MS}}_{xy}}$ sind in der Regel hoch. Wenn nicht anders angegeben, wird im Folgenden näherungsweise von $f_{\underline{\mathbf{MS}}_{xy}} \cong \infty$ ausgegangen und auf die theoretische Varianz-Kovarianzmatrix $\boldsymbol{\Sigma}_{\underline{\mathbf{MS}}_{xy}}$ anstatt $\mathbf{S}_{\underline{\mathbf{MS}}_{xy}}$ übergegangen. Die Position des MS wird wie folgt angegeben mit:

$$\underline{\mathbf{MS}}_{xy} = \begin{pmatrix} \underline{x} \\ \underline{y} \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{\underline{\mathbf{MS}}_{xy}} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} = E\left(\mathbf{S}_{\underline{\mathbf{MS}}_{xy}}\right). \tag{4.11}$$

Der Unterstrich „ _ “ verdeutlicht die zufällige Variabilität des Vektors und der Vektorkomponenten.

4.2.3.2 Fuzzy-Theorie

In der Fuzzy-Theorie werden die Unsicherheitsmaße mithilfe Zugehörigkeitsfunktionen charakterisiert. Die Unsicherheit der MS-Position kennzeichnen bislang die Gewichtswerte und deren Verteilung im xy-Koordinatenraum. Gesucht ist daher eine geeignete Abbildung, um die Gewichte in Zugehörigkeitsfunktionen zu überführen $g_{i,j} \in \Re \rightarrow m_{\widetilde{MS}_{xy}}(x,y) \in [0,1]$. In einem ersten Schritt wird dafür eine Skalierung der Gewichte in das Intervall $[0,1]$ (Normierung) durchgeführt $g_{i,j} \in \Re \rightarrow g^{\text{skaliert}}_{i,j} \in [0,1]$. Das zahlenmäßig kleinste Gewicht erhält den skalierten Wert Null und entsprechend das größte Gewicht den skalierten Wert Eins. Für die Skalierung der übrigen Gewichtswerte kann eine beliebige monotone Funktion Einsatz finden. In der vorliegenden Arbeit wird die folgende lineare Skalierungsfunktion verwendet:

$$g^{\text{skaliert}}_{i,j} = \frac{g_{i,j} - \min\left(g_{i,j}\right)}{\max\left(g_{i,j}\right) - \min\left(g_{i,j}\right)}, \quad g^{\text{skaliert}}_{i,j} \in [0,1]. \tag{4.12}$$

Mit Ausnahme des Kartenelementes mit dem kleinsten Gewichtswert, gilt nun für alle Kartenelemente des Sektors ein Gewichtswert $g^{\text{skaliert}} > 0$. Eine hieraus gebildete Zugehörigkeitsfunktion würde somit einen Träger zugeordnet bekommen, der (bis auf das Kartenelement mit $g^{\text{skaliert}} = 0$) die Größe des Sektors einschließt. Da

mit den Sektoren ausschließlich die Begrenzung des Suchgebietes und damit die Steigerung der Recheneffizienz anvisiert wird, würde der Träger der Zugehörigkeitsfunktion in Abhängigkeit von der Sektorgröße überschätzt (unplausibel zu groß) werden. Um eine derartige Überschätzung zu unterbinden, werden in der vorliegenden Arbeit nur diejenigen Kartenelemente für die Überführung in die Zugehörigkeitsfunktion verwendet, deren Gewichtswerte $g_{i,j}^{skaliert}$ einen Mindestwert überschreiten. Zur günstigeren Weiterverarbeitung sowie zur Gewährleistung einer konvexen Zugehörigkeitsfunktion wird die resultierende Menge der Gewichtswerte zudem durch einen achsparallelen Pyramidenstumpf im xy-Koordinatenraum approximiert. Aufgrund der Normierung der Gewichtswerte in das Intervall $[0,1]$ entspricht dieses Vorgehen einem α-Schnitt bei α_{min} für die untere Grundfläche des Pyramidenstumpfes $G_{\alpha_{min}}$ und einem weiteren α-Schnitt für die obere Grundfläche $G_{\alpha_{max}}$ bei α_{max}. Im Rahmen der Arbeit zeigte sich in empirischen Untersuchungen, dass $\alpha_{min} = 0{,}1$ und $\alpha_{max} = 0{,}8$ geeignete Anpassungen für die Form der Zugehörigkeitsfunktion an die Gewichtsmenge lieferten.

Entlang der x-Achse $\widetilde{MS}_x$ und entlang der y-Achse $\widetilde{MS}_y$ kann die Zugehörigkeitsfunktion nun durch eindimensionale Trapezfunktionen festgelegt werden. Das kartesische Produkt beider Trapezfunktionen $\widetilde{MS}_{xy} = \widetilde{MS}_x \times \widetilde{MS}_y \Rightarrow m_{\widetilde{MS}_{xy}}(x,y) \in [0,1]$ liefert schließlich eine konvexe Approximation der skalierten Gewichte im xy-Koordinatenraum. Je Koordinatenachse wird dazu der Kern durch die äußeren Positionen des α_{max}-Schnittes definiert mit:

$$\begin{aligned} m_x &= 0{,}5 \cdot \left(\min\left(G_{x,\alpha_{max}}\right) + \max\left(G_{x,\alpha_{max}}\right)\right), \quad m_y = 0{,}5 \cdot \left(\min\left(G_{y,\alpha_{max}}\right) + \max\left(G_{y,\alpha_{max}}\right)\right) \quad \text{Mittelpunkt} \\ r_x &= 0{,}5 \cdot \left(\max\left(G_{x,\alpha_{max}}\right) - \min\left(G_{x,\alpha_{max}}\right)\right), \quad r_y = 0{,}5 \cdot \left(\max\left(G_{y,\alpha_{max}}\right) - \min\left(G_{y,\alpha_{max}}\right)\right) \quad \text{Radius.} \end{aligned} \tag{4.13}$$

Der Träger wird analog dazu durch die äußeren Positionen des α_{min}-Schnittes definiert:

$$\begin{aligned} c_{l,x} &= m_x - r_x - \min\left(G_{x,\alpha_{min}}\right), \quad c_{l,y} = m_y - r_y - \min\left(G_{y,\alpha_{min}}\right) \quad \text{linke Spannweite} \\ c_{r,x} &= \max\left(G_{x,\alpha_{min}}\right) - m_x - r_x, \quad c_{r,y} = \max\left(G_{y,\alpha_{min}}\right) - m_y - r_y \quad \text{rechte Spannweite.} \end{aligned} \tag{4.14}$$

Grafische Interpretationen dazu zeigen die Abbildung 4-5a) und Abbildung 4-5b).

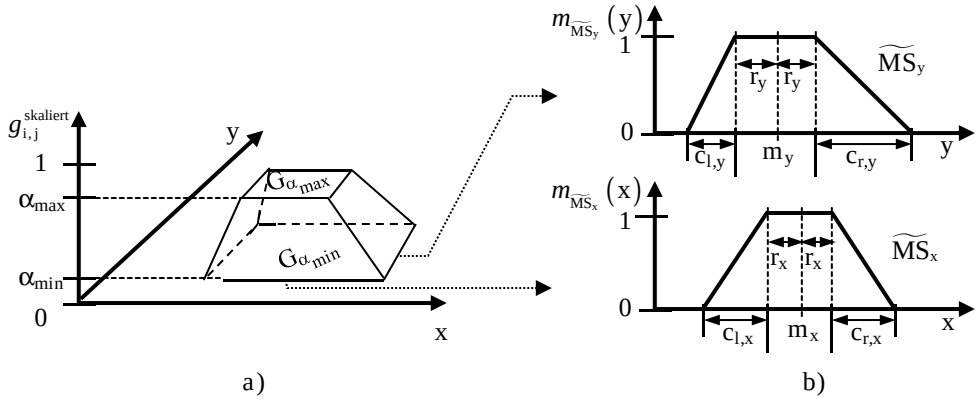


Abbildung 4-5: a) Pyramidenstumpf als geometrische Approximation der möglichen Aufenthaltsregion und b) dessen Formulierung als Verknüpfung zweier achsparalleler, trapezförmiger Fuzzy-Intervalle

Die Position des MS wird als zweidimensionales Fuzzy-Intervall angegeben mit:

$$\widetilde{MS}_{xy} := \left\{ \begin{bmatrix} m_x, r_x, c_{l,x}, c_{r,x} \\ m_y, r_y, c_{l,y}, c_{r,y} \end{bmatrix}_{LL}, m_{\widetilde{MS}_{xy}}(x,y) \;\middle|\; x,y \in \Re^2 \right\} \quad \text{mit: } m_{\widetilde{MS}_{xy}}(x,y) \in [0,1]. \tag{4.15}$$

Es sei angemerkt, dass $\widetilde{MS}_{xy}$ ausschließlich eine mengentheoretisch interpretierbare Approximation der aus den Gewichten abgeleiteten Aufenthaltsregionen ist.

4.2.3.3 Fuzzy-Randomness

Allein aus den gemessenen Signalpegeln und den prognostizierten Signalpegelkarten sind die systematischen Anteile Δ_d nicht bestimmbar, weshalb systematisch wirkende Unsicherheitsanteile nicht unmittelbar fortgepflanzt werden können. In den Abschnitten 4.2.3.1 und 4.2.3.2 wurde daher angenommen, dass die Verteilung der aus den Verbesserungsquadratsummen abgeleiteten Gewichte $g_{i,j}$ im xy-Raum, allein aufgrund ihrer zufälligen Variabilität erfolgt (Bestimmung der empirischen Varianz-Kovarianzmatrix). Die zufällige Variabilität der Summe $\varepsilon_d + \Delta_d$ (vgl. Abschnitt 4.2.3) wurde somit vorausgesetzt, was aufgrund der – in Vorzeichen und Größe unbekannten – systematisch wirkenden Anteile in Δ_d nur näherungsweise gerechtfertigt werden kann.

Eine den Eigenschaften der Unsicherheiten geeignetere Unsicherheitsfortpflanzung gelingt auf Basis der Fuzzy-Randomness-Theorie mittels FRV. Die zufälligen Anteile aus ε_d können darin mittels der Varianz-Kovarianzfortpflanzung berechnet werden. Die in Formel (4.7) mit Δ_d bezeichneten, systematisch wirkenden Fehleranteile werden mittels der Fuzzy-Theorie behandelt werden. Die Verteilung der aus den Verbesserungsquadratsummen abgeleiteten Gewichte $g_{i,j}$ im xy-Raum ist dann als Kombination aus zufälligen und systematisch wirkenden Anteilen interpretierbar und berechenbar. Für die vorliegende Anwendung ist es erforderlich, dass die Zugehörigkeitsfunktion von Δ_d geeignet bestimmt oder abgeschätzt wird, was mit Hilfe der Fuzzy-Theorie durch die Fuzzyfizierung von $\Delta_d \tilde{\rightarrow} \tilde{\Delta}_d$ gelingt.

Zur Konstruktion der Zugehörigkeitsfunktionen $m_{\tilde{\Delta}_d}$ wird angenommen, dass konstante systematische Einflüsse durch die Pegeldifferenzbildung vollständig eliminiert wurden. In empirischen Untersuchungen zur Pegelmessgenauigkeit in GSM-Netzen stellte JUNIUS (1995) fest, dass „*… die Messdaten der Mobilstationen im Mittel gut mit denen des Meßempfängers übereinstimmen…*“[10] JUNIUS (1995, S. 58). Diese Feststellung untermauert die zufällige Variation der Verbesserungen. Für den Gipfelpunkt gilt somit $\tilde{\Delta}_m = 0$. Ferner zeigte sich, dass die „*…Standardabweichungen…*“ der Einzelwerte „*…von ca. 1,5-3,5 dB um den Mittelwert…*“ (JUNIUS 1995, S. 58) variierten. Offensichtlich handelt es sich hier um einen in seinem Wert variierenden RMS. Die gemessenen Signalpegel bestehen demnach aus den Anteilen der zufälligen Variabilität mit einer Standardabweichung von ca. 1,5 dB und einem systematisch wirkenden Anteil, so dass Standardabweichungen von bis zu ca. 3,5 dB errechnet wurden. Aus diesen Feststellungen des variierenden RMS heraus werden die Spannweiten für alle Verbesserungen der Pegeldifferenzen näherungsweise mit $\tilde{\Delta}_c = \tilde{\Delta}_{c_l} = \tilde{\Delta}_{c_r} = 2\,\text{dB}$ und damit mittels einer symmetrischen dreieckförmigen Fuzzy-Zahl für $\tilde{\Delta}_d$ abgeschätzt, vgl. dazu die Abbildung 4-6. In dieser Abschätzung sind ebenfalls nicht modellierte oder nicht modellierbare Einflüsse aufgrund von Inkonsistenzen zwischen Daten (prognostizierten Signalpegel der Signalpegelkarten) und Realität, Rundungsfehler (z. B. aus der Datenübertragung im Messwertreport mit der Signalpegelauflösung von 1 dBm) sowie wegen Modellvereinfachungen vernachlässigte, gegebenenfalls bekannte, aber kleine Korrektionen enthalten.

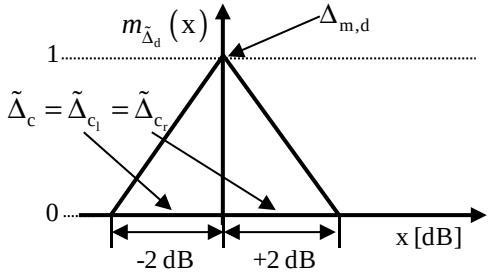


Abbildung 4-6: Fuzzy-Zahl zur Beschreibung der systematischen Abweichungen für die Verbesserungen des Signalpegel-Matchings mit Pegeldifferenzen

[10] Ein Messempfänger ist ein Prüfgerät, das als Referenzeinrichtung eine übergeordnete Messgenauigkeit des Signalpegels besitzt.

Durch die explizite Berücksichtigung der systematischen Anteile mit $\tilde{\Delta}_d$, können sie mit Δ^* aus den Verbesserungen der Pegeldifferenzen $v_{D,d,i,j} = \left(k_{i,j,0} - k_{i,j,d} + \overline{RX}_d - \overline{RX}_0\right) - \Delta^*$ herausgelöst werden. Da die Gipfelpunkte von $\tilde{\Delta}_d$ in Null liegen und die Spannweiten für alle Verbesserungen konstant gewählt wurden, wird jeder Verbesserungswert um $\Delta^* = 2$ dB reduziert. Formel (4.7) wird abgeändert in:

$$\min_{\substack{i=1\ldots n,\\ j=1\ldots m}} \tilde{v}^2_{\Sigma D,i,j} \rightarrow \widetilde{\underline{MS}}_{xy} \qquad \text{mit}: \tilde{v}^2_{\Sigma D,i,j} = \sum_{d=1}^{d_{max}} {}^{\oplus}\left(\tilde{v}_{D,d,i,j}\right)^2 = \sum_{d=1}^{d_{max}} \left(\left(\left(k_{i,j,0} - k_{i,j,d} + \overline{RX}_d - \overline{RX}_0\right) - \Delta^*\right) \oplus \tilde{\Delta}_d\right)^2. \quad (4.16)$$

Formel (4.16) kann nun α-schnittweise ausgewertet werden. Zur Berechnung der zufälligen Variabilität wird analog zum Ablauf in Abschnitt 4.2.3.1 mit den Formel (4.9) bis (4.11) der zufällige Mittelpunkt $\widetilde{\underline{MS}}_{xy,m}$ und die empirische Varianz-Kovarianzmatrix $\mathbf{S}_{\widetilde{\underline{MS}}_{xy}}$ bei $\alpha = 1$ berechnet.

Im Allgemeinen sind nun zur α-schnittweisen Unschärfefortpflanzung alle Kombinationsmöglichkeiten der systematischen Abweichungen $\tilde{\Delta}_{1,\alpha} * \ldots * \tilde{\Delta}_{d,\alpha} * \ldots * \tilde{\Delta}_{d_{max},\alpha}$ zu berechnen. Aus rechentechnischen Gründen werden in der vorliegenden Arbeit nur die Spannweiten bei $\alpha = 0$ und ausschließlich die ganzzahligen Werte berechnet, da sowohl die gemessenen Signalpegel als auch die Kartenpegel nur mit ganzzahligen Werten mit 1 dB Schrittweite zur Verfügung standen:

$$\tilde{\Delta}_{d,\alpha=0} = \{-2,-1,0,+1,+2\} \ \text{(jeweils in dB)}. \quad (4.17)$$

Mit Beschränkung auf $\alpha = 1$ und $\alpha = 0$ ergeben sich in Abhängigkeit von der Nachbarzellenanzahl d_{max} dennoch $1 + 5^{d_{max}}$ MS-Positionsberechnungen, so dass der Rechenaufwand bereits für $d_{max} > 2$ inakzeptabel ansteigt. Für diese Fälle wird eine weitere Vereinfachung vorgenommen, zu der angenommen wird, dass die Fortpflanzung der systematischen Abweichungen innerhalb eines α-Schnittintervalls linear verläuft, was aufgrund der quadratischen Terme nur näherungsweise gilt. Die Auswertung kann nun auf die Berechnung der Intervallgrenzen, also $\Delta_{d,\alpha=1} = \{0\}$ und $\Delta_{d,\alpha=0} = \{-2,+2\}$ beschränkt werden. Die Berechnungsanzahl reduziert sich dadurch von $1 + 5^{d_{max}=6} = 15626$ auf $1 + 2^{d_{max}=6} = 65$ MS-Positionsberechnungen. Diese MS-Positionsmenge spiegelt die Unschärfe im xy-Koordinatenraum und somit die Zugehörigkeitsfunktion der MS-Position wider:

$$m_{\widetilde{\underline{MS}}_{xy}} := \sup_{\alpha \in [0,1]} \min\left(m_{\tilde{\underline{v}}^2_{D\Sigma,i,j}}\left(\tilde{\underline{v}}^2_{D\Sigma,MS_{xy}}\right) \quad \middle| i \in 1\ldots n,\ j \in 1\ldots m\right). \quad (4.18)$$

Im Allgemeinen ist $m_{\widetilde{\underline{MS}}_{xy}}$ nicht konvex und wird für den weiteren Einsatz rechentechnisch ineffizient sein. Aus diesem Grund wird im Folgenden eine pyramidenförmige Approximation der Zugehörigkeitsfunktion verwendet, deren Gipfelpunkt $\widetilde{\underline{MS}}_{xy,\alpha=1}$ bei $\alpha = 1$ der zufällige Mittelpunkt ist und deren Spannweiten parallel zur x- und y-Richtung verlaufen. Die Spannweiten folgen aus den Ergebnissen des Schnittes bei $\alpha = 0$ mit:

$$\begin{aligned} c_{l,x} &= \widetilde{\underline{MS}}_{x,\alpha=1} - \min\left(\widetilde{\underline{MS}}_{x,\alpha=0}\right), \quad c_{r,x} = \max\left(\widetilde{\underline{MS}}_{x,\alpha=0}\right) - \widetilde{\underline{MS}}_{x,\alpha=1} \\ c_{l,y} &= \widetilde{\underline{MS}}_{y,\alpha=1} - \min\left(\widetilde{\underline{MS}}_{y,\alpha=0}\right), \quad c_{r,y} = \max\left(\widetilde{\underline{MS}}_{y,\alpha=0}\right) - \widetilde{\underline{MS}}_{y,\alpha=1}. \end{aligned} \quad (4.19)$$

Die FRV der MS-Position lautet:

$$\begin{aligned} \widetilde{\underline{MS}}_{xy} &:= \left(m_{\widetilde{\underline{MS}}_{xy}}(x,y), \mathbf{\Sigma}_{\widetilde{\underline{MS}}}\right) && : \text{FRV} \\ \text{mit:} \quad & m_{\widetilde{\underline{MS}}_{xy}}(x,y) \in [0,1] \quad \forall\, x,y \in \Re^2 \\ & m_{\widetilde{\underline{MS}}_{xy}}(x,y) = \left[\begin{matrix} m_x, r_x = 0, c_{l,x}, c_{r,x} \\ m_y, r_y = 0, c_{l,y}, c_{r,y} \end{matrix}\right]_{LL} && : \text{Anteil der Unschärfe (als Fuzzy-Zahl)} \\ & m_{\widetilde{\underline{MS}}_{xy}}(m_x, m_y) = 1 && : \text{Träger der zufälligen Variabilität} \\ & \mathbf{\Sigma}_{\widetilde{\underline{MS}}_{xy}} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} = E\left(\mathbf{S}_{\widetilde{\underline{MS}}_{xy}}\right) && : \text{Anteil der zufälligen Variabilität} \end{aligned} \quad (4.20)$$

Die zufällige Variabilität entspricht der Varianz-Kovarianzmatrix $\mathbf{S}_{\underline{\widetilde{MS}}}$ des zufälligen Mittelpunktes der MS-Position und wird analog zu Abschnitt 4.2.3.1 näherungsweise mit $\mathbf{\Sigma}_{\underline{\widetilde{MS}}_{xy}}$ anstatt mit $\mathbf{S}_{\underline{\widetilde{MS}}}$ und $f_{\underline{\widetilde{MS}}}$ angegeben. Die Unschärfe wird durch die Zugehörigkeitsfunktion $m_{\underline{\widetilde{MS}}_{xy}}(x,y)$ vom LL-Typ beschrieben, wobei die Referenzfunktionen L lineare Funktionen sind.

4.2.4 Datenpunktfusionierung mit FCM-Clustering

4.2.4.1 Motivation

Im Vergleich zur Größe der Koordinatenunsicherheiten von bis zu mehreren einhundert Metern ist der zeitliche Abstand zwischen den Positionsaktualisierungen mit 480 Millisekunden sehr gering. Ein Beispiel mit einer typischen MS-Positionsgenauigkeit von ca. 300 m veranschaulicht die Problemstellung: Werden MS mit einer Bewegungsgeschwindigkeit von ca. 50 km/h analysiert, dann sind nur alle ca. 700 m entfernt liegenden Positionen signifikant voneinander unterscheidbar. Zeitlich bedeutet das, dass nur alle ca. 50 Sekunden und damit nur ca. jede 105. Position von der Vorgängerposition signifikant unterscheidbar ist $(\alpha_{\text{Irrtum}} = 5\%)$. Der relative Informationsgewinn von einer MS-Position zur nächsten ist somit eher gering[11]. Mit Vorgriff auf die im späteren Verlauf der Arbeit erläuterten Identifizierungsverfahren *Korrelation zu ÖV-Fahrzeugen* in Abschnitt 4.5 und *kartengestützte Plausibilitätsprüfung* in Abschnitt 4.6, in denen umfangreiche Fahrplandaten und digitale Verkehrsnetzdaten je MS-Position verarbeitet werden, müssten sehr viele MS-Positionen mit sehr geringem Informationsgewinn ausgewertet werden. Der resultierende Rechenaufwand würde die Praxistauglichkeit der Identifizierungsverfahren mit heutigen Rechenanlagen infrage stellen. Mit dem Ziel die Effizienz der Identifizierungsverfahren dadurch zu steigern, dass möglichst nur MS-Positionen mit hohem räumlichen und zeitlichen Informationsgehalt auszuwerten sind, wird in den anschließenden Abschnitten ein Verfahren zur MS-Positionsreduzierung erarbeitet.

Ziel ist es, die Anzahl der Positionen einer MS-Positionsfolge zu reduzieren. Dabei sollen Positionen mit hoher räumlicher und zeitlicher Ähnlichkeit zu Positionsschwerpunkten zusammengefasst werden. In RAMM & SCHWIEGER (2007) werden beispielsweide Verfahren der Kalmanfilterung und der gleitenden Mittelung angewendet, deren Eignung für die vorliegende Anwendung kurz bewertet wird. Für die Mittelung wurde eine Filterlänge von 13 Sekunden für die dort untersuchte Positionsfolge empirisch ermittelt. Die Übertragbarkeit dieses Wertes auf andere als die untersuchte Positionsfolge und insbesondere eine Abhängigkeit von der Positionsgenauigkeit, vgl. das einleitende Beispiel, wird nicht behandelt. Mittels Kalmanfilter konnte die untersuchte Positionsfolge zwar geglättet werden, die Informationsdichte wird mit ihm jedoch nicht erhöht, so dass in einem nachgeschalteten – oder gegebenenfalls zusätzlich zu integrierenden - Selektionsverfahren die angestrebte Punktreduktion vorgenommen werden müsste. Die Übertragbarkeit eines der beiden Verfahren auf die Zielsetzung der Reduktion der Punkte, ist infolgedessen nicht unmittelbar gegeben.

Prinzipiell verfolgen die Verfahren der Clusterbildung unter anderem ebenfalls die Ziele, Ähnlichkeiten zwischen Eingangsdaten aufzudecken, um daraus Gruppen von Eingangsdaten gleicher oder ähnlicher Merkmale zu bilden. Aus diesen Gruppen können im Anschluss beispielsweise Schwerpunkte abgeleitet werden. Unter der Vielzahl der Clusterverfahren[12] wurde für die Lösung der Punktreduzierungsaufgabe das Verfahren des FCM-Clustering ausgewählt und auf die vorliegende Problemstellung angewendet.

4.2.4.2 Grundprinzip des FCM-Clustering

Die Clusterbildung oder auch das **Clustering** ist eine bekannte Methode der numerischen Datenanalyse (ANDERS 2003). Es wird darin versucht, signifikante Merkmale einer Datenmenge $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ zu bestimmen, die typische oder auch untypische Daten, Strukturen oder funktionale Abhängigkeiten sein können

[11] Weitere Abschätzungen für die Zusammenhänge zwischen Geschwindigkeit der MS und der Dauer, ab wann signifikante Positionsänderungen erkannt werden können, werden im nachfolgenden Abschnitt 4.3 gegeben.

[12] Unter anderem finden sich Überblicke und Erläuterungen zu Clusteralgorithmen in BACHER (2002), BERKHIN (2002), HÖPPNER ET AL. (1997), RUNKLER (1999)

(RUNKLER 1999, S. 355). Der Vektor $\mathbf{x}_i = \{x_{1,i} \quad \dots \quad x_{m,i}\}$ besteht aus mehreren Elementen m, die auch als Merkmale bezeichnet werden. Das Auffinden einer empirischen Klassifikation und somit die Bildung von homogenen Gruppen der Daten, die **Cluster** genannt werden, ist letztlich das Ziel aller Clusterbildungsverfahren (BACHER 2002, S. 1). Existieren im untersuchten Datensatz mindestens zwei und maximal $n-1$ unterscheidbare Cluster, ist eine clusteranalytische Auswertung berechtigt. Die Gruppierung der Daten mit Algorithmen der Clusterbildung bewirkt, dass Daten eines Clusters möglichst ähnliche Merkmale (Intra-Cluster-Homogenität) und Daten unterschiedlicher Cluster möglichst unähnliche Merkmale (Inter-Cluster-Heterogenität) aufweisen. Allgemein bleibt anzumerken, dass einer erfolgreichen Clusterbildung aller Clusterverfahren eine geeignete Parameterwahl im statistischen Modell und die geeignete Wahl des Modells zur Beschreibung der Merkmale der Cluster, wie z. B. Form, Größe oder Dichte, zugrunde liegt (BERKHIN 2002).

Im Folgenden steht die Suche nach Punkthäufungen, also die Suche nach Strukturen in den Daten, in deren Nähe sich eine Anzahl von Datenpunkten $\mathbf{x}_k$ mit $k = 1,2,\dots,n$ befindet, im Fokus. Ein klassisches Verfahren hierfür, das erstmalig in RUSPINI (1969) vorgestellt wurde, basiert auf dem **c-Means-Modell**, indem eine Unterteilung (**Partionierung**) der Daten bezüglich Zentroiden $V = \{v_1,\dots,v_n\}$ zu den *c* Clustern vorgenommen wird (MACQUEEN 1967; JAIN & DUBES 1988). Die mittlere euklidische Distanz der Datenpunkte zu den Clusterzentren soll dabei minimal werden. Die Aufzählung der zugehörigen Punkte zu einem Cluster $v_i = \{\mathbf{x}_{i,1},\dots,\mathbf{x}_{i,n_i}\}$ mit $n_i \in \{1,\dots,n-1\}$ und $i = 1,2,\dots,c$ kann mittels der Zugehörigkeit $u_{i,k} \in \{0,1\}$ der entsprechenden Punkte k zum Cluster i spezifiziert werden. Die Zugehörigkeiten werden in der sogenannten Partitionsmatrix **U** zusammengefasst. Insbesondere für die Analyse empirischer Daten führt diese scharfe Zuordnung in der Praxis zu Problemen (RUNKLER 1999, S. 356). Basierend auf der Theorie der unscharfen Mengen von ZADEH kann der Übergang auf eine unscharfe Zuordnung und damit auf eine unscharfe Partitionsmatrix der Form $\mathbf{U} \in [0,1]^{c \times n}$ vollzogen werden, worauf das **Fuzzy c-Means Modell** (FCM) aufbaut (BEZDEK 1981). Werte nahe Null $u_{i,k} \to 0$ bedeuten eine geringe und nahe Eins $u_{i,k} \to 1$ eine hohe Zugehörigkeit. Die Datenpunktzugehörigkeit $u_{i,k} \in [0,1]$ ist damit nicht mehr zwingend scharf, beinhaltet jedoch die Möglichkeit einer scharfen Zuordnung zu einem Cluster. Die aufgrund der Analogie zum normierten Wertebereich in der Wahrscheinlichkeitsrechnung bezeichnete **probabilistische Randbedingung** wird für beide hier besprochenen Varianten gefordert:

$$\sum_{i=1}^{c} u_{i,k} = 1 \quad \forall\, k = 1,2,\dots,n. \tag{4.21}$$

Sie besagt, dass jeder Datenpunkt mindestens einem Cluster, beim klassischen c-Means Modell genau zu einem Cluster, zugeordnet wird. Abweichungen davon werden in nichtprobabilistischen Clustermodellen, wie z. B. dem **possibilistischen c-Means-Modell** (PCM), behandelt (RUNKLER 1999; BACHER 2002). Sowohl für die scharfe als auch für die unscharfe Formulierung wird gefordert, dass jedes Cluster mindestens ein Element besitzt:

$$\sum_{k=1}^{n} u_{i,k} > 0 \quad \forall\, i = 1,2,\dots,c. \tag{4.22}$$

Für die Berechnung des FCM ist, mit Berücksichtigung der genannten Kriterien, die Minimierung der folgenden Kostenfunktion zu lösen (RUNKLER 1999, S. 357 Formel (16.6)):

$$J_{FCM}(\mathbf{U},V;\mathbf{X}) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{i,k}^{m_{fuzzy}} \cdot \left\|\mathbf{x}_k - \boldsymbol{v}_i\right\|^2. \tag{4.23}$$

Der Exponent m_{fuzzy} ist ein vom Benutzer festzulegender Parameter, der als **Fuzzyfizierer** (Verunschärfer) bezeichnet wird. In der Praxis werden häufig Werte von $m_{fuzzy} = 1\dots3,\ m_{fuzzy} \in \Re$ eingesetzt. Die Zugehörigkeit $u_{i,k}^{m_{fuzzy}}$ des k-ten Datenpunktes zum i-ten Cluster (Partition) folgt aus:

$$u_{i,k}^{m_{fuzzy}} = 1 \Big/ \sum_{j=1}^{c} \left(\frac{\left\| \mathbf{x}_k - \boldsymbol{v}_i \right\|}{\left\| \mathbf{x}_k - \boldsymbol{v}_j \right\|} \right)^{\frac{2}{m_{fuzzy}-1}}. \tag{4.24}$$

Mit $\|\bullet\|$ wird die Distanzfunktion zwischen Datenpunkt und Cluster bezeichnet. Im vorliegenden Fall wurde der euklidische Abstand d^E eingesetzt:

$$d^E(\mathbf{x}_k, \boldsymbol{v}_i) = \sqrt{(\mathbf{x}_k - \boldsymbol{v}_i)^T \cdot (\mathbf{x}_k - \boldsymbol{v}_i)}. \tag{4.25}$$

Die Clusterzentren berechnen sich aus dem gewichteten Mittel der Datenpunkte, wobei als Gewicht die in Formel (4.24) ermittelten Zugehörigkeiten genutzt werden:

$$\boldsymbol{v}_i = \frac{1}{\sum_{k=1}^{n} u_{i,k}^{m_{fuzzy}}} \cdot \sum_{k=1}^{n} \left(u_{i,k}^{m_{fuzzy}} \cdot \mathbf{x}_k \right). \tag{4.26}$$

Die Optimierung der Kostenfunktion aus Formel (4.23) wird alternierend gelöst, indem nach der Bestimmung einer neuen Partition mit Formel (4.24) neue Clusterzentren berechnet werden. Anschließend wird der Vorgang iteriert. Weitere Optimierungsmethoden des FCM sind z. B. *Relaxation*, *genetische Algorithmen*, *Reformulierung* oder *Artificial-Life-Methode* (RUNKLER 1999, S. 359; KAMEL & SELIM 1991; BABU & MURTY 1994; HATHAWAY & BEZDEK 1995; RUNKLER & BEZDEK 1997).

Mit der hier verwendeten alternierenden Methode nähern sich die Werte der Clusterzentren sukzessive einem Optimum. Der Prozess wird beendet, wenn die elementweise $\|\bullet\|_{\in}$ berechneten Abweichungen ε_V der Clusterzentren aufeinander folgender Iterationsschritte $\left\| \boldsymbol{V}^{(I)} - \boldsymbol{V}^{(I-1)} \right\| < \varepsilon_V$ genügend gering sind. Unter anderem wegen Rundungseinflüssen oder der Datenpunkteverteilung ist nicht zwingend von einem einzelnen Optimum auszugehen. Zum softwaretechnisch ordnungsgemäßen Abbruch wird daher eine maximale Iterationsanzahl I_{max} festgelegt. Den Ablauf des Iterationsverfahrens zeigt Abbildung 4-7.

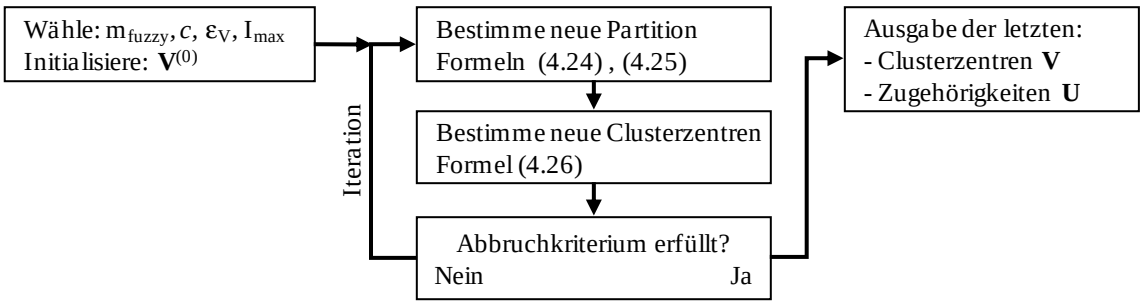


Abbildung 4-7: Ablauf für das Iterationsverfahren des FCM (nach RUNKLER (1999, S. 359 Bild 16.3))

Weisen die Clusterzentren $V^{(0)}$ unterschiedliche Werte auf $\boldsymbol{v}_i^{(0)} \neq \boldsymbol{v}_j^{(0)} \; \forall \, i, j = 1,2 \ldots c, \; i \neq j$, ist die initialisierende Wahl in den meisten empirisch erfassten Datensätzen als unkritisch anzusehen, gegebenenfalls werden einige wenige Iterationsschritte mehr benötigt. Beispielsweise bei symmetrisch angeordneten Punkten ist es jedoch möglich, dass die Clusterbildung aufgrund einer ungünstigen Initialisierung der Clusterzentren nicht möglich ist. RUNKLER (1999) empfiehlt daher eine den Daten angepasste Initialisierung der Clusterzentren.

4.2.4.3 Prinzip des GK-FCM-Clustering

Mit der Distanzfunktion nach Formel (4.25) des c-Means-Modells wird implizit eine kugelförmige Gestalt der Cluster angenommen (LESOT & KRUSE 2006). In der hier betrachteten Anwendung kann davon jedoch nur ausgegangen werden, wenn keine Bewegung stattfindet. Sobald das MS eine Bewegung vollzieht und mehrere

MS-Positionen zu einem Clusterschwerpunkt zusammengefasst werden sollen, sind kompliziertere Clusterformen erforderlich.

In GUSTAFSON & KESSEL (1979) wird eine Modifikation des FCM vorgeschlagen, wobei die Zugehörigkeit eines Datenpunktes zu einem Cluster mittels der **Mahalanobis Distanz** d^{M} berechnet wird:

$$u_{i,k}^{m_{fuzzy}} = 1 \Bigg/ \sum_{j=1}^{c} \left(\frac{d^{M}\left(\mathbf{x}_k, \boldsymbol{v}_i\right)}{d^{M}\left(\mathbf{x}_k, \boldsymbol{v}_j\right)} \right)^{\frac{2}{m_{fuzzy}-1}} \tag{4.27}$$

$$d^{M}\left(\mathbf{x}_k, \boldsymbol{v}_i\right) = \sqrt{\left(\mathbf{x}_k - \boldsymbol{v}_i\right)^{T} \cdot \mathbf{A}_i^{M} \cdot \left(\mathbf{x}_k - \boldsymbol{v}_i\right)} \quad , k = 1,2 \ldots n. \tag{4.28}$$

In der folgend mit **GK-FCM** (GUSTAFSON & KESSEL-FCM) bezeichneten Clusterbildung, wird mit der Matrix $\mathbf{A}^{M}$ die Clusterform dynamisch, d. h. für jeden Cluster individuell den Daten angepasst. $\mathbf{A}^{M}$ ist symmetrisch, positivdefinit und erlaubt die Berechnung ellipsoider Cluster (HÖPPNER ET AL. 1997, S. 41):

$$\mathbf{A}_i^{M} = \sqrt[p]{\rho_i \cdot \det\left(\mathbf{S}_i^{GK}\right)} * \left(\mathbf{S}_i^{GK}\right)^{-1} \quad , i = 1,2 \ldots c. \tag{4.29}$$

Die dynamische Anpassung der Cluster an die Daten wird mittels der empirischen Kovarianzmatrix $\mathbf{S}^{GK}$ vorgenommen, aus der letztlich $\mathbf{A}^{M}$ folgt. Sind die Größen oder die Größenverhältnisse der Cluster untereinander a priori bekannt, kann man mit den Konstanten ρ_i dieses Vorwissen mit der Forderung $\det\left(\mathbf{A}_i^{M}\right) = \rho_i$ in die Clusterbildung einführen. Für die im Rahmen der vorliegenden Arbeit zu analysierenden Daten existieren keine Vorkenntnisse diesbezüglich, weshalb $\rho_1 = \rho_2 \ldots \rho_c = 1$ gesetzt wird. Die empirische Kovarianzmatrix folgt aus der Gewichtung sowie Normierung der Datenpunkte $\mathbf{x}_k$ und den Clusterzentren $\boldsymbol{v}_i$ mit den Zugehörigkeiten $u_{k,i}$.

$$\mathbf{S}_i^{GK} = \frac{1}{\sum_{k=1}^{n} u_{k,i}^{m_{fuzzy}}} \cdot \sum_{k=1}^{n} u_{k,i}^{m_{fuzzy}} \cdot \left(\mathbf{x}_k - \boldsymbol{v}_i\right) \cdot \left(\mathbf{x}_k - \boldsymbol{v}_i\right)^{T}. \tag{4.30}$$

Im Vergleich zum FCM unterscheidet sich die Berechnung des GK-FCM, und somit die Minimierung der Kostenfunktion $J_{GK\text{-}FCM}\left(\mathbf{U}, \boldsymbol{V}; \mathbf{X}\right) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{i,k}^{m_{fuzzy}} \cdot \left\| \mathbf{x}_k - \boldsymbol{v}_i \right\|^2$, vgl. Formel (4.23), lediglich in der Berechnung der Zugehörigkeitswerte. Den Ablauf des GK-FCM zeigt Abbildung 4-8.

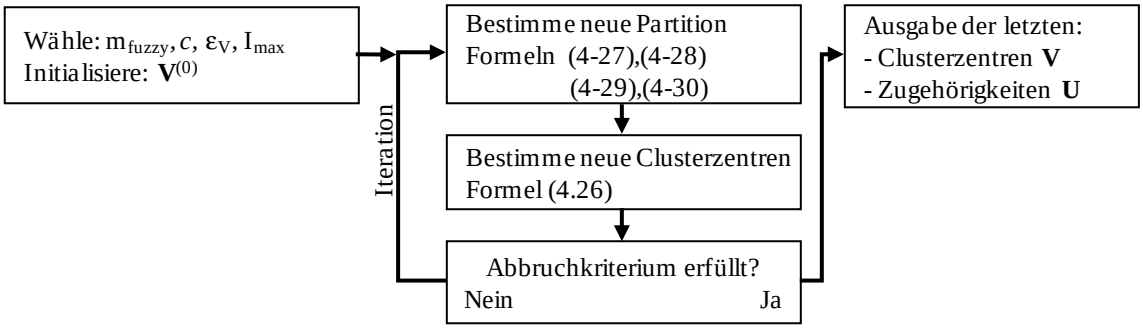


Abbildung 4-8: Ablauf für das Iterationsverfahren des GK-FCM

Das GK-FCM bildet, wie unter anderem das FCM auch, Cluster in Räumen gleicher Metrik. Kennzeichen der MS-Positionen ist hingegen die Ortsangabe im xy-Raum sowie die Zeitangabe im t-Raum. Die MS-Positionsangabe enthält somit Daten mit Merkmalen unterschiedlicher Metriken. Da für die vorliegende Anwendung eine Motivation zur Kombination der Metriken durch die Bildung der empirischen Kovarianzmatrix $\mathbf{S}_i^{GK}$ des Clusterzentrums $\boldsymbol{v}_i$ in Formel (4.30) fehlt, werden diese Kombinationen mittels eines Eingriffes in Formel (4.30) unterdrückt. Die Kovarianzen zwischen den unterschiedlichen Raummetriken werden Null

gesetzt: $s_{xt} = s_{tx} = s_{yt} = s_{ty} = 0$. Die geänderte empirische Kovarianzmatrix $\mathbf{S}_i^{GK_0}$ sowie die Matrix $\mathbf{A}^M$ des Clusterzentrums $\boldsymbol{v}_i$ werden wie folgt besetzt:

$$\mathbf{S}_i^{GK_0} = \left[\begin{array}{c|c|c} s_x^2 & s_{xy} & 0 \\ \hline s_{yx} & s_y^2 & 0 \\ \hline 0 & 0 & s_t^2 \end{array}\right] \quad \text{und} \quad \mathbf{A}_i^M = \left[\begin{array}{c|c|c} A_{xx} & A_{xy} & 0 \\ \hline A_{yx} = A_{xy} & A_{yy} & 0 \\ \hline 0 & 0 & A_{tt} \end{array}\right]. \tag{4.31}$$

Die Mahalanobis-Distanz zwischen dem Datenpunkt $\mathbf{x}_k$ und dem Clusterzentrum $\boldsymbol{v}_i$ wird berechnet mit:

$$\begin{aligned} d^M(\mathbf{x}_k, \boldsymbol{v}_i) &= \sqrt{\Delta x_{k,i}^2 \cdot A_{i,xx} + 2 \cdot \Delta x_{k,i} \cdot \Delta y_{k,i} \cdot A_{i,xy} + \Delta y_{k,i}^2 \cdot A_{i,yy} + \Delta t_{k,i}^2 \cdot A_{i,tt}} \\ \text{mit:} \quad \mathbf{x}_k = \begin{pmatrix} x_{x,k} \\ x_{y,k} \\ x_{t,k} \end{pmatrix}, \quad \boldsymbol{v}_i &= \begin{pmatrix} v_{x,i} \\ v_{y,i} \\ v_{t,i} \end{pmatrix} \quad \text{und} \quad \begin{aligned} \Delta x_{k,i} &= x_{x,k} - v_{x,i} \\ \Delta y_{k,i} &= x_{y,k} - v_{y,i} \\ \Delta t_{k,i} &= x_{t,k} - v_{t,i}. \end{aligned} \end{aligned} \tag{4.32}$$

Mit dem GK-FCM wird eine Reduzierung der Datenpunktanzahl vorgenommen, wobei für die Datenpunkte ein Zugehörigkeitsgrad entsprechend ihrer Lage zum Clusterzentrum berechnet wird. Für die grundlegende Clustergestalt wird im GK-FCM angenommen, dass sie ellipsenförmig ist. Da die Cluster eine Positionsfolge für die Bewegung der MS repräsentieren, ist diese Clustergestalt geeigneter als eine kugelförmige Clustergestalt des FCM anzusehen. Aber, mit dem gewogenen Mittel in Formel (4.26) wird grundsätzlich für eine Menge von Datenpunkten eine Zentrierung auf ein Clusterzentrum berechnet. Dieses Vorgehen widerspricht prinzipiell einer Bewegung.

Für die nachfolgende Bestimmung der Clusterunsicherheiten sind daher neben der Fortpflanzung der Datenpunktunsicherheiten auch die Diskrepanzen zwischen der Bewegung und der Zentrierung auf das jeweilige Clusterzentrum zu berücksichtigen.

4.2.4.4 Zufällige Variabilität der Clusterzentren

Zur Berechnung der zufälligen Variabilität der Clusterzentren wird zum einen zur Berücksichtigung des Bewegungsverhaltens die empirische Varianz-Kovarianzmatrix aus den nach Formel (4.26) gewichteten Datenpunkten bezüglich der Clusterzentren gebildet. Die dafür erforderlichen konstanten Zugehörigkeitswerte $u_{i,k}^{m_{fuzzy}}$ werden in der Funktionalmatrix $\mathbf{F}_i$ zusammengefasst. Die um Eins reduzierte Summe – für das Clusterzentrum als Unbekannte ist ein Freiheitsgrad abzuziehen – dieser Zugehörigkeitswerte, ergibt die Anzahl der Freiheitsgrade $f_{\underline{v}_i}$ des Clusterzentrums $\boldsymbol{v}_i$. Zum anderen sind die Varianz-Kovarianzmatrizen der Datenpunkte $\boldsymbol{\Sigma}_{\underline{\mathbf{MS}},k}$ auf die Clusterzentren fortzupflanzen. Aus der Kombination beider Anteile resultiert die empirische Varianz-Kovarianzmatrix der Clusterzentren:

$$\begin{aligned} \mathbf{S}_{\underline{v}_i \underline{v}_i} &= \frac{1}{f_{\underline{v}_i}} \cdot \underline{\mathbf{v}}_i \cdot \underline{\mathbf{v}}_i^T + \frac{1}{f_{v_i}} \cdot \mathbf{F}_i \cdot \boldsymbol{\Sigma}_{\underline{\mathbf{MS}}^{[\cdot,\cdot]}} \cdot \mathbf{F}_i^T \\ \mathbf{S}_{\underline{v}_i \underline{v}_i} &= \frac{1}{f_{\underline{v}_i}} \cdot \sum_{k=1}^{n} \underline{\mathbf{v}}_{i,k} \cdot \underline{\mathbf{v}}_{i,k}^T + \frac{1}{f_{v_i}} \cdot \mathbf{F}_i \cdot \boldsymbol{\Sigma}_{\underline{\mathbf{MS}}^{[\cdot,\cdot]}} \cdot \mathbf{F}_i^T \\ \mathbf{S}_{\underline{v}_i \underline{v}_i} &= \frac{1}{f_{\underline{v}_i}} \cdot \left(\sum_{k=1}^{n} \underline{\mathbf{v}}_{i,k} \cdot \underline{\mathbf{v}}_{i,k}^T + \sum_{k=1}^{n} u_{i,k}^{m_{fuzzy}} \cdot \boldsymbol{\Sigma}_{\underline{\mathbf{MS}},k} \right) \end{aligned} \qquad \begin{aligned} \text{mit:} \quad \underline{\mathbf{x}}_k &= \underline{\mathbf{MS}}_k, \qquad \underline{\mathbf{v}}_{i,k} = \sqrt{u_{i,k}^{m_{fuzzy}}} \cdot (\boldsymbol{v}_i - \underline{\mathbf{x}}_k), \\ f_{\underline{v}_i} &= \sum_{k=1}^{n} u_{i,k}^{m_{fuzzy}} - 1, \\ \mathbf{F}_i &= \left[\sqrt{u_{i,1}^{m_{fuzzy}}} \;\middle|\; \ldots \;\middle|\; \sqrt{u_{i,n}^{m_{fuzzy}}} \right], \\ \boldsymbol{\Sigma}_{\underline{\mathbf{MS}}^{[\cdot,\cdot]}} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\underline{\mathbf{MS}},1} & & 0 \\ & \ddots & \\ & & \boldsymbol{\Sigma}_{\underline{\mathbf{MS}},k} \\ 0 & & & \ddots \end{bmatrix}. \end{aligned} \tag{4.33}$$

Das Symbol $\underline{\mathbf{v}}_{i,k}$ kennzeichnet die Verbesserung in x, y und t zwischen Clusterzentrum und Datenpunkt, und $\mathbf{v}_i$ dementsprechend den elementweise zu behandelnden Vektor $\underline{\mathbf{v}}_i = \left[\underline{\mathbf{v}}_{i,1} \quad \underline{\mathbf{v}}_{i,2} \quad \ldots \right]^T$ dieser Verbesserungen. Die Bezeichnung *v* sowie *V* für die Clusterzentren, die sich in der Clustering-Literatur häufig findet, wird im

Folgenden aus Gründen der geeigneteren Darstellung aufgegeben. Stattdessen werden die Clusterzentren und deren Kennwerte der zufälligen Variabilität bezeichnet mit:

$$\underline{\mathbf{MS}} \Leftarrow \boldsymbol{v}, \qquad \mathbf{S}_{\underline{\mathbf{MS}}} \Leftarrow \mathbf{S}_{\underline{\boldsymbol{vv}}}, \quad f_{\underline{\mathbf{MS}}} \Leftarrow f_{\underline{\boldsymbol{v}}}. \tag{4.34}$$

4.2.4.5 Unschärfe der Clusterzentren

Im Falle der Fuzzy-Theorie kann das Bewegungsverhalten als zusätzlich wirkende Unschärfe betrachtet werden. Diese Unschärfe wird in der vorliegenden Arbeit mit Hilfe der Hilfsgröße $\tilde{h}_{i,k}$ aus den Abständen zwischen den Mittelpunkten der Datenpunkte $\tilde{x}_{m,k} = \widetilde{MS}_{m,k}$ und den Clusterzentren $\boldsymbol{v}_i$ als Fuzzy-Zahl abgeschätzt, wobei für den Mittelpunkt $\tilde{h}_{i,k,m} = 0$ und für den Radius $\tilde{h}_{i,k,r} = 0$ gelten. Mit Annahme linearer Referenzfunktionen LL für die Hilfsgröße, legen diese Abstände die Ausdehnung des Trägers und damit die Größe der Spannweiten c_l und c_r der Hilfsgröße fest:

$$\tilde{h}_{i,k} = \left[0, 0, c_l, c_r\right]_{LL} \qquad \text{mit: } c_l = \max\left(0, -\left(\boldsymbol{v}_i - \tilde{x}_{m,k}\right)\right) \text{ und } c_r = \max\left(0, \boldsymbol{v}_i - \tilde{x}_{m,k}\right). \tag{4.35}$$

Den Grad der Auswirkung von $\tilde{h}_{i,k}$ auf die Clusterzentren, gibt der Zugehörigkeitsgrad $u_{i,k}^{m_{fuzzy}}$ aus Formel (4.24) an. Mit Anwendung des Erweiterungsprinzips auf Formel (4.26) werden durch Kombination des Unschärfeanteils aus dem Bewegungsverhalten mit den Unschärfen aus den Datenpunkten selbst die Clusterunschärfen berechnet mit:

$$\tilde{v}_i = \underbrace{\frac{1}{\sum_{k=1}^{n} u_{i,k}^{m_{fuzzy}}} \odot \sum_{k=1}^{n}{}^{\oplus}\left(u_{i,k}^{m_{fuzzy}} \odot \tilde{h}_{i,k}\right)}_{\substack{\text{Unschärfen aufgrund} \\ \text{der Bewegung}}} \oplus \underbrace{\frac{1}{\sum_{k=1}^{n} u_{i,k}^{m_{fuzzy}}} \odot \sum_{k=1}^{n}{}^{\oplus}\left(u_{i,k}^{m_{fuzzy}} \odot \tilde{x}_k\right)}_{\substack{\text{Unschärfen aufgrund} \\ \text{der Datenpunkte}}}$$

$$\tilde{v}_i = \frac{1}{\sum_{k=1}^{n} u_{i,k}^{m_{fuzzy}}} \odot \sum_{k=1}^{n}{}^{\oplus}\left(u_{i,k}^{m_{fuzzy}} \odot \left(\tilde{h}_{k,i} \oplus \tilde{x}_k\right)\right) \tag{4.36}$$

$$\text{mit: } \tilde{x}_k = \widetilde{MS}_k \quad \text{und} \quad \tilde{h}_{i,k} := \begin{bmatrix} \left[0, 0, \max\left(0, -\left(\boldsymbol{v}_{x,i} - \tilde{x}_{x,m,k}\right)\right), \max\left(0, \boldsymbol{v}_{x,i} - \tilde{x}_{x,m,k}\right)\right] \\ \left[0, 0, \max\left(0, -\left(\boldsymbol{v}_{y,i} - \tilde{x}_{y,m,k}\right)\right), \max\left(0, \boldsymbol{v}_{y,i} - \tilde{x}_{y,m,k}\right)\right] \\ \left[0, 0, \max\left(0, -\left(\boldsymbol{v}_{t,i} - \tilde{x}_{t,m,k}\right)\right), \max\left(0, \boldsymbol{v}_{t,i} - \tilde{x}_{t,m,k}\right)\right] \end{bmatrix}_{LL.}$$

Aus Darstellungsgründen werden im Folgenden die Symboliken für die Clusterzentren angepasst:

$$\widetilde{MS} \Leftarrow \tilde{v}, \qquad m_{\widetilde{MS}}(x, y) \Leftarrow m_{\tilde{v}}(x, y). \tag{4.37}$$

4.2.4.6 Fuzzy-Randomness der Clusterzentren

Bei der Modellierung mittels des Fuzzy-Randomness sind sowohl die zufällige Variabilität als auch die Unschärfen der Datenpunkte für die Clusterzentren zu schätzen. Die beiden Unsicherheitsanteile werden für die Varianz-Kovarianzmatrix analog zu Abschnitt 4.2.4.4 sowie für die Unschärfen analog zu Abschnitt 4.2.4.5 auf die Clusterzentren $\underline{\tilde{\boldsymbol{v}}}$ fortgepflanzt:

zufällige Variabilität:

$$\mathbf{S}_{\underline{\tilde{\mathbf{v}}}_i\underline{\tilde{\mathbf{v}}}_i} = \frac{1}{f_{\underline{\tilde{\mathbf{v}}}_i}} \cdot \left(\sum_{k=1}^{n} \underline{\mathbf{v}}_{i,k} \cdot \underline{\mathbf{v}}_{i,k}^{T} + \sum_{k=1}^{n} u_{i,k}^{m_{fuzzy}} \cdot \mathbf{\Sigma}_{\underline{\tilde{\mathbf{x}}}\underline{\tilde{\mathbf{x}}},k} \right)$$

mit : $\underline{\tilde{\mathbf{x}}}_k = \widetilde{\underline{\mathrm{MS}}}_k$

$$\underline{\mathbf{v}}_{i,k} = \sqrt{u_{i,k}^{m_{fuzzy}}} \cdot \left(\boldsymbol{v}_i - \underline{\tilde{\mathrm{x}}}_{m,k} \right),$$

$$f_{\underline{\tilde{\mathbf{v}}}_i} = \sum_{k=1}^{n} u_{i,k}^{m_{fuzzy}} - 1$$

zufälliger Mittelpunkt und Unschärfe:

$$\underline{\tilde{\boldsymbol{v}}}_i = \frac{1}{\sum_{k=1}^{n} u_{i,k}^{m_{fuzzy}}} \odot \sum_{k=1}^{n}{}^{\oplus} \left(u_{i,k}^{m_{fuzzy}} \odot \left(\underline{\tilde{\mathrm{h}}}_{k,i} \oplus \underline{\tilde{\mathrm{x}}}_k \right) \right)$$

mit : $\underline{\tilde{\mathrm{x}}}_k = \widetilde{\underline{\mathrm{MS}}}_k$

$$\underline{\tilde{\mathrm{h}}}_{i,k} := \begin{bmatrix} \left[0,0,\max\left(0,-\left(\boldsymbol{v}_{x,i} - \underline{\tilde{\mathrm{x}}}_{x,m,k}\right)\right), \max\left(0,\boldsymbol{v}_{x,i} - \underline{\tilde{\mathrm{x}}}_{x,m,k}\right) \right] \\ \left[0,0,\max\left(0,-\left(\boldsymbol{v}_{y,i} - \underline{\tilde{\mathrm{x}}}_{y,m,k}\right)\right), \max\left(0,\boldsymbol{v}_{y,i} - \underline{\tilde{\mathrm{x}}}_{y,m,k}\right) \right] \\ \left[0,0,\max\left(0,-\left(\boldsymbol{v}_{t,i} - \underline{\tilde{\mathrm{x}}}_{t,m,k}\right)\right), \max\left(0,\boldsymbol{v}_{t,i} - \underline{\tilde{\mathrm{x}}}_{t,m,k}\right) \right] \end{bmatrix}_{LL} \quad (4.38)$$

Zur geeigneteren Darstellung in den nachfolgenden Abschnitten wird die Symbolik wie folgt angepasst:

$$\widetilde{\underline{\mathrm{MS}}} \Leftarrow \underline{\tilde{\boldsymbol{v}}}, \quad m_{\widetilde{\underline{\mathrm{MS}}}}(\mathrm{x,y,t}) \Leftarrow m_{\underline{\tilde{\boldsymbol{v}}}}(\mathrm{x,y,t}), \quad \mathbf{S}_{\widetilde{\underline{\mathrm{MS}}}} \Leftarrow \mathbf{S}_{\underline{\tilde{\boldsymbol{v}}}\underline{\tilde{\boldsymbol{v}}}}, \quad f_{\widetilde{\underline{\mathrm{MS}}}} \Leftarrow f_{\underline{\tilde{\boldsymbol{v}}}}. \quad (4.39)$$

4.2.4.7 Wahl der Clusteranzahl für das GK-FCM

Eine wesentliche Voraussetzung zur erfolgreichen Clusterbildung ist die a priori Wahl der Clusteranzahl *c*. Die Suche nach der optimalen Clusteranzahl, wenn sie bei empirisch erhobenen Daten überhaupt existiert, ist kein triviales Problem. Aus diesem Grund existiert eine große Anzahl von Lösungsvorschlägen, die auf den unterschiedlichsten Kriterien aufbauen. Im Wesentlichen bauen derartige Kriterien auf Beurteilungen des Clustermodells auf, deren Ausgangspunkt häufig die sogenannte **Streuungsquadratsumme** aus den übrig bleibenden Abweichungen zwischen den Datenpunkten und den Clustern ist (BERKHIN 2002, S. 38ff; BACHER 2002, S. 17ff, S. 308ff). Da jedoch auch die Clusteranzahl selbst das Clustermodell beeinflusst, ist ein Vergleich von Maßzahlen, die aus den Kriterien abgeleiteten wurden, nicht unmittelbar möglich. Vielmehr erfordern solche Vergleiche zahlreiche Versuche mit unterschiedlichen Clusteranzahlen, die rechentechnisch sehr ineffizient sind (BACHER 2002, S. 316ff). Insbesondere bei der Auswertung von großen Datenmengen, wie sie beispielsweise auch in der vorliegenden Arbeit zu behandeln sind, führt eine derartige Suche nach einer geeigneten Clusteranzahl zu inakzeptablen Rechenzeiten.

In der vorliegenden Arbeit wird die Clusteranzahl näherungsweise aus der zeitlichen Gesamtlänge des auszuwertenden Datensatzes $\Delta t_{MS_{xy,1},MS_{xy,n}}$ bestimmt mit:

$$c = \mathrm{int}\left(\frac{\Delta t_{MS_{xy,1},MS_{xy,n}}}{\Delta t_{MS}} + 1 \right), \quad (4.40)$$

wobei mit $\mathrm{int}(\bullet)$ das Runden auf die nächstliegende Ganzzahl symbolisiert wird. Der Parameter Δt_{MS} ist eine im Vorfeld zu bestimmende Konstante für den zeitlichen Mindestabstand zweier aufeinander folgender Clusterzentren, so dass sie sich bei einer – näherungsweise für alle MS angenommenen – durchschnittlichen MS-Geschwindigkeit $\bar{v}_{MS}$, signifikant voneinander unterscheiden. Mit angenommener Normalverteilung und der Standardabweichung σ_{MS} sowie der Sicherheitswahrscheinlichkeit *P* wird Δt_{MS} wie folgt berechnet:

$$\Delta t_{MS} = y_{P=1-\alpha_{Irrtum}} \cdot \sqrt{2 \cdot \sigma_{MS}^2} \cdot \frac{1}{\bar{v}_{MS}}. \quad (4.41)$$

Mit dem Ziel einen für die Verkehrserfassung im urbanen Gebiet repräsentativen Wert für $\bar{v}_{MS}$ näherungsweise zu bestimmen, wurde die Durchschnittsgeschwindigkeit aus einer Auswahl der in Abschnitt 6 verwendeten Versuchsfahrten berechnet. Ausgewählt wurden die Versuchsfahrten für Fußgänger, innerörtliche PKW-Fahrten, Fahrten mit Straßenbahnen sowie S-Bahnen und Regional-Bahnen. Versuchsfahrten auf Autobahnen und außerörtlichen Bundesstraßen wurden nicht berücksichtigt, da dort in der Regel höhere Geschwindigkeiten erzielt werden und somit sich zwei aufeinander folgende Clusterzentren bereits nach kürzen Zeitabständen signifikant voneinander unterscheiden. Die berechnete Durchschnittsgeschwindigkeit betrug ca. $\bar{v}_{MS} = 6{,}1\frac{m}{sec}\ \left(\approx 22\ \frac{km}{h}\right)$. Als Näherung für die Standardabweichung σ_{MS} wurde die in RAMM & SCHWIEGER (2007) ermittelte MS-Ortungsgenauigkeit von ca. 350 m für das Signalpegel-Matching verwendet. Für ein Signifikanzniveau von $P = 68{,}3\% \Rightarrow y_{68{,}3\%} = 0{,}47$ folgte nach Formel (4.41) ein durchschnittlicher zeitlicher Abstand zwischen den Clusterzentren von ca. 38 Sekunden.

Mit dieser Abschätzungsrechnung im Hintergrund wird der Wert für den zeitlichen Clusterabstand in der vorliegenden Arbeit näherungsweise mit $\Delta t_{MS} = 30$ Sekunden definiert. Die daraus berechneten Clusteranzahlen werden unabhängig von der verwendeten Unsicherheitstheorie für alle in Abschnitt 6 vorgenommenen Analysen gleichwertig verwendet. Für den zeitlichen Datenpunktabstand von 480 Millisekunden folgt mit $\Delta t_{MS} = 30$ eine Reduzierung der Datenpunktmenge auf eine Clusteranzahl, die ca. 1% der Datenpunktanzahl entspricht. Beispielsweise wird eine MS-Positionsfolge von 300 Sekunden Dauer durch $c = 11$ Cluster repräsentiert.

4.2.5 Verkehrsnetze und Kacheln digitaler Verkehrsnetze

Einige der in der vorliegenden Arbeit entwickelten Identifizierungsverfahren wenden sogenannte **digitale Verkehrsnetze** an. Die eingesetzten Verkehrsnetze bestehen im Wesentlichen aus den beiden Verkehrsnetzelementen: Verkehrsknoten und Verkehrskanten. Während beiden Verkehrsnetzelementen eine Geometriebeschreibung mittels kartesischer Koordinaten zugeordnet wird, besitzen die Verkehrskanten zusätzlich Informationen über die durchschnittliche Reisezeit, die metrische Länge und die Nutzungsart (z. B. Fußgänger, PKW, Straßenbahn). Seit Verabschiedung des GDF[13] 1.0 Standards im Oktober 1989 werden insbesondere die bestehenden Straßen- und Schienenwege in umfangreichen digitalen Datenbanken erfasst und fortgeführt. Hauptanwendungsgebiete finden diese Daten in Fahrzeugnavigationssystemen sowie allgemein in der Logistik und Reiseplanung. Zur mathematischen Beschreibung von Verkehrsnetzen eignet sich die **Graphentheorie** (DIESTEL 2006; KLEIN 2005). Ein Verkehrsnetz, das nun als **gerichteter Graph** $G = (\mathbf{V}, \mathbf{E})$ bezeichnet wird, besteht aus einer endlichen Menge $\mathbf{V} = \{V_1, V_2, \ldots\}$ Knoten (auch Ecke oder **V**ertex), den Verkehrsnetzknoten, und einer endlichen Menge $\mathbf{E} = \{E_1, E_2, \ldots\}$ Kanten (**E**dge), den Verkehrsnetzkanten. Einer Kante ist ein geordnetes Paar disjunkter Knoten $E = \left(V_{From,i}, V_{To,j}\right) \mid i \neq j$ zugeordnet, wobei V_{From} die Anfangskoten (**From node**) und V_{To} die Endknoten (**To node**) bezeichnen und damit die Befahrbarkeitsrichtung angeben. Aufgrund dieser topologisch gerichteten Bedingung werden real in beide Richtungen befahrbare Wege als zwei getrennte Kanten mit vertauschten Anfangs- und Endknoten verwaltet. Jedem Knoten und jeder Kante ist zudem eine geometrische Beschreibung mit $V = \{x, y\}$ und $E = \{(x, y), (x, y), \ldots\}$ zugeordnet.

Um effizient auf die Verkehrsnetzdaten zugreifen zu können, sind entsprechend des Anwendungsgebietes optimierte Techniken unumgänglich. Im Rahmen der vorliegenden Arbeit werden insbesondere die räumliche Objektsuche (Vorhandensein von Objekten im Umkreis) und die Selektion von Objekten (selektiere Objekte innerhalb eines Polygons) benötigt. Eine typische Technik derartiger effizienter Objektzugriffe ist die

[13] GDF: Geographic Data Files (dt. Geografische Datenbanken), Standard für den Datenaustausch vektorisierter Kartendaten, insbesondere für Verkehrsnetzdaten (z. B. Straßennetz) mit dem Einsatzziel in Navigationssystemen

Organisation der Verkehrsnetzelemente nach räumlich abgegrenzten, regelmäßigen Gebieten, sogenannten **Kacheln**. Die Abbildung 4-9 veranschaulicht ein verkacheltes Verkehrsnetz an einem Beispiel.

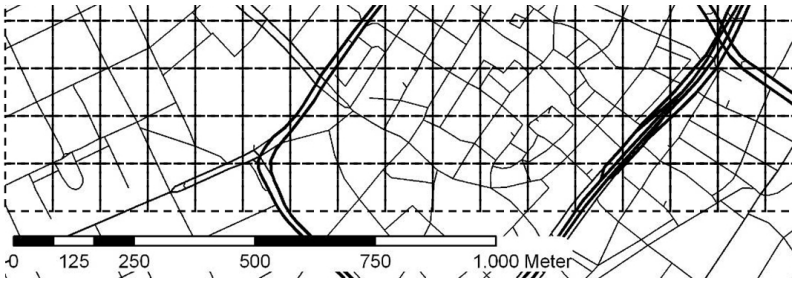


Abbildung 4-9: Ausschnitt eines in einem Kachelfeld organisierten Verkehrsnetzes mit a = 100 m

Mit Definition der Ursprungskoordinate $\left(x_{E_0}, y_{N_0}\right)$ sowie der **Kachelseitenlänge** a kann für eine beliebige Koordinate (x, y) die Kachelkoordinate (E, N) mit den Formeln (4.42) berechnet werden. Die Bezeichnung E steht für die Ostrichtung (***E***ast) und N für die Nordrichtung (***N***orth) in einem kartesischen Koordinatensystem. Mit E und N wird dann eine Kachel $K(E, N)$ im Kachelkoordinatensystem eindeutig beschrieben. Zur Vereinfachung der Datenverwaltung werden die Kachelkoordinaten zu einer impliziten Koordinatenzahl EN (***E***ast***N***orth) zusammengesetzt und in der vorliegenden Arbeit wie folgt berechnet:

$$E = \text{int}\left(\frac{x - x_{E_0}}{a} + 1\right) \quad \text{und} \quad N = \text{int}\left(\frac{y - y_{N_0}}{a} + 1\right)$$
$$EN = E \cdot 100\,000 + N. \tag{4.42}$$

Die Koordinatenzahl EN kann nun eindimensional verwaltet, sortiert und damit effizient in den Datenbeständen gefunden werden. Das Gleiche gilt natürlich für die zugehörigen Kartenobjekte. Die Abbildung 4-10 zeigt einen Ausschnitt der verwendeten Verkehrsnetze in Kacheldarstellung, wobei die Kante `Edge_ID` mit ihren Knoten `FromNode_ID` und `ToNode_ID` die Kachel `Kachel_EN` berührt oder vollständig in ihr liegt.

In_kacheln_edges_e100_n100 (7×3.001.968)

Kachel_EN	Edge_ID	FromNode_ID	ToNode_ID	Length	Form_Of_Way	AverageSpeed
79500360	326482	141345	141457	13,58	30	50
76600390	326483	141583	141584	19,92	30	80
76600390	326484	141584	141583	19,92	30	80

Abbildung 4-10: Kacheldarstellung des Verkehrsnetzes in einer Datenbanktabelle

Aus Formel (4.42) und dem Kachelmittelpunkt $P_{K,m} = \left(x_m = x_{E_0} + a \cdot E,\ y_m = y_{N_0} + a \cdot N\right)$ bezüglich der Ursprungskoordinate $\left(x_{E_0}, y_{N_0}\right)$ folgt unmittelbar die Zugehörigkeitsfunktion der Kachel im xy-Raum. Die Kachel kann danach als ein zweidimensionales Fuzzy-Intervall vom L-Typ geschrieben werden mit:

$$K := \begin{bmatrix} m_x, a, 0, 0 \\ m_y, a, 0, 0 \end{bmatrix}_L \Leftrightarrow m_K(x,y) = \begin{cases} 1, & m_x - a \le x \le m_x + a \ \text{ und } \ m_y - a \le y \le m_y + a \\ 0, & \text{sonst} \end{cases} \quad \forall\, x, y \in \Re^2. \tag{4.43}$$

In der vorliegenden Arbeit wird ausschließlich von scharfen Flanken der Zugehörigkeitsfunktionen der Kacheln ausgegangen, weshalb aus Vereinfachungsgründen, wenn nicht anders erwähnt, auf die Bezeichnung mit $\tilde{K}$ verzichtet wird. Zur geometrischen Beschreibung der Kacheln im zweidimensionalen Raum lassen sich daraus die vier Kacheleckpunkte $\left(P_{K,o,l}, P_{K,o,r}, P_{K,u,l}, P_{K,u,r}\right)$ leicht ableiten, vgl. Abbildung 4-11:

$$\begin{aligned} P_{K,o,l} &= (x_{min},\ y_{max}) \quad P_{K,o,r} = (x_{max},\ y_{max}) \quad \text{, mit:} \quad x_{max} = x_m + 0{,}5 \cdot a \quad y_{max} = y_m + 0{,}5 \cdot a \\ P_{K,u,l} &= (x_{min},\ y_{min}) \quad P_{K,u,r} = (x_{max},\ y_{min}) \qquad\quad x_{min} = x_m - 0{,}5 \cdot a \quad y_{min} = y_m - 0{,}5 \cdot a\,. \end{aligned} \tag{4.44}$$

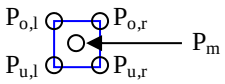


Abbildung 4-11: Kacheleckpunkte im xy-Raum

Kartenobjekte innerhalb des Umringspolygons $\left(P_{K,o,l}, P_{K,o,l}, P_{K,u,l}, P_{K,u,r}\right)$ gehören vollständig zur Kachel. Außerhalb liegende Kartenobjekte haben keine Zugehörigkeit zur Kachel.

Für die Darstellung der Verkehrsnetze im Kachelkoordinatensystem gelten als Grundprimitive der Verkehrsknoten Punkte und der Verkehrskanten Linien. Insofern Verkehrsknoten nicht auf dem Umringspolygon liegen, werden sie eindeutig einer Kachel zugewiesen. Liegen sie auf dem Umringspolygon, sind sie gleichzeitig Teil mehrerer Kacheln. Auf den Kachelseiten sind Verkehrsknoten somit Teil von zwei und auf den Kacheleckpunkten Teil von vier Kacheln. Analog zu Verkehrsknoten gehören vollständig innerhalb des Umringspolygons liegende Verkehrskanten eindeutig zu einer Kachel. Berühren oder scheiden Verkehrskanten das Umringspolygon, sind sie Teil mehrerer Kacheln. Bereits während der Verkachelung der Verkehrsnetze werden daher mehrere Kacheln schneidende Verkehrskanten in Unterkanten, sogenannten **Sub_Kanten** unterteilt, vgl. dazu die Abbildung 4-12. An den Schnittpunkten mit dem Umringspolygon, also den Kachelgrenzen, werden temporäre Verkehrsknoten, sogenannte **Sub_Knoten**, eingefügt. Die Sub_Kanten können nun „scharf“ und damit eindeutig als Element einer Kachel zugeordnet werden. Bis auf den Spezialfall, dass eine Sub_Kante auf dem Kachelrand verläuft, ist jede Sub_Kante nur einer Kachel zugeordnet. Verläuft sie auf dem Rand, ist sie ebenfalls Element der entsprechenden Nachbarkacheln.

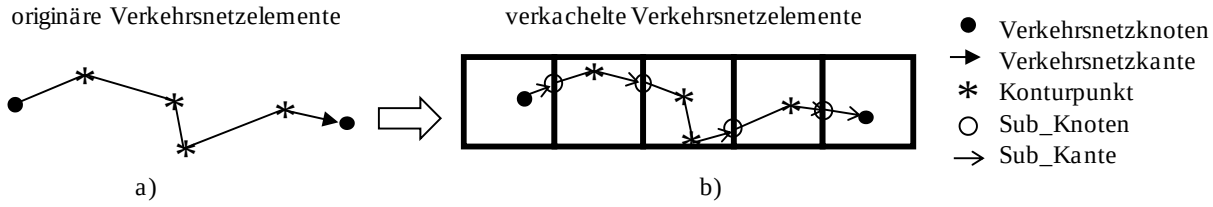


Abbildung 4-12: Verkachelung von Verkehrsnetzen mit Einführung von Sub_Knoten und Sub_Kanten

4.2.6 Unsicherheitsgrade für den Aufenthalt in Kacheln

Nach Einführung der Kacheln im vorherigen Abschnitt 4.2.5 soll nun der Frage nachgegangen werden, wie hoch der Grad des möglichen MS-Aufenthalts in einer Kachel ist. Dieser Grad soll unmittelbar aus den Unsicherheiten der MS-Positionen abgeleitet werden. Für die Kacheln und speziell die Kachelgeometrien wird vereinfacht davon ausgegangen, dass sie keine Unsicherheiten besitzen, wie durch die Definition als scharfe Menge K in Formel (4.43) bereits angedeutet. In den nachfolgenden Abschnitten werden alleinig die Unsicherheiten der MS-Position fortgepflanzt.

4.2.6.1 Zufällige Variabilität der Aufenthaltswahrscheinlichkeit in Kacheln

Für die Wahrscheinlichkeit, dass die MS in der Kachel K angetroffen wird, ist der Wert der Wahrscheinlichkeitsbeziehung $P_{\underline{\mathbf{MS}},\mathrm{K}} = P\left(\boldsymbol{\mu}_{\underline{\mathbf{MS}}} \in \mathrm{K}\right)$ mit $\boldsymbol{\mu}_{\underline{\mathrm{MS}}} = \mathrm{E}\left(\underline{\mathbf{MS}}\right)$ zu berechnen, der im Folgenden als **Aufenthaltswahrscheinlichkeit** bezeichnet wird. Die Aufenthaltswahrscheinlichkeit folgt aus, vgl. zur Definition von $\rho_{\mathrm{Gauß}}\left(\mathrm{x},\mathrm{y}\right)$ Formel (3.6) sowie für $\mathrm{x}_{\min}, \mathrm{x}_{\max}, \mathrm{y}_{\min}$ und $\mathrm{y}_{\max}$ Formel (4.44):

$$\begin{aligned} P\left(\underline{\mathrm{MS}} \in \mathrm{K}\right) &= \iint_{\mathrm{K}} \rho_{\mathrm{Gauß}}\left(\mathrm{x},\mathrm{y}\right)\cdot \mathrm{dx}\cdot \mathrm{dy} \\ &= \int_{\mathrm{x}_{\min}}^{\mathrm{x}_{\max}} \int_{\mathrm{y}_{\min}}^{\mathrm{y}_{\max}} \rho_{\mathrm{Gauß}}\left(\mathrm{x},\mathrm{y}\right)\cdot \mathrm{dx}\cdot \mathrm{dy} = P\left(\mathrm{x}_{\min} \le \underline{\mathrm{MS}}_{\mathrm{x}} \le \mathrm{x}_{\max}, \mathrm{y}_{\min} \le \underline{\mathrm{MS}}_{\mathrm{y}} \le \mathrm{y}_{\max}\right). \end{aligned} \tag{4.45}$$

Aus Gründen der Recheneffizienz wurden nur diejenigen Kacheln in die Berechnung mit einbezogen, die eine minimale Aufenthaltswahrscheinlichkeit von $P_{\underline{\mathbf{MS}} \in \mathrm{K}} \ge 0{,}1\%$ aufweisen können. Eine effiziente Ermittlung dieser Kacheln gelingt mit der Berechnung der Konfidenzellipse für $P_{\underline{\mathrm{MS}} \in \mathrm{K},\min} = 0{,}1\%$ und der anschließenden Berechnung der Aufenthaltswahrscheinlichkeit nur für die Kacheln, die innerhalb dieser Konfidenzellipse liegen, diese schneiden oder sie berühren. Mit Transformation der Varianz-Kovarianzmatrix $\mathbf{S}_{\underline{\mathrm{MS}}}$ in die Ellipsendarstellung $\underline{\mathbf{MS}}, \mathbf{S}_{\underline{\mathrm{MS}}} \rightarrow \mathrm{Ell}_{\underline{\mathrm{MS}}}\left(\mathrm{A}_{\underline{\mathrm{MS}}}, \mathrm{B}_{\underline{\mathrm{MS}}}, \Theta_{\underline{\mathrm{MS}}}\right)$, mit dem Ellipsenmittelpunkt in $\underline{\mathbf{MS}}_{\mathrm{m}}$, folgen die Kenngrößen der **Konfidenzellipse** (vgl. PELZER 1985, S. 161ff):

$$\left.\begin{aligned} \mathrm{A}_{\underline{\mathbf{MS}}} &= \sqrt{\chi^2_{2,P_{\underline{\mathbf{MS}},\mathrm{K},\min}}} \cdot \sqrt{\frac{1}{2}\cdot\left(\mathrm{s}_{\mathrm{x}}^2 + \mathrm{s}_{\mathrm{y}}^2 + w\right)} && \text{große Halbachse, mit: } w_{\underline{\mathbf{MS}}} = \sqrt{\left(\mathrm{s}_{\mathrm{x}}^2 - \mathrm{s}_{\mathrm{y}}^2\right)^2 + 4\cdot \mathrm{s}_{\mathrm{xy}}^2} \\ \mathrm{B}_{\underline{\mathbf{MS}}} &= \sqrt{\chi^2_{2,P_{\underline{\mathbf{MS}},\mathrm{K},\min}}} \cdot \sqrt{\frac{1}{2}\cdot\left(\mathrm{s}_{\mathrm{x}}^2 + \mathrm{s}_{\mathrm{y}}^2 - w\right)} && \text{kleine Halbachse} \\ \Theta_{\underline{\mathbf{MS}}} &= \frac{1}{2}\cdot \arctan\left(2\cdot \mathrm{s}_{\mathrm{xy}} \Big/ \left(\mathrm{s}_{\mathrm{x}}^2 - \mathrm{s}_{\mathrm{y}}^2\right)\right) && \text{Orientierungswinkel von } \mathrm{A}_{\underline{\mathbf{MS}}}. \end{aligned}\right\} \tag{4.46}$$

Als äußere Abgrenzung können nun beispielsweise nach der **Schnittgeradenmethode** (RAUBER 1993) die Kacheln berechnet werden, die die Konfidenzellipse berühren oder scheiden. Die eingeschlossenen Kacheln ergänzen diese Kachelmenge, vgl. Abbildung 4-13a).

Die Abbildung 4-13b) zeigt ein Beispiel für die Aufenthaltswahrscheinlichkeiten der Kacheln bezüglich einer MS-Position. Je dunkler die Kacheln eingefärbt sind, desto höher ist ihr Aufenthaltswahrscheinlichkeitswert. Mit anderen Worten: Je dunkler die Kachelfarbe ist, desto höher ist Wahrscheinlichkeit der Kachel, dass sie als mögliche Position der MS infrage kommt.

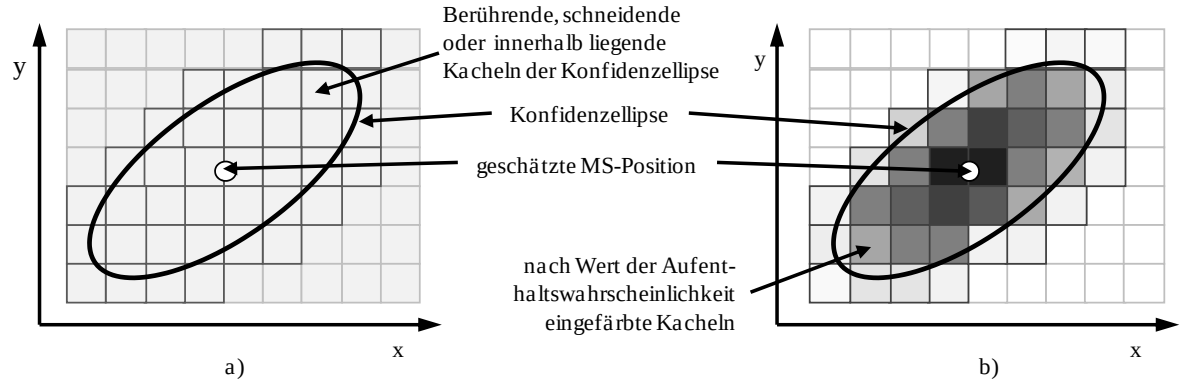


Abbildung 4-13: a) Berührende, schneidende oder innerhalb liegende Kacheln der Konfidenzellipse der geschätzten MS-Position, b) in Abhängigkeit vom Aufenthaltswahrscheinlichkeitswert eingefärbte Kacheln der geschätzten MS-Position

4.2.6.2 Unschärfe des Aufenthalts in Kacheln

Der Zugehörigkeitsgrad eines Punktes $H \in \Re^{(2)}$ zur Fuzzy-Menge $\widetilde{MS}$ ist bereits im Ansatz der Modellierung der Zugehörigkeitsfunktion enthalten, vgl. auch Formel (4.15) in Abschnitt 4.2.3.2. Daraus kann unmittelbar für beliebige Punkte, für die $H \in \left(K \cap \widetilde{MS}\right)$ gilt, mit $m_{K,\widetilde{MS}}(x,y) = m_{\widetilde{MS}}(x,y) \;\left|\forall x,y \in \left(K \cap \widetilde{MS}\right)\right.$ die Zugehörigkeit zu $\widetilde{MS}$ bestimmt werden. Für die als **Aufenthaltszugehörigkeit** $Z_{K \in \widetilde{MS}}$ bezeichneten Grade, sind für die vorliegende Arbeit die maximalen Zugehörigkeiten der Kachel K bezüglich der MS-Position $\widetilde{MS}$ von Interesse:

$$Z_{K \in \widetilde{MS}} = \max\left(m_{K,\widetilde{MS}}(x,y)\right) \quad , \forall x,y \in \left(K \cap \widetilde{MS}\right). \tag{4.47}$$

Die Kacheln, für die $Z_{K \in \widetilde{MS}} > 0$ gilt, bilden die Menge $M_{K \in \widetilde{MS}}$ aller zu $\widetilde{MS}$ gehörigen Kacheln. Für $M_{K \in \widetilde{MS}}$ kommen alle diejenigen Kacheln infrage, deren Schnittmenge mit $\widetilde{MS}$ nicht die leere Menge ist:

$$K \in M_{K \in \widetilde{MS}}, \;\left|\forall K \cap \widetilde{MS} \neq \emptyset.\right. \tag{4.48}$$

Mit den achsparallelen Definitionen der Zugehörigkeitsfunktion $m_{\widetilde{MS}_{xy}}$ aus Abschnitt 4.2.3.2 und der Kachelgeometrie in Abschnitt 4.2.5, vereinfacht sich die Berechnung von $Z_{K \in \widetilde{MS}_{xy}}$. Das Zugehörigkeitsmaximum ist in demjenigen Kachelpunkt $H \in K$ zu finden, der am dichtesten am Kern von $\widetilde{MS}$ liegt. Hierbei kann sich auf die Lotfußpunktberechnung ausgehend von den Eckpunkten des Kerns auf den Kachelrand sowie von den Kacheleckpunkten auf den Rand des Kerns von $\widetilde{MS}$ beschränkt werden. Mit Einsetzen des entsprechend dichtesten Kachelpunktes H in $m_{\widetilde{MS}}(x,y)$ wird das Zugehörigkeitsmaximum $Z_{K \in \widetilde{MS}} := m_{\widetilde{MS}}(H)$ der Kachel berechnet.

4.2.6.3 Fuzzy-Randomness des Aufenthalts in Kacheln

Im Falle des Fuzzy-Randomness sind die Anteile der zufälligen Variabilität und der Unschärfe der MS-Position für die Bestimmung der Aufenthaltswahrscheinlichkeit in der Kachel K zu berücksichtigen. Zur Berechnung des Aufenthaltswahrscheinlichkeitswertes seien für einen Moment die Unschärfen ausgeblendet. In diesem Fall entspricht der gesuchte Lösungsweg dem bereits in Abschnitt 4.2.6.1 beschriebenen Ablauf, worin aus der Wahrscheinlichkeitsbeziehung $P(\underline{MS} \in K)$ der Wert der Aufenthaltswahrscheinlichkeit berechnet wurde. Werden nun die Unschärfen wieder hinzugenommen, dann ist der gesuchte Wahrscheinlichkeitswert durch Fuzzy-Auswertung von $P\left(\underline{\widetilde{MS}} \in K\right)$ zu berechnen. Das Ergebnis wird im Allgemeinen eine Fuzzy-Menge $\tilde{P}_{\underline{\widetilde{MS}} \in K}$, also eine unscharfe Aufenthaltswahrscheinlichkeit sein. In der vorliegenden Arbeit werden ausschließlich die Maximalwerte (rechte Seite der Fuzzy-Menge) dieser unscharfen Aufenthaltswahrscheinlichkeit benötigt, weshalb im Folgenden ausschließlich auf deren Bestimmung eingegangen wird.

Mittels α-schnittweiser Auswertung der Wahrscheinlichkeitsbeziehung $\tilde{P}_{\underline{\widetilde{MS}} \in K,\alpha} = P\left(\tilde{\mu}_{\underline{\widetilde{MS}},\alpha} \in K\right)$ ist die Wahrscheinlichkeit dafür zu berechnen, dass der Erwartungswert $\tilde{\mu}_{\underline{\widetilde{MS}},\alpha}$ der α-Schnittfigur von $\underline{\widetilde{MS}}_{\alpha}$ Bestandteil der Kachel K ist. Mit Übergang auf die endlich abgeschlossenen Intervalle der unscharfen Wahrscheinlichkeit $\tilde{P}_{\underline{\widetilde{MS}} \in K,\alpha} := m_{\tilde{P},\alpha}(p) = \left[P_{\alpha,\min}, P_{\alpha,\max}\right], \forall \alpha \in [0,1]$ folgt prinzipiell:

$$\tilde{P}_{\underline{\widetilde{MS}} \in K,\alpha} := \left[P_{\alpha,\min} = P_{\min}\left(\tilde{\mu}_{\underline{\widetilde{MS}},\alpha}(x,y) \in K\right), \quad P_{\alpha,\max} = P_{\max}\left(\tilde{\mu}_{\underline{\widetilde{MS}},\alpha}(x,y) \in K\right)\right], \; \forall (x,y) \in \Re^2. \tag{4.49}$$

Diese Problemstellung lässt sich auf eine Nullstellensuche und damit beispielsweise auf Näherungsverfahren nach dem Bisektionsverfahren (JAULIN ET AL. 2001) oder iterativen Newton-Verfahren zurückführen. Da im vorliegenden Fall ausschließlich die maximale Aufenthaltswahrscheinlichkeit von Interesse ist, kann die Suche auf denjenigen Punkt der α-Schnittfigur von $\underline{\widetilde{MS}}_{\alpha}$ begrenzt werden, in dem der Erwartungswert μ der Normalverteilung zu legen ist, um $P_{\alpha,\max}$ für die Aufenthaltswahrscheinlichkeit in der Kachel zu erhalten. Zur Lösung sind die folgenden beiden Fälle zu unterscheiden:

1. $\widetilde{\underline{\mathbf{MS}}}_\alpha \cap \mathrm{K} = \varnothing$: Die α-Schnittfigur $\widetilde{\underline{\mathbf{MS}}}_\alpha$ und die Kachel haben keine gemeinsame Schnittmenge. $P_{\alpha,\max}$ kann in diesem Fall für einen Erwartungswert μ_i gefunden werden, bei dem sich μ_i auf dem Rand der α-Schnittfigur von $\widetilde{\underline{\mathbf{MS}}}_\alpha$ befindet, vgl. die Abbildung 4-14a). Aufgrund der geometrischen Definition der Zugehörigkeitsfunktionen von $\widetilde{\underline{\mathbf{MS}}}$ sind die Randseiten der Figur von $\widetilde{\underline{\mathbf{MS}}}_\alpha$ achsparallel, was die Auswahl der Hilfspunkte und damit die Suche weiter vereinfacht. Zum Auffinden dieses Maximums eignen sich iterative numerische Verfahren nach dem **Sekantenverfahren** (SCHWARZ & KÖCKLER 2011, S. 195; HERMANN 2011, S. 210). Entlang des Randes der α-Schnittfigur von $\widetilde{\underline{\mathbf{MS}}}_\alpha$ sind als Näherungen die Aufenthaltswahrscheinlichkeiten $P_{\mu_i} = P(\mu_i \in \mathrm{K})$ für die Hilfspunkte μ_i mit $i = 1, 2 \ldots n$ zu berechnen. Zur Berechnung von $P(\mu \in \mathrm{K})$ sei auf Abschnitt 4.2.6.1 verwiesen. Iterativ werden nun solange weitere Hilfspunkte eingeführt, bis das Abbruchkriterium ε der Näherung erreicht ist.

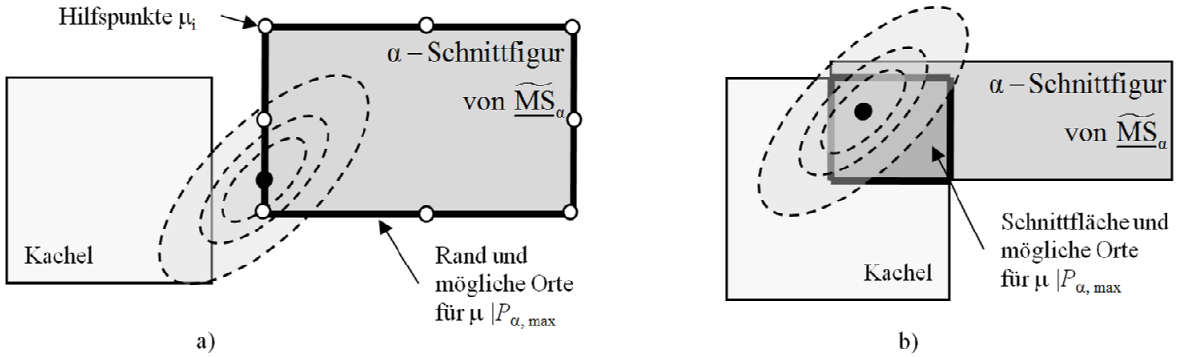


Abbildung 4-14: Fallunterscheidung für die Suche nach der maximalen Aufenthaltswahrscheinlichkeit
a) ohne gemeinsamer Schnittmenge und b) mit gemeinsamer Schnittmenge von Kachel und α-Schnittfigur

2. $\widetilde{\underline{\mathbf{MS}}}_\alpha \cap \mathrm{K} \neq \varnothing$: Die α-Schnittfigur $\widetilde{\underline{\mathbf{MS}}}_\alpha$ und die Kachel haben eine gemeinsame Schnittmenge. In diesem Fall kann $P_{\alpha,\max}$ durch Wahl eines Erwartungswertes μ_i gefunden werden, der sich innerhalb der Schnittfläche beider Figuren befindet $\mu_i \in \left(\tilde{\mu}_{\widetilde{\underline{\mathbf{MS}}},\alpha} \cap \mathrm{K}\right)$, vgl. die hervorgehobene Schnittfläche der beiden Figuren in Abbildung 4-14b). Aufgrund der achsparallelen Definitionen der Kacheln und der Unschärfen von $\widetilde{\underline{\mathbf{MS}}}$ haben die Schnittmengen ebenfalls eine achsparallele Figur, was die Suche vereinfacht. Prinzipiell lässt sich $P_{\alpha,\max}$ analog zum oben genannten iterativen Näherungsverfahren bestimmen, wobei eine (rechentechnisch nicht unerhebliche) Erweiterung auf die zweidimensionale Suche, z. B. auf ein **Mehrdimensionales Sekantenverfahren** oder auf ein **Newton-Verfahren im mehrdimensionalen Fall** (SCHWARZ & KÖCKLER 2011, S. 200f) mit numerischer Bestimmung der Ableitungen (Jacobi-Matrix), vorzunehmen ist. Nach jedem Iterationsschritt ist zu prüfen, ob das Abbruchkriterium ε für Annäherung an $P_{\alpha,\max}$ erreicht wurde.

Für die Auswertungen der vorliegenden Arbeit wurde als Abbruchkriterium in beiden Fällen $\varepsilon = 0{,}1\%$ für die Annäherung an $P_{\alpha,\max}$ festgesetzt. Abschließend kann mit Zusammenfassen aller $P_{\alpha,\max}$ die rechte Seite der Zugehörigkeitsfunktion für die unscharfe Aufenthaltswahrscheinlichkeit konstruiert werden mit:

$$\tilde{P}_{\underline{\mathbf{MS}} \in \mathrm{K},\max} = \sup_{\alpha \in [0,1]} \left[P_{\alpha,\max} \left(\widetilde{\mathbf{MS}}_\alpha \in \mathrm{K} \right) \right]. \tag{4.50}$$

4.3 Bewegungsidentifikation

Nach der Bereitstellung der für die Identifikationsverfahren wichtigsten Berechnungsgrundlagen, wird im aktuellen Abschnitt der Frage nach der Bewegung des MS nachgegangen. Messunsicherheiten bewirken durch die Mobilfunkortung scheinbare Positionsänderungen. Sind Art und Größe der auftretenden Unsicherheiten bekannt, liegt es nahe, geeignete Plausibilitätsprüfungen auf die Daten anzuwenden. In einem ersten Schritt soll eine qualitative Aussage in Form von: „MS steht“ oder „MS in Bewegung“ ermittelt werden. Es ist nicht das Ziel ein Bewegungsprofil entlang des Datensatzes zu erstellen.

Zur Lösung dieser Problemstellung können zwei wesentliche Modellansätze unterschieden werden: Zum einen die Annahme, dass die MS steht und zum anderen, dass es sich in Bewegung befindet. Geht man vom Stand der MS aus, können beispielsweise die räumlichen Abweichungen der MS-Positionsfolge bezüglich eines einzelnen Schwerpunktes (angenommener Standpunkt) untersucht werden, vgl. Abbildung 4-15a).

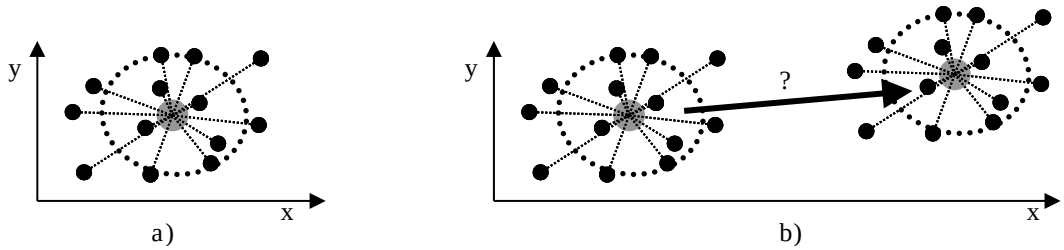


Abbildung 4-15: Annahme a) eines einzelnen Schwerpunktes für die MS-Positionen und b) einer Bewegung

Im Falle zufälliger Variabilität und Fuzzy-Randomness führt dies zu einem Ausgleichungsproblem, das z. B. mit einem GAUß-MARKOV-MODELL (HÖPCKE 1980; KOCH 1997; NIEMEIER 2002; WELSCH ET AL. 2000) beziehungsweise mit einem GAUß-MARKOV-MODELL BEI UNSCHÄRFE (KUTTERER 2002; NEUMANN 2009) gelöst werden kann. Ob die Modellannahme eines einzelnen Schwerpunktes gegenüber den MS-Positionen vereinbar ist, kann mit dem Globaltest der Ausgleichung geprüft werden. Werden die MS-Positionen als Fuzzy-Mengen beschrieben, ist die Existenz nicht-leerer Schnittmengen zwischen den MS-Positionen zu prüfen.

Die Modellannahme, dass die MS steht, wird für die Bewegungsidentifikation im Rahmen der vorliegenden Arbeit nicht näher vertieft. Stattdessen soll angenommen werden, dass sich die MS in Bewegung befindet. Wesentliches Merkmal einer Bewegung ist die Positionsänderung und der daraus resultierende zurückgelegte Weg, vgl. Abbildung 4-15b). Prinzipiell kann angenommen werden: Je größer der Weg im Verhältnis zu den wirkenden Messunsicherheiten ist, umso sicherer kann eine Bewegung schlussgefolgert werden.

Mit Annahme zufälliger Variabilität der MS-Postionen konnte gezeigt werden, dass aus der Bewegung resultierende Distanzen nur signifikant werden können, wenn hohe Bewegungsgeschwindigkeiten der MS vorliegen oder MS über lange Zeiträume beobachtet werden (SCHOLLMEYER 2008). Beispielsweise kann bei typischen Genauigkeiten der Mobilfunkortung von ca. 350 m (RAMM & SCHWIEGER 2007) und einer Messdatenlänge von mindestens 60 Sekunden, ein MS mit einer tatsächlichen Geschwindigkeit von mindestens 50 km/h mit einer Sicherheitswahrscheinlichkeit von $P = 95\%$ als „MS in Bewegung“ erkannt werden. Die Abbildung 4-16 veranschaulicht die Abhängigkeiten zwischen MS-Bewegungsgeschwindigkeit und der Zeit ab der eine Bewegung signifikant erkannt werden kann, für verschiedene Sicherheitswahrscheinlichkeiten. Ist die Bewegungsgeschwindigkeit des MS ca. 20 km/h, erhöht sich bei gleichen Rahmenbedingungen die notwendige Messdatenlänge auf über 150 Sekunden. Dieser Berechnung liegt die Annahme einer geradlinigen Bewegung zu Grunde.

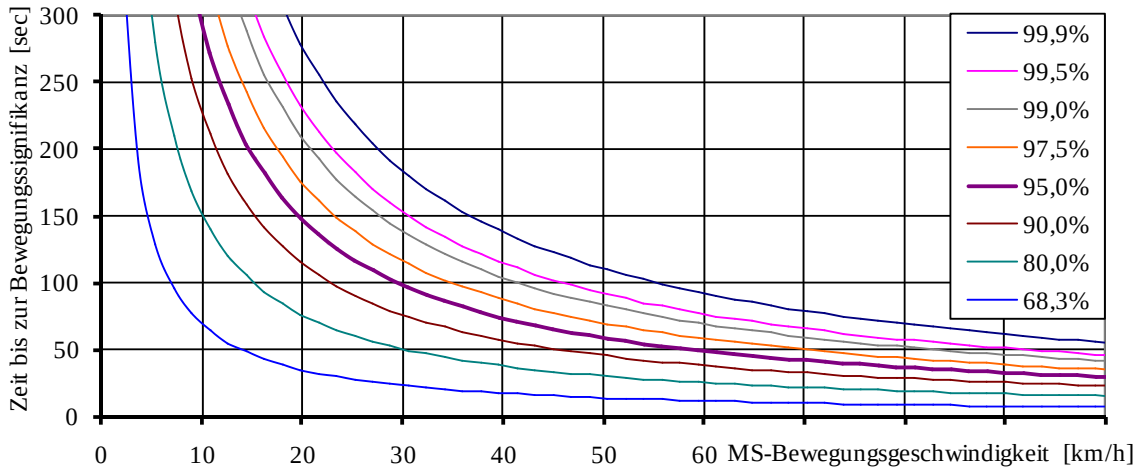


Abbildung 4-16: Abhängigkeit zwischen MS-Bewegungsgeschwindigkeit und der Zeit, ab der eine Bewegung signifikant wird, für verschiedene Sicherheitswahrscheinlichkeiten bei MS-Positionsgenauigkeiten von 350 m

Bei Kurvenfahrten oder Wendemanövern sind die räumlichen Distanzen zeitlich aufeinander folgender Positionen geringer, infolgedessen werden die Distanzen – bei sonst gleichen Rahmenbedingungen - erst bei höheren Messdatenlängen signifikant. Um derartige Fahrtrichtungsänderungen zu berücksichtigen, müssten beispielsweise paarweise alle Punktkombinationen der MS-Positionen auf Signifikanz untersucht werden. Empirische Untersuchungen zeigten jedoch, dass auch für kürzere Messdatenlängen „mit dem Auge“ ein Bewegungsverhalten erkennbar war. Zudem zeigte sich beim Vergleich mit Soll-Daten, dass lokale, größere Positionsabweichungen zwischen zeitlich benachbarten MS-Positionen häufig unrichtig ermittelte signifikante Distanzen verursachten. Unrichtige Identifizierungen von „MS in Bewegung“ waren die Folge. Messreihen übergreifende Signifikanzschwellen konnten nicht unmittelbar gefunden werden, so dass auch allgemeingültige Verfahren zur (statistischen) Ausreißersuche nicht zum Erfolg führten.

Zur Berücksichtigung von Fahrtrichtungsänderungen und einer möglichen Kompensation lokaler, größerer Positionsabweichungen wird in SCHOLLMEYER (2008) die Auswertung auf Basis von Positionsschwerpunkten, also die Clusterung der MS-Positionsfolge vorgeschlagen. Diese Vorgehensweise wird im Folgenden aufgegriffen.

4.3.1 Szenarien der Bewegung

Das auf die Daten angewendete Modell bezieht bereits Annahmen über deren Merkmale mit ein. Wird nur ein Cluster angenommen, wird implizit vom Stand ausgegangen. Werden mehrere Cluster angenommen, wird implizit eine Bewegung vorausgesetzt. Man kann daraus schlussfolgern, dass eine höhere Clusteranzahl auch die Schlussfolgerung auf eine Bewegung erhöht. Ob überhaupt eine Bewegung stattfand, ist jedoch Fragestellung der hier durchzuführenden Bewegungsidentifikation. Als entgegengesetztes Extremum zum Stand mit einem einzelnen Cluster, kann die Bewegung gesehen werden, in der jede der n MS-Positionen Träger der Bewegungsinformation und damit gleichzeitig Clusterschwerpunkt ist. Daraus kann ein Prüfverfahren entworfen werden, bei dem für jede angenommene Clusteranzahl c mit $1 \leq c \leq n$ die Bewegung nachzuweisen ist. Je größer die Clusteranzahl jedoch wird, desto rechenintensiver ist die Clusterbildung, weshalb die Rechenzeiten für ein derartiges Prüfverfahren unverhältnismäßig hoch sind. Um sich einer möglichen Lösung zu nähern, sei daher die hier zu analysierende Geometrieebene näher betrachtet.

Die Bewegung soll für zeitlich aufeinanderfolgende Positionen im zwei-dimensionalen kartesischen Raum geprüft werden. Für geradlinige Bewegungen genügen prinzipiell zwei Positionen zur Repräsentation der Bewegung, vgl. in Abbildung 4-17 das Szenario „**geradlinig**“.

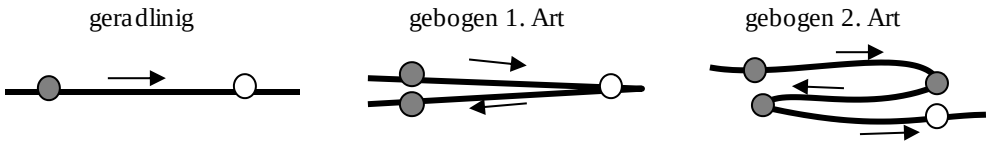


Abbildung 4-17: Szenarienformen zur Erkennung von Bewegungen nach SCHOLLMEYER (2008)

Für Bewegungen mit Hin- und Rückweg ist eine weitere Position erforderlich. In Abbildung 4-17 ist dieses Szenario mit „**gebogen 1. Art**“ gekennzeichnet. Bei Verwendung von nur zwei Positionen, könnten sie räumlich in unmittelbarer Nähe liegen, symbolisiert durch die beiden grauen Punkte, und dadurch eine mögliche Bewegung verbergen. Bei einer angenommen Bewegung mit zweifacher Richtungsänderung des Weges, wie es in Abbildung 4-17 im Szenario „**gebogen 2. Art**“ dargestellt ist, sind bereits mindestens vier Positionen erforderlich. Kompliziertere Bewegungsformen, bei denen in die Nähe früherer Aufenthaltsorte zurückgekehrt wird oder Wege gekreuzt werden, erfordern weitere Positionen. Die vorliegende Arbeit beschränkt sich jedoch auf die drei genannten Szenarien, die im Folgenden mit dem Index $\mathrm{Sz} \in \{1,2,3\}$ bezeichnet werden mit: 1=„geradlinig“, 2=„gekrümmt 1.Art“ und 3=„gekrümmt 2.Art“. Kann der Unterschied von mindestens zwei Positionen unter Berücksichtigung ihrer Unsicherheiten aus mindestens einem Szenario nachgewiesen werden, dann ist die Hypothese „MS steht“ zu verwerfen. Den Unsicherheitsmodellierungsarten entsprechende Prüfverfahren werden dafür in den folgenden Abschnitten beschrieben.

4.3.2 Bewegungsidentifikation mittels zufälliger Variabilität

Im Fall der zufälligen Variabilität ist zu prüfen, ob zwischen zwei aufeinander folgenden Clustern $\underline{\mathbf{MS}}_{\mathrm{i}}$ und $\underline{\mathbf{MS}}_{\mathrm{i+1}}$ eines Bewegungsszenarios eine signifikante Bewegung der MS nachgewiesen werden kann. Für eine Mindestdistanz $d_{\min}$, die sich die MS signifikant bewegt haben soll, lassen sich die Hypothesen formulieren mit:

$$\begin{aligned} H_0: &\quad \mu_{\Delta\underline{\mathbf{MS}}} = 0 &\rightarrow &\quad \text{„MS steht“} \\ H_{\mathrm{A}}: &\quad \mu_{\Delta\underline{\mathbf{MS}}} > d_{\min} &\rightarrow &\quad \text{„MS in Bewegung“.} \end{aligned} \tag{4.51}$$

Mit Annahme, dass die Komponenten des Zufallsvektors $\underline{\mathbf{MS}}$ mit der Varianz-Kovarianzmatrix $\mathbf{\Sigma}_{\underline{\mathbf{MS}},\underline{\mathbf{MS}}}$ normalverteilt sind, gilt für ihre geschätzten Koordinatendifferenzen:

$$\Delta\underline{\mathbf{MS}} \sim N\left(\underline{\boldsymbol{\mu}}_{\Delta\underline{\mathbf{MS}}}, \mathbf{\Sigma}_{\Delta\underline{\mathbf{MS}},\Delta\underline{\mathbf{MS}}}\right) \quad \text{mit}: \Delta\underline{\mathbf{MS}} = \underline{\mathbf{MS}}_{\mathrm{i+1}} - \underline{\mathbf{MS}}_{\mathrm{i}}, \tag{4.52}$$

sowie für ihre quadratische Form:

$$q = \left(\Delta\underline{\mathbf{MS}} - \boldsymbol{\mu}_{\Delta\underline{\mathbf{MS}}}\right)^{\mathrm{T}} \cdot \mathbf{\Sigma}^{-1}_{\Delta\underline{\mathbf{MS}},\Delta\underline{\mathbf{MS}}} \cdot \left(\Delta\underline{\mathbf{MS}} - \boldsymbol{\mu}_{\Delta\underline{\mathbf{MS}}}\right) \sim \underline{\chi}^2_f, \tag{4.53}$$

die der $\underline{\chi}^2$-Verteilung folgt mit $f = \mathrm{rang}\left(\mathbf{\Sigma}_{\Delta\underline{\mathbf{MS}},\Delta\underline{\mathbf{MS}}}\right)$ Freiheitsgraden. Für die vorliegende Untersuchung im xy-Raum ist $f = 2$. Für den Fall der Bewegung, also der Annahme der Alternativhypothese H_{A}, gilt zwischen der Testgröße $\underline{\chi}^2_f$ und dem Quantil $\underline{\chi}^2_{f,1-\alpha_{\mathrm{Irrtum}}}$ die Beziehung:

$$\underline{\chi}^2_f \overset{H_{\mathrm{A}}}{>} \underline{\chi}^2_{f,1-\alpha_{\mathrm{Irrtum}}}. \tag{4.54}$$

Die Wahl eines geeigneten Wertes für α_{Irrtum} ist prinzipiell mit dem Zutreffen der Nullhypothese H_0 und daher mit der Erkennung einer stehenden MS gekoppelt. Zur Erkennung der Bewegung ist hingegen das Zutreffen der Alternativhypothese zu prüfen (WIESER 2012):

$$P\left\{\underline{\chi}_f^2 > \underline{\chi}_{f,1-\alpha}^2 \middle| H_\text{A}\right\} = \gamma_{\min}, \tag{4.55}$$

worin $\gamma_{\min}$ die Testgüte ist, die die Wahrscheinlichkeit für das Verwerfen der Nullhypothese im Fall des Zutreffens der Alternativhypothese, also der Bewegung der MS angibt. Somit können Bewegungen gegenüber einer definierten Mindestdistanz $d_{\min}$ und einer Mindestwahrscheinlichkeit $\gamma_{\min}$ erkannt werden. Die Teststatistik im Falle H_A folgt einer nichtzentralen $\underline{\chi}_{\lambda,f}^2$-Verteilung, worin λ der Nichtzentralitätsparameter ist:

$$\lambda = \underline{\boldsymbol{\mu}}_{\underline{\Delta\mathbf{MS}}}^\text{T} \cdot \boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}^{-1} \cdot \underline{\boldsymbol{\mu}}_{\underline{\Delta\mathbf{MS}}}. \tag{4.56}$$

Die Hypothesen können nun wie folgt präzisiert werden (nach WIESER 2012 (2.9) bis (2.12)):

$$\begin{aligned}
\underline{\chi}_f^2 &= \Delta\underline{\mathbf{x}}^\text{T} \cdot \boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}^{-1} \cdot \Delta\underline{\mathbf{x}} = \delta\underline{\mathbf{x}}^\text{T} \cdot \boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}^{-1} \cdot \delta\underline{\mathbf{x}} \sim \underline{\chi}_f^2 \,\middle|\, H_0 \\
\underline{\chi}_f^2 &= \Delta\underline{\mathbf{x}}^\text{T} \cdot \boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}^{-1} \cdot \Delta\underline{\mathbf{x}} \\
&= \delta\underline{\mathbf{x}}^\text{T} \cdot \boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}^{-1} \cdot \delta\underline{\mathbf{x}} + 2 \cdot \delta\underline{\mathbf{x}}^\text{T} \cdot \boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}^{-1} \cdot \underline{\boldsymbol{\mu}}_{\underline{\Delta\mathbf{MS}}} + \underline{\boldsymbol{\mu}}_{\underline{\Delta\mathbf{MS}}}^\text{T} \cdot \boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}^{-1} \cdot \underline{\boldsymbol{\mu}}_{\underline{\Delta\mathbf{MS}}} \sim \underline{\chi}_{\lambda,f}^2 \,\middle|\, H_\text{A}
\end{aligned} \tag{4.57}$$

$$\text{mit:} \quad \Delta\underline{\mathbf{x}} = \delta\underline{\mathbf{x}} + \underline{\boldsymbol{\mu}}_{\underline{\Delta\mathbf{MS}}} \quad \text{und} \quad \delta\underline{\mathbf{x}} \sim N\left(0, \boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}\right)$$
$$\underline{\boldsymbol{\mu}}_{\underline{\Delta\mathbf{MS}}} = E\left\{\underline{\Delta\mathbf{MS}}\right\}.$$

Mit Minimum $\upsilon_{\min}$ und Maximum $\upsilon_{\max}$ der Eigenwerte von $\boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}$ sowie der Mindestdistanz $d_{\min}$ als Einflussfaktor der Nichtzentralität gelten:

$$\begin{aligned}
\frac{\mu_{\underline{\Delta\mathbf{MS}}}^2}{\upsilon_{\max}} \quad &\leq \quad \underline{\boldsymbol{\mu}}_{\underline{\Delta\mathbf{MS}}}^\text{T} \cdot \boldsymbol{\Sigma}_{\underline{\Delta\mathbf{MS}},\underline{\Delta\mathbf{MS}}}^{-1} \cdot \underline{\boldsymbol{\mu}}_{\underline{\Delta\mathbf{MS}}} \quad \leq \quad \frac{\mu_{\underline{\Delta\mathbf{MS}}}^2}{\upsilon_{\min}} \\
\frac{d_{\min}^2}{\upsilon_{\max}} = \lambda_{\min} \quad &\leq \quad \lambda \quad \leq \quad \lambda_{\max} = \frac{d_{\min}^2}{\upsilon_{\min}}.
\end{aligned} \tag{4.58}$$

Der kritische Wert, gegen den die Testgröße zu vergleichen ist, folgt aus:

$$\underline{\chi}_{f,1-\alpha_{\text{Irrtum}}}^2 = \underline{\chi}_{\lambda_{\min},f,1-\gamma_{\min}}^2. \tag{4.59}$$

Finden Bewegungen von mehr als $d_{\min}$ statt oder erfolgen diese Bewegungen in einer anderen als die durch $\upsilon_{\max}$ beschriebenen Richtung, dann sind die Wahrscheinlichkeiten einer korrekten Bewegungserkennung größer als $\gamma_{\min}$ (WIESER 2012). Derartige Bewegungen sind infolgedessen signifikant.

Die Alternativhypothese H_A und damit die Bewegung des MS ist anzunehmen, wenn die Testgröße $\underline{\chi}_f^2$ größer dem Quantil $\underline{\chi}_{\lambda_{\min},f,1-\gamma_{\min}}^2$ der $\underline{\chi}^2$-Verteilung für die definierte Mindestdistanz $d_{\min}$ und die Mindestbewegungswahrscheinlichkeit $\gamma_{\min}$ ist:

$$\underline{\chi}_{f,\text{Sz,i}}^2 \begin{Bmatrix} \leq \\ \forall \text{i}=1,\ldots,\text{Sz} \\ \text{mit Sz}\in\{1,2,3\} \\ > \\ \exists \text{i}\in 1,\ldots,\text{Sz} \\ \text{mit Sz}\in\{1,2,3\} \end{Bmatrix} \underline{\chi}_{\lambda_{\min},f,1-\gamma_{\min},\text{Sz,i}}^2 \quad \Rightarrow \begin{cases} H_0 \text{ annehmen} & \rightarrow \text{„MS steht“} \\ H_\text{A} \text{ annehmen} & \rightarrow \text{„MS in Bewegung“}. \end{cases} \tag{4.60}$$

Der Test ist für alle Bewegungsszenarien $\text{Sz}\in\{1,2,3\}$ und Koordinatendifferenzen $\Delta\underline{\mathbf{MS}}_{\text{Sz,i}} = \underline{\mathbf{MS}}_{\text{Sz,i+1}} - \underline{\mathbf{MS}}_{\text{Sz,i}}$ mit $\text{i}=1,\ldots,\text{Sz}$ durchzuführen, so dass sich die folgenden Tests ergeben:

$$\text{für}\ \left\{\begin{array}{c|c|l|l|l} \text{Sz} & \text{i} & \text{Tests für } \underline{\mathbf{\Delta MS}}_{\text{Sz,i}} & \text{Testgrößen und Quantile} & \text{Beschreibung} \\ \hline 1 & \{1\} & \underline{\mathbf{\Delta MS}}_{1,1} & \underline{\chi}^2_{f,1,1} \le \underline{\chi}^2_{\lambda_{\min},f,1,1} \,\Big|\, H_{\text{A}} & \text{geradlinig} \\ 2 & \{1,2\} & \underline{\mathbf{\Delta MS}}_{2,1},\ \underline{\mathbf{\Delta MS}}_{2,2} & \underline{\chi}^2_{f,2,1} \le \underline{\chi}^2_{\lambda_{\min},f,2,1} \,\Big|\, H_{\text{A}} & \text{gebogen 1. Art} \\ & & & \vee\ \underline{\chi}^2_{f,2,2} \le \underline{\chi}^2_{\lambda_{\min},f,2,2} & \\ 3 & \{1,2,3\} & \underline{\mathbf{\Delta MS}}_{3,1}, & \underline{\chi}^2_{f,3,1} \le \underline{\chi}^2_{\lambda_{\min},f,3,1} \,\Big|\, H_{\text{A}} & \text{gebogen 2. Art.} \\ & & \underline{\mathbf{\Delta MS}}_{3,2}, & \vee\ \underline{\chi}^2_{f,3,2} \le \underline{\chi}^2_{\lambda_{\min},f,3,2} & \\ & & \underline{\mathbf{\Delta MS}}_{3,3} & \vee\ \underline{\chi}^2_{f,3,3} \le \underline{\chi}^2_{\lambda_{\min},f,3,3} & \end{array}\right. \tag{4.61}$$

Ist für alle Distanzen - für alle sechs Tests - die Nullhypothese H_0 anzunehmen, dann ist „MS steht“ für die untersuchte Positionsfolge anzunehmen. Muss mindestens einmal H_0 verworfen werden, gilt „MS in Bewegung“.

Für die Wertewahl der Mindestdistanz wurde angenommen, dass MS identifiziert werden sollen, die Wege von mindestens $1\,\sigma$ der durchschnittlich erwartbaren Standardabweichung zurücklegen. In Anhang B sind ausgewählte Kennzahlen für die geschätzten Ortungsunsicherheiten zusammengefasst, die aus den der vorliegenden Arbeiten zur Verfügung stehenden Mobilfunkdaten ermittelt wurden. Die Kennzahlen der für die Bewegungsidentifizierung verwendeten geclusterten MS-Positionen, befinden sich in Tabelle B-2. Als Näherung an die Mittelwerte s_x und s_y der zufälligen Variabilität wurde für die Mindestdistanz $d_{\min} = 400\ \text{m}$ festgesetzt. Je nach Anwendung der Bewegungsidentifikation auf andere Ortungsdaten, ist dieser Wert entsprechend anzupassen.

Die Festlegung eines geeigneten Wertes für $\gamma_{\min}$ erfolgt im Rahmen der Kalibrierung in Abschnitt 6.3.1.

4.3.3 Bewegungsidentifikation mittels Unschärfe

Für die Bewegungsidentifikation im Rahmen der Fuzzy-Theorie kann die Problemstellung auf den Nachweis von nichtleeren Schnittmengen aufeinanderfolgender Cluster $\widetilde{\text{MS}}_{\text{Sz,i}}$ und $\widetilde{\text{MS}}_{\text{Sz,i+1}}$ zurück geführt werden. Für eine quantitative Bewertung wird dazu als Maßzahl der Übereinstimmungsindex $\delta_{\text{Sz,i}}$ ermittelt, der sich aus der Schnittmenge der Cluster $\widetilde{\text{MS}}_{\text{Sz,i}}$ und $\widetilde{\text{MS}}_{\text{Sz,i+1}}$ berechnet:

$$\begin{aligned} \delta_{\text{Sz,i}}\left(\widetilde{\text{MS}}_{\text{Sz,i}}, \widetilde{\text{MS}}_{\text{Sz,i+1}}\right) &= \frac{\text{card}\left(\tilde{\text{A}} \cap \tilde{\text{B}}\right)}{\text{card}\left(\tilde{\text{A}}\right)} \\ \text{mit}: \quad & \tilde{\text{A}} = \widetilde{\text{MS}}_{\text{Sz,i}} \ \text{ und } \ \tilde{\text{B}} = \widetilde{\text{MS}}_{\text{Sz,i+1}}, \quad \text{falls } \text{card}\left(\widetilde{\text{MS}}_{\text{Sz,i}}\right) \le \text{card}\left(\widetilde{\text{MS}}_{\text{Sz,i+1}}\right) \\ & \tilde{\text{A}} = \widetilde{\text{MS}}_{\text{Sz,i+1}} \text{ und } \tilde{\text{B}} = \widetilde{\text{MS}}_{\text{Sz,i}}, \quad \text{falls } \text{card}\left(\widetilde{\text{MS}}_{\text{Sz,i}}\right) > \text{card}\left(\widetilde{\text{MS}}_{\text{Sz,i+1}}\right) \\ & \text{i} = 1,\ldots,\text{Sz} \ \text{ und } \ \text{Sz} \in \{1,2,3\}. \end{aligned} \tag{4.62}$$

Der kleinste Übereinstimmungsindex $\delta_{H_0} \in [0,1]$ zwischen allen aufeinanderfolgenden Clustern aller Bewegungsszenarien zeigt den minimalen Möglichkeitsgrad an, dass die Hypothese „MS steht“ anzunehmen ist:

$$\delta_{H_0} = \min_{\substack{\text{i}=1,\ldots,\text{Sz} \\ \text{Sz}\in\{1,2,3\}}} \left(\delta_{\text{Sz,i}}\left(\widetilde{\text{MS}}_{\text{Sz,i}}, \widetilde{\text{MS}}_{\text{Sz,i+1}}\right)\right). \tag{4.63}$$

Je kleiner der Wert von δ_{H_0} ist, desto mehr ist davon auszugehen, dass sich die Cluster unterscheiden. Die abschließende binäre Entscheidung mit $\rho_{\text{Krit}} \in [0,1]$ folgt aus:

$$\delta_{H_0} \left\{\begin{array}{c}\ge\\<\end{array}\right\} \rho_{\text{Krit}} \Rightarrow \begin{cases} H_0 \text{ annehmen} & \rightarrow \ \text{„MS steht“} \\ H_0 \text{ verwerfen} & \rightarrow \ \text{„MS in Bewegung“}. \end{cases} \tag{4.64}$$

Die Wertebestimmung für ρ_{Krit} erfolgt in Abschnitt 6.3.1.

4.3.4 Bewegungsidentifikation mittels Fuzzy-Randomness

Für Bewegungsidentifikation mittels Fuzzy-Randomness ist das in Abschnitt 4.3.2 beschriebene Testverfahren der zufälligen Variabilität um die Anteile der Unschärfe zu erweitern. Dementsprechend lauten die Hypothesen für die Prüfung der Bewegung zwischen zwei MS-Positionen $\widetilde{\underline{\mathbf{\Delta MS}}}$ wie folgt:

$$\begin{aligned} H_0: &\quad \mu_{\widetilde{\underline{\mathbf{\Delta MS}}}} = 0 &&\rightarrow \quad \text{„MS steht“} \\ H_A: &\quad \mu_{\widetilde{\underline{\mathbf{\Delta MS}}}} > d_{\min} &&\rightarrow \quad \text{„MS in Bewegung“.} \end{aligned} \tag{4.65}$$

Mit Definition der Mindestdistanz $d_{\min}$ und der Mindestwahrscheinlichkeit $\gamma_{\min}$ für das Erkennen einer tatsächlichen Bewegung des MS sowie mit dem Maximum der Eigenwerte $\upsilon_{\max}$ von $\mathbf{\Sigma}_{\widetilde{\underline{\mathbf{\Delta MS}}}\widetilde{\underline{\mathbf{\Delta MS}}}}$ folgt das Quantil analog zu (4.59) aus:

$$\underline{\chi}^2_{\lambda_{\min},f,1-\gamma_{\min}} \qquad \text{mit: } \lambda_{\min} = \frac{d^2_{\min}}{\upsilon_{\max}}. \tag{4.66}$$

Prinzipiell ist eine Erweiterung von $d_{\min}$ und $\gamma_{\min}$ auf unscharfe Zahlen möglich, was ein unscharfes Quantil zur Folge hätte. Für die vorliegende Arbeit bestand keine Notwendigkeit einer unscharfen Formulierung, weshalb die Parameter mit scharfen Zahlen definiert wurden.

Aufgrund der Definition der MS-Positionen als FRV, ist die Testgröße als unscharfe Zahl zu bestimmen. Die Koordinatendifferenzen $\widetilde{\underline{\mathbf{\Delta MS}}}$ sind als erweiterte Subtraktion:

$$\widetilde{\underline{\mathbf{\Delta MS}}}_i = \widetilde{\underline{\mathbf{MS}}}_{i+1} \ominus \widetilde{\underline{\mathbf{MS}}}_i \qquad \text{mit: } i = 1,2,... \tag{4.67}$$

und die Varianz-Kovarianzmatrix entsprechend des Varianz-Kovarianzfortpflanzungsgesetzes mit:

$$\mathbf{\Sigma}_{\widetilde{\underline{\mathbf{\Delta MS}}}_i,\widetilde{\underline{\mathbf{\Delta MS}}}_i} = \mathbf{\Sigma}_{\widetilde{\underline{\mathbf{MS}}}_{i+1},\widetilde{\underline{\mathbf{MS}}}_{i+1}} + \mathbf{\Sigma}_{\widetilde{\underline{\mathbf{MS}}}_i,\widetilde{\underline{\mathbf{MS}}}_i} \tag{4.68}$$

zu berechnen. Die Auswertung von (4.53) und damit die Berechnung der unscharfen Testgröße kann α-schnittweise erfolgen, wobei je Figur des α-Schnittes $\widetilde{\underline{\mathbf{\Delta MS}}}_\alpha$ sowohl Minimum und Maximum von der Testgröße $\tilde{\underline{\chi}}^2_{f,\alpha}$ zu berechnen sind:

$$\tilde{\underline{\chi}}^2_{f,\alpha} = \left[\min\left(\widetilde{\underline{\mathbf{\Delta MS}}}^T_\alpha \cdot \mathbf{\Sigma}^{-1}_{\widetilde{\underline{\mathbf{\Delta MS}}},\widetilde{\underline{\mathbf{\Delta MS}}}} \cdot \widetilde{\underline{\mathbf{\Delta MS}}}_\alpha \right), \max\left(\widetilde{\underline{\mathbf{\Delta MS}}}^T_\alpha \cdot \mathbf{\Sigma}^{-1}_{\widetilde{\underline{\mathbf{\Delta MS}}},\widetilde{\underline{\mathbf{\Delta MS}}}} \cdot \widetilde{\underline{\mathbf{\Delta MS}}}_\alpha \right) \right]_\alpha . \tag{4.69}$$

Im Anschluss kann die Testgröße durch Zusammenführen der einzelnen α-Schnitte konstruiert werden mit:

$$\tilde{\underline{\chi}}^2_f = \sup_{\alpha\in[0,1]} \left[\tilde{\underline{\chi}}^2_{f,\alpha} \right]. \tag{4.70}$$

Für die Nullhypothese „MS steht“ wird $\mathrm{E}\left(\underline{\mu} = 0 \,|H_0\right)$ angenommen. Entsprechend gilt für die Alternativhypothese „MS in Bewegung“ $\mathrm{E}\left(\underline{\mu} \neq 0 \,|H_A\right)$. Daraus definieren sich der Annahmebereich mit $\tilde{\mathrm{A}} = \left[0, \underline{\chi}^2_{\lambda_{\min},f,1-\gamma_{\min}}\right]$ sowie der Verwerfungsbereich mit $\tilde{\mathrm{V}} = \tilde{\mathrm{A}}^c = \left(\underline{\chi}^2_{\lambda_{\min},f,1-\gamma_{\min}}, +\infty\right]$.

Zur Testentscheidung wird der Verwerfungsindex $\rho_{\tilde{\mathrm{V}}}\left(\tilde{\underline{\chi}}^2_f\right)$ aus dem Grad der Verwerfung $\gamma_{\tilde{\mathrm{V}}}\left(\tilde{\underline{\chi}}^2_f\right)$ und dem Grad der Annahme $\delta_{\tilde{\mathrm{A}}}\left(\tilde{\underline{\chi}}^2_f\right)$ berechnet:

$$\begin{gathered} \gamma_{\tilde{\mathrm{V}}}\left(\tilde{\underline{\chi}}^2_f\right) = \frac{\mathrm{card}\left(\tilde{\underline{\chi}}^2_f \cap \tilde{\mathrm{V}}\right)}{\mathrm{card}\left(\tilde{\underline{\chi}}^2_f\right)} \qquad \text{und} \qquad \delta_{\tilde{\mathrm{A}}}\left(\tilde{\underline{\chi}}^2_f\right) = 1 - \frac{\mathrm{card}\left(\tilde{\underline{\chi}}^2_f \cap \tilde{\mathrm{A}}\right)}{\mathrm{card}\left(\tilde{\underline{\chi}}^2_f\right)} \\ \rho_{\tilde{\mathrm{V}}}\left(\tilde{\underline{\chi}}^2_f\right) = \min\left(\gamma_{\tilde{\mathrm{V}}}\left(\tilde{\underline{\chi}}^2_f\right), \delta_{\tilde{\mathrm{A}}}\left(\tilde{\underline{\chi}}^2_f\right)\right). \end{gathered} \tag{4.71}$$

Mit Wahl eines geeigneten Wertes für den kritischen Wert ρ_{Krit} des Verwerfungsindexes $\rho_{\mathrm{Krit}} \in [0,1]$, eine mögliche Bestimmung wird in Abschnitt 6.3.1 erläutert, kann die Testentscheidung herbeigeführt werden:

$$\rho_{\tilde{V},Sz,i}\left(\underline{\tilde{\chi}}^2_{f,Sz,i}\right)\begin{Bmatrix}\leq \\ {\scriptstyle \forall i=1,\ldots,Sz \text{ mit } Sz\in\{1,2,3\}} \\ > \\ {\scriptstyle \exists i\in 1,\ldots,Sz \text{ mit } Sz\in\{1,2,3\}}\end{Bmatrix}\rho_{Krit}\in[0,1]\Rightarrow\begin{cases}H_0 \text{ annehmen} & \rightarrow\text{„MS steht“} \\ H_A \text{ annehmen} & \rightarrow\text{„MS in Bewegung“.}\end{cases} \tag{4.72}$$

Diese Prüfung ist für alle Koordinatendifferenzen $\widetilde{\underline{\mathbf{\Delta MS}}}_{Sz,i}$ mit $i=1,\ldots,Sz$ und $Sz\in\{1,2,3\}$ und damit für alle $\underline{\tilde{\chi}}^2_{f,Sz,i}$ durchzuführen. Gilt für alle $\widetilde{\underline{\mathbf{\Delta MS}}}_{Sz,i}$, dass H_0 anzunehmen ist, ist die Einstufung „MS steht“ schlusszufolgern. Ist H_A mindestens einmal anzunehmen, dann gilt „MS in Bewegung“.

Der Wert der Mindestdistanz wurde analog zum Verfahren der zufälligen Variabilität in Abschnitt 4.3.2 sowie aufgrund der näherungsweise gleichgroßen mittleren Standardabweichungen ebenfalls mit $d_{min} = 400\,\text{m}$ festgesetzt.

Die Wertebestimmung für γ_{min} wird in Abschnitt 6.3.1 im Rahmen der Kalibrierung erläutert.

4.4 Identifikation des Verkehrsmittels anhand der Geschwindigkeit

Die Bewegungsgeschwindigkeit einer MS kann auf die Nutzung einer spezifischen Fortbewegungsart und damit auf das genutzte Verkehrsmittel (VM) schließen lassen. In diesem Zusammenhang seien Fußgänger, obgleich sie im eigentlichen Sinne kein VM sind, im Folgenden der Einfachheit halber dennoch zu den VM zugeordnet. Alle betrachten VM haben gemeinsam, dass sie keine Minimalgeschwindigkeiten haben bzw. alle die Geschwindigkeit 0 km/h annehmen können. Ihre **Maximalgeschwindigkeiten** $v_{VM,max}$ sind jedoch verschieden. VM können daher ausgeschlossen werden, wenn die **Geschwindigkeit der MS** v_{MS} höher ist als diejenige des VM.

Hauptziel der nun behandelten Analyse ist es zu prüfen, ob zum einen ein Ausschluss von nicht dem MIV angehörigen VM aufgrund der MS-Geschwindigkeit signifikant ist. Zum anderen sollen diejenigen Mobilfunkdaten herausgefiltert werden, aus denen Geschwindigkeiten resultieren, die nicht mit den Geschwindigkeiten für VM des MIV vereinbar sind, beispielsweise aufgrund von Ortungsausreißern. Da in Deutschland für PKW und Motoräder kein allgemeines Tempolimit existiert (Stand Mai 2013), wurde für die Auswertungen im Rahmen dieser Arbeit eine Maximalgeschwindigkeit für Fahrzeuge des MIV von 250 km/h festgesetzt. Auf Straßen, die keine Autobahnen sind, gelten im Allgemeinen geringere zulässige Höchstgeschwindigkeiten.

Aus den Koordinaten- und Zeitinformationen der MS-Positionen werden die Bewegungsgeschwindigkeiten einschließlich ihrer Unsicherheiten ermittelt. Die Berechnung der Geschwindigkeit im zweidimensionalen kartesischen Raum basiert auf dem zurückgelegten Weg d_i innerhalb einer Zeitspanne dt_i:

$$v_i = \frac{d_i}{dt_i} = \frac{\sqrt{\left(MS_{x,i+1} - MS_{x,i}\right)^2 + \left(MS_{y,i+1} - MS_{y,i}\right)^2}}{t_{MS_{i+1}} - t_{MS_i}}. \tag{4.73}$$

In Abschnitt 4.3 wurde die Bewegung des MS nachgewiesen. Von nun an gilt daher die in Abschnitt 4.2.4.7 vorgenommene Abschätzung der Clusteranzahl für die Berechnung der Cluster MS_i.

Für die empirischen Analysen der vorliegenden Arbeit wurden VM-spezifische **Maximalgeschwindigkeiten** $v_{VM,max}$ festgelegt, die in Tabelle 4-1 zusammengestellt sind. Eine Häufigkeitsverteilung, z. B. wie viele Fahrzeuge einer VM-Art die Maximalgeschwindigkeit statistisch erreichen, wird im Rahmen dieser Arbeit nicht getroffen. Es wird ausschließlich der Grenzfall der Maximalgeschwindigkeit betrachtet, wie sie z. B. aus technischen/biologischen Einschränkungen oder auch aus administrativen Vorgaben folgt, wie z. B. durch den technischen Eingriff mittels Geschwindigkeitsbegrenzern ab einer Höchstgeschwindigkeit von 90 km/h bei LKW mit mehr als 3,5t zulässiger Gesamtmasse (StVZO 2009, §57c). Alle weiteren Maximalgeschwindigkeiten sind für das in Abschnitt 6 näher erläuterte Versuchsgebiet geeignet ausgewählte Schätzwerte.

Tabelle 4-1: Festgesetzte Maximalgeschwindigkeiten für den Ausschluss von Verkehrsmitteln

Verkehrsmittel	**Verkehrsmittelgruppe**	**Verkehrsnetz**	$v_{max,VM}$ **[km/h]**
Fußgänger	NMIV	Fußwege / $G^{Fußweg}$	6 (MID 2002, S. 92)
LKW	MIV	Straße / $G^{Straße}_{LKW}$	90 (StVZO 2009, §57c)
Reisebus		Straße / $G^{Straße}$	100 (StVZO 2009, §57c)
PKW / KRAD			250*
Linienbus	ÖV	Straße / $G^{Straße}$	80*
Straßenbahn		Schiene/ G^{Tram}	70*
Regionalbahn		Schiene / G^{DB}	140*
Fernbahn			300*

(* Schätzwerte für Versuchsgebiet)

4.4.1 Zufällige Variabilität des Verkehrsmittel-Ausschlusses

Im stochastischen Fall wird die Varianz der Geschwindigkeit $s^2_{\underline{v}_i}$ aus der Varianz der Strecke $s^2_{\underline{d}_i}$ und der Varianz der Zeitdifferenz $s^2_{\underline{dt}_i}$ abgeleitet:

$$\begin{aligned} s^2_{\underline{d}_i} &= \mathbf{F}\cdot\mathbf{S}_{\underline{\mathbf{MS}}_i,\underline{\mathbf{MS}}_{i+1}}\cdot\mathbf{F}^T \quad \text{mit:} \quad \mathbf{S}_{\underline{\mathbf{MS}}_i,\underline{\mathbf{MS}}_{i+1}} = \begin{bmatrix} \mathbf{S}_{\underline{\mathbf{MS}}_i} & 0 \\ 0 & \mathbf{S}_{\underline{\mathbf{MS}}_{i+1}} \end{bmatrix} \\ &\qquad \mathbf{F} = \begin{bmatrix} -\cos\phi_{\underline{\mathbf{MS}}_i,\underline{\mathbf{MS}}_{i+1}} & -\sin\phi_{\underline{\mathbf{MS}}_i,\underline{\mathbf{MS}}_{i+1}} & \cos\phi_{\underline{\mathbf{MS}}_{i+1},\underline{\mathbf{MS}}_i} & \sin\phi_{\underline{\mathbf{MS}}_{i+1},\underline{\mathbf{MS}}_i} \end{bmatrix} \\ &\qquad \phi_{\underline{\mathbf{MS}}_i,\underline{\mathbf{MS}}_{i+1}} = \arctan\left(\frac{\underline{\mathbf{MS}}_{y,i+1}-\underline{\mathbf{MS}}_{y,i}}{\underline{\mathbf{MS}}_{x,i+1}-\underline{\mathbf{MS}}_{x,i}}\right) \\ s^2_{\underline{dt}_i} &= s^2_{t_{\underline{MS},i}} + s^2_{t_{\underline{MS},i+1}}. \end{aligned} \tag{4.74}$$

Die Standardabweichung der Geschwindigkeit wird berechnet mit:

$$s_{\underline{v}_i} = \sqrt{\left(\frac{1}{\underline{dt}_i}\right)^2\cdot s^2_{\underline{d}_i} + \left(-\frac{\underline{d}_i}{\underline{dt}_i^2}\right)^2\cdot s^2_{\underline{dt}_i}} \qquad \text{und} \quad f_{\underline{v}_i} = f_{\underline{\mathbf{MS}}_i} + f_{\underline{\mathbf{MS}}_{i+1}}. \tag{4.75}$$

Zur Signifikanzprüfung dient der einseitige *t*-Test mit $H_0 : \mu_{\underline{v}_i} - v_{max,VM} \le 0 \sim t_{f_{\underline{v}_i},P}$ zur Annahme des VM und entsprechend $H_A : \mu_{\underline{v}_i} - v_{max,VM} > 0 \not\sim t_{f_{\underline{v}_i},P}$ für den Ausschluss des VM. Die Testgröße folgt aus:

$$\underline{T}_{VM,i} = \frac{\underline{v}_i - v_{VM,max}}{s_{\underline{v}_i}} \qquad \text{mit:} \quad \underline{v}_i = \frac{\underline{d}_i}{\underline{dt}_i} = \frac{\sqrt{\left(\underline{\mathbf{MS}}_{x,i+1}-\underline{\mathbf{MS}}_{x,i}\right)^2 + \left(\underline{\mathbf{MS}}_{y,i+1}-\underline{\mathbf{MS}}_{y,i}\right)^2}}{t_{\underline{\mathbf{MS}}_{i+1}} - t_{\underline{\mathbf{MS}}_i}}. \tag{4.76}$$

Das entsprechende VM kann nicht ausgeschlossen werden, falls für alle $\underline{\mathbf{MS}}_i, \underline{\mathbf{MS}}_{i+1}$ mit $i = 1,2\ldots n-1$ die Bedingung $\underline{T}_{VM,i} \le t_{f_{\underline{v}_i},P}$ bei der Sicherheitswahrscheinlichkeit *P* erfüllt wird. Die Geschwindigkeit der MS war in diesem Fall nicht-signifikant höher als es mit dem VM möglich ist. Wird hingegen bereits einmal $\underline{T}_{VM,i} > t_{f_{\underline{v}_i},P}$ festgestellt, dann bewegte sich die MS signifikant schneller als es mit dem VM möglich wäre:

$$\underline{T}_{VM,i} \begin{Bmatrix} \le \\ \forall i=1,\ldots,n-1 \\ > \\ \exists i\in 1,\ldots,n-1 \end{Bmatrix} t_{f_{\underline{v}_i},P} \Rightarrow \begin{cases} H_0 \text{ nicht verwerfen} & \rightarrow \text{„VM nicht ausschließen"} \\ H_0 \text{ verwerfen} & \rightarrow \text{„VM ausschließen".} \end{cases} \tag{4.77}$$

4.4.2 Unschärfe des Verkehrsmittel-Ausschlusses

Aus den unscharfen Clustern $\widetilde{MS}_i$ und $\widetilde{MS}_{i+1}$ folgt die unscharfe MS-Geschwindigkeit mittels der erweiterten Division ihrer unscharfe räumliche Distanz $\tilde{d}_i$ und ihrer unscharfe Zeitdifferenz $\widetilde{dt}_i$:

$$\tilde{v}_i\left(\widetilde{MS}_i, \widetilde{MS}_{i+1}\right) = \tilde{d}_i \oslash \widetilde{dt}_i \quad , \text{ mit: } 0 \notin \widetilde{dt}_i. \tag{4.78}$$

$\tilde{d}_i$ kann beispielsweise nach dem in Abschnitt 3.3.2.5 beschriebenen Verfahren berechnet werden. $\widetilde{dt}_i$ berechnet sich aus der erweiterten Subtraktion $\widetilde{dt}_i = \tilde{t}_{i+1} \ominus \tilde{t}_i$. Die Forderung $0 \notin \widetilde{dt}_i$ wird wegen dem kontinuierlichen Fortschreiten der Zeit auf der Zeitskala mit $\tilde{t}_i < \tilde{t}_{i+1}$ in der Regel erfüllt.

Annahmebereich und Verwerfungsbereich für die Testentscheidung, ob das VM mit der MS-Positionsfolge vereinbar ist, werden gebildet mit:

$$\tilde{A}_{VM} := \begin{cases} 1, & \text{für } v \le v_{max,VM} \\ 0, & \text{sonst} \end{cases} \qquad \text{und} \quad \tilde{V}_{VM} = 1 - \tilde{A}_{VM} := \begin{cases} 0, & \text{für } v \le v_{max,VM} \\ 1, & \text{sonst.} \end{cases} \tag{4.79}$$

Der Verwerfungsindex $\rho_{\tilde{V},VM}(\tilde{v}_i)$ folgt aus dem Grad der Verwerfung $\gamma_{\tilde{V},VM}(\tilde{v}_i)$ und dem Grad der Annahme $\delta_{\tilde{A},VM}(\tilde{v}_i)$:

$$\rho_{\tilde{V},VM}\left(\tilde{v}_i\right)=\min\left(\gamma_{\tilde{V},VM}\left(\tilde{v}_i\right)=\frac{\mathrm{card}\left(\tilde{v}_i\cap\tilde{V}_{VM}\right)}{\mathrm{card}\left(\tilde{v}_i\right)}\quad,\quad\delta_{\tilde{A},VM}\left(\tilde{v}_i\right)=1-\frac{\mathrm{card}\left(\tilde{v}_i\cap\tilde{A}_{VM}\right)}{\mathrm{card}\left(\tilde{v}_i\right)}\right). \tag{4.80}$$

Mit dem kritischen Wert für den Verwerfungsindex ρ_{Krit}, die Wahl geeigneter Werte für ρ_{Krit} wird in Abschnitt 6.3.2 gezeigt, folgt die binäre Testentscheidung:

$$\min_{i=1,\ldots,n-1}\left(\rho_{\tilde{V},VM}\left(\tilde{v}_i\right)\right)\begin{Bmatrix}\le\\>\end{Bmatrix}\rho_{Krit}\in[0,1]\Rightarrow\begin{cases}H_0\text{ nicht verwerfen} & \rightarrow\text{„VM nicht ausschließen"}\\ H_0\text{ verwerfen} & \rightarrow\text{„VM ausschließen".}\end{cases} \tag{4.81}$$

4.4.3 Fuzzy-Randomness des Verkehrsmittel-Ausschlusses

Im Falle des Fuzzy-Randomness können die Anteile der zufälligen Variabilität nach Formel (4.75) und die unscharfen Anteile nach Formel (4.78) auf die FRV der Geschwindigkeit $\underline{\tilde{v}}_i\left(\widetilde{\underline{\mathbf{MS}}}_i,\widetilde{\underline{\mathbf{MS}}}_{i+1}\right)$ fortgepflanzt werden. Wird für mindestens ein $\underline{\tilde{v}}_i$ der Positionsfolge die Maximalgeschwindigkeit $v_{max,VM}$ signifikant überschritten, ist die Nullhypothese $H_0:\mu_{\underline{\tilde{v}}_i}\le v_{max,VM},\left|\forall i=1,2\ldots,n\right.$ zu verwerfen. Die Freiheitsgrade $f_{\underline{\tilde{v}}_i}$ sowie die *t*-verteilte unscharfe Testgröße $\underline{\tilde{T}}_{v_i}$ für den einseitigen Test berechnen sich aus:

$$f_{\underline{\tilde{v}}_i}=f_{\widetilde{\underline{\mathbf{MS}}}_i}+f_{\widetilde{\underline{\mathbf{MS}}}_{i+1}}\qquad\text{und}\qquad\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}=\frac{\underline{\tilde{v}}_i-v_{max,VM}}{s_{\underline{\tilde{v}}_i}}. \tag{4.82}$$

Annahme- und Verwerfungsbereich werden mit dem Grenzwert $t_{f_{\underline{\tilde{v}},i},P=1-\alpha_{Irrtum}}$ festgelegt:

$$\tilde{A}:=\begin{cases}1, & \text{für } v\le v_{max,VM}+s_{\underline{\tilde{v}}_i}\cdot t_{f_{\underline{\tilde{v}},i},P}\\ 0, & \text{sonst}\end{cases}\qquad\text{und}\quad\tilde{V}=1-\tilde{A}:=\begin{cases}0, & \text{für } v\le v_{max,VM}+s_{\underline{\tilde{v}}_i}\cdot t_{f_{\underline{\tilde{v}},i},P}\\ 1, & \text{sonst.}\end{cases} \tag{4.83}$$

Aus dem Grad der Verwerfung $\gamma_{\tilde{V}}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)$ und dem Grad der Annahme $\delta_{\tilde{A}}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)$ wird der Verwerfungsindex $\rho_{\tilde{V}}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)$ berechnet mit:

$$\gamma_{\tilde{V}}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)=\frac{\mathrm{card}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\cap\tilde{V}\right)}{\mathrm{card}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)}\qquad\text{und}\qquad\delta_{\tilde{A}}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)=1-\frac{\mathrm{card}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\cap\tilde{A}\right)}{\mathrm{card}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)}$$
$$\rho_{\tilde{V}}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)=\min\left(\gamma_{\tilde{V}}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right),\delta_{\tilde{A}}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)\right). \tag{4.84}$$

Die binäre Testentscheidung folgt mit dem kritischen Wert für den Verwerfungsindex $\rho_{Krit}\in[0,1]$, die Wahl geeigneter Werte für ρ_{Krit} wird in Abschnitt 6.3.2 gezeigt, mit:

$$\min_{i=1,\ldots,n-1}\left(\rho_{\tilde{V}}\left(\underline{\tilde{T}}_{\underline{\tilde{v}}_i,VM}\right)\right)\begin{Bmatrix}\le\\>\end{Bmatrix}\rho_{Krit}\in[0,1]\Rightarrow\begin{cases}H_0\text{ nicht verwerfen} & \rightarrow\text{„VM nicht ausschließen"}\\ H_0\text{ verwerfen} & \rightarrow\text{„VM ausschließen".}\end{cases} \tag{4.85}$$

4.5 Korrelation zu ÖV-Fahrzeugen

Ziel des nachfolgend behandelten Verfahrens, ist die Identifikation von MS in ÖV-Fahrzeugen mit vordefinierten Fahrplänen. Diese Identifikation ist notwendig, um diejenigen MS herauszufiltern, deren Mobilfunkdaten in ÖV-Fahrzeugen generiert wurden. Würden diese MS nicht herausgefiltert werden, dann würden sie zu falschen Schlussfolgerungen der Verkehrslageerfassung für den MIV führen. Als Beispiel sei ein Schienenweg parallel zu einer Autobahn genannt. Aufgrund der räumlichen Nähe dieser beiden Verkehrsnetze und der Ortungsunsicherheiten kann eine räumliche Unterscheidung von MS auf dem einen oder dem anderen Verkehrsnetz nicht möglich sein. MS in einem schnell fahrenden Zug könnten daher von MS in Fahrzeugen auf der Autobahn nicht unterschieden werden. Infolgedessen würden beispielsweise Schätzungen der Verkehrsstärke oder der Geschwindigkeit für die betreffenden Streckenabschnitte negativ beeinflusst werden und ein Stau auf der Autobahn unter Umständen nicht erkannt werden.

Im Rahmen der vorliegenden Arbeit werden für die Identifizierung der ÖV-Fahrzeuge die Verkehrsmittel Linienbus (**Bus**), Straßenbahn (**Tram**) sowie die Regional- und Fernbahnen der Deutsche Bahn AG (**DB**) auf den Gleisen der DB Netz AG untersucht. Ausgangspunkt ist die Nutzung von **Fahrplandaten**, woraus die Zeitpunkte und Positionen der ÖV-Fahrzeuge abgeleitet werden. Die Untersuchung der Korrelation dieser räumlichen und zeitlichen Informationen mit denen der MS-Positionen gestattet die Identifikation von MS in ÖV-Fahrzeugen. Sind die Korrelationen genügend hoch, kann schlussgefolgert werden, dass die Mobilfunkdaten der MS in einem ÖV-Verkehrsmittel generiert wurden, was Ausgangspunkt für ein Verfahren zur Identifikation von ÖV-Fahrzeugen sein könnte. Da der Fokus auf die Identifikation von MS in MIV gerichtet ist, ist hingegen folgende Kernfragestellung zu beantworten:

Kann ausgeschlossen werden, dass das MS mit dem ÖV korreliert ist?

Um eine gesicherte Testentscheidung für diese Fragestellung treffen zu können, sind neben den Unsicherheiten der MS-Positionen auch die zwischen den Fahrplänen und den real durchgeführten Fahrten auftretenden **Fahrplanabweichungen** mit einzubeziehen. Bei der Modellierung der Fahrplanabweichungen beschränkt sich die vorliegende Arbeit auf die zeitlichen Abweichungen mit **Verfrühungen** und **Verspätungen**. Räumliche Abweichungen, also das Verlassen des geplanten Fahrwegs, werden nicht betrachtet. Im folgenden Abschnitt wird die Aufbereitung der (fehlerfreien) Fahrwegsinformationen erläutert. Im Anschluss werden jeweils unter Anwendung der drei Arten der Unsicherheitsmodellierung die Aspekte der Fahrplanabweichungen und ein Verfahren zur Beantwortung der Frage, ob eine Korrelation des MS mit dem ÖV ausgeschlossen werden kann, vorgestellt.

4.5.1 Georeferenzierung der Fahrplandaten

Mit den Fahrplänen der Verkehrsverbünde und -betriebe werden die Zu- bzw. Aussteigeorte, also die Haltepunkte der ÖV-Fahrzeuge angegeben. Die Ketten der Haltepunkte und gegebenenfalls weiteren geometrischen Zwischenpunkten ergeben die Routen der ÖV-Fahrzeuge. Auf diesen Routen, kann schließlich die ÖV-Fahrzeugposition prognostiziert werden, vgl. Abbildung 4-18.

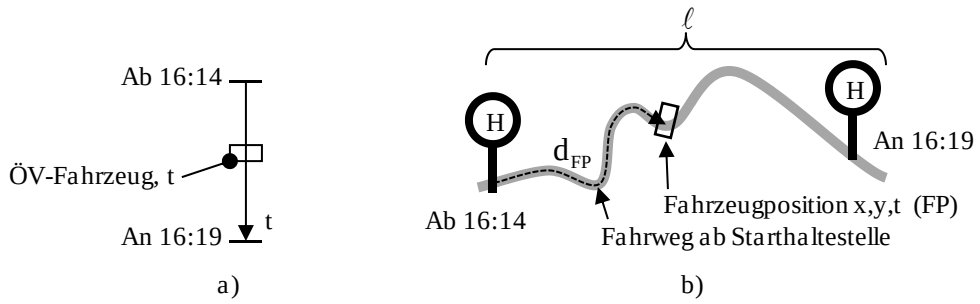


Abbildung 4-18: Schemadarstellung der Interpolation von ÖV Fahrzeugpositionen,
a) als zeitliche Kette und b) mit georeferenzierter Darstellung des Fahrweges und des Fahrplans

Grundlage zur geometrischen Beschreibung der Routen sind Verkehrsnetze. In die topologische Beschreibung der Verkehrsnetze als gerichtete Graphen $G=(\mathbf{V},\mathbf{E})$ mit den Knoten $\mathbf{V}=\{V_1,V_2,\ldots\}$ und den Kanten $\mathbf{E}=\{E_1,E_2,\ldots\}$ wurde bereits in Abschnitt 4.2.5 eingeführt. Während die Straßenbahnen sowie die Regional- und Fernbahnen eigene Schienenverkehrsnetze haben, nutzen Linienbusse im Wesentlichen das auch vom MIV verwendete Straßenverkehrsnetz $G^{Straße}$ (vgl. in Abschnitt 4.4 die Tabelle 4-1). Zur Verdeutlichung des ÖV-Verkehrsmittels, wird im vorliegenden Identifikationsverfahren die Bezeichnung G^{Bus} anstatt $G^{Straße}$ für das Straßenverkehrsnetz der Linienbusse verwendet. Mit G^{Tram} wird das Schienenverkehrsnetz der Straßenbahnen sowie mit G^{DB} das Schienenverkehrsnetz der Regional- und Fernbahnen der DB Netz AG bezeichnet.

Als Datenquelle des Straßenverkehrsnetzes eignen sich die digitalen Straßenkarten nach dem GDF-Standard (vgl. den Hinweis in Abschnitt 4.2.5). Kommerzielle Marktführer derartiger Kartenhersteller sind NAVTEQ® und TOMTOM® (NAVTEQ 2011; TOMTOM 2011). Auf Basis einer CC-BY-SA-Lizenz (Creative Commons Attribution-Share Alike) (CC-BY-SA-LICENCE 2010) werden im sogenannten OpenStreetMap-Projekt (OPENSTREETMAP 2011) weltweite Geodaten gesammelt und im OSM-Datenformat offen sowie kostenfrei verbreitet (RAMM & TOPF 2009). Die Verkehrsnetzsdaten der vorliegenden Arbeit entstammen den im Projekt DO-IT erarbeiteten Verkehrsnetzen, die auf digitalen Straßenkarten von NAVTEQ® für G^{Bus} und im Projekt DO-IT vom VuV[14] der Universität Stuttgart zusammengestellten Gleisnetzdaten für G^{Tram} und G^{DB} basieren.

Die **Haltepunkte H** der Fahrpläne wurden mittels ihrer Koordinaten auf die entsprechenden Kanten der Verkehrsnetzgraphen referenziert. Insofern die Haltepunkte nicht bereits identisch mit den in den Verkehrsnetzen vorhandenen Knoten waren, wurden sie als neue Knoten in die Verkehrsnetzkanten eingeschoben. Die Abbildung 4-19a) und Abbildung 4-19b) zeigen die Referenzierung auf Geometrieebene und die Abbildung 4-19c) die entsprechende topologische Darstellung.

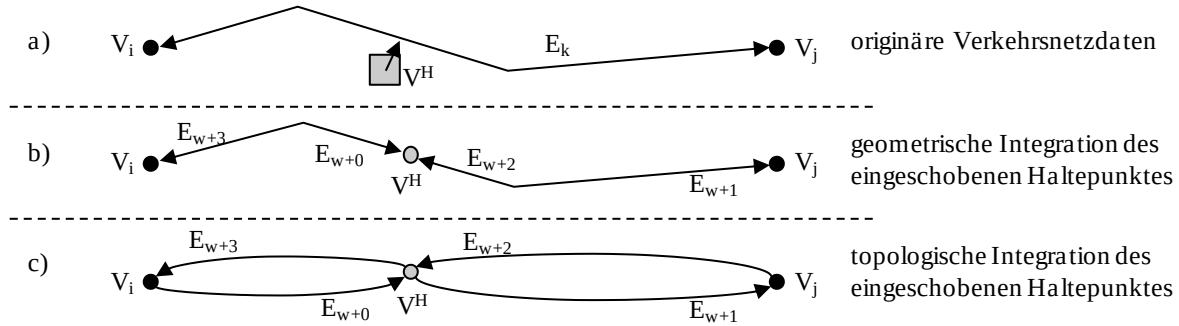


Abbildung 4-19: Referenzierung von Haltepunkten auf die Verkehrsnetze

[14] Der Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik (VuV) Universität Stuttgart ist einer von zwei Lehrstühlen des Instituts für Straßen- und Verkehrswesen (ISV) Universität Stuttgart.

Haltepunkte sind damit Bestandteil der Knotenmenge der Verkehrsnetzgraphen $H \xrightarrow[\text{auf Verkehrsnetz}]{\text{Referenzierung}} V^H$, besitzen jedoch neben den geometrischen und topologischen Eigenschaften auch die fahrplanmäßigen Ankunftszeiten $t_{V^H,An}$ sowie Abfahrzeiten $t_{V^H,Ab}$ der ÖV-Fahrzeuge. Die Menge aller Kanten und Knoten zwischen zwei aufeinander folgenden Haltepunkten V_i^H und V_j^H heißt **Streckenabschnitt** ℓ. Innerhalb eines Streckenabschnitts wird der zeitlich frühere Haltepunkt mit V_{From}^H und der zeitlich spätere mit V_{To}^H bezeichnet $\left(V_{From}^H, V_{To}^H\right) \in \ell$.

Um auf den Verkehrsnetzgraphen vollständige Routen von Anfangshaltestelle bis Endhaltestelle zu erstellen, werden aus den Fahrplandaten die Reihenfolgen der Haltepunkte entnommen. Eine derartige Route wird als ***Fahrt*** bezeichnet. Sie gibt den geometrischen und topologischen Verlauf des Fahrweges einschließlich der aus den Fahrplänen entnommenen Zeiten $t_{V_{To}^H,An}$ und $t_{V_{From}^H,Ab}$ für die Knoten der Haltepunkte für ein spezifisches ÖV-Fahrzeug an. Die Menge aller geometrisch und topologisch identischer *Fahrt*en heißt **Linie** L mit $\mathbf{L} = \{L_1, L_2, \ldots\}$, worin die Streckenabschnitte ℓ Bestanteile einer Linie sind $\ell \in L$. Aufgrund der Definition mit gerichteten Graphen, sind die *Fahrt* $A \in L_1$ mit Richtung von der Anfangshaltestelle V_i^H zur Endhaltestelle V_j^H und die *Fahrt* $B \in L_2$ von V_j^H nach V_i^H trotz geometrisch gleicher Route verschieden, und damit auch ihre Linie *Fahrt* $A \neq$ *Fahrt* $B \Rightarrow L_1 \neq L_2$. Das Tripel $\left(\mathbf{V}^H, \mathbf{E}, \mathbf{L}\right)$ wird **Netzgraph** NG genannt und enthält die Geometrieinformationen sowie die nach den Linien der Fahrpläne organisierten Fahrplanzeiten. Mit den neu in die Verkehrsnetzgraphen eingefügten, fahrplanmäßig bekannten Haltepunkten $\mathbf{V}^{H\text{-Bus}}$, $\mathbf{V}^{H\text{-Tram}}$ und $\mathbf{V}^{H\text{-DB}}$ heißen die Netzgraphen entsprechend ihrer Kategorisierung:

$$NG^{Bus} = \left(\mathbf{V}^{Bus}, \mathbf{E}^{Bus}, \mathbf{L}^{Bus}\right) \quad \text{mit: } \mathbf{V}^{Bus} \supseteq \mathbf{V}^{H\text{-Bus}} \tag{4.86}$$

$$NG^{Tram} = \left(\mathbf{V}^{Tram}, \mathbf{E}^{Tram}, \mathbf{L}^{Tram}\right) \quad \text{mit: } \mathbf{V}^{Tram} \supseteq \mathbf{V}^{H\text{-Tram}} \tag{4.87}$$

$$NG^{DB} = \left(\mathbf{V}^{DB}, \mathbf{E}^{DB}, \mathbf{L}^{DB}\right) \quad \text{mit: } \mathbf{V}^{DB} \supseteq \mathbf{V}^{H\text{-DB}}. \tag{4.88}$$

Ein wesentliches Kriterium für die Interpolation von Fahrzeugpositionen auf dem Linienweg ist der Zeitbedarf als Widerstandswert des bewerteten Graphen. Der Zeitbedarf $\Delta t_\ell \in L$ zwischen zwei aufeinander folgenden Knoten V_{From}^H und V_{To}^H eines Streckenabschnitts ℓ der Linie L, folgt aus der Fahrplanzeit als Differenz zwischen Abfahrtzeit und Ankunftszeit $\Delta t_\ell = t_{V_{To}^H,An} - t_{V_{From}^H,Ab}$ mit $\left(V_{From}^H, V_{To}^H\right) \in \ell$. Die Zeiten entstammen zumeist Optimierungsverfahren der Fahrplanerstellung, worin gegebenenfalls örtliche sowie tageszeitliche Einflüsse des sonstigen Verkehrs mit einfließen. Entlang des Linienweges können hieraus **Geschwindigkeitsprofile für die ÖV-Fahrzeuge** abgeleitet werden. In vielen Fällen ist der Detaillierungsgrad der Geschwindigkeitsprofile aus den originären Verkehrsnetzen jedoch höher, als der aus den Fahrplandaten. Um die Geschwindigkeitsprofile der originären Verkehrsnetze zu nutzen, werden der Zeitbedarf und damit die Geschwindigkeiten der Verkehrsnetzkanten v_k^* mit der fahrplanmäßigen Fahrzeit des Streckenabschnitts skaliert. Basierend auf der Geschwindigkeit v_k der Kanten des originären Verkehrsnetzes folgt die skalierte Geschwindigkeit v_k^* aus:

$$\left.\begin{aligned} v_k^* = m_\ell \cdot v_k \quad \text{mit:} \quad & m_\ell = \frac{\sum_{\forall E_k \in \ell} \Delta t_k}{\Delta t_\ell} && \left(\text{Skalierungsfaktor}\right) \\ & \Delta t_\ell = t_{V_{To}^H,An} - t_{V_{From}^H,Ab} \;,\; \left(V_{From}^H, V_{To}^H\right) \in \ell && \left(\text{fahrplanmäßige Fahrzeit}\right) \\ & \Delta t_k = \frac{d_k}{v_k} \;,\; \forall E_k \in \ell \;,\; \ell \in L && \left(\text{Fahrzeit aus originärem Verkehrsnetz}\right) \end{aligned}\right\} \tag{4.89}$$

Darin sind d_k die metrischen Längen und v_k die Durchschnittsgeschwindigkeiten der Kanten E_k des originären Verkehrsnetzes, für alle $E_k \in \ell$ des Streckenabschnitts $\ell \in L$ der Linie L. Zu beliebigen

Fahrzeugpositionen FP auf dem Weg der Linie L kann nun die Antreffzeit des ÖV-Fahrzeugs unter Berücksichtigung der Geschwindigkeitsprofile aus den originären Verkehrsnetzgraphen prognostiziert werden.

4.5.2 Prognose der Anwesenheitszeiten von ÖV-Fahrzeugen

Für die Bestimmung der Korrelation zu den ÖV-Fahrzeugen ist zu klären, ob räumliche und auch zeitliche Korrelationen zwischen (mindestens) einer Position der MS und den ÖV-Fahrzeugen bestehen. Mittels der Fahrpläne wurde festgelegt, zu welchen Zeiten sich die ÖV-Fahrzeuge an Haltepunkten aufhalten. Um auch Korrelationen zwischen MS und den ÖV-Fahrzeugen zwischen den Haltepunkten bestimmen zu können, ist die Prognose der Anwesenheitszeiten für (beliebige) Fahrzeugpositionen FP entlang der Fahrrouten erforderlich. In diesem Zusammenhang ist anzumerken, dass ein und derselbe Ort mehrfach vom selben ÖV-Fahrzeug, aber auch von mehreren unterschiedlichen ÖV-Fahrzeugen aufgesucht werden kann.

Zur Prognose des Zeitpunktes, wann sich ein ÖV-Fahrzeug an der Stelle FP auf der Fahrroute aufhält, vgl. Abbildung 4-18b), ist diejenige Kante E_k im Netzgraphen zu finden, auf der FP liegt (oder projiziert wird). Für alle Streckenabschnitte ℓ - unter Einbeziehung aller Linien –, in denen die Kante E_k enthalten ist, ist die Anwesenheit des jeweiligen ÖV-Fahrzeugs an FP zu bestimmen. Diese Streckenabschnitte werden hierfür zur **Menge aller Streckenabschnitte** Λ zusammengefasst.

Ausgehend von V^H_{From} über den geometrischen Verlauf der Kanten von ℓ bis FP, exklusive der Kante E_k, ist mit den Kantenlängen d und den skalierten Geschwindigkeiten v^* die Summe der Fahrzeiten $\Delta t_{\Sigma E}$ bis zur Kante E_k zu berechnen. Die Fahrzeit Δt_{FP} des übrigbleibenden Teilstückes auf der Kante E_k bis zur Stelle FP ist mit dem Fahrweg $d_{k,FP}$ und der skalierten Geschwindigkeit v^*_k dieser Kante zu berechnen. Die Anwesenheitszeit t_{FP}, zu der sich das ÖV-Fahrzeug an der Position FP befindet, berechnet sich aus der Summe dieser Fahrzeiten mit der fahrplanmäßigen Abfahrtzeit $t_{V^H_{From},Ab}$, mit $V^H_{From} \in \ell$, vgl. Formel (4.90). Beschleunigungs- und Bremszeiten werden im Rahmen dieser Arbeit nicht berücksichtigt. In Tabelle 4-2 ist der Ablauf dieser Berechnung zusammenfassend beschrieben.

$$
\left.
\begin{aligned}
& t_{FP} = t_{V^H_{From},Ab} + \Delta t_{\Sigma E} + \Delta t_{FP} \\
& \\
& \text{mit:} \quad t_{V^H_{From},Ab} \text{ mit } V^H_{From} \in \ell \qquad \left(\text{Abfahrtzeit von der Starthaltestelle des Streckenabschnitts } \ell\right) \\
& \qquad \Delta t_{\Sigma E} = \sum_{i=0}^{k-1} \frac{d_i}{v^*_i} \qquad \left(\begin{array}{l}\text{Summe des Zeitbedarfs aller Kanten} \\ \text{von } V^H_{From} \text{ bis } E_k, \text{ ohne die Kante } E_k\end{array}\right) \\
& \qquad \Delta t_{FP} = \frac{\Delta d_{k,FP}}{v^*_k} \qquad \left(\begin{array}{l}\text{Fahrzeit für das Teilstück auf der} \\ \text{Kante } E_k \text{ bis FP}\end{array}\right) \\
& \text{Es bedeuten außerdem:} \quad \Delta d_{FP} \quad \text{Fahrweg für das Teilstück auf der Kante } E_k \text{ bis FP} \\
& \qquad\qquad\qquad\qquad E_k \quad \text{Kante, auf der sich FP befindet.}
\end{aligned}
\right\} \qquad (4.90)
$$

Tabelle 4-2: Ablauf der Berechnung der Anwesenheitszeiten von ÖV-Fahrzeugen zu spezifischen Positionen

Schritt	Beschreibung
1.	Suche (auf Geometrieebene) die Kante $E_k \in NG$, für die $FP \in E_k$ gilt
2.	Bilde die Menge aller Streckenabschnitte Λ aus den Streckenabschnitten ℓ, in denen die Kante E_k enthalten ist, also wenn $FP \in E_k \wedge E_k \exists \ell$ gilt, dann ist $\ell \in \Lambda$
3.	Berechne $\forall \ell \in \Lambda$ die Summe der Kantenfahrzeit $\Delta t_{\Sigma E}$ bis zur Kante $E_k \in \ell$
4.	Berechne auf der Kante $E_k \in \ell$ die Fahrzeit Δt_{FP} bis zur Fahrzeugposition FP
5.	Berechne $\forall \ell \in \Lambda$ den Zeitpunkt $t_{FP} = t_{V^H_{From},Ab} + \Delta t_{\Sigma E} + \Delta t_{FP}$

Bezüglich der in Abschnitt 4.2.5 eingeführten Verkachelung der Verkehrsnetze, können die Linienwege nun als Ketten von Kacheln dargestellt werden, wobei die Einfahrzeit in die Kachel t_1 und die Ausfahrzeit aus der Kachel t_2 die prognostizierten Fahrplanzeiten jeweils an den Kachelübergängen widerspiegeln. Diese Übergänge von einer Kachel in die nachfolgende Kachel sind in Abbildung 4-20 mit Kreisen gekennzeichnet. Die Einfahrzeiten und Ausfahrzeiten an den Schnittpunkten FP_j mit $j = 1, 2, \ldots, n$ des Fahrwegs mit den Kachelgrenzen werden berechnet mit:

$$t_{1,K_{j+1}} = t_{2,K_j} = t_{FP_j} \quad \text{mit:} \quad \begin{array}{ll} K_j & \text{Kachel, die verlassen wurde,} \\ K_{j+1} & \text{Kachel, in die eingefahren wird,} \\ j = 1, 2 \ldots n & \text{mit n der Anzahl Schnittpunkte,} \\ h = 1, 2 \ldots m & \text{mit m der Anzahl Haltepunkte von L.} \end{array} \tag{4.91}$$

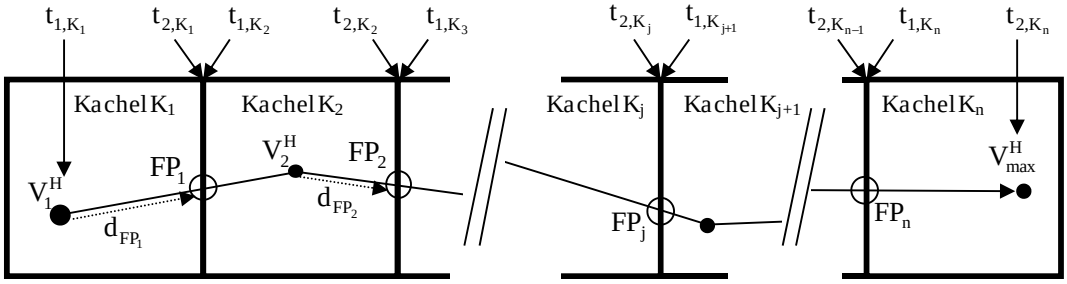


Abbildung 4-20: Transformation der Linienwege und Fahrplanzeiten ins Kachelkoordinatensystem

Die Einfahrzeit in die erste Kachel t_{1,K_1} entspricht der ersten fahrplanmäßig bekannten Zeit, in der Regel ist dies die Abfahrtzeit von der Starthaltestelle. Die Ausfahrzeit der letzten Kachel t_{2,K_n} entspricht der letzten fahrplanmäßig bekannten Zeit, welche in der Regel die Ankunftszeit an der letzten Haltestelle ist. Da dieselbe Kachel mehrfach überfahren werden kann, z. B. bei Kreisfahrten oder aufgrund der Kachelausdehnung, können mehrere K_j in Formel (4.91) die gleiche Kachelkoordinate *EN* besitzen. An diesen Kacheln sind demzufolge zu mehreren Zeitpunkten ÖV-Fahrzeuge anzutreffen. Beispielsweise ergibt sich für eine Kachel die dreimal durchquert wird, das in Abbildung 4-21 dargestellte Anwesenheitsmuster.

Abbildung 4-21: Anwesenheitsmuster von ÖV-Fahrzeugen für eine dreimal durchquerte Kachel

4.5.3 Modellierung von Fahrplanabweichungen

Wie in Abbildung 4-21 durch die senkrecht gepunkteten Linien angedeutet, fand die Modellierung bislang ausschließlich für „exakt“ pünktliche ÖV-Fahrzeuge statt. Im Allgemeinen ist davon nicht auszugehen, dass Unsicherheiten aus Verfrühungen und Verspätung unvermeidbar sind. Verfrühungen können z. B. durch Abfahrt vor der geplanten Abfahrtzeit oder wegen zu pessimistischen Geschwindigkeitsannahmen auf den Verkehrsnetzkanten entstehen. Verspätungen resultieren zumeist aus Kapazitätsengpässen, z. B. wegen vorausfahrenden Fahrzeugen, die nicht überholt werden können, oder auch an Haltestellen durch Verzögerungen beim Fahrgastwechsel. Die Ursachen für Fahrplanabweichungen sind vielschichtig.

Die nachfolgenden Verfahren sollen kleinere, betriebsbedingte Unsicherheiten der Fahrplanabweichungen modellieren. Die Modellierung von Unfällen oder anderen seltenen großen Katastrophen erscheint in diesem Rahmen nicht sinnvoll, weil sie im Allgemeinen weder vorhergesagt, noch abgeschätzt werden können.

Die Fahrplanabweichung wird als die zeitliche Differenz $\Delta t_{PlanAbw}$ an einem spezifischen Ort FP zwischen der geplanten t_{Plan} und der tatsächlichen Anwesenheit t_{Ist} eines ÖV-Fahrzeuges definiert mit:

$$\Delta t_{PlanAbw} = t_{Plan} - t_{Ist}. \tag{4.92}$$

Ist laut Fahrplan und einschließlich der Unsicherheiten zum Zeitpunkt t_{Ist} ein ÖV-Fahrzeug an FP anzutreffen, besitzt die MS eine Korrelation zum ÖV. Befindet sich hingegen laut Fahrplan kein ÖV-Fahrzeug an FP, dann ist eine Korrelation zum ÖV auszuschließen. Die für diese Prüfung zu berücksichtigenden Unsicherheiten von t_{Plan} wurden aus empirisch erhobenen Daten abgeleitet. Eine Zusammenstellung dieser Daten und der daraus gewonnenen Unsicherheitsparameterwerte befindet sich aufgeschlüsselt nach Linienbussen, Straßenbahnen und den Zügen auf den Gleisen der DB Netz AG in Anhang D.

4.5.3.1 Zufällige Variabilität der Fahrplanabweichungen

Im Falle der zufälligen Variabilität wird für die Fahrplanabweichungen eine Normalverteilung angenommen $\underline{\Delta t}_{PlanAbw} \sim N\left(\mu_{\underline{\Delta t}_{PlanAbw}}, \sigma^2_{\underline{\Delta t}_{PlanAbw}}\right)$, wobei auf die Wahl dieser Funktion am Ende des Abschnitts noch näher eingegangen wird. Die zufällige Variabilität für $\underline{t}_{FP,Ist}$ bildet sich daraufhin aus der Varianz $\sigma^2_{\underline{\Delta t}_{PlanAbw}}$ und der Varianz der MS-Position $s^2_{t_{\underline{MS}}}$ mit den Freiheitsgraden $f_{t_{\underline{MS}}}$ sowie der zufällige Mittelpunkt aus dem Erwartungswert $\mu_{\underline{\Delta t}_{PlanAbw}}$ und dem zufälligen Mittelpunkt $t_{\underline{MS},m}$:

$$\begin{aligned} s^2_{\underline{t}_{FP,Ist}} &= s^2_{t_{\underline{MS}}} + \sigma^2_{\underline{\Delta t}_{PlanAbw}} \quad \text{und} \quad f_{\underline{t}_{FP,Ist}} = f_{t_{\underline{MS}}} \\ \underline{t}_{FP,Ist} &= t_{\underline{MS},m} + \mu_{\underline{\Delta t}_{PlanAbw}}. \end{aligned} \tag{4.93}$$

Für eine Kachel K kann nun die Wahrscheinlichkeit $P_{\underline{FP},K,Fahrt}$ berechnet werden, dass $\underline{t}_{FP,Ist}$ innerhalb des geplanten Zeitfensters t_1 und t_2 des ÖV-Fahrzeugs *Fahrt* liegt. Dazu ist die Fläche der Wahrscheinlichkeitsdichtefunktion der Student-Verteilung zwischen t_1 und t_2 beim Freiheitsgrad $f_{\underline{t}_{FP,Ist}}$ zu bestimmen, vgl. Abbildung 4-22:

$$P_{\underline{FP},K,Fahrt} = P_{t_{f_{\underline{t}_{FP,Ist}}}}\left(t_1 \le \mu_{\underline{t}_{FP,Ist}} \le t_2\right). \tag{4.94}$$

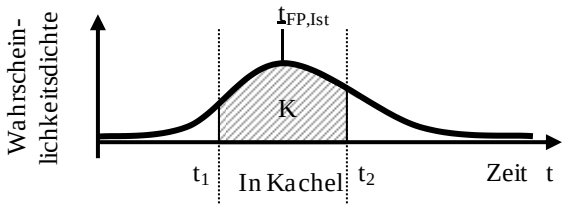


Abbildung 4-22: Bestimmung der Wahrscheinlichkeit für ein MS in der Kachel K eines ÖV-Fahrzeugs

Mit der in Abschnitt 4.2.6.1 bestimmten Aufenthaltswahrscheinlichkeit in der Kachel $P_{\underline{MS},K}$, wird die Wahrscheinlichkeit $P_{\underline{MS},K,Fahrt}$ berechnet, dass die MS zum Zeitpunkt $t_{\underline{MS}}$ mit dem ÖV-Fahrzeug *Fahrt* in der Kachel K sowohl zeitlich als auch räumlich korreliert ist:

$$P_{\underline{MS},K,Fahrt} = P_{\underline{MS},K} \cdot P_{\underline{FP},K,Fahrt}. \tag{4.95}$$

Diese Wahrscheinlichkeit ist im Allgemeinen für alle Kacheln – die Funktion der Normalverteilung ist im Bereich von $(-\infty, +\infty)$ definiert - zu berechnen, so dass die Summe aller $P_{\underline{MS},K,Fahrt}$ je *Fahrt* die Wahrscheinlichkeit angibt, dass das MS mit diesem ÖV-Fahrzeug zum Zeitpunkt $t_{\underline{MS}}$ sowohl zeitlich als auch räumlich korreliert ist:

$$P_{\underline{MS},Fahrt} = \sum_{k=1}^{n} P_{\underline{MS},K_k,Fahrt}. \tag{4.96}$$

Mit Definition der kritischen Grenze $P_{\text{ÖV,min}}$ erfolgt die binäre Entscheidung:

$$\begin{aligned} &\text{Fall I:} \quad P_{\underline{\mathbf{MS}},Fahrt} \geq P_{\text{ÖV,min}} \quad \text{„MS ist an dieser Position mit } \textit{Fahrt} \text{ korreliert"} \\ &\text{Fall II:} \quad P_{\underline{\mathbf{MS}},Fahrt} < P_{\text{ÖV,min}} \quad \text{„MS ist an dieser Position nicht mit } \textit{Fahrt} \text{ korreliert".} \end{aligned} \tag{4.97}$$

In der Praxis ist der Rechenbedarf für die Ermittlung von $P_{\underline{\mathbf{MS}},Fahrt}$ unverhältnismäßig hoch, da die Wahrscheinlichkeitswerte für alle Kacheln der Verkehrsnetze sowie jeweils für alle Kacheleinfahrzeiten und Kachelausfahrzeiten zu berechnen sind. Der dafür erforderliche Zeitbedarf ist – zumindest mit heutigen Rechnern - nicht akzeptabel. Aus diesem Grund wird räumlich nur auf die in Abschnitt 4.2.6.1 genannten MS - Positionskacheln mit der Mindestaufenthaltswahrscheinlichkeit $P_{\underline{\mathbf{MS}},\text{K,min}}$ zurückgegriffen. Zeitlich werden zudem nur diejenigen Zeitfenster verwendet, deren Zeitpunkte t_1 oder t_2 innerhalb eines mit t_0 und t_3 festgelegten Bereiches einer hohen Sicherheitswahrscheinlichkeit (z. B. $P_{\text{ÖV,max}} = 99{,}9\%$) liegen. Aufenthaltswahrscheinlichkeiten außerhalb dieser Zeitfenster, sind entsprechend gering (z. B. kleiner als 0,1%), weshalb sie vernachlässigt werden können. Für die beiden Zeitpunkte t_1 oder t_2 muss dann gelten:

$$\begin{aligned} &\left(t_1 < t_3\right) \wedge \left(t_0 < t_2\right) \quad \text{mit:} \quad t_0 = \underline{t}_{\text{FP,Ist}} - s_{\underline{t}_{\text{FP,Ist}}} \cdot t_{f_{\underline{t}_{\text{FP,Ist}}},P_{\text{ÖV,max}}} \\ &\qquad\qquad\qquad\quad t_3 = \underline{t}_{\text{FP,Ist}} + s_{\underline{t}_{\text{FP,Ist}}} \cdot t_{f_{\underline{t}_{\text{FP,Ist}}},P_{\text{ÖV,max}}}, \\ &\qquad\qquad\qquad\quad t_{f_{\underline{t}_{\text{FP,Ist}}},P_{\text{ÖV,max}}} : \text{Quantil der Student-Verteilung.} \end{aligned} \tag{4.98}$$

Die vier möglichen Fälle, um die Bedingung in Formel (4.98) zu erfüllen, sind in der folgenden Abbildung 4-23 grafisch dargestellt.

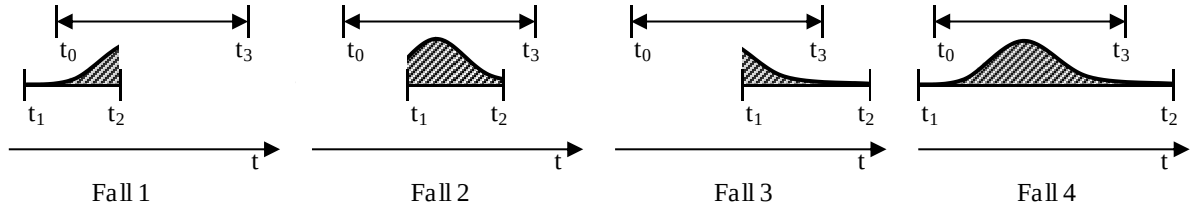


Abbildung 4-23: Grafisch dargestellte Fälle zur Auswahl derjenigen Zeitfenster, die einen möglichen relevanten Beitrag zur Berechnung der Aufenthaltswahrscheinlichkeit haben

Neben der hier verwendeten Normalverteilung werden in der Literatur weitere Verteilungsfunktionen für Fahrplanabweichungen vorgeschlagen. Während im Linienbusbetrieb nennenswerte Verfrühungen und Verspätungen auftreten (OETTING 2002, S. 181), beschränkt man sich in der Literatur für den gleisbezogenen ÖV vorrangig auf die Modellierung von Verspätungen. In der Praxis erfasste Fahrplanabweichungen bekräftigen dieses Vorgehen, da Verfrühungen im Eisenbahnverkehr durch spontan eingeschobene Wartezeiten vermieden werden. Die Gründe für die unterschiedlich starke Häufung, insbesondere der Verfrühungen im Linienbusbetrieb, können in der Fahrplanerstellung gefunden werden. Unter anderem fließen in deren Optimierung planmäßigen Umsteigewartezeiten, Investitionskosten, Ressourcenverbrauch und die Fahrzeuganzahl mit ein. Verkehrsbedingte Störungen sind auf Gleisen weniger zu erwarten als auf Straßen, da Straßen beispielsweise auch vom MIV genutzt werden. Auf den Gleisen können Pufferzeiten daher geringer ausfallen und die Maximalgeschwindigkeiten in die Planung mit einfließen. Der Aufbau von Verfrühungen ist damit deutlich geringer, hingegen Verspätungen bereits bei kleineren Störungen leicht möglich.

Mittels statistischer und numerischer Methoden und z. T. erheblichen Mengen empirisch erfasster Daten wurden Anpassungen an die Fahrplanabweichungen mittels Exponentialverteilungen (SCHWANHÄUSSER 1974), Weibullverteilung, Gammaverteilung (HERMANN 1996) oder auch grafisch begründete Anpassungen an Exponentialverteilungen (WEIGAND 1981) vorgenommen, um nur einige wenige zu nennen. In der Praxis und insbesondere auch bei der DB AG ist die Verteilungsfunktion nach einer Exponentialverteilung von

SCHWANHÄUSSER (1974) relevant (VAKHTEL 2002). Mit Ansätzen von stochastischen Modellen wird auf mathematischem Weg dieser Verteilungstyp von ENGELHARDT-FUNKE (2002) bestätigt. Die in der vorliegenden Arbeit eingesetzte Normalverteilung ist vom gleichen Funktionstyp $\left(f(x) = b \cdot e^{-a \cdot t}\right)$, weshalb ihr prinzipieller Einsatz dadurch untermauert wird.

In OETTING (2002) wird ebenfalls der Einsatz einer Normalverteilung motiviert. Dort wurde zudem beabsichtigt, die Pünktlichkeit aus den Toleranzgrenzen der Haltezeitdifferenz zwischen Ankunftszeit und Abfahrtzeit festzulegen. Diese Toleranzgrenzen sind daraufhin als Unstetigkeitsstellen in die Verteilungsfunktion zu integrieren. Abgesehen von den seltenen Fällen, dass ein Haltepunkt „exakt" auf einer Kachelgrenze liegt, sind Haltepunkte immer innerhalb einer Kachel und damit die Haltezeiten bereits mit dem Zeitfenster t_1 oder t_2 modelliert. Für die Verteilungsfunktion ist somit die Bestimmung der Parameterwerte μ und σ der Normalverteilung ausreichend. Unstetigkeiten werden nicht eingeführt. Aus den empirisch erhobenen Fahrplanabweichungen wurden die in Tabelle 4-3 zusammengefassten Parameterwerte bestimmt (vgl. dazu auch Anhang D).

Tabelle 4-3: Parameterwerte für die Normalverteilungsfunktionen der Fahrplanabweichungen

Verkehrsnetz	$\mu_{\Delta t_{PlanAbw}}$ [sec]	$\sigma_{\Delta t_{PlanAbw}}$ [sec]
G^{Bus}	-91,6	76,6
G^{Tram}	-27,6	61,8
G^{DB}	-192,8	341,3

4.5.3.2 Unschärfe der Fahrplanabweichungen

Die Übertragung der unscharfen Fahrplanabweichung $\widetilde{\Delta t}_{PlanAbw}$ auf die zeitliche Komponente der unscharfen MS-Position $t_{\widetilde{MS}}$ erfolgt nach den Regeln der erweiterten Addition:

$$\tilde{t}_{FP,Ist} = t_{\widetilde{MS}} \oplus \widetilde{\Delta t}_{PlanAbw}. \tag{4.99}$$

Für den Vergleich von $\tilde{t}_{FP,Ist}$ mit den Fahrplanzeiten werden die Kacheleinfahrtzeit t_1 und Kachelausfahrzeit t_2 als Fuzzy-Menge formuliert. Der Zeitraum, in dem sich das ÖV-Fahrzeug *Fahrt* laut Fahrplan innerhalb der Kachel K befindet, wird dazu als Fuzzy-Intervall $\widetilde{Fahrt}_K$ ausgedrückt mit:

$$\widetilde{Fahrt}_K := \left[m = \frac{t_1 + t_2}{2}, r = \frac{t_2 - t_1}{2}, c_l = 0, c_r = 0 \right]_L = m_{\widetilde{Fahrt}_K}(x) = \begin{cases} 1, & \text{für: } t_1 \le x \le t_2 \\ 0, & \text{sonst.} \end{cases} \tag{4.100}$$

Aus dem mengentheoretischen Schnitt zwischen $\tilde{t}_{FP,Ist}$ und $\widetilde{Fahrt}_K$ wird der Zugehörigkeitsgrad zwischen dem MS und dem ÖV-Fahrzeug *Fahrt* $Z_{\widetilde{MS}_t,K,Fahrt} \in [0,1]$ in der Kachel K zum Zeitpunkt $t_{\widetilde{MS}}$ berechnet mit:

$$Z_{\widetilde{MS}_t,K,Fahrt} = \text{height}\left(\tilde{t}_{FP,Ist} \cap \widetilde{Fahrt}_K\right). \tag{4.101}$$

Mit der in Abschnitt 4.2.6.2 bestimmten maximalen Aufenthaltszugehörigkeit des MS in dieser Kachel $Z_{K,\widetilde{MS}}$, folgt schließlich der Zugehörigkeitsgrad $Z_{\widetilde{MS},K,Fahrt}$, der sowohl die zeitliche als auch die räumliche Korrelation zwischen MS und *Fahrt* berücksichtigt. Da sowohl $Z_{K,\widetilde{MS}}$ als auch $Z_{\widetilde{MS},K,Fahrt}$ Singleton sind, besteht eine hohe Ähnlichkeit zum probabilistisch begründeten Rechenverfahren der Formel (4.95):

$$Z_{\widetilde{MS},K,Fahrt} = Z_{K,\widetilde{MS}} \cdot Z_{t_{\widetilde{MS}},K,Fahrt}. \tag{4.102}$$

Die Zugehörigkeitswerte $Z_{\widetilde{MS},K,Fahrt}$ sind damit „scharfe“ Zahlen, die für alle Kacheln je ÖV-Fahrzeug *Fahrt* der MS-Position zu berechnen sind. Von ihnen interessiert je *Fahrt* der maximale Zugehörigkeitswert:

$$Z_{\widetilde{MS},Fahrt} = \max_{\forall K}\left(Z_{\widetilde{MS},K,Fahrt}\right). \tag{4.103}$$

Für die binäre Entscheidung der Korrelation des MS mit *Fahrt* ist $Z_{\widetilde{MS},Fahrt}$ mit der kritischen Grenze $Z_{ÖV,min}$ zu vergleichen:

$$\begin{aligned} &\text{Fall I:} \quad Z_{\widetilde{MS},Fahrt} \geq Z_{ÖV,min} \quad \text{„MS ist an dieser Position mit } Fahrt \text{ korreliert"} \\ &\text{Fall II:} \quad Z_{\widetilde{MS},Fahrt} < Z_{ÖV,min} \quad \text{„MS ist an dieser Position nicht mit } Fahrt \text{ korreliert".} \end{aligned} \tag{4.104}$$

Zur Recheneffizienzsteigerung kann sich räumlich auf die Kacheln mit $Z_{K,\widetilde{MS}} > 0$ beschränkt werden, vgl. Abschnitt 4.2.6.2. Zeitlich genügen diejenigen Zeitfenster, die für t_1 und t_2 folgende Bedingung erfüllen:

$$\left(t_1 < t_3\right) \vee \left(t_0 < t_2\right) \text{ mit: } t_0 = \tilde{t}_{FP,Ist,min},\ t_3 = \tilde{t}_{FP,Ist,max}. \tag{4.105}$$

Die Parameterwerte der mit Fuzzy-Intervallen formulierten Zugehörigkeitsfunktionen $m_{\widetilde{\Delta t}_{PlanAbw}}$ für die eingangs eingeführten unscharfen Fahrplanabweichungen $\widetilde{\Delta t}_{PlanAbw}$ sind in der folgenden Tabelle 4-4 zusammengefasst. Wie für die Bestimmung der Funktionsparameterwerte der Normalverteilung in Abschnitt 4.5.3.1, wurde für die Bestimmung der Zugehörigkeitsfunktion von G^{DB} ebenfalls auf die in HERMANN (1996, S. 160f) ermittelten Parameter zurückgegriffen, vgl. dazu auch Anhang D. Mit Annahme, dass die Fahrplanabweichungen mit den dort aufgeführten Parameterwerten normalverteilt sind, wurde die Form der Zugehörigkeitsfunktion mittels zweier α-Schnitte näherungsweise an die Normalverteilungsfunktion angepasst. Aufgrund der Symmetrien haben daher für G^{DB} sowohl der Mittelpunkt $m_{\widetilde{\Delta t}_{PlanAbw}}$ der Zugehörigkeitsfunktion und der Erwartungswert $\mu_{\underline{\Delta t}_{PlanAbw}}$ der Normalverteilung als auch die Spannweiten $c_{l,\widetilde{\Delta t}_{PlanAbw}}$ und $c_{r,\widetilde{\Delta t}_{PlanAbw}}$ gleiche Werte.

Tabelle 4-4: Parameterwerte der Zugehörigkeitsfunktionen der Fahrplanabweichungen

Verkehrsnetz	$m_{\widetilde{\Delta t}_{PlanAbw}}$ [sec]	$r_{\widetilde{\Delta t}_{PlanAbw}}$ [sec]	$c_{l,\widetilde{\Delta t}_{PlanAbw}}$ [sec]	$c_{r,\widetilde{\Delta t}_{PlanAbw}}$ [sec]
G^{Bus}	-70,0	40,0	120,0	110,0
G^{Tram}	15,0	15,0	150,0	60,0
G^{DB}	-192,8	228,0	504,5	504,5

4.5.3.3 Fuzzy-Randomness der Fahrplanabweichungen

Im Rahmen der Unsicherheitsbetrachtung mit FRV können prinzipiell sowohl Anteile der zufälligen Variabilität als auch unscharfe Anteile für die Fahrplanabweichung modelliert werden. Eine derartige Aufspaltung der Unsicherheiten ist anhand der empirisch erhobenen Fahrplanabweichungen im Rahmen dieser Arbeit jedoch nicht genügend motivierbar. Korrektionen oder Reduktionen mit unscharfen Parametern, wodurch im Allgemeinen die Einführung von Unschärfeanteile einer FRV begründet wird, fanden ebenfalls nicht statt. Aus diesem Grund beschränkt sich die vorliegende Arbeit auf die Modellierung der Fahrplanabweichungen mittels der zufälligen Variabilität $\underline{\Delta t}_{PlanAbw} \sim N\left(\mu_{\underline{\Delta t}_{PlanAbw}}, \sigma^2_{\underline{\Delta t}_{PlanAbw}}\right)$, wie sie bereits in Abschnitt 4.5.3.1 eingeführt wurde. Die gemeinsame Fortpflanzung der Unsicherheitsanteile der FRV aus den MS-Positionen bleibt davon unberührt, so dass für die zeitliche Komponente der FRV $\tilde{\underline{t}}_{FP,Ist}$ folgt:

$$\begin{aligned} &s^2_{\tilde{\underline{t}}_{FP,Ist}} = s^2_{t_{\widetilde{\underline{MS}}}} + \sigma^2_{\underline{\Delta t}_{PlanAbw}} \quad \text{und} \quad f_{\tilde{\underline{t}}_{FP,Ist}} = f_{t_{\widetilde{\underline{MS}}}} && \text{Anteile der zufälligen Variabilität} \\ &m_{\tilde{\underline{t}}_{FP,Ist}}(t) = m_{t_{\widetilde{\underline{MS}}}}(t) \oplus \mu_{\underline{\Delta t}_{PlanAbw}} = \left[t_{\widetilde{\underline{MS}},m} + \mu_{\underline{\Delta t}_{PlanAbw}}, r_{t_{\widetilde{\underline{MS}}}}, c_{l,t_{\widetilde{\underline{MS}}}}, c_{r,t_{\widetilde{\underline{MS}}}} \right]_{LL} && \text{Anteile der Unschärfe.} \end{aligned} \tag{4.106}$$

Aus der α-schnittweisen Auswertung von $P\left(t_1 \le \tilde{\mu}_{\tilde{\underline{t}}_{FP,Ist},\alpha} \le t_2\right)_\alpha$ wird die unscharfe Wahrscheinlichkeit berechnet, dass $\tilde{\underline{t}}_{FP,Ist}$ innerhalb des geplanten Zeitfensters t_1 und t_2 des ÖV-Fahrzeugs *Fahrt* liegt. Da ausschließlich der maximale Wahrscheinlichkeitswert mit $P_{\widetilde{\underline{FP}},K,Fahrt}$ von Interesse ist, kann der Rechenweg mit Beschränkung auf $\tilde{\underline{t}}_{FP,Ist,min} = \min\left(\text{supp}\left(\tilde{\underline{t}}_{FP,Ist}\right)\right)$ und $\tilde{\underline{t}}_{FP,Ist,max} = \max\left(\text{supp}\left(\tilde{\underline{t}}_{FP,Ist}\right)\right)$ vereinfacht werden. Die maximale Wahrscheinlichkeit $P_{\widetilde{\underline{FP}},K,Fahrt}$, dass sich das MS im Zeitfenster des ÖV-Fahrzeugs in der Kachel K befindet, wird mit der Fläche der Wahrscheinlichkeitsdichtefunktion der Student-Verteilung zwischen t_1 und t_2 sowie dem Freiheitsgrad berechnet $f_{\tilde{\underline{t}}_{FP,Ist}}$ mit:

$$P_{\widetilde{\underline{FP}},K,Fahrt} = P_{t_{f_{\tilde{\underline{t}}_{FP,Ist}}}}\left(\frac{t_2 - \tilde{\underline{t}}_{FP,Ist,min}}{s_{\tilde{\underline{t}}_{FP,Ist}}}\right) - P_{t_{f_{\tilde{\underline{t}}_{FP,Ist}}}}\left(\frac{t_1 - \tilde{\underline{t}}_{FP,Ist,max}}{s_{\tilde{\underline{t}}_{FP,Ist}}}\right). \tag{4.107}$$

Dieser Wert ist für alle Kacheln K je *Fahrt* zu berechnen, woraufhin die weiteren Schritte analog zu den Formeln (4.95) und (4.96) sind. Mit der Aufenthaltswahrscheinlichkeit in der Kachel $P_{\widetilde{\underline{MS}},K}$ aus Abschnitt 4.2.6.3 folgt schließlich die Wahrscheinlichkeit $P_{\widetilde{\underline{MS}},K,Fahrt}$ für die zeitliche und räumliche Korrelation des MS mit dem ÖV-Fahrzeug *Fahrt* in der Kachel K. Die Wahrscheinlichkeit $P_{\widetilde{\underline{MS}},Fahrt}$, dass das MS mit dem *Fahrt* für die MS-Position zum Zeitpunkt $t_{\widetilde{\underline{\mathbf{MS}}}}$ korreliert ist, wird berechnet mit:

$$\begin{aligned} P_{\widetilde{\underline{MS}},K,Fahrt} &= P_{\widetilde{\underline{MS}},K} \cdot P_{\widetilde{\underline{FP}},K,Fahrt} \\ P_{\widetilde{\underline{MS}},Fahrt} &= \sum_{i=1}^{n} P_{\widetilde{\underline{MS}},K_i,Fahrt}. \end{aligned} \tag{4.108}$$

Mit der kritischen Grenze $P_{ÖV,min}$ erfolgt die binäre Entscheidung:

$$\begin{aligned} &\text{Fall I:} \quad P_{\widetilde{\underline{MS}},Fahrt} \ge P_{ÖV,min} \quad \text{„MS ist an dieser Position mit } \textit{Fahrt} \text{ korreliert"} \\ &\text{Fall II:} \quad P_{\widetilde{\underline{MS}},Fahrt} < P_{ÖV,min} \quad \text{„MS ist an dieser Position nicht mit } \textit{Fahrt} \text{ korreliert".} \end{aligned} \tag{4.109}$$

Da die Formeln (4.107) und (4.108) für über alle Kacheln und für alle Zeitfenster zu berechnen sind, analog zu Abschnitt 4.5.3.1, wird zur Effizienzsteigerung eine räumliche und zeitliche Beschränkung eingeführt. Räumlich werden nur diejenigen Kacheln mit Mindestaufenthaltswahrscheinlichkeit $P_{\widetilde{\underline{\mathbf{MS}}},K,min}$ betrachtet, vgl. Abschnitt 4.2.6.3. Zeitlich kann sich, analog zu Abschnitt 4.5.3.1 Formel (4.98), auf diejenigen Zeitfenster beschränkt werden, deren t_1 oder t_2 innerhalb einer erhöhten Sicherheitswahrscheinlichkeit mit z. B. $P_{K,ÖV,max} = 99{,}9\%$ liegen:

$$\begin{aligned} &\left(t_0 \le t_1\right) \vee \left(t_2 \le t_3\right) \quad \text{mit:} \quad t_0 = \tilde{\underline{t}}_{FP,Ist,min} - s_{\tilde{\underline{t}}_{FP,Ist}} \cdot t_{f_{\tilde{\underline{t}}_{FP,Ist}},P_{K,ÖV,max}} \\ &\qquad\qquad\qquad\qquad\quad t_3 = \tilde{\underline{t}}_{FP,Ist,max} + s_{\tilde{\underline{t}}_{FP,Ist}} \cdot t_{f_{\tilde{\underline{t}}_{FP,Ist}},P_{K,ÖV,max}}, \\ &\qquad\qquad\qquad\qquad\quad t_{f_{\widetilde{\underline{\mathbf{MS}}}},P_{K,ÖV,max}} : \text{Quantil der Student-Verteilung.} \end{aligned} \tag{4.110}$$

4.5.3.4 ÖV-Korrelation für eine MS-Positionsfolge

Die bisherigen Verfahrensschritte betrachteten ausschließlich einzelne MS-Positionen. Mit Übergang auf die MS-Positionsfolge ist von Interesse, ob eine festgestellte Korrelation zu einem ÖV-Fahrzeug *Fahrt* nur ein Einzelereignis ist oder, ob sie durch mehrfache Wiederholung bestätigt wird. Zur Lösung dieser Fragestellung werden die folgenden drei Szenarien herangezogen, deren grafische Interpretation Abbildung 4-24 zeigt. Die MS-Positionsfolge entstand, während:

A) das MS im ÖV-Fahrzeug war,

B) mit dem MS in das ÖV-Fahrzeug eingestiegen wurde,

C) mit dem MS aus dem ÖV-Fahrzeug ausgestiegen wurde.

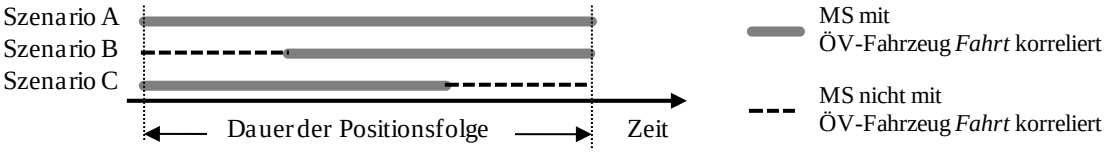


Abbildung 4-24: Szenarien für die Korrelationen der MS-Positionsfolge mit dem ÖV

Für die Szenarien B und C soll der Anteil der MS-Positionsfolge $\rho_{MS,Fahrt}$ mit Korrelation zu einem ÖV-Fahrzeug im Vergleich zur Gesamtdauer der Positionsfolge den kritischen Wert $\rho_{Krit.,MS,Fahrt} = 4/5$ nicht unterschreiten. Wird die Bedingung $\rho_{MS,Fahrt} > \rho_{Krit.,MS,Fahrt}$ für mindestens ein ÖV-Fahrzeug nachgewiesen, dann ist die Korrelation zum ÖV und speziell zum ÖV-Fahrzeug *Fahrt* genügend hoch, als dass ein Ausschluss des MS vom ÖV nicht gerechtfertigt werden kann. Das MS ist in diesem Fall mit dem ÖV korreliert und ist nicht gesichert vom ÖV unterscheidbar. Das MS ist somit für die Verkehrserfassung des MIV ungeeignet.

Gilt die Bedingung $\rho_{MS,Fahrt} \leq \rho_{Krit.,MS,Fahrt}$ für alle ÖV-Fahrzeuge, konnte kein ÖV-Fahrzeug nachgewiesen werden. Eine Korrelation zum ÖV kann ausgeschlossen werden. Das bedeutet, dass die Verkehrsmittel Linienbus, Straßenbahn sowie die Regional- und Fernbahnen als mögliche Transportfahrzeuge des MS auszuschließen sind:

$$\rho_{MS,Fahrt} \begin{Bmatrix} \leq \\ > \end{Bmatrix} \rho_{Krit.,MS,Fahrt} \quad \forall Fahrt \Rightarrow \begin{cases} \rightarrow \text{„Verkehrsmittel des ÖV ausschließen"} \\ \rightarrow \text{„Korrelation zum ÖV vorhanden, MS zur Verkehrserfassung des MIV ungeeignet".} \end{cases} \tag{4.111}$$

Umsteigeszenarien, also der Wechsel von einem in ein anderes ÖV-Fahrzeuge oder zwischen mehreren ÖV-Fahrzeugen, werden im Rahmen der vorliegenden Arbeit nicht betrachtet.

Die Überprüfung der Bedingungen in (4.111) wird für die drei Arten der Unsicherheitsmodellierung gleichermaßen angewendet.

4.6 Kartengestützte Plausibilitätsüberprüfung

Mit dem nachfolgenden Identifikationsverfahren wird geprüft, ob eine Einpassung der MS-Positionsfolge in Karten digitaler Verkehrsnetze unter Einhaltung der räumlichen, topologischen und zeitlichen Kriterien möglich ist. Der Grundgedanke des Identifikationsverfahrens stützt sich darauf, dass spezifische Verkehrsnetze von spezifischen Verkehrsmitteln genutzt werden (z. B. Straßennetze von PKW, Fußwege und für PKW gesperrte Wege von Fußgängern). Gelingt eine Einpassung der MS-Positionsfolge in ein Verkehrsnetz, kann für die MS die mögliche Nutzung mindestens eines dieser Verkehrsmittel und damit die Verkehrsmittelgruppe (z. B. MIV, NMIV) schlussgefolgert werden. Die Einpassung muss sowohl geometrisch, und damit auf dem Verkehrsnetz auch topologisch, als auch zeitlich anhand der möglichen Geschwindigkeiten in dem Verkehrsnetz gelingen. Gelingt keine Einpassung, können alle Verkehrsmittel des entsprechenden Verkehrsnetzes ausgeschlossen werden.

4.6.1 Ein Verfahren zur Trajektoriensuche mit Raum- und Zeitrestriktionen

Wesentliches Merkmal für eine gelungene Einpassung ist, dass auf dem Verkehrsnetz unter Einhaltung der Kriterien (z. B. die maximal erlaubte räumliche Abweichung der Positionsfolge von den Verkehrsnetzelementen oder die Übereinstimmung der resultierenden Topologie zwischen den gefundenen Verkehrsnetzelemente mit den Bedingungen des Verkehrsnetzes) eine **Route** gefunden wurde. Die räumliche und topologische Einpassung einer Positionsfolge als Route in ein Verkehrsnetz wird als **Map-Matching** bezeichnet (CZOMMER 2000; FRENCH 1989; ZHAO 1997). Die Aufgabenstellung des Map-Matchings wird in der vorliegenden Arbeit dadurch erweitert, dass die Verkehrsnetzelemente zusätzlich innerhalb vorgegebener Zeitfenster zu passieren sind. Diese Zeitfenster werden durch die Zeitabstände zwischen den aufeinanderfolgenden MS-Positionen definiert. Dieses zeitliche Kriterium unterscheidet die vorliegende Anwendung von denen des Map-Matching. Zur Abgrenzung vom räumlich und topologisch motivierten Begriff *Route* wird im Folgenden daher der Begriff **Trajektorie** verwendet. Es ist somit eine Trajektorie in das Verkehrsnetz einzupassen, die die Verkehrsnetzelemente je MS-Position innerhalb der, durch die MS-Positionen vorgegebenen, Zeitfenster passiert. Diese Aufgabenstellung wird mit dem Algorithmus nach DIJKSTRA (1959) gelöst, wobei die Zeitfensterprüfung zu ergänzen ist.

Ausgehend von einem vorgegebenen Startknoten löst der DIJKSTRA-Algorithmus das Problem der Suche nach den Pfaden mit den geringsten Widerständen, beispielsweise mit der geringsten Weglänge oder dem geringsten Zeitaufwand, zu einem oder allen Knoten in einem kantengewichteten Graphen. Übertragen auf die hier zu lösende Fragestellung, ist beginnend von der ersten MS-Position zwischen jeweils zwei aufeinanderfolgenden MS-Positionen bis zur letzten MS-Position eine Pfadsuche, also eine **Pfadsuche mit Zwischenpunkten** auszuführen. Berücksichtigt man zudem die MS-Positionsunsicherheiten, dann sind den jeweiligen MS-Positionen mehrere mögliche Start- und Zielknoten zuzuordnen. Ausgehend von m_1 Startknoten an der ersten MS-Position über jeweils m_i Knoten an den Zwischenpunkten $i = 2,\ldots,n-1$ bis hin zu m_n Zielknoten, wären bis zu $m_1 \cdot m_2 \cdot \ldots \cdot m_n$ Zwischenwege zu berechnen. Der resultierende Rechenaufwand ist inakzeptabel. Für die Anwendung des um die Zeitfensterprüfung ergänzten DIJKSTRA-Algorithmus, wird daher ebenfalls eine temporäre Veränderung des Verkehrsnetzes vorgeschlagen. Das auf diesen Vorgaben entwickelte **Verfahren zur Trajektoriensuche mit Raum und Zeitrestriktionen**, das prinzipiell zur Gruppe der Map-Matching-Verfahren zuzuordnen ist, wird im Folgenden näher erläutert.

Das Teilverkehrsnetz, das innerhalb der Zeitdifferenz zweier aufeinanderfolgender MS-Positionen MS_i und MS_{i+1} befahren werden kann, wird im Folgenden als **Korridor** i bezeichnet. Der Ablauf des Map-Matching-Verfahrens kann dann in die folgenden vier Schritte unterteilt werden:

Schritt 1: Zur temporären Veränderung des Verkehrsnetzes werden die aufeinander folgenden Korridore $i=1$ bis $i=n-1$ als neue eigene und somit nicht miteinander verknüpfte Verkehrsnetze betrachtet, so dass je Korridor der Teilgraph G_i aus dem Verkehrsnetzgraphen G selektiert wird $G_i \subseteq G \quad \forall i=1,\ldots n-1$, vgl. die durch die Korridore eingeschlossenen Verkehrsnetzteile in Abbildung 4-25a).

Schritt 2: Die Graphen der Korridore werden an denjenigen Knoten miteinander verknüpft, die jeweils der gleichen MS-Position angehören und deren (originärer) Ursprungsknoten derselbe war. Es werden also die Knoten der MS-Position i+1 im Korridor i mit den Knoten der MS-Position i+1 des Korridors i+1 topologisch mit Kanten $E_{V_p^{G_i} \times V_q^{G_{i+1}}} \left\{V_p^{G_i}, V_q^{G_{i+1}}\right\}$ verknüpft mit $V_p^{G_i} \times V_q^{G_{i+1}} \quad \forall V_p^{G_i} \in \left(G_i \cap MS_{i+1}\right)$ und $V_q^{G_{i+1}} \in \left(G_{i+1} \cap MS_{i+1}\right)$. Diese neu eingefügten Kanten erhalten die Fahrzeit Null $E_{V_p^{G_i} \times V_q^{G_{i+1}}} \left(\Delta t_E = 0\right) \quad \forall i=1,\ldots,n-1$, so dass aus den einzelnen Graphen G_i ein neuer, für die gesamte MS-Positionsfolge geltender Graph G_{MS} entsteht, vgl. die Korridore in Abbildung 4-25b) und deren schematisch angedeuteten Verknüpfungen durch die senkrechten gepunkteten Pfeile .

Schritt 3: Für die Zeitfenster des Korridors i gilt, dass dessen Knoten zwischen den Zeitpunkten der MS-Position i und der MS-Position $i+1$ „befahren“ werden dürfen. Unter Einbeziehung der MS-Positionsunsicherheiten ist der früheste Zeitpunkt $t_{i,\,von}$ und der späteste Zeitpunkt $t_{i,\,bis}$ zu bestimmen, deren Berechnung das Thema der nachfolgenden Abschnitte ist.

Schritt 4: Die Einpassung der MS-Positionsfolge erfolgt auf dem Graphen G_{MS}, wobei als Startknoten ein Knoten der ersten MS-Position MS_1, vgl. in der Abbildung 4-25b) die erste MS-Position, und als Zielknoten ein Knoten der letzten MS-Position MS_n, vgl. in der Abbildung 4-25b) die die vierte (letzte) MS-Position, für die gesuchte Trajektorie infrage kommt. Unter Einhaltung der Zeitfensterrestriktionen gilt die Einpassung dann als gelungen, wenn eine beliebig verlaufende Trajektorie zwischen mindestens einem Startknoten und einem Zielknoten gefunden werden kann. Zur Pfadsuche für die Trajektorie wird der um die Zeitfenster erweiterte DIJKSTRA-Algorithmus verwendet.

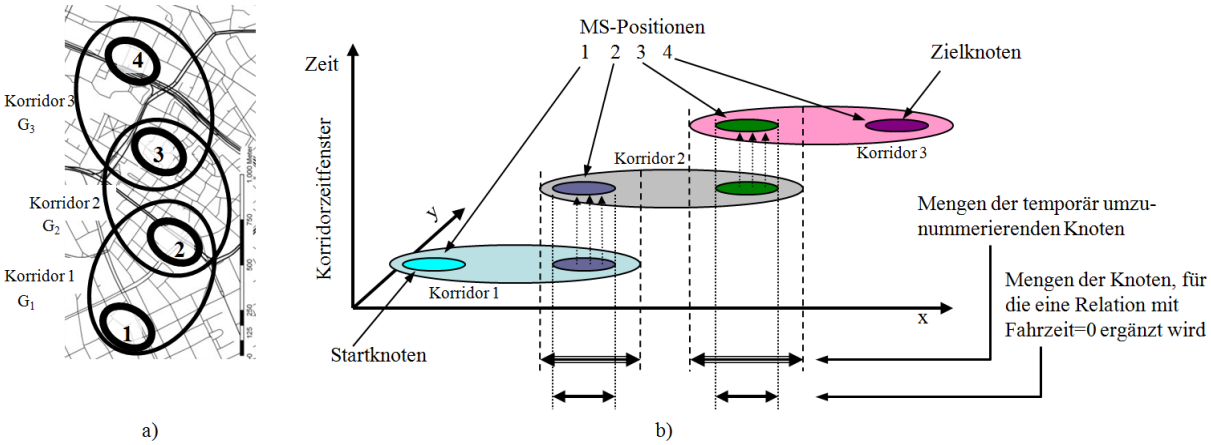


Abbildung 4-25: Beispiel für vier MS-Positionen a) eine geografische Darstellung des Verkehrsnetzes, der Ortungsflächen sowie der Korridore, und b) ein schematische räumliche und zeitliche Darstellung der Ortungsflächen und Korridore sowie Kennzeichnung der Überlappungsbereiche für die Umnummerierung selber Knoten und die Stellen für die Ergänzung der Relationen mit Fahrzeit Null der (geografisch) selben Knoten innerhalb der Ortungsflächen

Anmerkungen:

A) Mit der temporären Veränderung des Verkehrsnetzes kann die Trajektoriensuche zwischen den Startknoten m_1 von MS_1 zu den Zielknoten m_n von MS_n auf maximal $m_1 \cdot m_n$ Start-Zielknoten-Kombinationen reduziert werden. Der DIJKSTRA-Algorithmus baut ähnlich einer Baumstruktur die möglichen Pfade vom Startknoten zum Zielknoten auf, wonach der kürzeste dieser Pfade als Ergebnis

geliefert wird. Dazu wird nach jedem erreichten (Zwischen-)Knoten geprüft, ob er gleich dem gesuchten Zielknoten ist, um damit die Suche zu beenden. Für die vorliegende Anwendung kann diese Prüfung leicht gegen die Prüfung aller m_n Zielknoten von MS_n ersetzt werden. Die zu untersuchenden Start-Zielknoten-Kombinationen reduzieren sich damit auf die Anzahl m_1 der Startknoten.

B) Infolge der Bildung der unabhängigen Graphen G_i und der Zeitfenster sind auch Schleifen, also das mehrfache Überfahren derselben Verkehrsnetzelemente zu unterschiedlichen Zeiten möglich. Das schließt das Verbleiben des MS an Knoten über mehrere MS-Positionen hinweg mit ein, z. B. bei Stillstand.

Dieses Map-Matching-Verfahren ist für alle infrage kommenden Verkehrsnetze durchzuführen. Für die vorliegende Arbeit sind das die Verkehrsnetze $G^{Fußweg}$, $G^{Straße}_{LKW}$ und $G^{Straße}$, vgl. dazu auch die Tabelle 4-1 in Abschnitt 4.4. Abhängig vom Gelingen der Einpassung, können daraus die folgenden Schlüsse gezogen werden:

A) *Auf keinem Verkehrsnetz konnte eine Trajektorie gefunden werden*: Die entsprechenden Verkehrsmittel dieser Verkehrsnetze sind zu verwerfen. Die Mobilfunkdaten wurden entweder während der Benutzung eines unbekannten Verkehrsnetzes generiert oder die MS-Positionen spiegeln auch unter Berücksichtigung ihrer Unsicherheiten den Verlauf der Bewegung nicht genügend wider.

B) *Es konnte auf mindestens einem Verkehrsnetz eine Trajektorie gefunden werden:* Die von der MS verwendeten Verkehrsmittel, können auf diejenigen dieser Verkehrsnetze eingegrenzt werden. Die Verkehrsmittel der übrigen Verkehrsnetze sind zu verwerfen.

Die folgende Tabelle 4-5 zeigt die Wahrheitstabelle für die Entscheidung der Verkehrsmittelgruppe, die von der MS genutzt wurde, während die Mobilfunkdaten generiert wurden.

Tabelle 4-5: Wahrheitstabelle der möglichen Verkehrsmittelgruppen für die kartengestützte Plausibilitätsprüfung

Verkehrsnetz, auf dem mindestens eine Trajektorie gefunden wurde	von der MS mögliche verwendete Verkehrsmittelgruppe		
	NMIV	MIV	MIV außer LKW
(kein)	-	-	-
$G^{Fußweg}$	x	-	-
$G^{Straße}_{LKW}$	-	x	x
$G^{Straße}$	-	-	x
$G^{Fußweg} \wedge G^{Straße}_{LKW}$	x	x	x
$G^{Fußweg} \wedge G^{Straße}$	x	-	x
$G^{Straße}_{LKW} \wedge G^{Straße}$	-	x	x
$G^{Fußweg} \wedge G^{Straße}_{LKW} \wedge G^{Straße}$	x	x	x

Es bedeuten: - Ausschluss der Verkehrsmittelgruppe
x von der MS mögliche genutzte Verkehrsmittelgruppe

4.6.2 Auswahl der Teilgraphen und Bestimmung der Zeitfenster

Zur Bestimmung der Trajektorien werden im aktuellen Abschnitt die Bildung der Korridore, vgl. *Schritt 1*, und der Zeitfenster, vgl. *Schritt 3*, behandelt. Des Weiteren werden die Start- und Zielknoten der Korridore bestimmt.

Aus Gründen der höheren Arbeitsspeicher- und Recheneffizienz ist eine Beschränkung der zu behandelnden Verkehrsnetze auf diejenigen Teilgraphen empfehlenswert, in denen sich das MS (plausibel) aufgehalten haben

könnte. Mit der Annahme, dass sich ein Verkehrsteilnehmer maximal mit der im Verkehrsnetzgraph gestatteten Höchstgeschwindigkeit $v_{G,max}$ fortbewegt, kann hierfür die zu berücksichtigende **minimale Korridorgröße** bestimmt werden. Vereinfacht man dazu, dass die **maximal mögliche Weglänge** d_{max} von der ersten der beiden MS-Positionen des Korridors zur zweiten MS-Position über einen beliebigen Zwischenpunkt P mittels Luftlinienverbindungen verläuft, dann folgt die geometrische Figur des Korridors einer Ellipse, mit den beiden MS-Positionen in den Brennpunkten (KOHLEN 2004, S. 90f). Diese Ellipse wird im Folgenden als **Korridorellipse** $KoEll_i$ bezeichnet. Der Wert für d_{max} folgt aus der Zeitdifferenz zwischen den beiden MS-Positionen Δt mit:

$$d_{max} = v_{G,max} \cdot \Delta t. \tag{4.112}$$

Vereinfacht werden die räumlichen MS-Positionsunsicherheiten nachfolgend als Kreise approximiert und als **Ortungsflächen** OF_i mit den **Kreisradien** r_i und den **Kreismittelpunkten** $M_{OF,i}$ bezeichnet. Mit Berücksichtigung der räumlichen MS-Positionsunsicherheiten vergrößert sich die Korridorellipse. Die große Halbachse A_{KoEll_i} wird dementsprechend aus d_{max} und den Kreisradien berechnet, vgl. dazu zur Berechnung der weiteren Ellipsenparameter die Formeln in (4.113). Anderseits wird die Korridorellipse durch d_{max}, insbesondere bei kleinen Werten von d_{max}, wieder begrenzt (KOHLEN 2004, S. 90f). Ausgehend vom Rand der Ortungsflächen wird die maximal befahrbare Region durch die Distanz d_{max} je MS-Position begrenzt. Da es sich hierbei um eine Erweiterung der kreisförmigen Ortungsflächen handelt, wird die resultierende – ebenfalls kreisförmige – Fläche als **erweiterte Ortungsfläche** eOF_i mit **Kreisradius** $r_{OF,i}$ bezeichnet. Die aus dem Flächenschnitt von $KoEll_i$, eOF_i und eOF_{i+1} resultierende Figur wird als **Korridorfläche** KF_i bezeichnet und ist in Abbildung 4-26 fett umrahmt hervorgehoben, vgl. auch die Schnittmengenbeziehung in Formel (4.114). Die Verkehrsnetzelemente innerhalb der Korridorfläche bilden denjenigen Teilgraphen, der bei Höchstgeschwindigkeit des Verkehrsnetzgraphen vom MS innerhalb des Zeitabstandes zwischen MS_i und MS_{i+1} befahrbar ist. Dieser **Teilgraph** wird entsprechend mit G_i bezeichnet.

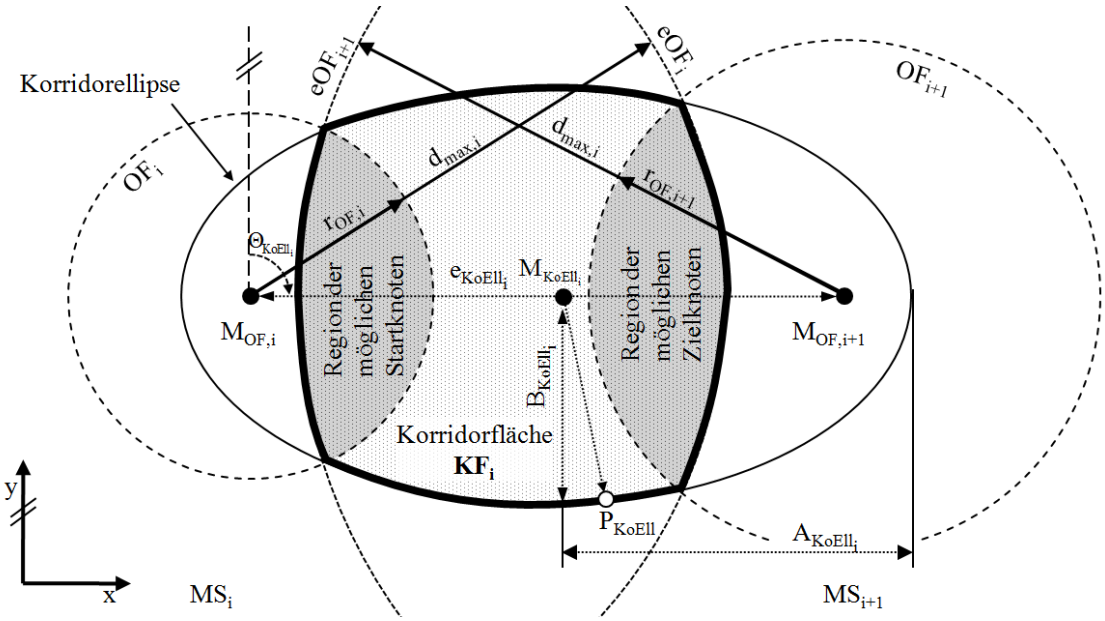


Abbildung 4-26: Korridorfläche mit Berücksichtigung der Ortungsunsicherheiten (nach KOHLEN 2004, S. 91)

$$\left.\begin{array}{lll} A_{KoEll_i} & = 0,5\cdot\left(d_{max,i} + r_{OF,i} + r_{OF,i+1}\right) & : \text{große Halbachse} \\ B_{KoEll_i} & = \sqrt{A_{KoEll}^2 - e_{KoEll}^2} & : \text{kleine Halbachse} \\ \Theta_{KoEll_i} & = \arctan\left(\left(M_{OF,y,i+1} - M_{OF,y,i}\right)/\left(M_{OF,x,i+1} - M_{OF,x,i}\right)\right) & : \text{Orientierung} \\ M_{KoEll_i}\begin{pmatrix} x \\ y \end{pmatrix} & = \begin{pmatrix} M_{OF,x,i} \\ M_{OF,y,i} \end{pmatrix} + e_{KoEll}\cdot\begin{pmatrix} \cos\Theta_{KoEll_i} \\ \sin\Theta_{KoEll_i} \end{pmatrix} & : \text{Ellipsenmittelpunkt} \\ P_{KoEll}\begin{pmatrix} x \\ y \end{pmatrix} & = \begin{pmatrix} M_{OF,x,i} \\ M_{OF,y,i} \end{pmatrix} + \dfrac{p_{KoEll_i}}{1-\varepsilon_{KoEll_i}\cdot\cos\left(\varphi-\Theta_{KoEll_i}\right)}\cdot\begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} & : \text{Punkt auf } KoEll_i \end{array}\right\} \quad (4.113)$$

$$\text{mit:}\quad e_{KoEll_i} = 0,5\cdot\sqrt{\left(M_{OF,x,i+1} - M_{OF,x,i}\right)^2 + \left(M_{OF,y,i+1} - M_{OF,y,i}\right)^2},$$
$$p_{KoEll_i} = B_{KoEll_i}^2 / A_{KoEll_i},\quad \varepsilon_{KoEll_i} = e_{KoEll_i} / A_{KoEll_i},\quad \varphi \in 0,\ldots 2\cdot\pi$$

In der Schnittfläche zwischen der Ortungsfläche OF_i mit der Korridorfläche KF_i befinden sich die möglichen Startknoten der Trajektorie auf diesem Teilgraphen G_i. Dementsprechend sind die Zielknoten in der Schnittfläche zwischen Korridorfläche und Ortungsfläche OF_{i+1} zu finden:

$$\left.\begin{array}{ll} \text{Korridorfläche}_i: & KF_i = eOF_i \cap \left(eOF_{i+1} \cap KoEll_i\right) \\ \text{Region der Startknoten}_i: & OF_i \cap KF_i \\ \text{Region der Zielknoten}_i: & OF_{i+1} \cap KF_i \end{array}\right\} \forall\, i = 1,\ldots n-1, \quad (4.114)$$

In der vorliegenden Arbeit dienen (aus Gründen der Recheneffizienz) die in Abschnitt 4.2.5 eingeführten Verkehrsnetzkacheln als kleinste räumliche Teilstücke zur räumlichen Selektion der Verkehrsnetzelemente. Verkehrsnetzelemente, deren Kacheln innerhalb der Korridorfläche liegen oder sie berühren, sind somit vollständig Bestandteil des Teilgraphen.

Die Zeitpunkte der Zeitfenster, die Zeitdifferenz Δt für die maximal mögliche Weglänge und die Ortungsflächenmittelpunkte sowie deren Radien sind unter Einbeziehung der MS-Positionsunsicherheiten zu bestimmen. Diese Themen werden in den nachfolgenden Abschnitten behandelt.

4.6.2.1 Zufällige Variabilität der Korridore

Im Falle der zufälligen Variabilität folgt der früheste Zeitpunkt $t_{i,von}$ und der späteste Zeitpunkt $t_{i,bis}$ aus:

$$t_{i,von} = t_{\underline{\mathbf{MS}},i} - t_{f_{\underline{\mathbf{MS}}_{t,i}},1-\alpha_{Irrtum}/2}\cdot s_{t_{\underline{\mathbf{MS}},i}} \quad \text{und} \quad t_{i,bis} = t_{\underline{\mathbf{MS}},i+1} + t_{f_{t_{\underline{\mathbf{MS}},i+1}},1-\alpha_{Irrtum}/2}\cdot s_{t_{\underline{\mathbf{MS}},i+1}}, \quad (4.115)$$

mit den Quantilen der *t*-Verteilung $t_{f_{\underline{MS}_{t,i}},1-\alpha_{Irrtum}/2}$ und $t_{f_{\underline{MS}_{t,i+1}},1-\alpha_{Irrtum}/2}$ für die Sicherheitswahrscheinlichkeit *P* der Zeitfenster. Der Maximalwert der Zeitdifferenz bei *P* folgt entsprechend mit der Varianzfortpflanzung aus:

$$\Delta t_i = \underline{\Delta t}_i + t_{f_{\underline{\Delta t}_i},1-\alpha_{Irrtum}/2}\cdot s_{\underline{\Delta t}_i} \quad \text{mit:} \quad \begin{array}{l} \underline{\Delta t}_i = t_{\underline{\mathbf{MS}},i+1} - t_{\underline{\mathbf{MS}},i} \\ s_{\underline{\Delta t}_i}^2 = s_{t_{\underline{\mathbf{MS}},i}}^2 + s_{t_{\underline{\mathbf{MS}},i+1}}^2 \\ f_{\underline{\Delta t}_i} = f_{\underline{\mathbf{MS}}_i} + f_{\underline{\mathbf{MS}}_{i+1}}. \end{array} \quad (4.116)$$

Für die gleiche Sicherheitswahrscheinlichkeit *P* werden auch die Konfidenzellipsen der jeweiligen MS-Positionen berechnet, vgl. dazu Abschnitt 4.2.6.1 Formel (4.46), aus denen die kreisförmig approximierten Ortungsflächen OF abgeleitet werden. Der Kreisradius sei dafür identisch mit der Länge der großen Halbachse der Konfidenzellipse $r_{\underline{OF},i} = A_{\underline{\mathbf{MS}}_i}$ und der Kreis- sowie der Ortungsflächenmittelpunkt identisch dem zufälligen Mittelpunkt $M_{\underline{OF},i}(x,y) = \left(\underline{\mathbf{MS}}_{x,i}, \underline{\mathbf{MS}}_{y,i}\right)$.

4.6.2.2 Unschärfe der Korridore

Bei unscharfen MS-Positionen folgen $t_{i,von}$ und $t_{i,bis}$ mit:

$$t_{i,von} = \min\left(t_{\widetilde{MS},i,\alpha=0}\right) = m_{t,\widetilde{MS}_i} - r_{t,\widetilde{MS}_i} - c_{l,t,\widetilde{MS}_i} \quad \text{und} \quad t_{i,bis} = \max\left(t_{\widetilde{MS},i+1,\alpha=0}\right) = m_{t,\widetilde{MS}_{i+1}} + r_{t,\widetilde{MS}_{i+1}} + c_{r,t,\widetilde{MS}_{i+1}} \quad (4.117)$$

Die maximal mögliche Zeitdifferenz Δt_i zwischen den MS-Positionen $\widetilde{MS}_i$ und $\widetilde{MS}_{i+1}$, folgt aus dem Maximum des Trägers der unscharfen Zeitdifferenz $\widetilde{\Delta t}_i$. Im Falle von LL-Fuzzy-Intervallen gilt dafür die Vereinfachung:

$$\Delta t_i = \max_{\forall\alpha\in[0,1]}\left(\widetilde{\Delta t}_i\right)_{LL} = \max\left(\tilde{t}_{\widetilde{MS},i+1,\alpha=0} \ominus \tilde{t}_{\widetilde{MS},i,\alpha=0}\right) = m_{t,\widetilde{MS}_{i+1}} - m_{t,\widetilde{MS}_i} + r_{t,\widetilde{MS}_i} + r_{t,\widetilde{MS}_{i+1}} + c_{l,t,\widetilde{MS}_i} + c_{r,t,\widetilde{MS}_{i+1}}. \quad (4.118)$$

Zur Bestimmung des approximierenden Kreises der jeweiligen MS-Position, dient der Flächenschwerpunkt des Trägers als Kreismittelpunkt $M_{OF,i}$. Für die verwendeten LL-Fuzzy-Intervalle der vorliegenden Arbeit lässt sich der Rechenweg auf die Bildung des Mittelwertes aus den Trägereckpunkten vereinfachen. Mit den maximal vier möglichen Eckpunkten der achsparallelen LL-Fuzzy-Intervalle folgt:

$$\left.\begin{aligned} &P_1(x,y) = \left(m_x - r_x - c_{l,x}, m_y - r_y - c_{l,y}\right),\ P_2(x,y) = \left(m_x - r_x - c_{l,x}, m_y + r_y + c_{r,y}\right) \\ &P_3(x,y) = \left(m_x + r_x + c_{r,x}, m_y + r_y + c_{r,y}\right),\ P_4(x,y) = \left(m_x + r_x + c_{r,x}, m_y - r_y - c_{l,y}\right) \end{aligned}\right\} \text{ Trägereckpunkte}$$

$$\left.\begin{aligned} M_{\widetilde{OF},i}(x,y) = \frac{1}{4}\cdot\sum_{j=1}^{4} P_j(x,y) = \left(m_x - \frac{1}{2}c_{l,x} + \frac{1}{2}c_{r,x} \quad , \quad m_y - \frac{1}{2}c_{l,y} + \frac{1}{2}c_{r,y}\right) \end{aligned}\right\} \text{ Schwerpunkt.} \quad (4.119)$$

Zwischen $M_{\widetilde{OF},i}$ und dem entferntesten Randpunkt zum Träger wird der Radius $r_{\widetilde{OF},i}$ des approximierenden Kreises aufgespannt. Aufgrund der zwischen $P_1,\ldots,P_4$ und $M_{\widetilde{OF},i}$ geltenden Symmetrien folgt:

$$r_{\widetilde{OF},i} = \max_{j=1,..,4}\sqrt{\left(x_{P_j} - x_{M_{OF,i}}\right)^2 + \left(y_{P_j} - y_{M_{OF,i}}\right)^2} = \sqrt{\left(x_P - x_{M_{OF,i}}\right)^2 + \left(y_P - y_{M_{OF,i}}\right)^2}, P \in P_1,\ldots,P_4. \quad (4.120)$$

4.6.2.3 Fuzzy-Randomness der Korridore

Im Falle des Fuzzy-Randomness wird sich bei der Sicherheitswahrscheinlichkeit *P* auf die zeitlichen und räumlichen Bereiche bei maximaler Unschärfe, also die Werte der Fuzzy-Intervalle bei $\alpha = 0$ beschränkt. Die früheste Abfahrzeit und späteste Ankunftszeit werden daraufhin aus der Kombination der jeweiligen Trägerenden des Fuzzy-Intervalls der Zeit mit den Grenzwerten aus der zufälligen Variabilität, wie sie bereits in Abschnitt 4.6.2.1 angewendet wurden, bestimmt:

$$\begin{aligned} t_{i,von} &= \underbrace{\min\left(\tilde{t}_{\underline{\widetilde{MS}},i,\alpha=0}\right)}_{\substack{\text{aus dem Unschärfeanteil be-}\\ \text{stimmter zufälliger Mittelpunkt}}} \underbrace{- t_{f_{\tilde{t}_{\underline{\widetilde{MS}},i,\alpha=0}},1-\alpha_{Irrtum}/2}\cdot s_{\tilde{t}_{\underline{\widetilde{MS}},i,\alpha=0}}}_{\substack{\text{für } P \text{ geltender Grenzwert des}\\ \text{Anteils der zufälligen Variabilität}}} \quad \text{und} \quad t_{i,bis} = \underbrace{\max\left(\tilde{t}_{\underline{\widetilde{MS}},i+1,\alpha=0}\right)}_{\substack{\text{aus dem Unschärfeanteil be-}\\ \text{stimmter zufälliger Mittelpunkt}}} \underbrace{+ t_{f_{\tilde{t}_{\underline{\widetilde{MS}},i+1,\alpha=0}},1-\alpha_{Irrtum}/2}\cdot s_{\tilde{t}_{\underline{\widetilde{MS}},i+1,\alpha=0}}}_{\substack{\text{für } P \text{ geltender Grenzwert des}\\ \text{Anteils der zufälligen Variabilität}}} \\ &= m_{\tilde{t}_{\underline{\widetilde{MS}},i}} - r_{\tilde{t}_{\underline{\widetilde{MS}},i}} - c_{l,\tilde{t}_{\underline{\widetilde{MS}},i}} - t_{f_{\tilde{t}_{\underline{\widetilde{MS}},i}},1-\alpha_{Irrtum}/2}\cdot s_{\tilde{t}_{\underline{\widetilde{MS}},i}} \qquad = m_{\tilde{t}_{\underline{\widetilde{MS}},i+1}} + r_{\tilde{t}_{\underline{\widetilde{MS}},i+1}} + c_{r,\tilde{t}_{\underline{\widetilde{MS}},i+1}} + t_{f_{\tilde{t}_{\underline{\widetilde{MS}},i+1}},1-\alpha_{Irrtum}/2}\cdot s_{\tilde{t}_{\underline{\widetilde{MS}},i+1}}. \end{aligned} \quad (4.121)$$

Mit der gemeinsamen Fortpflanzung der unscharfen und der zufälligen Unsicherheitsanteile folgt die FRV der Zeitdifferenz $\underline{\widetilde{\Delta t}}_i$ und daraus die maximal mögliche Zeitdifferenz $\Delta t_i = \max_{\alpha\in[0,1],P}\left(\underline{\widetilde{\Delta t}}_i\right)$ zwischen $\underline{\widetilde{MS}}_i$ und $\underline{\widetilde{MS}}_{i+1}$ bei *P*. Die Vorgehensweisen sind analog zu den Formeln (4.116) und (4.118):

$$\begin{aligned} \Delta t_i = \max_{\forall\alpha\in[0,1],P}\left(\tilde{t}_{\underline{\widetilde{MS}},i+1} \ominus \tilde{t}_{\underline{\widetilde{MS}},i}\right) &= \underbrace{\max_{\forall\alpha\in[0,1]}\left(\underline{\widetilde{\Delta t}}_i\right)_{LL}}_{\substack{\text{Anteil aus}\\ \text{Unschärfe}}} + \underbrace{t_{f_{\underline{\Delta t}_i},1-\alpha_{Irrtum}/2}\cdot s_{\underline{\Delta t}_i}}_{\substack{\text{Anteil aus}\\ \text{zufälliger Variabilität}}} \\ \text{mit}: \quad s^2_{\underline{\Delta t}_i} &= s^2_{\tilde{t}_{\underline{\widetilde{MS}},i}} + s^2_{\tilde{t}_{\underline{\widetilde{MS}},i+1}}, \quad f_{\underline{\Delta t}_i} = f_{\underline{\widetilde{MS}}_i} + f_{\underline{\widetilde{MS}}_{i+1}}, \\ \max_{\forall\alpha\in[0,1]}\left(\underline{\widetilde{\Delta t}}_i\right)_{LL} &= m_{\tilde{t}_{\underline{\widetilde{MS}},i+1}} - m_{\tilde{t}_{\underline{\widetilde{MS}},i}} + r_{\tilde{t}_{\underline{\widetilde{MS}},i+1}} + r_{\tilde{t}_{\underline{\widetilde{MS}},i}} + c_{r,\tilde{t}_{\underline{\widetilde{MS}},i+1}} + c_{l,\tilde{t}_{\underline{\widetilde{MS}},i}}. \end{aligned} \quad (4.122)$$

Zur Bestimmung der Ortungsflächen werden der Unsicherheitsanteil aus der zufälligen Variabilität analog zu Abschnitt 4.6.2.1 als approximierender Kreis der Konfidenzellipse bei *P* und der Unsicherheitsanteil aus der

Unschärfe als approximierender Kreis des Trägers kombiniert. Kreismittelpunkt $M_{\widetilde{OF},i}$ ist der Flächenschwerpunkt des Trägers, vgl. Formel (4.119):

$$\left.\begin{aligned} &P_1(x,y)=\left(m_x-r_x-c_{l,x},m_y-r_y-c_{l,y}\right),\ P_2(x,y)=\left(m_x-r_x-c_{l,x},m_y+r_y+c_{r,y}\right)\\ &P_3(x,y)=\left(m_x+r_x+c_{r,x},m_y+r_y+c_{r,y}\right),\ P_4(x,y)=\left(m_x+r_x+c_{r,x},m_y-r_y-c_{l,y}\right)\end{aligned}\right\}\ \text{Trägereckpunkte}$$

$$\left.\begin{aligned} M_{\widetilde{OF},i}(x,y)=\frac{1}{4}\cdot\sum_{j=1}^{4}P_j(x,y)=\left(m_x-\frac{1}{2}c_{l,x}+\frac{1}{2}c_{r,x}\quad,\quad m_y-\frac{1}{2}c_{l,y}+\frac{1}{2}c_{r,y}\right)\end{aligned}\right\}\ \text{Schwerpunkt.} \tag{4.123}$$

Der Kreisradius $r_{\widetilde{OF},i}$ setzt sich aus der großen Halbachse der Konfidenzellipse $A_{\widetilde{MS}_i}$, vgl. Abschnitt 4.2.6.1 Formel (4.46), sowie dem maximalen Abstand zwischen Kreismittelpunkt und dem entferntesten Trägereckpunkt, es geltenden die gleichen Symmetrien wie in Formel (4.120), zusammen:

$$r_{\widetilde{OF},i}=\underbrace{\sqrt{\left(x_P-x_{M_{\widetilde{OF},i}}\right)^2+\left(y_P-y_{M_{\widetilde{OF},i}}\right)^2}}_{\substack{P\in P_1,\ldots,P_4\\ \text{Approximation der Unschärfe mit dem Radius des Kreises, der den Träger vollständig einschließt}}}+\underbrace{A_{\widetilde{MS}_i}}_{\substack{\text{Approximation der zufälligen Variabilität mit dem Radius des Kreises, der die Konfidenzellipse vollständig einschließt = große Halbachse der Konfidenzellipse}}} \tag{4.124}$$

4.6.3 Anpassung der Maximalgeschwindigkeiten der Verkehrsnetzgraphen

Aus Gründen der Verkehrssicherheit und zur Verkehrssteuerung werden insbesondere in den Straßenverkehrsnetzen orts-, zum Teil auch (tages-)zeitabhängige zulässige Maximalgeschwindigkeiten administrativ festgelegt. Diese Vorinformationen sowie unter anderem auch empirisch bestimmte Geschwindigkeiten fließen bei der Anfertigung der digitalen Verkehrsnetzdaten in das Attribut der Maximalgeschwindigkeit der Kanten mit ein. Im Folgenden wird davon ausgegangen, dass sich ein überwiegender Teil der Verkehrsteilnehmer nicht schneller als mit maximal 20% Überschreitung dieses Attributwertes im entsprechenden Verkehrsnetz fortbewegt. Wenn im Folgenden von Maximalgeschwindigkeiten der Kanten gesprochen wird, dann ist implizit dieser Zuschlag berücksichtigt. Für die Maximalgeschwindigkeiten der verwendeten Verkehrsnetze folgen $v_{G^{Fußweg},max}=7,2\frac{km}{h}$, $v_{G^{Straße}_{LKW},max}=108\frac{km}{h}$ und $v_{G^{Straße},max}=300\frac{km}{h}$, vgl. die originären Werte der Tabelle 4-1 in Abschnitt 4.4. Die Maximalgeschwindigkeiten für $G^{Straße}_{LKW}$ und $G^{Straße}$ werden den jeweiligen maximalen Kantengeschwindigkeiten der Teilgraphen angepasst. Kantenspezifische Geschwindigkeiten sind für Fußgänger nicht ausreichend motivierbar. Daher wird für $G^{Fußweg}$ einheitlich für alle Kanten $v_{G^{Fußweg},E_v}=v_{G^{Fußweg},max}=7,2\frac{km}{h}$ als konstanter Wert angenommen.

4.6.4 Optimierungsverfahren zur Auswahl der Verkehrsnetzteilgraphen

Zur Bestimmung der Korridorellipse wurde eingangs als Maximalgeschwindigkeiten der maximal im Verkehrsnetz auftretende Geschwindigkeitswert $v_{G,max}$ angenommen. Nach der erstmaligen Berechnung der Korridorfläche kann dieser Wert mit dem höchsten Geschwindigkeitswert der Kanten E innerhalb des Korridors $v_{G,E_{v,max},i}=\max(E_v),\forall E_v\in Korridor_i$ ersetzt werden. Ist $v_{G,E_{v,max},i}$ kleiner als die zuvor verwendete Maximalgeschwindigkeit, dann wird auch $d_{max,i}$ und damit die Korridorfläche kleiner sein. Mit Wiederholung dieser Schritte ($v_{G,E_{v,max},i}$ in Korridor bestimmen, Korridorfläche neu berechnen) konvergiert dieses Verfahren gegen den kleinsten Wert, der höchsten Maximalgeschwindigkeit innerhalb des Teilgraphen, und damit gegen die kleinste Korridorfläche. Mit diesem in KOHLEN (2004, S. 92) vorgeschlagenen Verfahren, kann der Arbeitsspeicherbedarf sowie die Recheneffizienz zur Einpassung der MS-Positionsfolge in das Verkehrsnetz optimiert werden.

4.7 Verknüpfung der Identifizierungsergebnisse

Die in den vorherigen Abschnitten entwickelten Identifizierungsverfahren hatten das Ziel für eine MS-Positionsfolge nachzuweisen, dass sie eindeutige Merkmale des MIV aufweist. Konnte der Nachweis nicht oder nicht eindeutig erbracht werden, dann konnte das MS nicht als MIV identifiziert werden. Ausgangspunkt der Identifizierungsverfahren war der Nachweis für die Bewegung des MS in der Bewegungsidentifizierung. Mit der Feststellung das „MS steht“ wurde verbunden, dass eine Unterscheidung von MS, die beispielsweise der Gruppe „ohne Verkehrsbeteiligung“ (oVB) angehören, nicht mehr eindeutig vorgenommen werden kann. Dieser Ausschluss bedingt, dass beispielsweise in Verkehrsstaus stehende MS der Verkehrserfassung des MIV vorenthalten werden. Das Gleiche gilt für sich langsam bewegende MS, die aufgrund ihrer Geschwindigkeit nicht vom NMIV unterschieden werden können. Die Detektion von Verkehrsstaus ist jedoch eine relevante Aufgabenstellung der Verkehrslagebestimmung. Mit den bisherigen Darlegungen war somit eine objektive Identifizierung von MS in Verkehrsstaus nicht möglich. Das nachfolgend vorgestellte Verfahren setzt an dieser Stelle an und verwendet zur Identifikation derartiger MS einen subjektiven Ansatz nach SCHOLLMEYER & SCHWIEGER (2009).

Für MIV-Verkehrsmittel nutzende MS wird in diesem Ansatz angenommen, dass sie innerhalb eines mit $dt_{MIV,max}$ festgelegtem Zeitintervall keinen Verkehrsmittelwechsel vornehmen. „Stehend“ oder nicht eindeutig klassifizierte MS, können daraufhin ebenfalls als MIV eingestuft werden. In diesem Fall wird somit eine Übertragung der Identifikationsergebnisse von einer Positionsfolge auf eine zeitlich nachfolgende Positionsfolge und damit eine **Verknüpfung der Identifikationsergebnisse** vorgenommen. Die folgenden Bedingungen werden vorausgesetzt:

1. Für das MS muss eindeutig die Nutzung eines Verkehrsmittels des MIV in einer vorherigen Positionsfolge identifiziert worden sein. Die letzte Ortung dieser Positionsfolge muss innerhalb des mit $dt_{MIV,max}$ definierten Zeitintervalls vor der zu identifizierenden Positionsfolge liegen und
2. für gegebenenfalls zeitlich zwischenliegende Positionsfolgen darf die Nutzung keines anderen Verkehrsmittels identifiziert worden sein.
3. Die Position oder die Positionsfolge des MS muss (unter Einschluss der räumlichen Unsicherheiten) auf dem Verkehrsnetzgraph $G^{Straße}$ beziehungsweise $G_{LKW}^{Straße}$ liegen.

Die Abfolge der Identifizierungsverfahren mit den entsprechenden Verzweigungen zeigt Abbildung 4-27 für eine Epoche.

Wegen der Anonymität der Mobilfunkdaten ist eine objektive Bestimmung des Wertes für $dt_{MIV,max}$ nicht möglich, weshalb ein vom Auswerter subjektiv zu definierender Zeitwert zu wählen ist. In Abschnitt 6.4.3 werden in diesem Zusammenhang die Auswirkungen auf die Menge der Identifizierungsergebnisse für unterschiedlich hoch gewählte Zeitwerte von 0 Sekunden bis 10800 Sekunden untersucht.

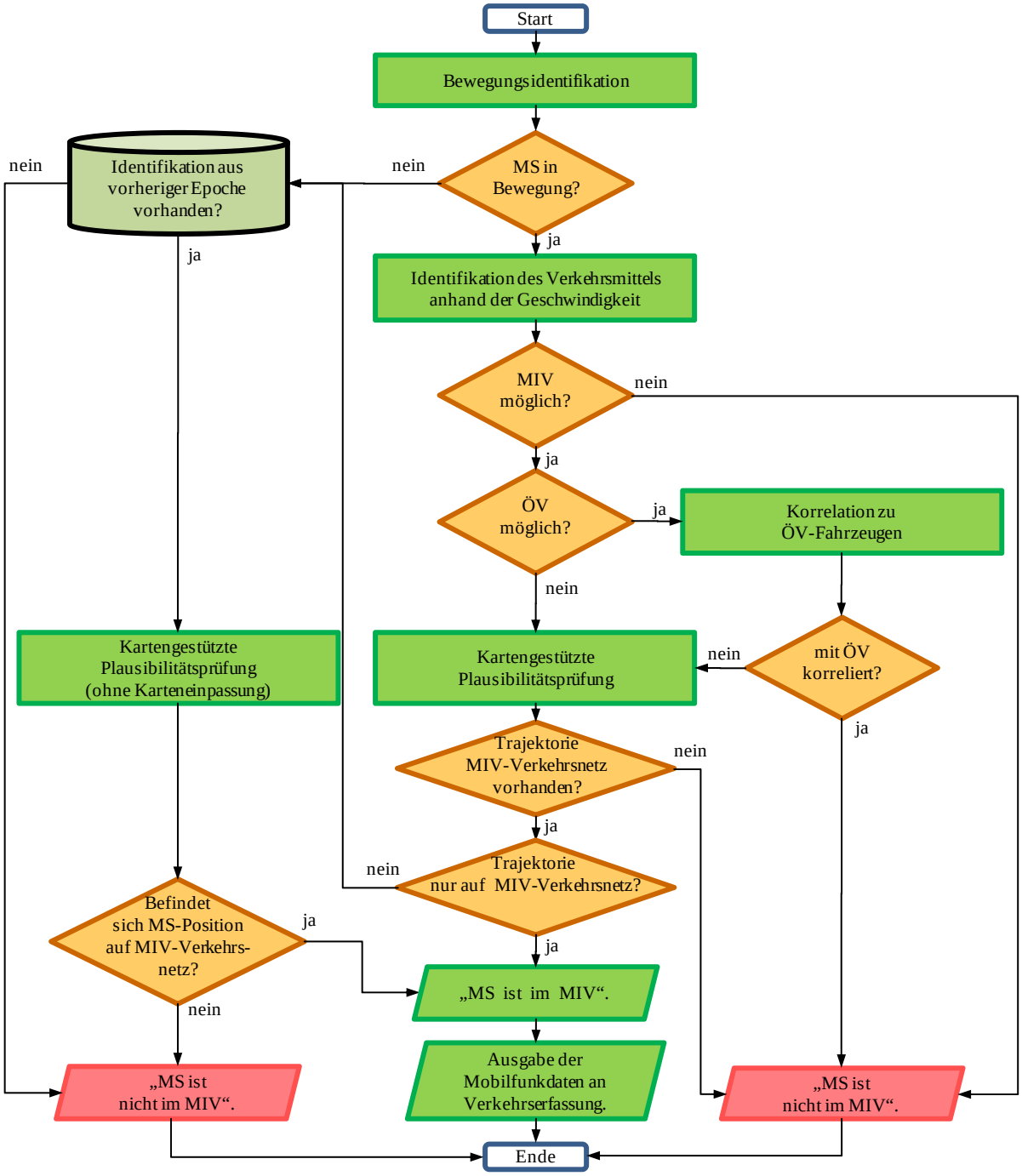


Abbildung 4-27: Ablauf der Identifizierung von Mobilfunkdaten mit Nutzung des MIV

5 Softwaretechnische Umsetzung

5.1 Erfassung und Anonymisierung der Mobilfunkdaten

Zur Validierung wurden die entwickelten Identifizierungsverfahren softwaretechnisch umgesetzt und in die im Projekt DO-IT erstellte Auswertekette zur Verkehrserfassung aus GSM-Mobilfunkdaten integriert. Die Abbildung 5-1 zeigt die Integration der Identifizierungsverfahren in den Datenfluss zur Gewinnung von Verkehrsdaten des MIV aus Mobilfunkdaten. Das Hauptaugenmerk zur softwaretechnischen Umsetzung lag in der Realisierung eines rechentechnisch selbstständig ablaufenden Softwareprogramms zur Identifizierung von Mobilfunkdaten des MIV. Die in der Abbildung 5-1 eingezeichneten Pfade „zufällige Variabilität“, „Fuzzy“ und „FRV“ verdeutlichen die in der vorliegenden Arbeit vorgenommenen wissenschaftlichen Untersuchungen.

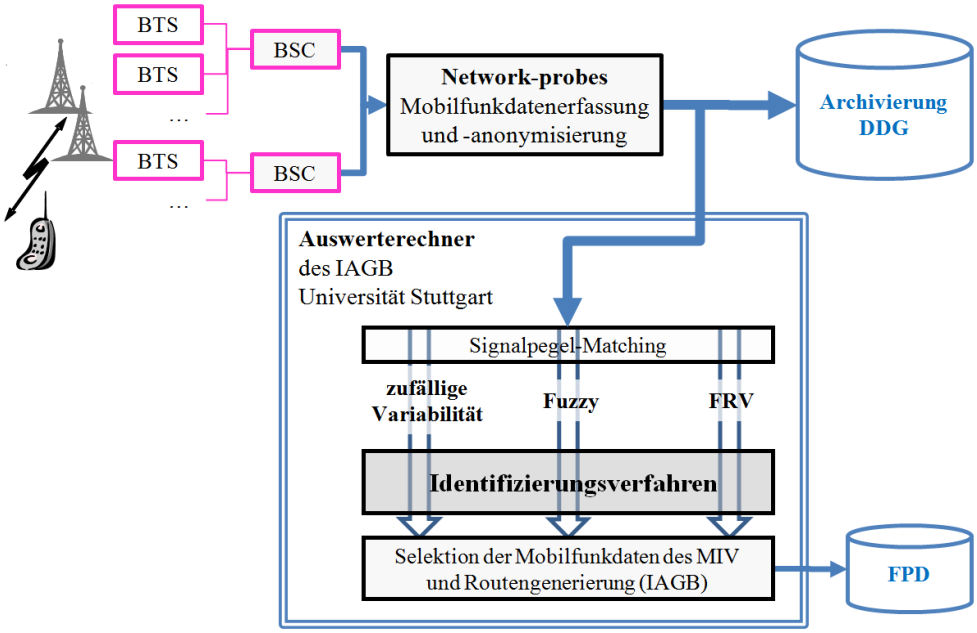


Abbildung 5-1: Übersicht der Softwareprogramme (schwarz eingerahmt) von der Mobilfunkdatenerfassung bis zur Erstellung der FPD

Zur Erfassung der Mobilfunkdaten wurden an den technischen Einrichtungen des GSM-Mobilfunknetzes an zwei ausgewählten BSC sogenannte **Network-Probes** installiert. Mit den Network-Probes wurden die Mobilfunkdaten der Abis- und der A-Schnittstelle des GSM-Mobilfunknetzes der T-Mobile Deutschland GmbH abgegriffen. Die Aufbereitung, Anonymisierung, Verarbeitung und Speicherung dieser Daten wurde im Rahmen des Projektes DO-IT mit speziellen, in der Programmiersprache C++ erstellten Softwareprogramme der Firma DDG[15] und deren Unterauftragnehmern vorgenommen. Diese Programme selektierten die für die Mobilfunkortung des Signalpegel-Matchings relevanten Mobilfunkdaten aus den Datenschnittstellen der Network-Probes und leiteten sie an die Auswerterechner des IAGB[16] der Universität Stuttgart weiter. Eine Zusammenstellung der für das eingesetzte Signalpegel-Matching erforderlichen Datentypen der Mobilfunkdaten befindet sich Anhang A.

[15] DDG Gesellschaft für Verkehrsdaten mbH

[16] Institut für Anwendungen der Geodäsie im Bauwesen, seit 2010 Institut für Ingenieurgeodäsie der Universität Stuttgart (IIGS)

5.2 Datenvorprozessierung und Datenhaltung

Zur Effizienzsteigerung der Identifizierungsverfahren wurden sowohl die verwendeten Fahrplandaten zur Korrelation mit ÖV-Fahrzeugen als auch das Verkehrsnetz für die kartengestützte Plausibilitätsprüfung im Vorfeld aufbereitet. Als wesentlicher Bearbeitungsschritt ist hierzu die Verkachelung der Verkehrsnetze zu nennen. Aufgrund der unterschiedlichen Suchanfragen bei der kartengestützten Plausibilitätsprüfung (z. B. die Suche nach den Knoten den Ortungsflächen, mit Knoten verknüpfte Nachbarknoten) wurde das Verkehrsnetz in einer relationalen Datenbank des Datenbankprogramms MySQL®[17] (MySQL DB) integriert. Zur weiteren Erhöhung der Effizienz wurden die Prognosen der Einfahrzeiten und der Ausfahrzeiten für die ÖV-Fahrzeuge ebenfalls im Vorhinein durchgeführt. Da die Datensuche auf die Suche nach Kachelnummern und anschließender Suche nach den zwischen t_0 und t_3 liegen Datensätzen beschränkt ist, wurden die prognostizierten Daten nach ihrer Kachelnummer sowie zeitlich sortiert und in Binärdateiform abgelegt. Der Datenumfang beträgt je prognostizierter Woche ca. 300 MB für alle Linienbusse und Straßenbahnen im Großraum Karlsruhe sowie Regional- und Fernbahnen im Nordwesten von Baden-Württemberg (vgl. die Beschreibung des Versuchsgebietes in Abschnitt 6.2).

Die Verwaltung der Mobilfunkdaten, der Ergebnisdaten der Mobilfunkortung sowie der Ergebnisse der Identifizierungsverfahren wurde in Datenbanken mit relationalen Tabellen in MySQL® vorgenommen. In Abbildung 5-2 ist die Organisation der Datenhaltung sowie die verwendete Speicherform für die in der vorliegenden Arbeit vorgenommene Auswertung der Mobilfunkdaten zusammenfassend dargestellt.

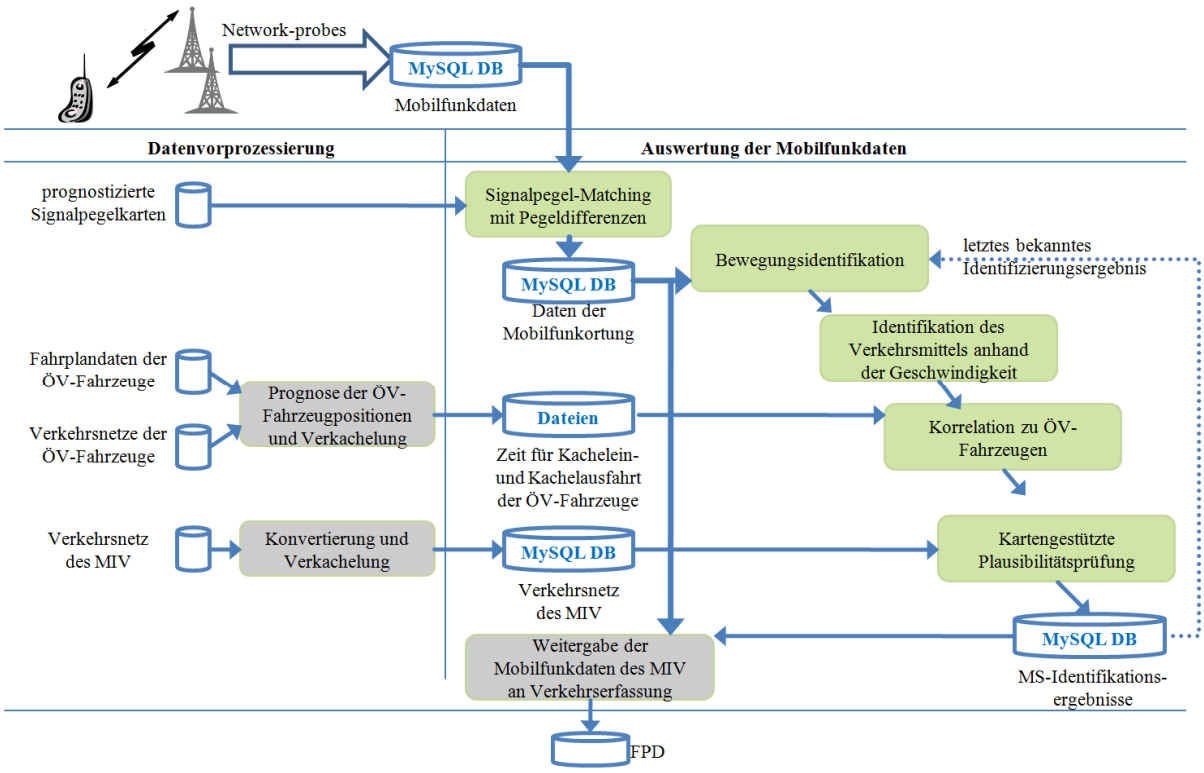


Abbildung 5-2: Verwaltung der Mobilfunkdaten und der für die Mobilfunkortung sowie die Identifizierung erforderlichen Daten einschließlich ihrer Ergebnisse

[17] Das Datenbankprogram MySQL® ist eine freie, unter der GPL (General Public License) (GNU GPL 2011) stehende Software.

5.3 Umsetzung der Mobilfunkortung und der Identifizierungsverfahren

Zur Mobilfunkortung wurde das Verfahren des Signalpegel-Matchings mit Pegeldifferenzen eingesetzt. Da dieses Verfahren im Wesentlichen Auswertungen auf Matrizenebene umfasst, wurde das für die Anwendung von Matrizen optimierte Programm MATLAB® The Mathworks™ in der „Version R2007a“ eingesetzt.

Für die softwaretechnische Umsetzung der Identifizierungsverfahren wurde die Programmierumgebung LabVIEW 8.5™ von National Instruments® gewählt. Die Routinen zur Berechnung des GK-FCM aus Abschnitt 4.2.4.7 und des Routings aus Abschnitt 4.6 wurden zur Erhöhung der Recheneffizienz in C und C++ umgesetzt und in das LabVIEW™ Programm eingebettet.

Die Auswertungen für die Validierung der Identifizierungsverfahren des nachfolgenden Abschnitts 6 wurden mit einem Desktop-PC durchgeführt (CPU: Intel® Core™2 Quad Q6600 @ 2,4 GHz, 8 GB RAM, Betriebssystem Microsoft Windows Vista™ 64 bit).

6 Validierung der Identifizierungsverfahren

6.1 Vorgehensweise

Grundlegendes Ziel der entwickelten Identifizierungsverfahren war die Selektion von Mobilfunkdaten, die in Verkehrsmitteln des MIV generiert wurden. Im vorliegenden Abschnitt werden die dazu angewendeten Unsicherheitsmodelle der zufälligen Variabilität, der Fuzzy-Theorie und des Fuzzy-Randomness gegenübergestellt und deren Identifikationsergebnisse beurteilt. Für den Nachweis der Praxistauglichkeit wurden in Feldversuchen Mobilfunkdaten erfasst und gleichzeitig Referenz-Informationen zu Ort und verwendetem Verkehrsmittel protokolliert. Zur Bestimmung der Qualität wurden die Identifikationsergebnisse (**IST**) mit den Referenz-Informationen (**SOLL**) verglichen. Als Qualitätsmaß der Identifikationsergebnisse diente die **Korrektheit**, die ein Maß derjenigen MS-Menge darstellt, die „korrekt“ ausgeschlossen wurden, da ihre Mobilfunkdaten nicht in Verkehrsmitteln des MIV generiert wurden:

$$\rho = \frac{n_{IST}^{\text{nicht MIV}}}{n_{SOLL}^{\text{nicht MIV}}}. \tag{6.1}$$

Mit n_{IST} wurde die Anzahl derjenigen identifizierten MS bezeichnet, die „korrekt“ nicht zur Verkehrsmittelgruppe des MIV zugeordnet wurden, da ihre Mobilfunkdaten nicht in Verkehrsmitteln des MIV generiert wurden. Mit n_{SOLL} wurde die verwendete Grundgesamtheit bezeichnet, die aus der tatsächlichen Menge (SOLL) der auszuschließenden MS bestand. Der Umfang der Grundgesamtheit war aufgrund der Feldversuche für jedes Identifizierungsverfahren unterschiedlich groß.

Aufgrund der geometrischen Unsicherheiten ist zu erwarten, dass mit steigender Korrektheit der Ergebnismenge der MS des MIV – die Menge der im Rahmen der Identifizierung klassifizierten MS, deren Mobilfunkdaten in Verkehrsmitteln des MIV generiert wurden –, die Größe dieser Ergebnismenge sinkt, da mehr MS zu Unrecht verworfen werden (Fehler 1. Art). Die Ursache ist darin zu finden, dass sich in Verkehrsmitteln des MIV generierte Mobilfunkdaten nicht genügend von denjenigen Mobilfunkdaten unterscheiden, die nicht in Verkehrsmitteln des MIV generiert wurden. Die Ergebnismenge der signifikant dem MIV zuzuordnenden MS sinkt somit bei höherer Korrektheit sowie höheren geometrischen Unsicherheiten. Dadurch sinkt gleichzeitig die **Effektivität** η der Identifizierungsverfahren, also die Menge der MS n_{IST}^{MIV}, die dem MIV zugeordnet wurden, im Verhältnis zur Menge der MS n_{SOLL}^{MIV}, deren Mobilfunkdaten tatsächlich in Verkehrsmitteln des MIV generiert wurden:

$$\eta = \frac{n_{IST}^{MIV}}{n_{SOLL}^{MIV}}. \tag{6.2}$$

In den vorgestellten Identifizierungsverfahren wurden kritische Grenzen für die Herbeiführung der Testentscheidungen eingeführt. Für die Anwendungen der zufälligen Variabilität wurden Sicherheitswahrscheinlichkeiten *P*, der Fuzzy-Theorie kritische Verwerfungsindizes ρ_{krit} und des Fuzzy-Randomness sowohl Sicherheitswahrscheinlichkeiten *P* als auch kritische Verwerfungsindizes ρ_{krit} eingesetzt. Ziel der nachfolgenden Analysen ist es daher, anhand von in Feldversuchen gewonnenen Mobilfunkdaten, geeignete Werte für diese Parameter zu finden, wodurch die Identifikationsverfahren bezüglich der in den Feldversuchen herrschenden Bedingungen **kalibriert** werden. Für die Kalibrierung der Identifikationsverfahren wurden die Parameterwerte so ausgewählt, dass Korrektheit und Effektivität der Identifizierungsergebnisse in einem ausgewogenen Verhältnis standen, also die gleichen Werte annahmen.

Bevor die Analysen der Identifikationsverfahren behandelt werden, werden in den nachfolgenden Abschnitten die zugrunde liegenden Daten und deren Unsicherheiten näher betrachtet. Das betrifft im Wesentlichen die Betrachtung der Mobilfunkpositionen und deren Positionsunsicherheiten einschließlich der aus dem Clustering erzielten Schwerpunkte. Im Anschluss werden die Kalibrierparameterwerte für die Identifikationsverfahren bestimmt und die damit erzielbaren Identifikationsergebnisse besprochen.

6.2 Versuchsanordnung

6.2.1 Versuchsgebiet

Zur Erfassung der Mobilfunkdaten wurden auf Seiten der Infrastruktur des Mobilfunkbetreibers T-Mobile Deutschland die anonymisierten Datenpakete der Abis-Schnittstelle der BTS abgegriffen. Eine Zusammenstellung der für das Signalpegel-Matching verwendeten Mobilfunkdatentypen befindet sich im Anhang A. Diese Daten werden im Wesentlichen dann generiert, wenn zwischen MS und Mobilfunkinfrastruktur eine Kommunikationsverbindung besteht (vgl. das Ortungsprinzip des Signalpegel-Matchings in Abschnitt 4.2.2). Mit den MS wurde daher während der Versuchsfahrten eine Sprachverbindung aufgebaut. Zusätzlich wurden die Positionen der MS mit einem Referenzortungssystem sowie das verwendete Verkehrsmittel protokolliert. Als Referenzortungssystem wurden GPS-Empfänger ohne Einsatz von Korrekturdaten verwendet. Die erzielten GPS-Ortungsgenauigkeiten lagen im Bereich von ca. 1-10 Metern.

Das Versuchsgebiet befand sich im Südwesten Deutschlands, im Großraum Karlsruhe und besaß eine Flächenausdehnung von ca. 345 km^2, vgl. dazu den Kartenausschnitt in Abbildung 6-1. In Rosa sind die Regionen dargestellt, an denen die zur Messdatenerfassung verwendeten BTS den maximalsten Signalpegel gegenüber benachbarten BTS aufwiesen. In der Regel wird die Datenfunkverbindung zwischen MS und Mobilfunkinfrastruktur zu derjenigen BTS hergestellt – bedienende BTS -, deren Signalpegel bei der MS am höchsten ist. Da von den bedienenden BTS die Daten der Abis-Schnittstelle abgegriffen werden, visualisieren die in Rosa hinterlegten Regionen näherungsweise die Erfassungsregionen der Mobilfunkdaten des untersuchten Versuchungsgebiets. Diese in Abbildung 6-1 als „GSM Bestserver Verfügbarkeit“ bezeichneten Regionen, wurden von der T-Mobile Deutschland mittels des Programmpakets PEGAPLAN® prognostiziert und visualisieren die Konfigurationen des Mobilfunknetzes der T-Mobile Deutschland im Herbst 2007.

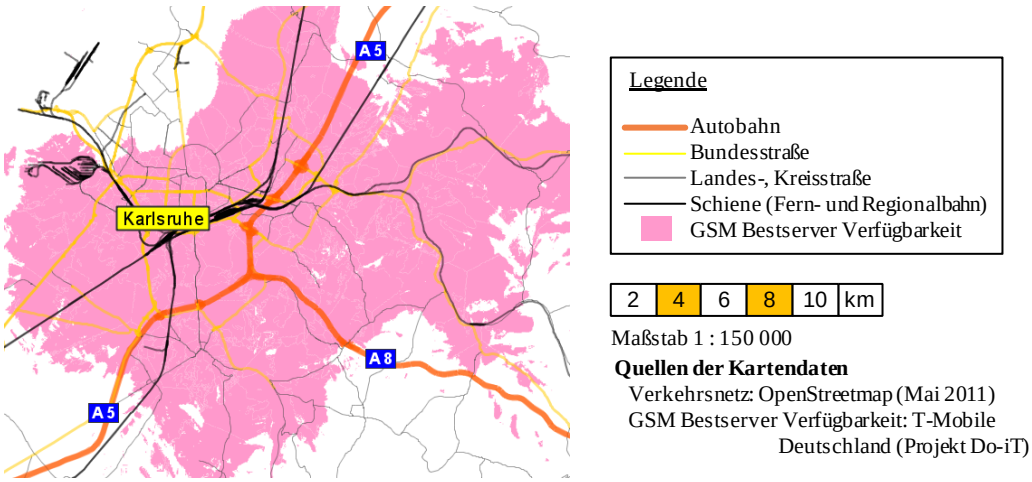


Abbildung 6-1: Versuchsgebiet zur Erfassung der Mobilfunkdaten im Raum Karlsruhe (Deutschland)

Insgesamt lagen den nachfolgenden Untersuchungen 89 Versuchsfahrten einschließlich Aufzeichnungen im Stand zugrunde, die im Rahmen des Projektes DO-IT (vgl. Abschnitt 2.3) erfasst wurden. Die Tabelle 6-1 zeigt

eine Zusammenstellung dieser Versuchsfahrten, die die Grundlage der nachfolgenden Analysen sind. Die Abbildung 6-2 veranschaulicht die Datenlängen der Versuchsfahrten in Sekunden und die Abbildung 6-3 zeigt korrespondierend dazu die metrischen Längen der Versuchsfahrten. Neben diesen Daten standen anonymisierte Mobilfunkdaten zur Verfügung, die im Versuchsgebiet am 09.07.2008 über einen Zeitraum von ca. 23 Stunden aufgezeichnet wurden. Die daraus berechneten 35 791 435 Mobilfunkortungen von 162 572 MS sind Gegenstand der Analysen in Abschnitt 6.4.

Tabelle 6-1: Versuchsfahrten mit Anzahl der Durchführung und Art des verwendeten Verkehrsmittels

Erfassungstag	Stand	Fußgänger	PKW	ÖV		
				Linienbus	Straßenbahn	DB
13.09.2007	8	-	41	-	-	-
07.11.2007	3	3	-	2	2	-
19.12.2007	-	12	-	-	-	11
27.02.2009	-	4	-	-	2	1
Summe:	**11**	**19**	**41**	**2**	**4**	**12**
Datenlänge*) [km]:	0,0	18,7	334,4	7,2	14,9	88,5
[Minuten]:	60,8	131,0	370,2	13,3	39,6	98,5

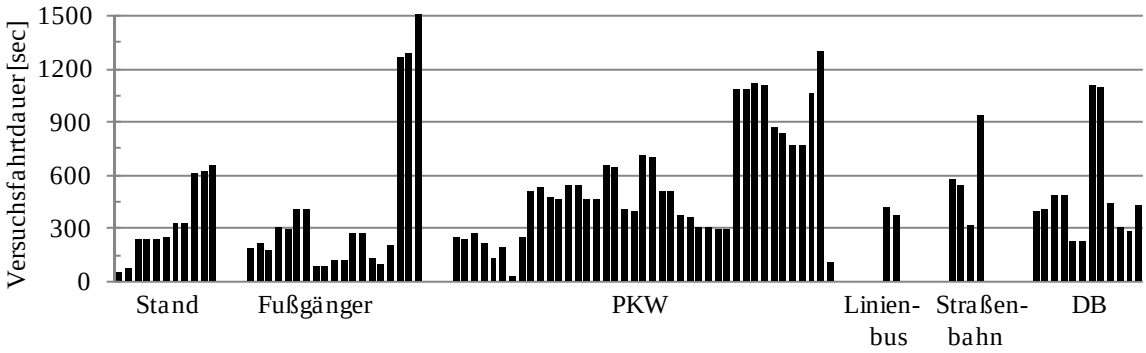


Abbildung 6-2: Dauer der Versuchsfahrten

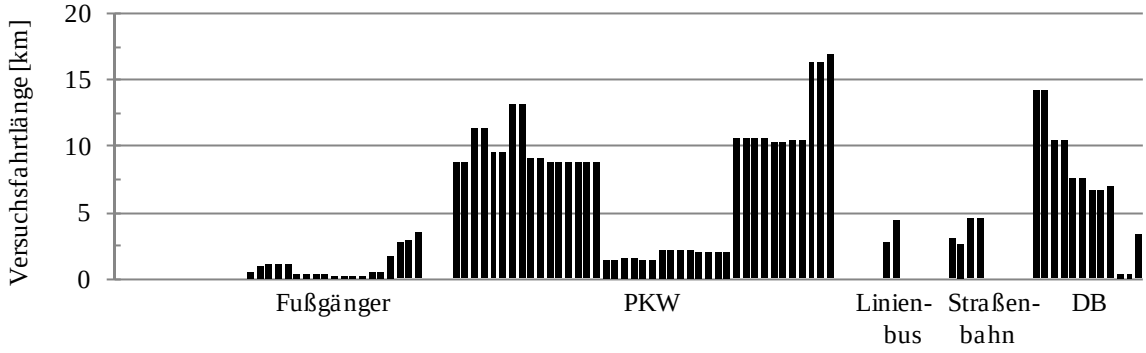


Abbildung 6-3: Metrische Längen der Versuchsfahrten

6.2.2 Geometrische Unsicherheiten des Signalpegel-Matchings

6.2.2.1 Prüfverfahren

Die Ergebnisse der Mobilfunkortung sollen die Positionen des MS und insbesondere die Unsicherheiten dieser Positionen geeignet widerspiegeln, um damit einen Hinweis auf die „wahre“ Position des MS zu geben. Zur Validierung der Mobilfunkortung wurden die geschätzten Positionen und deren Unsicherheiten aus dem Signalpegel-Matching mit Pegeldifferenzen des Abschnitts 4.2.2.4 sowie die geschätzten Positionen aus dem GK-FCM Clustering des Abschnitts 4.2.4 mit den GPS-Positionen verglichen. Die Unsicherheiten der GPS-Positionen wurden aufgrund ihrer deutlich geringen Größenordnung im Vergleich zu den Positionen der Mobilfunkortung und auch denjenigen des Clusterings im Folgenden vernachlässigt.

Für den Nachweis, ob sich die Mobilfunkortungsergebnisse zur Beschreibung der MS-Position eignen, wurde die Gültigkeit ihrer Unsicherheitsmaße überprüft. Hierzu wurden die Positionen aus der Mobilfunkortung und anschließend die Positionen aus dem Clustering mit den „wahren“ GPS-Positionen $\mathrm{MS}_{\mathrm{GPS}}(\mathrm{x},\mathrm{y})$ verglichen. Als Maß der Gültigkeit der Unsicherheitsangaben diente die Häufigkeit der Mobilfunkortungen, deren Positionsabweichungen bezüglich der „wahren“ Position des MS innerhalb ihrer geschätzten Unsicherheitsmaße lagen.

Im Falle der zufälligen Variabilität konnte die zu erwartende Häufigkeit für die Gültigkeit der Unsicherheitsangaben direkt aus der Sicherheitswahrscheinlichkeit P für den Signifikanztest der jeweiligen Abweichung des Ortungsergebnisses abgeleitet werden. Als Abweichungsmaß diente die Distanz d zwischen der MS-Position aus der Mobilfunkortung und der entsprechenden GPS-Position. Beispielweise sollten bei einer gewählten Sicherheitswahrscheinlichkeit von $P = 68{,}3\%$ mindestens $68{,}3\%$ der Abweichungen nicht-signifikant sein, wobei gleichzeitig bis zu $31{,}7\%$ der Mobilfunkortungen signifikante Abweichungen zur „wahren“ MS-Position aufweisen konnten. Werden für die gleichen MS-Positionen die Abweichungen für $P = 95\%$ geprüft, dann wäre zu erwarten, dass bis zu 5% der MS-Positionen höhere Abweichungen zu den GPS-Positionen aufweisen, vgl. dazu die Fälle 1 und 2 in Abbildung 6-4a)

Für unscharfe MS-Positionen wurde gefordert, dass die „wahre“ MS-Position innerhalb des Trägers von $\widetilde{\mathrm{MS}}$ lag. Die Unsicherheitsangabe war somit nur dann korrekt, wenn für die Mobilfunkortung die Bedingung $m_{\widetilde{\mathrm{MS}}}(\mathrm{MS}_{\mathrm{GPS}}) > 0$ erfüllt wurde.

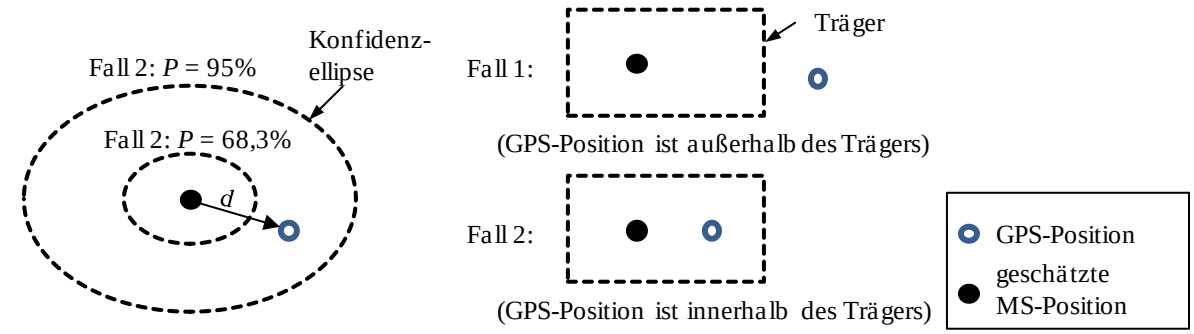


Abbildung 6-4: Darstellung der Abweichungen zwischen GPS-Position und MS-Position a) bei Wahl unterschiedlicher Sicherheitswahrscheinlichkeiten im Falle der zufälligen Variabilität und b) bezüglich des Trägers im Falle der Fuzzy-Theorie

Für die Anwendung des Fuzzy-Randomness wurde analog zur zufälligen Variabilität die zu erwartende Häufigkeit über die Wahl der Sicherheitswahrscheinlichkeit für den Signifikanztest der Distanz festgelegt. Als Abweichungsmaß dienten sowohl der minimale Distanzwert $\tilde{\underline{d}}_{\min} = \min\left(\mathbf{MS}_{GPS} - \widetilde{\underline{\mathbf{MS}}}\right)$ als auch der maximale Distanzwert $\tilde{\underline{d}}_{\max} = \max\left(\mathbf{MS}_{GPS} - \widetilde{\underline{\mathbf{MS}}}\right)$ zwischen der FRV der Mobilfunkortung und der GPS-Position. Die Abbildung 6-5a) zeigt an einem Beispiel die Figuren der Konfidenzellipse bei $P = 68,3\%$ und des Trägers der Zugehörigkeitsfunktion bei maximaler Unschärfe sowie die Kombination beider Figuren als erweiterten Konfidenzbereich. In dieser Abbildung ist zu erkennen, dass sich die GPS-Position außerhalb der erweiterten Konfidenzellipse befindet. Der Abstand d zwischen der GPS-Position (SOLL) und der MS-Position (IST) ist somit signifikant. Die Abbildung 6-5b) zeigt die gleiche Situation (ebenfalls auf die Eigenschaften für $\tilde{\underline{d}}_{\max}$ beschränkt), jedoch nun mit einer Sicherheitswahrscheinlichkeit von $P = 95\%$. Zu erkennen ist, dass die GPS-Position innerhalb der erweiterten Konfidenzellipse liegt und somit der Abstand zwischen GPS-Position und MS-Position nicht-signifikant ist.

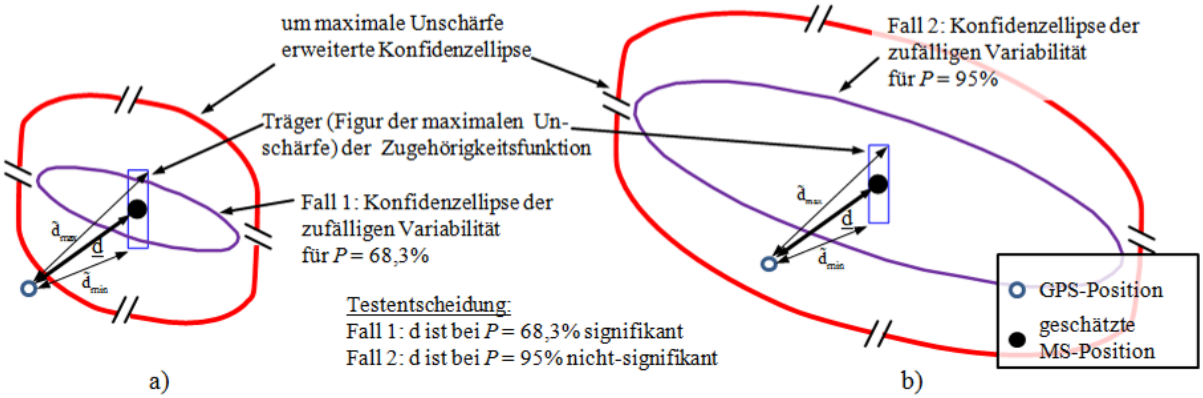


Abbildung 6-5: Darstellung der Abweichungen zwischen GPS-Position und MS-Position für unterschiedliche Sicherheitswahrscheinlichkeiten im Falle des Fuzzy-Randomness

6.2.2.2 Ergebnisse zur Prüfung der geometrischen Unsicherheitsangaben des Signalpegel-Matchings

Zur Überprüfung der Gültigkeit der Unsicherheitsangaben konnten insgesamt 68 554 Mobilfunkortungen mit den GPS-Positionen verglichen werden. Eine Unterscheidung nach Bewegungsverhalten oder Verkehrsmittelwahl fand im Vorfeld nicht statt.

Für die Positionen, deren Unsicherheiten mittels zufälliger Variabilität und Fuzzy-Randomness beschrieben wurden, zeigt die Abbildung 6-6 die Anzahl nicht-signifikanter Abweichungen $E\left(\underline{d}\right) = E\left(\left\|\mathbf{MS}_{GPS} - \underline{\mathbf{MS}}\right\|\right) = 0$ und $E\left(\tilde{\underline{d}}\right) = E\left(\left\|\mathbf{MS}_{GPS} - \widetilde{\underline{\mathbf{MS}}}\right\|\right) = 0$ in Abhängigkeit von der Sicherheitswahrscheinlichkeit. Eine durch die Mobilfunkortung zu optimistische Schätzung der Unsicherheit ist in den linken Bereichen dieser Abbildung zu erkennen, in denen die Anzahl nichtsignifikanter Abweichungen geringer als die erwartete Häufigkeit (schwarz eingezeichnete Linie) ist. Entsprechend zeigt eine höhere Anzahl nichtsignifikanter Abweichungen, dass die Unsicherheiten zu pessimistisch geschätzt wurden. Für beide Unsicherheitsmodellierungen sind sowohl zu optimistische, als auch zu pessimistische Schätzungen erkennbar.

Im Falle der zufälligen Variabilität zeigt sich aus dem Kurvenverlauf in Abbildung 6-6, dass Entscheidungen mittels Signifikanztests für Sicherheitswahrscheinlichkeiten von weniger als ca. $P = 83\%$ auf Grundlage zu optimistischen Varianzen getroffen werden. Werden Abstandsmaße aus den Positionen der zufälligen Variabilität bestimmt, um beispielweise die Gleichheit dieser Position zu prüfen, dann werden die Varianzen

dieser Abstandmaße zu optimistisch – die Zahlenwerte zu gering - geschätzt. Prinzipiell erhöht sich dadurch die Gefahr eines Fehlschlusses 1. Art. Bei der Signifikanzgrenze von 1σ $(P = 68{,}27\%)$ wurden ca. 39% der geschätzten Positionen mit zu optimistischen Kovarianzmatrizen angegeben. Bei Wahl höherer Sicherheitswahrscheinlichkeiten zeigt sich, dass die Varianzen zu pessimistisch geschätzt werden und damit die Gefahr eines Fehlschlusses 2. Art steigt. Bei der Signifikanzgrenze von 2σ $(P = 95{,}45\%)$ wurden für ca. 2% der Positionen zu pessimistische Kovarianzmatrizen geschätzt. Aus diesen Eigenschaften und auf Basis des Stichprobenumfangs kann angenommen werden, dass die geschätzten Positionen des Signalpegel-Matchings nur näherungsweise dem Funktionstyp der Normalverteilungsfunktion folgen.

Für die mit Hilfe der Fuzzy-Theorie formulierten unscharfen MS-Positionen befanden sich ca. 80,2% der GPS-Positionen innerhalb ihrer Träger. Dieser Wert ist zu Vergleichszwecken ebenfalls in Abbildung 6-6 dargestellt. Er ist jedoch nicht als von der Sicherheitswahrscheinlichkeit abhängige Größe zu verstehen. Es ist damit zu rechnen, dass ca. 20% der Positionen eine zu optimistische Unsicherheitsangabe (zu geringe Größe des Trägers) aufweisen. Durchschnittlich betrug der Zugehörigkeitswert der GPS-Positionen zu den Mobilfunkortungen 0,68.

Im Falle des Fuzzy-Randomness sind die durch die Unschärfen verursachten unterschiedlich hohen Häufigkeiten für die Minimalwerte $\underline{\tilde{d}}_{min}$ und Maximalwerte $\underline{\tilde{d}}_{max}$ der unscharfen Distanz in der Abbildung 6-6 deutlich erkennbar. Zur Beurteilung von Distanzen ist insbesondere der Minimalwert $\underline{\tilde{d}}_{min}$ von Interesse. Wird er im Rahmen von Signifikanztest (mit unscharfen Daten) eingesetzt, muss aufgrund einer zu optimistischen Schätzung der Unsicherheiten bei Sicherheitswahrscheinlichkeiten von weniger als $P = 76{,}5\%$ mit einer erhöhten Gefahr des Fehlschlusses 1. Art gerechnet werden. Bei der Signifikanzgrenze von 1σ besaßen ca. 19% der Positionen zu optimistische Unsicherheitsangaben. Die Gefahr des Fehlschlusses 2. Art ist bei Wahl größerer Sicherheitswahrscheinlichkeiten als $P = 76{,}5\%$ wegen der zu pessimistischen Unsicherheitsschätzung erhöht. Bei der Signifikanzgrenze von 2σ wurden für ca. 3% der Positionen zu pessimistische Unsicherheitsangaben bestimmt.

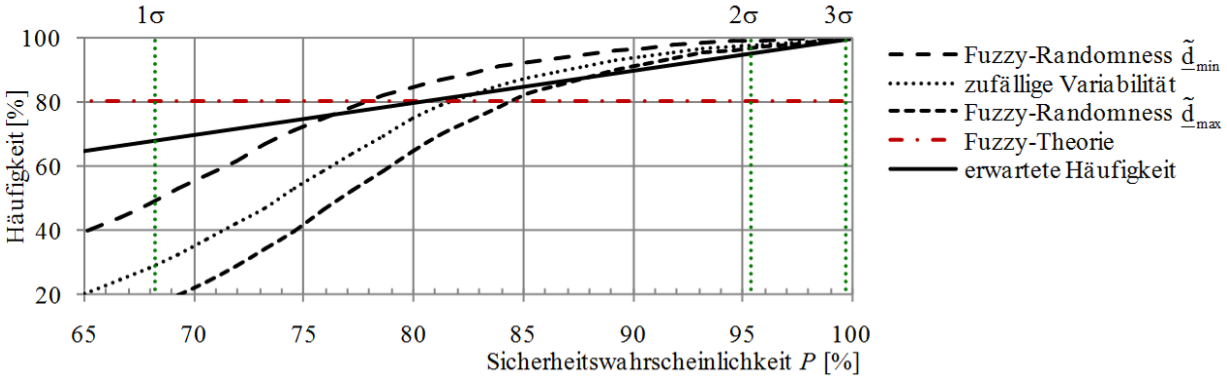


Abbildung 6-6: Häufigkeiten nicht-signifikanter Abweichungen der Positionen des Signalpegel-Matchings

6.2.2.3 Ergebnisse zur Prüfung der geometrischen Unsicherheitsangaben des GK-FCM-Clusterings

Für die Überprüfung der Unsicherheiten der MS-Clusterpositionen wurden die originären Positionen des Signalpegel-Matchings nach dem Clusteringverfahren des GK-FCM geclustert. Da aus mehreren Clustern bestehende Positionsfolgen nur für „MS in Bewegung“ gültig sind, wurde, mit Vorgriff auf die Untersuchungen des nachfolgenden Abschnitts, die *Bewegungsidentifikation* im Vorfeld durchführt. Daraufhin wurden für die Mobilfunkdaten mit der Identifizierung „MS steht“ ein einzelner Cluster und für Mobilfunkdaten mit der Identifizierung „MS in Bewegung“ eine Clusterfolge berechnet. Eine weiterführende Verkehrsmittel-unterscheidung fand nicht statt.

Im Falle der zufälligen Variabilität zeigt die Abbildung 6-7 die Häufigkeiten der nichtsignifikanten Abweichungen zu den GPS-Positionen. Die Grundgesamtheit bestand aus 1030 Positionsvergleichen. Im Vergleich zu den originären Positionen des Signalpegel-Matchings in Abbildung 6-6 erkennt man, dass die Unsicherheiten der zufälligen Variabilität zu optimistisch geschätzt wurden. An der oben genannten Grenze von ca. $P = 83\%$ zwischen den zu optimistischen und zu pessimistischen Unsicherheitsangaben waren nach der Clusterbildung nur noch ca. 70% der Cluster signifikant gleich mit den GPS-Positionen. Ca. 13% der Clusterpositionen hatten größere Abweichungen als erwartet, vgl. die eingezeichnete Stelle in Abbildung 6-7.

Eine Verbesserung zeigt sich im Falle der Fuzzy-Theorie (947 Positionsvergleiche), bei der sich ca. 93% der GPS-Positionen innerhalb der Träger der unscharfen Cluster befinden. Dieser Wert ist, wieder ohne Abhängigkeit zur Sicherheitswahrscheinlichkeit, in Abbildung 6-7 eingetragen. Im Vergleich zu den originären Positionen des Signalpegel-Matchings sind somit ca. 13% mehr Positionen mit den der Realität entsprechenden Trägerausmaßen geschätzt worden. Zu beachten ist, dass die hier durchführte Untersuchung ausschließlich zu klein (zu optimistisch) geschätzte Trägermaße aufzudecken vermag. Der durchschnittliche Zugehörigkeitswert der GPS-Positionen innerhalb der Träger beträgt ca. 0,834 und liegt ca. 15% höher als bei den originären Mobilfunkdaten.

Bei Anwendung des Fuzzy-Randomness (947 Positionsvergleiche) ist festzustellen, dass für einen Großteil der Clusterpositionen die Unsicherheiten im Vergleich zu den real auftretenden Abweichungen geeigneter geschätzt wurden als bei den originären Positionen des Signalpegel-Matchings, vgl. dazu den Verlauf der erwarteten Häufigkeiten und der Distanzminima $\tilde{d}_{min}$ in Abbildung 6-7 im Vergleich zu Abbildung 6-6. Die Anzahl der zu optimistischen Unsicherheitsangaben sank bei 1σ auf ca. 12% für die Clusterpositionen, vgl. die eingezeichnete Stelle an der 1σ Markierung in Abbildung 6-7. Bei 2σ wurden für ca. 2% der Clusterpositionen zu optimistische Unsicherheitsangaben geschätzt.

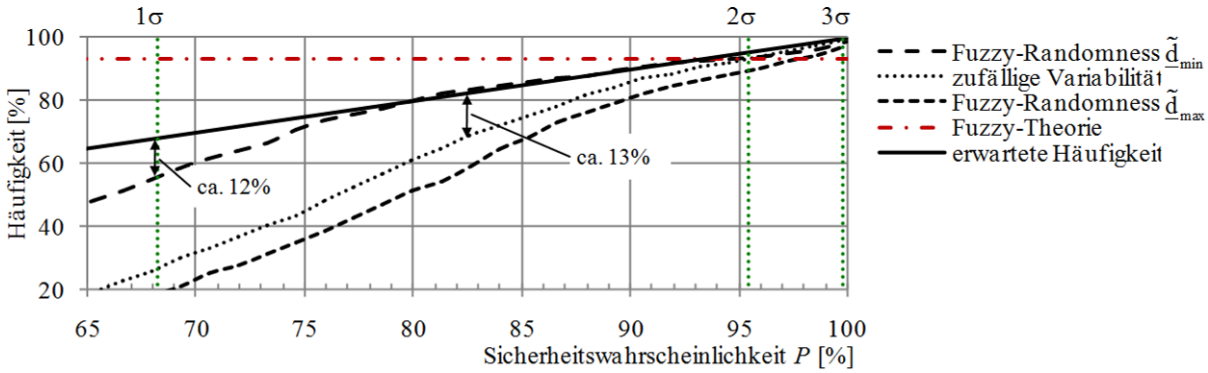


Abbildung 6-7: Häufigkeiten nichtsignifikanter Abweichungen der Positionen des GK-FCM-Clusterings

Durchschnittlich wurde eine Reduzierung der Anzahl der Positionen des Signalpegel-Matchings auf die Anzahl der Cluster auf ca. 1,6% erreicht. Neben der Entscheidung, ob das „MS steht“ (ein Cluster) oder in Bewegung ist (mehrere Cluster), wird dieser Wert im Wesentlichen durch Wahl der Clusteranzahl bestimmt (vgl. dazu Abschnitt 4.2.4.7).

6.3 Kalibrierung und Evaluation der Identifizierungsverfahren

6.3.1 Bewegungsidentifikation

Ziel der *Bewegungsidentifikation* ist es „MS in Bewegung“ zu identifizieren, um MS ohne Verkehrsbeteiligung (vgl. Verkehrsgruppe oVB in Abschnitt 2.1) herauszufiltern. Für diese Entscheidungsfindung ist die Festlegung der kritischen Grenzwerte ($\gamma_{\min,Bewegung}^{\text{zufällige Variabilität}}$ für die zufällige Variabilität, $\rho_{Bewegung}^{\text{Fuzzy-Theorie}}$ für die Fuzzy-Theorie sowie $\gamma_{\min,Bewegung}^{\text{Fuzzy-Randomness}}$ und $\rho_{Bewegung}^{\text{Fuzzy-Randomness}}$ im Falle des Fuzzy-Randomness) erforderlich. Anhand der Versuchsfahrten, vgl. Tabelle 6-1, werden im Folgenden geeignete Kalibrierparameterwerte für diese kritischen Grenzen bestimmt.

Insgesamt standen 89 Versuchsfahrten für die Bestimmung der Kalibrierparameterwerte zur Verfügung. Bei 11 Versuchsfahrten befand sich die MS im Stand $\left(n_{SOLL}^{\text{MS steht}}=11\right)$ und bei 78 Datensätzen wurden die Mobilfunkdaten während der Bewegung des MS $\left(n_{SOLL}^{\text{MS in Bewegung}}=78\right)$ erfasst. Die von der *Bewegungsidentifikation* korrekt als stehend identifizierten Versuchsfahrten, spiegelt die Anzahl $n_{IST}^{\text{MS steht}}$ wider. Die korrekt als bewegt identifizierten Versuchsfahrten, gibt $n_{IST}^{\text{MS in Bewegung}}$ an. Aus diesen Werten wurden die Korrektheit $\rho_{Bewegung}$ und die Effektivität $\eta_{Bewegung}$ wie folgt berechnet:

$$\rho_{Bewegung}=\frac{n_{IST}^{\text{MS steht}}}{n_{SOLL}^{\text{MS steht}}=11} \quad \text{und} \quad \eta_{Bewegung}=\frac{n_{IST}^{\text{MS in Bewegung}}}{n_{SOLL}^{\text{MS in Bewegung}}=78}. \tag{6.3}$$

In Abhängigkeit von der gewählten Mindestwahrscheinlichkeit zeigt die Abbildung 6-8a) im Falle der zufälligen Variabilität die Verläufe der Korrektheit und der Effektivität. Als Kalibrierparameterwert für die Sicherheitswahrscheinlichkeit wurde der Wert $\gamma_{\min,Bewegung}^{\text{zufällige Variabilität}}=56\%$ beim Schnittpunkt beider Kurven ausgewählt. An dieser Stelle wurden Werte für die Korrektheit und die Effektivität von ca. $\rho_{Bewegung}^{\text{zufällige Variabilität}}=\eta_{Bewegung}^{\text{zufällige Variabilität}}=81\%$ erreicht. Das heißt, etwa 19% der Mobilfunkdaten wurden zu Unrecht als „MS steht“ identifiziert. Aber das heißt auch, dass 19% der Mobilfunkdaten der tatsächlich stehenden MS als „MS in Bewegung“ identifiziert wurden.

Im Falle der Fuzzy-Theorie zeigt die Abbildung 6-8b) die Verläufe der Korrektheit und der Effektivität in Abhängigkeit vom kritischen Wert für den Verwerfungsindex $\rho_{Bewegung}^{\text{Fuzzy-Theorie}}$. Für den Kalibrierparameterwert wurde der Verwerfungsindex von $\rho_{Bewegung}^{\text{Fuzzy-Theorie}}=0,42$, wieder beim Schnittpunkt zwischen den Verläufen beider Kurven ausgewählt. Die Werte für die Korrektheit und die Effektivität betrugen bei diesem Schnittpunkt ca. $\rho_{Bewegung}^{\text{Fuzzy-Theorie}}=\eta_{Bewegung}^{\text{Fuzzy-Theorie}}=72\%$ und waren etwa 9% geringer als bei Anwendung der zufälligen Variabilität.

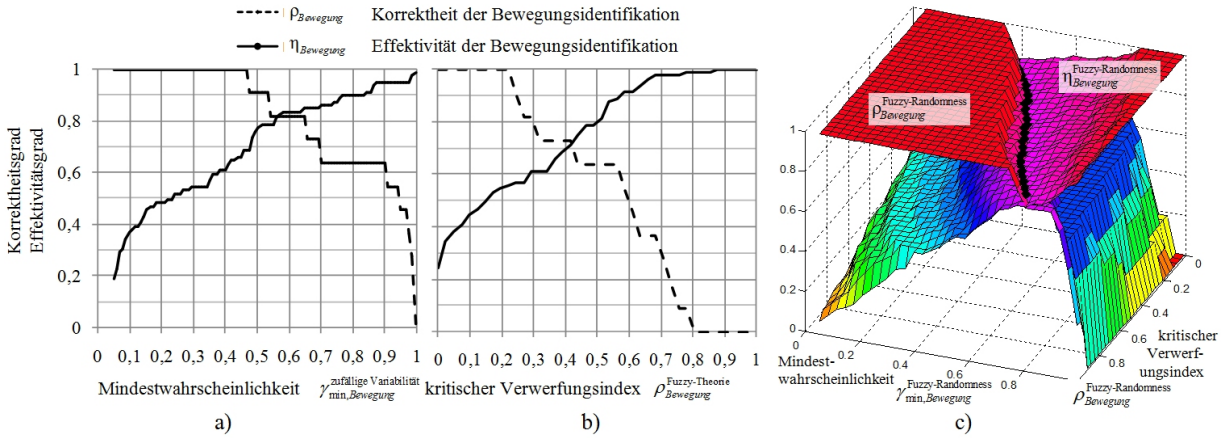


Abbildung 6-8: Verläufe der Korrektheit und Effektivität der Ergebnisse der Bewegungsidentifikation für die Anwendung a) der zufälligen Variabilität, b) der Fuzzy-Theorie und c) des Fuzzy-Randomness

Bei Anwendung des Fuzzy-Randomness ist die Testentscheidung sowohl vom Wert der Mindestwahrscheinlichkeit $\gamma_{\min,Bewegung}^{\text{Fuzzy-Randomness}}$ als auch vom Grenzwert des Verwerfungsindex $\rho_{Bewegung}^{\text{Fuzzy-Randomness}}$ abhängig. In Abhängigkeit von diesen Parametern zeigt die Abbildung 6-8c) die Verläufe der Korrektheit und der Effektivität. Analog zur Parameterwertebestimmung bei Anwendung der zufälligen Variabilität und der Fuzzy-Theorie wurde der Schnittpunkt beider Verläufe ausgewählt. Aufgrund der Abhängigkeit von Mindestwahrscheinlichkeit und Verwerfungsindex befindet sich dieser Schnittpunkt auf der Schnittlinie der beiden Kurven. Die in der Abbildung 6-8c) schwarz hervorgehobene Linie verdeutlicht diese Schnittlinie. Die Abbildung 6-9 zeigt eine verebnete Darstellung dieser Schnittlinie. Zu beachten ist, dass die Skala der Mindestwahrscheinlichkeit linear gewählt wurde, die Skalenwerte des Verwerfungsindex wurden hingegen aus den abgegriffenen Wertepaaren der Schnittpunkte generiert. Beim wertemäßigen Maximum für die Korrektheit und Effektivität auf dieser Schnittlinie bei ca. $\rho_{Bewegung}^{\text{Fuzzy-Randomness}} = \eta_{Bewegung}^{\text{Fuzzy-Randomness}} = 84\%$ wurden für die Kalibrierparameter die Werte $\gamma_{\min,Bewegung}^{\text{Fuzzy-Randomness}} = 61\%$ und $\rho_{Bewegung}^{\text{Fuzzy-Randomness}} = 0,50$ ausgewählt. Etwa 16% der Mobilfunkdaten wurden bei diesen Parameterwerten nicht als „MS steht“ erkannt (Korrektheit), aber ebenso viele zu Unrecht als „MS steht“ identifiziert (Effektivität). Die erzielten Korrektheitsgrade und Effektivtätsgrade waren ca. 3% höher als mit Anwendung der zufälligen Variabilität und ca. 12% höher als bei Anwendung der Fuzzy-Theorie. In Abbildung 6-9 ist zudem erkennbar, dass die Grade für die Korrektheit und Effektivität entlang der Schnittlinie näherungsweise gleich waren und somit Korrektheitsgrade von mindestens 82% ebenfalls bei anderen Wertepaaren erzielbar waren $\left(\text{z. B. } \left\{\gamma_{\min,Bewegung}^{\text{Fuzzy-Randomness}} ; \rho_{Bewegung}^{\text{Fuzzy-Randomness}}\right\} = \{0,45 ; 0,14\},\{0,55 ; 0,34\},\{0,75 ; 0,88\}\right)$.

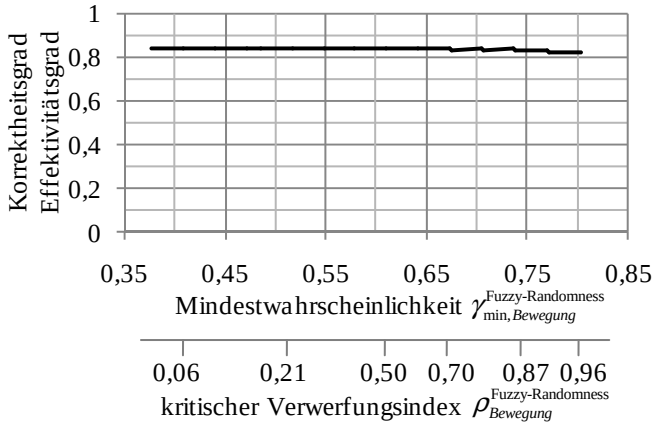


Abbildung 6-9: Verlauf der Korrektheit und Effektivität als verebnete Schnittlinie für die *Bewegungsidentifikation* im Falle des Fuzzy-Randomness

Augenscheinlich kann festgestellt werden, dass der Verlauf der Kurven in Abbildung 6-8a), b) und c) unstetig ist. Außerdem existieren Korrektheitsgrade von 1 und 0 über weite Parameterwertbereiche. Zurückzuführen sind diese Eigenschaften auf die zur Verfügung gestandenen Datenmengen für die Auswertungen im Rahmen der vorliegenden Arbeit. Insbesondere die geringe Anzahl von nur elf Versuchen für „MS steht“ wirkt sich hierbei auf die genannten Kurveneigenschaften aus.

6.3.2 Identifikation des Verkehrsmittels anhand der Geschwindigkeit

Bewegte sich das MS mit einer signifikant höheren Maximalgeschwindigkeiten als mit $v_{max,VM}$, dann wird mit der *Identifikation des Verkehrsmittels anhand der Geschwindigkeit* der Ausschluss des jeweiligen Verkehrsmittels schlussgefolgert. Die erforderlichen Werte für die verkehrsmittelspezifischen Maximalgeschwindigkeiten $v_{max,VM}$ wurden in Tabelle 4-1 im Abschnitt 4.4 festgelegt.

Für die Überprüfung dieses Identifikationsverfahrens wurden aus den GPS-Positionen der Versuchsfahrten die („wahren“) Maximalgeschwindigkeiten $v_{max,GPS}$ berechnet und damit die „wahre“ Menge der auszuschließenden Verkehrsmittel bestimmt. Für die Mobilfunkortungen derselben Versuchsfahrten wurde schließlich überprüft, ob mit ihnen, unter Einbeziehung ihrer Unsicherheiten, die gleichen Verkehrsmittel ausgeschlossen wurden. Beispielsweise besteht die Soll-Menge der auszuschließenden Verkehrsmittel für eine Versuchsfahrt mit $v_{max,GPS} = 127 \frac{km}{h}$ aus {Fußgänger ; Straßenbahn ; Linienbus ; LKW ; Reisebus}. Die SOLL-Menge der nicht-auszuschließenden Verkehrsmittel besteht in diesem Fall aus {PKW / KRAD ; Regionalbahn ; Fernbahn}.

Der Korrektheitsgrad für das korrekte Ausschließen der Verkehrsmittel und damit die Berechnung für die Korrektheit dieses Identifizierungsverfahrens $\rho_{Geschwindigkeit}$ folgte aus der Anzahl der korrekt ausgeschlossenen Verkehrsmittel $n_{IST}^{Ausschluss}$ und der „wahren“ Anzahl der auszuschließenden Verkehrsmittel $n_{SOLL}^{Ausschluss}$. Die Effektivität wurde aus der Anzahl der korrekt nicht ausgeschlossen Verkehrsmittel $n_{IST}^{Nicht\text{-}Ausschluss}$ und der „wahren“ Anzahl der nicht auszuschließenden Verkehrsmittel $n_{SOLL}^{Nicht\text{-}Ausschluss}$ berechnet mit:

$$\rho_{Geschwindigkeit} = \frac{n_{IST}^{Ausschluss}}{n_{SOLL}^{Ausschluss}} \quad \text{und} \quad \eta_{Geschwindigkeit} = \frac{n_{IST}^{Nicht\text{-}Ausschluss}}{n_{SOLL}^{Nicht\text{-}Ausschluss}}. \tag{6.4}$$

Für die Analysen wurden nur die Versuchsfahrten verwendet, die in der *Bewegungsidentifikation* mit „MS in Bewegung“ identifiziert wurden. Die Grundgesamtheiten $n_{SOLL}^{Ausschluss}$ und $n_{SOLL}^{Nicht\text{-}Ausschluss}$ waren daher je untersuchtes Unsicherheitsverfahren abhängig von den Ergebnissen der *Bewegungsidentifikation*.

Im Falle der zufälligen Variabilität zeigt Abbildung 6-10a) die Verläufe für die Korrektheit $\rho_{Geschwindigkeit}^{zufällige\ Variabilität}$ und die Effektivität $\eta_{Geschwindigkeit}^{zufällige\ Variabilität}$ in Abhängigkeit von der gewählten Sicherheitswahrscheinlichkeit $P_{Geschwindkeit}^{zufällige\ Variabilität}$. Die Grundgesamtheit bestand aus 64 Versuchsfahrten. Für den Kalibrierparameterwert wurde der Wert der Sicherheitswahrscheinlichkeit von $P_{Geschwindigkeit}^{zufällige\ Variabilität} = 67\%$ ausgewählt, der circa den Schnittpunkt beider Kurven widerspiegelt. An dieser Stelle betrugen die Korrektheit ca. $\rho_{Geschwindigkeit}^{zufällige\ Variabilität} = 0,78$ und entsprechend die Effektivität ca. $\eta_{Geschwindigkeit}^{zufällige\ Variabilität} = 78\%$.

Im Falle der Fuzzy-Theorie zeigt Abbildung 6-10b) die Verläufe der Korrektheit und der Effektivität. Die Grundgesamtheit bestand aus 54 Versuchsfahrten. Der ausgewählte Verwerfungsindex betrug $\rho_{Geschwindigkeit}^{Fuzzy\text{-}Theorie} = 0,87$, wieder beim Schnittpunkt beider Kurven. An dieser Stelle wurde ein Korrektheitsgrad von ca. $\rho_{Geschwindigkeit}^{Fuzzy\text{-}Theorie} = 0,74$ sowie entsprechend ein Effektivitätsgrad von ca. $\eta_{Geschwindigkeit}^{Fuzzy\text{-}Theorie} = 74\%$ erreicht. Beide waren damit etwa 4% geringer als bei Anwendung der zufälligen Variabilität.

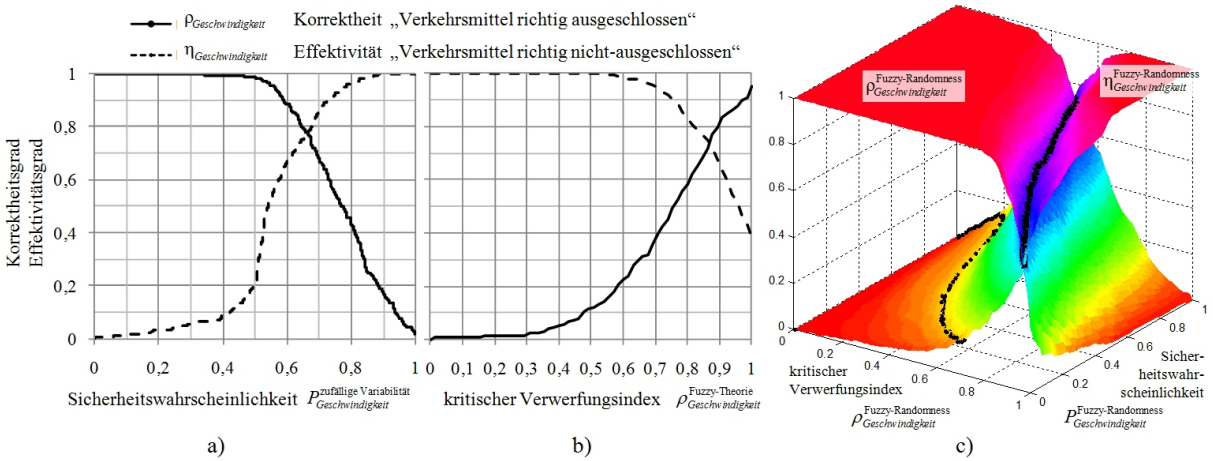


Abbildung 6-10: Korrektheitsgrade der *Identifikation des Verkehrsmittels anhand der Geschwindigkeit* für die Anwendung a) der zufälligen Variabilität, b) der Fuzzy-Theorie und c) des Fuzzy-Randomness

Bei Anwendung des Fuzzy-Randomness sind die Korrektheitsgrade von der Wahl der Sicherheitswahrscheinlichkeit und des Verwerfungsindex abhängig, deren Verläufe Abbildung 6-10c) zeigt. Die Grundgesamtheit bestand aus 64 Versuchsfahrten. Für die Wahl der Kalibrierparameter war die Stelle des Optimums aus der maximalen Korrektheit und der maximalen Effektivität zu bestimmen. Dieser Wert befindet sich auf der Schnittlinie beider Kurven, vgl. dazu die zwischen den Verläufen der Korrektheitsgraden schwarz eingezeichnete Linie in Abbildung 6-10c). Eine verebnete Darstellung dieser Schnittlinie zeigt Abbildung 6-11. Bei den ausgewählten Kalibrierparametern $P_{Geschwindigkeit}^{\text{Fuzzy-Randomness}} = 81\%$ und $\rho_{Geschwindigkeit}^{\text{Fuzzy-Randomness}} = 0,60$ befand sich näherungsweise das Maximum dieser Schnittlinie von ca. $\rho_{Geschwindigkeit}^{\text{Fuzzy-Randomness}} = 0,76$ und $\eta_{Geschwindigkeit}^{\text{Fuzzy-Randomness}} = 76\%$. Korrektheit und Effektivität waren somit ca. 2% geringer als bei Anwendung der zufälligen Variabilität und ca. 2% höher als bei Anwendung der Fuzzy-Theorie.

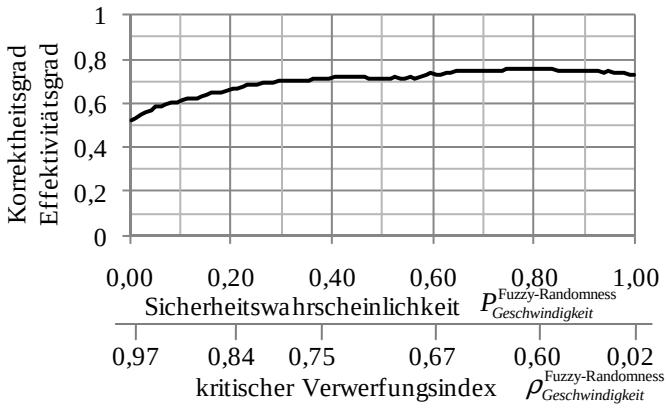


Abbildung 6-11: Verlauf der Korrektheit und Effektivität als verebnete Schnittlinie für die *Identifikation des Verkehrsmittels anhand der Geschwindigkeit* bei Anwendung des Fuzzy-Randomness

6.3.3 Korrelation zu ÖV-Fahrzeugen

Mit dem Identifizierungsverfahren der *Korrelation zu ÖV-Fahrzeugen* wird die räumliche und zeitliche Korrelation zwischen den Clusterpositionen und den ÖV-Fahrzeugen mit vordefinierten Fahrplänen berechnet. Um Fehlinterpretationen der Verkehrslage zu vermeiden, ist insbesondere das „sichere“ Ausschließen von MS dieser ÖV-Fahrzeuge von Interesse. Eine Korrelation zwischen MS und den ÖV-Fahrzeugen ist anzunehmen, wenn der Grenzwert für die Sicherheitswahrscheinlichkeit $P_{\underline{\text{ÖV}},\min}$ im Falle der zufälligen Variabilität, der Zugehörigkeitswert $Z_{\widetilde{\text{ÖV}},\min}$ bei Verwendung der Fuzzy-Theorie oder Grenzwert der Sicherheitswahrscheinlichkeit $P_{\underline{\widetilde{\text{ÖV}}},\min}$ bei Anwendung des Fuzzy-Randomness überschritten wurde.

Für die Analysen zur *Korrelation zu ÖV-Fahrzeugen* wurden alle der insgesamt 89 Versuchsfahrten ausgewertet, vgl. Tabelle 6-1. Bei 18 dieser Versuchsfahrten wurden die Mobilfunkdaten während der Mitfahrt in ÖV-Verkehrsmitteln generiert. Davon fanden 2 Versuchsfahrten in Linienbussen, 4 in Straßenbahnen und insgesamt 12 in S-Bahnen, Regionalbahnen und IC-Zügen der DB statt. Die restlichen 71 Versuchsfahrten wurden nicht in ÖV-Fahrzeugen durchgeführt, weshalb für sie keine Korrelation zum ÖV (nicht-ÖV) angenommen wird.

Der Korrektheitsgrad $\rho_{\ddot{O}V}$ für die korrekte Identifizierung der in ÖV-Fahrzeugen durchgeführten Versuchsfahrten und der Effektivitätswert $\eta_{\ddot{O}V}$ für die korrekte Identifizierung der nicht in ÖV-Fahrzeugen durchgeführten Versuchsfahrten wurden berechnet mit:

$$\rho_{\ddot{O}V} = \frac{n_{IST}^{\ddot{O}V}}{n_{SOLL}^{\ddot{O}V}} \quad \text{und} \quad \eta_{\ddot{O}V} = \frac{n_{IST}^{\text{nicht-}\ddot{O}V}}{n_{SOLL}^{\text{nicht-}\ddot{O}V}}. \tag{6.5}$$

Für die Analysen wurden ausschließlich die mit „MS in Bewegung“ identifizierten Versuchsfahrten ausgewertet. Die Grundgesamtheiten $n_{SOLL}^{ÖV}$ und $n_{SOLL}^{nicht\text{-}ÖV}$ waren daher abhängig von den Ergebnissen der *Bewegungsidentifikation*. Mit $n_{IST}^{ÖV}$ wurde die Anzahl der korrekt als ÖV identifizierten Versuchsfahrten und mit $n_{IST}^{nicht\text{-}ÖV}$ wurde die Anzahl der korrekt als nicht-ÖV identifizierten Versuchsfahrten bezeichnet.

Für die Anwendung der zufälligen Variabilität zeigt Abbildung 6-12a) die Verläufe der Korrektheit $\rho_{ÖV}^{\text{zufällige Variabilität}}$ und der Effektivität $\eta_{ÖV}^{\text{zufällige Variabilität}}$ in Abhängigkeit von der Sicherheitswahrscheinlichkeit $P_{ÖV}^{\text{zufällige Variabilität}}$. Die Grundgesamtheiten bestanden aus $n_{SOLL}^{ÖV} = 18$ und $n_{SOLL}^{nicht\text{-}ÖV} = 48$. Für den Kalibrierparameterwert wurde $P_{\underline{ÖV}}^{\text{zufällige Variabilität}} = 0,027\%$ beim Schnittpunkt beider Kurven ausgewählt. Dort wurde ein Korrektheitswert von ca. $\rho_{\underline{ÖV}}^{\text{zufällige Variabilität}} = 0,62$ und damit ein Effektivitätswert von ca. $\eta_{\underline{ÖV}}^{\text{zufällige Variabilität}} = 62\%$ erreicht.

Die Verläufe der Korrektheit und der Effektivität bei Anwendung der Fuzzy-Theorie zeigt Abbildung 6-12b). Die ausgewerteten Grundgesamtheiten bestanden aus $n_{SOLL}^{ÖV} = 18$ und $n_{SOLL}^{nicht\text{-}ÖV} = 44$. Als Kalibrierparameterwert wurde der Verwerfungsgrad von $\rho_{\widetilde{ÖV}}^{\text{Fuzzy-Theorie}} = 0,011$, wieder beim Schnittpunkt beider Kurven ausgewählt. An dieser Stelle wurden für die Korrektheit ca. $\rho_{\widetilde{ÖV}}^{\text{Fuzzy-Theorie}} = 0,66$ und entsprechend für die Effektivität ca. $\eta_{\widetilde{ÖV}}^{\text{Fuzzy-Theorie}} = 66\%$ erreicht. Korrektheit und Effektivität waren somit etwa 5% höher als bei Anwendung der zufälligen Variabilität.

Im Falle des Fuzzy-Randomness zeigt die Abbildung 6-12c) die Verläufe der Korrektheit $\rho_{\underline{\widetilde{ÖV}}}^{\text{Fuzzy-Randomness}}$ und Effektivität $\eta_{\underline{\widetilde{ÖV}}}^{\text{Fuzzy-Randomness}}$. Die Grundgesamtheiten bestanden aus $n_{SOLL}^{ÖV} = 18$ und $n_{SOLL}^{nicht\text{-}ÖV} = 42$. Für den Kalibrierparameterwert wurde $P_{\underline{\widetilde{ÖV}}}^{\text{Fuzzy-Randomness}} = 0,11\%$ etwa im Schnittpunkt dieser Kurven ausgewählt. Dort wurden Werte für die Korrektheit von ca. $\rho_{\underline{\widetilde{ÖV}}}^{\text{Fuzzy-Randomness}} = 0,78$ und für die Effektivität von ca. $\eta_{\underline{\widetilde{ÖV}}}^{\text{Fuzzy-Randomness}} = 78\%$ erzielt. Die Korrektheit und die Effektivität waren somit etwa 12% höher als bei Anwendung der Fuzzy-Theorie und etwa 16% höher als bei Verwendung der zufälligen Variabilität.

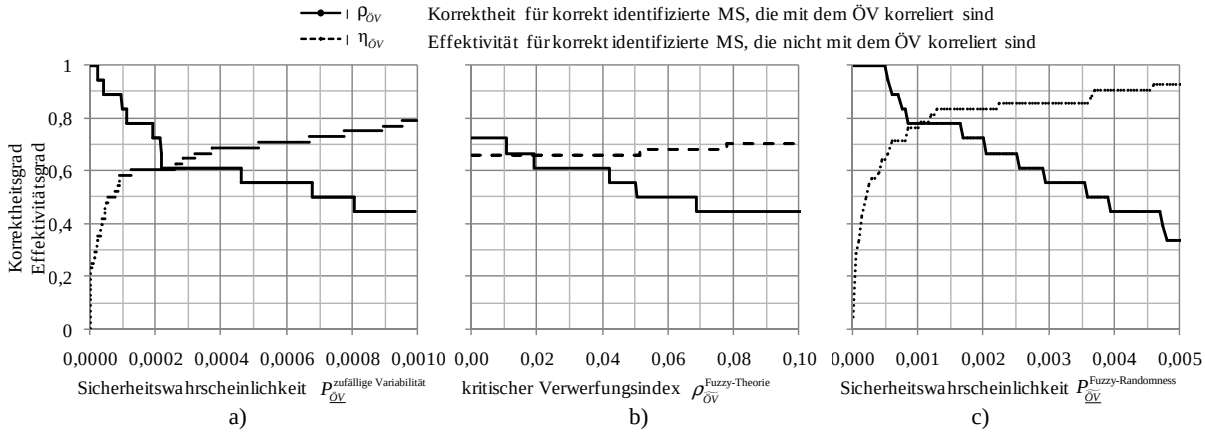


Abbildung 6-12: Verläufe der Korrektheit und der Effektivität für die Identifikation der *Korrelation mit ÖV-Fahrzeugen* bei Anwendung
a) der zufälligen Variabilität, b) der Fuzzy-Theorie und c) des Fuzzy-Randomness

Es ist anzumerken, dass die für die vorliegende Untersuchung zur Verfügung gestandene Anzahl an Versuchsfahrten in ÖV-Fahrzeugen gering war, was sich deutlich im stufenförmigen Verlauf der Kurven widerspiegelt. Dennoch konnte gezeigt werden, dass MS, die tatsächlich mit ÖV-Fahrzeugen korreliert waren, als solche identifiziert werden konnten. Konkret wurde gezeigt, dass mit den ausgewählten Kalibrierparameterwerten mehr als ca. 60%, mit Anwendung des Fuzzy-Randomness sogar mehr als dreiviertel der mit

dem ÖV korrelierten MS herausgefiltert werden können. Wie insbesondere der Abbildung 6-12a) und c) entnommen werden kann, sind mit geringen Effektivitätsverlusten zum Teil deutlich höhere Korrektheitsgrade erzielbar. Beispielsweise wären Korrektheitsgrade von ca. 90% im Falle der zufälligen Variabilität, das entspricht einer Steigerung von ca. 30%, mit einem Verlust von nur ca. 10% bei der Effektivität erreichbar (bei ca. $P_{\overline{\ddot{O}V}}^{\text{zufällige Variabilität}} = 0,009\%$). Im Falle des Fuzzy-Randomness würden bei gleicher Korrektheit ebenfalls ca. 10% mehr Mobilfunkdaten verworfen werden. Um die Ergebnisse der Unsicherheitsmodellierungsverfahren unter den eingangs definierten Kriterien der gewogenen Auswahl bewerten zu können, werden nachfolgend weiter die genannten Kalibrierparameter der Schnittpunkte der Korrektheits- und Effektivitätskurven verwendet.

Der verfahrensbedingte hohe Anteil der Versuchsfahrten, die zu Unrecht mit ÖV-Fahrzeugen korreliert identifiziert wurden, ist Ausdruck der wirkenden hohen Unsicherheiten. In diesen Fällen konnte eine Korrelation mit den ÖV-Fahrzeugen nicht gesichert ausgeschlossen werden. Hieraus wird deutlich, dass in der originären Menge der Mobilfunkdaten zahlreiche MS enthalten sind, die aufgrund möglicher Korrelationen zu ÖV-Fahrzeugen für die Verkehrserfassung des MIV ungeeignet sind. Mit dem vorgestellten Verfahren können diese Mobilfunkdaten herausgefiltert werden, um im Sinne der MIV-Verkehrslageerfassung qualitativ hochwertige Daten bereitstellen zu können.

6.3.4 Kartengestützte Plausibilitätsprüfung

Ziel der *kartengestützten Plausibilitätsprüfung* ist es, nachzuweisen, dass auf den Graphen des MIV für diejenigen MS Trajektorien existieren, die tatsächlich zur Gruppe der MIV gehören. Für den Nachweis der Plausibilität, dass ein MS ein Verkehrsmittel des MIV nutzte, genügt es auf dem Graphen $G^{\text{Straße}}$ eine Trajektorie zu finden. Insofern „Fußgänger“ nicht im Identifizierungsverfahren *Identifikation des Verkehrsmittels anhand der Geschwindigkeit* ausgeschlossen werden konnten, ist zudem die Existenz einer Trajektorie auf dem Graph $G^{\text{Fußweg}}$ zu prüfen. Wird auf diesem Graphen ebenfalls eine Trajektorie gefunden, kann keine eindeutige Entscheidung zwischen den Verkehrsmitteln des NMIV und des MIV erfolgen. In diesem Fall sind die Mobilfunkdaten für die Erfassung des MIV auszuschließen.

Im Sinne der Korrektheit, wie sie in den vorherigen Abschnitten bestimmt wurde, erzeugt das vorliegende Identifizierungsverfahren keine „unrichtigen“ Ergebnisse. Wurde keine Trajektorie auf $G^{\text{Straße}}$ gefunden oder eine Trajektorie auf $G^{\text{Fußweg}}$ gefunden, obwohl ein Verkehrsmittel des MIV genutzt wurde, dann verringert sich ausschließlich der Grad der Effektivität $\eta_{\text{Plausibilität}}$ für die *kartengestützte Plausibilitätsprüfung*. Maßgeblich ist somit die Qualität der Eingangsdaten, also der digitalen Verkehrsnetzkarten für die Korrektheit. In WILTSCHKO (2004) werden in Abhängigkeit von der Aktualisierungsrate Schätzungen für Korrektheitswerte digitaler Karten, insbesondere digitaler Straßenkarten gegeben. Mit Annahme, dass eine jährliche Aktualisierung der digitalen Verkehrsnetzkarten vorgenommen wird, wird, basierend auf den Schätzungen von WILTSCHKO (2004) S. 134ff, der Korrektheitswert von $\rho_{Plausibilität} = 98\%$ unabhängig von der Unsicherheits-modellierungsart und den Kalibrierparametern festgesetzt.

Bei den vorherigen Identifizierungsverfahren galt, dass nur die MS weiter betrachtet wurden, für die der MIV als mögliche Verkehrsmittelgruppe klassifiziert wurde. Im aktuellen Verfahren gilt zudem die Ausschließlichkeit, so dass nur diejenigen MS weiter betrachtet werden, die eindeutig als MIV klassifiziert wurden.

Der Wert der Effektivität $\eta_{Plausibilität}$ wurde aus den beiden Effektivitätswerten $\eta_{\text{MIV Trajektorie}}$ für den Nachweis der Plausibilität des MIV und $\eta_{\text{NMIV Ausschluss}}$ für den Nachweis des NMIV, obwohl ein Verkehrsmittel des MIV genutzt wurde, berechnet mit:

$$\eta_{Plausibilität} = \eta_{\text{MIV Trajektorie}} \cdot \eta_{\text{NMIV Ausschluss}} \quad \text{mit:} \quad \eta_{\text{MIV Trajektorie}} = \frac{n_{\text{IST}}^{\text{MIV}}}{n_{\text{SOLL}}^{\text{MIV}}} \qquad \eta_{\text{NMIV Ausschluss}} = 1 - \frac{n_{\text{IST}}^{\text{NMIV}}}{n_{\text{SOLL}}^{\text{MIV}}}. \tag{6.6}$$

Mit $n_{\text{SOLL}}^{\text{MIV}}$ wird die Grundgesamtheit der Versuchsfahrten bezeichnet, die im MIV durchgeführt wurden. Hierfür fanden 41 Versuchsfahrten im PKW statt, vgl. Tabelle 6-1 Spalte 4. Ein Teil dieser Versuchsfahrten wurde in der *Bewegungsidentifikation* als „MS steht“ identifiziert und ausgeschlossen. Die tatsächliche Anzahl der auswertbaren Datensätze war dadurch je nach Art der Unsicherheitsmodellierung geringer. Mit $n_{\text{IST}}^{\text{MIV}}$ wird die Anzahl der Versuchsfahrten in Verkehrsmitteln des MIV bezeichnet, für die Trajektorien – je Versuchsfahrt mindestens eine – auf $G^{\text{Straße}}$ gefunden wurden. Mit $n_{\text{IST}}^{\text{NMIV}}$ wird die Anzahl der Versuchsfahrten angegeben, für die eine Trajektorie auf $G^{\text{Fußweg}}$ gefunden wurde.

Im Falle der Fuzzy-Theorie wurde sich verfahrensbedingt zur Beschreibung der Ortungsflächen und der Zeitfenster auf den endlichen Support (sowohl im xy-Raum als auch für die Zeit) der jeweiligen MS-Positionen und damit auf die Berücksichtigung der maximalen Unschärfe beschränkt. Hierfür konnten von den $n_{\text{SOLL}}^{\text{MIV}} = 33$ auswertbaren Versuchsfahrten für $n_{\text{IST}}^{\text{MIV}} = 26$ Versuchsfahrten Trajektorien gefunden werden. Die Effektivität für den Nachweis der MIV-Plausibilität betrug somit ca. $\eta_{\underline{\text{MIV Trajektorie}}}^{\text{Fuzzy-Theorie}} = 79\%$. Auf dem Verkehrsnetz $G^{\text{Fußweg}}$ wurden für diese Versuchsfahrten $n_{\text{IST}}^{\text{NMIV}} = 8$ Trajektorien gefunden, weshalb die Effektivität für den Ausschluss von Verkehrsmittel des NMIV ca. $\eta_{\underline{\text{NMIV Ausschluss}}}^{\text{Fuzzy-Theorie}} = 76\%$ betrug. Bei Anwendung der Fuzzy-Theorie betrug die Effektivität der *kartengestützte Plausibilitätsprüfung* Identifizierungsverfahrens somit ca. $\eta_{\underline{Plausibilität}}^{\text{Fuzzy-Theorie}} = 60\%$.

Im Falle der zufälligen Variabilität und des Fuzzy-Randomness werden die Ausdehnung der Ortungsflächen und Zeitfenster durch die gewählte Sicherheitswahrscheinlichkeit $P_{\underline{Plausibilität}}^{\text{zufällige Variabilität}}$ beziehungsweise $P_{\underline{Plausibilität}}^{\text{Fuzzy-Randomness}}$ definiert. Die Verläufe der Effektivitätsgrade $\eta_{\text{MIV Trajektorie}}$ für die gefundenen Trajektorien der MIV-Versuchsfahrten und $\eta_{\text{NMIV Ausschluss}}$ für den Ausschluss von Trajektorien des NMIV sind in Abbildung 6-13a) und b) dargestellt. Zu erkennen ist, dass bei höheren Sicherheitswahrscheinlichkeiten mehr Trajektorien gefunden werden. Dieser Effekt ist Folge der größer werdenden Konfidenzbereiche und damit der größer werdenden Ortungsflächen sowie Zeitfenster, bei steigenden Sicherheitswahrscheinlichkeiten. Für den Grenzfall der Sicherheitswahrscheinlichkeit von Eins gehen die Konfidenzbereiche gegen unendlich, so dass sich alle Ortungsflächen aller MS-Positionen sowohl im xy-Raum als auch zeitlich überlappen. Die Möglichkeit, dass eine Trajektorie zwischen den MS-Positionen gefunden werden kann, steigt somit ebenfalls mit steigender Sicherheitswahrscheinlichkeit, da zum einen die Anzahl an Start- und Zielknoten je Ortungsfläche steigt und zum anderen längere Fahrzeiten für die Trajektorie gestattet werden, vgl. die Abschnitte 4.6.1 und 4.6.2. Infolgedessen kann erwartet werden, dass ab einem bestimmten Sicherheitswahrscheinlichkeitswert für alle MS-Positionsfolgen eine Trajektorie gefunden wird. Ab diesem Wert tritt eine **Sättigung** ein, da trotz steigender Sicherheitswahrscheinlichkeit die (maximale) Anzahl der MS, für die eine Trajektorie auf dem betreffenden Verkehrsnetz gefunden wurde, nicht mehr steigen kann.

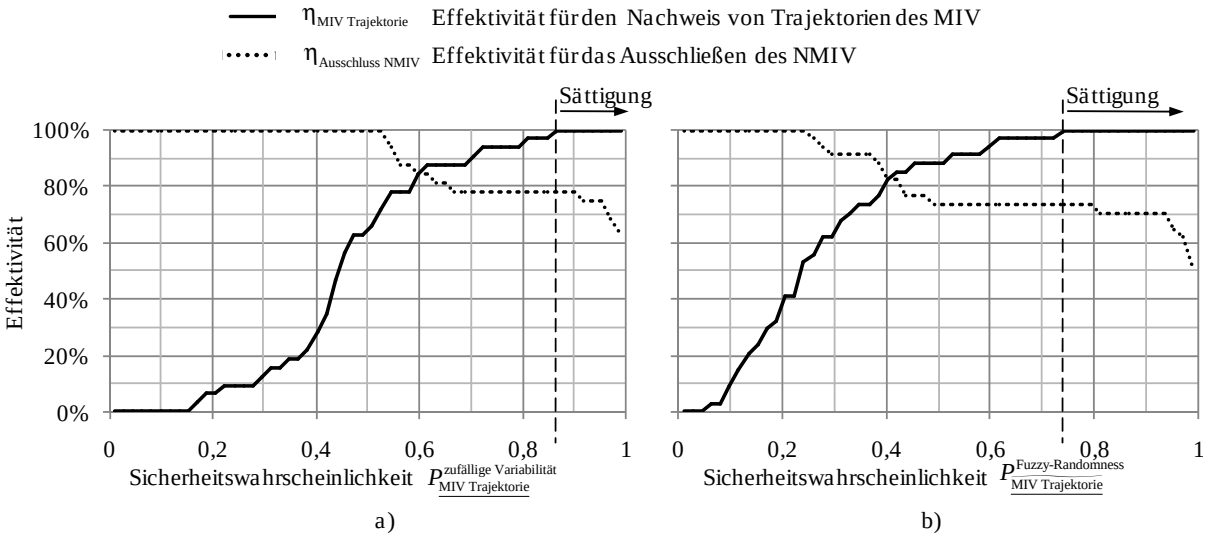


Abbildung 6-13: Effektivitätswerte der *kartengestützten Plausibilität* in Abhängigkeit von der Sicherheitswahrscheinlichkeit für die Anwendung a) der zufälligen Variabilität und b) des Fuzzy-Randomness

Die Grundgesamtheit bestand im Falle der zufälligen Variabilität aus $n_{SOLL}^{\text{MIV Trajektorie}} = 32$ auswertbaren Versuchsfahrten. Ab ca. $P_{\underline{Plausibilität}}^{\text{zufällige Variabilität}} \geq 87\%$ stellte sich die Sättigung ein, so dass ab diesem Wert für alle MIV-Versuchsfahrten Trajektorien auf dem Verkehrsnetzgraphen $G^{\text{Straße}}$ gefunden wurden und somit die Effektivität $\eta_{\text{MIV Trajektorie}}^{\text{zufällige Variabilität}} = 100\%$ erreicht wurde. Der Wert $P_{\underline{Plausibilität}}^{\text{zufällige Variabilität}} = 87\%$ wurde infolgedessen als Kalibrierparameterwert ausgewählt. Bei Wahl dieser Sicherheitswahrscheinlichkeit wurde für $n_{IST}^{\text{NMIV Trajektorie}} = 7$ Versuchsfahrten nachgewiesen, dass für sie auf dem Verkehrsnetz $G^{\text{Fußweg}}$ eine Trajektorien existierte, was einer Effektivität für den Ausschluss von Mobilfunkdaten des NMIV von ca. $\eta_{\text{Ausschluss NMIV}}^{\text{zufällige Variabilität}} = 78\%$ entspricht. Die Effektivität der kartengestützten Plausibilität im Falle der zufälligen Variabilität entsprach somit ca. $\eta_{\textit{Plausibilität}}^{\text{zufällige Variabilität}} = 78\%$. Sie ist damit ca. 18% höher als bei Anwendung der Fuzzy-Theorie.

Bei Anwendung des Fuzzy-Randomness wird unter Berücksichtigung der maximalen Unschärfen die Ausdehnung der Ortungsflächen und Zeitfenster, analog zur zufälligen Variabilität, durch die gewählte Sicherheitswahrscheinlichkeit $P_{\underline{\text{MIV Trajektorie}}}^{\text{Fuzzy-Randomness}}$ definiert. Die Verläufe der Effektivitätsgrade $\eta_{\text{MIV Trajektorie}}^{\text{Fuzzy-Randomness}}$ und $\eta_{\text{NMIV Ausschluss}}^{\text{Fuzzy-Randomness}}$ sind in Abbildung 6-13b) dargestellt. Insgesamt konnten $n_{SOLL}^{\text{MIV Trajektorie}} = 34$ Versuchsfahrten ausgewertet werden. Ab ca. $P_{\underline{\text{MIV Trajektorie}}}^{\text{Fuzzy-Randomness}} \geq 74\%$ stellte sich die Sättigung ein, so dass für alle MIV-Versuchsfahrten Trajektorien auf $G^{\text{Straße}}$ gefunden wurden, weshalb $P_{\underline{\text{MIV Trajektorie}}}^{\text{Fuzzy-Randomness}} = 74\%$ als Kalibrierparameterwert ausgewählt wurde. Bei diesem Wert wurden für $n_{IST}^{\text{NMIV Trajektorie}} = 9$ Versuchsfahrten Trajektorie auf $G^{\text{Fußweg}}$ gefunden, was einer Effektivität von ca. $\eta_{\text{NMIV Ausschluss}}^{\text{Fuzzy-Randomness}} = 74\%$ entspricht. Im Falle des Fuzzy-Randomness wurde somit ein Effektivität von $\eta_{\textit{Plausibilität}}^{\text{Fuzzy-Randomness}} = 74\%$ erzielt. Im Vergleich zur Anwendung der zufälligen Variabilität war die Effektivität ca. 4% geringer und zur Anwendung der Fuzzy-Theorie ca. 14% höher.

6.3.5 Zusammenfassung der Korrektheitsgrade und Effektivitätswerte

Im aktuellen Abschnitt werden die unter Verwendung der Versuchsfahrten aus Tabelle 6-1 ermittelten Korrektheitsgrade und Effektivitätswerte der Identifizierungsverfahren zusammengefasst dargestellt. Die Zahlenwerte beinhalten für die Anwendung der zufälligen Variabilität die Tabelle 6-2, für die Fuzzy-Theorie die Tabelle 6-3 und für das Fuzzy-Randomness die Tabelle 6-4.

Mit ρ_{MIV} wird der Korrektheitsgrad angegeben, dass die als MIV identifizierten Mobilfunkdaten tatsächlich in einem Verkehrsmittel des MIV generiert wurden. Und mit η_{MIV} wird die Effektivität der Identifizierungsverfahren angegeben, dass in Verkehrsmitteln des MIV generierte Verkehrsmittel, auch als Mobilfunkdaten des MIV identifiziert wurden. Die Markierung $^{*)}$ weist darauf hin, dass verfahrensbedingt keine Korrektheit berechenbar war. Sie wurde stattdessen aus den Annahmen zur Korrektheit digitaler Verkehrsnetzkarten festgesetzt.

Aus Darstellungsgründen wurden für die Bezeichnungen der Identifizierungsverfahren Abkürzungen verwendet: *Bewegung* $\triangleq$ *Bewegungsidentifikation*, *Geschwindigkeit* $\triangleq$ *Identifikation des Verkehrsmittels anhand der Geschwindigkeit*, *ÖV* $\triangleq$ *Korrelation zu ÖV-Fahrzeugen*, *Plausibilität* $\triangleq$ *kartengestützte Plausibilitätsprüfung*.

Die Berechnungen der Effektivität und Korrektheit in „Gesamt“ werden im nachfolgenden Abschnitt erläutert.

Tabelle 6-2: Kalibrierparameterwerte, Korrektheit und Effektivität bei Anwendung der zufälligen Variabilität

Identifikationsverfahren	Kalibrierparameterwert	Effektivität	Korrektheit
Bewegung	$\gamma_{\min,Bewegung}^{\text{zufällige Variabilität}} = 0,56$	0,81	0,81
Geschwindigkeit	$P_{Geschwindigkeit}^{\text{zufällige Variabilität}} = 0,67$	0,78	0,78
ÖV	$P_{\underline{\widetilde{ÖV}}}^{\text{zufällige Variabilität}} = 0,00027$	0,62	0,62
Plausibilität	$P_{\underline{Plausibilität}}^{\text{zufällige Variabilität}} = 0,87$	0,78	$0,98^{*)}$
Gesamt		**η_{MIV} = 0,306**	**ρ_{MIV} =0,798**

Tabelle 6-3: Kalibrierparameterwerte, Korrektheit und Effektivität bei Anwendung der Fuzzy-Theorie

Identifikationsverfahren	Kalibrierparameterwert	Effektivität	Korrektheit
Bewegung	$\rho_{Bewegung}^{\text{Fuzzy-Theorie}} = 0,42$	0,72	0,72
Geschwindigkeit	$\rho_{Geschwindigkeit}^{\text{Fuzzy-Theorie}} = 0,87$	0,74	0,74
ÖV	$\rho_{\widetilde{ÖV}}^{\text{Fuzzy-Theorie}} = 0,011$	0,66	0,66
Plausibilität	-	0,60	$0,98^{*)}$
Gesamt		**η_{MIV} = 0,211**	**ρ_{MIV} = 0,775**

Tabelle 6-4: Kalibrierparameterwerte, Korrektheit und Effektivität bei Anwendung des Fuzzy-Randomness

Identifikationsverfahren	Kalibrierparameterwert		Effektivität	Korrektheit
Bewegung	$\gamma_{\min,Bewegung}^{\text{Fuzzy-Randomness}} = 0,61$	$\rho_{Bewegung}^{\text{Fuzzy-Randomness}} = 0,50$	0,84	0,84
Geschwindigkeit	$P_{Geschwindigkeit}^{\text{Fuzzy-Randomness}} = 0,81$	$\rho_{Geschwindigkeit}^{\text{Fuzzy-Randomness}} = 0,60$	0,76	0,76
ÖV	$P_{\underline{\widetilde{ÖV}}}^{\text{Fuzzy-Randomness}} = 0,0011$	-	0,77	0,77
Plausibilität	$P_{\underline{\widetilde{\text{MIV Trajektorie}}}}^{\text{Fuzzy-Randomness}} = 0,74$	-	0,73	$0,98^{*)}$
Gesamt			**η_{MIV} = 0,359**	**ρ_{MIV} = 0,838**

6.3.6 Zusammenfassung zur Kalibrierung und Evaluation der Identifikationsverfahren

In den vorangegangenen Abschnitten wurden mit Hilfe real erfasster Mobilfunkdaten und Referenzinformationen zur Verkehrsmittelwahl Kalibrierparameterwerte (Grenzwerte für Sicherheitswahrscheinlichkeit und kritischem Verwerfungsindex) für die in der vorliegenden Arbeit entwickelten Identifizierungsverfahren bestimmt. Es zeigte sich, dass die Werteauswahl dieser Kalibrierparameterwerte sowohl die Korrektheit der identifizierten MS als auch die Effektivität der Identifizierungsverfahren beeinflusste. Die Korrektheit und die Effektivität wurden für jedes Identifizierungsverfahren separat ermittelt. Ziel des vorliegenden Abschnitts ist es, die Korrektheit und die Effektivität der Identifizierungsverfahren nach Ausführung aller vier Identifizierungsverfahren zu bestimmen.

Mit jedem Identifizierungsschritt werden weitere Mobilfunkdaten, die nicht mit den Merkmalen des MIV übereinstimmen, herausgefiltert. Die nach Anwendung aller Identifizierungsverfahren resultierende Korrektheit ρ_{MIV} der als MIV klassifizierten Mobilfunkdaten ist somit abhängig von den Korrektheiten der einzelnen Verfahren. Werte für die Korrektheiten wurden in den vorangegangenen Abschnitten auf Basis der Versuchsfahrten bestimmt. Erst nach Anwendung aller Verfahren ist eine abschließende Klassifizierung möglich. Jedes Verfahren trägt somit einen – in der vorliegenden Arbeit gleichgewichteten – Teil der Identifizierung bei, weshalb die Korrektheit aus dem Mittelwert der einzelnen Korrektheiten berechnet wurde mit:

$$\rho_{MIV} = \frac{1}{4} \cdot \left(\rho_{Bewegung} + \rho_{Geschwindigkeit} + \rho_{\ddot{O}V} + \rho_{Plausibilität} \right). \tag{6.7}$$

Da tatsächlich in Verkehrsmitteln des MIV generierte Mobilfunkdaten nicht oder nicht eindeutig identifiziert werden und deshalb – zu Unrecht – verworfen werden, reduziert sich in diesen Fällen die Anzahl der Mobilfunkdaten des MIV nach jedem Identifizierungsverfahren. Die resultierende Effektivität η_{MIV} nach Anwendung aller Identifizierungsverfahren wird infolgedessen berechnet mit:

$$\eta_{MIV} = \eta_{Bewegung} \cdot \eta_{Geschwindigkeit} \cdot \eta_{\ddot{O}V} \cdot \eta_{Plausibilität}. \tag{6.8}$$

Die Zahlenwerte für ρ_{MIV} und η_{MIV} sind in Tabelle 6-2 für die zufällige Variabilität, in Tabelle 6-3 für die Fuzzy-Theorie und in Tabelle 6-4 die Anwendung des Fuzzy-Randomness jeweils in der Zeile „Gesamt“ aufgeführt. Im Diagramm der Abbildung 6-14 werden die Zahlenwerte für die Korrektheit grafisch dargestellt. Folgendes kann der Abbildung entnommen werden:

- Für alle der drei Unsicherheitsmodellierungsarten konnten geeignete Kalibrierparameterwerte gefunden werden.
- Mit Anwendung des Fuzzy-Randomness identifizierte MS, wiesen mit ca. 83,8% die höchste Korrektheit auf. Die Korrektheit von MS, die mit Anwendung der zufälligen Variabilität identifizierten worden waren, war mit ca. 79,8% etwa 4% geringer, und mit Anwendung der Fuzzy-Theorie mit ca. 77,5% etwa weitere 2% geringer.
- Bei allen der drei Unsicherheitsmodellierungsarten war der Anteil der im Identifizierungsverfahren *Korrelation zu ÖV-Fahrzeugen* inkorrekt identifizierten MS am höchsten.

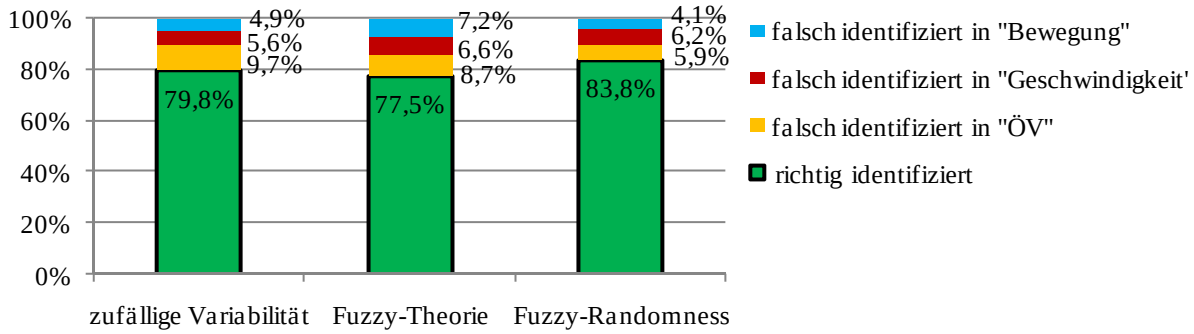


Abbildung 6-14: Korrektheiten für die den MIV nutzenden MS

Die Effektivitätswerte der Identifizierungsverfahren zeigt Abbildung 6-15. Bei Anwendung des Fuzzy-Randomness konnte die höchste Effektivität von ca. 35,9%, erzielt werden. Mit den Modellen der zufälligen Variabilität die zweithöchste Effektivität von ca. 30,6% und mit denen der Fuzzy-Theorie die geringste Effektivität mit ca. 21,1% erreicht.

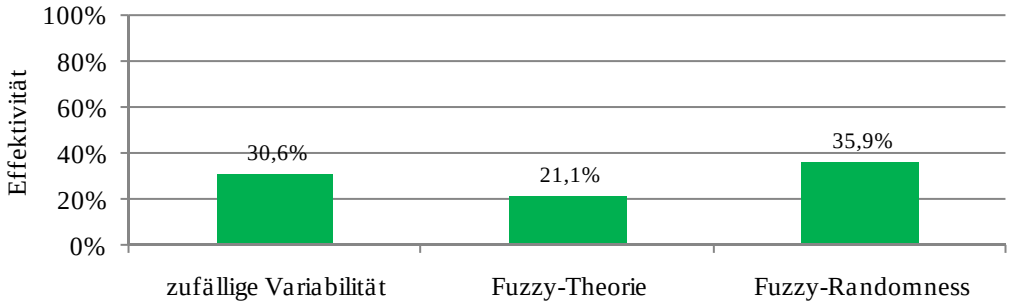


Abbildung 6-15: Zusammengefasste Effektivitätswerte der Identifizierungsverfahren

Zusammenfassend wurde festgestellt, dass sowohl für die Korrektheit als auch für die Effektivität mit Anwendung des Fuzzy-Randomness die besten Ergebnisse, im Falle der zufälligen Variabilität die zweitbesten Ergebnisse und mit Anwendung der Fuzzy-Theorie die drittbesten Ergebnisse erzielt wurden. Es konnte gezeigt werden, dass mittels ihrer Mobilfunkdaten diejenigen MS identifiziert werden können, die in Verkehrsmitteln des MIV transportiert wurden. Die identifizierten MS des MIV sind zudem qualitativ mittels der Korrektheitsgrade beschreibbar. Die (geringen) Effektivitätswerte der Identifizierungsverfahren sind als Resultat der hohen geometrischen Unsicherheiten der Mobilfunkortung zu sehen. Inwieweit Verbesserungen der geometrischen Unsicherheiten den Grad der Effektivität tatsächlich erhöhen können, bleibt nachfolgenden Arbeiten vorbehalten.

6.4 Evaluation anhand anonymisierter Mobilfunkdaten

6.4.1 Vorgehensweise

Wesentliches Merkmal der Verkehrserfassung des MIV mit Hilfe von Mobilfunkdaten ist, dass der Verkehr flächendeckend überall dort erfasst werden kann, wo Mobilfunkdaten durch MS in Verkehrsmitteln des MIV generiert werden. Zur Potentialabschätzung der entwickelten Identifizierungsverfahren und ihrer Unsicherheitsmodelle werden in den folgenden Abschnitten alle innerhalb eines Tages und einer Region erfassten anonymisierten Mobilfunkdaten ausgewertet und die Identifikationsergebnisse bewertet.

Im vorangegangenen Abschnitt 6.3 wurden für die Identifizierungsverfahren die Kalibrierparameterwerte bestimmt, die insbesondere eine qualitative Beurteilung der Identifizierungsergebnisse hinsichtlich der erwartbaren Korrektheit erlauben. Die Kalibrierung erfolgte mit Hilfe realen Daten (Mobilfunkdaten und Referenzdaten), die in derselben Region erfasst wurden. Für die nachfolgenden Analysen wird daher angenommen, dass die Kalibrierparameterwerte und die für die Identifizierungsergebnisse gefundenen Qualitätsmaße auf die Auswertung der anonymisierten Mobilfunkdaten übertragen werden können.

Nach der Beschreibung der ausgewerteten Mobilfunkdaten, ihrer Mobilfunkortung und deren Identifizierung im anschließenden Abschnitt 6.4.2, werden in Abschnitt 6.4.3 die Auswirkungen bei Wahl der Zeitspanne $dt_{MIV,max}$ bei der Verknüpfung der Identifizierungsergebnisse erläutert.

6.4.2 Identifizierung anonymisierter Mobilfunkdaten

Die anonymisierten Mobilfunkdaten wurden am 09.07.2008 im Versuchsgebiet Karlsruhe, vgl. Abschnitt 6.2.1 und Abbildung 6-1, aufgezeichnet. Für den Zweck der Mobilfunkortung wurden die erforderlichen Mobilfunkdaten, vgl. in Anhang A die Tabelle A-1, des Mobilfunknetzbetreibers T-Mobile Deutschland erfasst und für die post processing-Analyse gespeichert. Aufzeichnungsbeginn war um ca. 00 Uhr 12 Minuten und Aufzeichnungsende um ca. 23 Uhr 7 Minuten. Aus diesen Mobilfunkdaten konnten insgesamt 35 791 435 Mobilfunkortungen von 162 572 MS berechnet werden. Die Abbildung 6-16 zeigt ein Histogramm der georteten Mobilfunkdaten mit der Klassenbreite von einer Minute. In Anhang B Tabelle B-1 wurden ausgewählte Kennzahlen für die erzielten Ortungsunsicherheiten für die Anwendung der zufälligen Variabilität, der Fuzzy-Theorie und des Fuzzy-Randomness zusammengefasst.

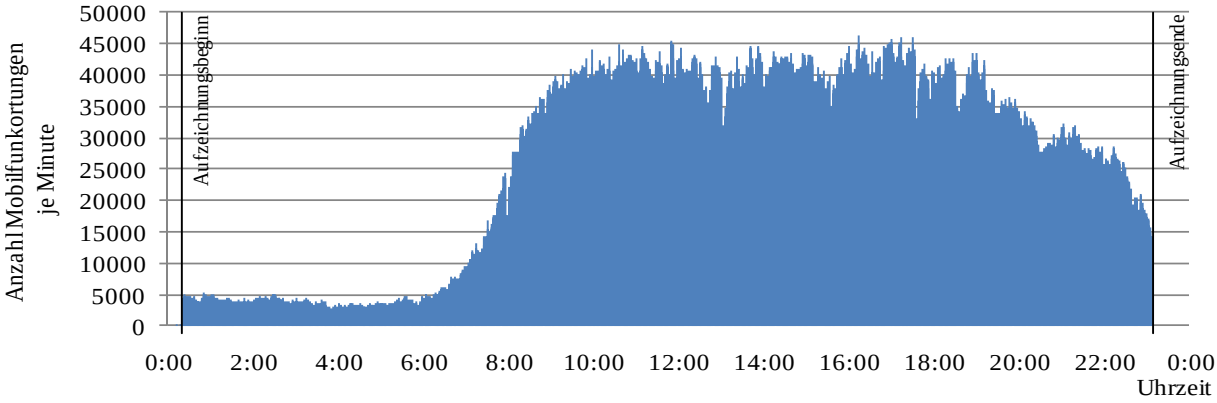


Abbildung 6-16: Histogramm für die georteten Mobilfunkdaten mit einer Klassenbreite von einer Minute

Die Identifizierung der anonymisierten Mobilfunkdaten wurde für die Anwendung der zufälligen Variabilität, der Fuzzy-Theorie und des Fuzzy-Randomness durchgeführt. Zur Bildung von Trajektorien wurden die

Positionsfolgen der 35 791 435 Mobilfunkortungen den jeweiligen Telefongesprächen (**call**) zugeordnet. Zur Erhöhung der Recheneffizienz des Clusterings wurde die maximale zeitliche Länge eines calls auf 30 Minuten begrenzt. Länger dauerende Telefongespräche wurde in mehrere Trajektorien zu je maximal 30 Minuten unterteilt. Insgesamt wurden die Mobilfunkortungen 426 191 Trajektorien zugeordnet. In Tabelle 6-5 sind die Anzahlen der als MIV identifizierten Trajektorien, die mit dem in Abschnitt 5.3 beschriebenen Computersystem benötigte Berechnungsdauer und die mittlere Anzahl an Trajektorien je Minute (Tr./Min.) zusammengefasst. Der Wert für Tr./Min. bezieht sich auf den Erfassungszeitraum von 22 Stunden und 55 Minuten zwischen Aufzeichnungsbeginn und Aufzeichnungsende.

Tabelle 6-5: Anzahlen und Berechnungsdauer der als MIV identifizierten Trajektorien

Unsicherheitsmodellierungsart	Anzahl	Berechnungsdauer*)	Tr./Min.**)
Zufällige Variabilität	9 243	7:44	6,72
Fuzzy-Theorie	15 206	10:34	11,06
Fuzzy-Randomness	15 519	19:23	11,68

*)Zeitangabe des Wertes „Berechnungsdauer“ mit [Stunden : Minuten]
**) bezüglich des Erfassungszeitraumes von 22 Stunden und 55 Minuten

Augenscheinlich erforderten die Berechnungen mit Anwendung der zufälligen Variabilität den geringsten Zeitbedarf (vgl. Spalte Berechnungsdauer). Da im Rahmen der Softwareimplementierung Rechenzeitoptimierungen der Algorithmen nur eine nebengeordnete Rolle spielten, sind Vergleiche der Unsicherheitsmodelle auf Ebene der Rechenzeiten nur bedingt aussagekräftig. Zukünftig sind z. B. mit moderneren Rechenkernen, Optimierungen des Programmcodes oder durch eine Parallelisierung der Auswerteschritte Verkürzungen bei der Berechnungsdauer erzielbar.

Ebenfalls fällt auf, dass die identifizierten relativen Mengen zwischen den Unsicherheitsmodellierungsarten nicht den relativen Anzahlen entsprechen, wie sie anhand der in Abschnitt 6.3.5 ermittelten Effektivitäten zu erwarten wären. Im Falle der zufälligen Variabilität ($\eta_{MIV} = 0{,}306$) wäre zu erwarten, dass mit ihr in etwa die gleiche Anzahl Trajektorien zu ermitteln sei, wie im Falle des Fuzzy-Randomness ($\eta_{MIV} = 0{,}359$). Die geringste Anzahl wäre bei Anwendung der Fuzzy-Theorie ($\eta_{MIV} = 0{,}211$) zu erwarten. Aufgrund der zum Teil geringen Anzahl an Referenzdaten (Versuchsfahrten) muss angenommen werden, dass die im Rahmen der Kalibrierung ermittelten Effektivitätswerte nur bedingt auf die Auswertung von Massendaten übertragen werden können.

Die Abbildung 6-17 zeigt die Mengenverhältnisse der verworfenen Trajektorien in Abhängigkeit vom Identifizierungsverfahren. Für diese Trajektorien konnten nicht oder nicht eindeutig klassifiziert werden, dass sie einem Verkehrsmittel des MIV angehören. In Tabelle 6-6 sind die jeweiligen Trajektorienanzahlen zusammengefasst. Man erkennt, dass der größte Anteil der Trajektorien in der *Bewegungsidentifikation* (Bewegung) verworfen wurde.

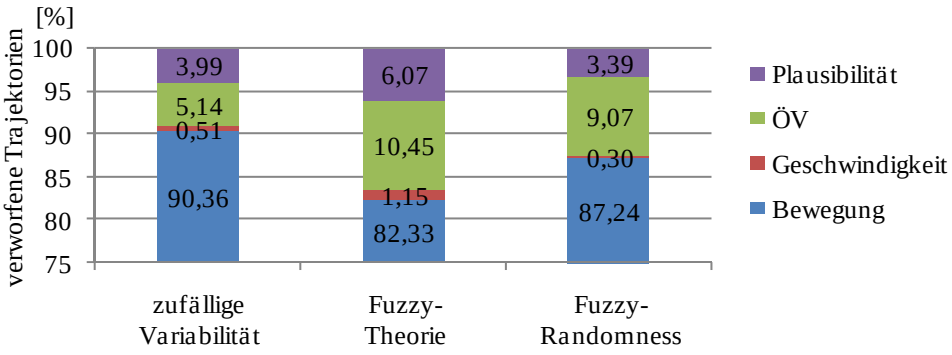


Abbildung 6-17: Mengenverhältnisse der verworfenen Trajektorien für die Identifizierungsverfahren

Tabelle 6-6: Anzahlen je Identifizierungsverfahren der verworfenen und als MIV klassifizierten Trajektorien

Anzahl verworfener Trajektorien bei:	Zufällige Variabilität	Fuzzy-Theorie	Fuzzy-Randomness
Bewegung	378 739	338 367	359 005
Geschwindigkeit	2 131	4 736	1 236
ÖV-Korrelation	21 538	42 936	37 313
Plausibilität	16 726	24 946	13 971
Als MIV klassifiziert	9 243	15 206	15 519

Der zweite höchste Anteil wurde im Verfahren der *Korrelation zu ÖV-Fahrzeugen* verworfen. Dieser Anteil kann damit erklärt werden, dass das Versuchsgebiet im Großraum der Großstadt Karlsruhe mit zahlreichen ÖV-Angeboten lag und die Menge der ÖV-Fahrzeuge infolgedessen hoch war. Zum einen kann angenommen werden, dass in diesem Anteil tatsächlich in ÖV-Fahrzeugen generierte Mobilfunkdaten enthalten sind. Zum anderen führen bei einer hohen Dichte von ÖV-Fahrzeugen die (hohen) Ortungsunsicherheiten dazu, dass Korrelationen zu ÖV-Fahrzeugen zu Unrecht nachgewiesen werden. Wie der Abbildung 6-17 und der Tabelle 6-6 entnommen werden kann, ist dieser Anteil im Falle der Fuzzy-Theorie im Vergleich zu den beiden anderen Unsicherheitsmodellierungsarten am höchsten. Im Vergleich zur zufälligen Variabilität sogar etwa doppelt so hoch. Das war aufgrund der Untersuchungen in Abschnitt 6.3.3 (die dort ermittelten Effektivitäten waren $\eta_{\underline{\overline{ÖV}}}^{\text{zufällige Variabilität}} = 62\%$, $\eta_{\underline{\overline{ÖV}}}^{\text{Fuzzy-Theorie}} = 66\%$, $\eta_{\underline{\overline{ÖV}}}^{\text{Fuzzy-Randomness}} = 78\%$) so nicht zu erwarten, da die festgestellten Effektivitäten mit 4% Unterschied zur zufälligen Variabilität in etwa gleich waren und zum Fuzzy-Randomness ca. 12% geringer waren. Im Rahmen der Arbeit durchgeführte empirische Untersuchungen bezüglich räumlicher Abhängigkeiten führten zu keiner eindeutigen Aussage. Korrelationen zu ÖV-Fahrzeugen wurden mit Anwendung der Fuzzy-Theorie im Vergleich die anderen Unsicherheitsmodellierungsarten nahezu ortsunabhängig generell häufiger festgestellt. Für detailliertere Untersuchungen und insbesondere für eine „gesichertere“ Bestimmung der Kalibrierparameter wären umfangreichere Referenzdaten (Versuchsfahrten in ÖV-Fahrzeugen) erforderlich, die der vorliegenden Arbeiten nicht zur Verfügung standen.

Der dritthöchste Anteil der verworfenen Trajektorien entstand in der *kartengestützten Plausibilitätsprüfung*. Ursache für das Verwerfen waren entweder, dass keine Trajektorie gefunden wurde oder nur auf Fußwegen ($G^{\text{Fußweg}}$) oder sowohl auf Fußwegen als auch auf Straßen ($G^{\text{Straße}}$) Trajektorien gefunden wurden. In Tabelle 6-7 sind die Mengen für die Suche nach Trajektorien auf $G^{\text{Fußweg}}$ und $G^{\text{Straße}}$ zusammengestellt. Aufgrund der vorgeschalteten Identifizierungsverfahren (Bewegung, Geschwindigkeit, ÖV-Korrelation) sind die Ausgangsmengen unterschiedlich groß, vgl. die Zeile *Summe* in Tabelle 6-7. Die Grundgesamtheiten waren somit nicht identisch. Auffällig ist, dass für Fuzzy-Randomness deutlich seltener keine Trajektorie gefunden wurde und im Gegenzug ca. 59% mehr Trajektorien als bei Anwendung der zufälligen Variabilität als MIV klassifiziert wurden.

Tabelle 6-7: Mengenverhältnisse für die Suche nach Trajektorien

Auf Verkehrsnetz gefundene Trajektorie	Zufällige Variabilität	Fuzzy-Theorie	Fuzzy-Randomness
weder $G^{\text{Straße}}$ noch $G^{\text{Fußweg}}$	15 988	12 497	4 508
$G^{\text{Straße}}$ und $G^{\text{Fußweg}}$	751	12 437	9 483
nur $G^{\text{Fußweg}}$	0	12	10
nur $G^{\text{Straße}}$ →MIV	9 243	15 206	15 519
Summe	25 982	40 152	29 520

Durch das Identifikationsverfahren *Identifikation des Verkehrsmittels anhand der Geschwindigkeit* (Geschwindigkeit) wurden die geringsten Anzahlen von Trajektorien verworfen. Stichprobenartige Untersuchen zeigten, dass in der Regel Ausreißer bei der Ortung Ursache für hohe Geschwindigkeiten waren, so dass

signifikant höhere Geschwindigkeiten als die in Abschnitt 4.4 festgesetzte $v_{max,VM} = 250$ km/h für Fahrzeuge des MIV berechnet wurden.

Die folgenden Abbildungen zeigen die MS-Positionen sowie deren Häufigkeit auf der bereits in Abbildung 6-1 verwendeten Karte. Für die Berechnung der Häufigkeiten wurden die MS-Positionen gezählt, die sich innerhalb derselben Kachel (Seitenlänge 100 m) befanden. Grundlagen dieser Zählungen bildeten im Fall der zufälligen Variabilität und des Fuzzy-Randomness die zufälligen Mittelpunkte sowie für die Anwendung der Fuzzy-Theorie die Mittelpunkte der Zugehörigkeitsfunktionen. Die Unsicherheiten der Ortungen sind aus den Abbildungen somit nicht unmittelbar ablesbar.

In Abbildung 6-18a) wurden alle 35 791 435 Mobilfunkortungen dargestellt, die mittels der Mobilfunkortung bei Anwendung der zufälligen Variabilität berechnetet wurden. Analog dazu zeigen Abbildung 6-18b) diejenigen mit Anwendung der Fuzzy-Theorie sowie Abbildung 6-18c) mit Anwendung des Fuzzy-Randomness. Zur Darstellung der Häufigkeitsunterschiede wurden unterschiedliche Grün-Töne gewählt, wobei dunklere Töne auf eine höhere Häufigkeit hinweisen. Man erkennt bei allen drei Abbildungen deutlich die höheren Häufigkeiten entlang der Verkehrsachsen und im Ballungszentrum (Bildmitte).

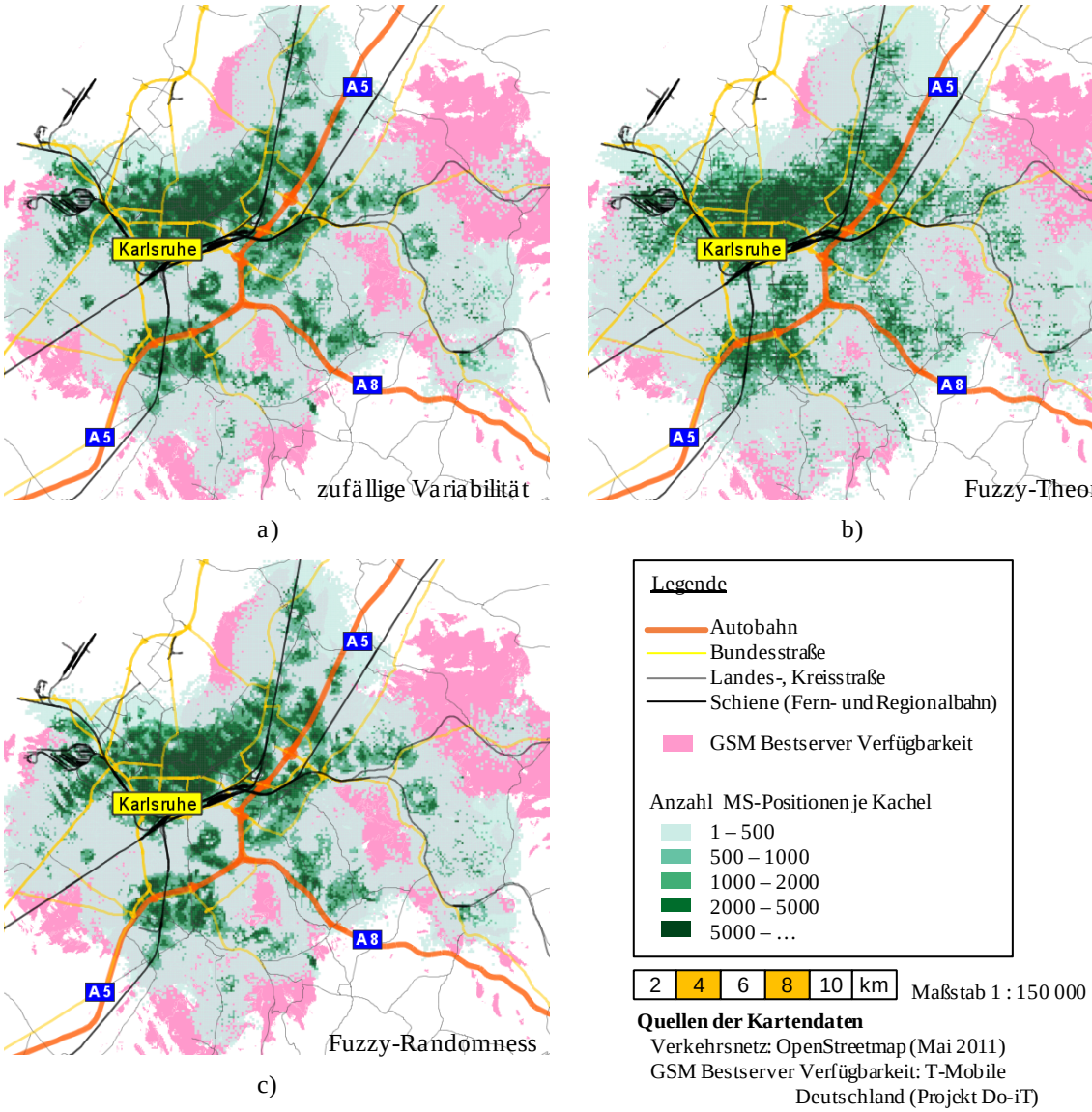


Abbildung 6-18: Geortete (noch nicht identifizierte) MS-Positionen und deren Häufigkeit in den Kacheln mit Anwendung a) der zufälligen Variabilität, b) der Fuzzy-Theorie und c) des Fuzzy-Randomness

Die Karten in Abbildung 6-19 zeigen die Clusterpositionen der MS, deren Trajektorien als MIV identifiziert wurden. Die Abbildung 6-19a) zeigt die entsprechenden MS-Positionen mit Anwendung der zufälligen Variabilität, die Abbildung 6-19b) mit Anwendung der Fuzzy-Theorie und die Abbildung 6-19c) mit Anwendung des Fuzzy-Randomness. Die verwendete Farb-Skala unterscheidet sich um den Faktor 100 im Vergleich zu den Karten in Abbildung 6-18 und wurde zur Verdeutlichung im blauen Farbton gewählt. Dunklere Blau-Töne zeigen eine höhere Häufigkeit im Vergleich zu helleren Blau-Tönen an. Wie in den vorherigen Abbildungen auch, wurden die Häufigkeiten der Kacheln ausschließlich für geschätzten Mittelpunkte im Falle der zufälligen Variabilität und des Fuzzy-Randomness sowie für die Mittelpunkte der Zugehörigkeitsfunktionen im Falle der Fuzzy-Theorie der Positionen berechnet. Unsicherheiten oder die tatsächlichen MS-Positionen sind daraus nicht unmittelbar ablesbar.

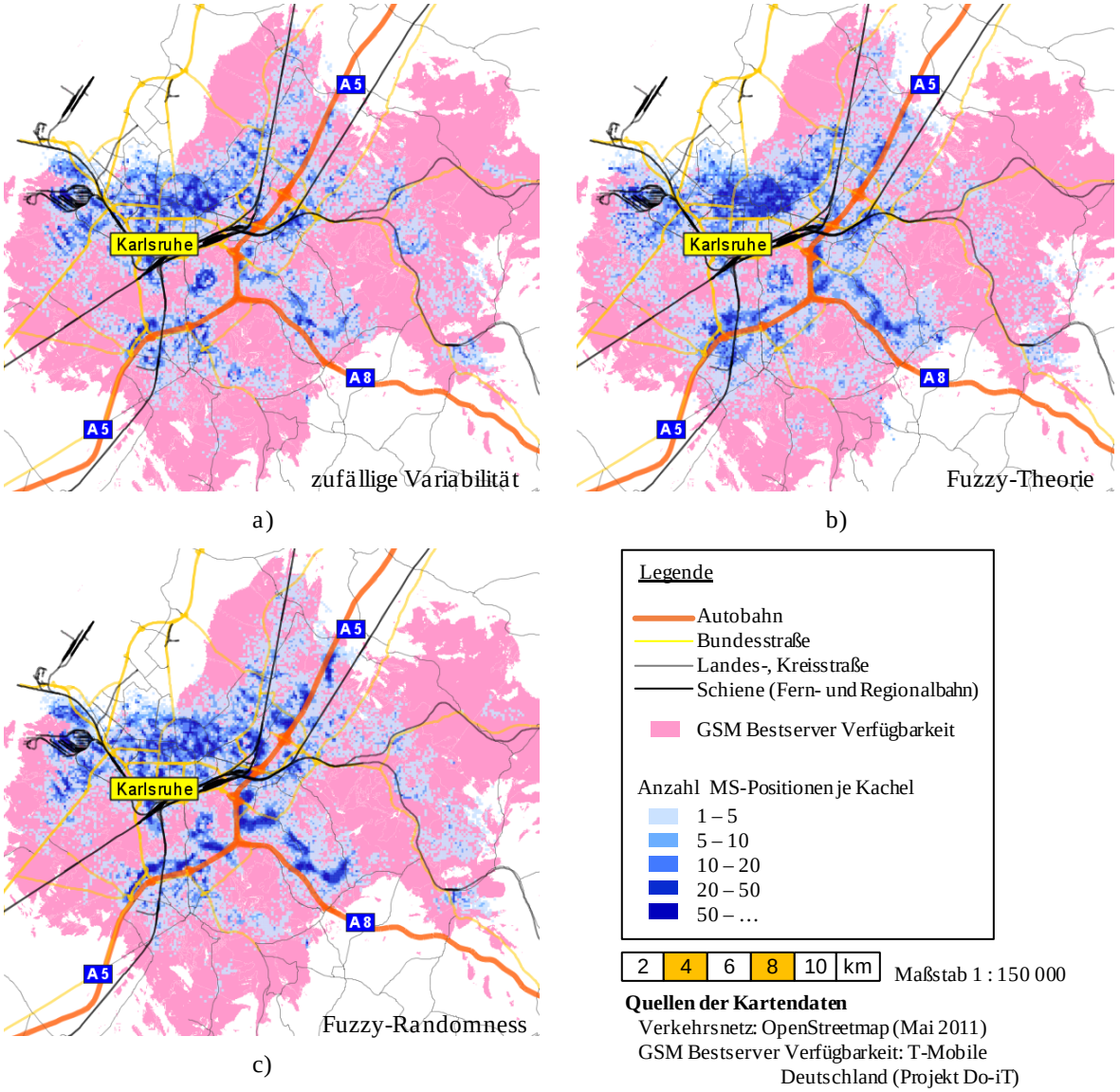


Abbildung 6-19: Clusterpositionen und deren Häufigkeiten in den Kacheln für die als MIV identifizierten Trajektorien bei Anwendung a) der zufälligen Variabilität, b) der Fuzzy-Theorie und c) des Fuzzy-Randomness

Zum Zweck der Verkehrslageschätzung können aus den Trajektorien Verkehrsdaten (beispielsweise Ort und Geschwindigkeit der MS) zu spezifischen Zeitpunkten berechnet werden. Unter der Voraussetzung, dass jede als MIV identifizierte Trajektorie einem anderen Fahrzeug des MIV angehört, ist für die Verkehrslageerfassung die Anzahl der zu gleichen Zeitpunkten erfassten Trajektorien von Interesse. Im Vergleich zu klassischen Messstellen wie beispielsweise den Induktionsschleifen, an denen an festen Standorten die Menge und die Geschwindigkeit der Fahrzeuge bestimmt wird, können anhand der Trajektorien an unterschiedlichen Orten - in Abhängigkeit der Fahrzeugpositionen - die Flussgeschwindigkeiten an diesen Orten ermittelt werden. Die Anzahl der gleichzeitig vorhandenen Trajektorien kann somit als Anzahl von Messstellen für die Geschwindigkeitsmessung im Verkehrsnetz verstanden werden, wie sie auch – jedoch nur an wenigen ausgewählten sowie festen Orten – mit den klassischen Sensoren (z. B. den Induktionsschleifen) vollzogen wird. Die Anzahl an Trajektorien, die zum selben Zeitpunkt vorhanden ist, wird im Folgenden als **Messwertaufkommen** bezeichnet.

Abbildung 6-20 zeigt die tageszeitliche Verteilung des Messwertaufkommens. Insbesondere zwischen ca. 9:00 Uhr und ca. 19:00 Uhr ist das höhere Messwertaufkommen im Falle des Fuzzy-Randomness (obere Kurve, grün) im Vergleich zur zufälligen Variabilität (blaue Kurve) und der Fuzzy-Theorie (rote Kurve) erkennbar. Durchschnittlich betrug das Messwertaufkommen in diesem Zeitraum bei Anwendung der zufälligen Variabilität ca. 49,6 Trajektorien, der Fuzzy-Theorie ca. 66,4 Trajektorien und des Fuzzy-Randomness ca. 77,3 Trajektorien. Analog den Erläuterungen zur Tabelle 6-5 auf Seite 128 fällt die geringe Anzahl an Trajektorien im Falle der zufälligen Variabilität auf. Anhand der in Abschnitt 6.3.5 bestimmten Effektivitätswerte, wäre für diese Unsicherheitsmodellierungsart in etwa die gleiche Anzahl zu erwarten gewesen, wie bei Anwendung des Fuzzy-Randomness.

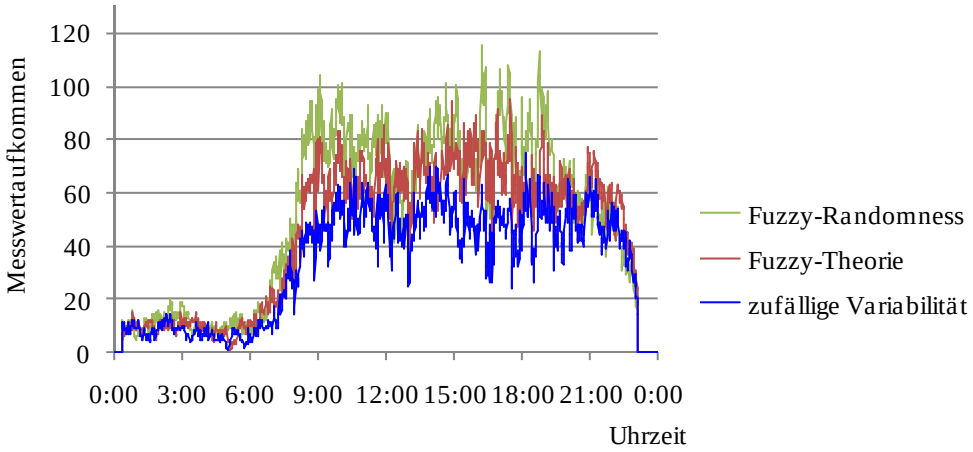


Abbildung 6-20: tageszeitliche Verteilung des Messwertaufkommens

Im Großraum der Stadt Stuttgart werden im Vergleich dazu ca. 200 stationäre Messstellen (infrastrukturgebundene Sensoren wie z. B. Induktionsschleifen und Infrarotsensoren) für die Erfassung der aktuellen Verkehrslage eingesetzt (Stand 2008). Zusätzlich sind in ca. 700 Taxis FCD Systeme installiert, die jedoch nicht alle gleichzeitig im Einsatz sind. Um ca. 50 neue Messstellen einzurichten, entstehen Kosten von 450 T€ (STGTZ 2008). In LAUFER 2011 wird mit qualitätsbeschreibenden Merkmalen gezeigt, dass Verkehrsstärken, die mittels FPD-Trajektorien auf Grundlage von Mobilfunkdaten ermittelt wurden, mit denen aus Induktionsschleifen vergleichbar sind, vgl. Abbildung 6-21. Berücksichtigt man zudem, dass für die vorliegenden Auswertung ausschließlich die Mobilfunkdaten der T-Mobile Deutschland verwendet wurden – in Deutschland sind insgesamt vier Mobilfunknetzanbieter T-Mobile Deutschland, Telefónica O2, Vodafone und E-Plus vertreten

(Stand 2013) – und dass die Durchdringungsrate von MS in der Bevölkerung sowie die Nutzung unter anderem durch Web-, Mail- und Multimedia-Funktionen (SERVICEREPORT 2013) zunimmt, dann zeigt sich, dass die Verkehrslageerfassung mittels Mobilfunkdaten bei ähnlichen Qualitätsmerkmalen einen erheblich finanziellen Gegenwert mit sich bringt. Die Aufwände für Wartung und Instandhaltung der infrastrukturgebundenen Sensoren sind in diesen Kosten noch nicht enthalten.

Das ca. 10% bis 25% höhere Messwertaufkommen im Falle der Anwendung des Fuzzy-Randomness im Vergleich zur Fuzzy-Theorie und zur zufälligen Variabilität kann auch unter dem Gesichtspunkt bewertet werden, dass durch Wahl dieses Unsicherheitsmodells bei ähnlichen Messwertanzahlen, Kosten für die Errichtung weiterer infrastrukturgebunden Sensoren oder auch Kosten für die Anbindung eines weiteren Mobilfunknetzes eingespart werden könnten.

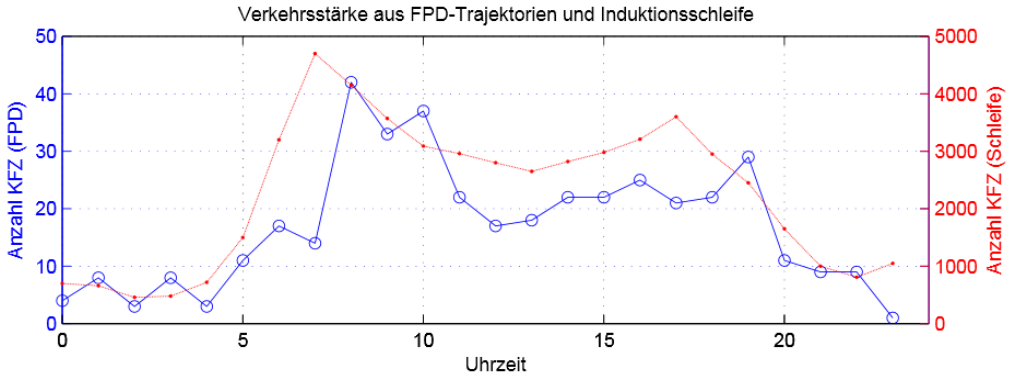


Abbildung 6-21: Vergleich der Verkehrsstärken aus FPD-Trajektorien (Mobilfunkdaten) mit denen aus einer Induktionsschleife (Quelle: LAUFER 2011)

6.4.3 Verknüpfung der Identifikationsergebnisse

In Abschnitt 4.7 wurde für im MIV transportierte MS angenommen, dass sie für eine Zeitspanne von $dt_{MIV,max}$ keine Verkehrsmittelwechsel vornehmen. Es wurde angenommen, dass MS innerhalb dieser Zeitspanne im jeweils selben Verkehrsmittel verbleiben, auch wenn keine Mobilfunkdaten vorliegen. Wurde somit eine Trajektorie einer MS als MIV klassifiziert, dann werden nachfolgende Mobilfunkdaten ebenfalls als MIV klassifiziert, und werden dadurch verknüpft. Das gilt, solange entweder für die Mobilfunkdaten der Ausschluss des MIV[18] nachgewiesen wird oder die Zeitspanne von $dt_{MIV,max}$ zwischen der letzten MS-Position der Vorgängertrajektorie und den zu identifizierenden Mobilfunkdaten überschritten wurden. Mobilfunkdaten einer stehenden MS können infolgedessen ebenfalls als MIV klassifiziert werden, wenn die MS bereits zuvor als MIV klassifiziert wurde und die aktuelle MS-Position auf dem Straßenverkehrsnetz - innerhalb der Unsicherheiten - liegt. Die Anzahl der als MIV klassifizierten Mobilfunkdaten vergrößert sich dadurch, da zum einen „MS im Stand“ und zum anderen in der kartengestützten Plausibilitätsprüfung uneindeutig – sowohl NMIV als auch MIV – klassifizierte Mobilfunkdaten als MIV klassifiziert werden können.

Der Zahlenwert für $dt_{MIV,max}$ ist subjektiv vom Auswerter im Vorfeld festzulegen. Um die Auswirkungen von $dt_{MIV,max}$ auf die Trajektorienanzahlen aufzuzeigen, wurden die Identifikationsergebnisse aus Abschnitt 6.4.2 für unterschiedlich gewählte Zeitspannen $dt_{MIV,max}$ miteinander verknüpft.

Abbildung 6-22 zeigt die resultierenden Trajektorienanzahlen. Die Ergebnisse der zufälligen Variabilität sind in Dunkelblau, die der Fuzzy-Theorie in Rot und die des Fuzzy-Randomness in Grün dargestellt.

[18] Der Ausschluss kann aufgrund zu hoher Geschwindigkeit, einer Korrelation zum ÖV oder weil keine Einpassung in das Straßenverkehrsnetz möglich ist, erfolgen.

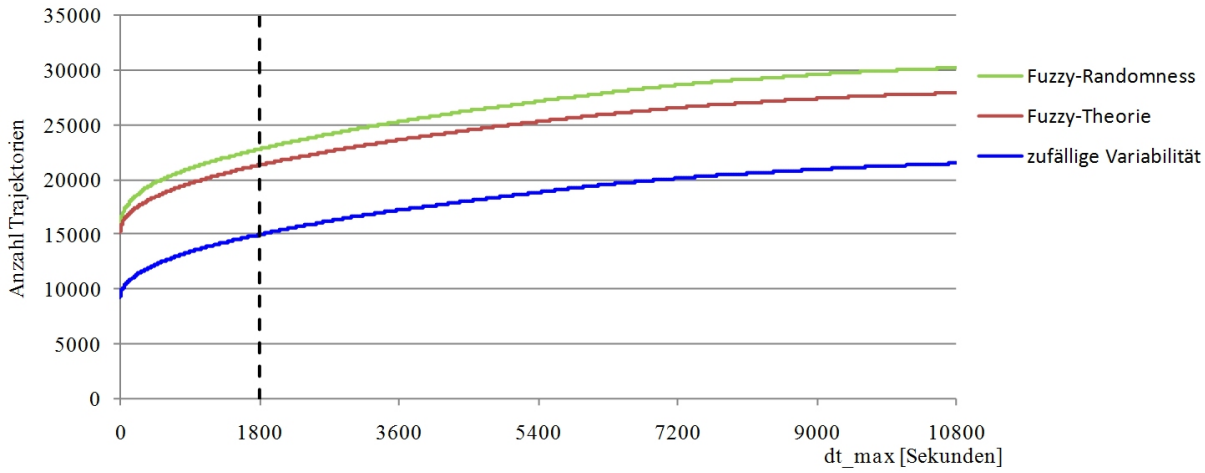


Abbildung 6-22: Trajektorienanzahlen in Abhängigkeiten von der gewählten Zeitspanne $dt_{MIV, max}$

Durch die Verknüpfung der Identifizierungsergebnisse werden zusätzliche – bislang nicht oder nicht eindeutig als MIV klassifizierte – Trajektorien der Verkehrsklasse des MIV zugeordnet. Das Messwertaufkommen erhöht sich infolgedessen ebenfalls.

Für eine Beispielauswertung mit $dt_{MIV,max} = 1800$ Sekunden zeigt die Abbildung 6-23 die tageszeitlichen Verläufe des Messwertaufkommens. Gegenüber den tageszeitlichen Verläufen ohne Verknüpfung der Identifizierungsergebnisse in Abbildung 6-20 sind das höhere Messwertaufkommen erkennbar. Durchschnittlich wurden prozentual die folgenden höheren Messwertaufkommen erzielt: Im Falle der zufälligen Variabilität ca. 15,4%, der Fuzzy-Theorie ca. 17,2% und des Fuzzy-Randomness ca. 11,0%. Im Zeitraum zwischen ca. 9:00 Uhr und ca. 19:00 Uhr, wie er für die Analyse der unverknüpften Identifizierungsergebnisse ausgewählt wurde, wurden Messwertaufkommen von durchschnittlich ca. 56,9 Trajektorien (unverknüpft ca. 49,6 Trajektorien) bei Anwendung der zufälligen Variabilität, ca. 78,0 Trajektorien (unverknüpft ca. 66,4 Trajektorien) der Fuzzy-Theorie und ca. 85,8 Trajektorien (unverknüpft ca. 77,3 Trajektorien) des Fuzzy-Randomness erreicht. Die geringere Anzahl im Falle der zufälligen Variabilität ist im Wesentlichen Folge der geringeren Anzahl unverknüpfter Trajektorie. Die aufgrund der Effektivitäten in Abschnitt 6.3.5 zu erwartenden relativen Anzahlen, im Vergleich zu den tatsächlich ermittelten Anzahlen, wurden bereits auf Seite 128 und auf Seite 132 erwähnt.

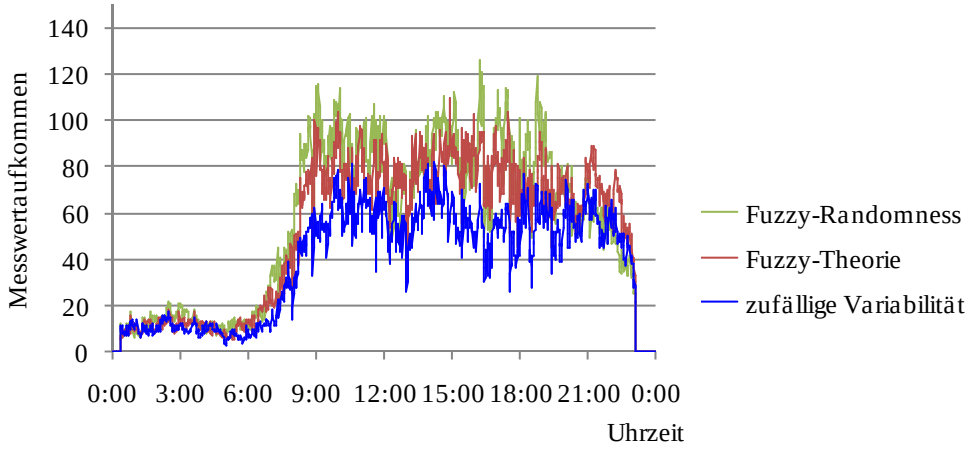


Abbildung 6-23: tageszeitliche Verteilung des Messwertaufkommens mit Verknüpfung der Identifizierungsergebnisse mit $dt_{MIV, max}$ = 1800 Sekunden

Mittels Verknüpfung der Identifizierungsergebnisse können somit zusätzliche Messwerte für die Verkehrslageerfassung gewonnen werden. Kostenintensive Installationen von stationären Sensoren, wie Induktionsschleifen oder Infrarotsensoren, sowie deren Wartung und Instandhaltung könnten infolgedessen vermieden werden.

Außerdem kann mit der Verknüpfung der Identifizierungsergebnisse die verfahrensbedingte Lücke der Bewegungsidentifikation geschlossen werden, dass ausschließlich sich in Bewegung befindliche MS als MIV klassifiziert werden können. Die insbesondere für die aktuelle Verkehrslage interessierenden Informationen, ob und an welchen Stellen im Verkehrsnetz Verkehrsstaus auftreten, werden somit prinzipiell auch mit den als MIV klassifizierten Mobilfunkdaten ermöglicht.

7 Zusammenfassung und Ausblick

7.1 Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Nutzung von Mobilfunkdaten des Global System for Mobile Communications (GSM) für die Verkehrslagerfassung des motorisierten Individualverkehrs (z. B. PKW, LKW). Wesentlicher Hintergrund ist, dass aus den Mobilfunkdaten Ortsinformationen (geografische Koordinaten) der Mobilfunkgeräte ableiten zu können. Von der Infrastruktur der Mobilfunknetze werden dazu anonymisierte Mobilfunkdaten abgegriffen. Aufgrund der hohen Durchdringungsrate in der Bevölkerung – jeder Bundesbürger besitzt statistisch betrachtet mindestens ein Mobilfunkgerät – wird dadurch eine kostengünstige Erfassung des Verkehrs ermöglicht. Diesen Vorteilen stehen hohe Ortungsunsicherheiten von bis zu mehreren hundert Metern sowie die Unkenntnis, ob und wenn ja in welchen Verkehrsmitteln die Mobilfunkgeräte transportiert wurden, gegenüber.

Ziel der vorliegenden Arbeit war es, geeignete Identifikationsverfahren und Modelle zur Behandlung der Unsicherheiten zu entwickeln, um aus anonymisierten Mobilfunkdaten diejenigen zu identifizieren, die in Verkehrsmitteln des motorisierten Individualverkehrs generiert wurden. Aus diesen selektierten Mobilfunkdaten resultieren die als Floating Phone Data bezeichneten Verkehrslagedaten. Die dafür eingesetzte Mobilfunkortung basiert auf dem Signalpegel-Matching, das die gemessenen Signalpegel der Mobilfunkgeräte mit Referenzpegelkarten vergleicht. Sowohl die Unsicherheiten der gemessenen Signalpegel als auch die Unsicherheiten der Referenzpegelkarten sind Resultat zahlreicher zufälliger, systematischer aber auch unbekannt wirkender Unsicherheiten. Um ihre Eignung und ihr Potential zu vergleichen, werden die Unsicherheiten der Mobilfunkortung mit Anwendung der zufälligen Variabilität, der Fuzzy-Theorie und des Fuzzy-Randomness beschrieben. Mit konsequenter Fortpflanzung der Unsicherheiten werden mathematische Unsicherheitsmodelle für folgende Identifizierungsverfahren, die softwaretechnisch umgesetzt wurden, vorgestellt:

- In der *Bewegungsidentifikation* wird der Frage nachgegangen, ob sich das Mobilfunkgerät in Bewegung befindet. Den Mobilfunkgeräten wird dazu hypothetisch eine Bewegung unterstellt. Aufgrund der hohen Positionsanzahlen aus der Mobilfunkortung – mit einer Rate von bis zu 2 Hz – im Vergleich zu ihren hohen metrischen Unsicherheiten wird ein Fuzzy c-Means Clustering angewendet. Das Zutreffen der hypothetisch angenommenen Bewegung wird abschließend überprüft.
- Bauartbedingt oder aus administrativen Gründen sind die Höchstgeschwindigkeiten von Verkehrsmitteln begrenzt. Mit dem Verfahren zur *Identifikation des Verkehrsmittels anhand der Geschwindigkeit* wird geprüft, ob die Geschwindigkeit des Mobilfunkgerätes höher als die Höchstgeschwindigkeit des untersuchten Verkehrsmittels war. Ist das der Fall, wird die Nutzung dieses Verkehrsmittel ausgeschlossen.
- Mit dem Identifizierungsverfahren *Korrelation zu ÖV-Fahrzeugen* werden die Positionen von Fahrzeugen des Öffentlichen Verkehrs (ÖV) mit den Positionen des Mobilfunkgerätes verglichen. Die Positionen der ÖV-Fahrzeuge werden aus Fahrplänen prognostiziert. Sind die Positionen des Mobilfunkgerätes und eines ÖV-Fahrzeugs (z. B. Linienbusse, Straßenbahnen) sowohl zeitlich als auch räumlich miteinander korreliert, kann die Nutzung dieses ÖV-Fahrzeugs nicht ausgeschlossen werden. Weisen Mobilfunkdaten eine derartige Korrelation auf, dann sind sie für die Verkehrserfassung des motorisierten Individualverkehrs ungeeignet und werden herausgefiltert.

- Mit der *kartengestützten Plausibilitätsprüfung* wird geprüft, ob für die Positionsfolge eines Mobilfunkgerätes eine Trajektorie auf dem Verkehrsnetzgraph des motorisierten Individualverkehrs gefunden werden kann. Nur wenn diese Positionsfolge in den Verkehrsnetzgraph räumlich, topologisch und zeitlich in Form einer Trajektorie eingepasst werden kann, gilt sie im Sinne der Erfassung für den motorisierten Individualverkehr als plausibel. Insofern für diese Positionsfolge aber auch eine Trajektorie auf dem Fußwegenetz existiert, ist die Identifizierung uneindeutig, da eine Trennung zwischen nicht-motorisiertem Individualverkehr und motorisiertem Individualverkehr nicht möglich ist. Derartige, gegebenenfalls von Fußgängern generierte, Mobilfunkdaten werden herausgefiltert.

Um Schlussfolgerungen aus den Mobilfunkdaten ziehen zu können, erfordern die Identifikationsverfahren Grenzwerte für die Parameter der Unsicherheitsmodelle. Es wurde gezeigt, dass die Wertewahl dieser Parameter sowohl die Korrektheit als auch die Anzahl der identifizierten Mobilfunkgeräte (als Maß der Effektivität) des motorisierten Individualverkehrs und damit die Effektivität der Identifizierungsverfahren beeinflusst. Im Rahmen einer Kalibrierung der Verfahren wurden die Parameterwerte unter Verwendung realer Mobilfunkdaten so ausgewählt, dass sowohl Korrektheit als auch Effektivität in einem ausgewogenen Verhältnis standen. Mit Anwendung des Fuzzy-Randomness wurden zum einen sowohl die höchste Korrektheit von ca. 83,8% als auch die höchste Effektivität von ca. 35,9% erzielt. Die zweitbesten Ergebnisse konnten mit den Modellen der zufälligen Variabilität mit einer Korrektheit von ca. 79,8% und einer Effektivität von ca. 30,6% sowie die drittbesten Ergebnisse mit Anwendung der Fuzzy-Theorie mit einer Korrektheit von ca. 77,5% und einer Effektivität von ca. 21,1% erzielt werden.

Für ein Versuchsgebiet mit der Flächengröße von ca. 350 km^2 wurden unter realen Bedingungen alle Mobilfunkdaten der T-Mobile Deutschland für einen Tag erfasst, anonymisiert, geortet, die im motorisierten Individualverkehr generierten Mobilfunkdaten mit den entwickelten Identifikationsverfahren identifiziert und die Menge der Verkehrsdaten (Messwertaufkommen) bestimmt. Das Messwertaufkommen spiegelt die Anzahl an Mobilfunkdaten wider, die für die Verkehrslageerfassung zu einem spezifischen Zeitpunkt genutzt werden können und wird in der Einheit Trajektorienanzahl angegeben. Aus einer einzelnen Trajektorie kann beispielsweise die Verkehrsgeschwindigkeit und aus der Anzahl an Trajektorien an einem Ort die Verkehrsdichte abgeleitet werden. Im Zeitraum von 9:00 Uhr bis 19:00 Uhr wurde mit Anwendung des Fuzzy-Randomness das höchste Messwertaufkommen mit einer durchschnittlichen Trajektorienanzahl von ca. 77 gleichzeitig verfügbarer Trajektorien ermittelt. Das zweithöchste Messwertaufkommen konnte mittels der Fuzzy-Theorie mit durchschnittlich ca. 66 und mit Anwendung der zufälligen Variabilität das dritthöchste Messwertaufkommen mit durchschnittlich ca. 50 gleichzeitig verfügbarer Trajektorien erzielt werden. Anhand der erwarteten Effektivitäten aus der Kalibrierung zeigte sich, dass sich aufgrund der geringen Anzahl der zur Verfügung gestandenen Referenzdaten die Effektivitätswerte nicht unmittelbar auf die Auswertung von realen Massendaten übertragbaren ließen.

Für stehende Mobilfunkgeräte ist in der Regel keine eindeutige Verkehrsmittelidentifikation möglich, da jedes Verkehrsmittel stehen kann. Eine Unterscheidung zwischen Mobilfunkgeräten in Verkehrsstaus und beispielsweise nicht am Verkehr beteiligten Mobilfunkgeräten (z. B. stehende Fußgänger) ist infolgedessen zunächst nicht möglich. Diese Problemstellung wurde durch die Verknüpfung zeitlich aufeinander folgender Identifizierungsergebnisse und damit über die Einbeziehung zeitlich früherer Identifizierungsergebnisse gelöst. Grundvoraussetzung für die Verknüpfung ist die Annahme, dass weiterhin dasselbe Verkehrsmittel vom Mobilfunkteilnehmer genutzt wurde. Für eine exemplarische Verknüpfung der Identifizierungsergebnisse von bis zu 1800 Sekunden wurde gezeigt, dass sich das Messwertaufkommen für das Fuzzy-Randomness um ca. 11%, für die Fuzzy-Theorie um ca. 17% und für die zufällige Variabilität um ca. 15% erhöht.

Es wurde gezeigt, dass aus anonymisierten GSM-Mobilfunkdaten mit den vorgestellten Identifizierungsverfahren diejenigen Mobilfunkdaten herausgefiltert werden können, die im motorisierten Individualverkehr generiert wurden. Aufgrund der durchgängigen Unsicherheitsmodellierung kann zum einen die Ergebnismenge anhand der Wertewahl der Unsicherheitsparameter in Bezug auf Effektivität (mehr Daten) oder Korrektheit (korrektere Daten) optimiert werden. Zum anderen können die Identifizierungsverfahren auf Ortungsdaten anderer Quelle als GSM übertragen werden. In den empirischen Analysen zeigte sich, dass mit Anwendung der Unsicherheitsmodelle des Fuzzy-Randomness sowohl die meisten als auch die korrektesten Informationen erzielbar waren.

7.2 Ausblick

Mit den entwickelten Identifikationsverfahren und ihren Unsicherheitsmodellen wurde gezeigt, dass aus anonymisierten Mobilfunkdaten die im motorisierten Individualverkehr transportierten Mobilfunkgeräte herausgefiltert werden können. Gleichzeitig bietet die vorliegende Arbeit mögliche Ausgangspunkte, an die zukünftige Forschungsarbeiten mit folgenden Themen anschließen könnten:

- Zur Schätzung der Unsicherheitswerte wurden rechenintensive Algorithmen eingesetzt. Die Formulierung – rechentechnisch einfacher – funktionaler Zusammenhänge zur effizienten Unsicherheitsschätzung könnte die Effizienz der Mobilfunkortung und der Identifizierung erhöhen.
- In die Entscheidungsfindung der Identifizierungsverfahren fließen maßgeblich die geometrischen Unsicherheiten der Mobilfunkortung mit ein. Es kann erwartet werden, dass mit Verbesserung der geometrischen Unsicherheiten auch höhere Effektivitätswerte der Identifizierungsverfahren folgen. Inwieweit geringere geometrische Unsicherheiten den Effektivitätsgrad tatsächlich erhöhen können, wird zu prüfen sein.
- Bereits heutige Mobilfunkgeräte nutzen Kommunikationsnetze mit unterschiedlichen Kommunikationstechniken (z. B. GSM, UMTS, WLAN) für den Datentransport. Um die kontinuierliche Nachverfolgung der Mobilfunkgeräte zu gewährleisten, ist eine Verknüpfung von Ortungsinformation der Kommunikationsnetze anzustreben.
- Verkehrsdaten, die von mehreren Mobilfunkgeräten desselben Fahrzeugs (z. B. ein Reisebus) generiert werden, können bei der Verkehrslageschätzung zu unrichtigen Ergebnissen führen. Ein Lösungsweg könnte die Suche nach gleichen Identifikationsmerkmalen in Raum und Zeit, beispielweise durch Clusterverfahren sein.
- Die Werteauswahl für die Grenzwerte der Identifizierungsverfahren im Rahmen der Kalibrierung zeigte, dass mit Auswahl anderer als den hier ausgewählten Werten, durchaus höhere Korrektheiten oder Effektivitäten erzielbar sind. Ziel nachfolgender Arbeiten könnte die Suche nach Methoden oder Verfahren sein, deren Parameterwerte unabhängiger von den Referenzdaten einer Kalibrierung sind. Denkbar wären auch Verfahren, die selbstoptimierend sind oder (zumindest für einige) unabhängig von Auswahlen von Parameterwerten wären.
- Die Identifizierung von Fahrrädern könnte, insbesondere wegen des zunehmenden (passiven) Datentransfers durch Web-, Mail- und Multimedia-Funktionen der Mobilfunkgeräte im Fokus stehen.
- Außerdem könnten Verfahren zur Detektion von Verkehrsmittelwechseln, zur Bestimmung der beteiligten Verkehrsmittel intermodaler Wegeketten, Ansatzpunkte nachfolgender Arbeiten sein.

Literatur

3GPP (2009): 3GPP TS 45.008 V9.0.0, http://www.3gpp.org/ftp/Specs/html-info/45008.htm (Stand 02/2022)

ALEFELD, G. (1968): *Intervallrechnung über den komplexen Zahlen und einige Anwendungen*. Karlsruhe

ALEFELD, G. und HERZBERGER, J. (1974): *Einführung in die Intervallrechnung*. Reihe Informatik, Band 12, Bibliographisches Institut, Mannheim/Wien/Zürich

ANDERS, K.-H. (2003): *Parameterfreies hierarchisches Graph-Clustering Verfahren zur Interpretation raum-bezogener Daten*. Dissertation, Stuttgart

ASSENMACHER, W. (2003): *Deskriptive Statistik*. Springer, Berlin

BABU, G. P. und MURTY, M. N. (1994): *Clustering with evolutionary strategies*. Pattern Recognition, 27(2): pp. 321-329

BACHER, J. (2002), *Clusteranalyse*. R. Oldenbourg Verlag, 2. Auflage, München ; Wien : Oldenbourg

BANDEMER, H. und GOTTWALD, S. (1990): *Einführung in Fuzzy-Methoden*. Akademie Verlag, Berlin

BANDEMER, H. und NÄTHER, W. (1992): *Fuzzy Data Analysis*. Kluwer, Dodrecht

BAST (2011): Bundesanstalt für Straßenwesen, http://www.bast.de (unter: Qualitätsbewertung /Prüfungen/Straßenbetrieb) (Stand 02/2011)

BECTOR, C. R. und CHANDRA, S. (2005): *Fuzzy Mathematical Programming and Fuzzy Matrix Games*. Springer-Verlag, Berlin

BEETZ, A.; CZOMMER, R.; LAUFER, R. und SCHWIEGER, V. (2009): *Qualitätsbewertung Map-Matching gestützter Positionsbestimmungsverfahren auf Abis und A-Ebene (AP4.6) sowie Evaluierung der Positionsbestimmung durch höherwertige Ortungsverfahren (AP10.3)*. (interner) Bericht des Projektes Do-iT VM2010, September 2009 (unveröffentlicht)

BEETZ, A.; CZOMMER, R.; und SCHWIEGER, V. (2011): *Signalstärken-Matching und Map-Aiding-Methoden zur Positionsbestimmung von Mobilfunkteilnehmern in Echtzeit für Verkehrsprognosen*. ZfV, Jahrgang 2011, Heft 3, S. 150-164

BERKHIN, P. (2002): *Survey Of Clustering Data Mining Techniques*. Accure Research Papers, Accrue Software

BEZDEK, J. C. (1981): *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York

BIEWER, B. (1997): *Fuzzy-Methoden : praxisrelevante Rechenmodelle und Fuzzy-Programmiersprachen*. Springer-Verlag, Berlin

BLAUSTEIN, N. und ANDERSEN, J. B. (2002): *Multipath Phenomena in Cellular Networks*. Boston London, Artech House

BOTHE, H.-H. (1995): *Fuzzy-Logik, Einführung in Theorie und Anwendungen*. Springer, Berlin

BUSCH, F.; GLAS, F. und BERGMANN, E. (2004): Dispositionssysteme als FCD-Quellen für eine verbesserte Verkehrslagerekonstruktion in Städten. Straßenverkehrstechnik, Ausgabe 9/2004

CATREIN, D.; HELLEBRANDT, M.; MATHAR, R. und SERRANO, M. (2004): *Location Tracking of Mobiles: A Smart Filtering Method and Its Use in Practice*. Proceedings: IEEE VTC Spring 2004, Milan, May 2004, CD., p. 2677-2681

CC-BY-SA-LICENCE (2010): CC-BY-SA-Licence 3.0, (Creative Commons Attribution-Share Alike 3.0): http://creativecommons.org/licenses/by-sa/3.0/de/ (Stand 08/2011)

CHARTA EU 2007/C 303/01 (2007): *Charta der Grundrechte der Europäischen Union*. Amtsblatt der Europäischen Union 14.12.2007, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2007:303:0001:0016:DE:PDF (Stand 08/2001)

CRAMÉR, H. (1946): *Mathematical models of statistics*. Princeton Mathematical Series, No. 9

CZOMMER, R. (2000): *Leistungsfähigkeit fahrzeugautonomer Ortungsverfahren auf der Basis von Map-Matching-Techniken*. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 535, München

CZOMMER, R. (2008): Qualitätsbewertung *Map-Matching gestützter Positionsbestimmungsverfahren*. (interner) Bericht des Projektes Do-iT VM2010, Oktober 2008 (unveröffentlicht)

DÄUBLER, W.; KLEBE, T; WEDDE, P. und WEICHERT, T. (2007): *Bundesdatenschutzgesetz, Basiskommentar zum BDSG*. Bund-Verlag GmbH, Frankfurt am Main

DDG (2011): DDG Gesellschaft für Verkehrsdaten mbH: http://www.ddg.de ; http://www.tmcpro.de (Stand 02/2011)

DESTATIS (2011a): Statistisches Bundesamt Deutschland. http://www.destatis.de/jetspeed/portal/cms/ (Stand 08/2011)

DESTATIS (2011b): Statistisches Bundesamt Deutschland. https://www-genesis.destatis.de, Stand (08/2011)

DETLEFSEN, J. und SIART, U. (2006): *Grundlagen der Hochfrequenztechnik*. Oldenbourg Verlag, München

DIESTEL, R. (2006): *Graphentheorie*. Springer Verlag, Berlin

DIJKSTRA, E.W. (1959), *A Note on Two Problems in Connexion with Graphs*. Numerische Mathematik Band 1, S. 269-271

DUBOIS, D. und PRADE, H. (1980): *Fuzzy Sets and Systems : theory and applications*. Academic Press, New York

ENGELHARDT-FUNKE, O. (2002): *Stochastische Modellierung und Simulation von Verspätungen in Verkehrsnetzen für die Anwendung bei der Fahrplanoptimierung*. Dissertation, PAPIERFLIEGER, Clausthal-Zellerfeld

EU.DESTATIS (2011): Europäische Kommission eurostat. http://epp.eurostat.ec.europa.eu, (Stand 08/2011)

FAHRMEIR, L.; KÜNSTLER, R.; PIGEOT, I. und TUTZ, G. (2010): *Statistik, Der Weg zur Datenanalyse*. Springer, Heidelberg

FASTENRATH, U. (2008): *Erfahrungen und Perspektiven mit FCD*. In: Verkehrsmanagement und Verkehrstechnologien Mobile Zukunft mit intelligenten Verkehrssysteme, Tagungsband Verkehrsmanagement 2010,
20. und 21. Mai 2008

FILZMOSER, P. und VIERTL, R. (2004): *Testing hypotheses with fuzzy data: The fuzzy p-value*. Metrika Vol. 59, Number 1/ Febr. 2004, pp. 21-29

FILZMOSER, P. und VIERTL, R. (2006): *Beschreibung und Analyse unscharfer Information, Statistische Methoden für unscharfe Daten*. Springer, Wien

FRIEDRICH, M.; JEHLICKA, P. und SCHLAICH, J. (2008a): *Automatic number plate recognition for the observance of travel behavior*. 8th International Conference on Survey Methods in Transport: Harmonisation and Data Comparability, Mai 2008, Annecy, Frankreich

FRIEDRICH, M.; JEHLICKA, P.; OTTERSTAETTER, T. und SCHLAICH, J. (2008b): *Mobile Phone Data for Telematic Applications*. Proceedings of International Multi-Conference on Engineering and Technological Innovation, IMETI 2008, Orlando, Florida, USA

GABLER (Hrsg.) (2013): *Stichwort: öffentlicher Personenverkehr*, Gabler Wirtschaftslexikon. Gabler Verlag, http://wirtschaftslexikon.gabler.de/Archiv/78689/oeffentlicher-personenverkehr-v5.html (Stand 04/2013)

GENG, N. und WIESBECK, W. (1998): *Planungsmethoden für die Mobilkommunikation - Funknetzplanung unter realen physikalischen Ausbreitungsbedingungen*. Springer Verlag, Berlin

GIACHETTI, R. E. und YOUNG, R. E. (1997): *A parametric representation of fuzzy numbers and their arithmetic operators*. Fuzzy sets and systems, vol. 91 pp. 185-202

GLEMSER, M. (2000): *Zur Berücksichtigung der geometrischen Objektunsicherheit in der Geoinformatik*. Dissertation, Universität Stuttgart

GNU GPL (2011): GNU General Public License: http://www.gnu.org/ (Stand 08/2011)

GOODMAN, I. R. und NGUYEN, H. T. (1985): *Uncertainty Models for Knowledge-Based Systems*. Amsterdam, North Holland

GUSTAFSON, E. E. und KESSEL, W. C. (1979): *Fuzzy Clustering with a Covariance Matrix*. In: IEEE International Conference on Decision and Control, San Diego, pp. 761-766

HARTUNG, J. und ELPELT, B. (1999): *Multivariate Statistik, Lehr- und Handbuch der angewandten Statistik*, Oldenbourg Verlag, München

HARTUNG, J.; ELPELT, B. und KLÖSENER, K.-H. (1995): *Statistik, Lehr- und Handbuch der angewandten Statistik.* Oldenbourg Verlag, München

HATA, M. (1980): *Empirical Formula for Propagation Loss in Land Mobile Radio Services*. IEEE Transaction on Vehicular Technology, Vol. VT-29, No. 3, pp. 317-325

HATHAWAY, R. J. und BEZDEK, J. C. (1995): *Optimization of clustering criteria by reformulation*. IEEE Transactions on Fuzzy Systems, 3(2): S. 241-245

HELLEBRANDT, M. und MATHAR, R. (1999): *Location Tracking of Mobiles in Cellular Radio Networks*. IEEE Transactions on Vehicular Technology, Vol. 48, No. 5, September 1999, p. 1558-1562

HELLEBRANDT, M.; MATHAR, R. und SCHEIBENBOGEN, M. (1997): *Estimating Position and velocity of Mobiles in a Cellular Radio Network*. IEEE Transactions on Vehicular Technology, Vol. 46, No. 1, February 1997, p. 65-71

HELTON, J. C. und OBERKAMPF, W. L. (2004): *Alternative representations of epistemic uncertainty*. Reliability Engineering and System Safety 85, pp. 1-10

HENNINGER, T.; HEINE-NIMS, T.;SCHNITTGER, S. und BETTERMANN, R. (2003): *RUDY-Regionale unternehmensübergreifende und dynamische Vernetzung von Auskunfts-, Betriebsleit- und Planungssystemen im ÖPNV und Taxigewerbe*. 19. Verkehrswissenschaftliche Tage 2003, http://www.trafficforum.ethz.ch/vwt_2003/beitraege/VWT19proceedings_contribution_9.1-9.19.pdf (Stand 08/2011)

HERMANN, M. (2011): *Numerische Mathematik*. München, Wien, Oldenbourg Verlag

HERMANN, U. (1996): *Untersuchung zur Verspätungsentwicklung von Fernreisezügen auf der Datengrundlage der rechnerunterstützten Zugüberwachung Frankfurt am Main*. Dissertation, Darmstadt

HOFMAN-WELLENHOF, B.; LEGAT, K. und WIESER, M. (2003): *Navigation. Principles of positioning and guidance*. Springer-Verlag, Wien

HÖPKE, W. (1980): *Fehlerlehre und Ausgleichsrechnung*. de Gruyter, Berlin

HÖPPNER, F.; KLAWONN, F. und KRUSE, R. (1997): *Fuzzy-Clusteranalyse, Verfahren für die Bilderkennung, Klassifikation und Datenanalyse*. Vieweg Verlag

JÄGER, R.; MÜLLER, T.; SALER, H. und SCHWÄBLE, R. (2005): *Klassische und robuste Ausgleichungsverfahren*. Herbert Wichmann Verlag, Heidelberg

JAIN, A. und DUBES, R. (1988): *Algorithms for Clustering Data*. Prentice Hall

JAULIN, L. ; KIEFFER, M. ; WALTER, E. (2001): *Applied Interval Analysis*. Springer Verlag, London

JUNIUS, M. (1995): *Leistungsbewertung intelligenter Handover-Verfahren für zellulare Mobilfunksysteme*. Dissertation, RTWH Aachen, Lehrstuhl für Kommunikationsnetze

KAHLERT, J. (1995): *Fuzzy Control für Ingenieure, Analyse, Synthese und Optimierung von Fuzzy-Regelungssystemen*. Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden

KAMEL, M. S. und SELIM, S. Z. (1991): *A thresholded fuzzy c-means algorithm for semi-fuzzy clustering*. Pattern Recognition, 24(9): pp. 825-833

KAUFMANN, A. und GUPTA, M. (1991): *Introduction to Fuzzy Arithmetic - Theory and Application*. Van Nostrand Reinhold, New York

KBA (2011): Kraftfahrt-Bundesamt, http://www.kba.de (Stand 08/2011)

KENNEMANN, O. (1997): *Lokalisierung von Mobilstationen anhand ihrer Funkmeßdaten*. Dissertation, RWTH Aachen, Lehrstuhl für Kommunikationsnetze

KHALAF-ALLAH, M. (2008): *Bayesian Algorithms for Mobile Terminal Positioning in Outdoor Wireless Environments*. Dissertation, Universität Hannover

KLEIN, R. (2005): *Algorithmische Geometrie - Grundlagen, Methoden, Anwendungen*. Springer, Berlin

KOCH, K. R. (1997): *Parameterschätzung und Hypothesentests in linearen Modellen*. Ferd. Dümmlers Verlag, Bonn

KOHLEN, R. (2004): *Routenidentifikation in Verkehrsnetzen auf der Grundlage unscharfer Ortungsinformationen (GSM)*. Dissertation, Schriftenreihe Heft 52, Shaker Verlag, Aachen

KRAMPE, S. und KIESLICH, W. (2005): *Nutzung von Floating-Traveller-Data (FTD) für mobile Lotsendienste im Verkehr*. Tagungsband der HEUREKA 2005, FGSV-Verlag, Köln, S. 365-383

KRAMPE, S. (2007): *Nutzung von „Floating Traveller Data (FTD)" für mobile Lotsendienste im Verkehr*. Dissertation an der TU Darmstadt

KUTTERER, H. (1994): *Intervallmathematische Behandlung endlicher Unschärfen linearer Ausgleichungsmodelle*. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 423, München

KUTTERER, H. (2002): *Zum Umgang mit Ungewissheit in der Geodäsie - Bausteine für eine neue Fehlertheorie* -. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 553, München

KUTTERER, H. (2004): *Statistical Hypothesis tests in case of imprecise data*. In: F. Sansò (Ed.): Proceedings of the 5. Hotine-Marussi-Symposium, International Association of Geodesy Symposia, Springer, Berlin

KWAKERNAAK, H. (1978): *Fuzzy random variables I: definitions and theorems*. Information Sciences vol. 15, pp. 1 - 29

KWAKERNAAK, H. (1979): *Fuzzy random variables II: algorithms and examples for the descrete case*. Information Sciences vol. 17, pp. 253 - 278

LACKRITZ, J. R. (1983): *Exact p values for chi-square tests, Procedings of the Section an Statistical Education*. American Statistical Association, p. 130-132

LACKRITZ, J. R. (1984): *Exact p values for F and t tests*. The American Statistician 38, p. 312-314

LAUFER, R. (2011): *Prozedurale Qualitätsmodellierung und -management für Daten* ingenieurgeodätische *und verkehrstechnische Anwendungen*. Dissertation, Universität Stuttgart

LEE, J. S. und MILLER, L. E. (1998): *CDMA Systems: Engineering Handbook*. Boston, London, Artech House

LEE, W. C. Y. (1997): *Mobile Communications Engineering: theory and applications*. 2nd edition, McGraw-Hill, New York

LEHNHOFF, N. (2005): *Überprüfung und Verbesserung der Qualität von automatisch erhobenen Daten an Lichtsignalanlagen*. Dissertation an der Universität Hannover

LESOT, M.-J. und KRUSE, R. (2006): *Gustafson-Kessel-like clustering algorithm based on typicality degrees*. In IPMU 02.07.-07.07.2006, Paris

LÜDERS, C. (2001): *Mobilfunksysteme*. Vogel Verlag, Würzburg

LUO, D. (2007*): Erfassung und Analyse von Fahrplanabweichungen ausgewählter ÖPNV-Angebote*. Studienarbeit am IAGB Universität Stuttgart (unveröffentlicht)

MACQUEEN, J. (1967): *Some methods for classification and analysis of multivariate observations*. in: Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 281-297

MAYER, G. (1989): *Grundbegriffe der Intervallrechnung*. in Kulisch, U. (Hrsg.) 1989: *Wissenschaftliches Rechnen mit Ergebnisverifikation - Eine Einführung* -. Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden, S. 101-117

MID (2002): Mobilität in Deutschland, Bundesministerium für Verkehr, Bau und Stadtentwicklung, Ergebnisbericht MiD 2002. http://www.mobilitaet-in-deutschland.de /03_kontiv2002/pdf/ergebnisbericht_mid_ende_144_punkte.pdf (Stand 08/2011)

MÖLLER, B. und BEER, M. (2004): *Engineering computation under uncertainty – Capabilities of non-traditional models.* Fuzzy Randomness -Uncertainty in Civil Engineering and Computational Mechanics-. Springer, Berlin

MOORE, R. E. (1966): *Interval Analysis*. Prentice Hall Inc., Englewood Cliffs, N. J.

NETZABDECKUNG T-MOBILE (2011): Telekom Deutschland GmbH, http://www.t-mobile.de/ funkversorgung/inland (Stand 02/2011)

NAVTEQ (2011): NAVTEQ®. http://www.navteq.com (Stand 08/2011)

NEUMANN, I. (2009): *Zur Modellierung eines erweiterten Unsicherheitshaushaltes in Parameterschätzung und Hypothesentests*. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 634, München

NEUMANN, I. und KUTTERER, H. (2005): *Geodetic Deformation Analysis with respect to observation imprecision and object fuzziness*. in International Symposium on Modern Technologies, Education And Professional Practice In Geodesy And Related Fields, Sofia, 03-04 November 2005, S. 147-156

NEUMANN, I. und KUTTERER, H. (2006a): *Geodetic Deformation Analysis with Respect to Observation Imprecision*. in Shaping the Change, XXIII FIG Congress Munich,
8-13 October 2006, München

NEUMANN, I. und KUTTERER, H. (2006b): *Congruence Tests and Outlier Detection in Deformation Analysis with Respect to Observation Imprecision*. in: Proceedings of the 3rd IAG Symposium on Geodesy for Geotechnical and Structural Engineering and the 12th FIG Symposium on Deformation Measurement, Baden, Österreich

NEUMANN, I. und KUTTERER, H. (2007): *Congruence tests and outlier detection in deformation analysis with respect to observation imprecision*. in International Journal of Applied Geodesy, Vol. 1 no. 1, de Gruyter, London

NEUMANN, I. und KUTTERER, H. (2008): *The probability of type I and type II errors in imprecise hypothesis testing*. in: Muhanna, R., L.; Mullen, R., L. (Eds.): Proceedings of the NSF Workshop on Imprecise Probability in Engineering Analysis and Design-. Savannah (Georgia), USA, pp. 137 - 154

NEUMANN, I.; KUTTERER, H. und SCHÖN, S. (2006): *Outlier Detection in Geodetic Applications with respect to Observation Imprecision*. in Muhanna R.L. and Mullen R.L. (Eds.) Proc. "REC 2006 - The NSF Workshop on Reliable Engineering Computing", pp. 75-90, Savannah, Georgia, USA

NIEMEIER, W. (2002): *Ausgleichungsrechnung, Eine Einführung für Studierende und Praktiker des Vermessungs- und Geoinformationswesens*. de Gruyter, Berlin

OETTING, A. (2002): *Angebotsabhängige Modellierung der reisezweckspezifischen Verkehrserzeugung und ihre Anwendung auf den städtischen Freizeitverkehr*. Dissertation, RWTH Aachen

OKUMURA, Y.; OHMORI, E.; KAWANO, T. UND FUKUDA, K. (1968): *Field Strength and It's Variability in VHF and UHF Land-Mobile Service*. Review of the Electrical Communication Laboratory, Vol. 16, pp. 825-873

OPENSTREETMAP (2011): http://www.openstreetmap.de (Stand 08/2011)

OUSSALAH, M. und DE SCHUTTER, J. (2003): *Approximated fuzzy LR computation*. Information Sciences, vol. 153, pp. 155-175

PAPULA, L. (2009): *Mathematik für Ingenieure und Naturwissenschaftler*. Band 1, Vieweg+Teubner, GWV Fachverlage GmbH, Wiesbaden

PARSONS, J. D. (2000): *The Mobile Radio Propagation Channel*. 2nd ed., John Wiley & Sons Ltd, Chichester, New York

PELZER, H. (HRSG.) (1985): *Geodätische Netze in der Landes- und Ingenieurvermessung*. Vorträge des Kontaktstudiums II, Konrad Wittwer Verlag, Stuttgart

RAMM, K. und SCHWIEGER, V. (2007): *Mobile positioning for traffic state acquisition*. Mobile Mapping Technologies Symposium, The 5th International Symposium on Mobile Mapping Technology MMT'2007, Padua - Italy

RAMM, F. und TOPF, J. (2009): *OpenStreetMap : die freie Weltkarte nutzen und mitgestalten*. Lehmanns Media, Berlin

RAMM, K.; CZOMMER, R. und SCHWIEGER, V. (2006): *Map-based Positioning using Mobile Phones*. Proceedings on XXIII International FIG Congress, München, 8.-13.10.2006

RAUBER, T. (1993): *Algorithmen in der Computergraphik*. Teubner, Stuttgart

RINNE, H. (2008): *Taschenbuch der Statistik*. Verlag Harry Deutsch, Frankfurt a. M.

ROTH, M. (2002): *Mobile Computing, Grundlagen, Technik, Konzepte*. dpunkt Verlag, Heidelberg

RUNKLER, T. A. (1999): *Probabilistische und Fuzzy Methoden für die Clusteranalyse, in Fuzzy Theorie und Stochastik*. Seising, R. (Hrsg.), Vieweg

RUNKLER, T. A. und BEZDEK, J. C. (1997): *Living clusters: An application of El Farol algorithm to the fuzzy c-means model*. In Proc. European Congress on Intelligent Techniques and Soft Computing, pp. 1678-1682, Aachen

RUSPINI, E. H. (1969): *A New Approach to Clustering.* Information and Control 15, 22-32

SACHS, L. (2004): *Angewandte Statistik, Anwendung statistischer Methoden*. 11. Auflage, Springer-Verlag, Berlin

SCHLAICH, J; OTTERSTÄTTER, T. und FRIEDRICH, M. (2010): *Generating Trajectories from Mobile Phone Data*. TRB 89th Annual Meeting Compendium of Papers, Transportation Research Board of the National Academies, Washington, D.C., USA

SCHMEINK, M. (2005): *Optimierungsalgorithmen zur automatisierten Funknetzplanung*. Dissertation an der RTWH Aachen, Shaker Verlag, Aachen

SCHOLLMEYER, R. (2008): *Bericht zur Identifizierung aktiver Verkehrsteilnehmern (AP3)*. (interner) Bericht des Projektes Do-iT VM2010 (unveröffentlicht)

SCHOLLMEYER, R. und SCHWIEGER, V. (2009): *Qualitätsbewertung zur Identifikation aktiver Verkehrsteilnehmer*. (interner) Bericht des Projektes Do-iT VM2010 (unveröffentlicht)

SCHOLLMEYER, R. und WILTSCHKO, T. (2007): *Classification of public transport vehicles using cellular mobile radio data*. 6th European Congress & Exhibition on Intelligent Transport Systems and Services, ITS Aalborg Congress 18-20 june 2007, Denmark

SCHÖN, S. (2003): *Analyse und Optimierung geodätischer Messanordnungen unter besonderer Berücksichtigung des Intervallansatzes*. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 567, München

SCHWANHÄUSSER, W. (1974): *Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn*. Veröffentlichungen des Verkehrswissenschaftlichen Instituts der RWTH Aachen, Heft 20

SCHWARZ, R. und KÖCKLER, N. (2011): *Numerische Mathematik*. Vieweg + Teubner Verlag, Wiesbaden

SCHWIEGER, V. (2007): *Positioning within the GSM Network*. 6th FIG Regional Conference San José, Costa Rica 12-15 November 2007

SCHWIEGER, V. (2008): *Generierung von Verkehrsdaten auf Basis von verfügbaren Mobilfunkdaten innerhalb der GSM-Netzinfrastruktur*. In: Verkehrsmanagement und Verkehrstechnologien Mobile Zukunft mit intelligenten Verkehrssysteme, Tagungsband Verkehrsmanagement 2010, 20. und 21. Mai 2008

SERVICEREPORT (2013): servicereport, http://servicereport.eu /datendienste-zunehmend-mobil/ (Stand Mai 2013)

SIWIAK, K. (1998): *Radiowave Propagation and Antennas for Personal Communications*. Boston London, Artech House

SOMMER, C. (2002): *Erfassung des Verkehrsverhaltens mittels Mobilfunktechnik; Konzeption und Akzeptanz eines neuen Erhebungsverfahrens*. Dissertation, Schriftenreihe des Instituts für Verkehr und Stadtbauwesen der TU Braunschweig, Heft 51, Shaker Verlag, Aachen

SPIES, M. (1993): *Unsicheres Wissen*. Spektrum Akademischer Verlag, Heidelberg

STEINMETZ, V. (1998): *Schließende Statistik mit unscharfen Stichproben*. Latex-Version der Nr. A9802, Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft, Universität des Saarlandes, http://www.uni-saarland.de/fak1/fr12/steinmetz/fustat.pdf (Stand 08/2011)

STGTZ (2008): http://content.stuttgarter-zeitung.de/stz/page/1891882_0_9223_-verkehrsleitzentrale-50-neue-messstellen-geplant.html (Stand 05/2013)

StVZO (2009): Straßenverkehrs-Zulassungs-Ordnung, http://www.gesetze-im-internet.de/bundesrecht/stvzo/gesamt.pdf (Stand 08/2011)

TELEKOM (2012): Telekom Deutschland GmbH, http://www.telekom.de/ (Stand 04/2013)

THE MATHWORKS (2013): http://www.mathworks.de/de/help/stats/multivariate-normal-distribution.html (Stand 06/2013)

TIZHOOSH, H. R. (1998): *Fuzzy Bildverarbeitung, Einführung in Theorie und Praxis*. Springer, Berlin

TOMTOM (2011): TOMTOM®, http://www.tomtom.com/ (Stand 08/2011)

TRUCANO, T. G.; SWILERA, L. P.; IGUSA, T.; OBERKAMPF, W. L. und PILCH, M. (2006): *Calibration, validation, and sensitivity analysis: What's what*. Reliability Engineering and System Safety 91, pp. 1331-1357

VAKHTEL, S. (2002): *Rechnerunterstützte Ermittlung der Kapazität von Eisenbahnnetzen*. Veröffentlichungen des Verkehrswissenschaftlichen Instituts der RWTH Aachen, Heft 59

VIERTL, R. (1996): *Statistical Methods for Non-Precise DATA*. CRC Press, Inc., Boca Raton, New York

VIERTL, R. (1997): *Einführung in die Stochastik : mit Elementen der Bayes-Statistik und Ansätzen für die Analyse unscharfer Daten*. Springer Lehrbuch Mathematik,
2. überarb. Aufl., Wien

VIERTL, R. (2003): *Einführung in die Stochastik : Mit Elementen der Bayes-Statistik und der Analyse unscharfer Informationen*. Springers Lehrbücher der Informatik,
3. Aufl., Wien

VIERTL, R. und HARETER, D. (2006): *Beschreibung und Analyse unscharfer Information - Statistische Methoden für unscharfe Daten*. Springer, Wien

VM2010 PROJEKT DO-IT (2011): *Projektergebnisse, Tagungsband*: http://www.vm2010.de/web/de/projekte/do-it.html (Stand 08/2011)

WALKE, B. (2001): *Mobilfunknetze und ihre Protokolle, 1: Grundlagen, GSM, UMTS und andere zellulare Mobilfunknetze*. Teubner Verlag, Stuttgart

WEIGAND, W. (1981): *Verspätungsübertragungen in Fernverkehrsnetzen*. ETR (30)

WELSCH, W.; HEUNECKE, O. und KUHLMANN, H. (2000): *Handbuch Ingenieurgeodäsie, Auswertung geodätischer Überwachungsmessungen*. Wichmann Verlag, Heidelberg

WERMUTH, M. ET AL. (2003): *TTS TeleTravel System - Telematiksystem zur automatischen Erfassung des Verkehrsverhaltens*. Schlussbericht zum BMBF Projekt TTS TeleTravel Services: http://www.ivs.tu-bs.de/cms/dokumente/forschungsberichte/schlussbericht-tts.pdf (Stand 02/2011)

WIESER, A. (2012): *Uncertainty of Motion Detection from Coordinate Differences*. Proceedings 3rd Conference on Machine Control & Guidance - MCG 2012, March 27-29, 2012, Stuttgart

WILTSCHKO, T. (2004): *Sichere Information durch infrastrukturgestützte Fahrerassistenzsysteme zur Steigerung der Verkehrssicherheit an Straßen-knotenpunkten*.
Dissertation, Fortschritt-Bericht VDI, Reihe 12, Nr. 570, VDI-Verlag, Düsseldorf

WILTSCHKO, T. und SCHOLLMEYER, R. (2006): *Potenzial von Geodaten bei Planung und Betrieb von öffentlichem Personennahverkehr*. http://www.newstix.de/downloads/fachbericht1707.pdf (Stand 08/2011)

WILTSCHKO, T. und SCHWIEGER, V. (2008): *Floating Phone Data – Ein Ansatz zur Generierung von Verkehrsdaten aus Mobilfunkdaten*. Beiträge zur HEUREKA 2008, FGSV-Verlag, Köln

WILTSCHKO, T.; SCHWIEGER, V. und MÖHLENBRINK, W. (2006): *Generating Floating Phone Data for Traffic Flow Optimization*. Proceedings on 3rd International Symposium Networks for Mobility 5.-6.10.2006, Stuttgart

WILTSCHKO, T.; SCHWIEGER, V. und MÖHLENBRINK, W. (2007): *Acquisition of traffic state information by mobile phone positioning*. 6th European Congress & Exhibition on Intelligent Transport Systems and Services. ITS Aalborg Congress 18-20 june 2007, Denmark

WITTE, B. und SCHMIDT, H. (2000): *Vermessungskunde und Grundlagen der Statistik im Bauwesen*. Verlag Konrad Wittwer, Stuttgart

ZADEH, L. A. (1965): *Fuzzy sets*. In: Information Control Vol. 8, pp. 338-353

ZHAO, Y. (1997): *Vehicle Location and Navigation Systems*. Intelligent Transportation Systems Artech House, Inc. Boston, London

ZIMMERMANN, H.-J. (2001): *Fuzzy set theory and its applications*. Kluwer

Abkürzungsverzeichnis

BDSG	Bundesdatenschutzgesetz
BSC	Base Station Controller
BTS	Base Tranceiver Station
card	Mächtigkeit einer Fuzzy-Menge (cardinality)
CCD	Charge-coupled Device
core	Kern einer Fuzzy-Menge
COX	Communication & Orientation eXpert, cox „Steuermann“
DO-IT	Datenoptimierung für integrierte Telematik
FCD	Floating Car Data
FCM	Fuzzy c-Means
FPD	Floating Phone Data
FRV	Fuzzy-Random-Variable
GK-FCM	Gustafson-Kessel Fuzzy c-Means
GPS	Global Positioning System
GSM	Global System for Mobile Communication
height	Höhe einer Fuzzy-Menge
KF	Korridorfläche
KoEll	Korridorellipse
KRAD	Kraftrad
L	Linke Referenzfunktion eines Fuzzy-Intervalls
LKW	Lastkraftwagen
MIV	Motorisierter Individualverkehr
MS	mobile station (Mobilfunkendgerät, Handy)
MSC	Mobile Switching Center
NMIV	Nicht-Motorisierter Individualverkehr
OF	Ortungsfläche
ÖV	Öffentlicher Verkehr
oVB	ohne Verkehrsbeteiligung
PCM	possibilistischen c-Means-Modell
PKW	Personenkraftwagen
R	Rechte Referenzfunktion eines Fuzzy-Intervalls
RX	Empfangener Signalpegel
sec	Sekunde
sup	Kleinste oberste Grenze (supremum)
supp	Träger einer Fuzzy-Menge (Support)
TA	timing advance
TMC	Traffic Message Channel
TTS	TeleTravel System
UMTS	Universial Mobile Telecommunications System
VM	Verkehrsmittel
vgl.	vergleiche
z. B.	zum Beispiel

Anhang

A Datentypen der Mobilfunkdaten für das Signalpegel-Matching

Im GSM-Mobilfunknetz wird während einer bestehenden Verbindung (z. B. eines Telefongesprächs) in der Regel im 480 ms Takt ein Datentelegramm (Messwertreport) vom MS an die Infrastruktur des Mobilfunknetzes gesendet. Die in diesem Datentelegramm enthaltenen Signalpegel sind die wesentlichen Messwertgrundlagen für die Mobilfunkortung des in der vorliegenden Arbeit eingesetzten Signalpegel-Matchings. In Tabelle A-1 sind dazu erforderlichen Datentypen zusammengefasst.

Tabelle A-1: Für das Signalpegel-Matching verwendete Datentypen

Abkürzung	**Datentyp**	**Beschreibung**
subscriber	integer	Anonymisierte Kennung des Mobilfunkteilnehmers
call_id	integer	Anonymisierte Kennung der Gesprächsverbindung
date	integer	Datum [JJJJMMDD]
time	double	Uhrzeit (Sekunden ab Mitternacht)
0CI	integer	Kennung der bedienenden Mobilfunkzelle (Cell Identity)
0RX	integer	(um Leistungsregelung korrigierter) Signalpegel der bedienenden Zelle
TA	integer	Wert des Timing Advance zur bedienenden Zelle
Ncell	integer	Anzahl der Nachbarzellen
1CI	integer	Kennung der Mobilfunkzelle der ersten Nachbarzelle
1RX	integer	Signalpegel der ersten Nachbarzelle
2CI	integer	Kennung der Mobilfunkzelle der zweiten Nachbarzelle
2RX	integer	Signalpegel der zweiten Nachbarzelle
…	…	…
6CI	integer	Kennung der Mobilfunkzelle der sechsten Nachbarzelle
6RX	integer	Signalpegel der sechsten Nachbarzelle

Hinweis: Die Werte für 0CI,1CI,…,6CI sind im GSM-Mobilfunknetz eindeutige Kennungen.

B Kennzahlen der Mobilfunkortung anonymisierter Mobilfunkdaten

In den folgenden Tabellen B-1 und B-2 sind die Kennzahlen der Mobilfunkortung und der geclusterten Mobilfunkortungen der am 09.07.2008 im Versuchsgebiet Karlsruhe des Mobilfunknetzes der T-Mobile Deutschland erfassten und anonymisierten Mobilfunkdaten zusammengefasst. Die originären Mobilfunkdaten umfassen 35 791 435 Mobilfunkortungen von 162 572 MS.

Tabelle B-1: Ausgewählte Kennzahlen für die Unsicherheitswerte der originären Mobilfunkortungen

	Zufällige Variabilität*)	Fuzzy-Theorie*)		Fuzzy-Randomness*)	
Minimum	$s_x = 27{,}6$ $s_y = 17{,}9$ $s_{xy} = 0{,}0$	$r_x = 0{,}0$ $c_{l,x} = 0{,}0$ $c_{r,x} = 0{,}0$	$r_y = 0{,}0$ $c_{l,y} = 0{,}0$ $c_{r,y} = 0{,}0$	$c_{l,x} = 0{,}0$ $c_{r,x} = 0{,}0$ $c_{l,y} = 0{,}0$ $c_{r,y} = 0{,}0$	$s_x = 28{,}4$ $s_y = 23{,}1$ $s_{xy} = 0{,}0$
Mittelwert	$s_x = 443{,}2$ $s_y = 416{,}3$ $s_{xy} = 83814{,}1$	$r_x = 242{,}87$ $c_{l,x} = 542{,}35$ $c_{r,x} = 370{,}84$	$r_y = 239{,}06$ $c_{l,y} = 462{,}00$ $c_{r,y} = 412{,}65$	$c_{l,x} = 96{,}78$ $c_{r,x} = 95{,}76$ $c_{l,y} = 89{,}44$ $c_{r,y} = 90{,}82$	$s_x = 450{,}68$ $s_y = 421{,}22$ $s_{xy} = 97722{,}2$
Maximum	$s_x = 2765{,}9$ $s_y = 2777{,}7$ $s_{xy} = 4214204{,}0$	$r_x = 3559{,}8$ $c_{l,x} = 6911{,}4$ $c_{r,x} = 6704{,}8$	$r_y = 3488{,}5$ $c_{l,y} = 6476{,}9$ $c_{r,y} = 6801{,}1$	$c_{l,x} = 2781{,}4$ $c_{r,x} = 2972{,}2$ $c_{l,y} = 3251{,}7$ $c_{r,y} = 3054{,}7$	$s_x = 2765{,}8$ $s_y = 2730{,}5$ $s_{xy} = 4473000{,}8$

*) für s_{xy} sind die Werte in der Einheit Meter^2, die restlichen Werte sind in der Einheit Meter angegeben

Tabelle B-2: Ausgewählte Kennzahlen für die Unsicherheitswerte der geclusterten Mobilfunkortungen

	Zufällige Variabilität*)	Fuzzy-Theorie*)		Fuzzy-Randomness*)	
Anzahl	643 619	717 333		893 885	
Minimum	$s_x = 0{,}0$ $s_y = 0{,}0$ $s_{xy} = 0{,}0$	$r_x = 0{,}0$ $c_{l,x} = 0{,}0$ $c_{r,x} = 0{,}0$	$r_y = 0{,}0$ $c_{l,y} = 0{,}0$ $c_{r,y} = 0{,}0$	$c_{l,x} = 0{,}0$ $c_{r,x} = 0{,}0$ $c_{l,y} = 0{,}0$ $c_{r,y} = 0{,}0$	$s_x = 0{,}0$ $s_y = 0{,}0$ $s_{xy} = 0{,}0$
Mittelwert	$s_x = 504{,}6$ $s_y = 479{,}0$ $s_{xy} = 96704{,}5$	$r_x = 271{,}39$ $c_{l,x} = 612{,}67$ $c_{r,x} = 430{,}63$	$r_y = 269{,}75$ $c_{l,y} = 518{,}22$ $c_{r,y} = 487{,}57$	$c_{l,x} = 147{,}0$ $c_{r,x} = 145{,}9$ $c_{l,y} = 138{,}0$ $c_{r,y} = 139{,}1$	$s_x = 424{,}0$ $s_y = 400{,}0$ $s_{xy} = 90361{,}6$
Maximum	$s_x = 3578{,}5$ $s_y = 3487{,}1$ $s_{xy} = 4630163{,}3$	$r_x = 3558{,}9$ $c_{l,x} = 6203{,}9$ $c_{r,x} = 6195{,}2$	$r_y = 3411{,}2$ $c_{l,y} = 5696{,}6$ $c_{r,y} = 5956{,}2$	$c_{l,x} = 2197{,}6$ $c_{r,x} = 2044{,}7$ $c_{l,y} = 3024{,}6$ $c_{r,y} = 2650{,}6$	$s_x = 3653{,}0$ $s_y = 3613{,}5$ $s_{xy} = 6766851{,}6$

*) für s_{xy} sind die Werte in der Einheit Meter^2, die restlichen Werte sind in der Einheit Meter angegeben

C Integral oder numerische Integration für die card-Funktion

Die folgende Tabelle C-1 fasst für ausgewählte Funktionen die Formeln zur Bestimmung des card-Funktionswertes (Integral, Flächenberechnung) im Falle normaler Fuzzy-Mengen zusammen. Es bedeuten: c_l linke Spannweite, c_r rechte Spannweite, c für gleiche Spannweiten und r Radius (Kern).

Tabelle C-1: Formeln für die Berechnung des Integrals ausgewählten Funktionen für Fuzzy-Sets

Funktionstyp	**Funktion** $m_{\tilde{A}}(x) =$	**Fläche (Integral)** $\mathrm{card}(\tilde{A}) =$
Dreiecktyp mit: $c > 0$	$= \begin{cases} 1 - \dfrac{x_m - x}{c} & , \quad x_m - c \le x \le x_m \\ 1 - \dfrac{x - x_m}{c} & , \quad x_m < x \le x_m + c \end{cases}$	$= c$
Dreiecktyp mit: $c_l \ge 0, c_r > 0$	$= \begin{cases} 1 - \dfrac{x_m - x}{c_l} & , \quad x_m - c_l \le x \le x_m \\ 1 - \dfrac{x - x_m}{c_r} & , \quad x_m < x \le x_m + c_r \end{cases}$	$= \dfrac{1}{2} \cdot (c_l + c_r)$
Trapeztyp mit: $c > 0$	$= \begin{cases} 1 - \dfrac{x_m - x - r}{c} & , \quad x_m - r - c \le x < x_m - r \\ 1 & , \quad x_m - r \le x \le x_m + r \\ 1 - \dfrac{x - x_m - r}{c} & , \quad x_m + r < x \le x_m + r + c \end{cases}$	$= 2 \cdot r + c$
Trapeztyp mit: $c_l > 0, c_r > 0$	$= \begin{cases} 1 - \dfrac{x_m - x - r}{c_l} & , \quad x_m - r - c_l \le x < x_m - r \\ 1 & , \quad x_m - r \le x \le x_m + r \\ 1 - \dfrac{x - x_m - r}{c_r} & , \quad x_m + r < x \le x_m + r + c_r \end{cases}$	$= 2 \cdot r + \dfrac{1}{2} \cdot (c_l + c_r)$
Konstruktion nach dem α- Schnittverfahren	$= \sup_{\alpha \in (0,1]} \alpha \cdot i_{\tilde{A}_\alpha}(x)$	$= \sum_{i=2}^{n_\alpha} \Delta\alpha \cdot (\Delta x_{i-1} + \Delta x_i)$ mit: $\Delta x_i = \tilde{A}_{\alpha=\Delta\alpha \cdot i,\max} - \tilde{A}_{\alpha=\Delta\alpha \cdot i,\min}$ $i = 1, 2 \ldots n_\alpha$ $\Delta\alpha = 1/(n_\alpha - 1)$

D Erhebung und Analyse von Fahrplanabweichungen

D.1 Linienbusse

Für die Bestimmung der Unsicherheitsparameterwerte zur Beschreibung der Fahrplanabweichungen von Linienbussen wurde auf Datenerhebungen im Raum Stuttgart von LUO (2007) zurückgegriffen. Dort wurden anhand der Fahrplanauskünfte und den tatsächlichen Antreffzeiten der Linienbusse an den Haltestellen, Fahrplanabweichungen empirisch erhoben. In Tabelle D-1 sind die Kennwerte dieser Daten zusammengefasst. Negative Fahrplanabweichungswerte kennzeichnen eine Verspätung, positive eine Verfrühung. Die Abbildung D-1 zeigt die Fahrplanabweichungen in einem Histogramm.

Die Kennwerte der Normalverteilung im Falle der zufälligen Variabilität sind für den Erwartungswert $\mu_{\Delta t_{PlanAbw}} = -91{,}6$ Sekunden und für die Standardabweichung $\sigma_{\Delta t_{PlanAbw}} = 76{,}6$ Sekunden.

Die Parameterwerte des Fuzzy-Intervalls im Falle der Fuzzy-Theorie wurden aus dem Histogramm mit den α-Schnitten der in das Intervall $[0,1]$ normierten Häufigkeiten bei $\alpha = 0{,}10$ für den Träger und $\alpha = 0{,}80$ für den Kern mit $m_{\widetilde{\Delta t}_{PlanAbw}} = -70$, $r_{\widetilde{\Delta t}_{PlanAbw}} = 40$, $c_{l,\widetilde{\Delta t}_{PlanAbw}} = 120$ und $c_{r,\widetilde{\Delta t}_{PlanAbw}} = 110$ jeweils in Sekunden berechnet.

Tabelle D-1: Kennwerte der empirisch erhobenen Fahrplanabweichungen für Linienbusse (LUO 2007)

	Wert
Anzahl:	1848
Minimum:	-302,0 Sekunden
Maximum:	215,0 Sekunden
arithmetischer Mittelwert:	-91,6 Sekunden
Standardabweichung:	76,6 Sekunden
Median:	-88,0 Sekunden

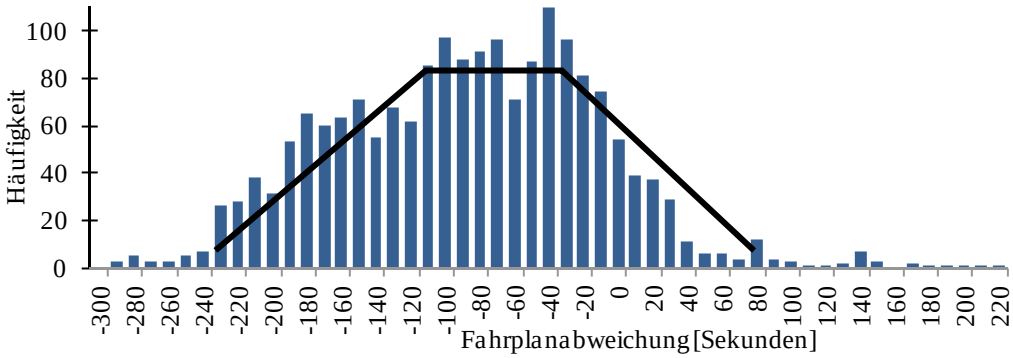


Abbildung D-1: Histogramm der Fahrplanabweichungen aus Versuchsfahrten in Linienbussen nach LUO (2007)

D.2 Straßenbahnen

Die Fahrplanabweichungen der Straßenbahnen wurden aus den Protokollen der Versuchsfahrten in den Straßenbahnen abgeleitet. Dazu wurden die Differenz zwischen den fahrplanmäßigen Abfahrtzeiten und den tatsächlichen Abfahrtzeiten berechnet. Die Kenndaten dieser Datenmenge sind in Tabelle D-2 zusammengefasst. Negative Fahrplanabweichungswerte kennzeichnen eine Verspätung, positive eine Verfrühung. Deren Histogramm zeigt Abbildung D-2. Im Vergleich zum Datenumfang für die Parameterwertebestimmung der Linienbusse und der Züge auf der Gleisen DB AG, ist die Grundgesamtheit der Fahrplanabweichungen für die Straßenbahnen als gering zu bewerten.

Die Kennwerte der Normalverteilung im Falle der zufälligen Variabilität sind für den Erwartungswert $\mu_{\underline{\Delta t}_{PlanAbw}} = -27,6$ Sekunden und für die Standardabweichung $\sigma_{\underline{\Delta t}_{PlanAbw}} = 61,8$ Sekunden.

Die Parameterwerte des Fuzzy-Intervalls im Falle der Fuzzy-Theorie wurden aus dem Histogramm mit den α-Schnitten der in das Intervall $[0,1]$ normiert Häufigkeiten bei $\alpha = 0,10$ für den Träger und $\alpha = 0,80$ für den Kern mit $m_{\widetilde{\Delta t}_{PlanAbw}} = 15$, $r_{\widetilde{\Delta t}_{PlanAbw}} = 15$, $c_{l,\widetilde{\Delta t}_{PlanAbw}} = 150$ und $c_{r,\widetilde{\Delta t}_{PlanAbw}} = 60$ jeweils in Sekunden berechnet.

Tabelle D-2: Kennwerte der empirisch erhobenen Fahrplanabweichungen für Straßenbahnen

	Wert
Anzahl:	32
Minimum:	-149,0 Sekunden
Maximum:	78,0 Sekunden
arithmetischer Mittelwert:	-27,6 Sekunden
Standardabweichung:	61,8 Sekunden
Median:	-11,0 Sekunden

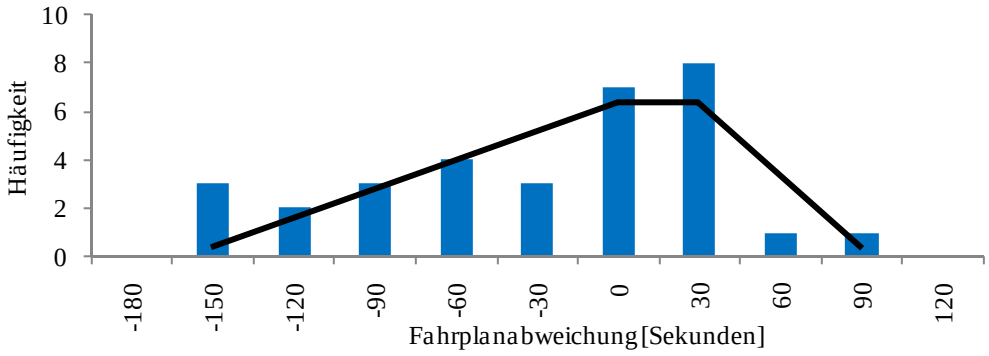


Abbildung D-2: Histogramm der Fahrplanabweichungen aus den Versuchsfahrten in Straßenbahnen

D.3 Züge der DB AG

Zur Bestimmung der Kennwerte für die Fahrplanabweichungen der Züge auf den Gleisen der DB-Netz AG wurden in HERMANN (1996, S. 160f) aufbereitete Daten übernommen. Dort wurden unter anderem umfangreiche Analysen zu Fahrplanabweichungen (dort wird der Begriff Einbruchsverspätung verwendet) für die Zuggattungen IR, D Züge und IC, EC und ICE der DB durchgeführt. Eine derartig detaillierte Unterteilung der Zuggattungen soll in der vorliegenden Arbeit nicht stattfinden. Stattdessen werden näherungsweise die nach Zuggattungen aufgeschlüsselten Mittelwerte und Varianzen gemittelt, wobei sie jeweils mit der Messanzahl (Anzahl) gewichtet wurden. Die von dort übernommenen Daten sind in Tabelle D-3 zusammengefasst und umfassen insgesamt 199015 Messungen. In der Quelle werden Verspätungen mit positiven Werten gekennzeichnet. Das Vorzeichen wurde für die Übernahme in die vorliegende Arbeit geändert. Negative Fahrplanabweichungswerte kennzeichnen eine Verspätung, positive eine Verfrühung.

Die Kennwerte der Normalverteilung im Falle der zufälligen Variabilität sind $\mu_{\underline{\Delta t}_{PlanAbw}} = -192,8$ Sekunden und $\sigma_{\underline{\Delta t}_{PlanAbw}} = 341,33$ Sekunden, vgl. Abbildung D-3.

Auf Basis dieser Normalverteilung wurden zur Beschreibung der Fahrplanabweichungen mittels Unschärfe die Parameterwerte des Fuzzy-Intervalls aus den Maximalwerten und Minimalwerten der α-Schnitte bei $\alpha_{min} = 0,1$ und $\alpha_{max} = 0,8$ berechnet. Die Wahrscheinlichkeitsdichtefunktion $f(x)$ der Normalverteilung wurde dafür in das Intervall $[0,1]$ normiert. Die Parameterwerte des Fuzzy-Intervalls lauten in Sekunden: $m_{\widetilde{\Delta t}_{PlanAbw}} = -192,8$, $r_{\widetilde{\Delta t}_{PlanAbw}} = 228,0$, $c_{l,\widetilde{\Delta t}_{PlanAbw}} = 504,5$ und $c_{r,\widetilde{\Delta t}_{PlanAbw}} = 504,5$.

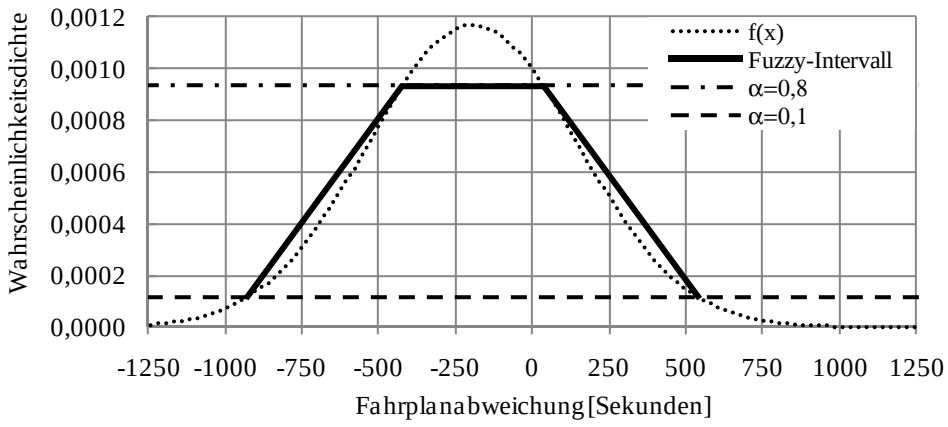


Abbildung D-3: Wahrscheinlichkeitsdichtefunktion f(x) der Normalverteilung und Fuzzy-Intervall zur Beschreibung der Fahrplanabweichungen der Züge auf den Gleisen der DB-Netz AG

Tabelle D-3: Fahrplanabweichungen für Züge auf dem Schienennetz der DB AG (HERMANN 1996)

Streckenbezeichnung	Gattung	Anzahl	Mittelwert	Varianz
47103N	IC/EC	14092	-2,896	24,477
47103N	IR/D	4577	-3,949	66,23
47103S	IC/EC	14067	-4,046	39,372
47103S	IR/D	4603	-4,12	75,044
61001N	ICE	13031	-2,91	26,177
61001S	ICE	13283	-2,046	17,662
61501N	ICE	8418	-4,999	25
61501N	IC/EC	2380	-5,848	39,47
61501N	IR/D	3100	-7,918	95,659
61501S	ICE	8635	-1,138	21,572
61501S	IC/EC	2364	-2,986	32,381
61501S	IR/D	1949	-2,781	61,984
62002N	IR/D	3284	-3,007	34,54
62002S	IR/D	3208	-2,728	27,015
63001N	IR/D	5710	-3,039	25,344
63001S	IR/D	5680	-3,29	29,262
64002N	IC/EC	4937	-1,889	20,221
64002N	IR/D	1252	-3,27	91,004
64002S	IC/EC	4922	-5,199	39,502
64002S	IR/D	1270	-6,591	106,197
64501N	IC/EC	2115	-1,084	6,496
64501S	IC/EC	2110	-0,501	1,959
64506N	IR/D	5710	-1,798	20,54
64506S	IR/D	5635	-4,251	29,613
64514N	IC/EC	6859	-3,607	35,663
64514N	IR/D	1508	-5,485	87,539
64514S	IC/EC	5363	-3,12	25,15
64514S	IR/D	1534	-3,712	84,994
65001N	IC/EC	3862	-4,729	35,868
65001N	IR/D	4559	-3,736	38,209
65001S	IC/EC	3879	-3,92	33,853
65001S	IR/D	4643	-4,482	48,48
65101N	IR/D	1932	-2,174	16,054
65101S	IR/D	2163	-3,943	62,25
65501N	ICE	8519	-1,819	12,728
65501S	ICE	8611	-3,135	25,983
80001N	ICE	4635	-1,161	14,782
80001S	ICE	4616	-2,141	21,558

Lebenslauf

Persönliche Daten

Name:	Ralf Borchers, geb. Schollmeyer
geboren am:	06. Juli 1976
Geburtsort:	Magdeburg (Sachsen-Anhalt, Deutschland)
Familienstand:	verheiratet mit Sylvia Borchers
Kinder:	Maren Sophie Borchers
	Beeke Anna Borchers

Berufliche Tätigkeit

seit 09/2011	Mitarbeiter der megatel Informations- und Kommunikationssysteme GmbH, Bremen
11/2001 – 03/2008	wissenschaftlicher Mitarbeiter am Institut für Anwendungen der Geodäsie im Bauwesen (IAGB) der Universität Stuttgart

Ausbildung

10/1996 – 11/2001	Studium Vermessungswesen und Geodäsie, Universität Hannover
10/1995 – 09/1996	Grundwehrdienst
07/1995 – 09/1995	Vorbereitendes Praktikum beim Öffentlich bestellten Vermessungsingenieur Klose in Genthin
09/1991 – 06/1995	Börde-Gymnasium Wanzleben mit dem Abschluss der Allgemeinen Hochschulreife
09/1983 – 09/1991	Grundschule und Polytechnische Oberschule Klein Wanzleben

Dank

An vorderster Stelle meiner Dankesworte bedanke ich mich bei meiner Frau Sylvia Borchers. Sie hat mit mir viele Höhen und Tief inmitten sowie am Rande dieser Arbeit und manchmal auch darüber hinaus durchlebt. Das Leben mit ihr ist mehr als Ruhm oder Ehre und bereichert mich am meisten.

Die vorliegende Arbeit ist thematisch während meiner Tätigkeit am Institut für Anwendungen der Geodäsie im Bauwesen der Universität Stuttgart (seit 2010 Institut für Ingenieurgeodäsie der Universität Stuttgart (IIGS)) entstanden. Herr Prof. Dr.-Ing Wolfgang Möhlenbrink eröffnete mir an seinem damaligen Institut die Möglichkeit, mich mit dieser Thematik zu beschäftigen, wofür ich ihm sehr dankbar bin. Leider ist er kurze Zeit später verstorben. Bereits während dieser Zeit und im Anschluss unterstützte mich Herr Prof. Dr.-Ing. habil. Schwieger mit fachlicher Expertise und immerwährender Geduld. Ihm sowie uneingeschränkt allen anderen Kollegen des Instituts danke ich für die kollegiale, freundliche und zuweilen stimmungsvolle Atmosphäre.

Ebenso bedanke ich mich bei Herrn Prof. Dipl.-Ing. Dr. techn. Andreas Wieser für die vielen wertvollen Anmerkungen zur Veröffentlichung sowie Herrn Prof. Dr.-Ing. Markus Friedrich, der mich schon früh zu diesem Thema motiviert hatte. Beiden danke ich für die Übernahme des Korreferats.

Bei meinen Eltern bedanke ich mich für die unentwegt und in allen Formen gegebene Unterstützung. Und dafür, dass sie mich meinen eigenen Weg laufen ließen.

Freunde sind mit die wichtigsten Begleiter, die man haben kann. Daher ein ganz spezielles Dankeschön an die Familie Könnecke, die irgendwie immer Zeiten hatten, wenn ich sie brauchte.