

# User Contributed City-Wide Wi-Fi Model to Improve Client Localization

Diploma Thesis  
Daniel Sevrin

RWTH Aachen University, Germany  
Chair for Communication and Distributed Systems

Advisors:

Dipl.-Inform. Nicolai Viol  
Prof. Dr.-Ing. Klaus Wehrle  
Prof. Dr. Bernhard Rumpe

Registration date: 14-12-2012  
Submission date: 30-07-2012

---



---

I hereby affirm that I composed this work independently and used no other than the specified sources and tools and that I marked all quotes as such.

Hiermit versichere ich, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Aachen, den 14. September 2012



## Kurzfassung

---

Das Ziel dieser Arbeit ist die Berechnung der Signalausbreitung von WLANs anhand von Smartphone-Messungen, OpenStreetMap und Referenz-WLANs, deren Position bekannt ist. Die Android-Smartphones zeichnen in einem WLAN-Scan die BSSID, die Signalstärke, die SSID und den WLAN-Kanal auf. Zudem werden für die Positionsermittlung die Werte der Latitude, der Longitude, die Angaben der Accuracy und der Lokalisierungsart Network und Provider gesammelt. Nachfolgend werden die Signalstärken der WLANs samt Position gespeichert. Es wurden insgesamt 124.500 Messpunkte in einer Million WLAN-Messungen aufgenommen. Da die Positionsdaten eine zufällige Messabweichung enthalten, wird ein Partikelfilter eingesetzt, um diese zu minimieren. Die durchschnittliche Abweichung des Partikelfilters von der richtigen Position ist 9 m kleiner als die durchschnittliche Abweichung von den Positionsdaten. Aus den Geodaten werden die Referenzpunkte ermittelt, welche die Straßen und die öffentlichen Plätze repräsentieren. Mit dem nächsten Nachbarn als Map Matching-Algorithmus werden die Messungen auf die Referenzpunkte aggregiert. Damit werden die Messungen geografisch zusammengefasst, sodass über deren Masse die zufällige Messabweichung statistisch minimiert wird. Aus diesen Daten werden die Eingaben für das Log Distance Model bestimmt und somit die Signalausbreitung berechnet.



# Danksagungen

An dieser Stelle ergreife ich die Gelegenheit, mich bei all jenen Menschen zu bedanken, die mich während der Anfertigung der vorliegenden Diplomarbeit begleitet und gefördert haben.

In erster Linie gebührt mein größter Dank gewiss meinem Betreuer Herrn Nicolai Viol und Herrn Prof. Dr. Klaus Wehrle für die interessante Themenstellung und für die kompetente Begleitung meiner Arbeit in Form von konstruktiven Gesprächen. Ihre wertvollen Ideen, ihr intensives Engagement und ihre unermüdliche Unterstützung sowie die Zurverfügungstellung des nötigen Materials haben entscheidend zum Gelingen dieser Arbeit beigetragen. Bei Herrn Prof. Dr. Bernhard Rumpe bedanke ich mich ebenfalls recht herzlich.

Natürlich spreche ich auch allen Menschen meinen Dank aus, die eifrig und tatkräftig Messungen durchgeführt haben. Ohne sie wäre die konkrete Bearbeitung dieses Themas nicht möglich gewesen.

Ein herzliches Dankeschön gilt ebenso meiner Familie für ihre Geduld und ihre stete Unterstützung.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Grundlagen</b>	<b>3</b>
2.1	Wireless Local Area Network . . . . .	3
2.1.1	Scannen . . . . .	4
2.1.1.1	Aktives Scannen . . . . .	4
2.1.1.2	Passives Scannen . . . . .	5
2.1.2	Dezibel . . . . .	6
2.2	Signalausbreitung . . . . .	7
2.2.1	Pfadverlust . . . . .	7
2.2.2	Einflussfaktoren . . . . .	7
2.2.3	Mehrwegeausbreitung . . . . .	9
2.2.4	Ausbreitungsmodell . . . . .	10
2.2.4.1	Deterministischer Ansatz . . . . .	10
2.2.4.2	Empirischer Ansatz . . . . .	10
2.3	Globales Navigationssatellitensystem . . . . .	12
2.3.1	Global Positioning System . . . . .	12
2.3.2	Koordinatensystem . . . . .	14
2.4	Trigonometrie . . . . .	14
2.4.1	Ebene Trigonometrie . . . . .	15
2.4.1.1	Berechnung des Geradenschnitts . . . . .	15
2.4.1.2	Berechnung der Lage eines Punktes . . . . .	15
2.4.2	Sphärische Trigonometrie . . . . .	15
2.4.2.1	Berechnung der Distanz zwischen zwei Punkten . . . . .	16
2.4.2.2	Berechnung der Richtung von zwei Punkten . . . . .	17
2.4.2.3	Berechnung der Zielkoordinate . . . . .	17

2.5	OpenStreetMap . . . . .	18
2.6	Messabweichung . . . . .	20
	2.6.0.4 Systematische Abweichung . . . . .	20
	2.6.0.5 Zufällige Abweichung . . . . .	20
2.7	Filter . . . . .	21
	2.7.1 Partikelfilter . . . . .	22
	2.7.2 Problembeschreibung . . . . .	22
	2.7.3 Monte-Carlo-Methode . . . . .	23
	2.7.3.1 Importance Sampling . . . . .	23
	2.7.3.2 Sequential Importance Sampling . . . . .	24
	2.7.3.3 Bootstrap-Filter . . . . .	25
2.8	Datenschutz . . . . .	27
2.9	Zusammenfassung . . . . .	28
<b>3</b>	<b>Verwandte Arbeiten</b>	<b>29</b>
	3.1 Positionierung . . . . .	30
	3.2 Signalausbreitung . . . . .	31
	3.3 Vorhandene Anwendungen . . . . .	32
	3.4 Zusammenfassung . . . . .	33
<b>4</b>	<b>Entwurf</b>	<b>35</b>
	4.1 Übersicht . . . . .	35
	4.2 Client-Server-Architektur . . . . .	36
	4.2.1 Client . . . . .	36
	4.2.2 Server . . . . .	38
	4.3 Positionierung . . . . .	39
	4.3.1 Rastermodell . . . . .	39
	4.3.1.1 Way . . . . .	40
	4.3.1.2 Area . . . . .	40
	4.3.2 Rasterdatenbank . . . . .	41
	4.3.3 Partikelfilter . . . . .	42
	4.3.3.1 Bestimmung des Suchraumes . . . . .	42
	4.3.3.2 Initialisierung . . . . .	43

4.3.3.3	Zustandsmodell . . . . .	44
4.3.3.4	Sensormodell . . . . .	45
4.3.3.5	Resampling . . . . .	45
4.3.3.6	Gewichteter Mittelwert aller Partikel . . . . .	46
4.3.4	Map Matching . . . . .	46
4.4	Signalausbreitungsmodell . . . . .	48
4.4.1	Log Distance Model . . . . .	48
4.4.1.1	Pfadverlustexponent . . . . .	49
4.4.1.2	Pfadverlust . . . . .	49
4.4.1.3	Distanz . . . . .	50
4.4.2	Multi Wall Model . . . . .	50
4.4.2.1	Umgebungsmodell . . . . .	50
4.4.2.2	Dämpfungsfaktor . . . . .	52
4.5	Visualisierung . . . . .	52
4.6	Datenschutz . . . . .	53
4.7	Zusammenfassung . . . . .	54
<b>5</b>	<b>Implementierung</b>	<b>55</b>
5.1	Externe Bibliotheken . . . . .	55
5.1.1	Android . . . . .	55
5.1.2	Python . . . . .	56
5.2	Messungen . . . . .	58
5.2.1	Android-Anwendung . . . . .	58
5.2.2	Server . . . . .	59
5.3	Modell . . . . .	60
5.4	Anwendung . . . . .	60
<b>6</b>	<b>Evaluation</b>	<b>61</b>
6.1	Szenario . . . . .	61
6.2	Messungen . . . . .	63
6.2.1	RSSI . . . . .	64
6.2.2	Accuracy von GPS . . . . .	65
6.2.3	Frequenzen . . . . .	66
6.3	Positionierung . . . . .	67
6.4	Ausbreitungsmodell . . . . .	67
6.4.1	Pfadverlust . . . . .	68
6.4.2	Referenz-Pfadverlust . . . . .	68

<b>7 Zusammenfassung</b>	<b>71</b>
7.1 Ausblick . . . . .	71
<b>Literaturverzeichnis</b>	<b>73</b>
<b>A Anhang</b>	<b>77</b>
A.1 Abkürzungsliste . . . . .	77
A.2 Wetterdaten . . . . .	77

# Kapitel 1

## Einleitung

In der heutigen Zeit steigt das Interesse an mobilen Geräten und deren Lokalisierung. Die Entwicklung der mobilen Geräte schreitet kontinuierlich voran, jedoch übersteigt die Anforderung an die Lokalisierung derzeit noch die technischen Möglichkeiten. So ist beispielsweise die Ortung per GPS zwischen zwei Hochhäusern schwer durchzuführen, sodass nach einer Alternativen in dieser Situation gesucht wird.

Die Abdeckung der WLAN-Signale in Städten bietet eine zusätzliche Möglichkeit, um (s)einen Standort zu bestimmen. Mehrere Modelle existieren, die beschreiben, wie sich die WLAN-Signale in der Umgebung ausbreiten. Diese lassen sich in ein deterministisches und ein empirisches Modell einteilen:

**Deterministisches Modell:** Dieses Modell berechnet die Ausbreitung der Signale, indem die physikalische Umgebung als digitale Fassung zur Berechnung genutzt wird. Dies benötigt jedoch eine lange Entwicklungszeit, vor allem bei der Umsetzung auf eine gesamte Stadt. Auftretende Effekte, wie z. B. Dämpfung und Reflexion bei Wänden, werden beachtet.

**Empirisches Modell:** Es besteht aus Messungen in der realen Welt, um daraus die Ausbreitung der Signale zu bestimmen. Gerade über eine große Anzahl von Messungen kann eine Messabweichung minimiert werden.

Die Hauptaufgabe der vorliegenden Diplomarbeit besteht in der Zusammenführung des deterministischen und des empirischen Modells, sodass ein semi-empirisches Modell entwickelt wird. Damit sollen die Nachteile des alleinigen deterministischen und des alleinigen empirischen Modells getilgt werden, die Vorteile jedoch voll ausgeschöpft werden.

Bei dem empirischen Modell werden mehrere Messungen, verteilt über ein Gebiet, in dem die Signalausbreitung bestimmt werden soll, ermittelt. Andererseits sollen mehrere Messungen am selben Ort durchgeführt werden, um die Messabweichung zu minimieren.

Eine Möglichkeit für die Datenerfassung der Messungen ist, dass das Entwickler-Team die Messungen selbst tätigt. Wie oben beschrieben, erfordert dies einen hohen

Aufwand, besonders dann wenn eine gesamte Stadt anhand mehrerer Messungen aufgezeichnet werden soll. Über einen längeren Zeitraum beobachtet, werden neue WLANs hinzukommen und andere verschwinden, z. B. wenn Personen umziehen und ihren Access Point mitnehmen. Daher sollten die Messungen in regelmäßigen Abständen wiederholt werden, um die Signalausbreitung der WLANs zu aktualisieren. Damit der Entwickler den Aufwand dieser Messungen in einer Stadt nicht erbringen muss, wird diese Aufgabe den Anwendern übertragen. Die Idee besteht darin, dass die Anwender ihre Position und die vorhandenen WLANs samt Signalstärke mittels ihrer mobilen Geräte während ihres Alltages, wie z. B. beim Einkaufen, aufzeichnen. Das entsprechende Programm soll dabei im Hintergrund laufen, sodass die Anwender in der Nutzung ihrer mobilen Geräte nicht eingeschränkt werden. Anhand dieser Daten wird die Signalausbreitung von WLANs bestimmt. Ein Vorteil gegenüber den vom Entwickler aufgezeichneten Messungen ist, dass der Prozess der Messaufzeichnung seitens der Anwender durchgehend läuft und somit die Signalausbreitung der WLANs auf dem neuesten Stand bleibt.

Wenn der Entwickler die Messungen durchführen würde, könnte er Messgeräte einsetzen, die nur eine kleine Messabweichung aufweisen. Da der Anwender jedoch nicht über solche Messgeräte, die zudem in den Alltag eingebunden werden sollen, verfügt, führt er die Messungen mit seinem eigenen mobilen Gerät durch. Dabei ist jedoch zu beachten, dass das mobile Gerät nicht als Messgerät konzipiert wurde und daher meist eine höhere Messabweichung beim Gerät selber, aber auch durch dessen Handhabung entsteht. So besteht die erste Problemstellung dieser Arbeit darin, die Messabweichungen zu minimieren und zu evaluieren, wie genau die Messungen im Vergleich zu den wirklichen Signalausbreitungen sind.

Wie oben beschrieben, wird beim deterministischen Modell eine digitale Version der Umgebung benötigt. Um möglichst übereinstimmende Ergebnisse im Vergleich zu der Signalausbreitung zu erzielen, muss präzises Wissen der Umgebung in das Modell einfließen. Wenn der Entwickler dieses Umgebungsmodell für eine Stadt erstellen soll, dann ist dies ein zeitaufwendiges Unterfangen. Wie beim empirischen Modell sollte er in regelmäßigen Zeitabständen das Modell aktualisieren, damit die Signalausbreitungen der WLANs stets auf dem aktuellen Stand sind, sodass z. B. dem Neubau eines Gebäudes Rechnung getragen wird.

Das Projekt OpenStreetMap stellt Kartenmaterial frei zur Verfügung. Aus den Geodaten kann ein vereinfachtes Umgebungsmodell einer Stadt erstellt werden. Diese Geodaten enthalten allerdings nur ein ungenügendes Wissen über die Umwelt einer Stadt, um eine genau Messung der Signalausbreitung durchzuführen, da z. B. die Höhe der Gebäude nicht bekannt ist. Hier soll das empirische Modell zum Einsatz kommen, um gemeinsam mit dem deterministischen Modell die Ergebnisse der Signalausbreitung zu verbessern. Dies führt allerdings zur zweiten Problemstellung, die darin liegt, wie das empirische Modell mit dem deterministischen Modell zu einem semi-empirischen Modell vereint werden kann.

Der Vorteil des semi-empirischen Modells besteht darin, dass die Messungen zu dem Kartenmaterial, welches aus OpenStreetMap generiert wurde, hinzugefügt werden. Diese Datenstruktur bildet ein Fingerprinting. Das Fingerprinting beschreibt das Aufzeichnen von SSIDs, welche die WLANs und deren Signalstärken sowie eine Position im Raum, wo diese Aufzeichnung stattgefunden hat, identifizieren. Damit ist eine Lokalisierung möglich, indem aus einem WLAN-Scan die SSIDs und die Signalstärken mit den Daten aus dem Fingerprinting überprüft werden. Dabei sollen die SSID aus dem Scan, das Fingerprinting und die Signalstärke des Scans mit der

---

kleinen Abweichung sowie die Signalstärke der SSID aus dem Fingerprinting übereinstimmen. Da die Daten aus dem Fingerprinting einer Position zugeordnet sind, kann die Position des Scans zurückverfolgt werden. Hieraus entsteht die dritte Problemstellung: Für das Fingerprinting muss das gesamte Gebiet nach WLANs gescannt werden. Bei einer Lücke können die Scans, die in ebendieser Lücke aufgezeichnet werden, keiner Position zugeordnet werden. Dies kann bei dem oben genannten empirischen Modell auftreten, bei dem die Anwender die Messungen durchführen. Die Anwender bewegen sich in gewissen Straßen statistisch seltener als in anderen, sodass dort eine Lücke entsteht oder die Daten im Fingerprinting veraltet sind. Um dieses Problem zu lösen, wird ein Signalausbreitungsmodell entwickelt und eingesetzt, sodass die Ausbreitung der WLAN-Signale in der Umgebung berechnet wird. Dafür werden die Messungen und die Geodaten von OpenStreetMap als Eingabe genutzt.

Eine weitere Anwendung dieser Arbeit ist die Feststellung der WLAN-Abdeckung in einer Stadt. So wird es einem Hotspot-Anbieter ermöglicht, Lücken in seiner WLAN-Abdeckung zu bestimmen und diese Lücken anschließend mittels weiterer Access Points zu schließen. Der Anbieter kann zudem Gebiete ermitteln, in denen sich mehrere Access Points überlagern und untereinander Störungen verursachen. Diese Access Points kann er konfigurieren, um die Störungen, z. B. durch die Nutzung verschiedener WLAN-Kanäle, zu beheben oder um einen Access Point zu entfernen. Mit diesem Werkzeug kann der Anbieter sein stadtweites WLAN planen und optimieren.

Nach der Einleitung im vorliegenden Kapitel 1 werden in Kapitel ?? die wichtigsten Grundlagen für diese Arbeit erläutert. Die Schwerpunkte stellen das WLAN, das GPS sowie deren Signalausbreitung, die OpenStreetMap und die Filter dar. Verwandte Arbeiten und deren Bezug zu dieser Arbeit werden in Kapitel ?? vorgestellt. Das Kapitel ?? beschreibt den Entwurf und schildert, wie die drei Problemstellungen gelöst werden. Die Implementierung in Kapitel ?? erklärt, wie bei der Programmierung des Entwurfs vorgegangen wurde. In Kapitel ?? werden die Algorithmen und die Messungen schließlich ausgewertet. Zuletzt fasst das Kapitel ?? die essentiellen Erkenntnisse dieser Arbeit zusammen.



# Kapitel 2

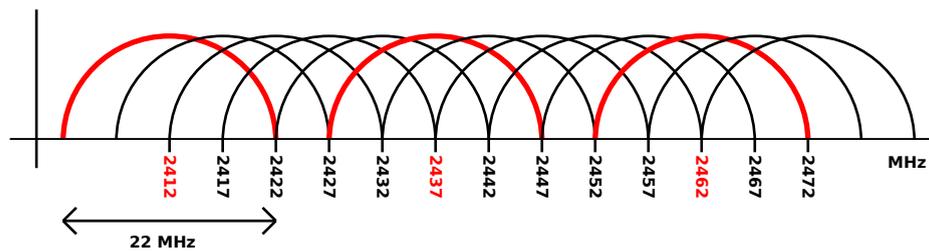
## Grundlagen

Die für diese Arbeit relevanten Begriffe, Methoden und Probleme werden zunächst erläutert. Zuerst wird im Unterkapitel 2.1 das WLAN erklärt und beschrieben, mit welchen Frequenzen es arbeitet und wie die Adressierung des WLANs funktioniert. Der nächste Punkt beschäftigt sich mit dem Auffinden des WLANs, dem sogenannten Scannen. Dabei werden zwei Arten, das aktive und das passive Scannen, unterschieden. Beim Scannen wird die Signalstärke ermittelt, die in der RSSI ausgedrückt wird. Das Dezibel gibt einen Indikator, ob zwei Pegel verstärkt oder abgeschwächt werden. Damit lässt sich die Dämpfung oder die Verstärkung der Signalstärken in einer kompakteren Darstellung angeben. Diese Grundlagen werden für das Auffinden, Identifizieren und Ermitteln der Signalstärken bei Messungen benötigt.

Im folgenden Unterkapitel ?? wird die Signalausbreitung erklärt. Das Kernelement ist die Beschreibung der physikalischen Faktoren, die auf die Signalausbreitung einen Einfluss haben. So werden Pfadverlust, Dämpfung, Streuung, Reflexion, Brechung und Beugung erläutert. Anschließend wird der Effekt der Mehrwegeausbreitung mehrerer Reflexionen von Funkwellen aufgezeigt. Zuletzt wird das Ausbreitungsmodell definiert: Es bietet die Berechnung, wie sich die Signale in der Umgebung ausbreiten, an. Damit werden Signalstärken auf eine Distanz ausgehend vom AP bestimmt. Als Modelle werden das Log Distance Model und das Multi Wall Model vorgestellt.

Das Unterkapitel ?? besteht aus dem Global Positioning System (GPS). Die Funktionsweise von GPS sowie deren Einflussfaktoren werden aufgezählt. Zusätzlich wird noch das Koordinatensystem WGS 84 beschrieben, welches von GPS verwendet wird. Mit GPS wird die Position der Messung bestimmt.

Das Unterkapitel ?? Trigonometrie befasst sich mit der Berechnung von Koordinaten im Raum. Hier wird zwischen ebener und sphärischer Trigonometrie unterschieden. Die Berechnung eines Schnittpunktes zweier Geraden und einer Distanz werden unter anderem dort erläutert. Dies wird für die Datenverarbeitung der Messungen genutzt. Das Unterkapitel 2.5 OpenStreetMap beschreibt das Projekt, welches freie Geodaten zur Verfügung stellt, die von jedem Benutzer bearbeitet werden können. Dieses Kartenmaterial kann so detailliert werden, dass nicht nur Straßen und Plätze bestimmt werden können, sondern auch Gebäude. Diese Daten sind später für das Rastermo-



**Abbildung 2.1** Schnitte der verschiedenen Kanäle von 802.11 im 2,4 GHz-Bereich. Die roten Kanäle sind überlappungsfrei.

dell und das Multi Wall Model relevant. Zudem wird die Datenstruktur von OSM erläutert.

Im nächsten Unterkapitel ?? wird die Messabweichung definiert und der Unterschied zwischen den zufälligen und den systematischen Messabweichungen dargelegt.

Im Unterkapitel 2.7 werden mehrere Filter, wie das Kalman-Filter, und die Problemstellung bei der Lösung von nicht linearen Modellen vorgestellt. Diese Filter können auf GPS-Daten angewendet werden, um die zufällige Messabweichung zu minimieren. Der Schwerpunkt ist das Partikelfilter sowie die Umsetzung anhand des Sequential Importance Samplings und des Bootstrap-Filters.

Zuletzt wird der Datenschutz in dem Unterkapitel 2.8 erklärt und bestimmt, was bei personenbezogenen Daten beachtet werden muss. Dies spielt eine Rolle, da die Messungen von den Anwendern ausgeführt und zur Verfügung gestellt werden, sodass dies mit dem Datenschutz vereinbar sein muss.

## 2.1 Wireless Local Area Network

Wireless Local Area Network (WLAN) ist eine kabellose Vernetzung von mehreren Computern über Funk. Für lokale Funknetzwerke gibt es mehrere Standards, wobei der gängigste Standard IEEE 802.11 ist. In den letzten Jahren sind auf Grund steigender Anforderungen, wie z. B. höherem Datendurchsatz, mehrere Erweiterungen entstanden. Die geläufigsten sind 802.11 a/b/g/n, die den 2,4 GHz- oder den 5 GHz-Bereich nutzen. In vielen Ländern ist das 2,4 GHz-Frequenzband lizenzfrei und muss bestimmte Regelungen, wie z. B. die maximal erlaubte Sendeleistung, erfüllen. In Deutschland ist diese maximale Sendeleistung auf 100 mW bei 2,4 Hz limitiert; für 5 GHz liegt sie bei 1.000 mW. Der 2,4 GHz-Frequenzbereich ist nach dem 802.11-Standard in 13 Kanäle eingeteilt. Der erste Kanal beginnt bei 2.412 MHz und jeder weitere Kanal ist in einem Abstand von 5 MHz eingeteilt. Der letzte, dreizehnte Kanal endet bei 2.472 MHz. Die Kanalbreite liegt bei 22 MHz, sodass sich fast alle Kanäle mit ihren Nachbarkanälen überlappen. Von diesen 13 Kanälen sind nur drei überlappungsfrei. Diese überlappungsfreien Kanäle sind Kanal 1, 6 und 11. Die Abbildung 2.1 stellt die Überlappung der Kanäle dar.

Die Vernetzung von mehreren Rechnern lässt sich in zwei Netzwerkformen umsetzen. Das Ad-hoc-Netzwerk ist eine Direktverbindung von zwei oder mehreren Rechnern. Der Vorteil ist, dass keine feste zentrale Infrastruktur, wie ein Access Point (AP), benötigt wird. Die einzige Voraussetzung ist, dass sich die Rechner in Reichweite der WLAN-Signale befinden und jeder Rechner den gleichen Kanal benutzt. Diese Netzwerkform wird meist für eine schnelle temporäre Vernetzung von mehreren

Rechnern genutzt.

Das Infrastruktur-Netzwerk ist eine zentrale Vernetzung von mehreren Rechnern über einen AP. Der AP kümmert sich um die gesamte Infrastruktur, die unter anderem aus der Datenadressierung besteht. Genau wie beim Ad-hoc-Netzwerk müssen sich alle Rechner sowie der AP im gleichen Kanal befinden. Alle Clients und der AP in Reichweite bilden im WLAN ein Basic Service Set (BSS). Die Basic Service Set Identification (BSSID) ist eine eindeutige Adressierung, die das BSS identifiziert. Das Format der BSSID ähnelt der 6 Byte langen MAC-Adresse. Die MAC-Adresse ist eine Adresse der Netzwerk-Hardware, die anhand dieser eindeutig identifiziert wird. Sie verändert sich nicht während der Lebensdauer des APs. Die BSSID eines APs im Infrastruktur-Modus wird der MAC-Adresse gleichgesetzt oder durch eine Zufallszahl erstellt. Meist handelt es sich jedoch um die MAC-Adresse, die häufig byteweise in hexadezimal geschrieben wird, wobei jedes Byte durch ein Zeichen, beispielsweise : oder -, getrennt wird. Ein Beispiel für eine Adresse lautet 00:17:df:a7:e8:e3. Die BSSID dient der Zuordnung der übertragenden Daten einer Funkzelle bzw. eines APs. Mehrere APs können untereinander zu einem WLAN verbunden werden. Da alle APs unterschiedliche BSSIDs haben, folgt die Zuordnung eines WLANs mittels der Service Set Identifier (SSID). So können mehrere APs zu einem Verbund definiert werden und der Client kann, z. B. anhand der höchsten Signalstärke, entscheiden, mit welchen APs er sich verbinden möchte. Die SSID ist ein Netzwerkname, der maximal 32 Byte lang sein kann.

### 2.1.1 Scannen

Bevor eine Verbindung mit einem WLAN aufgebaut werden kann, muss das Netzwerk zunächst gefunden werden. Diese Identifikation von WLANs wird Scannen genannt [6]. Mit Scanner sind alle Endgeräte gemeint, die WLANs suchen. Dabei wird das Scannen in zwei Arten, das aktive und das passive Scannen, eingeteilt.

#### 2.1.1.1 Aktives Scannen

Das aktive Scannen findet statt, wenn sich der Scanner im Kanal aktiv meldet und auf eine Antwort der APs wartet, die sich ihm zu erkennen geben möchten. Dafür wird ein Paket, der Probe Request Frame, verschickt, um WLANs in seiner Umgebung ausfindig zu machen oder um mehrere Informationen von anderen APs in seiner Reichweite abzurufen. Der Scanner schickt dieses Probe Request Frame in jeden Kanal und wartet auf einen Probe Response Frame eines APs. Die APs, die diesen Probe Request Frame erhalten, schicken eine Antwort samt Daten, wie Verschlüsselung, zurück und identifizieren sich gegenüber den anderen Endgeräten. Der AP kann das Versenden des Probe Request Frames verhindern, um sich vor dem Scanner zu verbergen. Der Vorteil ist die schnelle Suche nach WLANs. Dies verbraucht jedoch mehr Energie. Für aktives Scannen legt der Scanner die Parameter Probe Delay, MinChannelTime und MaxChannelTime fest. Das Probe Delay definiert ein Zeitfenster nach dem Kanalwechsel und vor dem Senden des ersten Probe Request Frames. Die MinChannelTime ist die minimale Zeit, die der Scanner auf Pakete von anderen Clients im Kanal wartet. Der Scanner kann somit feststellen, ob dieser Kanal benutzt wird. Die MaxChannelTime ist die maximale Zeit, die der Scanner

im Kanal wartet. In dieser Zeit erwartet er alle Probe Response Frames der APs in seiner Reichweite. Eine weitere Entscheidung betrifft die Reihenfolge, in welcher die Kanäle gescannt werden. Die Wahrscheinlichkeit, dass alle APs im Kanal ermittelt werden, kann durch die Reihenfolge der Kanäle beim Scannen vergrößert werden. Dieser Effekt entsteht durch die Kanalüberlappung, sodass der Scanner auch Beacons der Nachbarkanäle empfängt. Für die größtmögliche Wahrscheinlichkeit wechselt der Scanner in den nächsten überlappungsfreien Kanal. Der Vorgang des aktiven Scannens durchläuft folgende Schritte:

1. Der Scanner wechselt in einen Kanal und wartet, ob ein AP eine Antwort auf einen Probe Request Frame eines anderen Scanners verschickt. Sobald diese Antworten im Zeitfenster des Probe Delays ankommen, werden diese APs identifiziert. Nach dem Ablauf des Probe Delays geht er zu dem nächsten Schritt über.
2. Der Scanner schickt einen Probe Request Frame in den aktuellen Kanal hinaus.
3. Nun wartet er mindestens die MinChannelTime im Kanal ab und beobachtet diesen Kanal. Wenn in dieser Zeit keine Bewegung im Netzwerk festgestellt wird, dann geht der Scanner davon aus, dass sich keine APs in diesem Kanal befinden und wechselt in den nächsten Kanal. Wenn er jedoch Pakete von anderen Clients erkennt, wartet er die MaxChannelTime auf die Reaktionen, sprich auf einen Probe Response Frame.

Der Scanner kann anhand des Probe Request Frames anhand bestimmter Parameter seine Suche einschränken. So kann z. B. mittels BSS Type festgelegt werden, ob Ad-hoc-, Infrastruktur- oder alle Netzwerke gesucht werden.

### 2.1.1.2 Passives Scannen

Im Gegensatz zum aktiven Scannen wird beim passiven Scannen gewartet, bis sich der AP selbstständig meldet. Meist wird durch diesen alle 100 ms ein Beacon ausgesendet, um seine Präsenz anderen Geräten zu zeigen. Wie beim Probe Response Frame muss ein Beacon auch die nötigen Informationen für einen Verbindungsaufbau enthalten. Während eines Scans wartet der Scanner mindestens 100 ms in einem Kanal. Wenn er einen kürzeren Warteintervall hätte, könnte er genau zwischen zwei Beacons warten und somit das WLAN nicht erkennen. Während des Wartens zeichnet er alle Beacons auf. Danach führt er dies mit dem nächsten Kanal fort, bis er alle Kanäle gescannt hat. Im Vergleich zum aktiven Scannen ist der Batterieverbrauch durch das ausbleibende Senden des Probe Request Frames niedriger, da er kein Paket ausschickt und somit keine Energie für das Senden verbraucht. Dafür benötigt der Scanner jedoch eine längere Zeit, bis er alle APs festgestellt hat. Ein weiterer Vorteil ist, dass dieser Scan nicht erkennbar und nachweisbar ist, weil er eben kein Paket verschickt. Im Gegensatz zum aktiven Scan kann der passive auch verborgene SSIDs entdecken, da dieser die gesendeten Pakete zwischen dem verborgenen WLAN und dessen Clients mitverfolgen kann. Dabei beachtet er nicht nur die Beacons, sondern alle versendeten Pakete zwischen den APs und den anderen Clients.

### 2.1.2 Dezibel

Das Dezibel (dB) beschreibt einen relativen Vergleich zwischen zwei physikalischen Werten [7]. Es gibt an, ob zwischen den beiden Werten eine Verstärkung oder eine Dämpfung stattgefunden hat. Die positiven Werte von dB stellen eine Dämpfung dar und die negativen Werte eine Verstärkung der Leistung. Das Dezibel ist einheitslos und bezeichnet nur das Verhältnis zwischen zwei gleichartigen physikalischen Werten  $P_1$  und  $P_2$ .

$$L_{dB} = 10 \cdot \log_{10} \left( \frac{P_2}{P_1} \right) \quad (2.1)$$

Der Wert  $P_2$  ist die betrachtete Leistung, die mit der Bezugsgröße  $P_1$  verglichen wird. Es handelt sich um logarithmische Werte. Dadurch wird die Anzahl der Dezimalstellen im Vergleich zu den physikalischen Werten reduziert, wodurch eine vereinfachte, kompaktere Darstellung der Werte möglich wird. Ein weiterer Vorteil des Logarithmus ist die Vereinfachung der Berechnung, sodass eine Multiplikation der physikalischen Werte einer Addition sowie eine Division einer Subtraktion in Dezibel entspricht.

Wenn die Bezugsgröße  $P_1$  auf 1 mW gesetzt wird, ist die Größenangabe in dBm festgelegt. Beim Rechnen mit dB- und dBm-Werten müssen die folgenden Regeln beachtet werden [?]:

$$\begin{aligned} dB \pm dB &= dB \\ dBm \pm dB &= dBm \\ dBm - dBm &= dB \end{aligned} \quad (2.2)$$

Die Addition von zwei dBm-Werten kann nicht direkt berechnet werden. Dafür müssen diese Werte mit der Basis 10 potenziert sein und können dann erst addiert werden. Bei Verwendung als dBm wird die Summe wieder nach der Formel 2.1 mit  $P_2 = 1$  umgerechnet.

### RSSI

Die Received Signal Strength Indication (RSSI) ist ein Identifikator für die Empfangsleistung [19]. Er dient dazu, bei einem schlechten Empfang einer WLAN-Verbindung, diese Verbindung zu trennen und nach einem WLAN mit einer besseren Empfangsleistung zu suchen. Somit können die RSSI-Werte der gleichen Netzwerkhardware miteinander verglichen werden. Dies ist jedoch nicht möglich mit RSSI-Werten einer anderen Netzwerkhardware, da die RSSI-Angabe nicht auf eine physikalische Einheit nach einem Standard vereinbart wurde. So hat die RSSI nicht unbedingt eine Beziehung zu der dBm-Einheit. Dieser Identifikator besteht aus einem Byte. Den minimalen Wert und den maximalen Wert im Byte-Bereich bestimmt der Netzwerkhardware-Hersteller.

## 2.2 Signalausbreitung

In der Elektrotechnik besteht ein Funksignal aus elektromagnetischen Wellen [19, 25], die ein Zeichen oder eine Zeichenfolge im freien Raum übermitteln. Die Signalausbreitung ist eine Beschreibung der Funkwellen, wie sie sich im Raum ausbreiten. Eine

Vielzahl von Faktoren ist verantwortlich für die Ausbreitung, wobei diese Faktoren in die beiden Klassen Hardware und Einflussfaktoren im Funkfeld eingeteilt werden. Wichtige Eigenschaften der Hardware, die Frequenz und der Antennentyp, bestimmen die Ausbreitung der Signale. Die Wahl der Frequenz entscheidet mitunter über die Reichweite der Signale. Als Faustregel gilt, dass niedrige Frequenzen eine größere Reichweite als höhere Frequenzen erlangen. Die Antennen haben eine Ausstrahlungscharakteristik, die den Wirkungsgrad und deren Richtung beschreibt. So kann man entscheiden, welche Antenne benutzt und wie sie ausgerichtet wird. Eine Richtantenne hat z. B. einen kleinen Wirkungsgrad, daher sollte sie auf den Empfänger ausgerichtet werden.

In der zweiten Klasse befinden sich alle Faktoren, die im Funkfeld zwischen dem Sender und dem Empfänger liegen und die Ausbreitung beeinflussen. Dazu gehören der Pfadverlust in Kapitel ?? und die Einflussfaktoren in Kapitel 2.2.2.

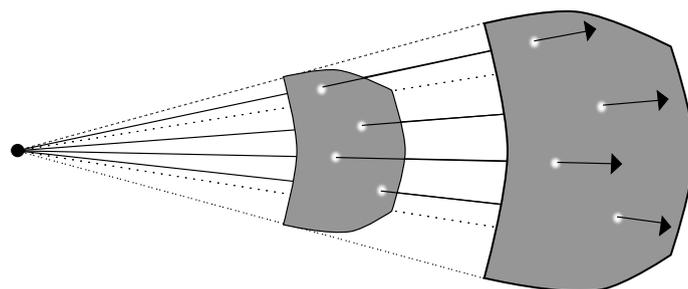
### 2.2.1 Pfadverlust

Ein Signal bildet eine gerade Linie zwischen dem Sender und dem Empfänger und wird in der Signalausbreitung als Sichtlinie bezeichnet. Wenn sich kein Hindernis in der Sichtlinie befindet, erfährt das Signal einen Pfadverlust [20]. Es repräsentiert die Signaldämpfung, die als positiver Wert in dB angegeben und aus der Differenz der Sendeleistung und Empfangsleistung berechnet wird [18]. Bei dieser Berechnung besteht die Möglichkeit, den Antennengewinn miteinzubeziehen. Über die Distanz zwischen dem Sender und dem Empfänger sinkt die Signalstärke, sodass die Empfangsleistung nicht gleich der Sendeleistung ist. Die Empfangsleistung  $P_r$  ist im Vakuum proportional zu  $\frac{1}{d^2}$  mit der Distanz  $d$  zwischen dem Sender und dem Empfänger. Diese Proportionalität ist als Abstandsgesetz bekannt. Dieses Gesetz ist in der Abbildung ?? dargestellt. Der Sender verschickt ein Signal, welches sich kugelförmig ausbreitet. Je größer die Distanz zwischen dem Sender und dem Signal ist, desto größer wird die Kugeloberfläche, die das Signal darstellt. Auf diese Kugeloberfläche verteilt sich die Sendeleistung gleichmäßig. Da die Kugeloberfläche immer größer wird, die Sendeleistung bei der Ausstrahlung aber konstant bleibt, verringert sich die Leistung pro Fläche. Diese Kugeloberfläche  $s$  steigt mit der Distanz vom Sender nach der Formel  $s = 4\pi d^2$ . Dadurch entsteht der Pfadverlust [20].

Das Vakuum ist das einzige Medium, in dem keine Beeinträchtigung der Signale im Vergleich zu anderen Medien, wie Luft, stattfindet, die unter anderem durch Feuchtigkeit und Temperatur unterschiedlich dämpft. Das Vakuum stellt einen idealen Raum dar. In der Realität beeinflussen allerdings mehrere Faktoren die Signalausbreitung.

### 2.2.2 Einflussfaktoren

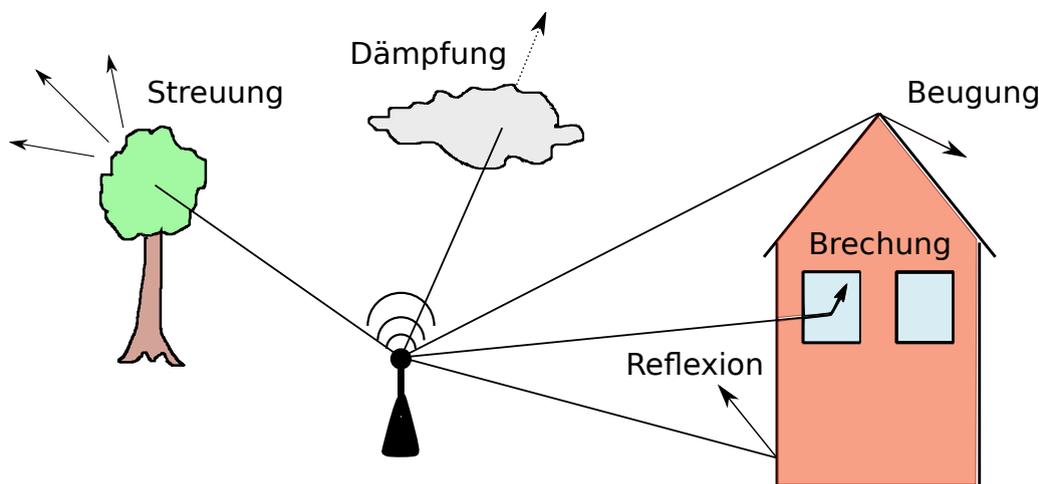
Die folgenden Einflussfaktoren verursachen Effekte, die die Ausbreitung der Funkwellen beeinträchtigen. Diese Faktoren befinden sich im Funkfeld zwischen dem Sender und dem Empfänger. Damit sind keine Hardwarefaktoren, wie Frequenz, Sendeleistung und Antennentyp, gemeint. Das einzige Funkfeld, in dem es keine Einflussfaktoren gibt, ist ein Vakuum-Raum. Die Abbildung 2.3 stellt die fünf Einflussfaktoren dar. Diese Faktoren sind [20, 25]:




---

 Abbildung 2.2 Pfadverlust über eine größere Distanz
 

---




---

 Abbildung 2.3 Darstellung der fünf Einflussfaktoren
 

---

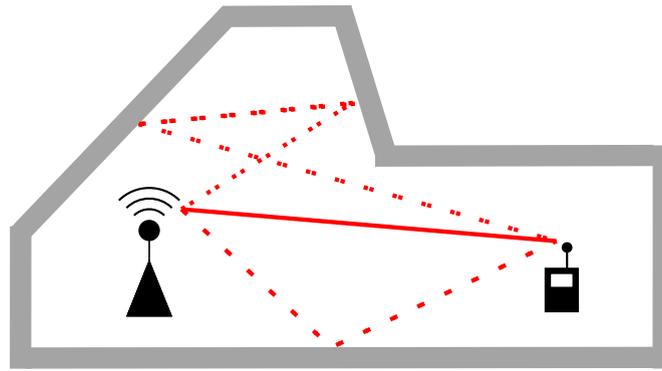
**Dämpfung:** Die Dämpfung ist eine Abschwächung der Signalstärke, wobei die Funkwelle in die gleiche Richtung ausstrahlt. Damit ist nicht der Pfadverlust gemeint, sondern jedes Medium, das eine Dämpfung hervorruft. Die Energie der Sendeleistung geht im Medium z. B. in Wärme über. Wie stark die Dämpfung ist, hängt von der Eigenfrequenz im Vergleich zu der Frequenz des Signals ab. So kann bei Ausbreitung im Wasser meist eine höhere Dämpfung als in der Luft entstehen. Auch Veränderungen am Medium selber, wie Temperatur und Luftfeuchtigkeit, führen zu einer anderen Dämpfung.

**Reflexion:** Wenn eine Funkwelle aus einem Medium auf ein anderes Medium trifft, wird ein Teil der Welle von dieser Fläche reflektiert und der andere Teil durchdringt dieses Medium gedämpft.

**Beugung:** Eine Welle kann gebeugt werden, wenn sie auf eine Kante eines Objektes trifft. Dadurch verändert sich die Richtung des Signals.

**Brechung:** Sobald die Wellen von einem Medium in ein anderes Medium eindringen, kann sich die Richtung des Signals ändern. Dafür muss die Welle schräg auf die Grenzfläche zweier Medien treffen. Diese Medien müssen eine unterschiedliche Dichte aufweisen, sodass die Geschwindigkeit der Signalausbreitung beider Medien verschieden ist. Im Gegensatz zur Reflexion breiten sie sich durch das Medium aus und werden nicht vom Medium zurückgeworfen.

**Streuung:** Alle Hindernisse, die gleich oder kleiner als die Wellen sind, können Wellen streuen. Wie bei der Reflexion wird die Welle vom Objekt zurückge-



**Abbildung 2.4** Darstellung der Mehrwegeausbreitung

worfen. Dazu kommt beim Zurückwerfen, dass sich die Welle in mehrere Wellen aufteilt, sodass deren Leistung in den gestreuten Wellen sinkt.

Beim sogenannten Body Shadow, dem Körperschatten, werden die Signale durch den eigenen Körper abgeschattet. Durch Reflexion und andere Faktoren kommen die Signale nicht nur über die Sichtlinie beim Empfänger an, sondern können sich auch über mehrere Umwege ausbreiten.

### 2.2.3 Mehrwegeausbreitung

Die Mehrwegeausbreitung beschreibt den Effekt, bei dem sich eine Funkwelle nicht nur über die Sichtlinie ausbreitet, sondern auf Grund der Einflussfaktoren, in Kapitel 2.2.2 beschrieben, über mehrere Umwege zum Empfänger gelangt [25]. Die Abbildung 2.4 vermittelt einen Eindruck, wie die Mehrwegeausbreitung funktioniert. Die rote Linie ist die Sichtlinie und somit der kürzeste Weg zwischen dem Sender und dem Empfänger. Dieses Signal kommt als erstes an und hat die größte Signalstärke, weil es nicht durch Reflexionen und eine größere Distanz abgeschwächt wurde. Die beiden gestrichelten roten Linien werden von der grauen Wand reflektiert. Es ist möglich, dass ein Signal mehrmals reflektiert wird, bis es bei dem Empfänger ankommt. Diese beiden Signale nehmen somit einen Umweg im Vergleich zur Sichtlinie. Dadurch sind die empfangenen Wellen phasenverschoben und verschiedene Wellen überlagern sich. Eine Welle folgt der Sichtlinie. Währenddessen sind die anderen Wellen über die Umwege noch nicht eingetroffen und kommen erst zeitversetzt an. Daher stimmt die Amplitude der Wellen zur gleichen Zeit nicht überein, sodass je nach Situation eine Verzerrung, eine Verstärkung oder eine Tilgung des Signals entstehen kann.

**Verzerrung:** Durch eine Verzerrung der Wellen kann das eigentliche Signal gedämpft werden. Gerade in realen Situationen können viele verschiedene Wellen eine Interferenz bilden.

**Verstärkung:** Eine Verstärkung tritt meist in Tunneln und in Straßenschluchten auf. Das Signal kann über die Umwege zur gleichen Zeit beim Empfänger ankommen, sodass es verstärkt wird.

**Tilgung:** Dabei kann die Signalstärke aller Wellen größer sein als die Signalstärke der Wellen über der Sichtlinie, dies führt zur Tilgung des Signals.

Diese Effekte definieren das Short Term Fading [20]. Im Gegensatz dazu ist das Long Term Fading [20] die Veränderung der durchschnittlichen Empfangsleistung, die durch Distanzveränderungen zwischen dem Empfänger und dem Sender sowie durch Hindernisse in der Umwelt entstehen.

## 2.2.4 Ausbreitungsmodell

Ein Ausbreitungsmodell ist eine empirische oder analytische Funktion, die die empfangene Signalstärke in einer vorgegebenen Distanz vom Sender berechnet [18]. Anstatt die Signalstärke für einen Punkt zu ermitteln, kann sie je nach Modell für einen Raum bestimmt werden. Diese Berechnungen können sehr komplex sein, da die Signalausbreitung von vielen Faktoren abhängig ist. Diese Faktoren sind in den Kapiteln ?? und 2.2.2 beschrieben. Daher sind die Modelle meist an ein Szenario wie Innen- und Außenbereich oder an Frequenzen für eine spezifizierte Funkübertragung wie GSM gebunden. Es gilt, den deterministischen und den empirischen Ansatz zu unterscheiden.

### 2.2.4.1 Deterministischer Ansatz

Ein deterministisches Modell beruht auf der Berechnung der Signalstärke von elektromagnetischen Wellen unter der Beachtung der physikalischen Einflussfaktoren [29], in Kapitel 2.2.2 erklärt. Die Eingaben für die Funktion sind die Position und die Sendeleistung des Access Points (AP) sowie eine digitale Repräsentation der Umwelt. In diesem Umgebungsmodell wird die Ausbreitung der Signale mit Hilfe der physikalischen Theorie der Wellenausbreitung simuliert. Die Berechnung beruht komplett auf der Simulation der Wellen und nutzt keine Messungen oder Daten, die vor Ort aufgezeichnet wurden. Die Genauigkeit der Ergebnisse ist von der Präzision des Umgebungsmodells im Vergleich zur Realität abhängig. Dies bedeutet einen zeitintensiveren Aufwand und stellt sich als ein großer Nachteil dieses Ansatzes heraus. Bekannte deterministische Modelle sind unter anderem Okumura-Hata, Electronic Communication Committee-33 (ECC-33) und Stanford University Interim (SUI) [23].

### 2.2.4.2 Empirischer Ansatz

Beim empirischen Ansatz werden Messungen der Signalstärke und deren Position vorgenommen, die eine statistische Repräsentation der Umwelt wiedergeben. Im Vergleich zum deterministischen Ansatz ist der empirische einfacher umzusetzen, da keine physikalischen Theorien der Wellenausbreitung auf Signale simuliert werden müssen. Ein weiterer Vorteil ist, dass er entgegen dem deterministischen Ansatz nicht ortsbezogen ist und somit kein Umgebungsmodell notwendig ist. Ein Nachteil des empirischen Ansatzes besteht darin, dass viele Messungen benötigt werden, bis die zufälligen Messabweichungen minimiert sind. Bekannte Vertreter sind das Log Distance Model und das Multi Wall Model, die nachfolgend genauer beschrieben werden.

### Log Distance Model

Das Log Distance Model (LDM) ist eine einfache empirische Berechnung des Pfadverlustes über eine Distanz [11, 29]. Es berechnet die Signalstärke, die ausgehend vom Sender beim Empfänger ankommt. Diese Signalstärke  $L_{LDM}$  ist in dB angegeben und wird wie folgt ermittelt:

$$L_{LDM} = L_0 + 10 n \log_{10} \left( \frac{d}{d_0} \right) \quad (2.3)$$

$L_0$  ist der Referenz-Pfadverlust in der Distanz  $d_0$  zwischen dem Sender und dem Empfänger in dB im Vakuum. Für kleinere Gebiete wird 1 m für  $d_0$  gewählt. Der Pfadverlustexponent  $n$  beschreibt, wie stark die Signalstärke über eine Distanz abfällt. Ein kleiner Pfadverlustexponent bedeutet, dass die Signalstärke über eine Distanz schwach gedämpft wird. Bei einem großen Wert des Exponenten wird das Signal über die gleiche Distanz stärker abgeschwächt. In einem Vakuum liegt der Wert bei 2 und in der städtischen Umgebung etwa um 5.  $d$  ist die Distanz vom Sender zum Empfänger in Metern. Die empirischen Parameter sind  $L_0$  und  $n$ , die aus Messungen bestimmt werden. Es braucht keine weiteren Details der Umwelt, wie z. B. ein 3D-Modell. Die Umsetzung erfolgt somit leicht aus den Messungen. Der Nachteil ist, dass nur der Pfadverlustexponent die ganze Umwelt und die daraus folgenden Ausbreitungsphänomene beschreibt. So kann in einer Straßenschlucht ein niedrigerer Exponent vorhanden sein als in einem Gebäude. Daher werden Reflexionen und Abschattungen von Wänden und Gebäuden nicht berücksichtigt, das bedeutet dass die Ergebnisse eine höhere Fehlerquote aufweisen.

### Multi Wall Model

Das Multi Wall Model (MWM) [11, 29] ist eine Erweiterung des LDMs. Dabei wird ein digitales Modell der Umwelt erstellt. Dämpfungsfaktoren werden jeder Wand und jedem Flur zugeordnet. Es handelt sich um ein semi-empirisches Modell, das Messungen mit einem Ray-Tracer kombiniert. Da Hindernisse in der Umwelt die Signale unterschiedlich dämpfen, werden alle Elemente des geometrischen Modells nach ihren Dämpfungsfaktoren klassifiziert. Die MWM-Formel lautet:

$$L_{MWM} = L_0 + 10 n \log_{10} \left( \frac{d}{d_0} \right) + \sum_{i=1}^N k_{wi} L_{wi} + k_f L_f \quad (2.4)$$

Sie beschreibt das LDM sowie dessen Erweiterung um die Anzahl der Schnitte durch Wände. Der Dämpfungsfaktor  $k_{wi}$  ist die Anzahl des  $i$ -ten Wandtyps und  $L_{wi}$  ist der Dämpfungsfaktor für den  $i$ -ten Wandtyp. Die Anzahl der Flure  $k_f$  wird mit ihrem Dämpfungsfaktor  $L_f$  multipliziert.  $N$  ist die Anzahl aller Wandtypen.

Die empirischen Parameter  $L_{wi}$  und  $L_f$  werden anhand der Messungen bestimmt. Die Anzahl der Schnitte eines Wandtyps oder Flurs wird durch eine Berechnung vom AP zum Empfänger anhand des geometrischen Modells ermittelt. Im Vergleich zum LDM gibt MWM die genauere Signalstärke an. Jedoch ist zu beachten, dass Reflexion und Beugung in dieser Berechnung nicht simuliert werden und somit eine Abweichung zu den Messungen vorliegt.

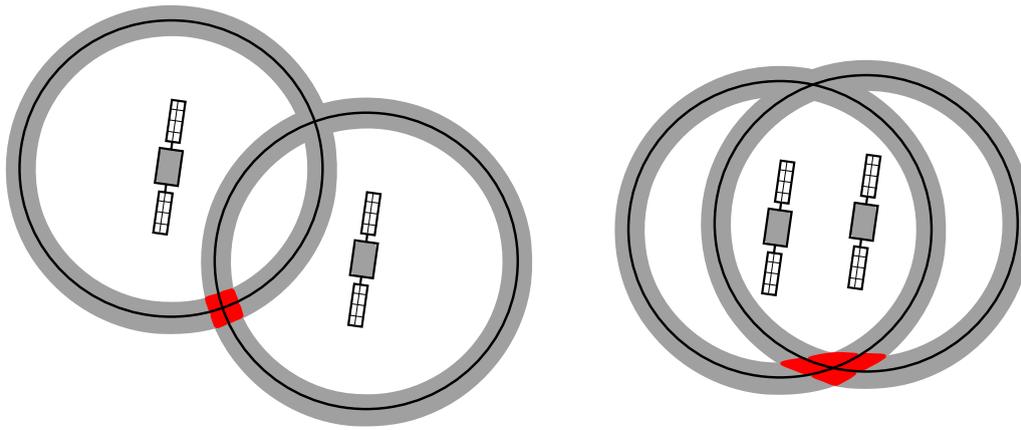


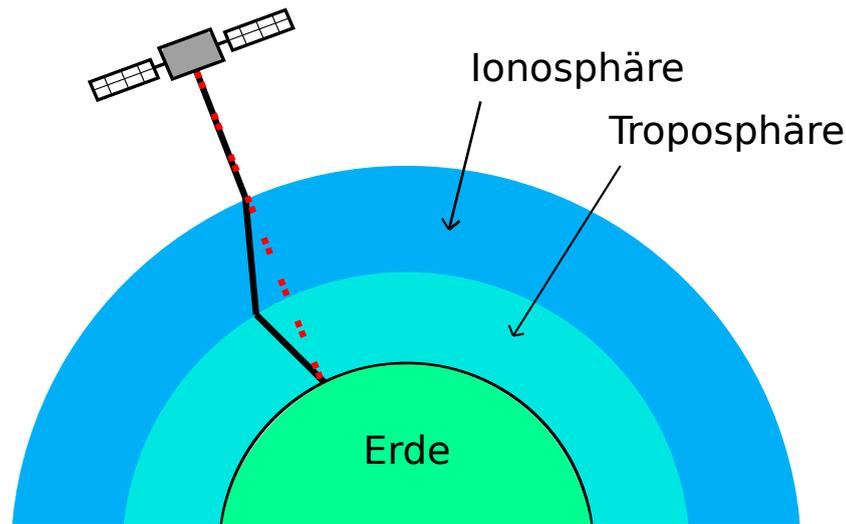
Abbildung 2.5 Schnitt der Laufzeitmessungen von mehreren Satelliten

## 2.3 Globales Navigationssatellitensystem

Heutzutage ist ein globales Navigationssatellitensystem (GNSS) die gängigste Methode für die Ermittlung der aktuellen Position auf der Erde. Das System ist ein Verbund von mehreren Satelliten, die die Erde umkreisen [3]. Diese Satelliten funken ihre Position im All und die Uhrzeit. Über die Entfernung zwischen dem Satelliten und dem Empfänger bekommt der Empfänger die Uhrzeit zeitversetzt übermittelt. Diese Laufzeitmessung ist proportional zur Distanz zwischen dem Empfänger und dem Satelliten und lässt sich daraus berechnen. Diese Technik ist auch als Time of Arrival bekannt. Für eine Positionierung müssen die Laufzeitmessungen von mindestens vier Satelliten empfangen werden. Die Abbildung 2.5 zeigt die Schnitte der Laufzeitmessung von zwei Satelliten. Da bei zwei Satelliten zwei Schnittpunkte entstehen, ist die Positionierung nicht eindeutig. Dies ist erst bei drei Satelliten möglich. Der vierte Satellit dient dazu, die genaue Uhrzeit zu ermitteln, um die Ungenauigkeit der Uhrzeit des Empfängers mit den Atomuhren des Satelliten zu korrigieren. Es gibt verschiedene GNSS, wie GPS aus den USA, GLONASS aus Russland, Galileo aus Europa usw. Das gängigste System im privaten Gebrauch ist das GPS. Diese Arbeit wird sich daher auf dieses System beschränken.

### 2.3.1 Global Positioning System

Das Global Positioning System (GPS) [3] ist ein GNSS, das unter der Leitung des US-amerikanischen Militärs steht. Es wurde 1977 zunächst für militärische Zwecke eingeführt und ist seit 1990 vollständig funktionsfähig. Ab 1. Mai 2000 wurde das künstliche Signalrauschen, Selective Availability, vom Militär entfernt, sodass eine genaue Positionierung für den zivilen Gebrauch bis auf 10 m ermöglicht wird. Das GPS arbeitet mit dem Koordinatensystem WGS 84, das in dem kommenden Kapitel 2.3.2 erläutert wird, und besteht aus 24 Satelliten. Die Konstellation dieser Satelliten ist so aufgebaut, dass zu jeder Zeit fünf bis acht Satelliten bei freier Sicht am Himmel empfangbar sind. Dadurch ist es theoretisch möglich, sich zu jeder Zeit an jedem Ort der Erde zu lokalisieren. Praktisch beeinflussen mehrere Faktoren die Signalausbreitung.



**Abbildung 2.6** Laufzeitfehler durch die Troposphäre und die Ionosphäre

## Einflussfaktoren

Mehrere Einflussfaktoren bestimmen die Berechnung der Positionierung und können daher einen Fehler verursachen. Ein absichtlicher Faktor war die Selective Availability. Diese Verschlechterung der Daten wurde vom US-Militär hinzugefügt, um die Nutzung des GPS für militärische Kräfte anderer Länder unbrauchbar zu machen. Heutzutage beeinflussen die physikalischen Einflussfaktoren den zivilen Gebrauch. Diese Faktoren bestehen aus [3]:

**Satellitenuhren:** Um die Laufzeitmessung zu ermitteln, müssen die Uhrzeit der Satelliten und die Uhrzeit des Empfängers synchronisiert werden. Bei dieser Synchronisierung kann eine Abweichung von einem Meter entstehen.

**Satellitenumlaufbahnen:** Durch physikalische Kräfte kreisen die Satelliten nicht immer konstant auf ihrer Bahn. Kontrollstationen auf der Erde berechnen diese Umlaufbahnen. Die Ortung kann um einen Meter abweichen.

**Ionosphärische Laufzeitverlängerungen:** Durch Sonneneinstrahlung in die Ionosphäre können Dämpfungen oder Reflexionen bei der Signalausbreitung entstehen. Diese verursachen eine Geschwindigkeits- und Richtungsveränderung der Signale, wie in Abbildung 2.6 erläutert. So kann eine Abweichung bis zu zehn Metern betragen.

**Troposphärische Laufzeitverlängerungen:** In der Troposphäre können Dämpfungen oder Reflexionen auf Grund von Wetterbedingungen, z. B. dichter Wolkendecke, auftreten. Wie in der Ionosphäre können sich die Geschwindigkeit und die Richtung des Signals ändern, wie in Abbildung 2.6 ebenfalls ersichtlich ist. Der große Unterschied ist, dass die Troposphäre eine niedrigere Abweichung von einem Meter bewirken kann.

**Mehrwegeeffekte:** Durch Reflexionen an Gebäuden, die sich in der Nähe des Empfängers befinden, kann das gleiche Signal über mehrere Wege zum Empfänger gelangen. Die Abweichung reicht bis 0,5 m. Dieser Effekt ist in Kapitel 2.2.3 ausführlich beschrieben.

**Satellitenkonstellation:** Die Position der Satelliten und deren Sichtbarkeit am Himmel verursacht auch eine Abweichung der Positionierung. Die Abbildung 2.5 zeigt zwei verschiedene Konstellationsmöglichkeiten. Der schwarze Kreis um den Satelliten ist das gesendete Signal und der graue Umkreis stellt die möglichen Abweichungen des Signals durch die vorher beschriebenen Einflussfaktoren dar. Der Schnitt beider Signale definiert die Position des Empfängers. Durch die Abweichungen liegt die wahre Position im roten Raum. Wenn die beiden Konstellationen verglichen werden, dann hat die rechte eine größere rote Fläche und kann somit eine höhere Ungenauigkeit aufweisen. Ideal ist es somit, wenn sich ein Satellit im Zenit und die anderen drei verteilt am Horizont befinden.

Wenn alle Fehler addiert werden, ergibt dies eine durchschnittliche Abweichung um 13,5 m. Dabei ist zu bedenken, dass die Einflussfaktoren nicht immer auftreten und sich ein Teil der Fehler durch Berechnungen beheben lassen. Dadurch liegt die Abweichung der Positionierung durchschnittlich unter 10 m.

### 2.3.2 Koordinatensystem

In dieser Arbeit geht es vor allem um das Referenzsystem World Geodetic System 1984 (WGS 84) für eine einheitliche Positionsangabe. Es beruht auf Polarkoordinaten, die als Koordinatensystem für die Positionierung auf der Erde dienen [3, 17]. Ein Punkt im Raum wird von zwei Winkeln und der Distanz zu ihrem Koordinatenursprung beschrieben. Diese Bezeichnung vereinfacht die Darstellung eines Punktes auf einer Kugeloberfläche, die die Erdoberfläche repräsentiert. Die Distanz von der Erdoberfläche zum Koordinatenursprung, dem mittleren Erdradius, ist auf 6.378.137 m (nach WGS 84) festgelegt. Da der Koordinatenursprung konstant ist, muss er nicht bei einer Positionierung angegeben werden. Daher beschreiben nur die beiden Winkel die Lage des Punktes auf der Erde. Die Winkel bestehen aus der geographischen Breite (Latitude) und der geographischen Länge (Longitude). Im späteren Verlauf der Arbeit werden die englischen Begriffe Latitude und Longitude verwendet. Eine weitere Festlegung ist die Winkelangabe in dezimal und nicht in sexagesimal<sup>1</sup>.

## 2.4 Trigonometrie

Dieses Kapitel befasst sich mit der Berechnung von geometrischen Koordinaten. Es wird in zwei Teilgebiete gegliedert. Die ebene Trigonometrie beschreibt die Berechnung von einem Punkt oder einer Punktmenge in einer Ebene. Die Formeln für die Bestimmung des Schnittpunktes zweier Geraden und die Ermittlung, ob sich ein Punkt auf einer Strecke befindet, werden erläutert.

Die sphärische Trigonometrie befasst sich mit der Berechnung von einem Punkt oder einer Punktmenge auf einer Kugel. Hier wird die Berechnung der Distanz zwischen zwei Punkten, der Richtung und der Zielkoordinate anhand einer Startkoordinate, einer Distanz und einer Richtung erklärt.

<sup>1</sup>Das Sexagesimalsystem ist ein Stellenwertsystem mit der Basis 60. Die Maßeinheit folgt in Grad °, Minuten ' und Sekunden ''.

## 2.4.1 Ebene Trigonometrie

Diese Trigonometrie befasst sich mit ebenen Flächen. Die beiden folgenden Formeln werden genutzt, um festzustellen, ob sich zwei Strecken schneiden. Jede Strecke besteht aus zwei Punkten mit je einem Koordinatenpaar. Zuerst werden die Strecken als Gerade aufgefasst und deren Schnittpunkt ermittelt. Danach wird bestimmt, ob sich der Schnittpunkt auf den beiden Strecken befindet, ob sich die zwei Strecken also schneiden. Diese Berechnung wird in Kapitel 4.4.2.1 gebraucht. Dort wird festgestellt, ob ein Strahl durch ein Gebäude oder eine Straße aus einem Umgebungsmodell verläuft.

### 2.4.1.1 Berechnung des Geradenschnittes

Eine Strecke besteht aus zwei Koordinatenpaaren. Für das erste Paar gilt  $p_1 = (x_1, y_1)$  und  $p_2 = (x_2, y_2)$  und für das zweite Paar  $p_1 = (x_3, y_3)$  und  $p_2 = (x_4, y_4)$ . Jedes Paar besteht aus einem Punkt, der anhand einer Latitude  $lat_i$  und einer Longitude  $lon_i$  im Raum definiert wird. Anhand dieser Daten wird der Schnittpunkt der beiden Geraden berechnet. Zuerst wird die Variable  $z$  berechnet:

$$z = \frac{(y_4 - y_3)(x_1 - x_2) - (y_1 - y_3)(x_4 - x_2)}{(y_2 - y_1)(x_4 - x_3) - (y_4 - y_3)(x_2 - x_1)} \quad (2.5)$$

Wenn  $z$  gleich null ist, dann sind die Geraden parallel zu einander und es entsteht kein Schnittpunkt. Bei allen anderen Werten von  $z$  entsteht ein Schnittpunkt  $p_{Schnitt} = (x_{Schnitt}, y_{Schnitt})$ , der wie folgt berechnet wird:

$$\begin{aligned} x_{Schnitt} &= x_1 + z \cdot (x_2 - x_1) \\ y_{Schnitt} &= y_1 + z \cdot (y_2 - y_1) \end{aligned} \quad (2.6)$$

### 2.4.1.2 Berechnung der Lage eines Punktes

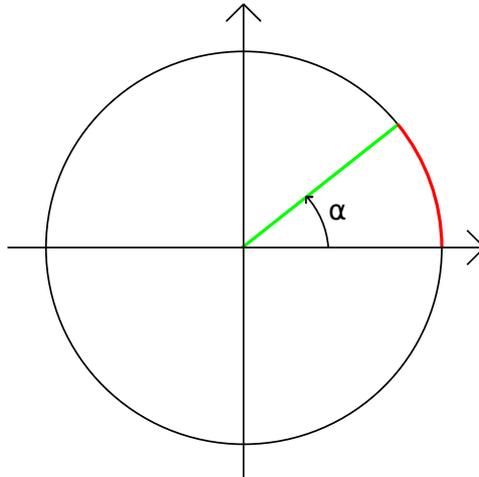
Hier wird beschrieben, ob sich ein Punkt auf einer Strecke befindet. Der Punkt  $p = (lat, lon)$  ist anhand einer Latitude und einer Longitude und die Strecke anhand zweier Punkte definiert. Es wird die Distanz  $dist$  zwischen den beiden Punkten  $p_1$  und  $p_2$  mit Hilfe der Haversine-Formel (siehe Kapitel 2.9) berechnet. Als Alternative kann auch der euklidische Abstand für die Distanzberechnung verwendet werden. Dann wird die Distanz  $dist_1$  zwischen dem ersten Punkt der Strecke  $p_1$  und dem Punkt  $p$  sowie die Distanz zwischen dem zweiten Punkt  $p_2$  der Strecke und dem Punkt  $p$  berechnet. Wenn die folgende Gleichung erfüllt wird, dann liegt der Punkt auf der Strecke:

$$dist \leq dist_1 + dist_2 \quad (2.7)$$

Wenn diese Ungleichung nicht erfüllt wird, liegt der Punkt folglich außerhalb der Strecke.

## 2.4.2 Sphärische Trigonometrie

Im vorherigen Kapitel wurde bezüglich des Koordinatensystems WGS 84 beschrieben, wie ein Punkt auf der Erdoberfläche dargestellt wird. Mit diesen Koordinaten




---

**Abbildung 2.7** Darstellung des Winkels  $\alpha$  und des Bogenmaßes in Rot
 

---

Latitude und Longitude nach WGS 84 lassen sich Berechnungen mit Formeln dieser Trigonometrie umsetzen. Die Koordinaten werden in der Einheit Grad ( $^{\circ}$ ) geschrieben. Diese Winkelangabe ist bei der Berechnung nicht hilfreich, da bei einem veränderten Radius bei gleichem Winkel die Bogenlänge nicht gleich bleibt. Die Abbildung 2.7 zeigt den Unterschied des Winkels  $\alpha$  und des Bogenmaßes, das mit roter Linie dargestellt ist. Daher werden die Winkelmaße der Koordinaten in die Einheit Radiant (rad) umgerechnet [?]. 1 rad ist ein Meter Bogenlänge bei einem Kreisradius von einem Meter. Ein Vollwinkel von  $360^{\circ}$  entspricht  $2\pi$  rad. Daraus lässt sich die Umrechnung bestimmen:

$$\text{Grad} = \text{Radiant} \cdot \frac{\pi}{180} \tag{2.8}$$

$$\text{Radiant} = \text{Grad} \cdot \frac{180}{\pi}$$

Eine Distanz in Metern wird auch in Radiant umgewandelt. Dieser Radiant-Wert ergibt sich aus der Distanz geteilt durch den Erdradius. Ein weiterer Punkt, der zu beachten ist, ist, dass eine Linie durch den kürzesten Punkt führt. Auf einer Karte wäre dies durch eine Gerade mit Hilfe eines Lineals einzuzeichnen. Diese Strecke ist eine Loxodrome. Die Karte ist eine flache Darstellung der Erde im Gegensatz zum Ellipsoid. Eine Loxodrome ist auf diesem Ellipsoid nicht der kürzeste Weg zwischen zwei Punkten mit Ausnahme des Äquators. Die Orthodrome definiert den kürzesten Weg entlang des Großkreises, auf dem sich die beiden Punkte befinden. Ein Großkreis ist der Kreis entlang der Erdoberfläche. Dabei entspricht der Mittelpunkt des Kreises dem Mittelpunkt der Erde. Somit teilt dieser Großkreis die Erde in zwei gleich große Hälften. Ein Großkreis, der durch zwei Punkte führt, bildet den kürzesten Weg auf dem Ellipsoid. Die unten stehende Berechnung bezieht sich auf die Loxodrome.

### 2.4.2.1 Berechnung der Distanz zwischen zwei Punkten

Die Haversine-Formel [22] dient der Berechnung von Distanz und Richtung anhand zweier Punkte mit je einer Latitude und einer Longitude auf der Erdoberfläche. Diese Formel existiert in mehreren Varianten. Eine Variante berechnet ein genaueres

Resultat bei längeren Distanzen, da die Erde als Ellipsoid als Referenz im WGS 84 gebraucht wird. Der Nachteil tritt bei kürzeren Distanzen auf, bei denen die Erdkrümmung so klein ist, dass diese sehr langen Dezimalstellen im Computer auf null abgerundet werden. Dadurch entsteht eine Division durch null und dies ist somit nicht berechenbar.

Die andere Variante ist eine schnellere Berechnung für kürzere Distanzen, welche die Kugelform der Erde als Referenz nutzt. Bei größeren Distanzen entstehen Abweichungen wegen der ungenauen Referenz der Erde [22]. In dieser Arbeit geht es meist um kürzere Distanzen von einem Meter bis 10 km. Daher beschränken wir uns auf die zweite Variante. Als Eingabe dienen zwei Koordinatenpaare mit je Latitude und Longitude. Als Formulierung für die Formel werden die erste Koordinate ( $lat1, lon1$ ) und die zweite Koordinate ( $lat2, lon2$ ) verwendet. Die Ausgabe ist die Distanz zwischen den beiden Punkten. Die Einheit der Ausgabe ist abhängig von der Einheit des Erdradius  $r$ . So erhält man die Ausgabe in Metern, wenn der Erdradius  $r$  auch in Metern ist. Zur einfacheren Übersicht der Formel werden  $dLat = lat2 - lat1$  und  $dLon = lon2 - lon1$  definiert. Es ist zu beachten, dass alle unten stehenden Variablen der Formeln 2.9 in die Einheit Radiant umgerechnet worden sind:

$$x = \left( \sin \left( \frac{dlat}{2} \right) \right)^2 + \cos(lat) \cdot \cos(lat2) \cdot \left( \sin \left( \frac{dlon}{2} \right) \right)^2 \quad (2.9)$$

$$dist = 2 \cdot \operatorname{atan} \left( \frac{\sqrt{x}}{\sqrt{1-x}} \right) \cdot \operatorname{radiusEarth}$$

Die Variable  $x$  dient einer übersichtlicheren Formel und die Variable  $dist$  präsentiert das Ergebnis der Berechnung.

#### 2.4.2.2 Berechnung der Richtung von zwei Punkten

Diese Formel berechnet die Richtung einer Linie, in die sie zeigt [?]. Die Richtung wird für die Berechnung der Zielkoordinate, in Kapitel 2.4.2.3 erläutert, benötigt. Die Eingabe ist die gleiche wie bei der Distanzberechnung, jedoch sollen die erste und zweite Koordinate nicht vertauscht werden, sonst wird das Ergebnis invertiert. Die erste Koordinate ist der Startpunkt und die zweite Koordinate ist der Punkt, der die Richtung ausgehend vom Startpunkt weist. Die Ausgabe *bearing* der Formel gibt die Himmelsrichtung zwischen  $0^\circ$  und  $360^\circ$  an. Es ist zu beachten, dass die Werte der Eingabe in Radiant angegeben werden:

$$bearing = \operatorname{atan} \left( \frac{\sin(lon2 - lon1) \cdot \cos(lat2)}{\cos(lat1) \cdot \sin(lat2) - \sin(lat1) \cdot \cos(lat2) \cdot \cos(lon2 - lon1)} \right) \quad (2.10)$$

Die Ausgabe *bearing* kann von Radiant in Grad umgewandelt werden.

#### 2.4.2.3 Berechnung der Zielkoordinate

Diese Formel berechnet die Zielkoordinate anhand einer Startkoordinate, Richtung und Distanz [?]. Von dieser Startkoordinate wird die Entfernung ermittelt. Sie ergibt einen Kreis mit der Startkoordinate als Mittelpunkt und dem Radius der entsprechenden Entfernung. Unter Einbeziehung der Richtung vom Startpunkt erhält man

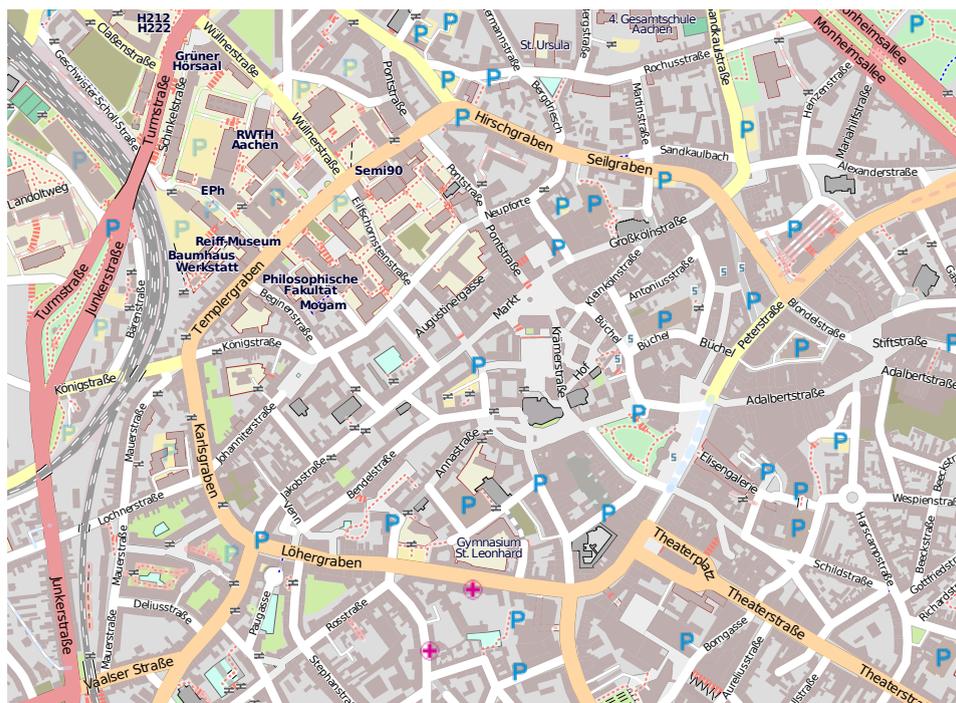


Abbildung 2.8 Karte von Aachen - Daten von OpenStreetMap, 2012, veröffentlicht unter CC-BY-SA 2.0

einen eindeutigen Punkt auf dem Kreis, der die Zielcoordinate darstellt. In der Formel ist  $d$  die Distanz in Metern und  $brng$  die Richtung in Grad angegeben. Die Eingabe erfolgt in Radiant:

$$\begin{aligned} newLat &= asin(\sin(lat) \cdot \cos(d) + \cos(lat) \cdot \sin(d) \cdot \cos(brng)) \\ newLon &= lon + atan\left(\frac{\sin(brng) \cdot \sin(d) \cdot \cos(lat)}{\cos(d) - \sin(lat) \cdot \sin(newLat)}\right) \end{aligned} \quad (2.11)$$

Die Ausgaben  $newLat$  und  $newLon$  können wieder in Grad umgerechnet werden.

## 2.5 OpenStreetMap

OpenStreetMap<sup>2</sup> (OSM) ist ein Projekt, welches sich zum Ziel gesetzt hat, freie Karten zur Verfügung zu stellen [17]. Gerade der Wert Frei ist der entscheidende Kernpunkt dieses Projektes. So bietet Google Maps auch kostenlose Karten an, jedoch ist der Entwickler an deren Lizenzierung gebunden. Bei einer Lizenzveränderung muss er diese akzeptieren. So muss er u. U. für einen Dienst zahlen oder auf einen anderen Datensatz wechseln.

Das Projekt OSM folgt dem Wikipedia-Gedanken, der dem Benutzer die Möglichkeit bietet, die Karten zu vervollständigen und zu verbessern. Es wird von der OpenStreetMap Foundation geführt und läuft unter der Creative Commons Attribution Share Alike-Lizenz 2.0<sup>3</sup> (CC-BY-SA). Die Daten von OSM liegen geobasierend auf

<sup>2</sup>OpenStreetMap, <http://www.openstreetmap.org/>

<sup>3</sup>Diese Lizenz erlaubt dem Anwender, die Daten frei zu verteilen und zu verändern unter der Berücksichtigung, den Autor zu nennen. Siehe <http://creativecommons.org/licenses/by-sa/2.0/>

einer Postgre-SQL-Datenbank. Die Geodaten sind in ein Dateiformat übertragen, das die Umgebung durch Punkte mit Latitude und Longitude nach WGS 84 beschreibt. Anhand dieser Daten werden die rasterbasierten Karten erstellt, die den meisten Anwendern bekannt sind, siehe Abbildung 2.8. Die Rasterdaten sind pixelbasierte Bilddateien. Neben den Rasterdaten können auch Vektordaten aus den Geodaten erstellt werden. Die Vektordaten benutzen ein zweidimensionales Koordinatensystem. Google Maps bietet ihr Kartenmaterial nur als pixelbasierte Daten mit einem Overlay an, bei dem das Pixel einer Koordinate zugeschrieben wird.

Bei weiteren Datenverarbeitungen in der vorliegenden Arbeit interessieren uns die geobasierenden Daten. Diese Daten werden als OSM XML in der Kodierung UTF-8 exportiert. Bei der Extensible Markup Language (XML) handelt es sich um eine Metasprache, mit der man eine eigene menschenlesbare Sprache definieren kann. Deren Ziel ist es, einen Informationsaustausch zwischen verschiedenen Systemen zu ermöglichen. Die Daten sind strukturell und hierarchisch gespeichert [?]. OSM hat mit Hilfe von XML ihre eigene Sprache OSM XML definiert. Sie basiert auf den drei Elementen Node, Way und Tag sowie deren Beziehung Relation zueinander.

**Node:** Ein Node ist ein geografischer Punkt, der durch die geografische Länge und Breite nach WGS 84 definiert ist. Er bekommt eine eindeutige Identifikationsnummer und kann mit einem oder mehreren Tags einer Bedeutung zugeordnet werden.

**Way:** Der Way ist eine geordnete Liste, die aus mindestens zwei Nodes besteht. Größere Ways mit über 2.000 Nodes werden in kleinere Ways aufgeteilt. Die Nodes beschreiben eine Linie, die offen oder geschlossen ist. Eine Linie ist geschlossen, wenn der Anfangspunkt und der Endpunkt einer Linie der gleiche Node ist. Bei einer offenen Linie sind der Anfangspunkt und der Endpunkt verschieden. Eine Area ist eine geschlossene Linie, die die Fläche des Polygons beschreibt. Um die Fläche von der Linie zu unterscheiden, wird bei der geschlossenen Linie das Tag *area* hinzugefügt.

**Tag:** Das Tag ist eine Kennzeichnung eines Nodes, Ways oder einer Relation und soll dem Element eine Bedeutung geben. Das Tag besteht aus einem Schlüssel und einem Wert. Ein Beispiel ist das Tag *building* mit dem Wert *yes*. Es beschreibt eine Fläche als ein Gebäude.

Eine Relation ist eine Liste von Nodes, Ways oder weiteren Relationen. Diese Liste ist geordnet und kann Tags enthalten. Die Relation beschreibt eine logische oder geografische Beziehung zwischen den Elementen der Liste. Damit sind komplexere Multipolygone möglich, deren Polygone zu einer Ordnung zusammengefasst sind. Ein Beispiel ist die Erstellung einer Exklave in einem Land, z. B. der Insel Sylt, die zum deutschen Festland gehört.

Neben den oben erklärten drei Elementen sowie Relation existiert noch das Changeset. Es beschreibt eine Menge von Veränderungen des Datensatzes als eine atomare Einheit. Damit soll verfolgbar sein, wer, wann und wo welchen Datensatz verändert hat. Das Changeset enthält noch weitere Tags, wie z. B. *comment*, das die Veränderung der Daten kommentieren kann. Anhand dieser vier Elemente Tag, Node, Way und Relation sind alle Objekte der Umwelt, wie z. B. die Straßen, Häuser, Zugstrecken usw., erstellt. Die Tags werden laufend erweitert, sodass die Nodes und Ways

einer neuen Bedeutung zugeordnet werden. Ein Beispiel ist das Tag *entrance*, das die Haustüre eines Gebäudes markiert. Durch die ständige Weiterentwicklung übertrifft OSM die Möglichkeiten einer gedruckten Karte und bietet eine detailliertere digitale Repräsentation der Umwelt.

## 2.6 Messabweichung

Das Ziel von Messungen ist eine genaue Aussage über eine unbekannte Größe. Leider tritt eine Abweichung bei den Messwerten auf, sodass sie nicht mit den wahren Werten übereintreffen [?]. Viele Messgeräte bieten keine genauen Messungen und sind zusätzlich von der Bedienung des Anwenders abhängig. So ist vor den Messungen die korrekte Handhabung zu beachten, während der Messungen sind Messfehler zu vermeiden, nach den Messungen ist die Größe des Fehlers zu bestimmen und, wenn möglich, der Messwert zu korrigieren. Dieses Kapitel erläutert die verschiedenen Arten von Messabweichungen und wie man sie möglichst minimieren kann. Zunächst wird die systematische und anschließend die zufällige Abweichung beschrieben.

### 2.6.1 Systematische Abweichung

Die systematischen Abweichungen sind Messfehler, die durch den Mittelwert einer Messreihe nicht beseitigt werden. Dies tritt z. B. bei falsch-justierten Messgeräten, falscher Bedienung der Geräte und Umwelteinflüssen auf: Über mehrere regnerische Tage werden z. B. RSSI-Werte eines APs gemessen. In dem Kapitel Einflussfaktoren unter 2.2.2 ist geschildert, dass der Regen eine Dämpfung der Signalausbreitung bewirkt. Diese Dämpfung führt zu niedrigeren RSSI-Werten als bei Messungen an einem trockenen Tage. So ist der Mittelwert auch niedriger und kann die Abweichung nicht korrigieren.

Systematische Abweichungen haben eine Tendenz in eine Richtung. Diese Tendenz lässt sich mit einer positiven oder negativen Konstanten darstellen. Diese Fehler sind schwer zu erkennen und lassen sich nur durch einen hohen Aufwand bei den Messungen und durch bessere Messtechniken verhindern. Eine Möglichkeit ist es, mit verschiedenen Messgeräten zu messen und die Messwerte zu vergleichen, um eine Abweichung zu überprüfen.

### 2.6.2 Zufällige Abweichung

Wenn der Mittelwert einer Messreihe die Abweichung korrigiert, dann handelt es sich um die zufällige Abweichung. Diese Fehler können durch ungenaue Messgeräte und wechselnde Bedingungen entstehen.

Ein Beispiel ist die Ungenauigkeit der Uhrzeit des GPS-Satelliten (siehe Kapitel 2.3.1). Durch diese Ungenauigkeit kann bei der Positionierung ein Fehler von 1 m auftreten. Bei mehreren Messungen reicht nicht immer die gleiche Fehlergröße auf. Dadurch entsteht eine Streuung der Messwerte um den wahren Wert. Der Mittelwert aus den gestreuten Messwerten ergibt etwa den wahren Wert. Je mehr Messwerte den Mittelwert bestimmen, desto genauer tritt der Mittelwert an den wahren Wert

heran.

Im Gegensatz zur systematischen Abweichung lässt sich diese Abweichung einfacher erkennen, jedoch nicht vollkommen korrigieren bzw. verhindern.

## 2.7 Filter

In dem Kapitel Einflussfaktoren des GPS (siehe 2.3.1) sind mehrere Faktoren beschrieben, die eine Auswirkung auf die Positionierung bewirken. Diese Faktoren sind meist nicht ersichtlich und können nicht direkt beobachtet werden. Daher bietet sich eine direkte Berechnung des Systems nicht an. Es müssen mehrere Messungen erstellt werden. Die Zustände sind verborgen und müssen aus der Wahrscheinlichkeitsdichte mehrerer Messungen geschätzt werden [4].

Als erster Algorithmus, um die Messabweichungen zu verkleinern, wurde das  $\alpha$ - $\beta$ - $\gamma$ -Filter [9] eingesetzt. Das  $\alpha$ -Filter ist eine eindimensionale, das  $\alpha$ - $\beta$ -Filter eine zweidimensionale und das  $\alpha$ - $\beta$ - $\gamma$ -Filter eine dreidimensionale Schätzung des wahren Wertes. Dieses lässt sich leicht analog implementieren und hat schnelle Laufzeiten. Der Schätzer basiert auf der mittleren quadratischen Abweichung.  $\alpha$ ,  $\beta$  und  $\gamma$  sind a priori Konstanten, die die Messungen glätten sollen. Die Wahl dieser Konstanten bestimmt die Güte der Ergebnisse. Die Resultate dieser Filter sind in manchen Bereichen ausreichend, jedoch sind sie meist ungenauer als das Kalman-Filter [9].

Heutzutage ist die meist genutzte Methode, um Messabweichungen von GPS-Daten zu korrigieren, das Kalman-Filter [21]. Es berechnet bei jeder Iteration einen Schätzwert, wo die Messungen gemacht wurden. Diese Berechnung basiert auf dem aktuellen GPS-Signal, dem Schätzwert aus der vorherigen Iteration und dem Kalman Gain. Der Gain gewichtet Messwerte mit hoher Varianz schwach und Messwerte mit geringer Varianz stark. Somit haben die stark gewichteten Messwerte eine höhere Beeinflussung, um den Schätzwert zu ermitteln. Wenn die Messwerte eine schwache Gewichtung aufweisen, so hat der Schätzwert aus der vorherigen Iteration eine höhere Beeinflussung auf den Schätzer. Die Schwierigkeit ist die Initialisierung. In der ersten Iteration wird der Schätzwert gleich dem Messwert gesetzt. Es braucht mehrere Iterationen, bis der Schätzwert gegen den wahren Wert konvergiert. Dieses Filter ist auf ein lineares Sensor- und Zustandsmodell angewiesen. Daher muss diese Wahrscheinlichkeitsdichte der Modelle gauß-verteilt sein.

Für nicht lineare Systeme wurde das Extended Kalman-Filter [?] entwickelt. Es handelt sich um eine Erweiterung des Kalman-Filters. Dabei werden die nicht linearen Modelle linear approximiert.

Eine andere Erweiterung ist das Unscented Kalman-Filter [?]. Dabei wird dem Schätzer eine statistische Näherung der nicht linearen Modelle hinzugefügt. Um die Nicht-linearität zu lösen, gibt es das gaußsche Summenfilter [?]. Dieses Filter basiert darauf, dass die Dichtefunktion der nicht linearen Modelle durch eine Summe von verschiedenen Gauß-Verteilungen dargestellt wird. Diese drei Ansätze führen meist zu schlechten Resultaten bei hoch dimensionalen Daten [4]. Das gitterbasierende Filter bietet genauere Ergebnisse. Jedoch ist dieses schwer zu implementieren und die Laufzeiten sind für die meisten praktischen Anwendungen zu lang [4].

Als Lösung wird das Partikelfilter [4] genutzt, das ausführlicher in dem nächsten Kapitel 2.7.1 erläutert wird. Das Zustand- und Messmodell kann nicht linear sein. Im Vergleich zu den vorherigen Filtern wird kein deterministischer Ansatz verwen-

det, sondern ein randomisierter Algorithmus. Das Prinzip besteht darin, dass der Schätzwert anhand eines Schwarmes von Partikeln ermittelt wird, die durch das Zustandsmodell vom vorherigen Zustand in den nächsten Zustand beschrieben werden. Das Messmodell gibt den Partikeln eine Gewichtung, wie gut diese mit der gemessenen Position übereinstimmen. Anhand eines Bootstrap-Filters werden schwach gewichtete Partikel gelöscht und stark gewichtete Partikel geklont. So bietet sich eine bessere Repräsentation der Dichtefunktion durch die Umverteilung der Partikel. Die Beschreibung des Bootstrap-Filters befindet sich in Kapitel 2.7.3.3.

### 2.7.1 Partikelfilter

Das Partikelfilter [16] gehört zu der Klasse der sequentiellen Monte-Carlo-Methode. Eine endliche Anzahl von Zuständen mit einer dazu gehörenden Gewichtung beschreibt die Wahrscheinlichkeitsdichtefunktion. Über einen Zeitraum verändert sich die Wahrscheinlichkeitsdichte und somit auch die an die Zustände angepasste Gewichtung. Diese Gewichtung dient als Beschreibung der Wahrscheinlichkeitsdichte, die unbekannt ist. Die Zustände sind eine Stichprobe dieser Dichte. Ein Partikel repräsentiert einen Zustand. Die endliche Partikelmenge  $P$  beinhaltet  $n$  Anzahl von Partikel  $x_i$  sowie dessen Gewichtung  $w_i$ :

$$P = \{(x_i, w_i) | i = 1, \dots, N\} \quad (2.12)$$

Bei der ersten Iteration werden die Partikel aus einer zufälligen Stichprobe festgelegt. Die Stärke dieses Filter ist, dass es sich auf nicht lineare und nicht gaußische Verteilungen anwenden lässt.

### 2.7.2 Problembeschreibung

Das Zustandsmodell beschreibt den Zustandsübergang von dem gegenwärtigen Zustand in den nächsten Zustand [16]. Dabei werden die Informationen von der begrenzten Anzahl von vorherigen Zuständen als Übergangswahrscheinlichkeit zur Entscheidung in den nächsten Zustand genommen. Der Zustand  $x_t$  gehört zu der Zustandsmenge  $\chi$  zu dem Zeitpunkt  $t \in \mathbb{N}_0$ :

$$\begin{aligned} P(x_0) \\ P(x_t | x_{t-1}) \text{ für } t \geq 1 \end{aligned} \quad (2.13)$$

Für den ersten Zustand  $x_0$  wird eine initialisierte Wahrscheinlichkeit von  $P(x_0)$  festgelegt. Die Wahrscheinlichkeit der Initialisierung und der Verteilung aller nachfolgenden Zustände beschreibt das Zustandsmodell. Der aktuelle Zustand  $x_t$  zu dem Zeitpunkt  $t$  kann nicht direkt beobachtet werden und muss aus einer Messung  $y_t$  mit einer Abweichung geschätzt werden. Die Messung  $y_t$  entsteht aus der Messungsmenge  $\mathcal{Y}$ :

$$P(y_t | x_t) \text{ für } t \geq 1 \quad (2.14)$$

Diese Verteilung stellt das Sensormodell dar. Als nächstes wird der Erwartungswert  $\mathbb{E}_{p(x_{0:t}|y_{1:t})}$  für die Funktion  $f_t : \chi^{t+1} \rightarrow \mathbb{R}^{n_{f_t}}$  bestimmt:

$$I(f_t) = \mathbb{E}_{p(x_{0:t}|y_{1:t})}(f_t(x_{0:t})) = \int f_t(x_{0:t}) p(x_{0:t}|y_{1:t}) dx_{0:t} \quad (2.15)$$

Diese Berechnung dient als eine Abschätzung, die rekursiv für jeden Zeitpunkt  $t$  ist. Die beiden Verteilungen  $P(x_0|y_{1:t})$  und  $P(x_{0:t}|y_{1:t})$  sowie der Erwartungswert  $I(f_t)$  werden abgeschätzt. Bei GPS-Daten wird für den Erwartungswert der wahrscheinlichste Ort der Messung samt Messabweichung genommen. Daraus ergäbe sich für  $f_t(x_{0:t}) = x_t$ . Die bedingte Varianz berechnet sich aus  $f_t(x_{0:t}) = x_t x_t^T - I(x_t) I^T(x_t)$ . Die Varianz gibt den Raum um den Wert  $x_t$  an, wo sich der wahre Wert bzw. die Position befindet. Die Posteriori-Verteilung  $p(x_{0:t}|y_{1:t})$  lässt sich durch das Bayes Theorem rekursiv zu dem Zeitpunkt  $t$  berechnen. Aus diesem Theorem entsteht die Formel:

$$p(x_{0:t}|y_{1:t}) = \frac{p(y_{1:t}|x_{0:t})p(x_{0:t})}{\int p(y_{1:t}|x_{0:t})p(x_{0:t})dx_{0:t}} \quad (2.16)$$

Damit lässt sich jede Verteilung zu jedem Zeitpunkt berechnen. Der nächste Schritt ist die rekursive Berechnung über die Zeit. Diese ergibt sich aus der Formel:

$$p(x_{0:t+1}|y_{1:t+1}) = p(x_{0:t}|y_{1:t}) \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{p(y_{t+1}|y_{1:t})} \quad (2.17)$$

Der Beweis zu dieser Formel befindet sich in Quelle [15]. Die Randverteilung  $p(x_t|y_{1:t})$  wird durch die beiden rekursiven Formeln *Prediction* und *Updating* ermittelt. Mit diesen Formeln können wir die Verteilungen  $P(x_0|y_{1:t})$  und  $P(x_{0:t}|y_{1:t})$  und den Erwartungswert  $I(f_t)$  abschätzen. Das Problem liegt bei der Berechnung von  $p(y_{1:t})$ , die durch ein komplexes Integral über viele Dimensionen bestimmt wird. Dies kann nicht berechnet werden. Daher folgt der Versuch, anhand der Monte-Carlo-Methode eine Approximation durchzuführen.

## 2.7.3 Monte-Carlo-Methode

Die Monte-Carlo-Methode [4, 15] dient dazu, nicht analytische Probleme mit Hilfe von Zufallszahlen zu lösen. Dabei sind die Ergebnisse approximiert. Ein Ergebnis kann eine größere Abweichung vom wahren Wert haben. Der Mittelwert von den Ergebnissen mehrerer Durchläufe bietet meist einen genaueren Wert. Es ist selten möglich, eine Stichprobe direkt aus einer Wahrscheinlichkeitsverteilung zu ziehen. Hier kommt das Importance Sampling zum Einsatz.

### 2.7.3.1 Importance Sampling

Die Importance-Funktion beschreibt eine Verteilung  $\pi(x_{0:t}|y_{1:t})$ , die anstelle der eigentlichen Wahrscheinlichkeitsverteilung  $P(x_{0:t}|y_{1:t})$  genutzt wird, da aus dieser nicht direkt Stichproben entnommen werden können. Die Menge der Punkte der Verteilung  $P(x_{0:t}|y_{1:t})$ , welche keine Null als Punkt enthält, ist eine Teilmenge der Menge der Punkte der Importance-Funktion, die auch keine Null als Punkt enthält. Daraus ergibt sich der Erwartungswert:

$$I(f_t) = \frac{\int f_t(x_{0:t})w(x_{0:t})\pi(x_{0:t}|y_{1:t})dx_{0:t}}{\int w(x_{0:t})\pi(x_{0:t}|y_{1:t})dx_{0:t}} \quad (2.18)$$

Der Erwartungswert enthält die Importance-Gewichtung  $w(x_{0:t})$ , die wie folgt berechnet wird:

$$w(x_{0:t}) = \frac{p(x_{0:t}|y_{1:t})}{\pi(x_{0:t}|y_{1:t})} \quad (2.19)$$

Aus der Verteilung  $\pi(x_{0:t}|y_{1:t})$  wählen wir eine endliche Anzahl  $N$  von unabhängigen Partikeln  $\{x_{0:t}^i, i = 1, \dots, N\}$ . Daraus lässt sich eine Abschätzung des Erwartungswertes  $I_N(f_t)$  ermitteln:

$$I_N(f_t) = \sum_{i=1}^N f_t(x_{0:t}^i) \bar{w}_t^i \quad (2.20)$$

Dabei ist die Gewichtung  $\bar{w}$  im Vergleich zur vorherigen Gewichtung  $w$  normalisiert und folgt aus:

$$\bar{w}_t^i = \frac{w(x_{0:t}^i)}{\sum_{j=1}^N w(x_{0:t}^j)} \quad (2.21)$$

Das aktuelle Gewicht  $w_t^i$  wird durch die Summe aller Gewichte geteilt. Der nächste Schritt zeigt, dass der Erwartungswert  $I_N(f_t)$  gegen den Erwartungswert  $I(f_t)$  konvergiert, wenn der Erwartungswert  $\mathbb{E}(\bar{w}_t)$  und  $\mathbb{E}(\bar{w}_t^2)$  kleiner als unendlich ist. Somit folgt der zentrale Grenzwertsatz von Lindeberg-Levy. Die Sampling-Funktion wird aus dieser Integrationsmethode festgelegt und bildet die Approximation

$$P_N(dx_{0:t}|y_{1:t}) = \sum_{i=1}^N \bar{w}_t^i \delta_{x_{0:t}^i}(dx_{0:t}) \quad (2.22)$$

für die Verteilung  $p(x_{0:t}|y_{1:t})$ . Daraus entsteht der approximierte Erwartungswert  $I_N(f_t)$  mit Hilfe der Sampling-Funktion und Integrationsmethode:

$$I_N(f_t) = \int f_t(x_{0:t}) P_n(dx_{0:t}|y_{1:t}) \quad (2.23)$$

Diese Methode benötigt alle Messungen  $y_{1:t}$  zu dem Zeitpunkt  $t$ , um die Berechnung der Verteilung  $P(x_{0:t}|y_{1:t})$  durchzuführen. Im nächsten Zeitpunkt muss die Berechnung mit den vergangenen und aktuellen Messungen neu erstellt werden. Somit steigt die Laufzeit dieses Algorithmus um jede Zeiteinheit. Um dies zu vermeiden, brauchen wir einen rekursiven Ansatz wie das Sequential Importance Sampling.

### 2.7.3.2 Sequential Importance Sampling

Die rekursive Variante des Importance Samplings ermöglicht eine schnellere Berechnung als die nicht rekursive Version. Das Grundprinzip des Importance Samplings wird übernommen und so angepasst, dass die Schätzung  $P_N(dx_{0:t}|y_{1:t})$  mit den unveränderten vergangenen Partikeln  $\{x_{0:t-1}^i : i = 1, \dots, N\}$  berechnet werden kann. Mit der Importance-Funktion zur Zeit  $t - 1$  und der Randverteilung wird es ermöglicht, die Importance-Funktion zur Zeit  $t$  zu bestimmen. Daraus ergibt sich die folgende Funktion:

$$\pi(x_{0:t}|y_{1:t}) = \pi(x_{0:t-1}|y_{1:t-1}) \pi(x_t|x_{0:t-1}, y_{1:t}) \quad (2.24)$$

Die sukzessive Approximation nach der Initialisierung, also ab der ersten Iteration, entsteht:

$$\pi(x_{0:t}|y_{1:t}) = \pi(x_0) \prod_{k=1}^t \pi(x_k|x_{0:k-1}, y_{1:k}) \quad (2.25)$$

Die rekursive Variante für die normalisierte Gewichtung ist

$$\bar{w}_t^i \sim \bar{w}_{t-1}^i \frac{p(y_t|x_t^i) p(x_t^i|x_{t-1}^i)}{\pi(x_t^i|x_{0:t-1}^i, y_{1:t})} \quad (2.26)$$

Dabei ist die aktuelle Gewichtung  $\bar{w}_t^i$  proportional zur Berechnung der vorherigen Gewichtung. Die Importance-Funktion spielt eine wichtige Rolle und zeigt, wie genau die Ergebnisse sind. Das Problem ist, dass über mehrere Iterationen die Streuung der Partikel immer größer wird. Wegen der Varianz bewegt sich der Erwartungswert weg vom wahren Wert [10]. Dadurch wird die Gewichtung für die meisten Partikel immer geringer, bis sie gegen null konvergiert. Um die Varianz zu minimieren, wird für die Importance-Funktion  $\pi(x_t|x_{0:t-1}, y_{1:t}) = P(x_t|x_{0:t-1}, y_{1:t})$  festgelegt. Der Beweis folgt Quelle [15]. Daraus ergibt sich die Importance-Funktion  $\pi(x_{0:t}|y_{1:t})$  sowie die Gewichtung  $\bar{w}_t^i$ :

$$\pi(x_{0:t}|y_{1:t}) = p(x_{0:t}) = p(x_0) \prod_{k=1}^t p(x_k|x_{k-1}) \quad (2.27)$$

$$\bar{w}_t^i \sim \bar{w}_{t-1}^i p(y_t|x_t^i)$$

Damit wird die Streuung minimiert, aber nicht ganz behoben. Hierfür kommt das Bootstrap-Filter zum Einsatz.

### 2.7.3.3 Bootstrap-Filter

Das Problem beim Sequential Importance Sampling sind die Fortpflanzungsfehler. Kleinere Abweichungen in den ersten Iterationen werden in den nächsten Iterationen fortgeführt und um eine weitere Abweichung erhöht. Dies führt dazu, dass bei steigender Iteration die Fehlerrate mitsteigt. Die meisten Importance-Gewichte sinken auf fast null ab, während ein einziges Partikel die stärkste Gewichtung mit fast 100 % bekommt. Somit enthält ein einziges Partikel die Information und nicht der Schwarm von Partikeln. Das Bootstrap-Filter ist eine Erweiterung des Algorithmus um ein Resampling [4]. Ein Resampling erstellt neue Stichproben als Partikel und ersetzt diese durch andere. Als Entscheidungskriterium für die Ersetzung wird die Gewichtung eingesetzt. Schwach gewichtete Partikel werden entfernt und an die Position von stark gewichteten Partikeln gesetzt. Diesem Grundprinzip folgt das Bootstrap-Filter. Die empirische Gewichtung  $P_N(dx_{0:t}|y_{1:t})$  in der Formel 2.22 wird durch eine ungewichtete Einheit ausgetauscht:

$$\bar{P}_N(dx_{0:t}|y_{1:t}) = \frac{1}{N} \sum_{i=1}^N N_t^i \delta_{x_{0:t}^i}(dx_{0:t}) \quad (2.28)$$

Dabei ist  $N_t^i$  die Anzahl der Nachfolger vom Partikel  $x_k^i$ .  $N$  ist die Gesamtzahl aller Partikel und folgt aus  $N = \sum_{i=1}^N N_t^i$ . Bei  $N_t^i = 1$  bekommt das Partikel keinen Nachfolger und bei  $N_t^i = 0$  wird das Partikel gelöscht.  $N_t^i$  wird am besten so gewählt, dass die Verteilung  $\bar{P}_N(dx_{0:t}|y_{1:t})$  gegen die Verteilung  $P_N(dx_{0:t}|y_{1:t})$  approximiert:

$$\int f_t(x_{0:t}) \bar{P}_N(dx_{0:t}|y_{1:t}) \approx \int f_t(x_{0:t}) P_N(dx_{0:t}|y_{1:t}) \quad (2.29)$$

Dies hat zur Folge, dass die Partikel nach  $P(x_{0:t}|y_{1:t})$  verteilt werden. Der nächste Schritt zeigt, wie sich der Fehler zwischen der empirischen Gewichtung und der ungewichteten Einheit verhält. Für die beschränkte Funktion  $f_t$  in

$$\mathbb{E} \left[ \left( \int f_t(x_{0:t}) \bar{P}_N(dx_{0:t}|y_{1:t}) - \int f_t(x_{0:t}) P_N(dx_{0:t}|y_{1:t}) \right)^2 \right] \leq \frac{C \|f_t\|^2}{N} \quad (2.30)$$

folgt, dass der quadratische Fehler bei unendlich vielen Partikeln gegen null konvergiert. Der Beweis ist in Quelle [2] zu finden.

**Initialisierung:**  
Für alle  $i = 1, \dots, N$ :  
    Erstelle Partikel  $x_0^i$  aus der Verteilung  $P(x_0)$   
 $t = 1$

**Sequential Importance Sampling:**  
Für alle  $i = 1, \dots, N$ :  
    Verschiebe das Partikel  $\bar{x}_t^i$  nach  $P(x_t | x_{t-1}^i)$   
    Füge das Element  $\bar{x}_t^i$  der Menge  $\bar{x}_{0:t}^i$  hinzu  
    Ermittle das Importance-Gewicht  $\bar{w}_t^i = p(y_t | \bar{x}_t^i)$   
Normalisiere die Gewichte nach  $\bar{w}_t^i = \frac{\bar{w}_t^i}{\sum_{j=1}^N \bar{w}_t^j}$

**Resampling:**  
Für alle  $i=1, \dots, N$ :  
    Ziehe aus der Menge  $\bar{x}_{0:t}$  mit Zurücklegen eines Partikels  
    Füge das gezogene Partikel der Menge  $\bar{x}_{t+1}$  hinzu  
 $t = t + 1$   
Gehe zu Sequential Importance Sampling

---

**Abbildung 2.9** Algorithmus des Bootstrap-Filters

---

### Algorithmus

Die Abbildung 2.9 stellt den Algorithmus des Bootstrap-Filters [4] in Pseudocode dar. Er ist in drei Schritte eingeteilt. Die Initialisierung kümmert sich um die Generierung der ersten Partikelmenge. Das Sequential Importance Sampling verschiebt die Partikel und berechnet die Importance-Gewichte. Das Resampling löscht die schwachen Partikel heraus und fügt starke Partikel mehrmals ein. Die Stärken des Algorithmus sind seine Einfachheit und seine Schnelligkeit. Je nach Problem muss nur die Importance-Gewichtung angepasst werden. Ein weiterer Vorteil ist, dass der Algorithmus parallelisierbar ist.

Die Abbildung 2.10 zeigt eine schematische Darstellung des Algorithmus. In Schritt (1) befinden sich die Partikel als Ergebnis aus dem vorherigen Schritt  $t - 2$ . Diese Partikel  $\bar{x}_{t-1}^i$  haben die ungewichtete Einheit  $\frac{1}{N}$ . Es handelt sich um eine Annäherung der Verteilung  $P(x_{t-1} | P(x_{t-2}))$ . Im nächsten Schritt (2) werden die Partikel mit  $\bar{w}_{t-1}^i$  gewichtet. Daraus folgt in Schritt (3) die Verteilung der Partikel  $x_t^i$ , also das Resampling, und dies geht dann in Schritt (4) in den nächsten Zeitpunkt  $t$  über. Dort fängt es von Neuem an und geht über in die Gewichtung, die in Schritt (5) dargestellt ist.

## 2.8 Datenschutz

In den letzten Jahren findet das Thema Datenschutz eine immer größere Aufmerksamkeit. Der Datenmissbrauch ist das nicht gestattete Aneignen und die unerlaubte Weitergabe von personenbezogenen Daten, der bei der maschinellen Datenverarbeitung häufiger und schneller vonstatten geht als bei der manuellen Datenverarbeitung mit einer Kartei. Aus diesen Dimensionen heraus bildet sich ein neues Verständnis

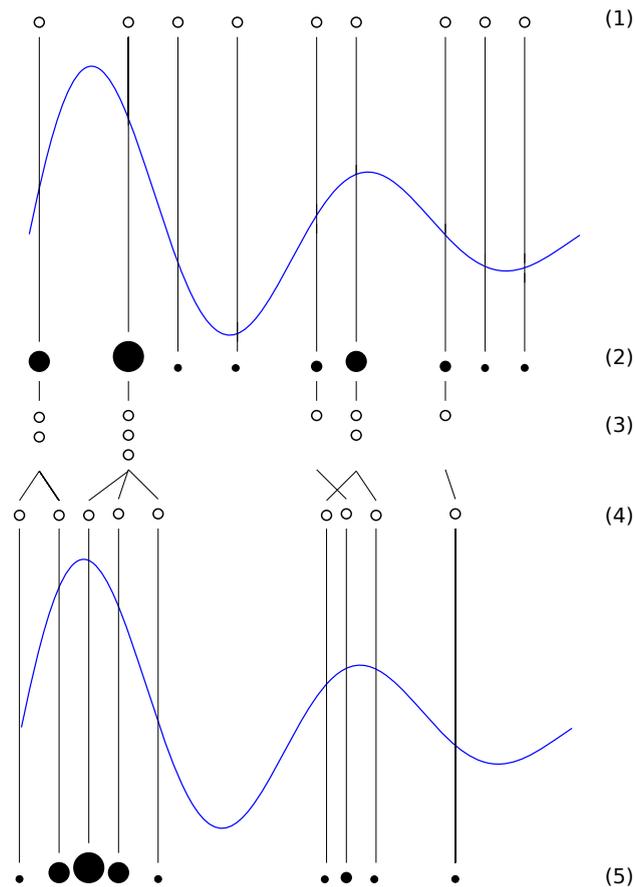


Abbildung 2.10 Darstellung des Bootstrap-Filters

für Daten. Mit Datensicherheit ist der Schutz von Daten gemeint, die von Dritten entwendet und verwendet wurden, ohne dass Kunde und Dienstleister sich darauf geeinigt haben. Es werden Datenschutzrechte benötigt. Der Datenschutz [28] beinhaltet die Einhaltung des Persönlichkeitsrechts und folgt Definition 1.

**Definition 1** (Datenschutz). „Schutz des Einzelnen vor Beeinträchtigung seines Persönlichkeitsrechts beim Umgang mit seinen personenbezogenen Daten“ nach §1 Abs. 1 des Bundesdatenschutzgesetzes [28]

Alle Daten, die einen Bezug zu einer natürlichen Person aufweisen, müssen das Persönlichkeitsrecht sowie die Datensicherheit einhalten. Anonymisierte Daten dagegen haben keinen Bezug zu einer Person und fallen somit nicht unter das Persönlichkeitsrecht und die Datensicherheit.

Ein anderer wichtiger Punkt in der vorliegenden Arbeit ist die Selbstbestimmung bei der Aufzeichnung und Verwendung der Daten. Es liegt in der Gewalt der Personen, wann, wo und welche Daten aufgenommen werden. Sie sind darüber informiert, welche Daten gespeichert werden, und entscheiden, ob die aufgezeichneten Daten gespeichert werden sollen.

Mit dem Einverständnis der Person und der Einhaltung des Datenschutzes kann die Datenverarbeitung mit der Rechtsgrundlage in Einklang gebracht werden.

## 2.9 Zusammenfassung

Die verschiedenen, für diese Arbeit relevanten Aspekte, wie WLAN, GPS, deren Signalausbreitung, OSM und unterschiedliche Filter, wurden eingehend vorgestellt. Ein besonderes Augenmerk wurde darauf gelegt, dass die Identifizierung des WLANs über die BSSID durchgeführt wird. Die Signalstärke wird in dBm angegeben. Jedoch muss beachtet werden, dass die meisten Geräte die Signalstärke als RSSI-Wert zurückgeben, der nicht unbedingt einer physikalischen Einheit folgt. Die Signalausbreitung wird durch den Pfadverlust und die Einflussfaktoren verändert. Das LDM berechnet die Signalausbreitung des Pfadverlustes. Das MWM beachtet zusätzlich die Dämpfung, die von Hindernissen ausgeht, sowie die niedrigen Dämpfungswerte in Gebieten mit starken Reflexionen, wie in z. B. Fluren. Das GPS bestimmt die Position anhand von Satelliten. Es beruht auf dem Koordinatensystem WGS 84, das eine Position auf der Erde mit Latitude und Longitude beschreibt. Um die zufälligen Messabweichungen von GPS zu verkleinern, wird das Bootstrap-Filter eingesetzt.



# Kapitel 3

## Verwandte Arbeiten

Nachdem die Grundlagen detailliert erläutert wurden, folgt in dem vorliegenden Kapitel eine sorgfältige Auseinandersetzung mit relevanten verwandten Arbeiten.

Zunächst wird auf die Positionierung und die Tracking-Algorithmen in dem Unterkapitel 3.1 eingegangen. Darin wird anhand von Untersuchungen aus der Arbeit von Behlendorf 2012 [1] aufgezeigt, wie genau eine Lokalisierung über Smartphones ist. Dabei spielt die Unterscheidung der verschiedenen Lokalisierungsarten über GPS-, WLAN- und GSM-Signale eine entscheidende Rolle, wie die Untersuchung von Von Watzdorf et al. 2010 [24] belegt.

Danach werden die Ergebnisse eines Partikelfilters mit einem Map Matching anhand der Arbeit von Fouque et al. 2012 [5] aufgezeigt und es wird erläutert, wie dies mit dem Empfang von wenigen GPS-Satelliten funktioniert. Es wird erklärt, wie das Partikelfilter mit GPS-Daten eingesetzt werden kann.

Zuletzt werden in dem Artikel von White et al. 2000 [26] vier Tracking-Algorithmen präsentiert und miteinander verglichen.

Im nächsten Unterkapitel 3.2 werden WLAN-Messungen aus der Stadt Nürnberg vorgestellt, die in der Arbeit von Meyer et al. 2008 [14] dargelegt werden. Die Autoren beschreiben verschiedene Charakteristiken von WLANs. Darunter ist eine Beschreibung der Häufigkeit von WLANs in verschiedenen Regionen, die Entwicklung der Häufigkeit von WLANs über 15 Monate und der Verlauf von RSSI-Werten an einem Tag. Diese Untersuchung besteht aus Messungen, die, wie oben kurz erwähnt, aus einem städtischen Gebiet hervorgehen.

Die nächste Studie von Meikle et al. 2012 [13] zeigt die Ergebnisse von 50 Millionen weltweit aufgezeichneten WLAN-Messungen. Anhand dieser Daten berechnen die Autoren den Pfadverlustexponenten für verschiedene Großstädte.

Ebenfalls genaue Angaben zu Pfadverlustexponenten, die allerdings in Bezug zu ihrer Umgebung gesetzt werden, finden sich in dem Buch von Rappaport 2002 [18].

Der Artikel von Martin et al. 2010 [12] schildert das Konzept einer Lokalisierung, die offlinebasiert ist.

In der Arbeit von Kan et al. 2012 [8] wird die Umsetzung von einem Fingerprinting auf einem Smartphone erklärt. Die Ergebnisse dieses Systems werden erläutert. Zu-

letzt wird die Lokalisierung über Trilateration mit Hilfe des Log Distance Models geschildert. Die verwandte Arbeit bietet einen Eindruck, wie das LDM mit Messungen des Smartphones entworfen wurde.

Schließlich wird das Crowdsourcing in dem Unterkapitel 3.3 als vorhandene Anwendung detailliert erläutert. Darin werden die Systeme von Apple und von Google beschrieben und es wird erklärt, wie WLAN-Messungen von Anwender durchgeführt werden, um ihr System zu verbessern. Dieses Crowdsourcing zeigt, wie vorhandene Anwendungen heutzutage genutzt werden und wie dementsprechende Unternehmen vorgehen.

## 3.1 Positionierung

Zuerst werden die Untersuchungen über die Genauigkeit der Lokalisierung von Smartphones präsentiert. In der Arbeit von Behlendorf 2012 [1] wurden die GPS-Empfänger von Smartphones mit einem Referenzempfänger verglichen. Die für den Vergleich genutzten Smartphones sind das Apple iPhone 4S und das Samsung Galaxy S II. Das iPhone weist eine leicht genauere Positionierung, mit einer durchschnittlichen Abweichung von 3,3 m, als das Galaxy S II, mit 4,8 m, auf im Vergleich zum Referenzgerät, das zudem selber auch einer Abweichung unterliegt. Größere Ungenauigkeiten gab es bei der Höhenangabe im Vergleich zum Referenzgerät. Die Ergebnisse der Messungen der verwandten Arbeiten dienen den Ergebnissen der Messungen in der vorliegenden Arbeit als Vergleichsmöglichkeit, sodass ein Verständnis für die unterschiedliche Genauigkeit der Lokalisierungsarten entwickelt wird.

In einer Untersuchung von Von Watzdorf et al. 2010 [24] wurden mehrere Apple-Produkte miteinander verglichen. Dabei wurde festgestellt, dass die Smartphones iPhone 3G und iPhone 3Gs mit um die 75 % die meisten Positionen in der Nähe des Referenzpunktes haben. Die Accuracy-Werte sind insgesamt höher im Vergleich zu einem iPod touch. Dies wird damit begründet, dass die Smartphones über Telefonmasten lokalisieren, wenn kein GPS- und WLAN-Signal verfügbar ist. Der iPod touch kann sich dagegen nur über WLAN lokalisieren.

Nach der Lokalisierung wird sich nun mit den Tracking-Algorithmen befasst. In der Arbeit von Fouque et al. 2012 [5] wird das Partikelfilter auf die Rohdaten von GPS-Messungen angewendet, um auf diesen ein Map Matching durchzuführen. Als Problemstellung werden Situationen in Häuserschluchten beschrieben, in denen nicht genügend GPS-Satelliten zur Verfügung stehen, um eine Position zu bestimmen. Das Partikelfilter dient dazu, die Rohdaten des GPS zu glätten und die Messabweichungen zu berichtigen, sodass wiederum eine Vergleichsmöglichkeit mit der vorliegenden Arbeit besteht. In der Untersuchung von Fouque et al. wird eine durchschnittliche Abweichung der Position um 12 bis 15 m bei guter Sicht zu den GPS-Satelliten festgestellt. Beim zweiten Versuch wird die Anzahl der GPS-Satelliten verringert. Bei drei Satelliten existiert keine große Beeinträchtigung, dagegen werden bei zwei Satelliten mindestens 50 % der Positionen im ungünstigsten Fall falsch gesetzt im Vergleich zu der Ortung bei drei Satelliten.

Der Artikel von White et al. 2000 [26] beschreibt mehrere Map Matching-Algorithmen für Personal Navigation Assistents. Die Gruppierung dieser Algorithmen ist das Point-to-Point Matching. Die gemessene Position wird zum nächstliegenden Punkt aus dem Kartenmaterial gesetzt. Die zweite Gruppe ist ein statischer Schätzer. Dabei

wird eine Kette der gemessenen Positionen gebildet und eine Kurvenanpassung auf diese Punkte erstellt. Diese zwei Gruppen werden kombiniert, sodass vier Algorithmen getestet werden. Das Szenario besteht aus vier verschiedenen Routen. Der erste Algorithmus setzt die gemessene Position zum nächstliegenden Kartenpunkt. Er ist einfach zu implementieren und hat schnelle Laufzeiten. Der zweite Algorithmus ist eine Erweiterung des ersten, sodass die Richtungsinformation hinzugenommen wird. Der dritte Algorithmus erweitert den zweiten um die topologischen Informationen, sodass nur zwischen zwei Punkten eine Verbindung im Kartenmaterial existiert. Der vierte Algorithmus ist der statische Ansatz, dabei wird die Kurvenanpassung der gemessenen Positionen mit der Kurve des Kartenmaterials abgeglichen. Der beste Algorithmus ist der zweite. Dieser setzt die Position zwischen 66 und 88 % richtig auf die Karte. Durchschnittlich sind die Punkte also zu 73 % richtig. Der dritte Algorithmus folgt sehr nah mit 71 % dem zweiten. Der erste und vierte Algorithmus liefern mit 60 % und 69 % schlechtere Ergebnisse. Im Hinblick auf die vorliegende Arbeit ist der wohl interessanteste Punkt der Vergleich des nächsten Nachbarn mit anderen Tracking-Algorithmen.

## 3.2 Signalausbreitung

In der Arbeit von Meyer et al. 2008 [14] werden die Veränderungen der RSSI in kurzer und langer Zeit beobachtet. Deshalb wurden 15 Monate lang Signalausbreitungen von WLANs aufgezeichnet. Diese Aufnahmen wurden in einem Szenario der Nürnberger Innenstadt getätigt. So wurden 2006 auf einem Referenzpunkt zwischen 5 und 7 WLAN-Stationen beobachtet. Dies steigerte sich 2008 auf 20 bis 35 WLAN-Stationen und wird als eine typische Steigerung von WLAN-Stationen bezeichnet. Durchschnittlich wurden 15 WLAN-Stationen pro Referenzpunkt empfangen.

In einer anderen Untersuchung von Meyer et al. 2008 [14] wurde festgestellt, dass in einer Siedlung maximal 15 APs entdeckt wurden. In der Industriezone wurde bemerkt, dass WLAN-Stationen sehr rar vorhanden sind, sodass große Flächen frei von WLAN-Empfang sind. In einem städtischen Szenario wurden 50 bis 60 WLANs gezählt. Es ist selten, dort Flächen zu finden, an denen die Anzahl WLANs stark absinkt. Bei dem Szenario der vorliegenden Arbeit, das sich in Aachen befindet, handelt es sich demnach auch um ein städtisches Gebiet, so dass eine Vergleichsmöglichkeit besteht.

Des Weiteren wurde eine Messung von Meyer et al. 2008 [14] während 48 Stunden in der Innenstadt von Nürnberg durchgeführt, bei der die Anzahl der WLAN-Stationen und deren RSSI aufgezeichnet wurden. Das entsprechende Messgerät stand in einem Geschäft, von dem aus diese Messungen erfolgten. Dabei ist klar erkennbar, dass über Nacht die empfangenen Signale stabil bleiben, aber während der Öffnungszeiten des Geschäfts größere Schwankungen auftreten.

WiEye ist ein WLAN-Sniffer, der für Android programmiert wurde. Dieser Sniffer zeichnet die RSSI sowie die BSSID auf. Diese werden einer GPS-Position zugeordnet. WiEye wird über den Android Market vertrieben, von dem 30.000 Anwender jenen Sniffer heruntergeladen haben. Es wurden über 50 Millionen Messungen durchgeführt. In der Arbeit von Meikle et al. 2012 [13] wurde eine Studie über weltweite WLAN-Messungen erläutert. Dafür wurde der Datensatz von WiEye genutzt. Diese Messungen werden in drei Regionen aufgeteilt: in Geschäftsviertel, in Wohnsiedlung

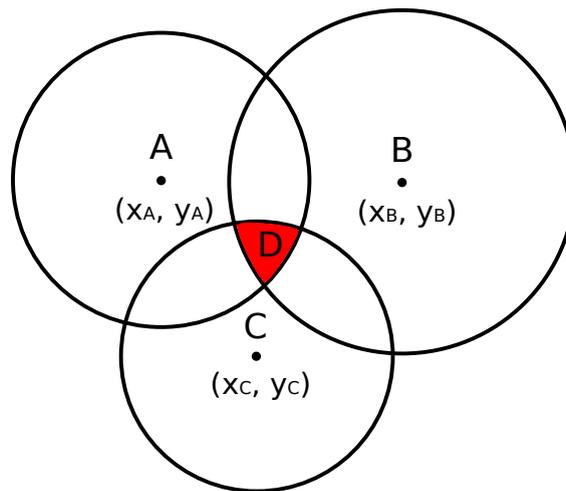
Umgebung	Pfadverlustexponent
Vakuum	2
städtisches Gebiet	2,7 - 3,5
städtisches Gebiet mit Straßenschluchten	3 - 5

**Abbildung 3.1** Pfadverlustexponent für verschiedene Umgebungen im Mobilfunk [18]

und in ländliche Gebiete. Jene Einteilungen in Regionen wurden aus OSM erhoben. Anhand der WLAN-Messungen in der Region sollte der Pfadverlustexponent für ebendieses Gebiet ermittelt werden. Dies wurde anhand des LDMs berechnet, das in Kapitel 2.2.4.2 näher erläutert ist. Die Parameter des LDMs wurden aus der Datenbank ermittelt. Dabei wurden folgende Ergebnisse ermittelt: In dem Geschäftsviertel lag der Pfadverlustexponent bei 4,5, in Wohnsiedlungen betrug er 4 und in der ländlichen Gegend schwankte er zwischen 2 und 3. Diese Messungen beziehen sich auf Greater London. Zusammenfassend kann man sagen, dass der Pfadverlustexponent zwischen 2 und 5 in städtischen Gebieten liegt. Aus dieser Datenbank wurde auch ein Bewegungsmodell der Anwender erstellt. Dabei wurden die Messungen in Cluster zusammengefasst. Das Hervorragende dieser Untersuchung für die vorliegende Arbeit sind die Masse der Messungen und die Ergebnisse des Pfadverlustexponenten in verschiedenen Städten.

In dem Buch von Rappaport 2002 [18] wird eine Tabelle mit typischen Pfadverlustexponenten in verschiedenen Umgebungen aufgezeigt. Diese Exponenten befinden sich in der Tabelle 3.1 und entsprechen etwa den Werten aus der vorherigen Arbeit. In dem Artikel von Martin et al. 2010 [12] wurde eine offlinebasierte Lokalisierung über WLAN, GSM und Accelerometer für Android-Smartphones entwickelt. Zuerst wurde eine Onlinephase ausgeführt, um die Fingerprintdatenbank aufzubauen. Dabei wurde ein Weg abgelaufen, der auch bei den Messungen abgegangen wird. Der Accelerometer und der Kompass wurden genutzt, um die Richtung der Messungen zu bestimmen und somit den Fehler durch den Körperschatten zu minimieren. Durchschnittlich wurden 25 WLANs pro Messeinheit ermittelt. Dies ergibt eine Abweichung von 1,5 m.

Anstatt die Lokalisierung über Fingerprinting durchzuführen, wird in der Arbeit von Kan et al. 2012 [8] eine Trilateration mit Hilfe des Log Distance Models umgesetzt. Das LDM berechnet die Signalstärke für eine bestimmte Distanz zu einem WLAN. Die genauere Beschreibung des LDMs befindet sich in dem Kapitel Grundlagen unter 2.2.4.2. Die Parameter stammen von Multi-Sensoren aus einem System namens Smart Space. Bekannt sind die Positionen der WLANs und die Signalausbreitung dank des LDMs. Die gemessene Signalstärke und die Distanz zwischen dem WLAN und dieser Signalstärke sind negativ korrelativ. Daher kann die Distanz zum WLAN aus der Signalstärke geschätzt werden. Das Smartphone sucht sich aus einem WLAN-Scan die drei besten Signalstärken aus und führt eine Trilateration aus. Anhand der geschätzten Distanzen zu den drei WLANs sowie deren bekannten Positionen kann die Position der Messung berechnet werden. Diese Position ist nicht immer eindeutig. Durch die Ungenauigkeit der gemessenen Signalstärke kann sich ein Raum bilden, in dem sich die Messung befinden soll. Die Abbildung 3.2 stellt die Vorgehensweise dar. Die rot markierte Fläche beschreibt den Raum, in dem die Messung vermutet wird. 50 Messungen wurden in einem Szenario mit Hindernissen aufgenommen. Mit dem



**Abbildung 3.2** Darstellung der Trilateration mit den drei WLANs A, B und C sowie deren Ergebnis, die rote Fläche D

oben beschriebenen Algorithmus wird bei der Lokalisierung ein durchschnittlicher Fehler von 1,6 m mit festen Parametern für das LDM im Vergleich zu der wahren Position festgestellt. Die Implementierung in der vorliegenden Arbeit wird ebenfalls auf einem Smartphone umgesetzt. Die Positionierungsfehler werden bestimmt.

### 3.3 Vorhandene Anwendungen

Die Grundidee des Crowdsourcings ist, dass Aufgaben nicht vom Unternehmen selber durchgeführt, sondern den Anwendern übergeben werden. Die Smartphones der Anwender zeichnen im Hintergrund die WLAN-Signalstärken mit GPS-Positionen auf, die anschließend dem Entwickler übertragen werden. Somit muss der Entwickler nicht den hohen Aufwand an Messungen erbringen. Dieses Konzept wird schon in der Wirtschaft angewendet. Apple nennt sein eigenes System *Crowdsourcing*. Dieses System existiert bereits in Smartphones, wie iPhones, und in dem mobilen Mediaplayer iPod touch. Auf der Webseite<sup>1</sup> erklärt Apple das grobe Prinzip und schildert kurz, aus welchem Grund die Daten gesammelt werden. In regelmäßigen Zeitabständen werden auf dem iPhone Messungen von WLAN und Mobilfunk zu einer GPS-Position aufgezeichnet. Das Ziel ist es, ihre Crowdsourcing-Datenbank mit Positionen der WLANs und des Mobilfunks zu erweitern bzw. zu aktualisieren. Dies ermöglicht es, bei fehlendem GPS-Empfang über die Signale von WLAN und Mobilfunk zu lokalisieren. Der Anwender hat die Option, in den Geräten diese Funktion *Ortungsdienste* zu unterbinden. Apple versichert, dass diese Daten keine Rückschlüsse auf den Besitzer des Smartphones zulassen.

Diese Lokalisierung ist ebenfalls in Googles Android-Smartphone enthalten und wird dort als *Location Service*<sup>2</sup> bezeichnet. Über die Funktion *Use wireless network* wird diese Lokalisierungsmethode aktiviert. Dabei stimmt der Benutzer gleichzeitig zu, seine Positionsdaten sowie WLAN- und Mobilfunkdaten an ihren Dienst zu schicken. Auf dieser Webseite wird auch bekanntgegeben, dass das WLAN anhand seiner

<sup>1</sup>Apple's Crowdsourcing, [http://support.apple.com/kb/HT4995?viewlocale=de\\_DE](http://support.apple.com/kb/HT4995?viewlocale=de_DE)

<sup>2</sup>Google's Location Service, <http://support.google.com/maps/bin/answer.py?hl=de&answer=1725632>

MAC-Adresse und der BSSID identifiziert wird. Google wirbt damit, dass die Lokalisierung schnell ist und dass kein großer Energieverbrauch entsteht. Es wird ebenfalls versichert, dass die Daten des Smartphone-Besitzers anonymisiert sind und der Besitzer des WLANs nicht ermittelt wird. Wenn der Besitzer nicht möchte, dass seine WLAN-Daten in das Location Service aufgenommen werden, so fügt er am Ende der SSID des WLANs das Kürzel *\_nomap* hinzu. Dieses WLAN wird dann bei den Messungen aus der Datenbank gelöscht. Jene Methode wird nicht von allen Anbietern anerkannt, doch Google hofft, dass sich dieses Kürzel in der Branche durchsetzt. Beide Unternehmen geben ihre Algorithmen allerdings nicht bekannt, sondern stellen diese unter Geheimhaltung. Es ist nur verbreitet, welche Daten aufgenommen werden und wofür sie benutzt werden. Alles, was zwischen der Eingabe und der Ausgabe passiert, können wir als Endverbraucher, Entwickler und Forscher nicht nachvollziehen.

### 3.4 Zusammenfassung

Aus diesen Arbeiten können folgende wichtige Punkte hinsichtlich der vorliegenden Arbeit herauskristallisiert werden: Die Genauigkeit der Positionierung ist stark von der Lokalisierungsart abhängig. Das GPS ist die genaueste Lokalisierung mit einer Abweichung von etwa 5 m im Vergleich zum Referenzgerät, welches selber ebenfalls einer Abweichung unterliegt. Ein guter Map Matching-Algorithmus besteht dann, wenn die Position zu der kürzesten Distanz zum nächsten Kartenpunkt mit der Richtungsinformation abgeglichen wird. Ein Ansatz wie ein Partikelfilter, um Messabweichungen zu minimieren, samt eines Map Matchings wurde mit einer durchschnittlichen Abweichung von 13 m vorgestellt. Bei der Signalausbreitung wurden mehrere Studien von WLAN-Messungen mit Smartphones präsentiert, dabei wurden in einem städtischen Szenario in Nürnberg pro Messung um die 50 WLANs festgestellt. Aus den Messungen der WiEye-Anwendung wurde der Pfadverlustexponent für Greater London um 4,5 festgestellt. Die Lokalisierung eines Smartphones mit Fingerprinting kommt auf einen Positionierungsfehler von 1,5 m und die Lokalisierung über die Trilateration mit LDM hat eine Abweichung von 1,6 m im Vergleich zur wahren Position. Diese Anwendungen werden heute schon in Apple- und Google-Smartphones genutzt. Jedoch sind deren Algorithmen nicht bekannt.

Im Vergleich mit den oben beschriebenen verwandten Arbeiten hat die vorliegende Arbeit das Ziel, eine große Anzahl repräsentativer Messungen anhand von Smartphones durchzuführen, die für das Ausbreitungsmodell in einem Stadtszenario genutzt werden können. Dabei dienen die in den verwandten Arbeiten erhaltenen Angaben als Anhaltspunkt.

# 4

Kapitel

## Entwurf

Die vorliegende Arbeit lässt sich in vier Bereiche aufgliedern, wie in der Abbildung 4.1 ersichtlich ist: in Eingabe, Positionierung, WLAN-Ausbreitungsmodell und Ausgabe.

Der erste Bereich Eingabe besteht aus drei großen Datengruppen: Die erste Datengruppe setzt sich aus den Anwendern und deren beigetragenen Messungen anhand ihrer Smartphones zusammen. Diese Messdaten bestehen aus WLAN-Scans, GPS-Positionen und Sensordaten. Diese Daten werden auf dem Smartphone zwischengespeichert. Die Messungen dienen als empirischer Ansatz für das WLAN-Ausbreitungsmodell. Der Anwender verfügt über deren Hochladen in die Scandatenbank. Die Scandatenbank dient der zentralen Sammlung aller Messungen. Die folgende große Datengruppe besteht aus OSM. Dabei wird ein Kartenausschnitt der OpenStreetMap gewählt. Dieser Ausschnitt wird von OSM nach OSM XML exportiert. Dies wird für den späteren Rasteralgorithmus benötigt. Aus den Geodaten der OSM können wir die Umwelt definieren, sodass ein deterministischer Ansatz für das Ausbreitungsmodell genutzt werden kann. Die letzte große Datengruppe ist das Referenz-WLAN. Es handelt sich um eine Liste von WLANs, die mindestens durch die BSSID und die Position des WLANs beschrieben wird. Daher ist es uns möglich, anhand eines WLAN-Scans die genaue Position des WLANs und die Distanz zwischen dem WLAN und dem Smartphone zu erkennen.

Der zweite Bereich Positionierung befasst sich mit der Minimierung des GPS-Fehlers. Dafür nutzen wir die Geodaten aus OSM, um ein Map Matching durchzuführen. Diese Geodaten lassen wir durch den Rasteralgorithmus laufen, um einen Referenzpunkt auf öffentlichen Straßen und Plätzen zu setzen. Daraus entwickeln wir die Rasterdatenbank mit ihren Referenzpunkten. Das Partikelfilter wird eingesetzt, um die systematische Abweichung der GPS-Position zu verkleinern. Diese Ergebnisse werden den Referenzpunkten in der Rasterdatenbank hinzugefügt.

Der dritte Bereich WLAN-Ausbreitungsmodell befasst sich mit dem Log Distance Model. Anstatt des Fingerprintings wird das Log Distance Model genutzt, damit die Signalstärke bei allen Referenzpunkten berechnet wird. Dieses Modell besteht aus drei Variablen, die bestimmt werden müssen: aus Pfadverlustexponent, Pfadverlust

und Distanz. Es definiert die Signalstärke auf einem Referenzpunkt für ein WLAN und gibt dies in KML (Keyhole Markup Language) aus, sodass wir sie visualisieren können. Dies ist der letzte Bereich der Anwendung, also die Ausgabe. Im Folgenden wird dieses Kapitel in fünf Unterkapitel aufgeteilt: in Client-Server-Architektur, Positionierung, Signalausbreitungsmodell und Visualisierung sowie Datenschutz. Die Client-Server-Architektur beschreibt in dem Unterkapitel ?? das Smartphone und die Scandatenbank, die in der Übersicht Abbildung 4.1 als Eingabe dargestellt sind. In dem nächsten Unterkapitel ?? Positionierung werden die Eingabe der Geodaten von OSM und der Scandatenbank erläutert, um die Rasterdatenbank mit Referenzpunkten und Messungen zu füllen.

Danach folgt in dem Unterkapitel ?? der Entwurf der Signalausbreitungsmodelle LDM und MWM. Es wird beschrieben, wie sie aus den Messungen und Referenz-WLANs die Ausbreitung der WLAN-Signale berechnen.

Das nächste Unterkapitel ?? beschäftigt sich mit der Visualisierung der Ergebnisse der Signalausbreitungsmodelle in KML.

Zum Schluss wird in dem Unterkapitel ?? erläutert, dass die Messungen, die die Anwender durchführen, gemäß des Datenschutzes aufgezeichnet und verwendet werden.

## 4.1 Client-Server-Architektur

Die Client-Server-Architektur besteht aus mehreren Smartphones als Clients und einem Datenbankserver. Es beschreibt die erste Eingabegruppe, wie die Smartphones die Messungen durchführen und danach die Messungen zentral auf einen Datenbankserver speichern. Die Clients führen die Messungen also aus und speichern sie lokal. Der Anwender kann diese Messungen später per Button auf den Server hochladen. Der Server dient dazu, alle Messungen der Clients in einer Datenbank zusammenzufassen.

### 4.1.1 Client

Eine der wichtigsten Eingaben für diese Arbeit sind WLAN-Messungen und die Bestimmung deren Position mit Hilfe von mobilen Geräten. Dabei kommen nur mobile Geräte in Frage, welche die Anwender im Alltag bei sich tragen und die im Hintergrund die Messungen durchführen. Nach diesen Kriterien entschieden wir uns für Smartphones. Im Vergleich zu anderen mobilen Geräten, wie z. B. Laptops und Tablets, trägt der Anwender im Alltag eher ein Smartphone bei sich. Es gewährleistet eine Plattform, auf der man Anwendungen erweitern kann, sodass wir über eine selbst programmierte Anwendung Messungen durchführen können. Hinzu kommt, dass es meist über die gesamte wichtige Hardware verfügt, wie z. B. WLAN und GPS. In Deutschland existieren genügend Smartphone-Besitzer, die Messungen durchführen können. Laut einer repräsentativen ARIS-Umfrage<sup>1</sup> trägt jeder dritte Anwender in Deutschland ein Smartphone bei sich. Ein Hauptproblem des Smartphones ist, dass es im eigentlichen Sinne kein Messgerät ist und dass die Messwerte daher abweichen

<sup>1</sup>ARIS-Umfrage von 2012 im Auftrag von BITKOM, weitere Informationen unter [http://www.bitkom.org/de/presse/8477\\_71854.aspx](http://www.bitkom.org/de/presse/8477_71854.aspx)

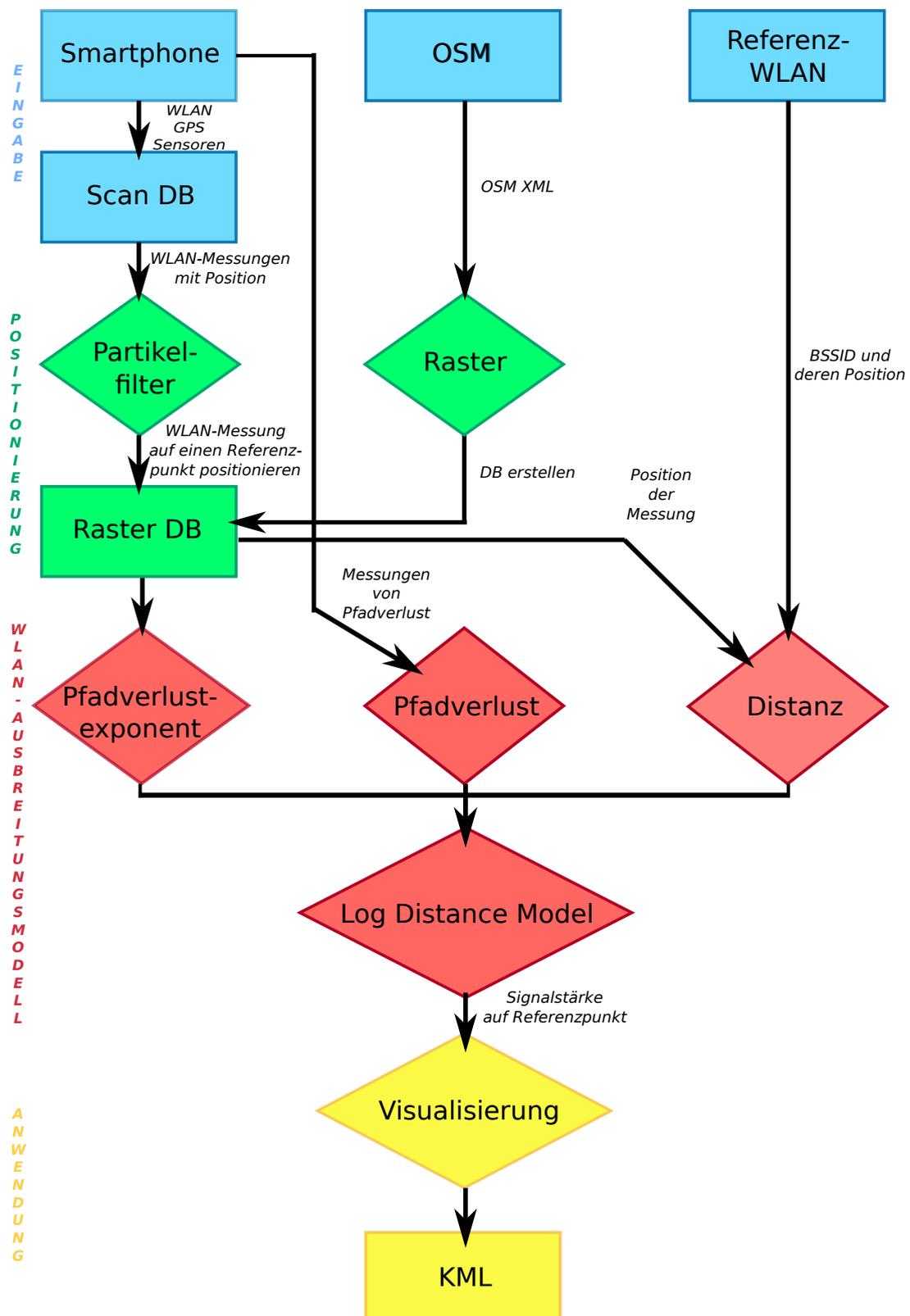


Abbildung 4.1 Übersicht des Entwurfs: Die Rechtecke sind Daten, die Rauten sind Algorithmen und die jeweilige Farbe ist eine Klassifizierung nach Themen.

können. Einzelne Messungen würden zu einer systematischen Messabweichung (siehe Kapitel 2.6.0.4) führen. Daher erfolgt der Versuch, über eine große Anzahl von Messungen die zufälligen Abweichungen und über verschiedene Smartphones die systematische Abweichung zu minimieren.

Auf dem heutigen Markt existieren vier verschiedene Betriebssysteme für Smartphones. Ein bekannter Anbieter ist iOS von der Firma Apple. Ihr Betriebssystem ist laut einer comScore MobiLens-Umfrage<sup>2</sup> mit 22 % der Smartphone-Plattformen weitverbreitet. Jedoch handelt es sich um ein proprietäres System, der Entwickler kann nur auf das ihm zur Verfügung gestellte API zugreifen. Apple bietet offiziell keine Funktion auf dem iPhone an, um die Signalstärke eines WLANs auszugeben. Die einzige Möglichkeit ist der Zugriff auf ein privates Framework, um die genaue Angabe zu erlangen. Das private Framework trägt den Namen Apple80211<sup>3</sup>. Apple verbietet diese Vorgehensweise und verbannt entsprechende Anwendungen aus dem App-Store. Eine weitere Möglichkeit wäre, dass iPhone außerhalb ihrer Kontrolle zu jailbreaken, das bedeutet dass man einen eigenen Code ohne die Signatur von Apple ausführen kann. Dies wollen wir den Anwendern jedoch nicht zumuten. Daher wird in dieser Arbeit auf das Betriebssystem iOS verzichtet.

Die beiden anderen Betriebssysteme sind Symbian von Nokia, mit 24 % ebenfalls gut vertreten, und Windows Phone von Microsoft, mit 7 % nur rar vertreten, sodass sich der Aufwand im Verhältnis zum Nutzen bei Letzterem nicht rentiert.

Der letzte Anbieter ist das Betriebssystem Android von Google. Die Wahl fällt auf dieses System, da es mit 40 % den höchsten Marktanteil hat und sämtliche APIs zur Entwicklung dieser Anwendung verfügbar sind. Ein weiterer Vorteil ist, dass Google dieses System als Open Source anbietet.

Das Ziel der Messungen mit Smartphones ist die Positionierung der Signalstärke eines WLANs auf einer Karte. Für Messungen werden die BSSID und deren RSSI von einem WLAN-Scan vorausgesetzt, mit der das Netzwerk identifizierbar wird und die empfangene Signalstärke bestimmt werden kann. Dazu wird die Position des Smartphones mit Hilfe von Mobilfunk, des WLANs und des GPS bestimmt. Auf Letzterem wird sich hauptsächlich basiert. Eine weitere Information liefert die Herstellerangabe sowie das Modell des Smartphones, sodass eine Klassifizierung der Geräte ermöglicht wird.

Für die Durchführung der Messungen wird eine Anwendung für Smartphones entwickelt. Der Anwender kann diese Anwendungen nach Bedarf eigenhändig starten, um dann eine Messung durchzuführen. Dabei werden alle Informationen von der Hardware abgefragt und in einer Datenbank abgespeichert.

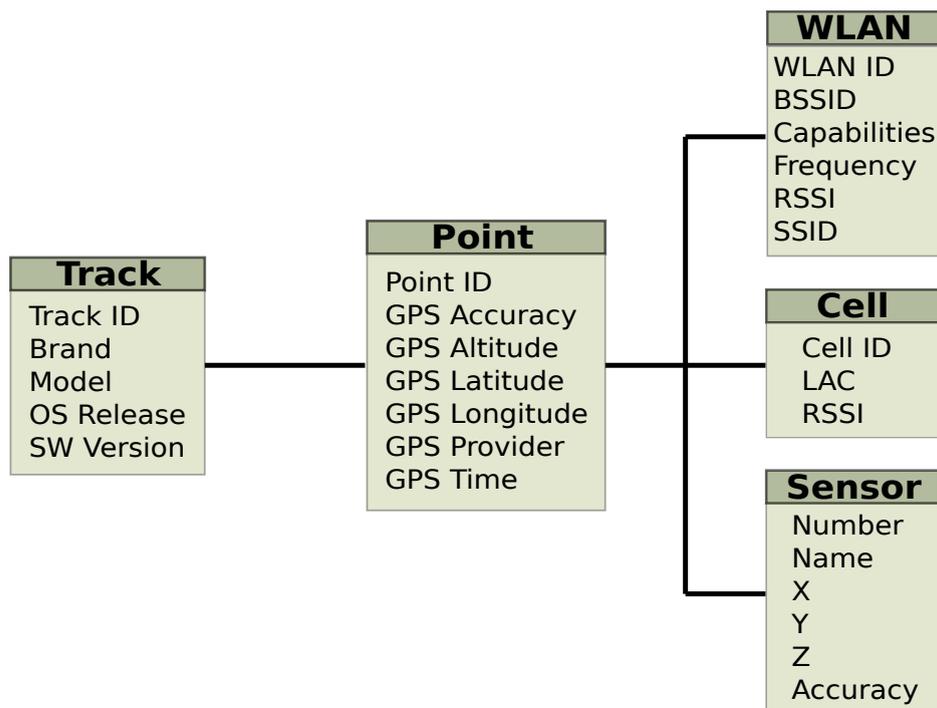
Die Anwendung sollte im Hintergrund laufen, sodass der Anwender sein Smartphone weiterhin ungestört benutzen kann. Er kann den Messvorgang jederzeit beenden. Er verfügt über die Möglichkeit, die gesammelten Daten auf eine Datenbank, den Server, hochzuladen.

## 4.1.2 Server

Die Struktur der Datenbank ist in Abbildung 4.2 dargestellt. Sie beinhaltet die fünf Tabellen Track, Point, WLAN, Cell und Sensor. Der Datensatz der Haupttabelle

<sup>2</sup>comScore-Umfrage von 2012 im Auftrag von BITKOM, weitere Informationen unter [http://www.bitkom.org/de/presse/30739\\_72316.aspx](http://www.bitkom.org/de/presse/30739_72316.aspx)

<sup>3</sup>Apple80211Scan, <http://code.google.com/p/iphone-wireless/wiki/Apple80211Scan>



**Abbildung 4.2** Struktur der Scan-Datenbank der Client-Messungen

Track stellt je einen kompletten Messvorgang dar. Dieser beinhaltet die Identifizierung des Smartphone-Herstellers und des Modells sowie die Version des Betriebssystems. Ein weiterer Wert ist die Softwareversion der Smartphone-Anwendung.

Dieser Messvorgang enthält mehrere Messungen, die in der Tabelle Point zusammengefasst sind.

Jede einzelne Messung speichert ihre Daten in den vier Tabellen Point, WLAN, Cell und Sensor ab. Die Tabelle Point gibt die Position der Messung an. Sie setzt sich aus Altitude, Latitude und Longitude sowie aus der Genauigkeit der Position, der Uhrzeit und der Quelle der Positionsbestimmung, über WLAN oder GPS, zusammen. Damit ist der Ort der Messungen nachverfolgbar, sodass die Messungen geografisch aggregiert werden können.

Von dieser Tabelle aus folgen die drei restlichen Tabellen: Die Tabelle WLAN beinhaltet die Daten eines WLAN-Scans. Mit der BSSID aus dieser Tabelle ist die Identifikation des WLANs möglich. Das Attribut RSSI ist die Empfangsleistung von einem WLAN. Diese Empfangsleistung spielt eine wichtige Rolle für das Ausbreitungsmodell. Die nächste Tabelle Cell erfasst alle empfangbaren Mobilfunkzellen der Umgebung. Mit diesen Daten ist es möglich, ein Ausbreitungsmodell für Mobilfunkzellen zu erstellen. Die letzte Tabelle Sensor ist eine Aufzeichnung der Daten aller Sensoren und deren Identifizierung, um welchen Sensor es sich handelt. Diese Daten können beim Ausbreitungsmodell helfen, um die WLAN-Messungen z. B. nach einem digitalen Kompass zu klassifizieren.

Alle Verbindungen in der Tabelle sind eine 1:n-Beziehung. Diese Struktur ist hierarchisch zu der Wurzeltabelle Track.

## 4.2 Positionierung

Die Positionierung verfolgt das Ziel, die Ungenauigkeit der Lokalisierung zu minimieren und die Position auf einen Referenzpunkt auf öffentlichen Straßen oder Plätzen zu setzen. Die erste Etappe ist die Erstellung der Referenzpunkte für die Rasterdatenbank. Dies bietet eine Metrik der Daten, um WLAN-Messung geografisch zu aggregieren. Diese Vorgehensweise wird Rastermodell genannt. Die zweite Etappe, das Minimieren des Lokalisierungsfehlers, wird erreicht, indem das Partikel-filter und das Map Matching auf diese Messdaten angewendet werden. Mit Hilfe des Map Matchings werden die Messdaten zu einem Referenzpunkt in die Rasterdatenbank hinzugefügt. Mit dem Partikelfilter soll die zufällige Messabweichung von GPS verkleinert werden. Das Map Matching fasst die Messungen geografisch zusammen, damit die Minimierung der zufälligen Messabweichung von WLAN-Signalstärken später einfacher ist.

### 4.2.1 Rastermodell

Dieses Modell befasst sich damit, wie oben kurz erwähnt, von allen öffentlichen Flächen ein Raster von Referenzpunkten zu erstellen. Die Information von diesen öffentlichen Flächen wird in dieser Arbeit den Geodaten der OSM entnommen, siehe Kapitel 2.5. Diese Flächen werden in zwei Kategorien eingeteilt: in Way und Area. Unter Way versteht man das Setzen von Referenzpunkten auf eine Linie. Area steht für das Setzen der Referenzpunkte als Gitterstruktur auf diese Fläche.

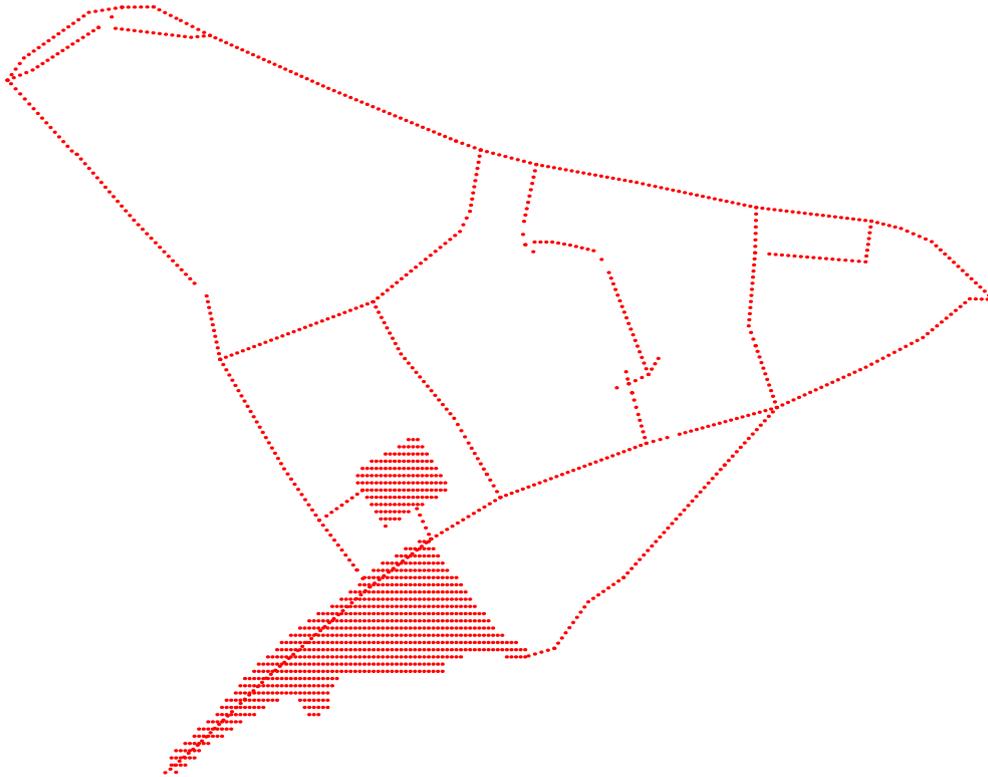
Für die Ausarbeitung wurden folgende Werte festgelegt: Die Referenzpunkte liegen im Abstand von 3 Metern zueinander. Die Abbildung 4.3 zeigt die berechneten Referenzpunkte der OSM in einem Ausschnitt der Innenstadt von Aachen.

#### 4.2.1.1 Way

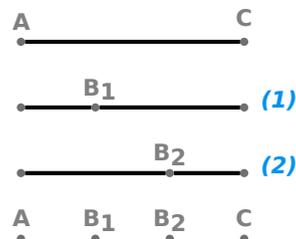
Hier wird eine Linie von Referenzpunkten zwischen zwei Nodes erstellt, das heißt dass wir aus den Geodaten den Start- und den Endpunkt der Linie entnehmen. Alle dazwischen liegenden Punkte, die auch als Referenzpunkte gelten, werden zur Rasterdatenbank hinzugefügt. Der Startpunkt ist als  $(sLat, sLon)$  und der Endpunkt als  $(eLat, eLon)$  definiert.  $countPart$  ist die Anzahl der Teilstrecken, die durch diese Teilung  $\frac{distance}{3}$  mit dem Abstand von 3 m berechnet wird. Die Berechnung einer Distanz mit zwei Koordinatenpaaren befindet sich in Kapitel 2.4.2.1.  $part1$  ist die erste Teilstrecke, die  $(i = 1, \dots, countPart - 1)$  entspricht und  $part2$  ist die restliche Teilstrecke der gesamten Linie und ergibt sich aus  $countPart - part1$ .

$$\begin{aligned} newLat &= (1/countPart) \cdot (part2 \cdot sLat + part1 \cdot eLat) \\ newLon &= (1/countPart) \cdot (part2 \cdot sLon + part1 \cdot eLon) \end{aligned} \quad (4.1)$$

Die Koordinaten  $newLat$  und  $newLon$  bilden einen Referenzpunkt. Diese Berechnung wird  $countPart - 1$ -mal wiederholt und ermittelt alle Punkte zwischen dem Start- und dem Endpunkt. Die Abbildung 4.4 zeigt die Berechnung für eine Linie, die 9 m lang ist. Daraus folgt, dass  $countPart$  gleich 3 ist. In der ersten Iteration (1) ist  $part1 = 1$ , also die Teilstrecke zwischen den Punkten  $A$  und  $B_1$ , und  $part2 = 2$ ,



**Abbildung 4.3** Referenzpunkte in Aachen von dem Markt, dem Karlshof, von der Pontstraße, dem Driescher Gässchen, dem Hirschgraben, dem Seilgraben, der Großkölstraße, der Kleinkölstraße, dem Büchel, der Neupforte, der Mostardstraße und der Minoritenstraße, vergleichbar mit der Abbildung 6.1



**Abbildung 4.4** Berechnung der Rasterung einer Linie

also die Teilstrecke zwischen  $B_1$  und  $C$ . In der zweiten Iteration (2) ist  $part1 = 2$ , die Teilstrecke zwischen  $A$  und  $B_2$ , und  $part1 = 1$ , die Teilstrecke zwischen  $B_2$  und  $C$ . Daraus ergeben sich die vier Punkte, die als Referenzpunkte in die Rasterdatenbank hinzugefügt werden.

#### 4.2.1.2 Area

Mit Area sind alle öffentlichen Straßen und Plätze gemeint. Im Gegensatz zu einem oben genannten Way handelt es sich um einen geschlossenen Way, der mit dem Tag *area* markiert ist. Dieser geschlossene Way bildet die Randlinie der Fläche. Zur Berechnung des Rasters, welches möglichst gleichmäßig über die Fläche verteilt werden soll, werden die minimale und maximale Latitude sowie Longitude bestimmt; sie formen den Raum, in dem das Raster erstellt wird. Dieser Raum bildet ein Rechteck, sodass die gegenüberliegenden Seiten gleich lang sind und die Winkel  $90^\circ$  betragen.

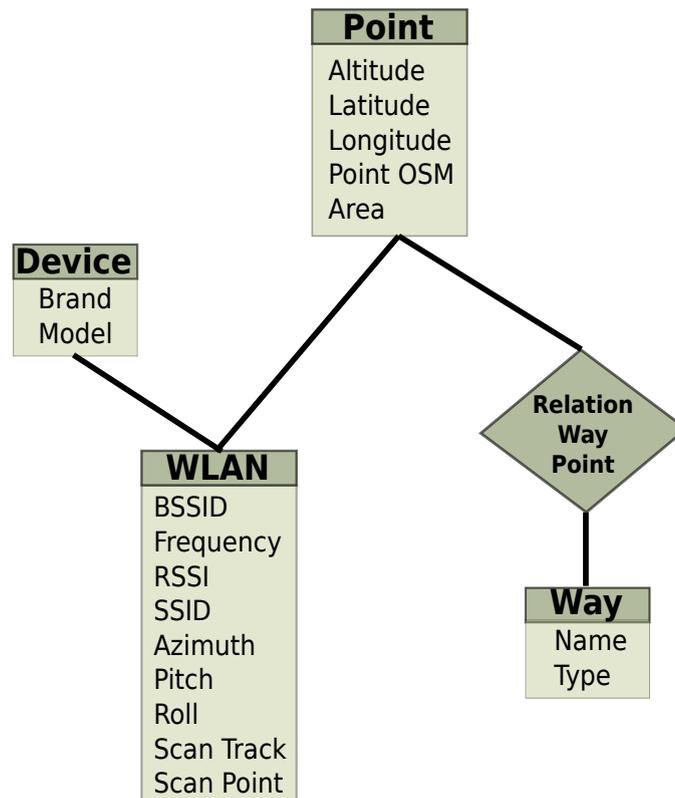


Abbildung 4.5 Datenbankstruktur vom Raster

Wir nehmen die untere (bot = bottom) und die linke Strecke (left) des Rechtecks. Auf diesen beiden Strecken wird die Berechnung aus der Formel 4.1 durchgeführt. Wir erhalten die Menge der Punkte für die untere Strecke

$$botPoints = \{(botLat_1, botLon_1), \dots, (botLat_{countBotPart}, botLon_{countBotPart})\}$$

und die Menge der Punkte für die linke Strecke

$$leftPoints = \{(leftLat_1, leftLon_1), \dots, (leftLat_{countLeftPart}, leftLon_{countLeftPart})\}$$

Um die Menge der Punkte des Gitters zu erlangen, führen wir folgende Multiplikation durch:

$$\begin{aligned}
 botPoints \times leftPoints = & \{(botLat_i, leftLon_j) | \\
 & (botLat_i, botLon_i) \in botPoints; \\
 & (leftLat_j, leftLon_j) \in leftPoints; \\
 & i = 1, \dots, countBotPart; \\
 & j = 1, \dots, countLeftPart\}
 \end{aligned} \tag{4.2}$$

Diese Punktmenge enthält Punkte in dem definierten Raum und somit auch außerhalb der Area. Daher wird jeder Punkt noch überprüft. Wenn er außerhalb der Fläche liegt, wird er entfernt.

## 4.2.2 Rasterdatenbank

Die Abbildung 4.5 stellt die Datenbankstruktur der Rasterdatenbank dar. Die Haupttabelle Point beinhaltet die Referenzpunkte, die durch die Position auf der Erde

definiert sind. Point OSM und Area sind Boolean-Werte. Point OSM sagt aus, ob der Referenzpunkt ein OSM-Punkt oder ein berechneter Rasterpunkt ist. Area beschreibt den Punkt als einen geschlossenen Way oder als einen Punkt in einer Fläche. Die Tabelle Device greift die Smartphone-Hersteller und die Modelle der Smartphones auf.

Die nächste Tabelle WLAN ergibt sich aus den Tabellen Point und Device. Sie beschreibt einen WLAN-Scan und den Orientation Sensor sowie eine Referenz zum Track, Point und zur GPS-Accuracy aus der Scandatenbank.

Die Relation Way-Point verbindet die Tabellen Point und Way zu einer n:m-Beziehung. Mit Ausnahme der Relation sind alle Verbindungen eine 1:n-Beziehung.

### 4.2.3 Partikelfilter

Die theoretischen Grundlagen sowie die Beschreibung des Algorithmus befinden sich in den Kapiteln 2.7 und 2.7.3.3. In diesem Kapitel beschäftigen wir uns mit dem Partikelfilter, um die zufälligen Messabweichungen von GPS zu minimieren sowie ein Map Matching anzuwenden, das die Messpunkte auf öffentlichen Flächen setzt. Neben der Minimierung der zufälligen Messabweichung von GPS ermöglichen es die beiden Algorithmen, eine Aggregation der WLAN-Messungen geografisch zu bestimmen. Damit werden nah beieinander liegende Messungen zusammengefasst. Dies bietet später eine einfachere Handhabung, um die zufälligen Messabweichungen von WLAN-Messungen zu verkleinern. Es wird davon ausgegangen, dass sich die Anwender während ihrer Messungen auf öffentlichen Straßen und Plätzen befinden. Anhand des Parameters GPS-Provider kann festgestellt werden, ob die Lokalisierung per GPS, WLAN oder Mobilfunk erfolgt. In dieser Arbeit wird die Lokalisierung mit Hilfe des GPS bevorzugt, weil es die genaueste Positionsbestimmung von den drei oben genannten Lokalisierungsmöglichkeiten bietet und freie Sicht zum Himmel hat. Gerade der letzte Punkt erteilt eine weitere Information und gibt an, ob sich der Anwender im Außen- oder Innenbereich befindet. Wenn die Messung ein GPS-Signal empfängt, kann davon ausgegangen werden, dass er sich im Außenbereich aufhält. Umgekehrt ist die Aussage nicht möglich, da z. B. durch hohe Straßenschluchten nicht mindestens vier GPS-Satelliten empfangbar sein können.

#### 4.2.3.1 Initialisierung

Die Initialisierung des Algorithmus für das Partikelfilter und das Map Matching werden in die beiden Unterkapitel Suchraum und Partikelfilter eingeteilt. Das erste Unterkapitel besteht aus der Bestimmung des Suchraumes. Sie bietet eine schnellere Berechnung, als eine Kalkulation über den gesamten Suchraum beanspruchen würde. Das zweite Unterkapitel beschreibt die Initialisierung und die erste Iteration des Partikelfilters.

#### Bestimmung des Suchraumes

Ein Teil der Initialisierung ist die Minimierung des Suchraumes auf die möglichen Referenzpunkte eines Tracks. Der vollständige Track ist in der Scandatenbank abge-

speichert, sodass ein Offline-Algorithmus angewendet wird. Der Vorteil des Offline-Algorithmus ist, dass er für die Berechnung einer aktuellen Position sowohl die Vergangenheit, das heißt wo er sich befand, als auch die Zukunft, das heißt wo er noch hingehet, nutzt. Der Online-Algorithmus kann die Informationen dagegen nur der Vergangenheit entnehmen.

Die Grundidee ist, dass wir für jeden GPS-Punkt einen Raum mit Hilfe der GPS-Accuracy bilden. Der gesamte Suchraum  $S$  ist eine Menge mit Referenzpunkten  $r_{lat,lon} \in S$ , die durch die Position  $lat$  und  $lon$  definiert ist. Der Track  $T$  besteht aus 2 Tupeln einer GPS-Position  $g_{lat,lon}$  und dem Accuracy-Wert  $a$  in Metern. Dies ergibt  $T = \{(g_{lat,lon}, a)\}$ . Die Accuracy beschreibt eine Distanz in Metern, die von der GPS-Position ausgeht. Dies bildet einen Kreis, in dem sich die wahre Position befindet. Jedes Tupel  $t \in T$  bildet diesen Raum mit dem Mittelpunkt  $r_{lat,lon}$ , in dem der Radius  $a$  definiert ist. Mit der Menge  $R_t$  sind alle Referenzpunkte in dem Kreis von  $t$  bestimmt. Der minimierte Suchraum  $\bar{S}$  ist die Vereinigung aller Referenzpunkte  $r_{lat,lon}$  nach:

$$\bar{S} = \{r_{lat,lon} | r_{lat,lon} \in R_t; \forall t \in T; r_{lat,lon} \in S\} \quad (4.3)$$

Das Problem, das bei großen Accuracy-Werten auftritt, ist, dass der Suchraum deutlich vergrößert wird im Vergleich zur Berechnung, die für einen Track benötigt wird. Daher stellt sich die Frage, wie groß der maximale Accuracy-Wert sein muss, sodass sich der vollständige Track im Suchraum befindet. Daher wird der minimierte Suchraum um eine obere Schranke  $a_{bound}$  für die Accuracy erweitert:

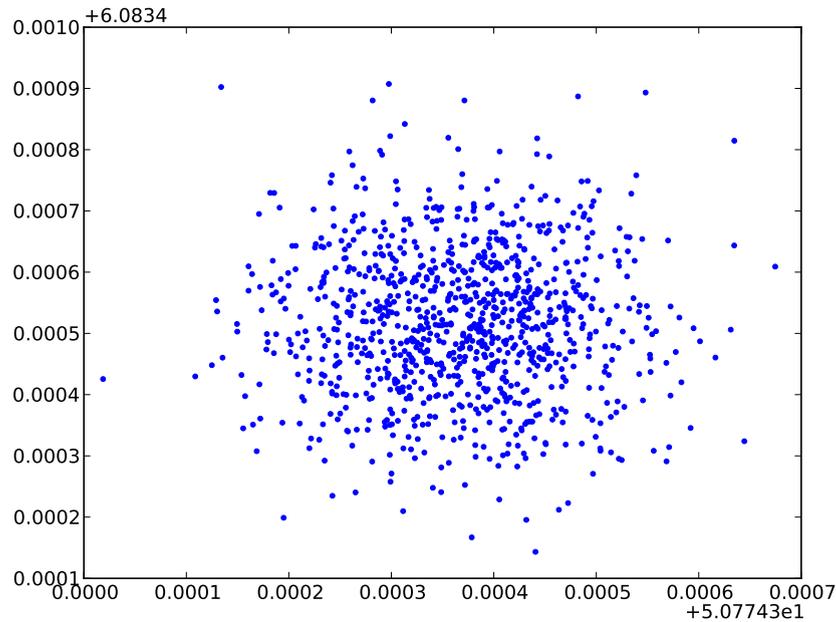
$$\bar{S} = \{r_{lat,lon} | r_{lat,lon} \in R_{\bar{t}}; \forall \bar{t} \in \bar{T}; r_{lat,lon} \in S\} \quad (4.4)$$

Der beschränkte Track  $\bar{T}$  ist die Menge  $\{(g_{lat,lon}, a) | a \leq a_{bound}; (g_{lat,lon}, a) \in T\}$ . In dem Kapitel Evaluation 6.2 sind die Accuracy-Werte des GPS als Histogramm (siehe Abbildung 6.5) dargestellt. 70 m werden als Grenzwert gewählt und festgestellt, dass 98 % der Accuracy-Werte unter 70 m bleiben. Damit sind die meisten GPS-Positionen eingeschlossen. Es können alle Positionen mit Hilfe von WLAN und Mobilfunk miteinbezogen werden, solange diese nicht den Accuracy-Wert von 70 m überschreiten. Somit wird die obere Schranke  $a_{bound}$  auf 70 m festgelegt. Ein weiterer Vorteil für die Minimierung des Suchraumes ist, dass die Position mit einem hohen Accuracy-Wert nicht im kreisförmigen Suchraum mit Hilfe der Accuracy gesucht wird, sondern im minimierten Suchraum. Der minimierte Suchraum enthält nur die Referenzpunkte mit einem niedrigen Accuracy-Wert. Dieser Raum ist kleiner und dessen Referenzpunkte liegen näher an der wahren Position.

## Initialisierung des Partikelfilters

Neben der Suchraumbestimmung werden in der Initialisierung die Anzahl  $N$  Partikel festgelegt und  $N$  Partikel aus  $P(x_0)$  gezogen. Aus den Grundlagen ist bekannt, dass, wenn  $N$  gegen unendlich geht, der quadratische Fehler gegen null konvergiert. Wenn  $N$  zu klein gewählt wird, steigt der quadratische Fehler deutlich an. Daher sollte  $N$  eine genügend große Anzahl von Partikeln sein, sodass die Größe der Stichprobe repräsentativ für die unbekannte Wahrscheinlichkeitsdichte des aktuellen Zustandes ist. Für die vorliegende Arbeit wurde die Partikelanzahl auf 100 festgelegt.

Für initialisierende Ziehungen der  $N$  Partikel stehen die GPS-Position und der



**Abbildung 4.6** 1.000 zufällig gewählte Positionen aus der Gauß-Verteilung mit der Position (50.774668,6.083927) und dem Accuracy-Wert von 10 m

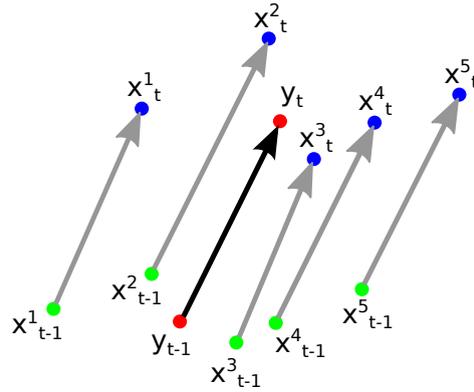
Accuracy-Wert als Information zur Verfügung. Die Position ist der vermutete Zustand und der Accuracy-Wert beschreibt den Raum, in dem sich der wahre Zustand befindet. Anstatt dass die Ziehung der Positionen zufällig im kompletten Raum bestimmt wird, wird die zufällige Position aus einer zweidimensionalen Gauß-Verteilung gezogen. Es wird angenommen, dass die GPS-Position der Erwartungswert  $\mu$  ist und der Accuracy-Wert die Standardabweichung  $\sigma$  bildet. Daher werden die Eingaben für die zweidimensionale Gauß-Verteilung erhalten: Sie ergeben sich aus der Position  $\mu = (lat, lon)$  und der Kovarianzmatrix  $\mathcal{K} = \begin{bmatrix} cLat & 0 \\ 0 & cLon \end{bmatrix}$ . Die Variablen  $cLat$  und  $cLon$  werden auf Basis der Offset-Koordinaten berechnet, wobei die Erde als flache Scheibe betrachtet wird [?]:

$$dLat = \frac{dist}{radiusEarth \cdot \cos(rLon)} \quad (4.5)$$

$$dLon = \frac{dist}{radiusEarth \cdot \cos(rLat)}$$

$rLat$  und  $rLon$  sind die radianten Werte von  $lat$  und  $lon$  der GPS-Position. Die beiden Ergebnisse werden in Grad umgewandelt und ergeben die  $cLat$ - und  $cLon$ -Werte.

Anhand dieser zweidimensionalen Gauß-Verteilung lässt sich eine Zufallsposition ziehen. Diese Position hat die Eigenschaft der Verteilung, sodass die meisten Ziehungen nah am Erwartungswert liegen und sie von der Häufigkeit her weniger zum kreisförmigen Rand werden, der durch den Accuracy-Wert in Metern definiert ist und vom Erwartungswert entfernt liegt. Diese Eigenschaften sind in der Abbildung 4.6 ersichtlich. Damit ist die Anzahl Partikel gleich  $N$  und es wurden  $N$  zufällig gewählte Partikel bestimmt, um die GPS-Position zu ermitteln. Diese Partikel bekommen



**Abbildung 4.7** Aus den roten Messpositionen  $y_{t-1}$  und  $y_t$  werden die Richtung und die Distanz bestimmt. Auf alle Partikel  $x_{t-1}^i$  werden diese Richtung und diese Distanz mit einer Abweichung angewendet.

als Initialisierung eine Gleichgewichtung. Jedes Partikel  $x_0^i$  erhält als Gewichtung  $w_0^i = \frac{1}{N}$ . Diese Gewichtungen sind schon normalisiert.

#### 4.2.3.2 Zustandsmodell

Das Zustandsmodell beschreibt die Übergangswahrscheinlichkeit  $P(x_t|x_{t-1})$  vom vorherigen Zustand in den aktuellen Zustand. Die vorherige Position und die aktuelle Position stehen als Information für den Zustandsübergang zur Verfügung. Die Idee für den Zustandsübergang wird der folgenden Publikation [27] entnommen. Aus den beiden Positionen werden die Richtung und die Distanz berechnet. Die Berechnung der Distanz *dist* wird in dem Kapitel Grundlagen unter 2.4.2.1 ausführlich beschrieben. Das Gleiche gilt auch für die Berechnung der Richtung *brng* aus zwei Positionen, die sich in Kapitel 2.4.2.2 befindet. Anhand dieser beiden Werte führen alle Partikel die gleiche Bewegung aus. Dieser Ablauf ist in der Abbildung 4.7 verfolgbar. Die Berechnung der Zielposition *newLat* und *newLon* anhand einer Startposition mit Richtung und Distanz ist in Kapitel 2.4.2.3 enthalten.

$$\begin{aligned} lat_t &= newLat(lat_{t-1}, dist, brng) \\ lon_t &= newLon(lon_{t-1}, lat_{t-1}, lat_t, dist, brng) \end{aligned} \quad (4.6)$$

Die Eingaben und Ausgaben liegen in Radiant vor. Wenn die Koordinaten  $lat_t$  und  $lon_t$  in Grad umgewandelt sind, wird ein künstliches Rauschen mit Hilfe der oben beschriebenen zweidimensionalen Gauß-Verteilung hinzugefügt, um zu einer besseren Verteilung der Partikel zu gelangen. Der Erwartungswert  $\mu$  ist gleich null und die Kovarianzmatrix wird so gewählt, dass eine kleine Abweichung um einen Meter entsteht. Mit diesem künstlichen Rauschen wird vermieden, dass Partikel auf der gleichen Position durch das Zustandsmodell auf die nächste gleiche Position gesetzt werden. Dadurch werden diese Partikel zufällig um eine Abweichung von der nächsten Position versetzt. Mehrere Partikel auf der gleichen Position entstehen bei stark gewichteten Partikeln, die das Bootstrap-Filter durchlaufen. Da ein Partikel stark gewichtet ist, wird dieses mehrfach geklont. Mit diesem Rauschen wird eine bessere Repräsentation der Wahrscheinlichkeitsdichte erreicht.

### 4.2.3.3 Sensormodell

Das Sensormodell stellt die Verteilung  $P(y_t|x_t)$  dar. Diese Verteilung gibt eine Abschätzung des aktuellen Zustandes durch eine Messung an. Das Sensormodell wird für die Gewichtung der Partikel  $\bar{w}_t^i = p(y_t|\bar{x}_t^i)$  genutzt. Die Messungen geben eine Position mit einer Abweichung von der wahren Position an. Anhand dieser Position und des Accuracy-Wertes wird eine Gauß-Funktion erstellt, welche die Gewichtung der Partikel ist. Die Berechnung der eindimensionalen Gauß-Funktion wird mit den folgenden Eingaben vorgenommen: Der Erwartungswert  $\mu$  ist gleich null. Die Varianz  $\sigma^2$  ist gleich dem Accuracy-Wert  $acc$ . Die Distanzen  $dist$  zwischen den Partikeln und den Messpositionen ist die Zufallsvariable  $\mathcal{X}$ , die zu der Gauß-Verteilung approximiert. Daraus ergibt sich die Gauß-Funktion  $\mathcal{N}$ :

$$w_t^i = \mathcal{N}(dist_t^i|0, acc_t^i) \sim \mathcal{X} \quad (4.7)$$

Je näher sich das Partikel an der Messposition befindet, desto größer ist sein Gewicht. Diese Gewichtung  $w_t^i$  muss noch normiert werden nach  $\bar{w}_t^i = \frac{w_t^i}{\sum_{j=1}^N w_t^j}$ . Daraus folgt für die Summe aller Gewichtungen  $\sum_{i=1}^N w_t^i = 1$  zu dem Zeitpunkt  $t$ .

### 4.2.3.4 Resampling

In den Grundlagen ist das Prinzip des Resamplings in Kapitel 2.7.3.3 erklärt. Schwach gewichtete Partikel sollen entfernt werden und dafür stark gewichtete Partikel geklont werden. Hier wird entschieden, welches Resampling-Verfahren gewählt wird. Wie oft ein Partikel geklont wird, wird von  $n_t^i$  bestimmt:

$$n_t^i = N \cdot w_t^i \quad (4.8)$$

Der Wert  $n_t^i$  wird abgerundet. Durch die Abrundung bleibt die Anzahl der Partikel nicht genau  $N$  über mehrere Iterationen, sondern approximiert gegen  $N$ . Wenn  $n_t^i$  null ist, wird das Partikel entfernt. Beträgt der Wert 1, bleibt das Partikel unverändert. Ist der Wert größer als 2, bleibt das Partikel erhalten und wird  $n_t < i$ -mal vervielfacht.

Dieses Resampling muss nicht bei jeder Iteration  $t$  durchgeführt werden [16]. Zur Entscheidung, ob ein Resampling folgt, wird das Maß  $effN_t$  genutzt:

$$effN_t = \frac{1}{\sum_{i=1}^N (w_t^i)} \quad \text{mit } 1 \leq effN_t \leq N \quad (4.9)$$

Damit lässt sich bestimmen, ob alle Partikel eine gute Repräsentation der Verteilung  $P(y_t|x_t)$  sind. Wenn alle Partikel gleichmäßig über das Sensormodell verteilt sind, ist die Größe  $effN_t \sim N$ . Wenn dagegen ein einziger Partikel  $x_t^i$  mit der Gewichtung  $w_t^i \sim 1$  ist und die anderen  $w_t^i \sim 0$  sind, dann ist die Größe  $effN_t \sim 1$ . Als Schranke haben wir  $N_{threshold} = \frac{14 \cdot N}{15}$  gewählt. Bei Unterschreiten dieser Schranke wird ein Resampling auf die Partikel durchgeführt.

### 4.2.3.5 Gewichteter Mittelwert aller Partikel

Bei jeder Iteration  $t$  entsteht ein Schwarm von Partikeln, der die Dichtefunktion darstellt. Um aus allen Partikeln die wahrscheinlichste Position zu ermitteln, wird der gewichtete Mittelwert auf alle Positionen der Partikel angewendet:

$$\begin{aligned} \text{meanLat}_t &= \frac{\sum_{j=1}^N (\text{lat}_t^j \cdot w_t^j)}{\sum_{k=1}^N w_t^k} \\ \text{meanLon}_t &= \frac{\sum_{j=1}^N (\text{lon}_t^j \cdot w_t^j)}{\sum_{k=1}^N w_t^k} \end{aligned} \quad (4.10)$$

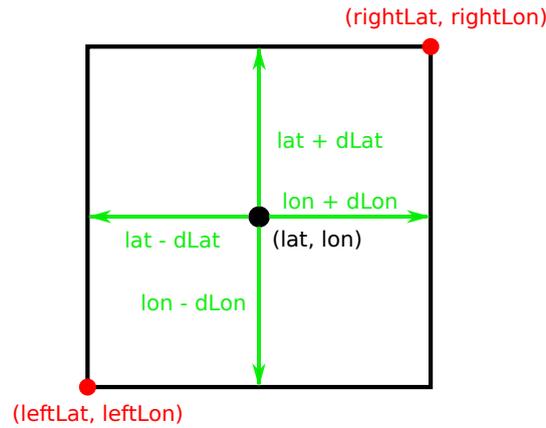
Die Position  $\text{lat}_t^j$  und  $\text{lon}_t^j$  ist in Grad angegeben, genauso wie die berechnete Position  $\text{meanLat}_t$  und  $\text{meanLon}_t$ . Der gewichtete Mittelwert hat den Vorteil gegenüber dem einfachen Mittelwert, dass Positionen von Partikeln mit einer starken Gewichtung, die somit nah an das Sensormodell herankommen, eine stärkere Beeinflussung der Bestimmung der wahrscheinlichsten Position haben.

Bei jeder Iteration  $t$  wird diese Position berechnet und als Ausgabe des Algorithmus angesehen. Alle Daten aus der Messung zum Zeitpunkt  $t$  in der Scandatenbank werden mit der neuen Position in der Rasterdatenbank abgespeichert.

## 4.2.4 Map Matching

Das Map Matching ist ein Abgleich der Position der Messungen auf einer digitalen Karte. Die Infrastruktur wurde mit der Rasterdatenbank festgelegt, sodass Referenzpunkte auf öffentlichen Flächen erstellt wurden. Das Ziel ist, die Messpunkte zu einem Referenzpunkt zu aggregieren, sodass die wahre Position auf der Karte ermittelt wird. Diese Datenstruktur erleichtert die Auswertung und die Aggregation führt zu einer schnelleren Berechnung.

Die kürzeste Orthodrome zwischen einem Messpunkt und einem Referenzpunkt (siehe Kapitel ??) bestimmt, auf welchen Referenzpunkt der Messpunkt gesetzt wird. Die Berechnung der Orthodrome befindet sich in Kapitel 2.4.2.1. Damit nicht alle Orthodrome von dem Messpunkt zu allen Referenzpunkten berechnet werden, findet eine Begrenzung auf eine kleinere Menge von Referenzpunkten statt. Damit werden kürzere Laufzeiten ermöglicht. Wir gehen von der Annahme aus, dass sich der Messpunkt in der Nähe eines Referenzpunktes befindet, da die wahre Position per GPS freie Sicht zum Himmel braucht und sie sich somit auf einer öffentlichen Fläche mit Referenzpunkten befindet. Daher soll die kürzeste Orthodrome im minimierten Suchraum enthalten sein. Die Frage wird gestellt, wie groß dieser Suchraum sein muss. In der Evaluation in Kapitel ?? zeigt das Histogramm 6.5, dass sich etwa 86% aller Messpunkte unter 30 m des Accuracy-Wertes befinden. Diese 30 m werden als erste Distanz  $\text{dist}_1$  festgelegt. Wenn in dieser Distanz  $\text{dist}_{i-1}$  kein Referenzpunkt in diesem Raum gefunden wird, wird die Distanz um  $\text{dist}_i^{i*2}$  für die  $i$ -te Iteration vergrößert. Um eine Endlosschleife zu vermeiden, wird eine obere Schranke hinzubestimmt. Mit dieser Distanz lassen sich vom Messpunkt die zwei Positionen  $\text{leftLat}$  und  $\text{leftLon}$  sowie  $\text{rightLat}$  und  $\text{rightLon}$  festlegen. Die Formeln 4.6 werden für die Offset-Koordinaten  $dLat$  und  $dLon$  gebraucht und berechnen die Distanz  $\text{dist}_i$



**Abbildung 4.8** Der schwarze Punkt ist die Messposition. Sie wird mit den Offsets  $dLat$  und  $dLon$  addiert bzw. subtrahiert, um die äußeren Positionen, die roten Punkte, zu bestimmen.

in einen Offset um. Damit lassen sich die zwei Positionen durch eine Addition und Subtraktion definieren:

$$\begin{aligned} rightLat &= lat + dLat & leftLat &= lat - dLat \\ rightLon &= lon + dLon & leftLon &= lon - dLon \end{aligned} \quad (4.11)$$

Die Ergebnisse werden in Grad umgewandelt und ergeben die äußeren Punkte des Suchraumes. Die Berechnung und die Definition des Raumes stellt die Abbildung 4.8 dar. Um aus den beiden Positionen den Suchraum zu definieren, gilt Folgendes:

$$\begin{aligned} S_m &= \{r_{lat,lon} \mid lat \geq leftLat; lat \leq rightLat; \\ &\quad lon \geq leftLon; lon \leq rightLon; \\ &\quad r_{lat,lon} \in S\} \end{aligned} \quad (4.12)$$

Der Vorteil der Bestimmung dieses minimierten Suchraumes ist, dass sich die Menge mit Database Management System (DBMS) auf eine Datenbank anwenden lässt und dies somit zeitlich gesehen effizient berechnet wird. Aus dieser Menge der Referenzpunkte werden die Distanzen zur Messposition berechnet. Die Messposition wird bei der minimalen Distanz auf diesen Referenzpunkt gesetzt. Wir gehen davon aus, dass die Abweichung der Messposition zur wahren Position vom Partikelfilter so minimiert wurde, dass sich der nächstliegende Referenzpunkt der wahren Position nähert.

Dieses Map Matching wird direkt im Partikelfilter angewendet, sodass jedes Partikel auf einem Referenzpunkt liegt. Der Vorteil ist, dass durch das künstliche Rauschen eine Streuung bei den Partikeln entsteht. Sie werden mit Map Matching wieder zurück auf eine öffentliche Fläche gesetzt. Sonst kann ein Partikel mitten in einem Gebäude liegen, in dem sich die Person per GPS nicht aufgehalten hat. An den folgenden Stellen im Partikelfilter wird das Map Matching angewendet:

**Initialisierung:** Sobald ein Partikel zufällig über die Gauß-Verteilung gezogen ist, wird es auf den nächstliegenden Referenzpunkt gesetzt.

**Zustandsmodell:** Die Partikel führen zunächst die gleiche Bewegung wie die beiden GPS-Positionen aus. Da sie nicht unbedingt auf einem Referenzpunkt liegen, wird das Map Matching ausgeführt.

**Gewichteter Mittelwert:** Aus den Positionen aller Partikel wird die Position durch den gewichteten Mittelwert berechnet. Diese Position muss auch einem Referenzpunkt zugeordnet werden, da gerade diese Position als Ausgabe des Algorithmus dient.

Genau an den Stellen im Partikelfilter, an denen sich die Position des Partikels verändert, wird das Map Matching ausgeführt und somit dessen Position korrigiert.

## 4.3 Signalausbreitungsmodell

Bis ein komplettes Gebiet mit Messungen abgedeckt wird, vergeht eine längere Zeit. Daher gibt es mehrere Referenzpunkte, die keine Messungen enthalten. Über diese Referenzpunkte kann keine Aussage getroffen werden. Um dies zu vermeiden, wird ein Signalausbreitungsmodell eingesetzt. Anhand der Berechnung des Modells wird die Ausbreitung der Signale über das Sendegebiet gemessen. Dadurch werden die Lücken, wo keine Messungen stattgefunden haben, anhand des Modells vervollständigt. Das erste Modell ist das Log Distance Model, das den Pfadverlust über eine Distanz berechnet. Bei dem zweiten Modell handelt es sich um das Multi Wall Model. Es ist eine Erweiterung des LDMs. Dabei wird jeder Strahl eines Signals zwischen dem Sender und dem Empfänger anhand eines Umgebungsmodells berechnet. Es wird gezeigt, ob der Strahl ein anderes Medium schneidet. Wenn ein Strahl z. B. durch eine Wand gesendet wird, so wird der Dämpfungsfaktor für diese Wand mitberechnet. In den nächsten beiden Unterkapiteln wird die Ermittlung der Eingaben für beide Modelle, LDM und MWM, aus der Rasterdatenbank erläutert.

### 4.3.1 Log Distance Model

In den Grundlagen in Kapitel 2.2.4.2 ist das Log Distance Model ausführlich erklärt. Dieses empirische Modell braucht drei Eingaben für die Berechnung der Signalstärke auf der Distanz. Diese Eingaben sind der Pfadverlustexponent, der Pfadverlust und die Distanz. Im Folgenden werden diese einzeln genau beschrieben und es wird gezeigt, wie deren Berechnung aus den Messungen erfolgt.

#### 4.3.1.1 Pfadverlustexponent

Der Pfadverlustexponent  $n$  beschreibt die Dämpfung der Umwelt. In einem Vakuum-Raum ist der Pfadverlustexponent gleich 2. Befinden sich Gebäude im Weg der Signalausbreitung, so wird der Pfadverlustexponent größer als 2. Bei Straßenschluchten, in denen die Signalstärke durch Mehrwegeausbreitung verstärkt wird, kann der Exponent dagegen unter 2 fallen. In einer Innenstadt kann der Wert über Smartphone-Messungen um 4 liegen [13]. Aus der LDM-Formel 2.3 lässt sich Folgendes umwandeln:

$$n = \frac{L_{LDM} - L_0}{10 \log(d)} \quad (4.13)$$

Der Wert  $L_0$ , und wie dieser bestimmt wird, ist in dem Kapitel ?? Pfadverlust beschrieben. Die Berechnung der Distanz ist in Kapitel 4.4.1.3 zu finden. Die einzige

Variable, die offen bleibt, ist die Empfangsleistung  $L_{LDM}$  in dB. Von den Referenz-WLANs wissen wir, dass deren Sendeleistung  $tp^i$  in dBm für ein WLAN  $i$  beträgt und dass aus den Messungen die Empfangsleistung  $rss_i^i$  in dBm ist. Den Rechenregeln für das Dezibel zufolge, in Kapitel 2.2 erläutert, wird die Empfangsleistung bestimmt:

$$L_{LDM}^i = tp^i - rssi^i \quad (4.14)$$

Mit diesen drei Werten lässt sich der Pfadverlustexponent berechnen. Aus der Rasterdatenbank werden Messungen für eine Region entnommen und für jede Messung  $n$  ausgewertet. Aus allen  $n$  wird der Mittelwert bestimmt und dieses Ergebnis als Pfadverlustexponent für diese Region festgelegt.

#### 4.3.1.2 Referenz-Pfadverlust

Der Referenz-Pfadverlust  $L_0$  beschreibt den Leistungsverlust eines Signals in einem Vakuum über eine vorgegebene Distanz  $d_0$ . Für  $d_0$  wird meist ein Meter festgelegt. Die Ermittlung dieses Pfadverlustes kann durch folgende Möglichkeiten bestimmt werden: Die erste Möglichkeit ist eine Berechnung über die Rasterdatenbank. Daraus wird die Distanz zwischen dem Empfänger und dem AP für die gegebene Empfangsleistung bestimmt. Die Sendeleistung des APs ist bekannt. Mit deren Hilfe und der Hilfe der Empfangsleistung kann der Signalverlust berechnet werden. Über die Distanz und den Signalverlust wird der Pfadverlust ermittelt. Das Problem bei diesen Ergebnissen ist, dass es sich nicht um einen Vakuum-Raum handelt. Die meisten APs befinden sich innerhalb eines Gebäudes und das Signal muss mindestens eine Wand nach draußen durchdringen. Außerhalb des Gebäudes können Wetterbedingungen, wie z. B. hohe Luftfeuchtigkeit, das Signal weiter dämpfen. Der eigene Körperschatten oder andere Personen können das Signal beeinflussen. Dies sind alle Faktoren, die Einfluss auf das Signal haben können und die sich nicht durch Berechnungen feststellen lassen.

Das Ziel ist es, einen genauen Wert für den Referenz-Pfadverlust über einen Meter zu erhalten, um eine große Abdeckung von Smartphones zu ermitteln, die zur Messung genutzt wurden. Daher wird die zweite Möglichkeit gewählt, bei der Messungen mit Hilfe mehrerer Smartphones von einem WLAN in einem Abstand von einem Meter durchgeführt werden. Die Auswertung dieser Messungen befindet sich in dem Kapitel Evaluation ???. Dort wird begründet, wie sich die Signalstärken von verschiedenen Smartphones untereinander verhalten und wie die Signalstärke über ein Zeitfenster variiert.

#### 4.3.1.3 Distanz

Die Distanz  $d$  ist die Entfernung zwischen dem AP und der Position, in der die Empfangsleistung berechnet wird. Diese Distanz ist in der Einheit Meter angegeben. Um sie überhaupt berechnen zu können, muss die Position des APs bekannt sein. Daher ist die Liste der Referenz-WLANs eine Eingabe dieses Entwurfs. In dieser Liste werden die WLANs, die mit der BSSID identifiziert werden, an der wahren Position beschrieben. Anhand der Haversine-Formel (siehe Kapitel 2.9) kann für jeden Punkt in der Umgebung die Distanz und somit auch die Empfangsleistung berechnet werden. Für die Visualisierung bietet sich der umgekehrte Fall. Anstatt

die Empfangsleistung für eine Position zu berechnen, suchen wir die Distanz für eine vordefinierte Leistung:

$$d = 10^{\left(\frac{L_{LDM} - L_0}{10n}\right)} \quad (4.15)$$

Dafür geben wir eine Liste mit Empfangsleistungen vor, z. B.: [-50,-60,-70,-80,-90]. Die Ergebnisse sind die passenden Distanzen für ein WLAN. Diese lassen sich als Kreis gut darstellen und ermöglichen einen direkten Vergleich mit anderen WLANs und geben an, welche Distanz sie bei gleicher Empfangsleistung erreichen.

### 4.3.2 Multi Wall Model

Das nächste Signalausbreitungsmodell, das hier vorgestellt wird, ist das Multi Wall Model. Genau wie das LDM wird es in dem Kapitel Grundlagen unter 2.2.4.2 genauer beschrieben. Bei dem MWM handelt es sich um eine Erweiterung des LDMS. Es enthält somit die Eingaben für die Dämpfungsfaktoren für Wände und Flure. Im Vergleich zu LDM werden die Effekte, die durch Hindernisse, wie z. B. Wände und Flure, entstehen, berücksichtigt. Das MWM beachtet, ob sich ein Hindernis, wie z. B. eine Wand oder ein Gebäude, in der Sichtlinie befindet. Wenn es sich in der Sichtlinie befindet, dann wird ein Dämpfungsfaktor von diesem Hindernis hinzugerechnet. Dabei werden diese Hindernisse in verschiedenen Typen mit deren passenden Dämpfungsfaktoren eingeteilt. In Fluren und Straßenschluchten können Reflexionen entstehen, welche die Signalstärke verstärken. Dieser Effekt wird mit einem eigenen Dämpfungsfaktor angegeben. Somit werden mehr Einflussfaktoren im MWM in Betracht gezogen als im LDM. Die Ergebnisse von MWM sind meist näher an der wirklichen Signalausbreitung als die Ergebnisse des LDMS [29].

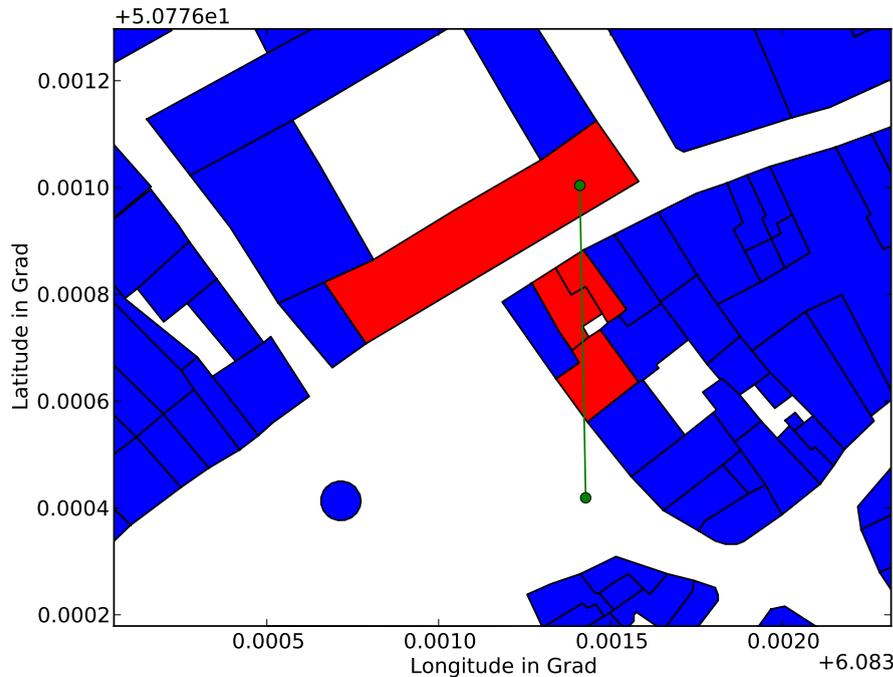
Hier wird beschrieben, wie das Umgebungsmodell aus OSM erstellt wird und wie anhand dieses Modells bestimmt wird, ob ein Strahl ein Medium des Modells schneidet. Statt Wände und Flure im Innenbereich beim MWM werden im Außenbereich Gebäude und Straßen festgelegt. Diese beiden Medien werden im Umgebungsmodell aus den Geodaten umgesetzt. Anhand der Messungen aus der Rasterdatenbank sowie der Referenz-WLANs kann der Strahl vom AP zur Messung bestimmt werden. Die Dämpfungsfaktoren für Gebäude und Straßen werden anhand der Messungen festgelegt. Die anderen Eingaben des Modells basieren auf dem LDM, das in Kapitel 4.4.1 beschrieben ist. Die vollständige Implementierung und die daraus folgende Evaluierung übersteigen allerdings den Rahmen dieser Arbeit.

#### 4.3.2.1 Umgebungsmodell

In diesem Unterkapitel geht es um die Bestimmung der Anzahl Schnitte von Gebäuden  $k_w$  und Straßen  $k_f$ . Diese Werte  $k_w$  und  $k_f$  ergeben einen Teil der Eingaben für die MWM-Formel aus dem Kapitel 2.4. Aus OSM werden die Daten der Gebäude und Straßen ausgelesen.

#### Gebäude

Ein Gebäude ist in den Geodaten von OSM ein geschlossener Way, bei dem das Tag *building* gleich *yes* gesetzt wird. So stehen die Daten des Gebäudes als eine



**Abbildung 4.9** Darstellung eines Schnittes durch mehrere Gebäude: Die grünen Punkte sind der Sender und der Empfänger und die grüne Linie ist der Strahl zwischen den beiden. Die Polygone stellen Gebäude dar, wobei die roten Gebäude nach der Berechnung geschnitten werden und die blauen Gebäude nicht.

geordnete Liste von WGS 84-Koordinaten zur Verfügung, die ein Polygon bilden. Bei jeder Kante des Polygons wird überprüft, ob ein Schnitt des Strahls durch die Kante des Polygons verläuft. Wenn mindestens ein Schnitt vorhanden ist, wird das Polygon vom Strahl geschnitten und der Wert  $k_w$  um eins erhöht. Es müssen nicht notwendigerweise mindestens zwei Schnitte vorhanden sein, da sich der Sender bzw. der AP in einem Gebäude des Modells befinden kann, sodass der Dämpfungsfaktor dieses Gebäudes in der MWM-Formel hinzugezählt wird.

Damit stellt sich die Frage, wie ein Schnitt von zwei Strecken bestimmt wird. Für die Berechnung werden die Strecken als Geraden aufgefasst. In dem Kapitel Grundlagen in 2.4.1.1 befindet sich die Erläuterung, wie der Schnittpunkt von zwei Geraden bestimmt wird. Wenn die beiden Geraden parallel zueinander sind, existiert kein Schnitt und es wird kein Schnittpunkt ausgegeben. Wenn ein Schnittpunkt existiert, muss überprüft werden, ob sich der Schnittpunkt auf den beiden Strecken befindet. Die Berechnung, ob sich ein Punkt auf einer Strecke befindet, wird in Kapitel 2.4.1.2 beschrieben. Wenn sich der Schnittpunkt auf beiden Strecken befindet, existiert ein Schnitt dieser Strecken.

Die Abbildung 4.9 stellt einen Schnitt eines Strahls durch die Gebäude dar. Daraus ist zu entnehmen, dass der Strahl durch vier Gebäude sendet und somit die Variable  $k_w$  gleich vier ist.

## Straße

Ob ein Strahl durch eine Straße läuft, wird nach dem gleichen Prinzip, dem Schnitt von zwei Strecken, berechnet, wie in dem vorherigen Unterkapitel Gebäude beschrieben ist. Wenn der Strahl die Straße schneidet, dann wird der Wert  $k_f$  um eins erhöht.

### 4.3.2.2 Dämpfungsfaktor

Wenn ein Strahl durch ein Medium sendet, dann wird er je nach Medium unterschiedlich gedämpft. Der Dämpfungsfaktor beschreibt, wie stark der Strahl zwischen dem Eintreten in das Medium und dem Austreten aus dem Medium gedämpft wird.

Die Berechnung des Dämpfungsfaktors folgt diesem Prinzip. Die Formel wird dem Kapitel Grundlagen unter 2.4 entnommen. Dabei wird diese Formel vereinfacht und der Gebäudetyp auf einen einzigen beschränkt.

$$L_{MWM} = L_{LDM} + k_b L_b + k_s L_s \quad (4.16)$$

Die Dämpfungsfaktoren,  $L_b$  für Gebäude und  $L_s$  für Straßen, werden gesucht und bilden die beiden Unbekannten in der Gleichung. Die restlichen Werte sind Konstanten und werden auf folgende Weise bestimmt: Die Position der Messung und des Referenz-WLANs sind bekannt. Aus den Geodaten können wir die Straßen und Gebäude ermitteln, sodass die Anzahl der Gebäude  $k_b$  und der Straßen  $k_s$ , die der Strahl zwischen dem Referenz-WLAN und dem Smartphone durchläuft, berechnet werden kann. Aus den Messungen wird die Signalstärke für eine Position von einem Referenz-WLAN ausgelesen. Diese Signalstärke wird mit der Variablen  $L_{MWM}$  gleichgesetzt. Sie soll das Ergebnis eines MWMs darstellen. Es gilt die Annahme, dass die Berechnung des MWMs der realen Messung entspricht. Mit dem LDM können wir die Signalausbreitung für die Position der Messung vom Referenz-WLAN berechnen.

Die Annahme ist, dass die Formel ?? angibt, dass es sich um ein lineares Gleichungssystem mit zwei Unbekannten handelt. Dafür muss die Konstante  $L_{LDM}$  auf die andere Seite der Formel gesetzt werden.

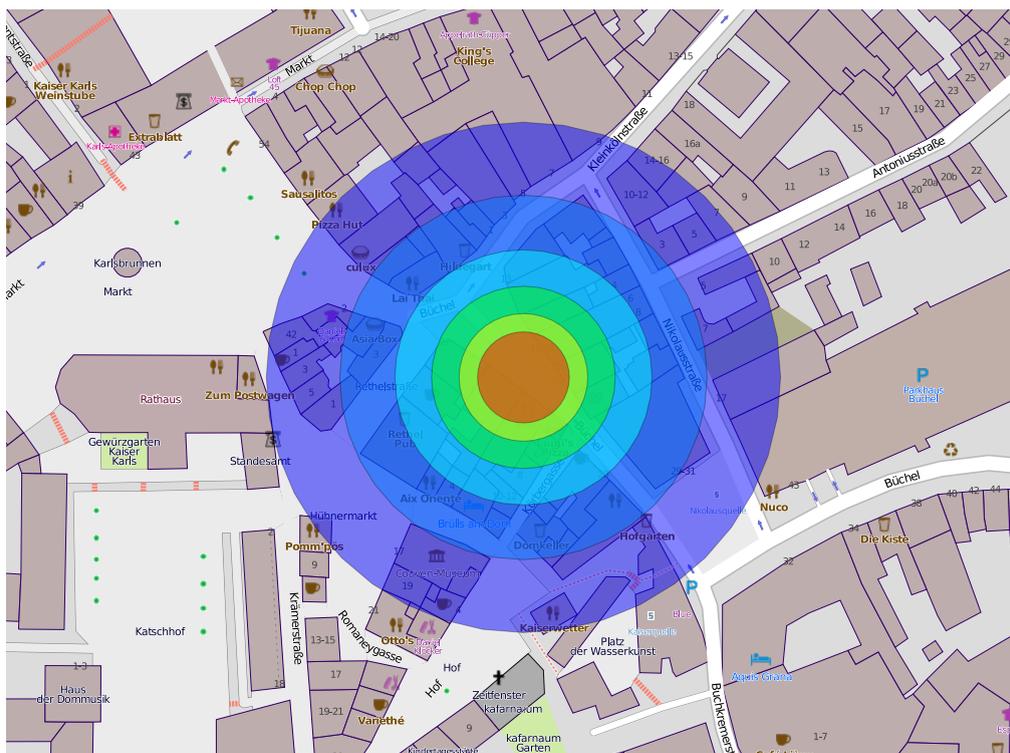
$$L_{MWM} - L_{LDM} = k_b L_b + k_s L_s \quad (4.17)$$

Für das Lösen von linearen Gleichungssystemen gibt es mehrere Möglichkeiten, wie das gaußsche Elementierungsverfahren, das Jacobi-Verfahren usw.

Da der Fall eintreten kann, dass mehrere Messergebnisse als Unbekannte existieren, wird das Gleichungssystem mehr Gleichungen als Unbekannte haben. Dadurch kann das Gleichungssystem keine Lösung bieten. In diesem Fall wird die Methode der kleinsten Quadrate eingesetzt. Damit wird eine Lösung gefunden, die die Gleichung des Systems nicht genau löst, aber zumindest sich dem wahren Wert nähert.

## 4.4 Visualisierung

Für die Darstellung, wie die Signale in die Umgebung ausstrahlen, wird die Keyhole Markup Language (KML) genutzt. Es handelt sich um eine definierte XML-Sprache,



**Abbildung 4.10** Darstellung der Signalausbreitung anhand des LDMs - Daten von OpenStreetMap, 2012, veröffentlicht unter CC-BY-SA 2.0

die Geodaten enthält. Das LDM beschreibt, wie sich die Signalstärke kreisförmig ausbreitet. So kann die Idee aus dem Kapitel 4.4.1.3 übernommen werden, sodass für eine Liste von Empfangsleistungen die passenden Distanzen berechnet werden. Diese Distanzen werden als Kreise mit der Position des APs als Mittelpunkt in der entsprechenden Farbe der Signalstärke gezeichnet bzw. gefärbt. Die Abbildung 4.10 zeigt ein Beispiel dieser Darstellung von LDM.

## 4.5 Datenschutz

In dem Kapitel Grundlagen ist der Datenschutz eingehend erklärt (siehe 2.8). Er wird beim Entwurf für die Aufnahmen von Messungen fortwährend beachtet. Folgende zwei Kriterien kennzeichnen daher die Messdaten:

1. Die Daten sind nicht personenbezogen. Dies bedeutet, dass in den Messungen nicht nachvollziehbar ist, von welcher Person welche Messung stammt. Es wird den Personen zudem keine eindeutige anonymisierte Identifikation gegeben. Damit fallen die Messungen nicht unter das Persönlichkeitsrecht und die Datensicherheit. Die Datenbankstruktur in Kapitel 4.2.1 ist ebenfalls nach diesen Kriterien aus Punkt 1 aufgebaut.
2. Der Anwender ist informiert und in Kenntnis gesetzt, welche Daten aufgenommen werden. Er hat die vollständige Kontrolle, wann und wo die Daten aufgezeichnet werden. Im Nachhinein hat er die Möglichkeit, diese Daten zu löschen. Mit dem Upload zeigt er sich damit einverstanden, die Daten dieser

Arbeit zur Verfügung zu stellen. Diese Kriterien werden in der Implementierung der Android-Anwendung umgesetzt. Eine Homepage mit allen elementaren Informationen bringt den Anwender stets auf den neuesten Stand.

Diese Anwendung wurde nur einer geschlossenen Gruppe zur Verfügung gestellt, nicht der gesamten Öffentlichkeit.

## 4.6 Zusammenfassung

Dieser Entwurf beinhaltet alles, um eine Signalausbreitung des WLANs mit Hilfe von Smartphone-Messungen, Referenz-WLANs und OSM durchzuführen. Dafür wurde eigens eine Android-Anwendung entworfen, die die wichtigsten Daten, wie WLAN-Signalstärke, Position und Sensordaten, aufzeichnet und lokal zwischenspeichert. Daher ist es möglich, diese Messdaten auf einen Server hochzuladen und einer Datenbank hinzuzufügen. Deshalb wurde eine einheitliche Datenstruktur auf dem Smartphone und in der Datenbank entwickelt. Der Rasteralgorithmus ermöglicht die Erstellung der Referenzpunkte mit Hilfe von OSM. Schließlich kann dies auf jegliches Gebiet, das in OSM kartografiert ist, angewendet werden, sodass die Rasterdatenbank mit Referenzpunkten gefüllt werden kann und Messungen aus diesem Gebiet zu den Referenzpunkten aggregiert werden können. Die Aggregation wird mit Hilfe des Partikelfilters und des Map Matchings realisiert. Aus der Rasterdatenbank können die Eingaben für das LDM und das MWM berechnet werden, um ein solches Ausbreitungsmodell durchzuführen. Aus diesen Ergebnissen wird die Signalausbreitung für jeden AP anhand von KML visualisiert.

# Kapitel 5

## Implementierung

Die Implementierung beschreibt die Umsetzung des Entwurfs in ein Programm. Hier werden die wichtigsten Konzepte und deren Funktionsweise dargelegt. Die drei Unterkapitel sind in Module, Android und Python aufgeteilt.

Das Unterkapitel ?? Module behandelt die Struktur der Anwendungen. Die Struktur ist in drei Kategorien unterteilt: Die erste Kategorie ist die Android-Anwendung, die die Messungen auf dem Smartphone durchführt und später auf den Datenbankserver hochlädt. Dieser Server stellt die zweite Kategorie dar. Die dritte Kategorie entspricht der Anwendung der Algorithmen auf die Daten.

Das zweite und dritte Unterkapitel setzt sich aus der Erläuterung von Android (??) und Python (??) sowie deren genutzte Bibliotheken zusammen.

### 5.1 Module

Wie schon in vorherigen Kapitel 4.1 Entwurf aufgezeigt, sollen die Messungen auf einem mobilen Gerät, einem Smartphone, durchgeführt werden. In jenem Kapitel wurde auch die Begründung für die Wahl des Betriebssystems Android dargelegt. Im folgenden Unterkapitel ?? wird die Implementierung der Android-Anwendung beschrieben. Es wird erläutert, wie Messungen vorgenommen und diese später auf einem Server hochgeladen werden.

Im nächsten Unterkapitel ?? wird die Umsetzung der Serveranwendung dargelegt. Der Schwerpunkt ist die Aufnahme der Smartphone-Messdaten in die Datenbank.

Im letzten Unterkapitel ?? wird das Modell beschrieben, das aus der Implementierung des Partikelfilters, dem Map Matching und dem Ausbreitungsmodell besteht.

#### 5.1.1 Android-Anwendung

Die Hauptprogrammiersprache für eine Anwendung unter Android sowie für Android selber ist Java. Java ist objektorientiert und plattformunabhängig. Der Programmcode wird in einen Java-Bytecode umgewandelt, der von einer virtuellen Maschine

ausgeführt wird, die diesen Maschinencode auf die Hardware umsetzt. Der Vorteil besteht darin, dass sich viele Bibliotheken aus der Java-Umgebung auf Android umsetzen lassen.

Die Anwendung besteht aus zwei Komponenten. Die erste Komponente, das Activity, beinhaltet das User Interface und dient als Steuerung für einen Messvorgang sowie für die Ausgabe einer Messung und das spätere Hochladen aller gespeicherten Messungen.

Die zweite Komponente, der Service, führt die Messungen aus, indem er die aktuelle Position aus der Klasse *LocationManager* abfragt sowie gleichzeitig einen Wi-Fi-Scan aus der Klasse *WifiManager* durchführt und die Abfrage der Sensoren aus *SensorManager* gewährleistet. Dieser Service ist unabhängig vom User Interface und läuft somit im Hintergrund. Dies ermöglicht den Benutzern, diese Anwendung zu schließen und ihr Smartphone anderweitig zu gebrauchen. Währenddessen werden weiterhin Messungen im Service aufgenommen, solange der Anwender diese nicht unterbricht. Dieser Messvorgang wird jede Sekunde automatisch durchgeführt, alle Messdaten werden zusammengefasst und mit Hilfe von Gson in dem JSON-Format auf der SD-Karte abgespeichert. Die externe Bibliothek Gson wird in Kapitel ?? genauer erläutert. Das JSON-Format baut auf der Datenstruktur der Scandatenbank in Kapitel 4.2 auf, sodass nach dem Hochladen der Daten auf dem Server die JSON-Daten geparkt und in die Datenbank eingefügt werden. Das Auslesen der JSON-Daten von der SD-Karte sowie das Hochladen dieser Daten per HTTP-Post werden mit Hilfe der Funktionen *FileEntity* und *DefaultHttpClient* ausgeführt. Diese beiden Funktionen sind in der Bibliothek Apache HttpComponents enthalten und in Kapitel ?? beschrieben. Nach dem Hochladen werden die Daten auf der SD-Karte gelöscht. Der Anwender hat die Möglichkeit, die Daten eines Tracks einzeln zu löschen, und kann somit frei wählen, ob er diese Daten vernichten oder dem Server zur Verfügung stellen möchte.

Schlussfolgernd hat sich Android als eine adäquate Lösung für Messanwendungen herausgestellt. Es bietet alle wichtigen Funktionen in seinen Bibliotheken an, um diese Anwendungen umzusetzen. Als problematisch erwies sich allerdings, dass manche Hersteller die Umsetzung der Smartphones nicht nach der Android-Dokumentation realisiert hatten. So stimmte die Rückgabe von Funktionen nicht mit der Dokumentation überein. Das Samsung Nexus S gab z. B. bei der Funktion *WifiManager.getScanResults()*<sup>1</sup> null aus, wenn kein WLAN gefunden wurde, anstatt eine leere Liste auszugeben. Dies führte dazu, dass die Messanwendung bei manchen Smartphones abstürzte, obwohl sie bei anderen einwandfrei funktionierte. Um dieses Problem zu lösen, wurden die Smartphones daher vom Entwickler eingehend getestet. Bei einem Absturz wurde der Fehler gesucht und unmittelbar behoben.

## 5.1.2 Server

Der Server besteht aus zwei Kernelementen: CherryPy und SQLite. CherryPy ist ein minimalistischer HTTP-Server, der in Kapitel ?? genauer erklärt wird. Der Anwender startet das Hochladen seiner Messdaten. Die Webanwendung von CherryPy nimmt die Messdaten per Multipart HTTP-Post vom Smartphone entgegen. Diese

<sup>1</sup>Android Funktion *WifiManager.getScanResults()*, [http://developer.android.com/reference/android/net/wifi/WifiManager.html#getScanResults\(\)](http://developer.android.com/reference/android/net/wifi/WifiManager.html#getScanResults())

JSON-Daten werden geparkt und in die Datenbank eingefügt. Die Webanwendungen werden unter anderem auch dazu genutzt, eine Homepage bereitzustellen, die einerseits den Anwendern Informationen über diese Arbeit zur Verfügung stellt und andererseits eine Anleitung zur Installation der Android-Anwendung beinhaltet, um durch Messungen das Projekt zu unterstützen. Um den Anwendern einen Anreiz zu verschaffen, werden Statistiken über alle Messungen veröffentlicht.

Das zweite Kernelement ist die SQLite-Datenbank. Es handelt sich um eine relationale Datenbank, die zum Teil auf dem SQL-Standard beruht. Die SQLite-Bibliothek ist gemeinfrei; sie fällt somit unter kein Urheberrecht. Diese Bibliothek eignet sich für Prototyping, sodass sie später ohne großen Aufwand auf andere Datenbanksysteme wie PostgreSQL und Oracle umgesetzt werden kann, die für größere Datenmengen konzipiert sind. Diese Bibliothek befindet sich in der Standardausstattung von Python und nennt sich *sqlite3*. Die Datenbank auf dem CherryPy-Server besteht aus der Datenbankstruktur, die näher in dem Kapitel 4.2 beschrieben ist. Es handelt sich um die erwähnte Scandatenbank, welche die Messungen der Smartphones zentral speichert.

Der Server läuft auf Ubuntu 10.10 mit Python 2.6.6. Da alle Anwendungen auf dem Server komplett auf Python basieren, sind sie plattformunabhängig.

### 5.1.3 Modell

Der nächste Schritt besteht aus der Datenverarbeitung der Scandatenbank, die im vorherigen Kapitel ?? erklärt ist. Die Verarbeitung ist in drei Elemente, Rastermodell, Partikelfilter und Ausbreitungsmodell, eingeteilt.

Im Rastermodell, dem ersten Element, erstellt der Rasteralgorithmus die Rasterdatenbank anhand von OSM. Die Erläuterung zum Rasteralgorithmus befindet sich in Kapitel 4.3.1. Die Rasterdaten bestehen aus SQLite, wie die Scandatenbank ebenfalls. Deren Datenbankstruktur wird in Kapitel ?? beschrieben. Die Geodaten von OSM werden mit Hilfe von Imposm, siehe Kapitel ??, geparkt. Dort werden die Straßen und freien Flächen entnommen, auf denen der Rasteralgorithmus angewendet wird. Als Ergebnis entstehen die Referenzpunkte, die der Rasterdatenbank hinzugefügt werden.

Das zweite Element ist das Partikelfilter. Die Messungen werden aus der Scandatenbank ausgelesen und als Eingabe in dem Partikelfilter ausgeführt. Dieses Filter hat das Ziel, die Abweichung der Positionsmessung zu minimieren. Die genaue Beschreibung ist in Kapitel 4.3.3 zu finden. Gleichzeitig wird ein Map Matching ausgeführt, bei dem die Messungen den Referenzpunkten zugewiesen werden. Der Rasteralgorithmus und das Partikelfilter aus Kapitel 4.3.1 sind komplett in Python geschrieben. Als externe Bibliotheken wurden NumPy und SciPy eingesetzt, da die Berechnung kürzere Laufzeiten im Vergleich zu Python hat. Matplotlib wurde für statistische Darstellungen wie Histogramme genutzt und die visuelle Darstellung der Positionen auf OSM wurde anhand des heatmap-Skripts erstellt. Beide wurden als visuelle Verifikation des Algorithmus genutzt. Diese Bibliotheken und Skripte sind in Kapitel ?? näher erläutert.

Das dritte Element ist das Ausbreitungsmodell sowie die Visualisierung. Das LDM ist nach der Formel in Kapitel 2.3 in Python programmiert. Die Eingaben der Berechnung werden der Rasterdatenbank entnommen und sind in Kapitel ?? erläutert. Für das Multi Wall Model wird die Bibliothek shapely (siehe ??) genutzt. Die Geodaten

von Gebäuden aus OSM werden als Polygone mit shapely erstellt. Damit werden die Polygone überprüft und es wird festgestellt, ob ein Signal ein Gebäude schneidet. Diese Algorithmen und die folgenden Schritte sind auf dem Server ausführbar, können aber auch auf einem anderen PC ausgeführt werden, solange man Zugriff auf die Scandatenbank hat. Meist wurden sie für die vorliegende Arbeit auf einem MacBook mit Ubuntu 11.04 durchgeführt. Auf einem Windows 7 Computer mit Portable Python<sup>2</sup> wurden sie ebenfalls mehrfach getestet. Portable Python ist eine portable Version von Python für ein USB-Speichermedium. Der Vorteil ist, dass die benötigten Bibliotheken NumPy, SciPy und matplotlib vorkonfiguriert mitinstalliert werden.

## 5.2 Android

Android ist ein Betriebssystem für mobile Geräte wie Smartphones, Tablets usw. Es basiert auf dem Linux-Kernel und ist unter den Lizenzen Apache 2.0 und GPL v2 zugelassen. Dieses System ist quelloffen, das heißt dass der Quellcode jedem zugänglich ist. Für die Entwicklung von Android-Anwendungen stehen die Programmiersprachen Java und C bzw. C++ zur Verfügung. Eine Anwendung selber besteht meistens aus Java. Zeitaufwendige Berechnungen können in C oder C++ programmiert werden. Die vorliegende Arbeit basiert für die Android-Anwendung vollständig auf Java.

Neben den internen Bibliotheken Java und Android für die Android-Anwendung werden die externen Bibliotheken Gson und Apache HttpComponents importiert.

### Gson

Gson<sup>3</sup> ist eine Java-Bibliothek von Google und überführt Java-Objekte in JavaScript Object Notation (JSON), bzw. wandelt JSON in Java-Objekte um. Diese Bibliothek läuft unter der Apache License 2.0.

Die Wahl richtete sich gegen die interne Java-Bibliothek org.json, da diese die Messungen nicht phasenweise auf die SD-Karte speichert. Sie muss die Daten erst im Arbeitsspeicher sammeln, ehe sie sie dann als Gesamtpaket auf die SD-Karte schreibt. Dies hat für diese Arbeit zwei Nachteile: Zunächst kann das Betriebssystem die Anwendung wegen des hohen Arbeitsspeicherverbrauchs auffordern, sich zu beenden. Schließlich können alle Daten bei einem Absturz verloren gehen.

Die Bibliothek Gson bietet diesen phasenweisen Schreibvorgang an.

### Apache HttpComponents

Die nächste externe Bibliothek ist die Apache HttpComponents<sup>4</sup> Bibliothek, die die wichtigsten HTTP-Komponenten vereint. Auch diese Bibliothek läuft unter der

---

<sup>2</sup>Portable Python, <http://www.portablepython.com/>

<sup>3</sup>Gson, <http://code.google.com/p/google-gson/>

<sup>4</sup>Apache HttpComponents, <http://hc.apache.org/>

Apache License 2.0. Mit Hilfe der Bibliothek können wir Nachrichten über HTTP-Protokoll senden und empfangen. Apache wird für die Kommunikation zwischen dem Smartphone und dem Server genutzt. Die Stärke dieser Bibliothek kommt vor allem beim Hochladen von größeren Dateien zum Tragen. Mit der Funktion *FileEntity* wird die Datei phasenweise eingelesen und über einen Multipart HTTP-Post gesendet. Größere Dateien müssen nicht komplett in den Arbeitsspeicher eingelesen werden, um gesendet zu werden.

## 5.3 Python

Die komplette Server-Architektur ist in Python geschrieben. Python ist eine höhere universelle Programmiersprache, deren Ziel es ist, dem Quellcode eine höhere Lesbarkeit zu bieten. Es gibt mehrere Programmierparadigmen wie objektorientierte, imperative und funktionale Programmierung. Die Stärken liegen in der voll dynamischen Sprache und in dem automatischen Speichermanagement. Der Nachteil durch das voll dynamische System besteht in der längeren Ausführungszeit im Vergleich zu anderen Programmiersprachen. Neben den internen Bibliotheken von Python wurden die folgenden externen Bibliotheken importiert.

### CherryPy

CherryPy<sup>5</sup> ist ein minimalistisches Python Web Framework, welches als HTTP-Server fungiert. Damit ist es möglich, ein Python-Programm als Webanwendung in kürzester Zeit umzusetzen. CherryPy läuft unter der BSD Lizenz. Da CherryPy in Python programmiert ist, übernimmt es auch deren Nachteile. Dies sind einerseits die schlechte Performance und andererseits daraus folgend die Begrenzung der maximalen gleichzeitigen Seitenzugriffe auf den Server. Für die vorliegende Arbeit sind diese Nachteile jedoch nicht relevant, da nur eine geschlossene Gruppe von Anwendern auf den Server zugreift.

### NumPy/SciPy/matplotlib

Diese drei Pakete bilden eine größere Bibliothek für höhere mathematische Funktionen. NumPy<sup>6</sup> kümmert sich um große mehrdimensionale Datenstrukturen, die als Felder und Matrizen nutzbar sind. Die Stärke von NumPy ist, dass sie schneller ausführbar ist als eine vergleichbare Berechnung der Datenstruktur in Python. NumPy ist wie SciPy<sup>7</sup> unter der BSD Lizenz zugelassen.

SciPy beinhaltet die mathematischen Berechnungen auf den Datenstrukturen von NumPy. SciPy basiert wie NumPy auf der Programmiersprache C und bietet demnach eine schnelle Performance.

Matplotlib<sup>8</sup> ist eine Bibliothek, die die Daten von der Datenstruktur von NumPy

---

<sup>5</sup>CherryPy, <http://www.cherrypy.org/>

<sup>6</sup>NumPy, <http://numpy.scipy.org/>

<sup>7</sup>SciPy, <http://scipy.org/>

<sup>8</sup>Matplotlib, <http://matplotlib.sourceforge.net/>

in einen Plot umgesetzt. Dadurch können die Daten graphisch, z. B. als Histogramm, dargestellt werden. Sie ist unter der matplotlib Lizenz veröffentlicht, die der BSD Lizenz ähnelt.

Es gibt keine vergleichbaren Pakete, die über den gleichen Umfang und die gleiche Effizienz verfügen.

## shapely

Shapely<sup>9</sup> ist eine Bibliothek für die Manipulation von planaren geometrischen Objekten. Diese geometrischen Objekte basieren unter anderem auf den geometrischen Elementen Point, Line, Polygone, wie sie auch in OSM vorhanden sind. Shapely ist nicht auf ein festgelegtes Koordinatensystem und auf ein Datenformat beschränkt. Sie läuft unter der BSD Lizenz, wurde in C programmiert und bietet somit schnelle Laufzeiten.

## Imposm

Imposm<sup>10</sup> ist eine Bibliothek, um Geodaten von OSM in Form von XML in eine Datenbank zu überführen. Wir nutzen nur einen Teil dieser Bibliothek, den *imposm.parser*, womit sich die OSM XML leicht parsen lässt. Die Stärke liegt darin, dass die Daten automatisch nach einem Kriterium gefiltert werden. Dies ermöglicht es, die Geodaten auf das Nötigste zu beschränken, wie Straßen, Plätze und Häuser. Sie ist unter der Apache License 2.0 veröffentlicht.

## kmlcircle

KML bietet keine Funktion an, die ein Kreiselement zeichnet. Die einzige Möglichkeit für die Darstellung eines Kreises in KML ist ein über 20-seitiges Polygon. Kmlcircle<sup>11</sup> ist ein Algorithmus, der anhand einer Position und eines Radius das über 20-seitige Polygon berechnet und in KML abspeichert. Es basiert auf der MIT Lizenz.

## heatmap

Heatmap<sup>12</sup> ist ein Python-Skript, das eine Position mit Latitude- und Longitude-Koordinaten auf eine Karte plottet. Im Vergleich zu KML sind die Ausgaben hier pixelbasiert. Das Kartenmaterial kann von OSM ausgewählt werden. Eine Liste von Koordinaten lässt sich als Video, das aus aufeinander folgenden Kartenbildern besteht, darstellen. Das Heatmap-Skript basiert auf der GNU General Public License in Version 3. Dieses Skript nutzen wir, um die Ergebnisse eines Algorithmus visuell darzustellen und anhand dieser Darstellung zu verifizieren. Die vier Abbildungen in dem Kapitel ?? wurden mit diesem Skript erstellt.

---

<sup>9</sup>Shapely, <http://toblerity.github.com/shapely/>

<sup>10</sup>Imposm, <http://imposm.org/>

<sup>11</sup>Kmlcircle, <http://code.google.com/p/kmlcircle/>

<sup>12</sup>Heatmap, <http://sethoscope.net/heatmap/>

## 5.4 Zusammenfassung

In diesem Kapitel wurde die Struktur der Module der Android-Anwendung, des Datenbankservers und der Algorithmen erklärt. Das Android-Betriebssystem eignet sich gut für eine Messanwendung. Alle wichtigen Funktionen sind in den Android-Bibliotheken vorhanden.

Als problematisch gestaltete sich, dass manche Hardwarehersteller das Betriebssystem auf ihren Smartphones nicht nach der Android-Dokumentation umgesetzt hatten. Daher musste eine Fehlerbehebung bei manchen Geräten durchgeführt werden, obwohl diese Fehler bei anderen Geräten gar nicht erst auftraten. Sie funktionierten einwandfrei.

Die Python-Sprache bietet eine schnelle Umsetzung und eignet sich daher für das Prototyping.

Die Bibliotheken Numpy und Scipy benötigen eine kürzere Laufzeit, besonders im Vergleich zu einer Programmierung in Python. Gerade die Vielfalt der Bibliotheken in Python hat eine große Zeitersparnis bei dem hohen Programmieraufwand erbracht.



## Evaluation

Das Kapitel Evaluation ist in die folgenden vier Unterkapitel Szenario, Messungen, Positionierung und Ausbreitungsmodell aufgeteilt. In dem Unterkapitel ?? Szenario wird zunächst das Gebiet erklärt, in welchem ein großer Teil der Messungen durchgeführt wird. Die Messungen werden anschließend auf ein kleineres Gebiet begrenzt, um den Fokus und damit auch die Auswertung der Messungen auf ein abgestecktes Gebiet genauer legen zu können.

Im Unterkapitel 6.2 werden alle Messungen, also auch diejenigen, die nicht nur auf das Szenario beschränkt sind, in der Scandatenbank ausgewertet. Diese Evaluation soll zeigen, wie repräsentativ die Scandatenbank ist. Dazu werden die RSSI-Werte, die GPS-Accuracy und die Frequenzen betrachtet. Es wird festgestellt, in welchem Bereich diese liegen. Diese Auswertungen werden für das Definieren von Schranken wie beim Partikelfilter benötigt.

Das Unterkapitel ?? Positionierung beinhaltet die Auswertung der Anzahl Messungen pro Referenzpunkt. Es stellt die Repräsentativität der Rasterdatenbank dar. Die Evaluation beschreibt die Abweichungen von den Positionsdaten, dem nächsten Nachbarn und dem Partikelfilter im Vergleich zu den richtigen Positionen. Die Daten mit der kleinsten durchschnittlichen Abweichung stellen das beste Ergebnis dar, welches der richtigen Position am nächsten kommt.

Das Ausbreitungsmodell in dem Unterkapitel ?? erläutert den Nachweis, dass der Pfadverlust aus den Messungen erkennbar ist. Beim Referenz-Pfadverlust werden 100 WLAN-Messungen von einem AP im Abstand von einem Meter durchgeführt. Dabei werden die Abweichung von den 100 Messungen sowie die Abweichung von vier verschiedenen Smartphones ausgewertet.

Zuletzt wird das LDM berechnet und mit der Kurvenanpassung verglichen. Damit soll bewiesen werden, dass keine komplette Abdeckung für die Signalausbreitung mit Hilfe der Kurvenanpassung gebraucht wird, sondern die Berechnung der Signalausbreitung ausreichend ist.

## 6.1 Szenario

Am Anfang der Arbeit bestand lediglich die Vorgabe der Beschränkung auf den räumlichen Bereich, sodass die Messungen in Aachen stattfinden. Der Vorteil ist, dass dadurch ein Gesamteindruck von verschiedenen APs dargestellt werden kann. Es bietet die Möglichkeit, die verschiedenen Frequenzen und die Häufigkeit der Signalstärken über ein größeres Gebiet besser zu erfassen, um die repräsentative Darstellung der Varianz zu ermitteln. Ein Nachteil entstand dadurch, dass pro Referenzpunkt gar keine oder nur wenige Messungen vorhanden waren. So ließen sich die Fehler der Positionierung und der WLAN-Scans auf Grund der fehlenden Anzahl Messungen nicht ausreichend minimieren.

Dafür wurde ein Szenario entwickelt, sodass das Gebiet für Messungen eingegrenzt wurde. Das Ziel war es, durch die räumliche Beschränkung die Anzahl der Messungen pro Referenzpunkt zu steigern. Dieses Gebiet befindet sich in der Innenstadt von Aachen und wird von den folgenden Straßen umschlossen, die hier im Uhrzeigersinn aufgeführt werden: Markt, Pontstraße, Driescher Gässchen, Hirschgraben, Seilgraben, Großkölnstraße, Kleinkölnstraße und Büchel. Die Abbildung 6.1 stellt in Blau die Linien und Flächen des Szenarios dar. Auf diesen Flächen und Linien wurden die Referenzpunkte dieses Szenarios erstellt. Dies ist in der Abbildung 4.3 veranschaulicht.

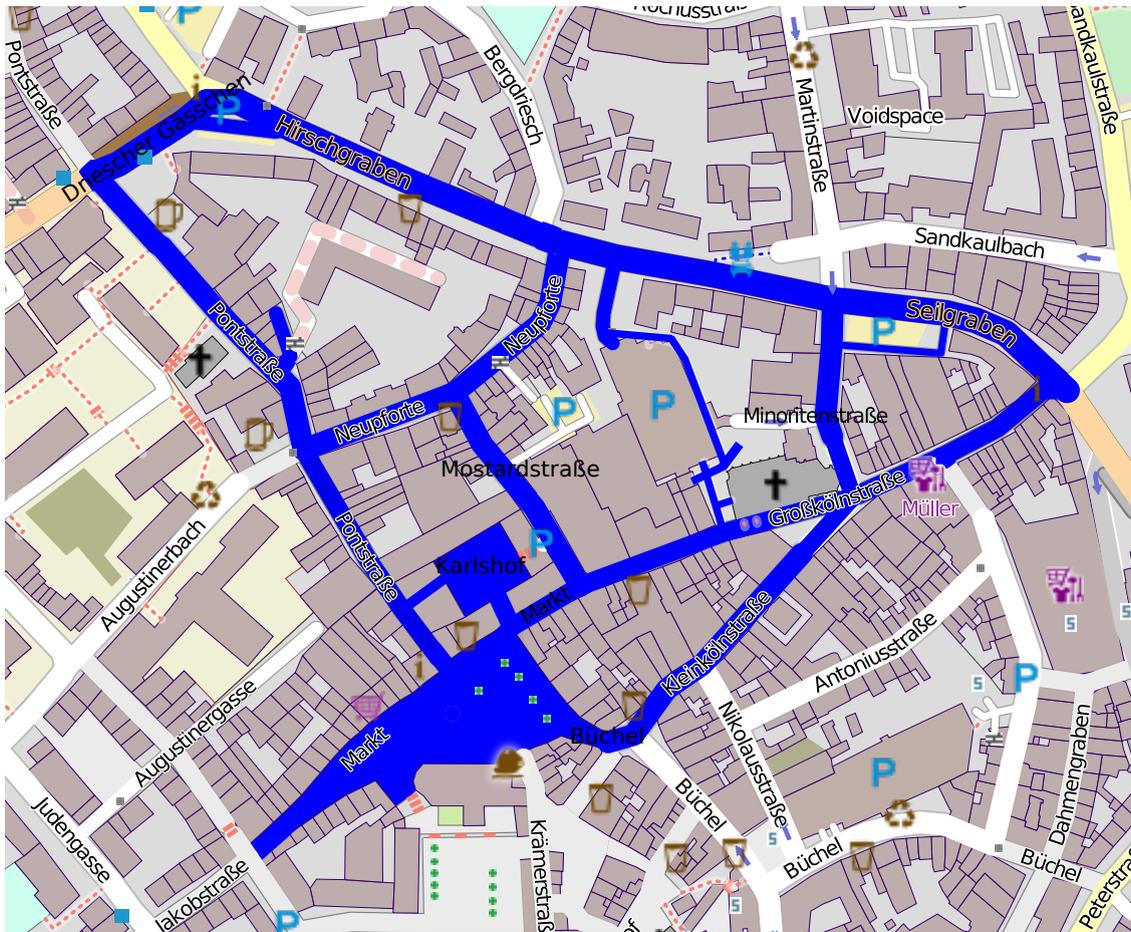
Das Szenario umfasst eine Fläche von 95.555 m<sup>2</sup> und hat einen Umfang von 1.528 m. Die Summe der Längen aller Straßen im Szenario beträgt 2.666 m. Das Szenario ist eine Mischung aus Wohn- und Geschäftsanlage. Der Bereich beinhaltet sehr wenige Hochhäuser. Die meiste Fläche ist zugebaut, nur der Markt sowie der Innenhof Karlsruhof bilden frei zugängliche Flächen. Die Abbildung 6.2 gibt einen Eindruck der Gebäudestruktur des Szenarios anhand eines Satellitenbildes.

Die Anwender wurden darauf hingewiesen, dass sie sich für die Messungen nur im Außenbereich, also auf öffentlichen Straßen und Plätzen, jedoch nicht in Gebäuden aufhalten sollen. Ein Schwerpunkt dieser Arbeit ist, dass die Anwender die Messungen während ihres Alltages durchführen. Sie kümmern sich nicht aktiv um das Smartphone, wenn die Aufzeichnung läuft. Daher wurde den Anwendern keine Anweisung gegeben, wo sie ihr Smartphone aufbewahren sollen. Bei Nachfrage wurden die Anwender angewiesen, ihr Smartphone in ihrer Jacken- oder Hosentasche aufzubewahren. Dies kommt der Handhabung während des Alltages näher, im Vergleich dazu, wenn sie ihr Smartphone während der Messungen in der Hand halten.

In dem Aachener Szenario wurden 55 verschiedene Messungen aufgenommen, die 31.852 Messpunkte der gesamten Datenbank betragen. Es wurden dabei 1.935 verschiedene WLANs entdeckt. Daraus ergibt sich, dass sich in diesem Gebiet durchschnittlich 20 WLANs pro km<sup>2</sup> befinden.

Neben den Messungen sind die Referenz-WLANs eine weitere Eingabequelle für das System. Die Referenz-WLANs bestehen aus einer Liste des Hotspot-Anbieters Öcher WLAN<sup>1</sup> in Aachen. Diese Hotspots werden überwiegend durch das Routermodell Linksys WRT54G zur Verfügung gestellt. In der Standardkonfiguration der Router beträgt die Sendeleistung 16 dBm. Wenn die Kunden die Routerkonfiguration nicht verändern, dann ist mit derselben Sendeleistung und Antennencharakteristik zu rechnen. In dieser WLAN-Liste sind über 2.000 APs mit deren Position als WGS 84-Koordinatensystem, deren Hausadresse und ihrer MAC-Adresse aufgeführt. Auf

<sup>1</sup>Öcher WLAN, <http://www.oecher-wlan.de>



**Abbildung 6.1** Darstellung des Szenarios anhand von blauen Linien und Flächen - Daten von OpenStreetMap, 2012, veröffentlicht unter CC-BY-SA 2.0, vergleichbar mit der Abbildung 4.3



**Abbildung 6.2** Satellitenbild der Szenario-Fläche am Markt, das die Geschäfts- und Wohnanlage dokumentiert und die Gebäudestruktur zeigt - Bild von Bing Microsoft Corporation, 2012

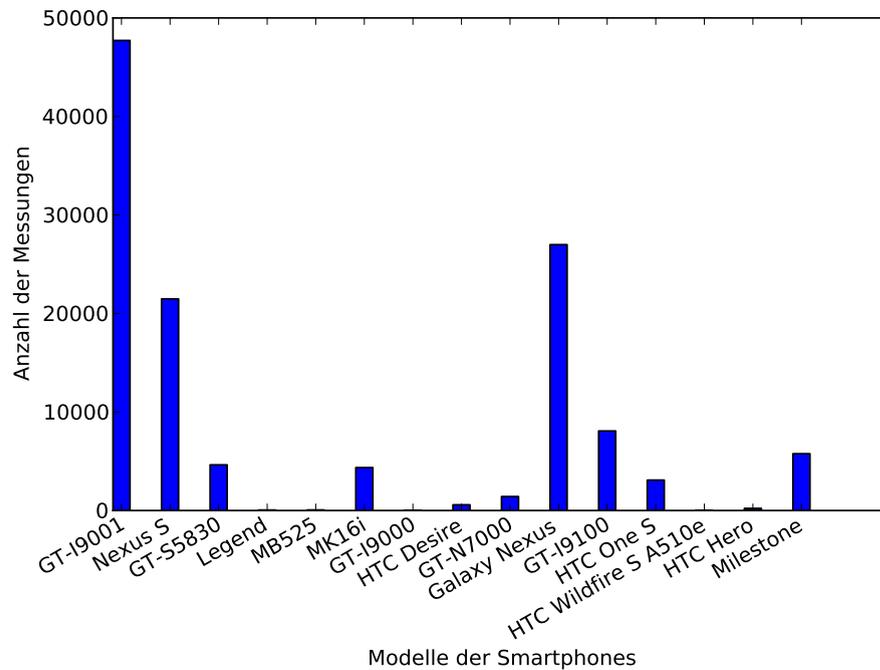
Grund dieser Angaben können die Messungen den APs zugeordnet werden, deren Position bekannt ist. Im Szenario wurden insgesamt 67 verschiedene Referenz-WLANs entdeckt, deren Messungen für das Ausbreitungsmodell genutzt wurden.

## 6.2 Messungen

Um einen Eindruck zu vermitteln, wie repräsentativ die Scandatenbank ist, werden folgende Zahlen mitgeteilt. Alle Daten und Auswertungen beziehen sich dabei auf die gesamte Scandatenbank und somit auch auf Messungen außerhalb des Szenarios. Die Scandatenbank besteht aus 147 Tracks, die 124.452 Messpunkte beinhalten. Insgesamt wurde ein Weg von etwa 338,85 km zurückgelegt. Es wurden 11.825 verschiedene WLANs entdeckt. Diese WLANs sind in über einer Million Messungen aufgezeichnet. 15 verschiedene Smartphones führten die Messungen durch, wovon drei Smartphones 76 % aller Messpunkte ausmachen. Diese Smartphones sind das Samsung Galaxy S Plus (GT-I9001) mit 38 %, das Samsung Galaxy Nexus mit 22 % und das Samsung Nexus S mit 16 % der Messpunkte. Die Abbildung 6.3 stellt die Übersicht der Smartphonemodelle im Vergleich zu der Anzahl der Messpunkte dar. Die Messdaten wurden zu unterschiedlichen Zeitpunkten von Januar 2012 bis Juli 2012 aufgezeichnet und umfassen somit eine Zeitspanne von sieben Monaten. Während der Messungen betrug die durchschnittliche Temperatur  $7,97^\circ$ . Im Januar lag sie bei  $3,9^\circ$  und im Juni ist sie auf  $14,7^\circ$  (siehe Abbildung A.1) gestiegen. Die durchschnittliche Luftfeuchte lag bei 79,62% (siehe Abbildung A.2) und der Mittelwert des Bedeckungsgrades des Himmels betrug 5,6 Achtel (siehe Abbildung A.3). Alle Wetterdaten wurden in Aachen-Orsbach erhoben. Diese drei Faktoren sind als Abbildungen mit einem zeitlichen Verlauf in Anhang A.2 hinzugefügt.

### 6.2.1 RSSI

Nach der Szenariobeschreibung bietet die Auswertung der Messungen interessante Ergebnisse. Für ein Ausbreitungsmodell ist der Verlauf aller empfangenen Signalstärken der APs ein bedeutsamer Punkt zur Evaluierung. Er beschreibt die Minimal- und die Maximalwerte, wo sich die Daten befinden und wie die Daten über diesen Bereich verteilt sind. Die Verteilung der Signalstärken ist in der Abbildung 6.4 dargestellt. Es ist zu erkennen, dass sich die meisten gemessenen Signalstärken in einem Bereich befinden: 83 % aller Messungen liegen zwischen -95 dBm und -75 dBm. Dabei sind der Mittelwert bei -85,5 dBm und der Median bei -87 dBm zu finden. Zusätzlich ist festzustellen, dass es sich um eine rechtsschiefe Verteilung handelt. Der niedrigste Wert ist -104 dBm und der höchste Wert beträgt -22 dBm. Aus diesen Beobachtungen ist ersichtlich, dass die meisten Messungen ein schwaches Signal um -85 dBm aufweisen. Das Signal wird von mehreren Faktoren, wie Distanz, Gebäuden usw., beeinflusst, sodass es abschwächt. Die genauen Faktoren für die Abschwächung sind nicht zu ermitteln. So sind kaum Signalstärken unter -70 dBm wahrnehmbar. Dies stimmt mit der Erwartung überein. Die APs sind meist in Gebäuden positioniert und müssen mindestens eine Wand durchstrahlen. Diese Wand dämpft das Signal und verhindert eine Sichtlinie zwischen dem AP und dem Smartphone. So stimmt



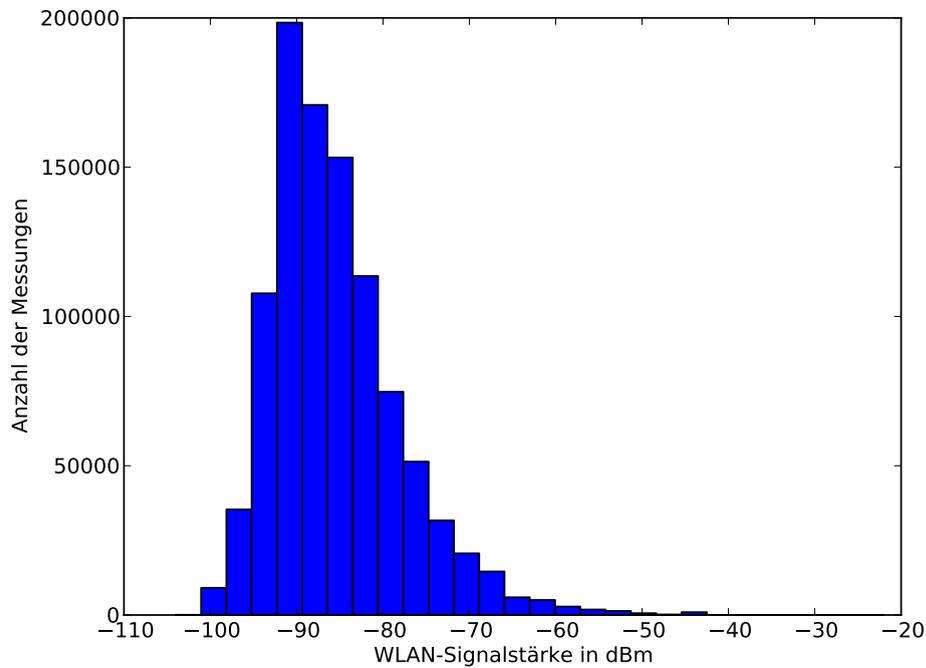
**Abbildung 6.3** Histogramm der Smartphonemodelle im Vergleich zu der Anzahl der Messungen

dies mit der vorher geäußerten Erwartung überein, dass die meisten Messungen eine Schwankung von 20 dB im niedrigen Wertebereich haben.

## 6.2.2 GPS-Accuracy

Neben den WLAN-Scans wurde auch die Positionierung aufgenommen. Die Abbildung 6.5 zeigt ein Histogramm der Accuracy-Werte bis 90 m, das in GPS und Network eingeteilt ist. Die Positionierung über GPS ist in Kapitel 2.3.1 ausführlich erklärt. Die Positionierung über Network wird in Android über die Lokalisierung der Signale der GSM-Zellen und der WLAN-Signale ermittelt. In der Android-Dokumentation<sup>2</sup> wird Network als zumeist ungenauere Lokalisierung im Vergleich zu GPS beschrieben. Dies lässt sich auch anhand der Accuracy-Werte belegen. Der Mittelwert aller GPS-Positionen liegt bei 18 m, wobei er mittels Network bei 349 m liegt. Dies spricht für die GPS-Positionen. Jedoch sollten die Network-Positionen nicht komplett fallen gelassen werden, sondern nach dem Accuracy-Wert die Güte der Position klassifizieren. Vor allem da der Median von Network 60 m beträgt, würden viele Messungen wegfallen. Daher sollte nach einer oberen Schranke gesucht werden, die unter anderem auch für  $a_{bound}$  in Kapitel 4.3.3.1 genutzt wird. Die Verteilung  $P(X < a_{bound})$  wird betrachtet und  $a_{bound}$  so gewählt, dass sie einen großen Bereich aller Accuracy-Werte des GPS und des Networks abdeckt. Mit GPS kann sichergestellt werden, dass die Messungen unter dem freien Himmel erfolgten und

<sup>2</sup>Android Location Strategies, <http://developer.android.com/guide/topics/location/strategies.html>



**Abbildung 6.4** Histogramm der WLAN-Signalstärken aller Messungen

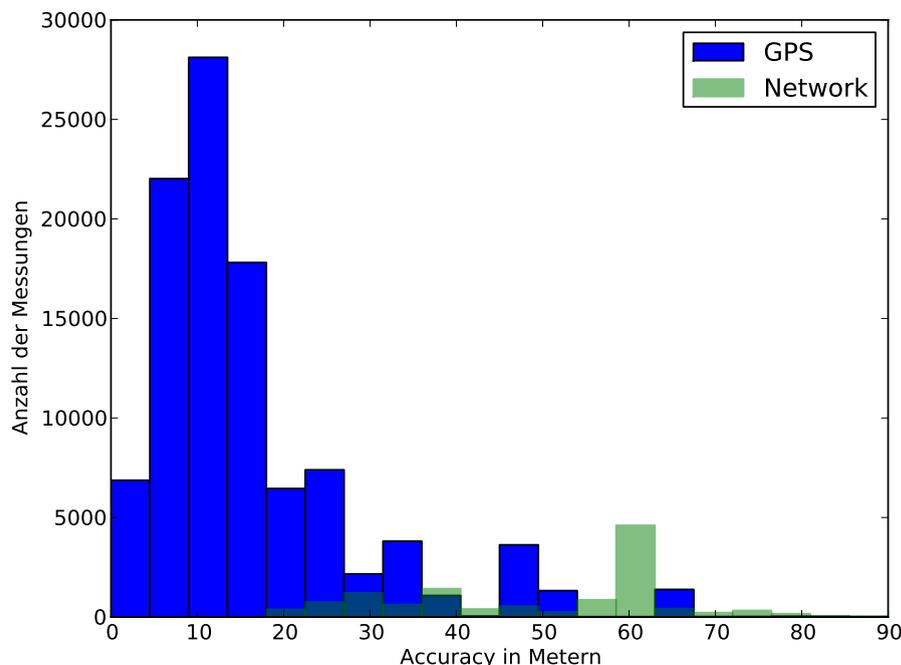
einen niedrigeren Accuracy-Wert im Vergleich zu dem Network haben. Außerdem stehen mehr GPS-Positionen, bei 84 % positioniert, als Network-Positionen, auf 16 % festgelegt, zur Verfügung. Es wird der Wert 70 für die Schranke  $a_{bound}$  gewählt. Damit liegen die Werte in der Verteilung  $P(X < 70)$  zu 98 % der GPS-Positionen im Vergleich zu der Gesamtverteilung  $P(X)$  und bieten somit eine gute Abdeckung aller GPS-Positionen.

Im Map Matching, in Kapitel 4.3.4 erläutert, wurde ein Initialisierungswert  $dist_1$  für die erste Iteration gesucht. Diese sollte einen größeren Teil der GPS-Positionen beinhalten, sodass mehrere Iterationen vermieden werden. Andererseits sollte die Schranke nicht zu groß sein, sodass bei einer Iteration zu viele Berechnungen durchgeführt werden. Daher wird beschlossen, den Wert von 30 m zu nehmen. Somit liegt die Verteilung  $P(X < 30)$  bei 86 % der gesamten Verteilung  $P(X)$ .

### 6.2.3 Frequenzen

Die Häufigkeit der genutzten WLAN-Kanäle ist auch ein interessanter Punkt hinsichtlich der Signalausbreitung. Dies wird in der Abbildung 6.6 dargestellt. Anstatt einer Angabe in Kanälen wird sie in den passenden Frequenzen angezeigt. Das Markanteste ist die stark unterschiedliche Verteilung im 2,4 GHz- und 5 GHz-Bereich. 99 % aller WLAN-Messungen liegen in der Frequenz 2,4 GHz. Dies hat zwei Gründe: Der erste Grund ist die geringe Anzahl von APs, die mit dem veralteten 802.11a- oder dem ganz neuen 802.11n-Standard arbeiten.

Der zweite Grund ist geringe Anzahl von Smartphones, die diese beiden Standards unterstützen und diese somit scannen können. Von den 15 Smartphones unterstützen drei Smartphones den 802.11a-Standard und elf Smartphones den 802.11n-Standard.



**Abbildung 6.5** Histogramm der Accuracy-Werte von GPS und Network bis 90 m

Da gerade nur 1 % der Messungen im 5 GHz stattfinden, werden diese Messungen herausgefiltert. Daher wird der Pfadverlust über einem Meter nur für den 2,4 GHz-Bereich bestimmt. Ein anderer Schwerpunkt in dem Histogramm sind die Frequenzen 2.412 MHz, 2.437 MHz und 2.462 MHz, die dreimal häufiger auftreten als die anderen Frequenzen. Dies liegt daran, dass viele Anwender die überlappungsfreien Kanäle auswählen. Es bietet den Anwendern einen störungsfreien Empfang von WLAN-Signalen. Dieser Hinweis ist in Kapitel 2.1 beschrieben.

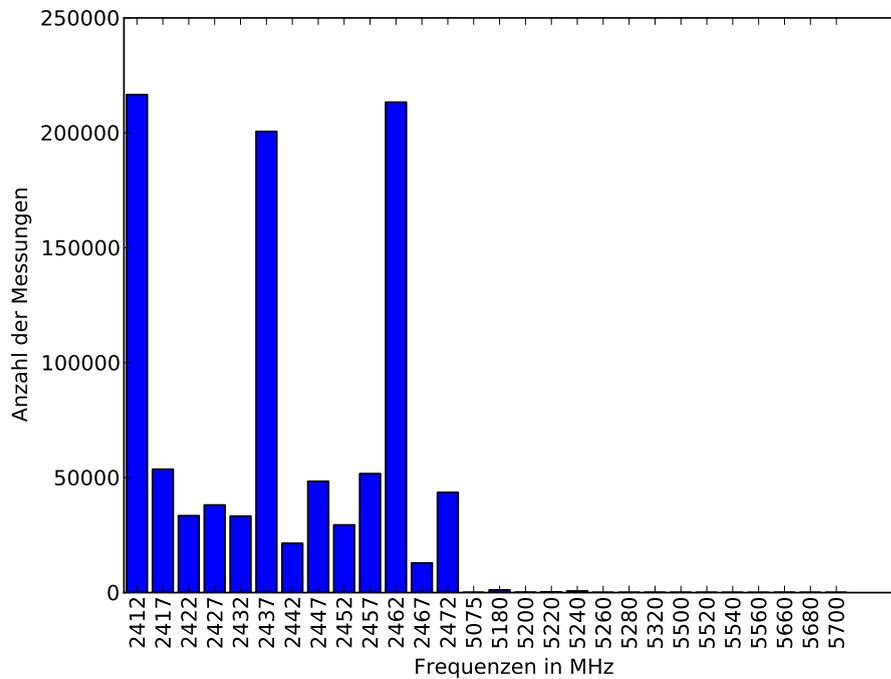
## 6.3 Positionierung

Bei der Positionierung wird die Lokalisierungsabweichung der Messungen mit Hilfe des Partikelfilters minimiert und das Map Matching auf die Referenzpunkte aggregiert. Daher ist die folgende Auswertung in zwei Unterkapitel aufgeteilt. Das erste Unterkapitel ?? beschäftigt sich damit, wie aussagekräftig die Referenzpunkte sind. Dabei wird die Anzahl Messungen, die sich auf einem Referenzpunkt befinden, ermittelt.

Das zweite Unterkapitel ?? befasst sich mit dem Vergleich von Positionsdaten, dem nächsten Nachbarn und dem Partikelfilter zu den richtigen Positionen. Dabei soll festgestellt werden, ob eine Verbesserung der Algorithmen zu den richtigen Positionen entsteht.

### 6.3.1 Messung pro Referenzpunkt

In Kapitel 6.2 wurde dargelegt, wie aussagekräftig die Scandatenbank ist. Hier wird nun die Aussagekraft der Rasterdatenbank anhand der Anzahl Messungen pro Re-



**Abbildung 6.6** Histogramm der WLAN-Frequenzen im Vergleich zu der Anzahl der Messungen

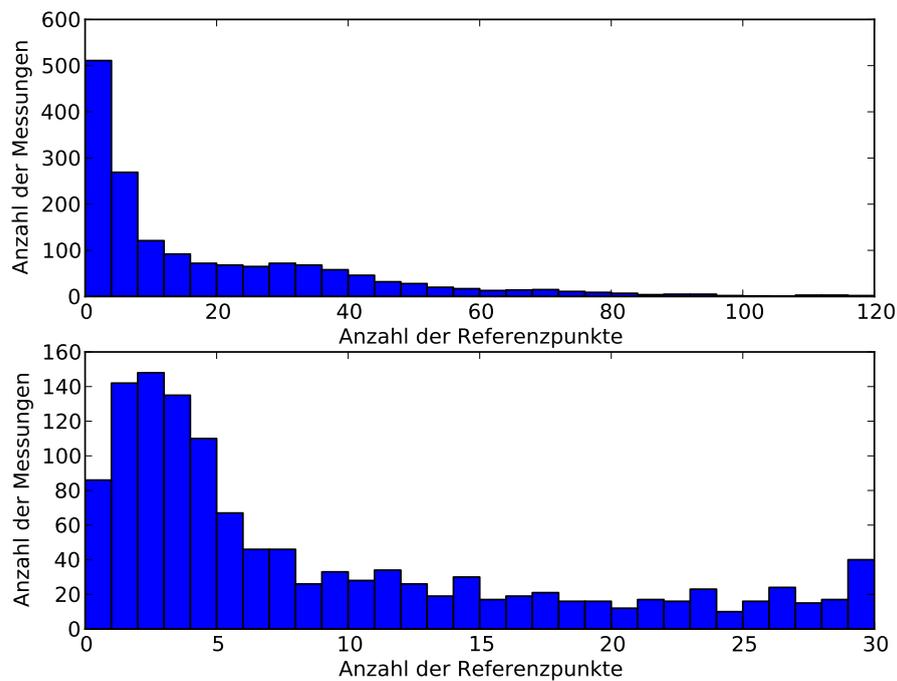
ferenzpunkt dargestellt. Dabei werden nur die Referenzpunkte aus dem Szenario berücksichtigt. Wie schon in der Szenariobeschreibung (siehe Kapitel 6.2) geschildert, ist es das Ziel, pro Referenzpunkt eine hohe Anzahl von WLAN-Messungen zu erreichen, jedoch nicht die komplette Abdeckung der Stadt Aachen zu erwirken. Je mehr Messungen pro Referenzpunkt vorhanden sind, desto eher ist es möglich, die zufällige Abweichung der WLAN-Signalstärken zu minimieren.

Die Abbildung ?? zeigt auf, wie viele Messungen pro Referenzpunkt existieren. Diese Abbildung ist in zwei Histogramme eingeteilt. Das obere Histogramm stellt den gesamten Raum dar. Das untere Histogramm fokussiert die Anzahl der Referenzpunkte auf den Bereich von 0 bis 30. Daraus lässt sich Folgendes schlussfolgern: Der Mittelwert ist 19. Der Median liegt bei 9. Somit befinden sich 50 % der Referenzpunkte vor und 50 % nach dem Wert 9.

Die Messungen, die im Szenario durchgeführt wurden, können eine höhere Abweichung bei der Position haben. Somit werden diese mit Hilfe des Partikelfilters auf die Nachbarstraßen gesetzt, welche außerhalb des definierten Szenarios liegen. So werden einige Messpunkte pro Referenzpunkt nicht mitgezählt, wodurch diese Anzahl niedriger scheint, als sie in Wirklichkeit ist.

### 6.3.2 Positionierungsfehler

In diesem Kapitel folgt der Vergleich der Algorithmen, um die Messabweichungen zu minimieren und das Map Matching adäquat durchzuführen. Dafür werden die gemessenen Positionsdaten, der nächste Nachbar und das Partikelfilter miteinander verglichen. Beim nächsten Nachbarn werden die gemessenen Positionen zu den nächstliegenden Referenzpunkten gesetzt. Es handelt sich nicht um ein Filter wie beim



**Abbildung 6.7** Histogramm der Anzahl Scans pro Referenzpunkt

Partikelfilter (siehe Kapitel 4.3.3), sondern um einen Map Matching-Algorithmus. Durch den Vergleich der Daten wird ermittelt, wie sich die gefilterten Positionen zu den gemessenen Positionen verhalten. Es ist jedoch nicht nachvollziehbar, ob sich die festgestellte Veränderung der wahren Position annähert oder ob sie sich von dieser entfernt. Dafür werden die richtigen Positionen bestimmt, um so einen angemessenen Vergleich der Daten durchzuführen. Es handelt sich um die wahre Position mit einer sehr geringen Abweichung, die jedoch für diesen Vergleich unerheblich ist. Ein Kriterium besteht darin, die richtigen Positionen zu ermitteln, sodass sie unabhängig von den gemessenen Positionen erstellt werden. Es soll vermieden werden, dass die gleiche Abweichungscharakteristik wie bei den Messungen entsteht.

Die Positionsmessungen werden jeweils in einer festgelegten Distanz aufgezeichnet. Dieser Abstand wird mit einem analogen Messgerät, einem Messrad, bestimmt. Über vordefinierte Orientierungspunkte werden bestimmte Messpunkte sowohl in der Wirklichkeit, als auch auf der OSM-Karte bestimmt. Somit können diese Messungen aus der Wirklichkeit auf Grund der Orientierungspunkte auf die Karte gesetzt werden. Alle Messungen zwischen den Orientierungspunkten werden in demselben Abstand gemessen. Damit werden diese Messungen gleichmäßig zwischen zwei Orientierungspunkten auf den Referenzpunkt gesetzt.

Ein neues Szenario wurde hierfür bestimmt und eine Route festgelegt: von dem Ausgangspunkt der Kreuzung Pontstraße-Neupforte, über das Driescher Gässchen, den Hirschgraben und die Neupforte zurück zum Ausgangspunkt der Kreuzung Neupforte-Pontstraße in ebendieser Richtung. Die Strecke beträgt 580 m. Dieser Rundgang wird in der Abbildung ?? (a) verdeutlicht. Folgende Orientierungspunkte wurden auf dieser Strecke bestimmt:

1. Ecke Pontstraße-Neupforte

2. Ecke Pontstraße-Driescher Gässchen
3. Ecke Driescher Gässchen-Hirschgraben: Beginn der Parkeinbuchtung
4. Ecke Driescher Gässchen-Hirschgraben: Ende der Parkeinbuchtung
5. Ecke Hirschgraben-Neupforte
6. Ecke Neupforte-Mostardstraße
7. Ecke Neupforte-Pontstraße (entspricht dem Ausgangspunkt, Punkt 1)

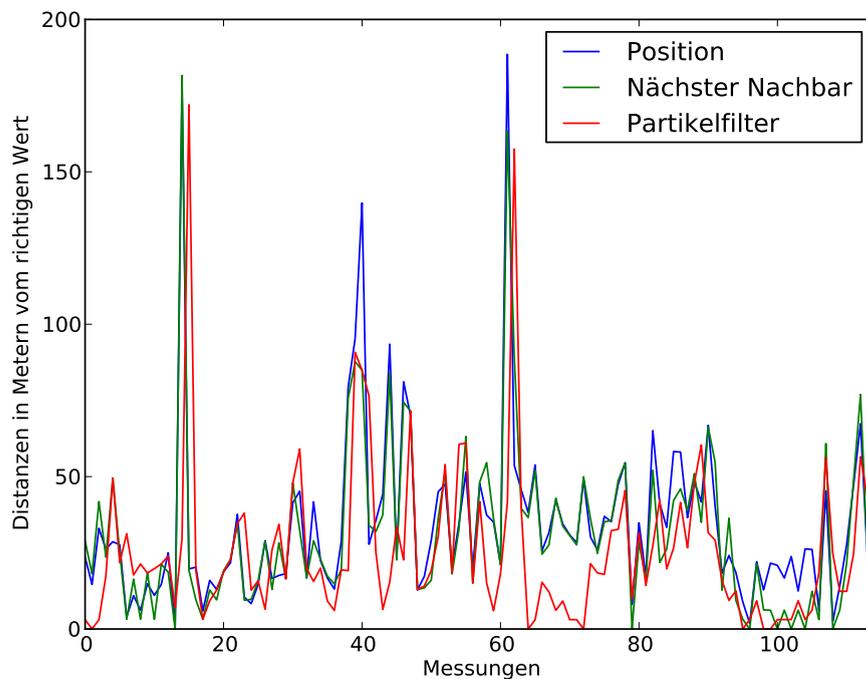
Vom Ausgangspunkt an der Ecke Pontstraße-Neupforte mit Blick in Richtung Driescher Gässchen wird sich stets an der äußeren Kante des sich rechts befindenden Bürgersteiges fortbewegt. Alle 5 m, die anhand eines Messrades ermittelt werden, wird ein Halt eingelegt. Dann wird jeweils ein Scan mit einem Smartphone, dem Samsung Galaxy Nexus, gestartet, eine Position aufgezeichnet und wieder gestoppt. Die Aufzeichnung wird so fortgesetzt, bis der Ausgangspunkt an der Ecke Neupforte-Pontstraße wieder erreicht wird. Es ist zu erwähnen, dass bei jeder Positionsaufzeichnung ein Start und ein Stopp der Messanwendung erfolgt. Dadurch können insgesamt schlechtere Ergebnisse erzielt werden, als wenn diese Aufzeichnung fortlaufend durchgeführt würde. Eine fortlaufende Aufzeichnung hätte zwar den Vorteil, dass die Positionierung durch mehrere GPS Fix verbessert würde. Diese Vorgehensweise war jedoch nicht möglich, da die Aufzeichnung ansonsten nicht mehr einem Referenzpunkt zuordbar wäre, was in diesem Fall allerdings sehr relevant ist. Zusätzlich wird die Anzahl der bis dahin stattgefundenen Messungen an den jeweiligen Orientierungspunkten notiert. Die Abbildungen ?? zeigen die richtigen Positionen, die gemessenen Positionen, den nächsten Nachbarn und das Partikelfilter des begrenzten Szenarios auf der OSM-Karte.

Aus den Messungen werden die Positionen ausgelesen, worauf das Partikelfilter und der nächste Nachbar ausgeführt werden. Mit Hilfe der Orientierungspunkte und des Messrades werden die Referenzpunkte der Rasterdatenbank entnommen, die als richtige Position festgelegt werden. Somit ist es möglich, die Abweichung zwischen den richtigen Positionen und den Messungen sowie die Ergebnisse der Algorithmen zu bestimmen. Die Abbildung ?? stellt die Abweichung von der richtigen Position dar, also den Referenzpunkt aus den Rasterdaten, an dem die Messung stattfand, zu den Positionsdaten des Smartphones, des nächsten Nachbarn und des Partikelfilters angewendet auf die Positionsdaten. Die Positionsdaten bestehen aus der Lokalisierung von GPS, WLAN und Mobilfunk. Daher sind vereinzelt größere Sprünge bei den Positionsdaten zu beobachten, die auch die Ergebnisse der Algorithmen beeinflussen. Der nächste Nachbar mit einer durchschnittlichen Abweichung von 32 m kommt der Abweichung der Positionsdaten mit durchschnittlichen 34 m sehr nah. Vereinzelt sind die Abweichungen sogar kleiner, da das Ergebnis des nächsten Nachbarn auf einen Referenzpunkt einer Straße gesetzt wird, die näher an der richtigen Position liegt. Das Partikelfilter kommt insgesamt auf eine kleinere durchschnittliche Abweichung von 25 m. Gerade im Bereich der 60. und der 80. Messung ist ersichtlich, dass eine Verbesserung von 10 m entsteht.

Die Abweichungen in der Grafik ?? sind nicht vergleichbar mit den Accuracy-Werten des Smartphones, da es sich bei der Abweichung in der Grafik um die richtige Position handelt, die also die wahre Position mit der Abweichung vom Messrad sowie



Abbildung 6.8 Darstellung der Messdaten für den Positionsvergleich



**Abbildung 6.9** Darstellung der euklidischen Distanz vom richtigen Wert der gemessenen Positionen, vom nächsten Nachbar und vom Partikelfilter

die Abweichung der OSM-Karten beinhaltet. Die Accuracy-Werte beschreiben die Abweichung, an der die wahre Position vermutet wird.

## 6.4 Ausbreitungsmodell

Dieses Kapitel beschreibt die Evaluation der beiden Parameter des LDMs sowie die Abweichung des Modells von einer Messreihe. Dies bedeutet, dass jede einzelne Eingabe für dieses Modell anhand der Messungen ausgewertet wird. Das Ergebnis wird mit einem Track des Szenarios verglichen. Die Abweichung der Signalstärke wird untersucht. Die erste Untersuchung ist der Nachweis für einen Pfadverlust aller Messungen. Damit soll die Abhängigkeit von Distanz und Signalstärke anhand der Messung nachgewiesen werden. Dies ist essentiell, da das LDM die Berechnung auf den Pfadverlust bezieht [18]. Dies wird in Kapitel ?? anhand der Messungen aller Referenz-WLANs sowie anhand eines einzigen Referenz-WLANs dargestellt. Der Verlauf der Distanzen aller Messungen ist in dem Streudiagramm ?? erkennbar.

Im nächsten Unterkapitel wird der Pfadverlust  $L(d_0)$  von einem Meter  $d_0$  Abstand zwischen dem Smartphone und dem AP in der Sichtlinie ermittelt. Dabei werden die Schwankungen der Signalstärke eines Smartphones über mehrere Messungen und die Abweichung der Signalstärke von vier verschiedenen Smartphones visualisiert. Der letzte Parameter, der Pfadverlustexponent, wird als Histogramm aller Messungen berechnet. Anhand der Abbildung ?? wird der Pfadverlust über eine Distanz in dem Szenario festgestellt. Dieses Ergebnis wird mit der Auswertung der anderen Arbeiten verglichen. Zuletzt wird die Abweichung des Ergebnisses des LDMs mit einer

Messreihe ermittelt, um eine Güte des LDM-Algorithmus mit realen Messdaten zu erhalten.

### 6.4.1 Pfadverlust

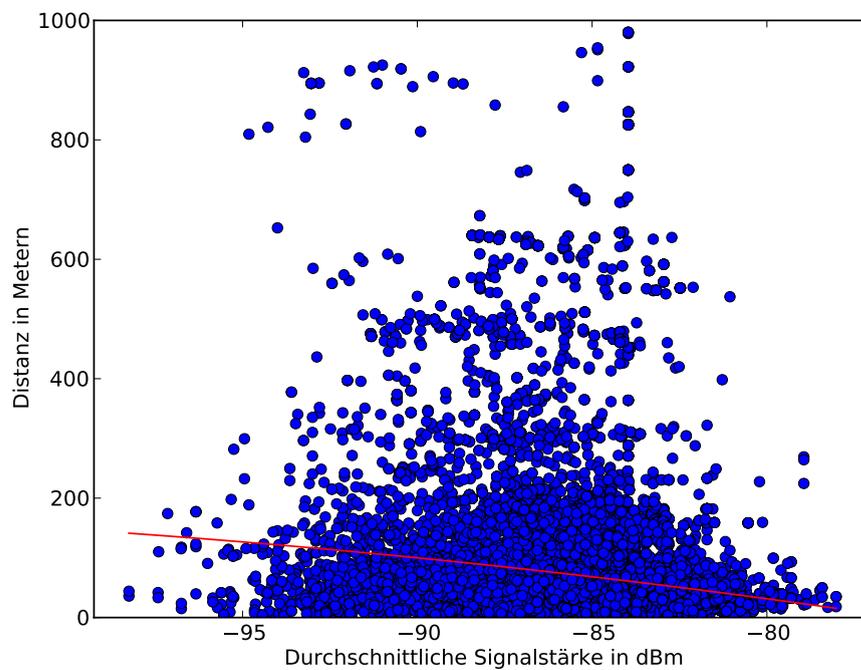
Diese Auswertung soll nachweisen, ob ein Pfadverlust in den Messungen überhaupt erkennbar ist. Die Erwartung ist, dass bei steigender Distanz die Signalstärke zwischen dem AP und der Messung schwächer wird. Dieser Nachweis wird für die Ausbreitungsmodelle benötigt, da die Berechnung auf dem Pfadverlust beruht. Dafür werden die Messungen für alle Referenz-WLANs durchgeführt und alle Messungen für ein einziges Referenz-WLAN evaluiert. Die Vorgehensweise ist bei beiden die gleiche. Mit Hilfe des Partikelfilters wird die Abweichung der GPS-Position in den Messungen minimiert. Mittels Map Matching wird das Partikel auf den passenden Referenzpunkt gesetzt. Der Algorithmus des Partikelfilters befindet sich in dem Kapitel 4.3.3. Das Map Matching wird in dem Kapitel 4.3.4 näher erklärt.

Die beiden Abbildungen ?? und 6.7 stellen die durchschnittliche Signalstärke mehrerer Messungen pro Referenzpunkt dar. Die andere Achse der Abbildungen ist die Distanz zwischen dem Referenzpunkt und dem AP, deren Signalstärke gemessen wird. Aus der entstandenen Punktwolke ist es schwer, eine Aussage zu treffen, daher folgt eine Kurvenanpassung (Curve Fitting) auf diese Datenwolke. Es handelt sich um eine Schätzung der unbekannt Parameter aus den Messdaten für das Modell, das die Daten repräsentieren soll. Diese Berechnung basiert auf der Methode der kleinsten Quadrate. Daraus ergibt sich, dass die Summe aller Abweichungen, die sich zwischen den Mess- und Modelldaten befinden, minimal ist. Das Ergebnis liefert eine Linie, die die Beziehung zwischen den Variablen Distanz und Signalstärke visualisiert. Die Erwartung ist, dass die Abhängigkeit der beiden Variablen bei einer großen Anzahl von Messungen erkennbar ist. Die Abbildung ?? ist die Darstellung des Pfadverlustes anhand der Messungen von allen APs. Der Großteil der Referenzpunkte befindet sich in dem Signalstärkebereich von -95 bis -80 dBm und in der Distanz von 1 bis 200 m. In diesem Bereich verläuft auch die Kurvenanpassung. Dabei ist festzustellen, dass die Kurvenanpassung bei steigender Distanz die Signalstärke abschwächt. Damit ist der Pfadverlust anhand der Kurvenanpassung nachgewiesen.

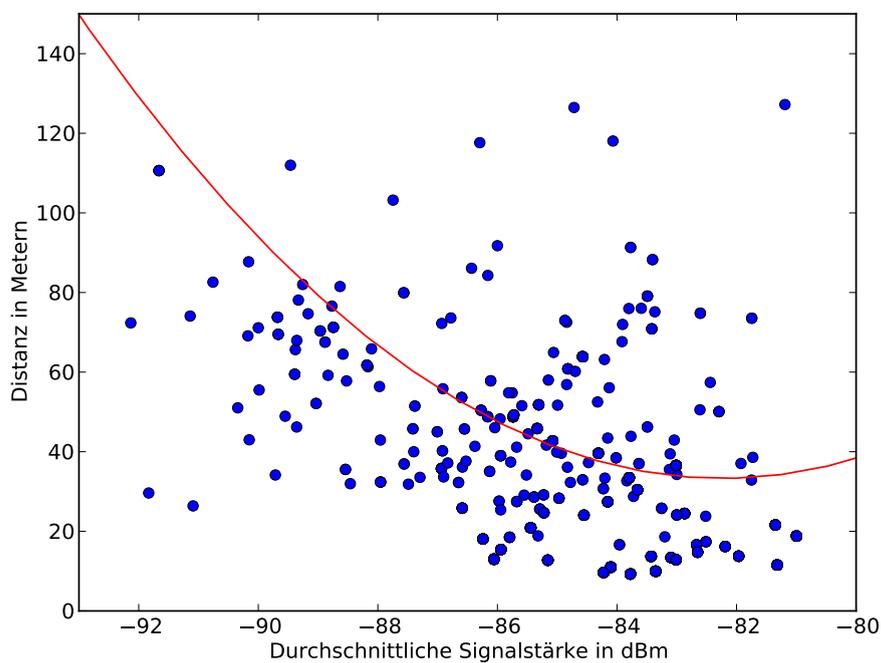
Die Abbildung 6.7 stellt den Pfadverlust anhand von Messungen für einen einzigen AP dar. Dieser AP befindet sich in der Pontstraße und liegt somit auch im Szenario. Es handelt sich um ein Referenz-WLAN, wo Messungen an 822 verschiedenen Referenzpunkten durchgeführt wurden. In dem Bereich von -93 bis -82 dBm sinkt die Signalstärke jenes APs bei steigender Distanz. Damit ist für diesen AP der Pfadverlust auf Grund der gemachten Messungen nachgewiesen.

### 6.4.2 Referenz-Pfadverlust

Diese Auswertung hat das Ziel, den Pfadverlust  $L(d_0)$  der Smartphones über einen Meter  $d_0$  zu ermitteln. Er beschreibt die Dämpfung des Signals über eine festgelegte Distanz von einem Meter im Vakuum [18]. Dabei soll festgestellt werden, ob der Verlauf der Signalstärke bei denselben sowie bei unterschiedlichen Smartphones



**Abbildung 6.10** Darstellung der durchschnittlichen Signalstärke eines Referenzpunktes im Vergleich zu der Distanz zwischen dem Referenzpunkt und dem AP aus den Messungen. Die rote Linie ist die Kurvenanpassung der Punktwolke

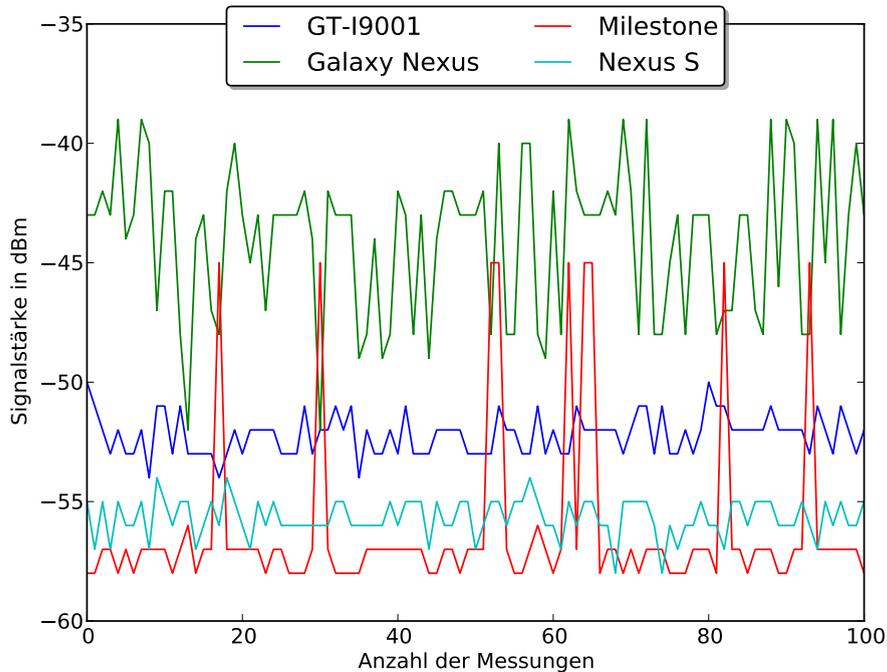


**Abbildung 6.11** Darstellung der durchschnittlichen Signalstärke eines Referenzpunktes im Vergleich zu der Distanz zwischen dem Referenzpunkt und dem AP aus den Messungen eines APs. Die rote Linie ist die Kurvenanpassung der Punktwolke

gleich ist. Anhand der Messungen der unterschiedlichen Smartphones kann die systematische Messabweichung im Vergleich zu anderen Smartphones dargestellt werden. Diese Ermittlung des Pfadverlustes ist auch über die Rasterdatenbank möglich. Es wurde sich jedoch gegen diese Möglichkeit entschieden, weil die meisten Messungen nicht die Sichtlinie zwischen dem AP und dem Smartphone erreichen, da sich die meisten APs in einem Gebäude und die Smartphones außerhalb des Gebäudes befinden. Diese Ermittlung soll in einer weiteren Messreihe durchgeführt werden, damit die Smartphones dieselben Bedingungen haben.

So wurde eine Messung von vier Smartphones aufgezeichnet, die je 100 WLAN-Scans mit deren Signalstärken aufweisen. Diese Smartphones sind das Samsung Nexus S, das Samsung Galaxy Nexus, das Motorola Milestone und das Samsung Galaxy S Plus (GT-I9001). Während der Messung befanden sie sich in einem Abstand von einem Meter zum AP und lagen mit der Displayseite in Richtung Boden. Der AP ist ein Linksys WRT54G WLAN-Router, der üblicherweise als Router beim Öcher WLAN zur Verfügung gestellt wird. Somit entspricht dies den meisten Referenz-WLANs, die zur Berechnung der Ausbreitungsmodelle genutzt wurden. Dieser AP läuft unter der Standardkonfiguration mit einer Sendeleistung von 16 dBm. Während der Messung war er auf der Frequenz 2.437 MHz im sechsten Kanal eingestellt. Der Verlauf der Signalstärke ist in der Abbildung ?? dargestellt. Obwohl alle Smartphones die Signalstärken unter denselben Bedingungen aufzeichneten, sind die durchschnittlichen Signalstärken verschieden zueinander. Dies kann unterschiedliche Gründe haben. Eine Annahme ist z. B., dass die Smartphones unterschiedliche eingebaute Antennen mit verschiedener Richtcharakteristik haben. Das Galaxy Nexus hatte den besten Empfang mit einer durchschnittlichen Signalstärke von -44 dBm. Das GT-I9001 wies -52 dBm auf. Das Nexus S und das Milestone lagen je bei -56 dBm. Daraus entsteht eine Abweichung von 12 dB zwischen dem kleinsten und dem größten Wert der unterschiedlichen Smartphones. Der Verlauf der Signalstärke bei dem GT-I9001 und dem Nexus S ergibt eine Standardabweichung von knapp 0. Sie geben eine fast konstant gleiche Signalstärke in den 100 Messungen aus. Das Milestone und das Galaxy Nexus haben einen höheren Wert bei der Standardabweichung, der bei 3,35 liegt. In der Abbildung ?? ist dies bei dem Milestone gut erkennbar: Sieben Sprünge führen zu einem höheren Wert. Somit entsteht eine Abweichung von 38 dB zwischen dem höchsten Wert von -45 dBm und dem Median von -57 dBm. Diese sieben Sprünge außer Acht gelassen, erreicht das Milestone ansonsten auch einen konstanten Wert mit einer kleineren Abweichung. Beim Galaxy Nexus schwanken die Signalstärken selbst über einen kurzen Zeitraum. Der niedrigste Wert ist -52 dBm und der höchste Wert ist -39 dBm, sodass eine Abweichung von 13 dBm bei demselben Smartphone entsteht. Für weitere Berechnungen beträgt der Pfadverlust einen Meter, sodass dieser dem durchschnittlichen Wert von -50 dBm, der bei den vier Smartphones erhalten wurde, gleichgesetzt wird.

In dem Kapitel Grundlagen unter ?? wird der Pfadverlust als positiver Wert der Differenz von der Sendeleistung und der Empfangsleistung beschrieben. Die Sendeleistung der Referenz-WLANs ist 16 dBm. Daraus resultiert die Rechnung  $|16 \text{ dBm} - 50 \text{ dBm}| = 34 \text{ dB}$ . Das Vorzeichen des Ergebnisses fällt weg, da der Pfadverlust immer positiv ist. Der Wert 34 dB wird für die folgenden Berechnungen des Referenz-Pfadverlustes  $L_0$  eingesetzt.



**Abbildung 6.12** WLAN-Messungen der Signalstärke verschiedener Smartphones im Abstand von einem Meter zum AP

### 6.4.3 Pfadverlustexponent

Für das LDM ist der Pfadverlustexponent eine wichtige Eingabe. Er bestimmt wie stark die Signalstärke über die Distanz abfällt. Um den Exponenten zu bestimmen, wird folgende Formel ?? eingesetzt. Der Referenz-Pfadverlust  $L_0$  beträgt 34 dB und wurde in dem Kapitel ?? beschrieben. Diese Distanz  $d$  wird aus der Position der Messung und des Referenz-WLANs berechnet. Die Variable  $L_{LDM}$  ist der Pfadverlust über die Distanz und beschreibt die Dämpfung der Sendeleistung zur Empfangsleistung. Dieser Verlust wird in dB angegeben. Daraus folgt, dass  $L_{LDM} = \text{Sendeleistung} - \text{Empfangsleistung}$  ist. Die Sendeleistung ist mit 16 dB bei den Referenz-WLANs bekannt und die Empfangsleistung ist die gemessene Signalstärke in dBm. Mit diesen Eingaben wird die oben erwähnte Formel berechnet, deren Ergebnis der Pfadverlustexponent ist. Dies wurde mit allen Messungen und den Öcher WLANs durchgeführt und als Histogramm in dem Kapitel ?? dargestellt. Aus dieser Abbildung ist zu entnehmen, dass viele Werte knapp über 2 liegen. Dieser Effekt kann durch Reflexionen in Straßenschluchten entstehen, so dass die Signalstärke verstärkt wird. Der zweithöchste Punkt liegt knapp über 4 mit einer größeren Varianz zum vorherigen Punkt. Diese Werte können durch den Pfadverlust und durch die Dämpfung von Hindernissen wie Gebäuden entstehen. Der durchschnittliche Pfadverlustexponent beträgt 3,75.

In den verwandten Arbeiten in Kapitel 3.2 wurden Pfadverlustexponenten aus anderen Untersuchungen vorgestellt. Unsere Ergebnisse entsprechen den Angaben aus diesen wissenschaftlichen Arbeiten.

Der Wert von 3,75 für den Pfadverlustexponenten  $n$  wird für die folgende Berechnung des LDMs eingesetzt.

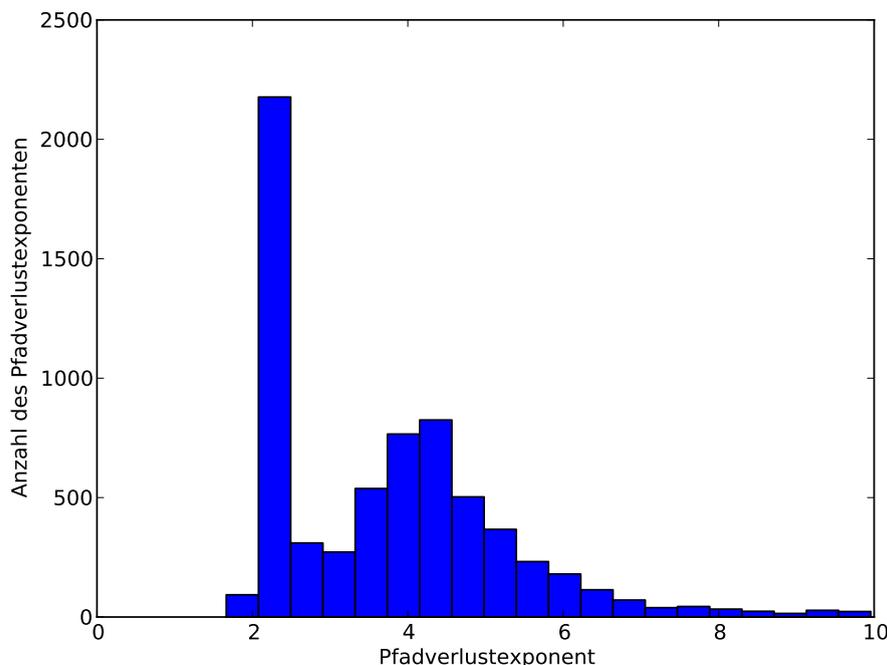


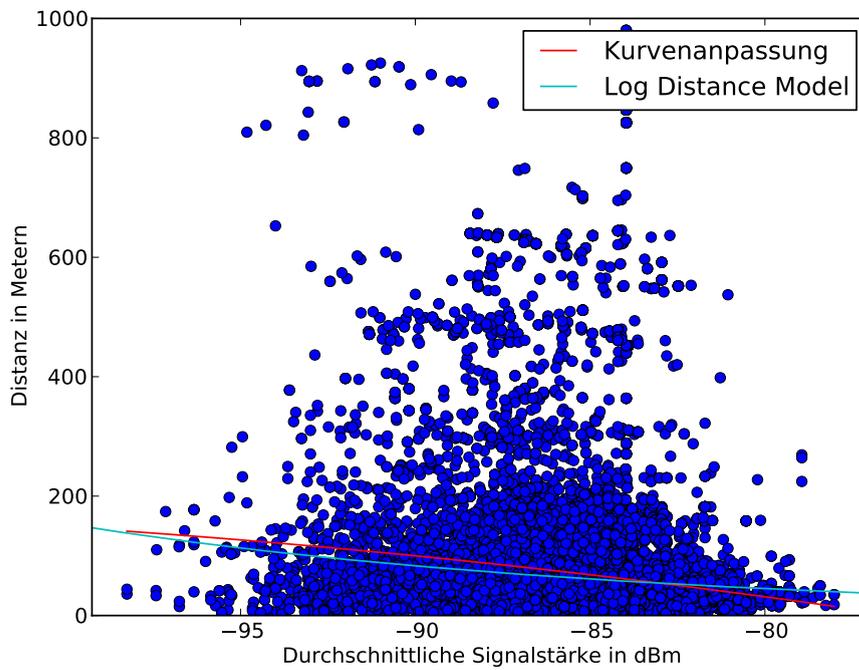
Abbildung 6.13 Histogramm des Pfadverlustexponenten

#### 6.4.4 LDM

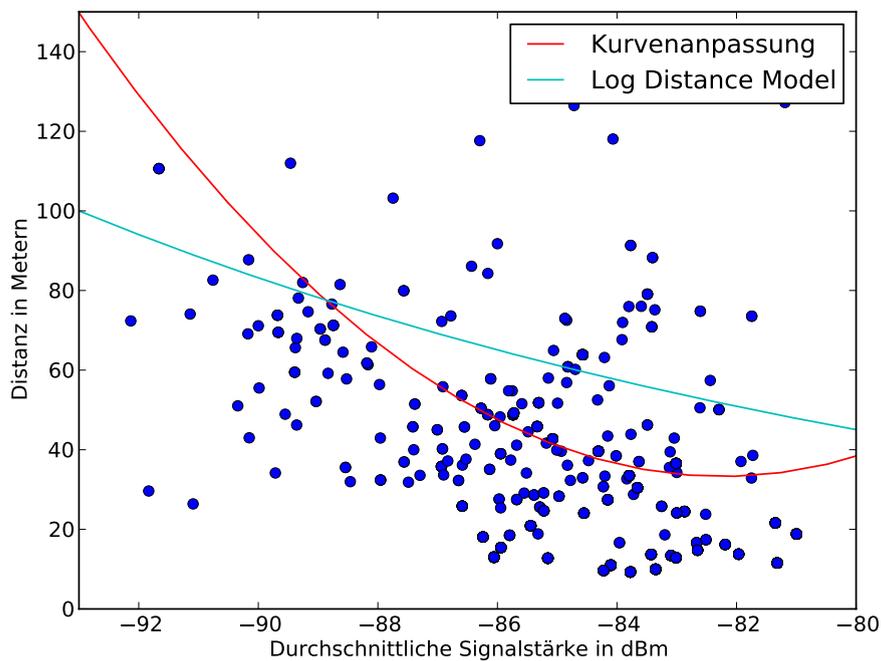
In diesem Kapitel wird das Log Distance Model ausgewertet und mit den Messungen verglichen. Die Formel für LDM stammt aus dem Kapitel 2.3. Als Eingabe für die Formel wird der Pfadverlustexponent  $n$  auf den Wert 3,75 aus dem Kapitel ?? gesetzt und der Referenz-Pfadverlust  $L_0$  wird auf den Wert 34dB aus dem Kapitel ?? gesetzt. Die Distanz wird über die Position des Referenz-WLANs und über die Position der Messung errechnet. Der Vergleich des LDMs wird nicht an einem Track durchgeführt, da hier die systematischen Messabweichungen von einem einzigen Smartphone des Tracks die Abweichungen des Vergleichs verfälschen würden. Es ist keine Unterscheidung zwischen der systematischen Messabweichung und der Abweichung des Vergleiches möglich. Daher werden die Ergebnisse des LDMs mit der Kurvenanpassung aller Messungen verglichen. Die Abbildung ?? stellt die Kurvenanpassung (Linie in Rot) und die Ergebnisse des LDMs (Linie in Cyan) dar. Diese Abbildung basiert auf der Abbildung ?. Aus der Abbildung ist zu entnehmen, dass die Ergebnisse des LDMs nah an der Kurvenanpassung liegen. Die durchschnittliche Abweichung der beiden Kurven liegt bei 8,5 m bei gleicher Signalstärke.

Die Abweichung lässt sich auch für die Signalausbreitung eines WLANs bestimmen. Dies stellt die Abbildung ?? dar und basiert auf Kapitel 6.7. Hier entsteht sogar eine kleine durchschnittliche Abweichung von 6,6 m.

Somit ist nachgewiesen, dass keine komplette Abdeckung von Messungen eines Gebietes benötigt wird, um die Signalausbreitung anhand der Kurvenanpassung zu bestimmen. Es reicht aus, den Pfadverlustexponenten für dieses Gebiet zu bestimmen und die Signalausbreitung mit Hilfe des LDMs zu berechnen.



**Abbildung 6.14** Darstellung der durchschnittlichen Signalstärke eines Referenzpunktes im Vergleich zu der Distanz zwischen dem Referenzpunkt und dem AP aus den Messungen eines APs



**Abbildung 6.15** Darstellung der durchschnittlichen Signalstärke eines Referenzpunktes im Vergleich zu der Distanz zwischen dem Referenzpunkt und dem AP aus den Messungen eines APs

## 6.5 Zusammenfassung

Das Szenario wurde auf einen vorher definierten Teil der Innenstadt von Aachen begrenzt. Es handelt sich um eine Mischung aus Geschäfts- und Wohnanlagen. In diesem Gebiet wurden knapp 2.000 verschiedene WLANs entdeckt; dies ergibt 20 WLANs pro km<sup>2</sup>. Die Scandatenbank besteht aus knapp 124.500 Messpunkten mit einer Million WLAN-Messungen. Diese Messungen zeichneten 15 Smartphones während sieben Monaten auf, wovon drei Smartphones 76 % der Messungen durchführten. Die RSSI-Werte liegen meist in den niedrigen Bereichen von -95 dBm bis -75 dBm. Die Lokalisierung von GPS bietet mit einer durchschnittlichen Abweichung von 18 m eine genauere Lokalisierung im Vergleich zu Network mit einer durchschnittlichen Abweichung von 349 m. Die WLAN-Messungen entstammen zu 99 % dem 2,4 GHz-Bereich.

Der Mittelwert der Anzahl Messungen pro Referenzpunkt in der Rasterdatenbank beträgt 16. Da die Rasterdatenbank mit Hilfe des Partikelfilters erstellt wurde, ist dieser Algorithmus mit den Positionsdaten und dem nächsten Nachbarn verglichen worden. Die Abweichungen dieser Daten zu den richtigen Werten wurden ermittelt. Es wurde festgestellt, dass das Partikelfilter die gleiche Abweichung zum nächsten Nachbarn aufweist.

Für die Ausbreitungsmodelle LDM und MWM wurde der Pfadverlust für alle Messungen und für einen einzigen AP dargestellt. Es ist erkennbar, dass sich die Signalstärke über die steigende Distanz abschwächt. Bei der Untersuchung für den Referenz-Pfadverlust wurden Abweichungen über 100 Messungen mit einer Standardabweichung von 0 bis 3,4 ermittelt. Auch Unterschiede zwischen verschiedenen Smartphones sind erkennbar, die maximal bis 13 dBm abweichen.

Der Vergleich von LDM und Kurvenanpassung zeigt auf, dass die durchschnittliche Abweichung bei etwa 8,5 m liegt. Damit ergibt sich die Möglichkeit, das LDM zu nutzen und nicht an eine komplette Abdeckung des Gebietes mit Messungen gebunden zu sein.



# Kapitel 7

## Zusammenfassung

In dieser Arbeit wurde eine Android-Messanwendung vom Entwickler konzipiert, um repräsentative Messungen anhand von Smartphones durchzuführen, die während des WLAN-Scans die BSSID, die Signalstärke, die SSID und den WLAN-Kanal ermitteln und bezüglich der Positionsermittlung die Werte der Latitude, der Longitude, die Accuracy sowie die Lokalisierungsarten Network und Provider aufzeichnen. Diese Messungen wurden von Endanwendern in der Stadt Aachen und dort insbesondere in einem vordefinierten Szenario durchgeführt. Anschließend wurden diese Messdaten dem Entwickler zwecks Auswertung zur Verfügung gestellt. Die Daten wurden in eine eigens dafür erstellte Scandatenbank hochgeladen. Da die Ermittlung der Position anhand von GPS und Network eine zufällige Messabweichung enthält, wurde das Partikelfilter auf die Positionsdaten angewendet. Dadurch konnte die Messabweichung minimiert werden. Der nächste Schritt bestand in der Erstellung von Referenzpunkten, die öffentliche Straßen und Plätze repräsentieren, anhand der Geodaten aus OSM. Diese Referenzpunkte bilden das Grundgerüst der Rasterdatenbank. Aus den Messungen der Scandatenbank wurde eine Aggregation der WLAN-Messungen zu den entsprechenden Referenzpunkten durchgeführt und in die Rasterdatenbank eingefügt. Für das Ausbreitungsmodell LDM wurden die Distanzen zwischen dem WLAN und der Messung, dem Pfadverlustexponenten und dem Referenz-Pfadverlust ermittelt. Für die Berechnung der Distanzen zwischen Referenz-WLANs und den darauf basierenden Messungen steht eine Liste mit Referenz-WLANs und deren Position vom Hotspot-Anbieter Öcher WLAN zur Verfügung. Der Pfadverlustexponent wurde anhand der Daten aus der Rasterdatenbank und der Referenz-Pfadverlust anhand eigener Messungen ermittelt. Mit diesen drei Elementen wurde die Berechnung des LDMs durchgeführt, deren Ergebnisse für jeden AP visualisiert wurden. Für die Evaluation wurden folgende Werte zwecks Repräsentativität der Scandatenbank ermittelt: 147 Tracks von 15 verschiedenen Smartphones haben 124.500 Messpunkte aufgezeichnet, wobei eine Million WLANs gescannt wurden. Insgesamt wurde eine Strecke von etwa 340 km in der Stadt Aachen zurückgelegt. Um die Anzahl der Messpunkte pro Referenzpunkt zu steigern, wurde ein Szenario auf eine Fläche von 95.555 m<sup>2</sup> beschränkt. In diesem Szenario, welches eine Mischung aus

Geschäfts- und Wohnanlage darstellt, wurden 20 WLANs pro km<sup>2</sup> aufgezeichnet. Die Aggregation der Messpunkte auf die Referenzpunkte für das Szenario ergeben durchschnittlich 19 Messungen pro Referenzpunkt. Diese Messungen zeigen die Repräsentativität der Rasterdatenbank.

Aus diesen Daten ist ersichtlich, dass die meisten Messungen eine Signalstärke in dem niedrigen Bereich von -95 dBm bis -80 dBm aufweisen und die meisten Messungen der Signalstärken zwischen 1 m und 200 m vom AP aufgezeichnet wurden. Es ließ sich auf Grund dessen nachweisen, dass sich die Signalstärke bei steigender Distanz abschwächt, so dass der Pfadverlust anhand der Messungen erkennbar ist. Dieser Nachweis ist relevant, damit das empirische Modell LDM auf die Messungen anwendbar ist.

In der Auswertung der Messungen für den Referenz-Pfadverlust wurde festgestellt, dass die systematische Messabweichung 13 dBm bei vier verschiedenen Smartphones beträgt.

Während der Messungen mit den Smartphones entstand eine zufällige Messabweichung der Signalstärke. Die Smartphones Samsung Galaxy S Plus (GT-I9001) und Samsung Nexus S hatten bei der systematischen Abweichung eine Standardabweichung von nahezu null und gaben somit fast konstant die gleiche Signalstärke über mehrere Messungen aus. Bei den weiteren Smartphones Motorola Milestone und Samsung Galaxy Nexus lag die Standardabweichung mit 3,35 höher als bei den oben genannten Geräten. So lag der Mittelwert bei -50 dBm, der als Eingabe für das LDM verwendet wurde.

Die Abweichung vom LDM und der Kurvenanpassung der Messungen liegt bei 8,5 m. Mit dieser kleinen Abweichung bietet sich das LDM eher als Kurvenanpassung an. Der Vorteil ist, dass keine komplette Abdeckung der Umgebung mit Messungen benötigt wird. Insgesamt zeigte sich, dass die in der vorliegenden Arbeit entwickelte Messanwendung für Android-Smartphones, durchgeführt von Anwendern, qualitativ hochwertige Messdaten liefert. Die Auswertung der repräsentativen Messdaten aus der Scandatenbank ergaben einen Beleg bereits genannter Ansätze in den verwandten Arbeiten sowie weitere Erkenntnisse auf Grund des Einsatzes eines Partikelfilters, die dem Signalausbreitungsmodell in einem Stadtszenario zu Gute kamen.

## 7.1 Ausblick

Dieses Unterkapitel beschreibt mehrere Ansätze für zukünftige Forschungsarbeiten, mit denen diese Diplomarbeit fortgesetzt werden kann. Diese Ansätze werden in folgende drei Themen unterteilt: Messungen, Positionierung und Signalausbreitungsmodell.

Bei den Messungen geht es grundsätzlich darum, wie eine größere Fülle an Informationen aus diesen extrahiert und in die Positionierung und das Signalausbreitungsmodell eingesetzt werden kann:

- Bei den Messungen wurden alle Daten der verfügbaren Sensoren im Sekundentakt aufgezeichnet sowie alle Werte vom Orientation Sensor gesammelt. Der Orientation Sensor beschreibt eingehend die Ausrichtung des Smartphones im Raum. Mit den Daten des Orientation Sensors kann eine Klassifizierung der WLAN-Messungen bestimmt werden. Dafür muss bei der Lokalisierung die

eigene WLAN-Messung mit den WLAN-Messungen aus der Rasterdatenbank verglichen werden, welche die gleiche räumliche Ausrichtung bei den Smartphones aufweisen. Durch die gleiche Ausrichtung des Smartphones wird angenommen, dass die Messungen mit der gleichen Antennenausrichtung aufgezeichnet wurden und einen ähnlichen Verlauf in den Signalstärken nehmen.

- Eine andere Klassifizierungsmöglichkeit ist die Aufteilung der Messungen anhand der Modelle der Smartphones. In Kapitel ?? wurde die systematische Messabweichung von vier verschiedenen Smartphones gemessen. Durch eine Klassifizierung können Smartphones in eine Kategorie zusammengefasst werden, die das Ausgabekriterium der gleichen Signalstärke der WLANs unter gleichen Bedingungen aufweist. Damit sollte die systematische Messabweichung minimiert werden können.
- Aus Erhebungen der Scandatenbank kann der Effekt der Fluktuation veranschaulicht werden und aufgezeigt werden, wie viele APs in einem Gebiet verschwinden und wie viele neue APs erscheinen. Diese Auswertung kann ein Indiz dafür sein, wie oft Messungen aus diesem Gebiet benötigt werden, um die Datenbank aktuell zu halten.

Bei der Positionierung kann die zufällige Messabweichung der Positionsdaten noch weiter minimiert werden. Zudem kann eine weitere Optimierung des Partikelfilters erreicht werden:

- Eine mögliche Optimierung des vorhandenen Partikelfilters ist die Erweiterung um eine Offline-Variante. Diese Variante ist bekannt als BackTracking und wurde in verschiedenen Arbeiten bereits umgesetzt und evaluiert.
- In der Arbeit [13] et al. 2012 wurde beschrieben, dass aus WLAN-Messungen und deren Position eine Lokalisierung des gemessenen APs entwickelt wurde. Mit dieser Methode kann die Eingabe von Referenz-WLANs vermieden werden. Dabei muss allerdings evaluiert werden, welche Abweichungen zwischen der Position eines Referenz-WLANs und der berechneten Position entstehen.

Bei dem Thema Signalausbreitungsmodell kann das Konzept des Multi Wall Models weiter verfolgt und die Idee eines detaillierten Umgebungsmodells erforscht werden:

- Das Konzept des MWMs wurde in Kapitel 4.16 erläutert. Das MWM ist eine Erweiterung des LDMs, bei dem mit Hilfe von Dämpfungsfaktoren zusätzlich die Dämpfung von Hindernissen und Reflexionen in Straßen berücksichtigt werden. Es beschreibt, wie das Umgebungsmodell aus OSM erstellt wird und wie viele Schnitte eines WLAN-Signals durch Häuser entstehen. Der Dämpfungsfaktor für Häuser und Straßen wird anhand der Messungen bestimmt. Daraus lässt sich das MWM berechnen. So könnten genauere Signalstärken im Vergleich zu Messungen berechnet werden [29].
- Das Projekt OSM wird ständig weiter entwickelt. Zunehmend mehr Informationen der Umgebung werden in Geodaten digitalisiert. Bei Gebäuden gibt es beispielsweise inzwischen ein Tag für die Gebäudehöhe. Mit der Höhe und dem Polygon der äußeren Hauswand kann ein primitives 3D-Modell erstellt werden, sodass weitere Daten für die Signalausbreitung gesammelt werden können.

Neben diesen Schwerpunkten nimmt die Client-Lokalisierung einen wichtigen Stellenwert ein. Aus den Resultaten des LDMS kann die Position anhand von Lokalisierungsalgorithmen mit einem WLAN-Scan bestimmt werden. Diese Lokalisierung kann sowohl auf einem Smartphone, als auch auf einem Server ausgeführt werden.

# Literaturverzeichnis

- [1] BEHLENDORF, C. Analyse der gps-empfänger in smartphones. Bachelor's thesis, Universität der Bundeswehr München, 2012.
- [2] CRISAN, D., AND DOUCET, A. A survey of convergence results on particle filtering methods for practitioners. *IEEE TRANSACTIONS ON SIGNAL PROCESSING vol. 50, issue 3* (2002), 736–746.
- [3] DE LANGE, N. *Geoinformatik in Theorie und Praxis*, 2nd ed. Springer Verlag, 2006.
- [4] DOUCET, A., DE FREITAS, N., AND GORDON, N. *Sequential Monte Carlo Methods in Practice - Statistics for Engineering and Information Science*, 1st ed. Springer Verlag, 2001.
- [5] FOUQUE, C., AND BONNIFAIT, P. Matching raw gps measurements on a navigable map without computing a global position. *Intelligent Transportation Systems, IEEE Transactions on 13* (2012), 887–898.
- [6] GAST, M. S. *802.11 Wireless Networks - The Definitive Guide - Creating and Administering Wireless Networks*. O'Reilly Associates, 2002.
- [7] HERTER, E., AND LÖRCHER, W. *Nachrichtentechnik: Übertragung - Vermittlung - Verarbeitung*, 9th ed. Carl Hanser Verlag, 2004.
- [8] KAN, Z. Research and implementation of intelligent mobile phone location based on rssi in smart space. In *Systems and Informatics (ICSAI), 2012 International Conference on* (2012), pp. 1635–1639.
- [9] KLAUS, F. *Einführung in Techniken und Methoden der Multisensor-Datenfusion*. PhD thesis, Universität Siegen, 1999.
- [10] KONG, A., LIU, J. S., AND WONG, W. H. Sequential imputations and bayesian missing data problems. *Journal of the American Statistical Association vol. 89, no. 425* (1994), 278–288.
- [11] LOTT, M., AND FORKEL, I. A multi-wall-and-floor model for indoor radio propagation. In *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS* (2001), vol. 53rd vol., pp. 464–468.
- [12] MARTIN, E., VINYALS, O., FRIEDLAND, G., AND BAJCSY, R. Precise indoor localization using smart phones. In *Proceedings of the international conference on Multimedia* (2010).

- [13] MEIKLE, R., AND CAMP, J. A global measurement study of context-based propagation and user mobility. In *Proceedings of the 4th ACM international workshop on Hot topics in planet-scale measurement* (2012), pp. 21–26.
- [14] MEYER, S., VAUPEL, T., AND HAIMERL, S. Wi-fi coverage and propagation for localization purposes in permanently changing urban areas. In *IADIS International Conference Wireless Applications and Computing* (2008).
- [15] MÜLLER, S. P. Monte Carlo-Methoden - Angewandt in der Roboterlokalisierung. Master's thesis, Universität Stuttgart, 2007.
- [16] OBST, M. Untersuchungen zur kooperativen Fahrzeuglokalisierung in dezentralen Sensornetzen. Master's thesis, Technische Universität Chemnitz, 2009.
- [17] RAMM, F., AND TOPF, J. *OpenStreetMap - Die freie Weltkarte nutzen und mitgestalten*, 3rd ed. Lehmanns Media, 2010.
- [18] RAPPAPORT, T. S. *Wireless Communications: Principles and Practice - Second Edition*, 19th ed. Prentice Hall PTR, 2010.
- [19] RECH, J. *Wireless LANs: 802.11-WLAN-Technologie und praktische Umsetzung im Detail*, 1st ed. Heise, 2004.
- [20] SCHILLER, J. H. *Mobile Communications*, 2nd ed. Addison-Wesley, 2003.
- [21] SCHUMANN, T. Genauigkeitserhöhung bei der fahrzeugnavigation mittels map matching und dgps. Master's thesis, Technische Universität Ilmenau, 1996.
- [22] SINNOTT, R. W. Virtues of the haversine. *Sky and Telescope vol. 68, no. 2* (1984), 158.
- [23] VARELA, F., CERCAS, F., AND VELEZ, F. J. Unified propagation model for wi-fi, umts and wimax planning in mixed scenarios. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on* (2010), pp. 81–86.
- [24] VON WATZDORF, S., AND MICHAHELLES, F. Accuracy of positioning data on smartphones. In *Proceedings of the 3rd International Workshop on Location and the Web* (2010).
- [25] WALKE, B. *Mobilfunknetze und ihre Protokolle 1: Grundlagen, GSM, UMTS und andere zellulare Mobilfunknetze*, 3rd ed. Vieweg+Teubner Verlag, 2001.
- [26] WHITE, C. E., BERNSTEIN, D., AND KORNHAUSER, A. L. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies 8* (2000), 91–108.
- [27] WIDYAWAN, KLEPAL, M., WEYN, M., WIBOWO, S., AND NAJIB, W. Locon deliverable 5.1, 2009.
- [28] WITT, B. C. *Datenschutz kompakt und verständlich: Eine praxisorientierte Einführung mit Online-Service*, 2nd ed. Vieweg+Teubner Verlag, 2010.

- 
- [29] ZVANOVEC, S., PECHAC, P., AND KLEPAL, M. Wireless lan network design: Site survey or propagation modeling? *Radioengineering vol. 12, no. 4* (2003), 42–49.



Anhang

# A

## Anhang

### A.1 Abkürzungsliste

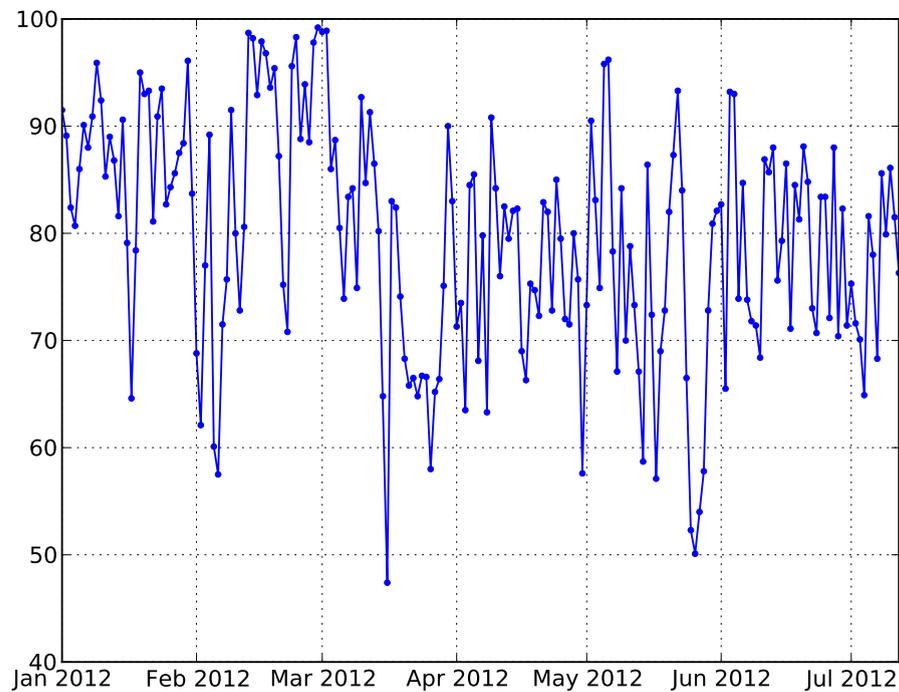
3D	drei Dimensionen
AP	Access Point
API	Application Programming Interface
BSD	Berkeley Software Distribution
BSS	Basic Service Set
BSSID	Basic Service Set Identification
CC-BY-SA	Creative Commons by Share Alike
dB	Dezibel
dBm	Dezibel mit der Bezugsgröße 1 mW
DBMS	Database Management System
GHz	Gigahertz
GLONASS	Globalnaja Nawigazionnaja Sputnikowaja Sistema
GNSS	Globales Navigationssatellitensystem
GNU	GNU's Not Unix
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HTTP	Hypertext Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
JSON	JavaScript Object Notation
KML	Keyhole Markup Language
LDM	Log Distance Model
MAC	Media Access Control
MHz	Megahertz
MIT	Massachusetts Institute of Technology
mW	Milliwatt
MWM	Multi Wall Model
OS	Operating System
OSM	OpenStreetMap

rad	Radiant
RSSI	Received Signal Strength Indication
SD	Secure Digital
SSID	Service Set Identifier
UTF-8	8 Bit Universal Character Set Transformation Format
WGS 84	World Geodetic System 1984
Wi-Fi	Synonym für WLAN
WLAN	Wireless Local Area Network
XML	Extensible Markup Language

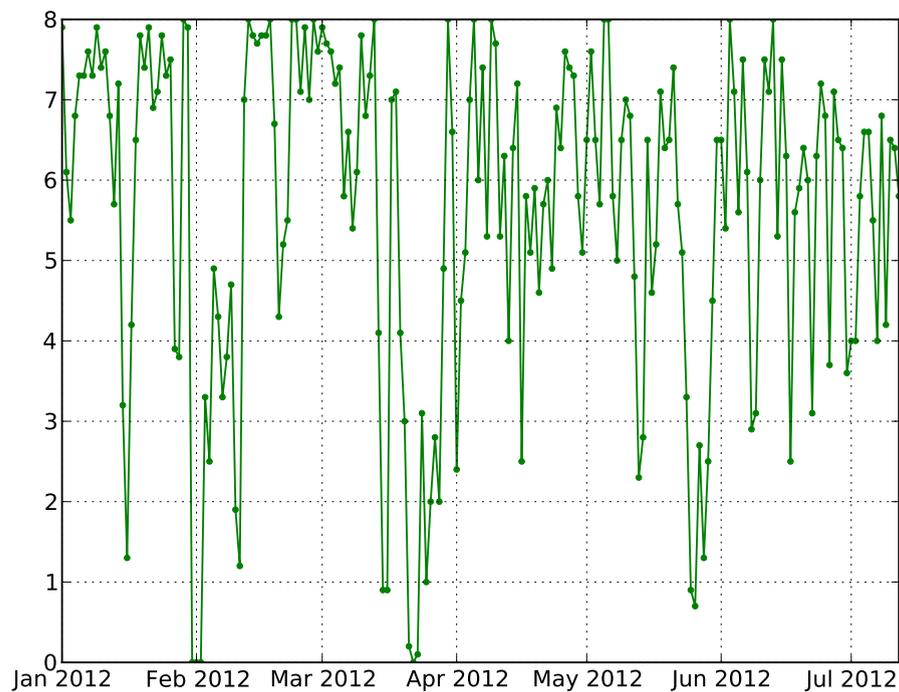
## A.2 Wetterdaten



**Abbildung A.1** Mittelwert der Temperatur in 2 m über dem Erdboden (in Grad Celsius) pro Tag in Aachen-Orsbach - Daten vom Deutschen Wetterdienst



**Abbildung A.2** Mittelwert der relativen Feuchte (in %) pro Tag in Aachen-Orsbach - Daten vom Deutschen Wetterdienst



**Abbildung A.3** Mittelwert des Gesamtbedeckungsgrades (in Achtel) pro Tag in Aachen-Orsbach - Daten vom Deutschen Wetterdienst