
Qualitätsparameter mobil erfasster Feinstaubsensordaten

INAUGURALDISSERTATION

zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften (Dr. rer. nat.)

an der

FAKULTÄT FÜR ANGEWANDTE INFORMATIK

UNIVERSITÄT AUGSBURG

von

CHRISTIAN RÖGER

2021

Gutachter: Prof. Dr. Jukka M. KRISP
Prof. Dr. Sabine TIMPF
Prof. Dr. Gerd BUZIEK

Datum der Disputation: 29.10.2021

Danksagung

An dieser Stelle möchte ich allen meinen großen Dank aussprechen, die mich bei der Anfertigung meiner Dissertation unterstützt haben.

Besonders danken möchte ich meinem Doktorvater, Prof. Dr. Jukka Krisp, für die hervorragende Betreuung bei der Durchführung meiner Arbeit. Die regelmäßigen Treffen, seine fachliche Kompetenz und seine Erfahrung halfen mir sehr bei der Anfertigung meiner Dissertation. Außerdem bedanke ich mich bei Prof. Dr. Sabine Timpf für die Begutachtung der Dissertation und die Unterstützung während meiner kompletten Zeit an der Universität Augsburg. Des weiteren möchte ich mich bei Prof. Dr. Gerd Buziek bedanken für die Begutachtung meiner Dissertation.

Ein besonderer Dank gilt auch meiner Kollegin Dr. Eva Nuhn, die mir während meiner Promotion immer mit guten Ratschlägen zur Seite stand. Außerdem bedanke ich mich bei Maja Kalinic, die einen großen Anteil am Gelingen dieser Dissertation trägt. Ohne die unzähligen Kaffeepausen, Konferenzreisen, Diskussionen und gemeinsam verfassten Publikationen hätte ich es deutlich schwerer gehabt, diese Dissertation anzufertigen. Darüber hinaus möchte ich mich bei den Mitgliedern meines Erfolgsteams Selina Thanheiser, Sally Jahn, Reiner Schwandt und Erik Petersen bedanken, mit denen ich während dem Schreiben meiner Dissertation häufig im Austausch stand.

Nicht zuletzt bedanke ich mich bei meinen Freunden, die mich stets in allen Lebenslagen unterstützt haben. Besonderer Dank gilt Andreas Wagner und Andre Brenner, die schon seit langer Zeit an meiner Seite stehen.

Meinen Eltern und meiner Schwester möchte ich ganz herzlich danken für die aufmerksame und vielseitige Unterstützung vor und während der Verfassung dieser Dissertation. Meiner Freundin Irada Ismayilova danke ich von ganzem Herzen für ihre uneingeschränkte Unterstützung und Motivation.

Zusammenfassung

Die vorliegende Dissertation zeigt, wie durch Geoinformationsanalysen mit Qualitätsfaktoren hochdynamischer Daten umgegangen werden kann. Die Qualität von Geodaten wird von unterschiedlichen Parametern beeinflusst. Mobil erfasste Feinstaubsensordaten stehen jedoch im Einfluss weiterer Faktoren. Deshalb wird mithilfe einer experimentellen Analyse überprüft, ob die Beschreibung der Genauigkeit dieser Art von Daten mithilfe von Standards zur Datenqualität von Geo- und Sensordaten möglich ist, oder ob dafür neue Kategorien definiert werden müssen. Außerdem werden statische und dynamische Dimensionen von Einflussfaktoren analysiert. In einem Experiment werden dafür Verkehrs- und Sensordaten erfasst. Analysiert werden die Einflussfaktoren *Positionierungsgenauigkeit*, *Sensorspezifikationen*, *Fahrgeschwindigkeit*, *Sensorausrichtung*, *Witterungsbedingungen* und *Umgebungsverkehr*. Die Methodik zur Analyse dieser Faktoren umfasst die Fachbereiche der angewandten Geoinformatik, der Geostatistik und des computerbasierten Sehens. Im Zuge einer experimentellen Fallstudie in Augsburg werden in einem Zeitraum von fünf Monaten drei Sensordatensätze aufgezeichnet. Mithilfe von sechs Feinstaubsensoren, einem Temperatur- und Luftfeuchtigkeitssensor, einem GNSS-Sensor und einer Videokamera werden Verkehrs- und Sensordaten erfasst. Die Ergebnisse der Analysen zeigen, dass die *Positionierungsgenauigkeit* sowohl einen statischen, als auch einen dynamischen Einfluss auf die Datenqualität nimmt. Statischen Einfluss nehmen die *Sensorspezifikationen* und die *Sensorausrichtung*. Von der *Fahrgeschwindigkeit*, den *Witterungsbedingungen* und dem *Umgebungsverkehr* geht ein dynamischer Einfluss aus. Mithilfe verschiedener Modelle aus dem maschinellen Lernen und den Einflussfaktoren als Prediktorvariablen kann darüber hinaus die Feinstaubkonzentration in der Luft vorhergesagt werden, was die Präsenz des Einflusses auf die Datenqualität unterstreicht. Die untersuchten Einflussfaktoren stellen sich als Ergänzung zu den bereits existierenden Kategorien der Datenqualität heraus. Im Zuge der Entwicklung neuer vernetzter Technologien in den Bereichen des autonomen Fahrens oder intelligenter Transportsysteme ist die Auseinandersetzung mit der Datenqualität hochdynamischer Sensordatensätze ein sehr wichtiger Punkt.

Summary

We show how to deal with highly dynamic factors of data quality in geoinformation analysis. Multiple parameters affect the quality of geodata. We examine, if there is a need of modifying common standards and guidelines for describing the quality of mobile recorded particulate matter data. Furthermore, we investigate static and dynamic dimensions of data quality factors. For our analysis, we take into account traffic and sensor data. Analysed factors include the *positioning accuracy*, *sensor specifications*, *driving velocity*, *sensor alignment*, *weather conditions* and *ambient traffic*. These factors are analysed with methods from applied geoinformatics, geostatistics and computer vision. We record three datasets within five months in Augsburg, Germany. Therefore, we use six particulate matter sensors, one temperature and humidity sensor, one *GNSS* module and a video camera. Findings indicate, that the *positioning accuracy* is influenced by static and dynamic factors. The *sensor specifications* and the *sensor alignment* are solely static. The *driving velocity*, *weather conditions* and the *ambient traffic* influence data quality dynamically. By implementing different machine learning models, we show that it is possible to predict particulate matter concentrations based on the analysed quality factors. We conclude that these factors are a complement to already existing data quality categories. With the implementation of novel, intelligent and integrated technologies in the fields of autonomous driving and Intelligent Transportation Systems (ITSs) comes the necessity of dealing with highly dynamic geodata.

Inhaltsverzeichnis

Danksagung	i
Zusammenfassung	ii
Summary	iii
Tabellenverzeichnis	viii
Abbildungsverzeichnis	x
Abkürzungsverzeichnis	xvii
1 Einleitung	1
1.1 Hinführung und Motivation	1
1.2 Relevanz des Themas	5
1.3 Fragestellungen und Ziele	8
1.4 Aufbau der Arbeit	10
2 Statisch und mobil erfasste Sensordaten	12
2.1 Statische Ansätze zur Erfassung von Verkehrsdaten	15
2.1.1 Induktive Zählschleifen (ILD)	15
2.1.2 Radarmessungen	17
2.1.3 Verkehrsüberwachungskameras	18
2.1.4 Bluetooth und WiFi	18
2.2 Mobile Erfassung von Verkehrsdaten	19
2.2.1 Floating Car Daten (FCD)	19
2.2.2 Erweiterte Floating Car Daten (xFCD)	24
2.2.3 Floating Car Observer (FCO)	25
2.3 Erfassung und Visualisierung von Feinstaubsensordaten	31
2.3.1 Citizen Science	31
2.3.2 Datenerfassung mit mobilen Feinstaubsensoren	32
2.3.3 Visualisierung von Sensordaten	36
3 Stand der Forschung zur Qualität von Sensordaten	40
3.1 Standards zur Beschreibung der Qualität von Sensordaten	40
3.2 Teilfaktoren der Qualität von mobil erfassten Feinstaubsensor- daten	43
3.3 Erfassung und Eliminierung von Positionierungsungenauigkeiten	45

3.4	Messunterschiede identischer und verschiedener Feinstaubsen- soren	48
3.4.1	Kalibrierung von Feinstaubsensoren	48
3.4.2	Gewichtete Interpolation unterschiedlicher Messwerte .	50
3.5	Witterungsbedingte Einflüsse auf die Datenqualität	51
3.6	Einfluss der Fahrgeschwindigkeit und der Sensorausrichtung auf die Datenqualität	52
3.7	Modellierung von Umwelt-Verkehr Beziehungen	53
3.8	Floating Car Observer zur Bestimmung des Umgebungsverkehrs	54
3.8.1	Feinstaub-basierte Floating Car Observer	56
3.8.2	Objektdetektionsalgorithmen für bildbasierte Floating Car Observer	56
4	Ausgewählte Methoden zur Analyse der Qualitätsparameter von Sens- ordaten	60
4.1	Methoden der Geostatistik zur Analyse von Sensordaten	62
4.1.1	Korrelationsanalyse zur Untersuchung von Zusammen- hängen verschiedener Datenreihen	63
4.1.2	Methoden des maschinellen Lernens	64
4.1.3	Hidden Markov Modelle in Mapmatching Algorithmen	69
4.1.4	Beschreibung und Vergleich von Trajektorien	71
4.2	Computerbasiertes Sehen (CV) in der Verkehrsdatenanalyse . .	73
4.2.1	Gefaltete Neuronale Netzwerke (CNN) in CV Anwen- dungen	74
4.2.2	You Only Look Once (YOLO) zur Objektdetektion in Bil- dern	76
5	Auswahl der Qualitätsparameter	78
6	Experimentelle Analyse von Einflüssen auf die Messgenauigkeit bei der mobilen Erfassung von Feinstaubsensordaten	80
6.1	Sensoren und Testgebiet der Fallstudie	80
6.1.1	Verwendete Sensoren zur Datenerfassung	80
6.1.2	Räumliche Abgrenzung des Testgebietes	83
6.2	Versuchsaufbauten für die Datenerfassung	84
6.2.1	Versuchsaufbau 1: statische Messungen mit Feinstaub- sensoren	84
6.2.2	Versuchsaufbau 2: statische und mobile Vergleichsmes- sungen	86

6.2.3	Versuchsaufbau 3: mobile Erfassung von Feinstaubsensordaten	90
6.3	Datenaufbereitung und Überblick über die resultierenden Datensätze	92
6.4	Analyse der Positionierungsgenauigkeit	100
6.4.1	Analyse der horizontalen und vertikalen Positionierungsabweichung	100
6.4.2	Anwendung von Mapmatching-Algorithmen	101
6.4.3	Vergleich von Trajektorien	103
6.5	Statische Analyse von Messunterschieden	106
6.6	Analyse der Fahrgeschwindigkeit und Sensorausrichtung	110
6.6.1	Vergleich von statisch und mobil erfassten Daten	110
6.6.2	Vergleich verschieden ausgerichteter Sensoren	113
6.7	Analyse witterungsbedingter Einflüsse	115
6.8	Analyse des Umgebungsverkehrs	119
6.9	Vorhersage der Feinstaubkonzentration mithilfe von Einflussfaktoren der Datenqualität	127
6.10	Einflüsse auf die Qualität mobil erfasster Feinstaubdaten	132
7	Diskussion und Limitationen	139
7.1	Auswahl der Einflussfaktoren auf die Qualität von mobil erfassten Feinstaubsensordaten	139
7.2	Kritische Auseinandersetzung mit der Einordnung in statische und dynamische Faktoren	142
7.3	Kalibrierung und Korrektur von Feinstaubsensordaten auf Basis von Qualitätsfaktoren	143
7.4	Qualitätsfaktoren im Zusammenhang mit Standards und Richtlinien	145
7.5	Temporale und räumliche Übertragbarkeit der Fallstudie	147
7.6	Limitationen der in der Fallstudie durchgeführten Experimente	150
7.6.1	Limitationen der Genauigkeit bei der Fahrzeugdetektion aus Verkehrsvideos	150
7.6.2	Bestimmung des Einflusses von Bewegungsgeschwindigkeit und Sensorausrichtung	154
7.6.3	Limitationen der Modell- und Variablenauswahl für die Vorhersage der Feinstaubkonzentration	155
7.7	Umgebungsverkehr - Störfaktor oder wichtiger Bestandteil von Feinstaubmessungen?	156

8	Schlussfolgerungen und Ausblick	158
8.1	Schlussfolgerungen	158
8.2	Zukünftige Arbeiten	161
	Literaturverzeichnis	165

Tabellenverzeichnis

1	(+)Vor- und (-)Nachteile des motorisierten Individualverkehrs (MIV), Taxiverkehrs, öffentlichen Personennahverkehrs (ÖPNV) und Lieferverkehrs für die Erfassung von Floating Car Daten. Ausgeglichenere Eignung ist mit einem (o) gekennzeichnet. . . .	21
2	Bewegungsparameter von Floating Car Daten nach Ranacher et al. (2016)	23
3	Optimale Samplingraten für Floating Car Daten nach Ranacher et al. (2016)	24
4	In <i>SAEJ1979</i> definierte auslesbare Fahrzeugparameter, ihre PIDs und ihre Einheit	25
5	Sensoren für die Implementierung eines Floating Car Observers, ihre Funktionsweise und ihr Blickfeld	26
6	Wachstumsfunktionen für die Luftfeuchtigkeitskorrektur von Feinstaubsensordaten nach Streibl (2017)	52
7	Spezifikationen des <i>DHT22</i> -Sensors	81
8	Spezifikationen des <i>Plantower PMS3003</i> Sensors und des <i>AlphaSense OPC-N3</i> Sensors	82
9	Präsenz (+) und Abwesenheit (-) von Datentypen bei Versuchsaufbau 1, 2 und 3	92
10	Übersicht über die Spaltennamen der aus der Datenaufbereitung resultierenden datei mit Datentypen und Beispielen (* = nur Versuchsaufbau 2 und 3, ** = nur Versuchsaufbau 3)	95
11	Messunterschiede bei zeitgleicher statischer und mobiler Datenerfassung bei unterschiedlichen Geschwindigkeiten. Aufschlüsselung nach Sensorausrichtung <i>a</i> (frontal, seitlich und entgegen der Fahrtrichtung) und Partikeldurchmesser PM_x (PM_1 , $PM_{2.5}$, PM_{10}). Farbliche Hervorhebung des jeweils niedrigsten und höchsten Wertes einer Kombination.	112
12	Genauigkeit, Sensitivität und Spezifität für die Splitoptionen 50/50, 60/40 und 70/30 eines RF Modells zur Vorhersage von Umgebungsverkehr mithilfe von Feinstaubwerten als Prediktoren	124
13	Genauigkeit, Sensitivität und Spezifität für die Splitoptionen 50/50, 60/40 und 70/30 eines RF Modells zur Vorhersage von Umgebungsverkehr mithilfe von Feinstaubwerten als Prediktoren mit $m_{try} = 4$ und $n_{tree} = 200$	126

14	RMSE-Werte für <i>Random Forest</i> , <i>Gradient Boost</i> und <i>Lineare Regression</i> bei der Vorhersage von PM_1 , $PM_{2.5}$ und PM_{10}	129
15	Einordnung der (nicht-)numerisch beschreibbaren Einflussfaktoren in die Kategorien <i>Genauigkeit</i> und <i>Vertrauen</i>	146
16	Trajektorien mit ihrer Anzahl an Observationen sowie Datenpunkten mit Umgebungsverkehr (identifiziert und händisch bestimmt)	153

Abbildungsverzeichnis

1	Überblick über die verwandten Fachbereiche (äußerer Ring), verwendeten Methoden (mittlerer Ring) und genutzten Datentypen (innerer Ring) der Dissertation	5
2	Aufbau der Arbeit, dargestellt als Flussdiagramm	10
3	Schematische Darstellung der Erfassung von Floating Car Daten	20
4	Darstellung einer mithilfe der <i>myTracks</i> - Applikation erstellten Fahrtrajektorie	21
5	Schematische Darstellung eines auf Taxis basierenden FCD-Systems nach Brockfeld et al. (2007)	23
6	Schematische Darstellung der Abstandsmessung mit Infrarotsensoren in FCO-Anwendungen	27
7	Schematische Darstellung der Abstandsmessung mit Laserscannern in FCO-Anwendungen	27
8	Schematische Darstellung der Abstandsmessung mit Ultraschallsensoren in FCO-Anwendungen	28
9	Schematische Darstellung der mobilen Verkehrszählung mit Videokameras in FCO-Anwendungen	29
10	Schematische Darstellung der mobilen Verkehrszählung mit Bluetooth- und WiFi Sensoren in FCO-Anwendungen	29
11	Schematische Darstellung der mobilen Verkehrszählung mit Feinstaubsensoren in FCO-Anwendungen	30
12	CO ₂ -Emissionen von Fahrzeugen in Mönchengladbach, dargestellt als Kartogramm. Quelle: Röger et al. (2019)	37
13	Qualitätsfaktoren einer mobilen Sensormessung, dargestellt als <i>Star Plot Map</i> . Quelle: Röger (2019)	39
14	Anwendungsgebiete des Umweltmonitoring in der Geoinformatik	61
15	Schematische Darstellung eines <i>Hidden Markov Modells</i> für zwei Status (S_1, S_2) und drei Observationen (Y_1, Y_2, Y_3)	70
16	Verschiedene Ansätze zum Vergleich von Kurven: <i>Partial Curve Mapping</i> (oben), <i>Discrete Fréchet</i> (mitte) und die <i>Fläche zwischen den Kurven</i> (unten)	72

17	Schematischer Aufbau eines gefalteten neuronalen Netzwerkes. Einer Eingabeebene werden im <i>Convolutional Layer</i> Filter angewandt. Im <i>Pooling Layer</i> werden die Ergebnisse aggregiert. Nach beliebig vielen <i>Convolutional</i> - und <i>Pooling</i> - Ebenen folgt ein <i>Fully Connected</i> -Layer. Die Ausgabebene legt die finalen Klassen fest.	75
18	Einflussfaktoren auf die Datenqualität mobil erfasster Feinstaubsensordaten	78
19	<i>NEO-6M</i> -Sensor zur Positionsbestimmung	80
20	<i>DHT22</i> -Sensor zur Messung der Luftfeuchtigkeit und der Temperatur	81
21	<i>Plantower PMS3003</i> Sensor (links) sowie <i>Alphasense OPC-N3</i> Sensor (rechts)	82
22	Karte mit der räumlichen Abgrenzung des Studiengebietes	83
23	Feinstaubsensor im Winkelrohr zur Reduzierung von witterungsbedingten Einflüssen	85
24	Anordnung von fünf <i>Alphasense OPC-N3</i> Sensoren sowie von einem <i>PMS3003</i> Sensor in Winkelrohren	85
25	Vergleichsmessung von fünf <i>OPC-N3</i> sowie einem <i>PMS3003</i> Feinstaubsensor	86
26	Ausrichtung der Feinstaubsensoren entgegen, mit und seitlich der Fahrtrichtung	87
27	Schematische Zeichnung mit den Positionen der Feinstaubsensoren (rote Kreise), des <i>GNSS</i> -Moduls (grüner Kreis) sowie der zentralen Stromquelle (gelbes Rechteck)	87
28	Architektur des Messsystems zu Versuchsaufbau 2	88
29	Testgebiet zur Überprüfung des Einflusses von Sensorausrichtung und Fahrtgeschwindigkeit auf Messwerte von Feinstaubsensoren	89
30	Schematische Zeichnung mit den Positionen der Feinstaubsensoren (rote Kreise), des <i>GNSS</i> -Moduls (grüner Kreis), der zentralen Stromquelle (gelbes Rechteck) sowie der Videokamera zur Aufzeichnung des Umgebungsverkehrs (violetter Kreis)	90
31	Architektur des Messsystems zu Versuchsaufbau 3	91
32	Flowchart zur Datenaufbereitung und Detektion der Fahrzeuge aus den Verkehrsvideos	94
33	Temperatur- und Luftfeuchtigkeitsverlauf während der Erfassung von Datensatz 1	96

34	Überblick über die mit Versuchsaufbau 3 erfassten Daten mit 23.725 Observationen	97
35	Histogramme für die Geschwindigkeit und die Feinstaubwerte (PM_1 , $PM_{2.5}$, PM_{10}) der einzelnen Observationen	98
36	Histogramme für die Temperatur und Luftfeuchtigkeit der einzelnen Observationen	99
37	Histogramm für die Anzahl der Fahrzeuge im Umgebungsverkehr	100
38	Histogramme für die vertikalen ($VDOP$) und horizontalen ($HDOP$) Abweichungen der Positionierungsgenauigkeit im aufgenommenen Datensatz	101
39	Verschiedene Ansätze zum Mapmatching, angewandt auf die Datenpunkte der Fallstudie: originale Punkte (a), Nutzung des <i>Snapping</i> -Tools mit zehn Metern Suchradius (b), Nutzung des <i>Snapping</i> -Tools mit 20 Metern Suchradius (c), <i>nearest</i> -Option der <i>noiseplanet</i> -Umgebung (d) und <i>hidden markov model</i> -Option der <i>noiseplanet</i> Umgebung (e)	102
40	Balkendiagramme für die Parameter PCM (links) und DF (rechts) der einzelnen Trajektorien beim Vergleich der originalen Kurve mit der korrigierten Kurve	104
41	Balkendiagramme für den Unterschied der <i>Kurvenlängen</i> (links) und die <i>Fläche zwischen den Kurven</i> (rechts) der einzelnen Trajektorien beim Vergleich der originalen Kurve mit der korrigierten Kurve	105
42	Ergebnis des Sensorenvergleichs für PM_1 Feinstaub (stündliche Mittelwerte)	106
43	Korrelationsmatrizen für fünf $OPC-N_3$ sowie einen PMS_{3003} Sensor für PM_1 , $PM_{2.5}$ und PM_{10} Feinstaub (stündliche Mittelwerte)	107
44	Ergebnis des Sensorenvergleichs für $PM_{2.5}$ Feinstaub (stündliche Mittelwerte)	108
45	Ergebnis des Sensorenvergleichs für PM_{10} Feinstaub (stündliche Mittelwerte)	109
46	Ablaufdiagramm zur Analyse der Messunterschiede statischer und bewegter Sensoren unter Berücksichtigung der Fahrgeschwindigkeit und der Sensorausrichtungen	110

47	Vergleichsmessung (PM_{10}) von zwei statischen Sensoren (Sensor 1 und Sensor 2, rote Punkte) und drei bewegten Sensoren (frontal, seitlich und entgegen der Fahrtrichtung) bei einer Geschwindigkeit von 20 km/h	111
48	Korrelationen der Messwerte aus dem nach vorne ausgerichteten Sensor zu den dazugehörigen Werten aus dem seitlich ausgerichteten Sensor	114
49	Korrelationen der Messwerte aus dem nach vorne ausgerichteten Sensor zu den dazugehörigen Werten aus dem seitlich ausgerichteten Sensor bei $v > 90\text{km/h}$	114
50	Vergleich der PM_1 , $PM_{2.5}$ und PM_{10} Feinstaubwerte mit dem Luftfeuchtigkeitsverlauf von Datensatz 1	115
51	Ablaufdiagramm für die Luftfeuchtigkeitskorrektur der Feinstaubwerte	117
52	Vergleich der Feinstaubwerte ($PM_{2.5}$) nach Anwendung unterschiedlicher Wachstumsfunktionen (Hänel, Soneja, Combo, Skupin) mit den originalen Werten für Trajektorie 17 bei einer durchschnittlichen Luftfeuchtigkeit von 61%	117
53	Vergleich der Feinstaubwerte ($PM_{2.5}$) nach Anwendung unterschiedlicher Wachstumsfunktionen (Hänel, Soneja, Combo, Skupin) mit den originalen Werten für Trajektorie 25 bei einer durchschnittlichen Luftfeuchtigkeit von 20%	118
54	Ablaufdiagramm für die Implementierung eines FCO auf Basis von Feinstaubwerten	120
55	Korrelationsnetzwerk für die Prediktoren des Random Forest Modells	122
56	<i>Feature Importance</i> Analyse für die Prediktoren des Random Forest Modells	123
57	Hyperparameter-Tuning für den m_{try} -Wert: Genauigkeitswerte für $m_{try} = 1, \dots, m_{try} = 15$	125
58	Hyperparameter-Tuning für den n_{tree} -Wert: Genauigkeitswerte für $n_{tree} = 100, \dots, n_{tree} = 1000$	125
59	ROC-Kurven für die Splitoptionen 50/50, 60/40 und 70/30	126
60	Ablaufdiagramm für die Vorhersage der Feinstaubkonzentration mithilfe von Einflussfaktoren der Datenqualität	128

61	Vorhersage der PM_1 Feinstaubkonzentration mithilfe von Luftfeuchtigkeit, Bewegungsgeschwindigkeit und der Anzahl von Fahrzeugen im Umgebungsverkehr mit <i>Random Forest</i> , <i>Gradient Boost</i> und <i>Linearer Regression</i> zwischen Punkt-ID 800 und 900 .	129
62	Vorhersage der $PM_{2.5}$ Feinstaubkonzentration mithilfe von Luftfeuchtigkeit, Bewegungsgeschwindigkeit und der Anzahl von Fahrzeugen im Umgebungsverkehr mit <i>Random Forest</i> , <i>Gradient Boost</i> und <i>Linearer Regression</i> zwischen Punkt-ID 100 und 200 .	130
63	<i>Variable Importance</i> Analyse für die Prediktorvariablen der genutzten Modelle	131
64	Zusammenfassende Unterteilung der Einflussfaktoren <i>Positionierungsgenauigkeit</i> , <i>Sensorspezifikationen</i> , <i>Messgleichheit identischer Sensoren</i> , <i>Witterung</i> , <i>Bewegungsgeschwindigkeit</i> , <i>Sensorausrichtung</i> und <i>Umgebungsverkehr</i> in statische und dynamische Faktoren .	133
65	Verbindungen der in der Studie untersuchten Einflussfaktoren mit den Kategorien der Qualität von Geodaten nach <i>ISO 19157:2013</i>	147
66	Detektion unterschiedlicher Verkehrssituationen: korrekte Objektdetektion (a), doppelte Detektion gleicher Fahrzeuge (b), korrekte Detektion mit falschen Klassifizierungen (c), Detektion von parkenden Fahrzeugen (d), nicht funktionierende Detektion bei starkem Regen (e) und nicht funktionierende Detektion bei schlechten Lichtverhältnissen (f)	151

Abkürzungsverzeichnis

- aGPS** assisted Global Positioning System (assistiertes globales Positionsbestimmungssystem)
- AUC** Area Under the Curve (Fläche unter der Kurve)
- AUROC** Area Under the Receiver Operating Characteristic (Fläche unter der ROC-Kurve)
- CNN** Convolutional Neural Network (gefaltetes neuronales Netzwerk)
- CO** Carbon Monoxide (Kohlenstoffmonoxid)
- CO₂** Carbon Dioxide (Kohlenstoffdioxid)
- dGPS** Differential Global Positioning System (differenzielles globales Positionsbestimmungssystem)
- DLR** deutsches Zentrum für Luft-und Raumfahrt
- DOP** Dilution Of Precision (Verringerung der Genauigkeit)
- FCD** Floating Car Data (Floating Car Daten)
- FCO** Floating Car Observer
- FPN** Feature Pyramid Network
- FPS** Frames per Second (Bilder pro Sekunde)
- GB** Gradient Boost
- GNSS** Global Navigation Satellite System (globales Navigationssatellitensystem)
- GPR** Gaussian Process Regression (Gaußprozess-Regressionsmodell)
- GPS** Global Positioning System (Globales Positionsbestimmungssystem)
- GPU** Graphics Processing Unit (Grafikprozessor)
- HDOP** Horizontal Dilution Of Precision (horizontale Verringerung der Genauigkeit)
- HMM** Hidden Markov Model
- ILD** Inductive Loop Detector (Induktive Zählschleife)

IoT Internet of Things (Internet der Dinge)

KI Künstliche Intelligenz

LBS Location-Based Service

LKW Lastkraftwagen

LoRa Long Range Wide Area Network

LR Lineare Regression

MAC Media-Access-Control

MIV Motorisierter Individualverkehr

ML Maschinelles Lernen

OBD On Board Diagnostics (On-Board Diagnose)

ÖPNV öffentlicher Personennahverkehr

OPC Optical Particle Counter (optischer Partikelzähler)

PANet Path Aggregation Network

PKW Personenkraftwagen

PM Particulate Matter (Feinstaub)

RCNN Region-based Convolutional Neural Networks (Regionen-basierende gefaltete neuronale Netzwerke)

RF Random Forest

RMSE Root Mean Square Error (Mittlere quadratische Abweichung)

ROC Receiver Operating Characteristic (Operationscharakteristik eines Beobachters)

SAM Spatial Attention Module

SAT Self-Adversarial Training

SDG Sustainable Development Goals (Ziele zur nachhaltigen Entwicklung)

SSD Single Shot Multibox Detector

UAV Unmanned Aerial Vehicle (unbemanntes Luftfahrzeug)

VDOP Vertical Dilution Of Precision (vertikale Verringerung der Genauigkeit)

xFCD extended Floating Car Data (erweiterte Floating Car Daten)

YOLO You Only Look Once

1 Einleitung

Some air quality indicators, such as [...] Particulate Matter, may see increased adoption and use of low-cost sensors in the coming decade. [...] These activities need to be paralleled by efforts to raise awareness of the current data quality assurance mechanisms. - Fritz et al. (2019)

1.1 Hinführung und Motivation

Beim Einsatz von Sensoren zur Datenerfassung ist die Überprüfung der Datenqualität ein nicht zu vernachlässigender Punkt. Informationen darüber werden in der Regel in den Metadaten gespeichert. Dafür existieren international anerkannte Standards wie die *ISO 19115-2:2019* Richtlinie (ISO Central Secretary 2019). Bestandteile der Datenqualität von Geodaten werden dagegen in der *ISO 19157:2013* festgelegt. Diese definiert Qualitätsparameter wie die *Vollständigkeit* eines Datensatzes, die *logische Konsistenz*, die *Positionsgenauigkeit*, die *temporale Genauigkeit*, die *thematische Genauigkeit* und die *Benutzerfreundlichkeit* (ISO Central Secretary 2013). Sensordaten verstehen sich als eine Art von Geodaten. Die in der Regel multivariaten Datensätze bringen Anforderungen an die Datenqualität mit sich, die über die bestehenden ISO-Standards hinaus gehen. Folglich definieren Klein et al. (2009) die fünf Parameter *Genauigkeit*, *Vertrauen*, *Vollständigkeit*, *Volumen* und *Aktualität*, mit denen die Qualität von Sensordaten beschrieben werden können. Ein besonderes Augenmerk liegt hierbei auf der *Genauigkeit* und dem *Vertrauen*, die den Einfluss statischer und dynamischer Einflussfaktoren umfassen (Klein et al. 2009).

Welche statischen Parameter nehmen Einfluss auf die Genauigkeit und welche dynamischen Parameter nehmen Einfluss auf das Vertrauen? Erste Indizien zur Einordnung finden sich in Röger (2021). Dort werden die Sensorspezifikationen als wichtiger Punkt genannt. Auch wenn hochpräzise Messgeräte zur Datenbeschaffung zur Verfügung stehen, liefern diese nur eine Annäherung an die Realität und können damit nie ganzheitlich genau sein. Dass vergleichsweise günstige Messgeräte noch stärker davon betroffen sind, lässt sich in der Regel durch den Vergleich von Sensorspezifikationen nachvollziehen. Zu den entscheidenden Parametern gehören hier unter anderen die Präzision, die Messverzögerung und der mögliche Temperaturbereich. Einige Hersteller bieten darüber hinaus auch Diagramme mit Vergleichsmessungen des jeweiligen Messgerätes mit hochpräzisen Referenzgeräten an. Es zeigt sich, dass nicht nur die Sensorspezifikationen einen Qualitätsfaktor von erfassten Sens-

ordaten ausmachen, sondern auch externe Einflüsse wie die Witterung oder Störquellen wie der Umgebungsverkehr eine Rolle spielen. Darüber hinaus werden auch die Parameter der Fahrgeschwindigkeit und der Sensorausrichtung als Einflüsse genannt (Röger 2021). In der Literatur finden sich zahlreiche Methoden, mithilfe deren etwaige Einflüsse auf die Datenqualität minimiert werden können. Dazu gehören die Luftfeuchtigkeitskorrektur nach Di Antonio et al. (2018) und die Kalibrierung kostengünstiger Sensoren mithilfe von Referenzgeräten (Wang et al. 2019). Auch nach der Anwendung von Korrekturalgorithmen darf nicht davon ausgegangen werden, dass die bearbeiteten Werte der in der Natur vorkommenden Feinstaubkonzentration entsprechen. Maag et al. (2018) unterstreichen, dass aufgrund der Komplexität und Dynamik der in der Natur vorkommenden Messvoraussetzungen auch kalibrierte Sensoren abweichende Werte liefern können. Hagler et al. (2018) sehen die Anwendung immer komplexer werdender Korrekturalgorithmen kritisch, vor allem wenn langfristig erfasste Sensordaten mit kurzfristig gemessenen Vergleichswerten kalibriert werden. Allgemein gilt sowohl für korrigierte, als auch für unkorrigierte Werte, dass die Datenqualität bei jeglicher Analyse beachtet und diskutiert werden sollte (Fritz et al. 2019).

Die öffentliche Wahrnehmung von Effekten hoher Feinstaubkonzentrationen im urbanen Raum nimmt zu. Yang et al. (2017) observieren Korrelationen zwischen Feinstaub-bedingten Lungenkrebsfällen und damit verwandten Anfragen in Suchmaschinen in 31 chinesischen Städten. Neben Atemwegserkrankungen können hohe Feinstaubwerte unter anderem auch zu Schäden der DNA, Herzproblemen, Entzündungen, Frühgeburten und Atherosklerose führen (Mukherjee et al. 2017b). Entscheidend für die menschliche Aufnahme von Feinstaub ist der Partikeldurchmesser. Typischerweise wird zwischen PM_{10} ($\varnothing < 10 \mu m$), $PM_{2.5}$ ($\varnothing < 2,5 \mu m$) und PM_1 ($\varnothing < 1 \mu m$) unterschieden (Ariola et al. 2006). Dabei dringen Partikel mit fünf bis zehn μm Durchmesser bis in den Nasen-Rachenraum vor und Partikel mit Durchmessern von drei bis fünf μm bis in die Luftröhre. Ein Angriffsort für ein bis drei μm große Partikel sind die Bronchien und Bronchiolen. Besonders feine Feinstaubpartikel dringen bis zu den Lungenbläschen vor ($0,1$ bis $1 \mu m$) (Kelly et al. 2012). Hinsichtlich der Entstehung von Feinstaub unterscheidet man zwischen primärem und sekundärem Feinstaub. Partikel, die direkt an der Emissionsquelle entstehen, werden als primär bezeichnet. Bilden sich Feinstaubpartikel erst durch weitere Prozesse und Reaktionen in der Atmosphäre, spricht man von sekundärem Feinstaub (Parvez et al. 2017). Quellen der Partikelemission können sowohl natürlich, als auch menschlich sein. Natürliche Quellen von

Feinstaub sind beispielsweise Vulkanausbrüche, Bodenerosion und die Entstehung von Aerosolen. Menschengemachter Feinstaub entsteht primär bei Verbrennungsprozessen. Dementsprechend stellen der Verkehr, die Landwirtschaft und die Industrie wichtige Faktoren dar. Ein besonderes Augenmerk liegt dabei auf dem Verkehr. Hier entsteht nicht nur Feinstaub durch die Verwertung fossiler Brennstoffe, sondern auch durch Brems- und Reifenabrieb sowie durch Aufwirbelungseffekte beim Befahren einer Straße. Im Jahr 2018 sind die Hauptverursacher von PM_{10} -Feinstaub mit 45,4 Prozent die Industrieprozesse, mit 18,5 Prozent der Verkehr, mit 14,5 Prozent die Landwirtschaft und mit 11,6 Prozent Haushalte und Kleinverbraucher¹.

Die Tatsache, dass die Verminderung der Luftverschmutzung durch Feinstaub Teil der 17 Ziele zur nachhaltigen Entwicklung (SDG) der vereinten Nationen (UN) ist, unterstreicht die Relevanz des Themas (Fritz et al. 2019). Herausforderungen stellen sich dabei allerdings nicht nur hinsichtlich der Verringerung der Feinstaubemissionen, sondern auch im Monitoring des *status quo* der Luftverschmutzung. Traditionelle Datenströme sehen dabei vor, Erkenntnisse über Regierungsbehörden, nationale Statistikbüros und internationale Organisationen zu erlangen. Aufgrund der großen Nachfrage nach Daten rücken jedoch auch nicht-traditionelle Datenströme zur Informationsbeschaffung in den Fokus. Dazu gehören die Errichtung von Sensornetzwerken, die Nutzung von Methoden aus der Fernerkundung und Citizen Science (Fritz et al. 2019). Vorrangig das in der letzten Dekade stark gestiegene Angebot an kostengünstigen Feinstaubsensoren eröffnet neue Möglichkeiten der Nutzung von nicht-traditionellen Datenströmen zum Monitoring von Feinstaubkonzentrationen. Morawska et al. (2018) sprechen in dem Zusammenhang gar von einer revolutionären Wende in der Erfassung von Feinstaubdaten. Portale wie luftdaten.info² oder enviroCar³ ermöglichen es Bürgern, eigene Sensordaten in eine zentrale Datenbank zu übertragen und somit Wissenschaftler bei der Datenbeschaffung zu unterstützen.

Die Datenerfassung mit Feinstaubsensoren kann auf zwei grundlegende Wege erfolgen. Zum einen gibt es die Möglichkeit der statischen Messung (Qin et al. 2020). Dabei werden die vorgesehenen Messgeräte fest an definierten Punkten installiert. Folglich werden dabei Daten erfasst, die die Feinstaubkonzentration einer bestimmten Geolokation abbilden. Ein prominenter Vertreter des statischen Messansatzes sind die offiziellen Messstationen innerhalb Deutsch-

¹<https://de.statista.com/statistik/daten/studie/1090888/umfrage/verteilung-der-der-luftverschmutzung-nach-verursachergruppen-in-deutschland/>, 05/21

²<https://luftdaten.info>, 05/21

³<https://envirocar.org>, 05/21

lands. Wird eine höhere räumliche Abdeckung erwünscht, kann das über die Ausdehnung des bestehenden Sensornetzwerkes erreicht werden. Da dieser Vorgang als kostenintensiv gilt und sich die Suche nach potentiellen neuen Standorten einer Messstation oft schwierig darstellt, rücken auch mobile Messansätze immer weiter in den Fokus. Diese Art der Datenerfassung liegt vor, wenn sich ein Sensor durch den Raum bewegt (Qin et al. 2020). Über die Art der Fortbewegung kann dabei vom Nutzer selbst entschieden werden. Beispiele in der Literatur umfassen den Transport von Sensoren unter anderem mithilfe von Personenkraftwagen, Bussen und Fahrrädern (Mazzoleni et al. 2004; Goyal et al. 2018; Liu et al. 2015). Trotz des Vorteiles der größeren räumlichen Abdeckung mobiler Messansätze gibt es für einen flächendeckenden Einsatz der Technologie noch Einschränkungen wie die bislang weitestgehend wenig beforschten Einflüsse auf die Datenqualität (Röger 2021). In anderen Fachbereichen wie der Verkehrsanalyse sind mobile Wege der Datenbeschaffung dagegen schon etabliert. Auf Floating Car Daten (FCD) basierende Ansätze nutzen Fahrzeuge als bewegliche Sensoren, die in einer definierten Frequenz Informationen über ihre aktuelle Position abspeichern. Durch die Kombination mehrerer mit Sensoren ausgestatteter Fahrzeuge können so Parameter wie der Verkehrsfluss berechnet werden. Die hohe räumliche Abdeckung macht aus FCD eine leistungsfähige Alternative zu herkömmlichen Ansätzen wie Radarsensoren oder induktive Schleifen (De Fabritiis et al. 2008).

Faktoren der Datenqualität mobil erfasster Sensordaten sind das übergeordnete Thema der vorliegenden Dissertation. Der gewählte ingenieurwissenschaftliche Ansatz sieht die Nutzung diverser Datentypen und Methoden aus unterschiedlichen mit der angewandten Geoinformatik verwandten Fachbereichen vor.

Abbildung 1 zeigt eine Einordnung der Arbeit in verwandte Fachbereiche, Methoden und Daten. Die Dissertation ist primär in der angewandten Geoinformatik angesiedelt. Diese bildet den Schnittpunkt zwischen der Geographie und der (angewandten) Informatik. Ebenfalls adressiert werden Teilbereiche der Umwelt- und Verkehrswissenschaften. Aus den angesprochenen Fachbereichen ergeben sich auch die Methoden, die in der Fallstudie dieser Arbeit zur Datenanalyse gewählt werden. Das maschinelle Lernen, die Geostatistik und die räumliche Analyse sind häufig genutzte Werkzeuge in der (Geo-)Informatik. Floating Car Observer produzieren Daten über den Umgebungsverkehr eines Fahrzeuges. Analysiert in der Dissertation werden darüber hinaus (erweiterte) Floating Car Daten sowie Feinstaubsensordaten.

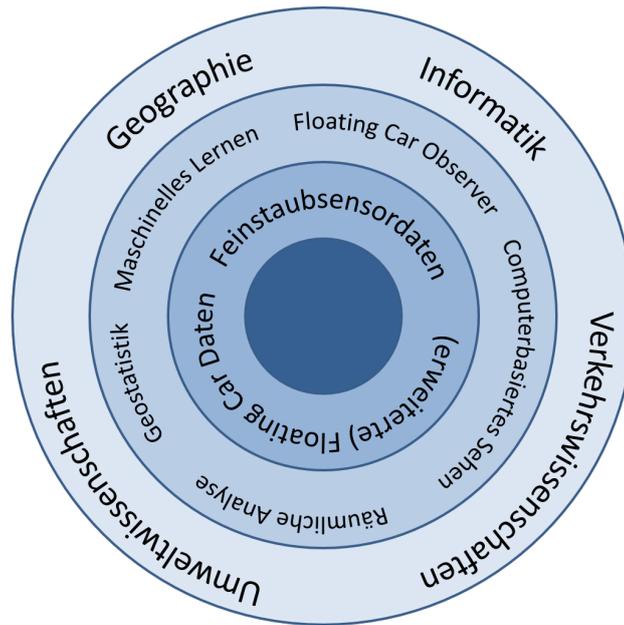


Abbildung 1: Überblick über die verwandten Fachbereiche (äüßerer Ring), verwendeten Methoden (mittlerer Ring) und genutzten Datentypen (innerer Ring) der Dissertation

1.2 Relevanz des Themas

Warum ist die Erfassung von Feinstaubdaten wichtig? Die Relevanz von Feinstaubmessungen zeigt sich unter anderem in den gesundheitlichen Risiken hoher Luftverschmutzung (Fritz et al. 2019). Bekannte Folgen hoher Feinstaubkonzentrationen können unter anderem Entzündungen, Lungenkrebs und Herzprobleme sein (Mukherjee et al. 2017b). Es zeigt sich allerdings auch, dass Feinstaubpartikel zur Übertragung von Viren beitragen. Tung et al. (2021) sehen bei hohen PM-Konzentrationen ein erhöhtes Risiko für die Infektion mit dem Coronavirus SARS-CoV-2. Feinstaub sorgt dabei für einen Anstieg des Angiotensin-konvertierenden Enzym 2 (ACE2) in der Lunge. Das vereinfacht es Coronaviren, über die Atemwege aufgenommen zu werden. Darüber hinaus lassen sich Coronaviren in PM_{10} -Proben aus Bergamo (Italien) feststellen (Setti et al. 2020). Yao et al. (2020) finden Korrelationen zwischen der SARS-CoV-2 Sterberate und der Feinstaubkonzentration in einer Studie in 49 chinesischen Städten (Setti et al. 2020; Yao et al. 2020; Tung et al. 2021).

Warum ist die Erforschung von Qualitätsfaktoren mobil erfasster Feinstaubsensordaten relevant? Sensornetzwerke revolutionieren die Überwachung der

Luftqualität. Die steigende Verfügbarkeit von Sensoren sorgt dafür, dass eine hohe Quantität an Daten teils frei verfügbar ist. Reddy et al. (2007) sehen in der Überwachung der Datenqualität eine der größten Herausforderungen bei der Implementierung von Sensornetzwerken. Campbell et al. (2013) unterstreichen in diesem Zusammenhang die Notwendigkeit von gemeinschaftsweiten Standards, damit die Qualität dieser Daten eingesehen und verglichen werden kann. Die Implementierung von internationalen Standards ist wichtig, denn ohne diese ist es unmöglich, die Qualität verschiedener Produkte oder Datensätze miteinander zu vergleichen (Graves 2000). Nach Klein et al. (2009) geht die Beschreibung der Qualität von Sensordaten über die standardisierten Anforderungen für Geodaten in *ISO 19157:2013* hinaus. Sensordaten gelten zwar als Geodaten, unterscheiden sich jedoch in verschiedenen Merkmalen. Sensoren können fehleranfällig sein, was in unpräzisen Messungen und Datenlücken resultiert. Dazu kann es durch die Interpolation von Messwerten aufgrund von fehlenden oder fehlerhaften Werten zu einem verstärkten Auftritt von Datenrauschen kommen, was die Qualität noch weiter beeinträchtigt. Klein et al. (2009) unterstreichen, dass die Identifikation von Faktoren der Datenqualität daher ein sehr wichtiger Punkt ist. Zur Beschreibung des numerischen Gesamtfehlerwertes der Qualität eines Datensatzes nennen die Autoren die Parameter der *Genauigkeit* und des *Vertrauens*. Die Genauigkeit umfasst statische Einflüsse, während das Vertrauen dynamische Einflüsse beschreibt (Klein et al. 2009; Campbell et al. 2013).

Inwiefern unterscheiden sich mobil und statisch erfasste Feinstaubsensordaten hinsichtlich ihrer Datenqualität? Statisch erfasste Feinstaubsensordaten unterscheiden sich bezüglich ihrer Qualitätsparameter maßgeblich von mobil erfassten Feinstaubsensordaten. Über die Einflussfaktoren der Genauigkeit und des Vertrauens statisch erfasster Feinstaubsensordaten existiert bereits eine Reihe an Literatur und Methodik. Labor- und Feldevaluationen von Feinstaubsensoren zeigen hierbei oft größtenteils identische Ergebnisse. Nach Mukherjee et al. (2017a) spielen die Witterungsverhältnisse und die Genauigkeit der verwendeten Messgeräte eine tragende Rolle bei der Frage nach der Datenqualität. Zu diesem Schluss kommen ebenfalls Bulot et al. (2019) und Bai et al. (2020). Um den Einfluss beider Faktoren zu minimieren, verweisen einige Studien auf Korrekturalgorithmen. So lassen sich beispielsweise durch die Luftfeuchtigkeit auftretende Messungenauigkeiten nach Chakrabarti et al. (2004), Streibl (2017) sowie Di Antonio et al. (2018) durch die Anwendung von Wachstumsfunktionen minimieren. Auch für den Ausgleich von Messungenauigkeiten, die durch die Spezifikationen von vergleichsweise günstigen

Feinstaubsensoren auftreten, existieren Lösungen. Mögliche Kalibrierungsansätze nutzen häufig Methoden des maschinellen Lernens und sind unter anderem in Wang et al. (2015), Wang et al. (2019) und Si et al. (2020) einzusehen. Aus zwei Gründen ist die Beschreibung der Qualität mobil erfasster Feinstaubsensordaten eine Forschungslücke. Zum einen ist bisher weitestgehend unerforscht, welche Teilfaktoren Einfluss auf die Genauigkeit und das Vertrauen der Daten nehmen. Zum anderen existiert daraus folgend ein Mangel an Methoden, um den Einfluss möglicher Quellen von Messungenauigkeiten durch die Anwendung von Korrekturalgorithmen oder Kalibrierungsansätzen zu minimieren. Feinstaubdaten statischer und mobiler Erfassungsansätze unterscheiden sich maßgeblich durch die räumliche Dimension. Während fest installierte Messgeräte Feinstaubkonzentrationen durchgehend am selben Ort erfassen, erreichen bewegte Sensoren eine deutlich höhere räumliche Abdeckung eines Gebietes. Der mobile Messansatz hat allerdings nicht nur Vorteile gegenüber statisch installierten Messgeräten. So observieren beispielsweise Goyal et al. (2018) plötzliche und extreme Anstiege der Feinstaubwerte, sobald sich ihr für die Messung vorgesehener Bus in Bewegung setzt. Aus Ermangelung von Ursachenforschung in verwandten Studien schließen die Autoren die Verwendung der Daten, die während der Fahrt des Messfahrzeuges erfasst werden, für ihre Studie aus. Neben der Bewegungsgeschwindigkeit identifiziert Röger (2021) darüber hinaus noch weitere Faktoren, die Einfluss auf die Datenqualität bewegter Feinstaubsensoren nehmen. Die Bewegungsgeschwindigkeit nimmt vorrangig Einfluss auf die Qualität mobil erfasster Feinstaubdaten, da sie für einen instabilen Luftfluss in der Umgebung eines Sensors sorgt (Gao et al. 2016). Konsequenterweise spielt dabei auch die Ausrichtung und Positionierung eines Messgerätes eine Rolle, denn der Luftfluss an einer bestimmten Stelle eines Fahrzeuges wird von den aerodynamischen Eigenschaften der Objekte in der direkten Umgebung des Messgerätes mitbestimmt (Sharma et al. 2013). Nach Röger (2021) ist die Positionierungsgenauigkeit ein weiterer Einflussfaktor auf die Qualität von mobil erfassten Feinstaubdaten. Dieser Gesichtspunkt ist jedoch schon weitestgehend im Zusammenhang von mobilen Verkehrsdaten, wie zum Beispiel Floating Car Daten, erforscht (Cerqueira et al. 2018). Letztendlich spielt auch die Präsenz von Störquellen in der direkten Umgebung eines Feinstaubsensors eine Rolle. Während bei statischen Messungen der Sensor so platziert werden kann, dass er sich außerhalb der Reichweite von Feinstaubemittlern befindet, sind mobile Sensoren stark von dem Umgebungsverkehr während einer Messung beeinflusst (Röger 2021).

Während Röger (2021) einen Überblick über bereits bekannte und aus anderen Fachbereichen hergeleitete Einflussfaktoren auf die Datenqualität mobil erfasster Feinstaubsensoren gibt, mangelt es immer noch an Methoden zur Identifizierung und Korrektur dieser Faktoren. Die Bewegungsgeschwindigkeit ist ein gutes Beispiel dafür. Der Einfluss dieses Parameters wird zwar in einigen Studien wie Goyal et al. (2018) erwähnt, Arbeiten zur Untersuchung dieses Phänomens sind jedoch schwer auffindbar. Ähnlich gestaltet sich dieses Problem beim Faktor des Umgebungsverkehrs. Ein Einfluss wird zwar in Studien wie Röger et al. (2020a) indiziert, allerdings finden sich in der Literatur nur wenige damit verwandte Studien. Darüber hinaus fehlt es an der Diskussion, inwiefern sich der Umgebungsverkehr als ein Störfaktor oder als ein dazugehöriger Teil der natürlichen Feinstaubkonzentration darstellt.

1.3 Fragestellungen und Ziele

Das übergeordnete Ziel der Arbeit ist, mobil erfasste Feinstaubsensordaten hinsichtlich ihrer Einflussfaktoren auf die Datenqualität zu untersuchen. Aus den Erkenntnissen daraus wird beurteilt, ob gängige Standards und Richtlinien zur Beschreibung der Qualität von Geodaten auch für mobil erfasste Feinstaubsensordaten ausreichen, oder ob dafür die Definition weiterer Kategorien notwendig ist. Die Hypothese der Arbeit lautet daher:

Qualitätsparameter mobil erfasster Feinstaubsensordaten gehen über die Standards zur Beschreibung der Qualität von Geodaten hinaus.

Die Erfassung und Analyse mehrerer Datensätze soll dabei helfen, die Hypothese zu überprüfen. Dabei werden ebenfalls eine Reihe an Fragestellungen bearbeitet:

Mit welchen Standards und Richtlinien kann die Qualität von Geodaten beschrieben werden?

Die Beschreibung der Qualität von Geodaten ist von hoher Bedeutung, denn die Qualität der verwendeten Geodaten entscheidet auch über die Aussagekraft einer Analyse. Wenn aus einer Datenanalyse heraus Schlüsse gezogen werden, sollten dem Anwender durch den Datenbestand existierende Limitationen bewusst sein. Deshalb werden im Rahmen der Arbeit gängige

Standards und Richtlinien zur Beschreibung der Qualität von Geodaten dokumentiert.

Welche Einflussgrößen wirken auf die Qualität mobil erfasster Feinstaub-sensordaten?

Einflussfaktoren auf die Datenqualität statisch erfasster Sensordaten sind weitestgehend bekannt. Konsequenterweise stellt sich die Frage, welche weiteren Faktoren eine Rolle spielen, wenn sich Sensoren im Raum bewegen. Dabei werden sowohl sensorspezifische als auch externe Einflüsse wie die Luftfeuchtigkeit untersucht.

Welche Einflussgrößen sind statisch, und welche Einflussgrößen ändern sich dynamisch während einer Messung?

Die detektierten Einflussfaktoren auf die Qualität von mobil erfassten Sensordaten werden auf ihre statischen und dynamischen Dimensionen hin überprüft. Außerdem wird untersucht, welche der Einflüsse mithilfe von Sekundärdaten (wie zum Beispiel der Luftfeuchtigkeit) näherungsweise erklärbar sind und welche der Faktoren unvorhersehbar sind.

Welche der Einflussgrößen können mithilfe von Datenkorrektur und Kalibrierung eliminiert werden?

Die Kalibrierung kostengünstiger Feinstaubsensoren mithilfe von hochpräzisen Referenzgeräten ist bereits gängige Praxis in der Analyse von Sensordaten. Dabei stellt sich die Frage, welche der detektierten Einflussgrößen auf die Qualität von mobil erfassten Daten ebenfalls kalkulatorisch minimiert werden können.

Ist der Umgebungsverkehr während einer mobilen Erfassung von Feinstaubdaten ein Störfaktor oder ein Teil der natürlichen Feinstaubkonzentration?

Fahrzeuge im Straßenverkehr gelten als einer der primären Emittenten von Feinstaub. Werden Fahrzeuge allerdings als Träger einer Messstation eingesetzt, befinden sich diese direkt an der Quelle der Feinstaubemission. Gemessene Werte können somit kurzzeitig und kleinräumig deutlich höher er-

scheinen, als sie in Wahrheit sind. Konsequenterweise stellt sich die Frage, ob der Umgebungsverkehr eine Störquelle darstellt, oder ob dieser trotz des gegebenen Einflusses auf die gemessenen Werte ein Teil der natürlichen Feinstaubkonzentration ist.

1.4 Aufbau der Arbeit

Die Monographie ist in mehrere Kapitel gegliedert. Ein Flussdiagramm zur Erörterung des Ablaufes der Arbeit wird in Abbildung 2 gezeigt.

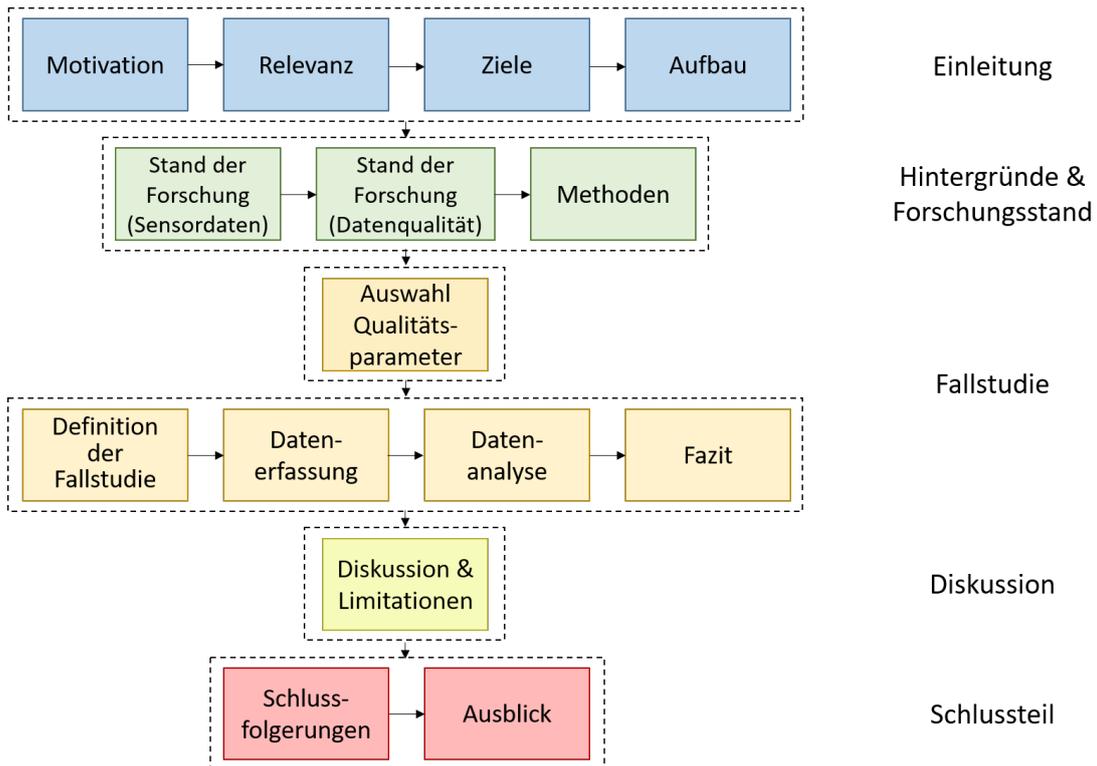


Abbildung 2: Aufbau der Arbeit, dargestellt als Flussdiagramm

Abbildung 2 zeigt den Aufbau der Arbeit. Die Subkapitel der Einleitung sind in hellblau dargestellt. Dazu gehören die Motivation des Autors und eine Abhandlung zur Relevanz des Themas. Außerdem werden die Hypothese und die Forschungsfragen sowie die Ziele der Arbeit genannt. Am Ende der Einleitung wird ein Überblick über den Aufbau der Arbeit gegeben. Die grünen Kästen zeigen die Subkapitel zu Hintergründen, Daten und Methoden der Arbeit. Außerdem wird Einblick über den aktuellen Stand der Forschung gegeben. Im Anschluss daran folgt eine experimentelle Analyse zur Überprüfung der Einflussparameter auf die Qualität von mobil erfassten Feinstaubdaten, bei der zunächst in einem separaten Kapitel die in der Arbeit untersuchten Qualitätsparameter definiert werden. Daraufhin wird der Ablauf der

Fallstudie beschrieben. Es folgen Informationen über das Studiengebiet und die Auswahl der dazu verwendeten Sensoren. Danach wird die Erfassung von drei Datensätzen zur Überprüfung der Qualitätsparameter beschrieben. Die Daten werden mithilfe mehrerer Methoden analysiert. Am Ende der Fallstudie steht ein Fazit, wo die Erkenntnisse bezüglich Faktoren der Datenqualität zusammengefasst werden. Im darauf folgenden Kapitel werden diese Erkenntnisse diskutiert. Dabei werden die Ergebnisse der Fallstudie im Zusammenhang mit der gestellten Hypothese und den Forschungsfragen kritisch betrachtet. Rot dargestellt sind die Subkapitel des Schlussteils. Zum letzten Teil der Arbeit gehören die Schlussfolgerungen, wobei noch einmal auf Hypothese und Forschungsfragen eingegangen wird. Zum Abschluss folgt ein Ausblick, der einen Überblick über zukünftige Arbeiten gibt.

2 Statisch und mobil erfasste Sensordaten

Sensoren sind Geräte, die auf Begebenheiten in ihrer physischen Umgebung reagieren. Sie sind dazu in der Lage, quantifizierbare Werte auszugeben, die weiter verarbeitet werden können. Sensoren können unterschiedliche Parameter erfassen. Luftschadstoffe und Verkehrsparameter gehören zu den häufig mit Sensoren untersuchten Gebieten. Verkehrsdaten geben Aufschluss über Kennzahlen wie das Verkehrsvolumen, den Verkehrsfluss, Längen von Staus und Geschwindigkeiten von individuellen Fahrzeugen. Zu den Luftschadstoffen zählen unter anderem die Feinstaubkonzentration, Kohlenstoffdioxid (CO_2) und Ozon. Sensoren können auf unterschiedliche Arten betrieben werden. Diese unterscheiden sich im Wesentlichen darin, ob die verwendeten Sensoren statisch installiert sind oder Daten aufzeichnen, während sie sich durch den Raum bewegen.

Sensordaten können für unterschiedliche Anwendungsgebiete genutzt werden. Intelligente Transportsysteme (ITSs) stellen eine Auswahl an Analysemöglichkeiten bereit. Sie gelten als wichtiger Bestandteil moderner Verkehrsplanung und -überwachung. Die steigende Anzahl an Fahrzeugen innerhalb und außerhalb von Städten führt mancherorts zu Überbelastungen der Verkehrswege, was zu einer größeren Anzahl und Häufigkeit von Staus führen kann. Eine weitere Konsequenz daraus ist die Schwierigkeit der Reaktion auf Not- sowie Unfälle. Sind Verkehrswege überlastet, ist die Anfahrt für Rettungsfahrzeuge nur erschwert zu bewerkstelligen. Darüber hinaus bedeutet ineffiziente Verkehrsplanung und -führung für Teilnehmer des Verkehrs (Fahrzeuge sowie Fußgänger und Radfahrer) ein erhöhtes Unfallrisiko. Effekte auf die Umwelt spiegeln sich in erhöhten Schadstoffwerten auf und neben Verkehrswegen wider (Dimitrakopoulos et al. 2010).

ITSs verstehen sich als Kombination von Verkehrsplanung und -management mit Verkehrsdaten, die bestenfalls in Echtzeit bereitgestellt werden. Auf Basis der Datenlage kann somit das Verkehrsgeschehen gesteuert werden, was Verkehrsabläufe sicherer und effizienter macht. Zu den mit ITSs verwandten Anwendungsgebieten gehören:

- die Verkehrseinschätzung und das Verkehrsmanagement,
- das Management der Fahrzeug- und Verkehrssicherheit,
- Fahrermodellierungstechniken,
- das Notfallmanagement,

- und die Reduzierung der Verkehrsemissionen.

Die Verkehrseinschätzung basiert primär auf Sensornetzwerken, die Daten über aktuelle Verkehrszustände erfassen. Das Verkehrsmanagement wird durch den Einsatz von Verkehrszentralen bewerkstelligt. Die Fahrzeug- und Verkehrssicherheit basiert auf der Analyse von individuellen Fahrstilen unterschiedlicher KFZ-Fahrer. Fahrermodellierungstechniken beschäftigen sich mit kognitiven Prozessen von Fahrern in (semi-)autonomen Fahrzeugen. Im Notfallmanagement werden zum einen Staus, ausgelöst durch Unfälle, analysiert. Zum anderen wird sich mit der Lösung solcher Notfallsituationen beschäftigt. Ein weiteres Ziel von ITSs ist die Reduzierung der durch den Verkehr auftretenden Emissionen wie von Feinstaub oder Kohlenstoffdioxid (Dimitrakopoulos et al. 2010).

Die Implementierung von ITSs ist an diverse Anforderungen geknüpft, die von Dimitrakopoulos et al. (2010) diskutiert werden:

- *Wahrnehmung multikontextualer Situationen*: Akteure innerhalb eines ITS müssen sich auf unterschiedlich komplexe Situationen adaptieren können.
- *Personalisierung*: Ein ITS soll nicht nur den Verkehr im Gesamtkontext betrachten, sondern auch Lösungen für Einzelakteure, zum Beispiel beim Routing von Rettungsfahrzeugen, bereitstellen.
- *Vernetzte Systeme*: Akteure und Sensoren innerhalb eines ITS müssen miteinander über verschiedene Kommunikationsstandards vernetzt sein.
- *Uneingeschränkter Zugriff auf Netzwerkelemente*: Die Kommunikation von Akteuren und Sensoren innerhalb eines ITS sollte lückenlos funktionieren.
- *Skalierbarkeit*: ITSs sollten sich an die aktuelle Verkehrssituation anpassen.

Die Ermittlung von Verkehrsstatus ist die Basis vieler Anwendungen von ITSs. Diese wird in Altintasi et al. (2017) in drei Kategorien unterteilt:

- Die Kalkulation von Durchschnittsgeschwindigkeiten,
- die Detektion von Engpässen in einem Straßennetz,
- und die Kalkulation von Verkehrsflussparametern.

Die Ermittlung von Verkehrsstatus auf Basis von Durchschnittsgeschwindigkeiten basiert auf Fahrzeiten auf bestimmten Straßensegmenten. Daraus können in einem nächsten Schritt kritische beziehungsweise stauanfällige Punkte innerhalb eines Straßennetzes detektiert werden. Verkehrsdichte, Verkehrsfluss sowie Geschwindigkeit stellen Verkehrsflussparameter dar (Altintasi et al. 2017).

In der Literatur finden sich einige Beispiele für die Ermittlung von Verkehrsstatus auf Basis von Verkehrsdaten. Kerner et al. (2005) präsentieren in ihrer Arbeit eine auf Durchschnittsgeschwindigkeiten basierende Methode zur Ermittlung von Verkehrsstatus. Dieser wird hier durch Fahrzeiten von Fahrzeugen auf bestimmten Streckenabschnitten beschrieben. Dabei soll die zu übermittelnde Datenmenge möglichst gering gehalten werden, um unnötigen Datenverkehr zu vermeiden. Des Weiteren reicht bei Benutzung dieser Methode bereits die Ausstattung eines Bruchteiles der sich im Verkehr befindlichen Fahrzeuge mit FCD-Sensoren aus, um Verkehrsstatus eines Straßennetzes zu bestimmen. Die am FCD-System teilnehmenden Fahrzeuge ermitteln ihre Fahrzeit $R_k^{(v)}$ auf bestimmten Streckenabschnitten. Sobald ein Fahrzeug einen ausgewählten Streckenabschnitt passiert hat, vergleicht es die Fahrzeit mit einer von einer Zentrale vorgegebenen Referenzzeit $R_k^{(c)}$. Liegt diese außerhalb einer definierten Toleranz $\Delta R_k^{(c)}$, wird sie an die Zentrale weitergegeben. Sollte sie innerhalb der Toleranz liegen, wird sie nicht an die Zentrale gesendet. Wenn die Zentrale eine neue Fahrzeit für einen bestimmten Streckenabschnitt zugesandt bekommt, wird die Referenzzeit $R_k^{(c)}$ des jeweiligen Abschnittes daran angepasst und an alle Fahrzeuge im FCD-System weitergegeben. Kerner et al. (2005) berechnen, dass bereits bei einer FCD-Ausstattung von 1.5% aller sich im Gesamtverkehr befindlichen Fahrzeuge 65% aller Änderungen im Verkehrsaufkommen auf bestimmten Streckenabschnitten vorhersagen lassen.

Zhang et al. (2018) präsentieren einen Ansatz zur Detektion von Engpässen innerhalb von Straßennetzen mithilfe von Floating Car Daten. Dafür verwenden sie einen Floating Car Datensatz von Peking, China. Als Engpass wird der Abschnitt vom Beginn eines Staus bis zum Ende eines Staus bezeichnet. Zhang et al. (2018) observieren Geschwindigkeitsunterschiede, um die Anfangspunkte von Engpässen zu kalkulieren. Aus den Geschwindigkeitsdifferenzen werden zunächst die Zeitfenster zwischen Start und Ende der Ereignisse berechnet. Anschließend werden die Durchschnittsgeschwindigkeiten der berechneten Zeitfenster mit den regulären Durchschnittsgeschwindigkeiten der Straßenabschnitte verglichen, um sicherzustellen, dass es sich bei den detektierten

Ereignissen tatsächlich um den Beginn von Engpässen handelt. Die folgenden Kapitel beschäftigen sich mit unterschiedlichen Sensordaten sowie den Messgeräten, mit denen sie erfasst werden können.

2.1 Statische Ansätze zur Erfassung von Verkehrsdaten

Ist ein Sensor zur Verkehrsdatengenerierung an einem Punkt fixiert, spricht man von einem statischen Datenerfassungsansatz. Die Position des Sensors ändert sich während einer Messung nicht. Höhere Auflösungen innerhalb einer Messkampagne lassen sich durch den Einsatz mehrerer Sensoren bewerkstelligen. Zu den Vertretern statischer Ansätze zur Verkehrsdatenerfassung gehören induktive Zählschleifen, die Radardetektion, bildgebende Verfahren wie die Nutzung von Videokameras sowie die Nutzung von Signaldetektoren (Bluetooth, WiFi).

2.1.1 Induktive Zählschleifen (ILD)

Induktive Zählschleifen werden unter dem Straßenbelag an definierten Orten innerhalb von Straßennetzen installiert. Dabei werden Kabel verlegt, die mit sinusförmigem Wechselstrom versorgt werden. Die Breite einer induktiven Zählschleife entspricht typischerweise der Breite der Straße. Die Längen der Aufbauten können variieren und befinden sich im Bereich von bis zu 20 Metern. Der Wechselstrom hat eine Frequenz von wenigen bis über 100 Kilohertz. Der Strom erzeugt ein Magnetfeld in der Umgebung der induktiven Schleife. Wenn ein metallisches Objekt, beispielsweise ein Fahrzeug, die Zählschleife passiert, treten zwei Effekte auf. Zum einen verhalten sich magnetische Elemente im Fahrzeug (beispielsweise die Felgen der Reifen) beim Überfahren einer Zählschleife wie ein magnetischer Kern. Dadurch ändert sich kurzzeitig die Induktivität der Zählschleife. Zum anderen treten durch Wechselwirkungen des magnetischen Feldes der Schleife und den Fahrzeugmaterialien Wirbelströme auf. Diese Wirbelströme sind messbar über die Induktivität und den Widerstandswert innerhalb der Zählschleife (Gajda et al. 2012).

Es existieren unterschiedliche Ansätze zur Datenauswertung induktiver Zählschleifen. Ein Boole'scher Ansatz wandelt die Änderungen innerhalb des Magnetfeldes in die Werte 0 (keine Änderung des Magnetfeldes) und 1 (Änderung des Magnetfeldes) um. Bei einem Wert von 1 wird davon ausgegangen, dass ein Fahrzeug die induktive Zählschleife passiert hat. Komplexere Ansätze zur Datenauswertung definieren die individuelle Wellenform des Indukti-

vitätswertes einer Zählschleife als Fahrzeugsignatur (Oh et al. 2002).

Jedes Fahrzeug hinterlässt eine unterschiedliche Signatur. Fahrzeugsignaturen haben unterschiedliche Charakteristika, auf Basis derer weitere Informationen als nur die Präsenz und Absenz eines Fahrzeuges beim Passieren einer Zählschleife abgeleitet werden können. Diese Charakteristika umfassen:

- die maximale Magnitude a ,
- die magnetische Länge $d + e$,
- die Wellenform $d/(d + e)$,
- die Signatur f ,
- die Anstiegsrate c ,
- sowie die Dauer b .

Die Anstiegsrate c und die Dauer b hängen dabei vom Gesamtverkehr ab, während die Magnitude a , die magnetische Länge $d + e$, die Wellenform $d/(d + e)$ und die Signatur f als fahrzeugspezifische Parameter gelten. Deren Charakteristika werden beispielsweise von der Fahrzeuglänge beeinflusst. Als Einflussgrößen des Gesamtverkehrs gelten die Fahrtgeschwindigkeit der Fahrzeuge, der Verkehrsfluss und das Verkehrsvolumen (Oh et al. 2002).

Aus der Fahrzeugsignatur einer induktiven Schleife kann nicht nur bestimmt werden, ob ein Fahrzeug die Zählstation passiert. Auch die Geschwindigkeit ist ein Faktor, der mithilfe von induktiven Schleifen bestimmt werden kann. Möglich machen das entweder die Installation zweier Schleifen nacheinander, die bei Doppeldetektion eines Fahrzeuges die Geschwindigkeit mit dem Zeitunterschied der Messungen triangulieren, sondern auch die Analyse von Fahrzeugsignaturen einzelner Zählschleifen. Dabei korreliert die Dauer b negativ mit der Fahrtgeschwindigkeit, während die Anstiegsrate c proportional mit der Geschwindigkeit in Relation steht. Auch eine Fahrzeugklassifikation ist möglich durch Analyse der Signatur. Mithilfe der fahrzeugspezifischen Parameter Magnitude a , magnetische Länge $d + e$, Wellenform $d/(d + e)$ und Signatur f können Fahrzeuge in unterschiedliche Klassen wie Automobile, Trucks und Busse aufgeteilt werden (Oh et al. 2002).

2.1.2 Radarmessungen

Radarmessgeräte sind beliebte Sensoren in der Verkehrsüberwachung. Sie können sowohl zur Verkehrszählung als auch zur Geschwindigkeitsüberwachung eingesetzt werden. Radarsensoren werden in der Regel neben oder über Straßen platziert. Sie emittieren Lichtsignale, deren Reflexionen Aufschluss über sich auf der Straße bewegende Objekte geben. In der Verkehrsüberwachung verwendete Radarsysteme operieren in der Regel im X-Band (mit 10 Gigahertz Frequenz) oder im K-Band (24 Gigahertz). Unterschiedliche Radar Systeme wie Pulsdoppler-Radars (PD), unmodelierte Dauerstrich-Radars und modulierte Dauerstrich-Radars (FMCW) kommen zum Einsatz (Fang et al. 2007). Ein Beispiel für einen häufig eingesetzten Radar Sensor zur Verkehrsüberwachung ist der auf FMCW basierende *Road Traffic Microwave Sensor* (RTMS), der das Verkehrsvolumen, die Fahrzeugpräsenz, die Fahrspurbelegung, die Fahrzeuggeschwindigkeit und Staulängen bestimmen kann (Wang et al. 1992).

Verschiedene Verkehrssituationen produzieren unterschiedliche Signaturen in den Daten. Diese Signaturen entstehen, wenn einzelne detektierte Punkte eines Objektes zu einer Polylinie verbunden werden (Garcia et al. 2009).

Signaturen lassen Rückschlüsse über die aufgenommenen Objekte zu. Dazu gehören:

- die L-förmige Signatur,
- kleine Objekte,
- fixierte Objekte,
- Straßenbegrenzungen,
- sowie bewegliche Objekte.

Dabei haben geparkte Fahrzeuge in der Regel eine L-förmige Signatur, wenn die reflektierten Punkte zu einer Linie verbunden werden. Fußgänger stellen sich in der Signatur als besonders kleine Objekte dar. In die Gruppe der fixierten Objekte fallen alle Detektionen, die sich nicht in eine der anderen Kategorien zuordnen lassen. Dazu gehören beispielsweise Straßenlaternen oder Verkehrsschilder. Straßenbegrenzungen verlaufen meist parallel zu Bewegungsvektoren von Fahrzeugen, Radfahrern oder Fußgängern. Zur Gruppe der beweglichen Verkehrsobjekte gehören Fahrzeuge (Garcia et al. 2009).

2.1.3 Verkehrsüberwachungskameras

Verkehrsüberwachungskameras sind an definierten Punkten innerhalb eines Straßennetzes installiert. Im Gegensatz zu induktiven Schleifen und Radarlösungen gilt die kameragestützte Verkehrsüberwachung nicht als *blind* - sie kann komplexe Zusammenhänge innerhalb des Verkehrs darstellen. Dazu gehört die Verkehrssicherheit sowie die Überwachung der Qualität des Straßenbelages (Li et al. 2013). Der Verkehrsfluss, die Geschwindigkeit passierender Fahrzeuge, die Länge von Staus und die Verkehrsdichte gehören zu Parametern, die ebenfalls von Videokameras überwacht werden können. Neben der Produktion von Verkehrsstatistiken können von Videokameras produzierte Daten unter anderem dazu verwendet werden, digitale Verkehrsschilder mit variablen Geschwindigkeitsbegrenzungen zu implementieren. Dazu sind zunächst Prozessierungsschritte notwendig. Die Aufbereitung von Videodaten sieht vor, Beleuchtungsunterschiede im Video auszugleichen, Schatten von Fahrzeugen zu entfernen sowie witterungsbedingte Einflüsse auf das Bild zu eliminieren. Darauf folgend werden die Videodaten analysiert. Algorithmen des maschinellen Lernens wie neuronale Netzwerke, Entscheidungsbäume und Kaskadenklassifikatoren können dabei helfen, Fahrzeuge sowie Objekte wie Begrenzungsstreifen zu detektieren. Durch die Bewegung eines detektierten Fahrzeuges durch ein Bild lassen sich weitere Parameter wie die Geschwindigkeit eines Objektes feststellen (Alpatov et al. 2018).

2.1.4 Bluetooth und WiFi

Viele Verkehrsteilnehmer tragen Geräte mit sich, die digitale Signale emittieren. Dazu gehören Mobiltelefone und ähnliche Geräte wie Tablet-Computer. Darüber hinaus verfügen einige Fahrzeuge bereits über Module, die in der Lage sind, die Fahrzeugelektronik drahtlos mit externen Geräten wie Mobiltelefonen zu verbinden. Drahtlose Verbindungen werden dabei häufig mit Übertragungsstandards wie Bluetooth oder WiFi hergestellt. Zur Datenübertragung verwendete Geräte haben eine eindeutige MAC-Adresse, über die eine Verbindung hergestellt werden kann. Durch die Dokumentation sämtlicher detektierter MAC-Adressen können Rückschlüsse auf den Verkehr in der Umgebung eines Signalscanners gezogen werden. (Friesen et al. 2015). Die Reichweite der Geräte ist stark abhängig vom verwendeten Übertragungsstandard. Bluetooth-Signale mobiler Geräte haben in der Regel eine Reichweite von bis zu zehn Metern. Der WiFi-Übertragungsstandard hat mit bis zu

35 Metern innerhalb von Gebäuden und bis zu 100 Metern außerhalb von Gebäuden eine wesentlich höhere Reichweite. Auch bezüglich der Verbindungszeit, die benötigt wird um eine Adresse abzugreifen, unterscheidet sich die Bluetooth-Technik maßgeblich vom WiFi-Standard. Während Bluetooth-Geräte dafür eine Verbindungszeit von 10.21 Sekunden benötigen, können WiFi Scanner Adressen fremder Geräte innerhalb von einer Sekunde bestimmen (Abedi et al. 2013).

Im Feld der intelligenten Transportsysteme können Scanner genutzt werden, um diese individuellen Adressen zu sammeln. Das lässt Rückschlüsse über das derzeitige Verkehrsvolumen zu. Darüber hinaus lassen sich Parameter wie der Verkehrsfluss und die Verkehrsdichte abschätzen (Friesen et al. 2015). Über die Installation von mehreren vernetzten Signalscannern sind auch Reisezeiten (Erkan et al. 2016), mittlere Fahrtgeschwindigkeiten (Puckett et al. 2010) sowie der Ausgangs- und Bestimmungsort (Blogg et al. 2010) einer Fahrt berechenbar.

2.2 Mobile Erfassung von Verkehrsdaten

Im Gegensatz zu statischen Ansätzen, sind die verwendeten Sensoren bei mobilen Lösungen zur Verkehrsdatenerfassung in Bewegung. Dadurch ist eine größere Abdeckung gegenüber statischen Ansätzen erreichbar, da ein Sensor mehrere Messpunkte bedient. Zur Datengenerierung werden unterschiedliche Sensoren verwendet. Diese können beispielsweise Videokameras oder Laserscanner sein. Auch ein Fahrzeug selbst kann als Sensor fungieren. Floating Car Daten sowie erweiterte Floating Car Daten sind Beispiele für mobil erzeugte Verkehrsdaten. Floating Car Observer erfassen xFCD auf Basis von Observationen des Umgebungsverkehrs.

2.2.1 Floating Car Daten (FCD)

Floating Car Daten sind mobil erfasste Daten, die aus den Bewegungen von Fahrzeugen generiert werden. Hierbei werden in einem definierten Intervall Datentupel, die aus einer ID, der aktuellen Position und dem aktuellen Zeitstempel eines Fahrzeuges bestehen, gespeichert (De Fabritiis et al. 2008). Abbildung 3 zeigt den Prozess der Erfassung von Floating Car Daten schematisch.

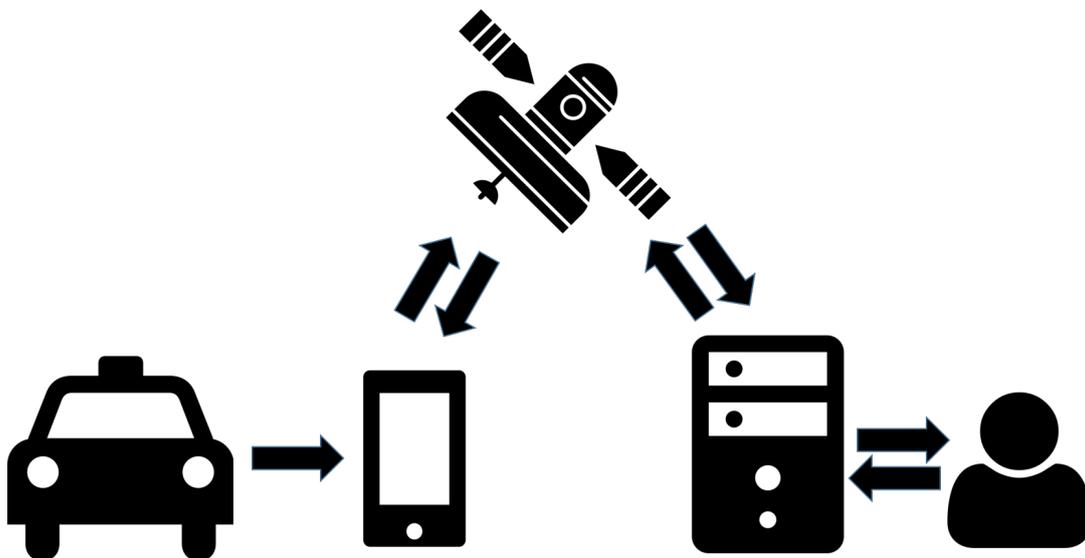


Abbildung 3: Schematische Darstellung der Erfassung von Floating Car Daten

Abbildung 3 zeigt schematisch, wie Floating Car Daten erfasst werden. Hierbei fungieren Fahrzeuge als mobile Sensoren, die ihre ID, einen aktuellen Zeitstempel sowie ihre aktuelle Position regelmäßig über ein mobiles Endgerät auf einen zentralen Server übertragen. Hier können die Daten im Anschluss von Endnutzern observiert und analysiert werden.

Zur Erfassung von Floating Car Daten werden ein Positionierungssystem, ein Zeitnehmer sowie ein Übertragungs- oder Speichermedium benötigt. Raumzeitliche Daten lassen sich über GPS-Empfänger generieren. Diese können über die Pseudo-Laufzeiten von Signalen, die von Satelliten aus gesendet werden, ihren aktuellen Standort sowie die aktuelle Datumszeit bestimmen. Zur Übertragung von Daten bieten sich verschiedene Standards wie Mobilfunknetze, WiFi-Applikationen, Bluetooth oder *Long Range Wide Area Networks* (LoRa) an. Die Speicherung von Daten lässt sich über Elemente wie Festplatten, USB-Speichermedien oder Speicherkarten bewerkstelligen (De Fabritiis et al. 2008).

Viele der sich unter der Bevölkerung befindlichen mobilen Endgeräte beinhalten sowohl Positionierungs-, Übertragungs-, sowie Speicherfunktionen. Aus diesem Grund werden Mobilfunkgeräte oft für die Erfassung von Floating Car Daten verwendet. Applikationen wie *myTracks*⁴, *GPS-Tracker Pro*⁵ oder *GPS-Logger für Android*⁶ bieten eine benutzerfreundliche Plattform zum Aufzeichnen und Verschicken von GPS-Logs, die als Floating Car Daten fungieren

⁴<https://www.mytracks4mac.info/index.php/de/>, 03/2020

⁵<https://apps.apple.com/de/app/gps-tracker-pro/id984920064>, 03/2020

⁶<https://play.google.com/store/apps/details?id=com.mendhak.gpslogger&hl=de>, 03/2020

können. Abbildung 4 zeigt eine Fahrzeugtrajektorie, die auf Basis von Floating Car Daten aus der Applikation *myTracks* generiert wurde.

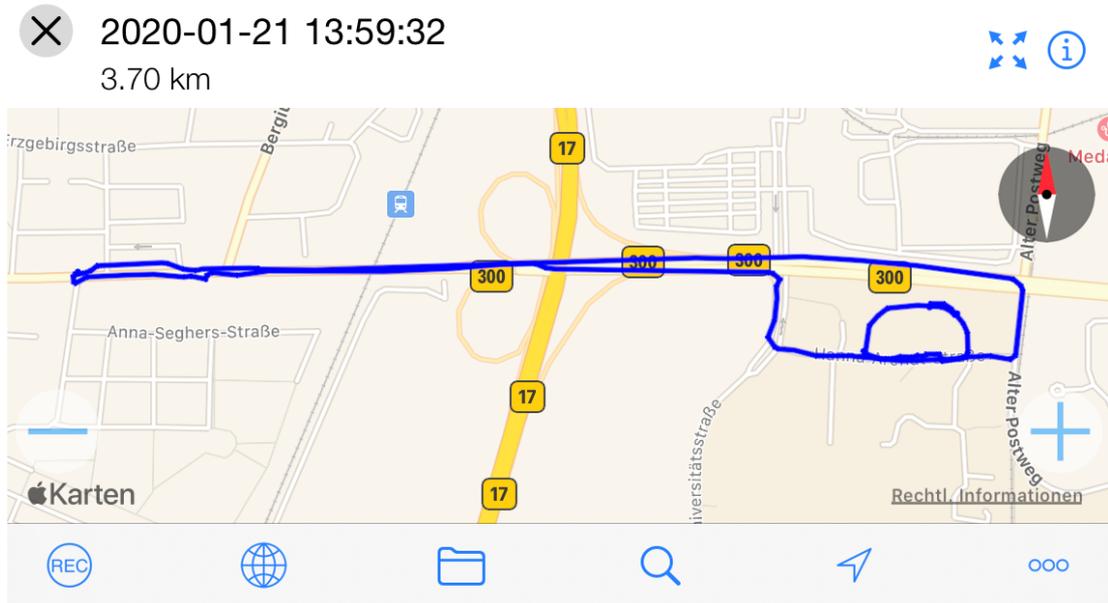


Abbildung 4: Darstellung einer mithilfe der *myTracks*- Applikation erstellten Fahrtrajektorie

Abbildung 4 zeigt ein Bildschirmfoto aus der Applikation *myTracks*, auf dem eine Fahrtrajektorie eines Fahrzeuges zu sehen ist. Die Trajektorie besteht aus mehreren aufgenommenen Punkten, die auch als Rohdaten einsehbar sind.

Geräte zum Erfassen von Floating Car Daten lassen sich auf unterschiedlichen Fahrzeugarten montieren. Zu den am häufigsten genutzten Fahrzeugarten gehören unter anderen der motorisierte Individualverkehr (MIV), Taxis, der öffentliche Personennahverkehr (ÖPNV; zum Beispiel Busse oder Trams) sowie der Lieferverkehr. Tabelle 1 vergleicht diese unterschiedlichen Fahrzeugarten hinsichtlich ihrer Vor- und Nachteile zur Erfassung von FCD anhand von verschiedenen Faktoren. Dazu gehören die Gesamtabdeckung des Straßennetzes, die produzierte Datenmenge sowie die Einfachheit der Installation von FCD Kollektoren.

Fahrzeugart	Abdeckung	Datenmenge	Einfachheit Installation
MIV	+	+	-
Taxis	+	+	+
ÖPNV	-	+	+
Lieferverkehr	+	+	o

Tabelle 1: (+)Vor- und (-)Nachteile des motorisierten Individualverkehrs (MIV), Taxiverkehrs, öffentlichen Personennahverkehrs (ÖPNV) und Lieferverkehrs für die Erfassung von Floating Car Daten. Ausgeglichene Eignung ist mit einem (o) gekennzeichnet.

Tabelle 1 vergleicht unterschiedliche Fahrzeugarten auf Basis von deren Eignung zur Nutzung als FCD-Kollektoren. Der motorisierte Individualverkehr deckt Straßennetze vollumfänglich ab. Die potentiell erfassbare Datenmenge ist durch die große Abdeckung und das große Verkehrsvolumen ebenfalls hoch. FCD-Kollektoren müssen jedoch in den meisten Fahrzeugen nachinstalliert werden, da nur wenige moderne Fahrzeuge mit Positionierungssensoren ausgestattet sind (Röger et al. 2018). Taxis erreichen ähnlich hohe Abdeckungen wie der MIV, da sie sich auf großen Teilen des Straßennetzes bewegen. Während den Tagesstunden sind durchschnittlich 30% aller Taxis gleichzeitig im Dienst (Lorkowski et al. 2003). Dadurch ist die potentiell erfassbare Datenmenge ebenfalls auf einem hohen Niveau. Viele Taxis sind darüber hinaus bereits mit Fahrtenschreibern ausgestattet, die aktuelle Bewegungen einer Flotte protokollieren (Brockfeld et al. 2007). Der öffentliche Personennahverkehr bewegt sich auf definierten Routen, die nur selten von ihren vorgesehenen Straßen abweichen. Außerdem bewegen sich die Fahrzeuge des ÖPNV teils auf eigens konzipierten Straßen, wodurch FCD-Messungen nicht die Verkehrssituation aller sich im Straßennetz befindlicher Verkehrsteilnehmer widerspiegeln. Ähnlich wie bei Taxis werden jedoch die Bewegungen von ÖPNV-Fahrzeugen bereits aufgezeichnet und verarbeitet (Gökaşar et al. 2017). Der Lieferverkehr deckt hingegen größere Teile des Straßennetzes ab, wodurch die Erfassung einer großen Datenmenge möglich ist. Einige der sich im Gesamtverkehr befindlichen Lieferfahrzeuge sind bereits mit Positionierungsinstrumenten ausgestattet (Ebendt et al. 2010).

Taxis werden in vielen Fällen als FCD-Kollektoren verwendet. Durch die hohe Abdeckungsrate und eine große verfügbare Datenmenge sind einige auf Taxis basierende Floating Car Datensätze verfügbar (Brockfeld et al. 2007).

Abbildung 5 zeigt den schematischen Ablauf einer auf Taxidaten basierenden FCD-Erfassung. Hierbei sind die Taxiflotten mit Positionierungssystemen ausgestattet. Über Funktechnologien wie Radiowellen oder Mobilfunknetze werden die aktuelle Position sowie die aktuelle Zeit an eine Taxizentrale übertragen. Dort werden die Daten dann zusammen mit einer eindeutig zuweisbaren Fahrzeug-ID an einen Server übertragen, der die Daten reinigt und filtert. Im Anschluss werden die Floating Car Daten in einer Datenbank gespeichert, damit Endnutzer zur Analyse darauf zugreifen können (Brockfeld et al. 2007). Floating Car Daten haben unterschiedliche Bewegungsparameter, die die Fahrt eines Fahrzeuges beschreiben. Dazu gehören der Pfad, die zurückgelegte Distanz, die Geschwindigkeit des Fahrzeuges sowie die aktuelle Fahrtrichtung (Ranacher et al. 2016). Diese Parameter sind für die später vorgestellten Ana-

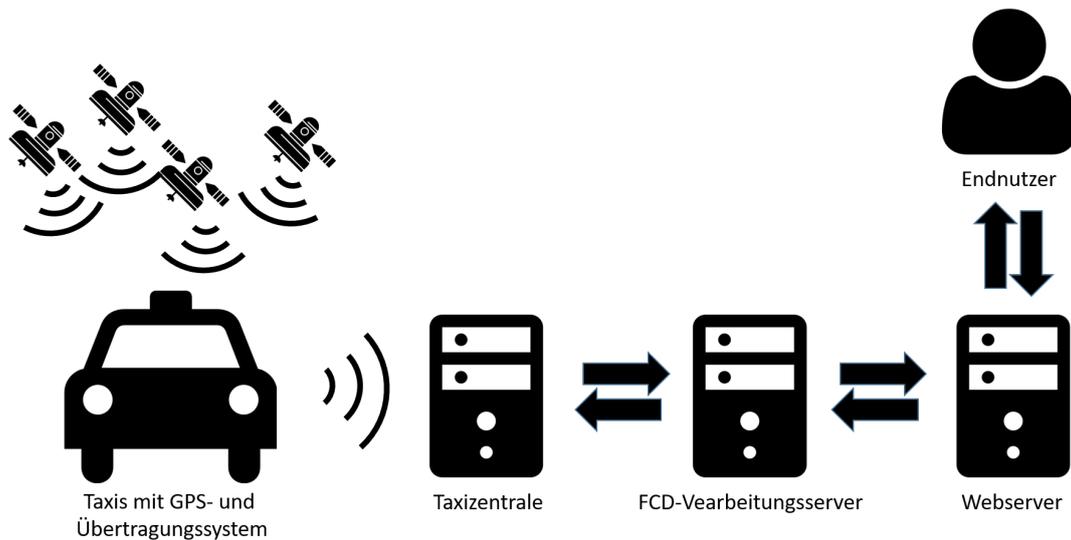


Abbildung 5: Schematische Darstellung eines auf Taxis basierenden FCD-Systems nach Brockfeld et al. (2007)

lysewerkzeuge relevant. Tabelle 2 zeigt die unterschiedlichen Bewegungsparameter von Floating Car Daten sowie deren Definition.

Bewegungsparameter	Variable	Definition
Pfad	II	$II = \langle P_1, \dots, P_n \rangle$
Distanz	d	$d = d(P_i, P_j)$
Geschwindigkeit	v	$v = \frac{d}{\Delta t}$
Richtung	θ	$\theta = \angle P_i P_j$

Tabelle 2: Bewegungsparameter von Floating Car Daten nach Ranacher et al. (2016)

Tabelle 2 zeigt die vier Bewegungsparameter von Floating Car Daten (Pfad, Distanz, Geschwindigkeit und Richtung). Des Weiteren werden deren Variablen sowie deren Definition dargestellt. Der Pfad II gibt an, auf welcher Trajektorie sich ein Fahrzeug befindet. Er wird als Vektor der sich auf der Trajektorie befindlichen Punkte P_1, \dots, P_n definiert. Die Distanz d gibt Aufschluss über den Weg zwischen zwei Punkten P_i und P_j . Die Geschwindigkeit v beschreibt das Tempo eines Fahrzeuges auf einer definierten Strecke zwischen zwei Punkten. Die Richtung θ gibt an, in welchem Winkel sich das jeweilige Fahrzeug bewegt (Ranacher et al. 2016).

Die Samplingrate gibt an, mit welcher Frequenz FCD Punkte aufgenommen werden. Diese wird in der SI-Einheit $Hertz(Hz) = \frac{1}{s}$ gemessen. Eine Samplingrate von einem Hertz entspricht hierbei einer Frequenz von einem Datenpunkt pro Sekunde. In der Literatur werden unterschiedliche Samplingraten diskutiert. Ranacher et al. (2016) definieren in ihrer Studie optimale Erfassungsfrequenzen. Die Berechnungen basieren auf der Eliminierung von Feh-

lerwerten bei der Berechnung der vier Bewegungsparameter Pfad, Distanz, Geschwindigkeit und Richtung. Tabelle 3 zeigt die resultierenden optimalen Samplingraten für die Erfassung von Floating Car Daten.

Bewegungsparameter	Variable	Samplingrate
Pfad	l	$\frac{1}{3}Hz - \frac{1}{5}Hz$
Distanz	d	$\frac{1}{5}Hz - \frac{1}{10}Hz$
Geschwindigkeit	v	$1Hz - \frac{1}{2}Hz$
Richtung	θ	$\frac{1}{3}Hz - \frac{1}{5}Hz$

Tabelle 3: Optimale Samplingraten für Floating Car Daten nach Ranacher et al. (2016)

Tabelle 3 zeigt optimale Samplingraten für die Erfassung von Floating Car Daten. Für die Berechnung von Pfaden empfehlen Ranacher et al. (2016) eine Frequenz von $\frac{1}{3}Hz$ bis $\frac{1}{5}Hz$. Für die Berechnung von Distanzen aus Floating Car Daten wird eine Samplingrate von $\frac{1}{5}Hz$ bis $\frac{1}{10}Hz$ empfohlen. Für Geschwindigkeitsanalysen resultiert eine Frequenz von $1Hz$ bis $\frac{1}{5}Hz$ in den niedrigsten Fehlerwerten. Die Fahrtrichtung von Fahrzeugen lässt sich am optimalsten mit einer Frequenz von $\frac{1}{3}Hz$ bis $\frac{1}{5}Hz$ berechnen (Ranacher et al. 2016).

2.2.2 Erweiterte Floating Car Daten (xFCD)

Werden Floating Car Daten mit weiteren Fahrdaten wie der aktuellen Geschwindigkeit oder dem momentanen Kraftstoffverbrauch ergänzt, spricht man von erweiterten Floating Car Daten. Erweiterte Daten müssen in digitaler Form vorliegen und können sowohl von fahrzeuginternen als auch von externen Sensoren stammen. Seit 1993 besitzen alle in Europa produzierten Fahrzeuge ein elektronisches System zum Auslesen von Fehlercodes oder Fahrzeugparametern wie der Kraftstoffeinspritzung. Mit der Einführung des *Euro III* Standards⁷ im Jahr 2001 (in den USA bereits 1996) haben sich die Fahrzeughersteller dazu verpflichtet, ein grundlegendes Set von Parametern über ein standardisiertes System bereitzustellen. Dieses System beinhaltet ein standardisiertes Kommunikationsprotokoll sowie ein weltweit einheitliches Steckersystem. Während der *SAEJ1979*-Standard⁸ ein Protokoll zum Verbinden und Auslesen eines Fahrzeugcomputers definiert, legt der *SAEJ1962*-Standard⁹ die Form und den Aufbau eines On-Board-Diagnostics (OBD)-Steckers fest (Messelodi et al. 2009; Alessandrini et al. 2012; Prasad et

⁷<https://www.eea.europa.eu/help/glossary/eea-glossary/euro-iii>, 06/21

⁸https://www.sae.org/standards/content/j1979_201702/, aufgerufen 05/21

⁹https://www.sae.org/standards/content/j1962_201607/, aufgerufen 05/21

al. 2019).

Erweiterte Floating Car Daten setzen sich aus drei Hauptbestandteilen zusammen. Dazu gehören:

- **B** - Basisinformationen,
- **V** - Fahrzeuginformationen,
- und **X** - erweiterte Informationen.

Zur Kategorie **B** gehören FCD-Informationen wie die Fahrzeug-ID, der Zeitstempel und Positionsdaten. Kategorie **V** besteht aus fahrzeuginternen Informationen wie dem aktuellen Gang, der Motorlast und der Kühleremperatur. Zu Kategorie **X** gehören Daten externer Sensoren (Messelodi et al. 2009). Dazu zählen unter anderem Messungen mit Feinstaubsensoren (Röger et al. 2020a), Kohlenstoffmonoxid-Sensoren (Hagler et al. 2010) und CO₂-Sensoren (Lee et al. 2017).

In Kategorie **V** sind einige Parameter definiert, die mithilfe eines OBD-Steckers in einem bestimmten Intervall ausgelesen werden können. Tabelle 4 zeigt eine Auswahl dieser Parameter, ihre Parameter IDs (PIDs) und ihre Einheit.

Fahrzeugparameter	PID	Einheit
Geschwindigkeit	0D	km/h
Kraftstoffdruck	0A	kPa
Ansauglufttemperatur	0F	°C
Barometrischer Druck	33	kPa
Position des Gaspedals	11	%
Lufttemperatur	46	°C
Luftflussrate	10	g/s

Tabelle 4: In *SAEJ1979* definierte auslesbare Fahrzeugparameter, ihre PIDs und ihre Einheit

Tabelle 4 zeigt eine Auswahl der im Standard *SAEJ1979* definierten Parameter mit den dazugehörigen PIDs (im hexadezimalen Format) und Einheiten. Einige Fahrzeughersteller definieren darüber hinaus noch weitere Modi, in denen zusätzliche Fahrzeugparameter ausgelesen werden können (Schäffer 2013).

2.2.3 Floating Car Observer (FCO)

Eine Schwachstelle jeglicher auf Floating Car Daten basierenden Technologien zur Verkehrsdatenanalyse ist, dass mit FCD-Technik ausgestattete Fahrzeuge in der Regel nur Daten über sich selbst aufzeichnen. Wenn ein Fahrzeug beispielsweise bremst, ist dem Betrachter der Datengrundlage häufig

unklar, ob die Verringerung der Geschwindigkeit auf Basis von auftretendem Verkehr vor dem analysierten Fahrzeug oder aufgrund des Willens des Fahrers geschieht. Floating Car Observer lösen dieses Problem, indem sie FCD-Datentupel um Verkehrsobservationen aus dem Umfeld des Fahrzeuges anreichern. Dafür werden externe Sensoren verwendet, die zusätzlich zur FCD-Technik am fahrenden Fahrzeug angebracht werden. Es gibt unterschiedliche Sensoren, die für die Observation des Umgebungsverkehrs eines Fahrzeuges nutzbar sind. Die unterschiedlichen Sensortypen werden in Tabelle 5 dargestellt.

Sensorart	Messtechnik	Blickfeld
Infrarotsensor	Abstandsmessung zum nächsten Objekt	
Laserscanner	Abstandsmessung zum nächsten Objekt	
Ultraschallsensor	Abstandsmessung zum nächsten Objekt	
Videokamera	Bildprozessierung (Detektion von Fahrzeugen)	
Signal-detektor	Empfang von Bluetooth & WIFI Signalen anderer Fahrzeuge	
Feinstaubsensor	Messung des Straßenabriebes von vorausfahrenden Fahrzeugen	

Tabelle 5: Sensoren für die Implementierung eines Floating Car Observers, ihre Funktionsweise und ihr Blickfeld

Tabelle 5 zeigt Sensoren für FCO-Anwendungen, ihre Eigenschaften hin-

sichtlich der Messtechnik und ihr Blickfeld auf den Umgebungsverkehr. Zu den Sensoren gehören Abstandssensoren (Infrarot, Ultraschall, Laserscanner), Videokameras, Signaldetektoren (Bluetooth, WiFi) sowie Feinstaubsensoren. Abbildung 6 zeigt schematisch, wie die Abstandsmessung mit Infrarotsensoren funktioniert.

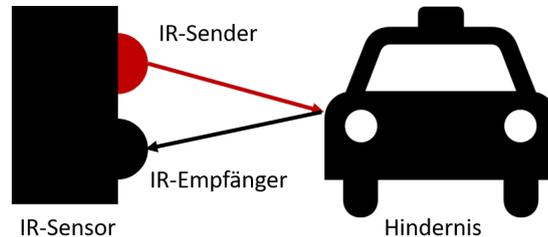


Abbildung 6: Schematische Darstellung der Abstandsmessung mit Infrarotsensoren in FCO-Anwendungen

Abbildung 6 zeigt die Funktionsweise eines Infrarotsensors zur Abstandsmessung. Infrarotsensoren nutzen eine Leuchtdiode, deren Lichtsignale an Hindernissen reflektiert werden und von einem Empfänger wieder aufgespürt werden. Durch die Eigenschaften des Infrarotlichts (zum Beispiel die Wellenlänge), die sich durch die Reflexion geändert haben, kann der Abstand zum jeweiligen Hindernis abgeschätzt werden. Die Reichweite eines herkömmlichen Infrarotsensors geht von wenigen Zentimetern bis zu 20 Metern. Die Anschaffungskosten eines Infrarotsensors bewegen sich im niedrigen Bereich und stellen somit eine kostengünstige Lösung zur Implementierung eines FCO dar. Andererseits ist die Detektion von Objekten bestimmter Materialien wie zum Beispiel Glas mit Infrarotsensoren nicht möglich (Wolf et al. 2008a).

Laserscanner haben eine ähnliche Funktionsweise wie Infrarotsensoren. Allgemein unterscheiden sie sich in der Art des emittierten Signales sowie dessen Strahlungsrichtung. Abbildung 7 zeigt deren Funktion schematisch.

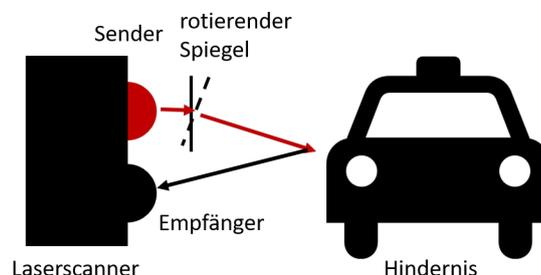


Abbildung 7: Schematische Darstellung der Abstandsmessung mit Laserscannern in FCO-Anwendungen

Abbildung 7 zeigt die Abstandsmessung mit Laserscannern schematisch. Ein emittierter Laserstrahl trifft auf einen rotierenden Spiegel, der den Strahl mehrmals pro Sekunde in unterschiedliche Richtungen abwirft. Der Strahl wird an einem Hindernis reflektiert und dann vom Empfänger wieder erkannt. Dieser kann durch die Änderung der Wellenlänge und der Intensität des ausgesendeten Lasersignals den Abstand zu einem Objekt bestimmen. Typische Reichweiten von Laserscannern gehen von wenigen Zentimetern bis zu mehreren hundert Metern. Da die Anschaffungskosten oft in einem hohen Bereich liegen, ist ein Laserscanner nur bedingt zur nachträglichen Aufrüstung eines Fahrzeuges zum Floating Car Observer geeignet. Laserscanner sind jedoch hochpräzise Werkzeuge, die zum Teil schon in modernen Fahrzeugen im Zusammenhang mit dem autonomen Fahren vorhanden sind (Wender et al. 2008; Wolf et al. 2008a).

Super- sowie Ultraschallsensoren nutzen die Ausbreitung von Schallwellen zur Abstandsmessung. Abbildung 8 zeigt die Funktionsweise dieser Sensoren schematisch.

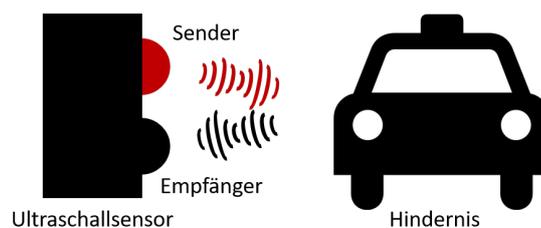


Abbildung 8: Schematische Darstellung der Abstandsmessung mit Ultraschallsensoren in FCO-Anwendungen

Abbildung 8 zeigt das Messprinzip von Super- beziehungsweise Ultraschallsensoren. Ein Sender emittiert Schallwellen, die an einem Hindernis zurückgeworfen werden und von einem Empfänger wieder erkannt werden. Über die Übertragungszeit des Signals und die Geschwindigkeit des emittierten Schalls kann der Sensor berechnen, wie weit das Hindernis entfernt ist. Während die Anschaffungskosten von Super- beziehungsweise Ultraschallsensoren als gering gelten, umfasst deren maximale Reichweite nur mehrere Meter (Wolf et al. 2008a; Weisheit et al. 2011).

Videokameras sind dazu in der Lage, Umgebungsverkehr in einem großen Winkel zu erfassen. Abbildung 9 zeigt deren Prinzip.

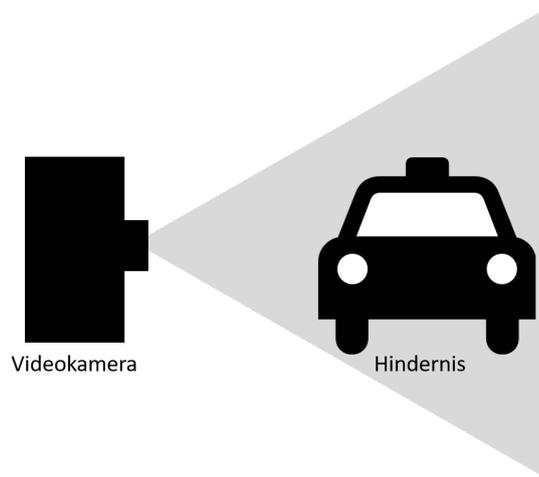


Abbildung 9: Schematische Darstellung der mobilen Verkehrszählung mit Videokameras in FCO-Anwendungen

Abbildung 9 zeigt die Verkehrsdatenaufnahme mit Videokameras. Kameras sind in einigen modernen Fahrzeugen bereits installiert. Dort können sie beispielsweise die Verkehrsschilder erkennen und aktuelle Verkehrshinweise auf den Fahrermonitor projizieren. Auch als Verkehrsüberwachungskameras an der Frontscheibe von Fahrzeugen werden Videokameras genutzt (sogenannte *Dashcams*). Um als Floating Car Observer zu fungieren, muss der Videostream einer Kamera mithilfe von Bildprozessierungswerkzeugen auf wiederkehrende Muster untersucht werden. Dafür geeignet sind beispielsweise Nummernschilder von Fahrzeugen. Aufgrund ihrer Verbreitung in Fahrzeugen, günstigen Nachrüstungskosten sowie guten Erkennungsraten des Verkehrs sind Videokameras sehr beliebt zur Implementierung von Floating Car Observern. Datenschutzrechtliche Fragen sowie schwierige Belichtungsbedingungen gehören zu den Limitationen von Videokameras als FCO-Sensor (Wolf et al. 2008b; Weisheit et al. 2011).

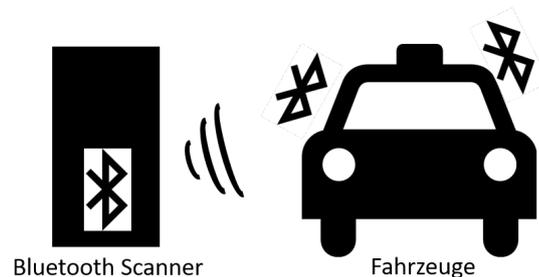


Abbildung 10: Schematische Darstellung der mobilen Verkehrszählung mit Bluetooth- und WiFi Sensoren in FCO-Anwendungen

Signaldetektoren zählen den Umgebungsverkehr, indem sie ihre direkte Umgebung auf Bluetooth- oder WiFi-Signale abtasten. Abbildung 10 zeigt die Nutzung von Bluetooth- und WiFiscannern zur Zählung von Fahrzeugen im Umgebungsverkehr schematisch. Ein Großteil der Verkehrsteilnehmer trägt mobile Geräte mit sich, zum Beispiel Mobiltelefone. Diese und weitere Geräte, wie einige moderne Kraftfahrzeuge selbst, verfügen über Technologien zur Datenübertragung. Die Bluetooth- und WiFitechnologie sind prominente Vertreter der Übertragungstechniken mobiler Geräte. Bei beiden Technologien besitzt jedes Endgerät eine eindeutig zuweisbare ID, die für FCO-Anwendungen genutzt werden kann. Werden alle individuellen IDs im Umfeld eines Signaldetektors gezählt, können Rückschlüsse auf das Volumen des Umgebungsverkehrs gezogen werden. Ein großes Problem und Bestandteil aktueller Forschung dabei ist, dass es Fahrzeuge gibt, die keinen Emittenten von digitalen Signalen besitzen. Andere Fahrzeuge hingegen könnten mehrere Bluetooth- oder WiFi-Signale absondern, was das Ziehen von Rückschlüssen auf das Volumen des Umgebungsverkehrs eines Fahrzeuges limitiert (Gurczik 2016; Gurczik 2017).

Die Messung des Straßenabriebes von vorausfahrenden Fahrzeugen zur Abschätzung des Verkehrsvolumens, wie in Abbildung 11 gezeigt, ist ein neues Forschungsgebiet im Bereich der Floating Car Observer.



Abbildung 11: Schematische Darstellung der mobilen Verkehrszählung mit Feinstaubsensoren in FCO-Anwendungen

Wie in Abbildung 11 zu sehen ist, nutzen auf Feinstaubsensoren basierende FCO-Systeme den Straßenabrieb der vorausfahrenden Fahrzeuge, um Rückschlüsse auf deren Anzahl zu ziehen. Jedes Fahrzeug emittiert beim Fortbewegen Schadstoffe. Diese variieren je nach Motorart. Gleichbleibend für jegliche Antriebsarten (Ottomotoren, Gasmotoren, Dieselmotoren, Elektromotoren) ist jedoch der Feinstaubausstoß durch Aufwirbelungen von Partikeln auf dem Fahrbahnbelag. Außerdem sorgen Brems- und Reifenabrieb für eine weitere Quelle von Feinstaub. Durch die richtige Platzierung von Feinstaubsensoren an einem Fahrzeug ist es möglich, durch kurzfristige Änderungen der Feinstaubwerte Rückschlüsse auf Fahrzeuge in der direkten Umgebung

zu ziehen (Röger et al. 2020a).

2.3 Erfassung und Visualisierung von Feinstaubsensordaten

Feinstaubdaten können auf verschiedene Arten erfasst werden. Gegenstand aktueller Forschung sind sowohl Ansätze, die statische Sensoren verwenden, als auch die Nutzung mobiler Sensoren. Bei der Datenerfassung mit beweglichen Sensoren ist primär die Fahrzeugart, auf der die Sensoren installiert werden, ein wichtiger Punkt. Bei Messkampagnen jeglicher Art ist darüber hinaus eine wichtige Fragestellung, welche Messparameter aufgezeichnet werden sollen und welche Sensoren dafür verwendet werden. Feinstaubsensoren sind mittlerweile zu vergleichsweise günstigen Preisen zu erwerben. Viel diskutiert ist daher aktuell der Einbezug bürgerlich erfasster Messdaten für die Bearbeitung wissenschaftlicher Fragestellungen.

2.3.1 Citizen Science

Auf *Citizen Science* basierende Messkampagnen nutzen häufig kostengünstige Feinstaubsensoren, die von bürgerlichen Hilfskräften installiert und betrieben werden. Es existiert eine Reihe an Anwendungen, bei denen Messdaten primär von nicht-wissenschaftlichen Akteuren aufgezeichnet werden. Benabbas et al. (2019) präsentieren einen auf *Citizen Science* basierenden Ansatz zur Messung der Feinstaubkonzentration innerhalb eines definierten Testgebietes. Sie sehen die steigende Nachfrage von Luftschadstoffdaten als Chance, das bereits existierende Messnetz durch eine höhere Anzahl an kostengünstigen Sensoren zu erweitern. In diesen Prozess sollen auch bürgerliche Hilfskräfte mit einbezogen werden. Dazu werden die Sensorenbaukästen von *luftdaten.info*, basierend auf einem *Nova SDS01* Feinstaubsensor, verwendet. Um auch mobile Messungen zu ermöglichen, erweitern die Autoren den Sensor um ein *LoRa*-Modul. Die Leistungsfähigkeit der Sensoren wird in einem Experiment evaluiert. Dafür werden drei Sensoren an unterschiedlichen Standorten neben offiziellen Messstationen ausgewählt, um deren Messwerte mit denen hochpräziser Sensoren zu vergleichen. Die Datenanalyse zeigt, dass günstige, von bürgerlichen Hilfskräften betriebene, Feinstaubsensoren durchaus dazu in der Lage sind, bestehende Messnetzwerke zu erweitern. Limitationen sehen Benabbas et al. (2019) im Einfluss von Witterungsbedingungen (speziell der Luftfeuchtigkeit) sowie der zwingenden Implementierung von Autokalibrierungsverfahren, um Messungenauigkeiten auszugleichen (Benabbas et al. 2019).

Snik et al. (2014) stellen einen Sensor für Citizen Science Messkampagnen vor, der mit herkömmlichen Mobiltelefonen betrieben werden kann. Typischerweise werden derartige Messkampagnen mit kostengünstigen Feinstaubsensoren durchgeführt, deren Messwerte mithilfe von weiteren Modulen zu einer zentralen Datenbank übertragen werden. Der Ansatz von Snik et al. (2014) sieht dagegen vor, das Smartphone selbst als Übertragungsmedium zu benutzen, was *Citizen Science* Kampagnen noch einfacher durchzuführen macht. Der vorgestellte *iSPEX* Sensor wird dabei vor der Smartphonekamera angebracht. Er misst den *Grad der linearen Polarisierung* (DoLP). Des Weiteren kann der Lichtstreuungswinkel beim Schwenken des Smartphones bestimmt werden. Die zum *iSPEX* Sensor gehörende Applikation kann aus diesen Kenngrößen Rückschlüsse auf den Aerosolgehalt der Luft ziehen. In einer Messkampagne testeten die Autoren das System in den Niederlanden. Dabei werden 8000 Sensoren unter der Bevölkerung verteilt. Die Ergebnisse zeigen vielversprechende Ansätze, allerdings muss dafür eine sinnvolle räumliche Verteilung der Sensoren bewerkstelligt werden (Snik et al. 2014).

Um mehr Bewusstsein in der Gesellschaft für das Thema Luftverschmutzung zu schaffen, präsentieren Mahajan et al. (2020) eine auf *Citizen Science* basierende Lösung. Mithilfe einer Kombination aus lernbasierten Spielen, Workshops und Messungen mit Feinstaubsensoren helfen die Autoren Bürgern dabei, das Thema Luftverschmutzung besser zu verstehen. Ihr Ansatz basiert auf drei Säulen: *Inklusion*, *Kollaboration* und *kritische Diskussion*. Hierbei soll das *Interactive Air Pollution Quiz* Teilnehmern der Studie auf spielerische Weise Grundlagenwissen zum Thema Luftverschmutzung und insbesondere Feinstaub vermitteln. In Workshops wird Mitwirkenden beigebracht, wie sie mithilfe von Feinstaubsensoren eigene Messungen durchführen können. In einer anschließenden Messkampagne werden die Teilnehmer ermutigt, Daten zu sammeln und die Ergebnisse anschließend mit anderen Studienteilnehmern zu teilen. Die Autoren schließen, dass *Citizen Science* Projekte eine gute Möglichkeit sind, Wissen über komplexe Themen wie die Luftverschmutzung und Feinstaubmessungen der Gesellschaft zugänglich zu machen (Mahajan et al. 2020).

2.3.2 Datenerfassung mit mobilen Feinstaubsensoren

Gozzi et al. (2016) geben in ihrer Arbeit einen Überblick über den *Status Quo* der mobilen Datenerfassung mit Feinstaubsensoren und diskutieren Chancen sowie Limitationen des Ansatzes. Sie sehen den Grund für den Anstieg der Popularität bewegter Sensoren primär in der gestiegenen Verfüg-

barkeit von Systemen unterschiedlicher Genauigkeit und Preiskategorie. Vornehmlich optische Partikelzähler (OPC), die Feinstaubwerte durch Lichtbrechung in einer Messkammer bestimmen, werden sehr häufig genutzt in mobilen Messkampagnen. Die Art, wie dabei Sensoren bewegt werden, ist sehr unterschiedlich. Gozzi et al. (2016) machen Rucksäcke, Autos, Fahrzeuge im ÖPNV, Fahrräder sowie Multikopter als beliebte Trägersysteme mobiler Messkampagnen aus. Gegenüber herkömmlichen Datenströmen wie beispielsweise offiziellen Messstationen haben mobil bewegte Feinstaubsensoren einige Vor- und Nachteile. Generell hängt die Genauigkeit der Messungen stark von den benutzten Messgeräten ab. Portable Sensoren geben in der Regel weniger akkurate Ergebnisse aus als große, hochpräzise Messgeräte. Darüber hinaus haben Feinstaubwerte in urbanen Räumen eine sehr hohe räumliche und zeitliche Variabilität, die stark von Umgebungsfaktoren wie Witterungsbedingungen und dem Verkehrsvolumen abhängt. Diese kann nur durch eine hohe Anzahl an Messungen ausgeglichen werden. Dagegen ist die erreichbare Messabdeckung bei mobilen Messungen deutlich höher als bei herkömmlichen, statischen Ansätzen. Des Weiteren ermöglichen mobile Ansätze auch Messungen in ländlichen, abgelegenen Gebieten, in denen die Abdeckung hochpräziser statischer Messstationen häufig nicht gut ist. Das gilt beispielsweise auch für Messungen in Entwicklungsländern, in denen finanzielle Mittel zur Errichtung von präzisen Sensornetzwerken häufig nicht vorhanden sind. Mobile Sensoren haben darüber hinaus auch den Vorteil, in kurzfristig auftretenden Messstandorten eingesetzt werden zu können. Dies ist beispielsweise der Fall bei plötzlich auftretenden Waldbränden in Gebieten, die nicht über statische Messstationen verfügen. Die Autoren schließen, dass die mobile Datenerfassung mit Feinstaubsensoren ein wichtiges Feld zukünftiger Forschung ist. Die Datenqualität ist jedoch ein Faktor, der beim Messen und Analysieren Beachtung geschenkt werden muss (Gozzi et al. 2016).

Die Nutzung mobiler Sensoren findet in einigen Studien Anwendung. Anjomshoaa et al. (2018) stellen ihre mobile Messplattform *City Scanner* vor. Urbane Räume verfügen über eine Reihe von messbaren Parametern, die unterschiedlichsten Kategorien zuzuordnen sind. Beispiele hierfür sind akustische Parameter, die Helligkeit sowie die Luftqualität. Die Autoren gehen bei ihrer Messung deshalb nach der *FEELS* Klassifikation vor. Die Buchstaben des Begriffes stehen für unterschiedliche physische Aspekte der Umwelt:

- *Fluid*: Bestandteile der Umgebungsluft. Hier spielt vorrangig die Zusammensetzung und Schadstoffbelastung der direkten Umwelt eine Rolle.

- *Electromagnetic*: Elektromagnetische Strahlung, die von mobilen Geräten wie Mobiltelefonen oder Sendemasten emittiert wird.
- *Envelope*: Kategorie für den Grad der Bebauung, zum Beispiel Häuser am Straßenrand.
- *Light*: Messung von sichtbarem und unsichtbarem Licht.
- *Sound*: Kategorie für akkustische Parameter, zum Beispiel die Lärmverschmutzung.

Die Parameter der *FEELS* Klassen werden von Anjomshoaa et al. (2018) mithilfe von mobilen Messeinheiten erfasst. In einem Experiment installieren sie den *City Scanner* an mehreren Fahrzeugen der städtischen Müllabfuhr von Cambridge (Massachusetts, USA). Der Sensoraufbau beinhaltet Videokameras, Thermalkameras, Beschleunigungssensoren, Schadstoffsensoren, Lichtintensitätssensoren, Positionierungssensoren, WiFi Scanner sowie Temperatur- und Luftfeuchtigkeitmessgeräte. Der resultierende Datensatz besteht aus 1,6 Millionen Observationen. Teil der Datenanalyse ist die Bestimmung des Wärmeflusses der Gebäude am Straßenrand mithilfe der Aufnahmen der Thermalkameras. Mithilfe der Luftschadstoff-Messgeräte wird außerdem eine Karte der Schadstoffausbreitung innerhalb des Testgebietes erstellt. Die Autoren der Studie sehen großes Potential in der mobilen Nutzung von Sensoren. Sie heben hervor, dass mobile Sensoren gut auf bereits existierenden Fahrzeugflotten, wie zum Beispiel Fahrzeugen der Müllabfuhr, installiert werden können. Dadurch müssen Messfahrzeuge nicht zusätzlich definierte Routen abfahren. Herausforderungen sehen die Autoren im Datentransfer, der Bereitstellung von Energie für mobile Sensoren sowie in der Interpretation von Messergebnissen (Anjomshoaa et al. 2018).

Gao et al. (2016) sehen ebenfalls Potential in der Installation von Sensoren auf bereits bestehenden Fahrzeugflotten. Sie wählen Busse einer Stadt in China als Träger von Feinstaubsensoren zur Messung von $PM_{2.5}$ Feinstaub. Durch eine Literaturrecherche kommen die Autoren zum Schluss, dass der instabile Luftfluss in der Umgebung von mobilen Feinstaubsensoren ein Problem darstellt. Luftstau und sich dynamisch ändernde Fließbedingungen der Umgebungsluft fälschen Messergebnisse maßgeblich ab. In ihrer Studie setzen Gao et al. (2016) deshalb auf eine Sensorbox, die durch ein schräg verbautes Luftrohr innerhalb der Box Luft zu einem *Dylos* Feinstaubsensor leitet. Für Labortests werden darüber hinaus Windsensoren eingesetzt. Der Sensoraufbau erhält

den Namen *MOSAIC*. Als Testgebiet wählen die Autoren eine neun Quadratkilometer große Fläche aus. Um eine möglichst gute Abdeckung des Gebietes zu erlangen, werden darüber hinaus bestimmte Buslinien für die Installation der Sensoren definiert. Die betroffenen Busse werden über einen Algorithmus bestimmt. Gao et al. (2016) kommen zum Schluss, dass mobile Messsysteme zur Bestimmung der Feinstaubkonzentration wie *MOSAIC* großes Potential haben, da mit niedrigen Investitionen bereits akzeptable Messgenauigkeiten und gute Gebietsabdeckungen der Messungen möglich sind (Gao et al. 2016). *BlueAer*, vorgestellt von Hu et al. (2016), ist ein Konzept zur dreidimensionalen Modellierung von $PM_{2.5}$ Feinstaub. Die Autoren merken an, dass die weitläufig verwendeten Systeme zur Bestimmung von Feinstaubkonzentrationen (wenige und hochpreisige Messstationen sowie ein dichtes Netz günstiger Sensoren) einen Schwachpunkt in der Sensorverteilung haben. In Städten, in denen viele hohe Gebäude stehen, sind herkömmliche zweidimensionale Messansätze (mit Sensoren am Boden) nicht ausreichend. *BlueAer* sieht daher vor, herkömmliche Sensornetzwerke mit mobilen Messungen in unterschiedlichen Höhen zu ergänzen. Mobile Messungen werden durch Taxis bewerkstelligt, die mit Feinstaubsensoren ausgestattet sind. Für Messungen in größeren Höhen kommen Hexakopter zum Einsatz, die ebenfalls Feinstaubsensoren mitführen. Die Messergebnisse werden mithilfe einer dreidimensionalen probabilistischen Konzentrationsabschätzung (3D-PCEM) interpoliert, sodass die Feinstaubkonzentration verschiedener Höhenstufen sichtbar wird. Die Autoren testen das *BlueAer* System in einer Fallstudie in der chinesischen Stadt Hangzhou. Für ein Jahr werden Datensätze aufgezeichnet. Für den Zeitraum von sechs Monaten werden zusätzliche Vergleichsmessungen zur Evaluierung aufgenommen. Daraus ergibt sich eine Fehlerrate von 5,669 Prozent. Die Autoren schließen, dass mobile Ansätze wie *BlueAer* vielversprechende Ergebnisse zur Aufzeichnung der städtischen Feinstaubkonzentration liefern. Darüber hinaus sehen sie Potential in der dreidimensionalen Modellierung von Feinstaubdaten in Städten (Hu et al. 2016).

Budde et al. (2019) gehen bei der Aufzeichnung von Feinstaubdaten einen anderen Weg als mit herkömmlichen Messgeräten. *FeinPhone* ist ein System, das Nutzer zur Messung von Feinstaub direkt auf ihrem Mobiltelefon befestigen können. Das Messgerät ist günstiger und einfacher als herkömmliche Messgeräte zu handhaben und damit gut für mobile Messungen im Zusammenhang mit *Citizen Science* Kampagnen anzuwenden. Herkömmliche, auf Lichtstreuung basierende, Feinstaubsensoren besitzen eine Lichtquelle sowie eine Diode zur Detektion von durch Partikel gebrochenes Licht. Im Fall von *FeinPhone*

fungiert das Blitzlicht der Smartphonekamera als Lichtquelle und die Fotolinse des Mobiltelefons als Lichtempfänger. Das als Prototyp 3D-gedruckte Gehäuse zum Befestigen an einer Mobiltelefonkamera fungiert als Messkammer. Um die Strahlung der Lichtquelle umzuleiten, ist darüber hinaus noch ein Spiegel in den Aufbau integriert. Mithilfe einer auf dem Mobiltelefon laufenden Applikation können aus den erzeugten Bildern Informationen über die Menge und Größenverteilung der Partikel abgeleitet werden. Zur Evaluierung der Genauigkeit des Systems nutzen Budde et al. (2019) einen Labortest. Dabei werden fünf *FeinPhone* Einheiten in einer luftdichten Kammer eingeschlossen. In diese wird im Anschluss Ammoniumsulfat injiziert. Zeitgleich wird durch die Messgeräte *TSI Aerodynamic Particle Sizer* (APS) und *Scanning Mobility Particle Sizer* (SMPS) ein Referenzwert bestimmt, anhand dessen die Genauigkeit der *FeinPhone* Partikelzähler abgeleitet wird. Die Ergebnisse zeigen Korrelationen von 94 bis 98 Prozent für die fünf getesteten Sensoren. Die Autoren sehen Potential für die Nutzung des *FeinPhone* Ansatzes zum Messen von Feinstaub und wollen in zukünftigen Arbeiten den Sensoraufbau optimieren (Budde et al. 2019).

2.3.3 Visualisierung von Sensordaten

Möglichkeiten, Sensormesswerte zu visualisieren, gibt es viele. Je nach Verarbeitung der Daten sind Repräsentationen als Punktwolken, Linien, Polygone und Raster möglich. Auch in der Literatur findet man diese Visualisierungsarten wieder. Werden Punkte zur Repräsentation verwendet, findet in der Regel kein Interpolationsprozess statt. Das reduziert den Grad der Abschätzung in Darstellungen, weil keine Werte, die sich zwischen den Messpunkten befinden, berechnet werden müssen. Punkte als Darstellung für Sensormesswerte finden sich in Pant et al. (2019) und Ford et al. (2019).

Mithilfe von Polylinien können Messtrajektorien nachvollzogen werden. Bei mobilen Messungen erscheint diese Art der Visualisierung sehr sinnvoll, denn sie enthält neben den Sensormesswerten zusätzlich noch Informationen über die Fahrtwege der benutzten Fahrzeuge. Ebenfalls häufig angewandt zur Visualisierung werden Polylinien, wenn die Messdaten auf Basis der Straßennetze dargestellt werden. Dabei werden Sensordaten räumlich mit Straßendaten verbunden. Polylinien zur Visualisierung von Sensordaten finden sich unter anderem in Báthory et al. (2019), Masic et al. (2017) und Röger et al. (2020b). Polygone finden sich in der Regel in Arbeiten, die Teilgebiete bestimmter Charakteristika bestimmen. Dazu zählen beispielsweise Feinstaub-Hotspots oder Kreuzungen mit besonders hohem Verkehrsaufkommen. Beispiele in der Lite-

ratur für Sensorwerte, die als Polygone dargestellt werden, finden sich unter anderem in Keler et al. (2015), Pan et al. (2018) und Howerton et al. (2020). Raster als Darstellung von Sensordaten sind in der angewandten Geoinformatik häufig vertreten. Viele Analysemethoden haben Rasterdaten als Ausgabedatentyp. Dazu zählen unter anderem Kerndichteschätzungen (KDE) und verschiedene Interpolationsalgorithmen. Im Gegensatz zu diskreten Vektordaten wie Punkte, Linien und Polygone sind Rasterdaten kontinuierlich. Sie verfügen für eine bestimmte Ausdehnung über flächendeckende Daten in ihren Rasterzellen. Makowski (2018), Reddy et al. (2020) und Liang et al. (2021) benutzen für die Visualisierung ihrer Sensordaten Raster.

Für die Darstellung von Sensordaten werden mitunter auch komplexere Methoden angewandt. Röger et al. (2019) nutzen Kartogramme, um erweiterte Floating Car Daten zu visualisieren. Kartogramme sind eine Abwandlung herkömmlicher Choroplethenkarten. Dabei werden unterschiedliche Werte verschiedener Teilgebiete der Karte nicht nur farblich voneinander getrennt. Diese Form der Visualisierung verzerrt die Geometrien von Formen so, dass ein bestimmter Sensorwert der Größe dessen zugeordneten Polygons entspricht Tingsheng et al. (2019). Das führt dazu, dass Werte innerhalb eines Kartogramms alleine durch Betrachtung der Formen abschätzbar sind.

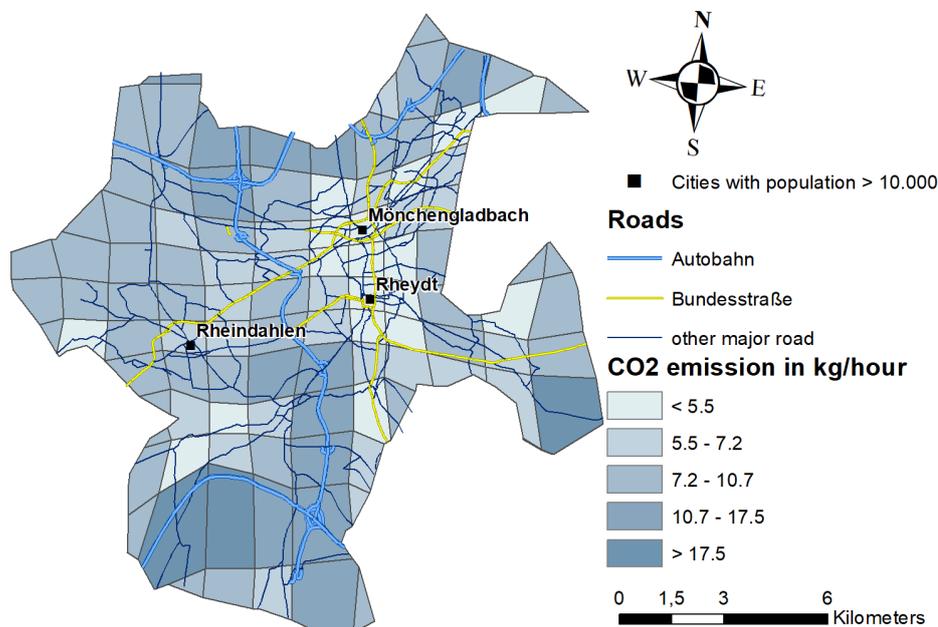


Abbildung 12: CO₂-Emissionen von Fahrzeugen in Mönchengladbach, dargestellt als Kartogramm. Quelle: Röger et al. (2019)

Abbildung 12 zeigt ein Kartogramm mit CO₂ Emissionen im Raum Mönchengladbach. Das Gebiet ist in Zellen von zwei auf zwei Kilometern unterteilt, die je nach Größe der Werte kleiner oder größer dargestellt werden. Es ist zu sehen, dass entlang der mit blauen Linien dargestellten Autobahnen die Emissionswerte vergleichsweise hoch sind. Kleinere Gitterzellen finden sich beispielsweise in den Stadtzentren von Mönchengladbach, Rheydt und Rheindahlen. Röger et al. (2019) schließen, dass Kartogramme eine gute Möglichkeit sein können, Sensorwerte zu visualisieren. Dafür muss allerdings eine homogene Punktedichte pro Zelle vorliegen, damit die Ergebnisse nicht durch Ausreißerwerte verfälscht werden (Röger et al. 2019).

Die Herausforderung beim Visualisieren von Qualitätsfaktoren mobiler Sensordaten ist, dass es sich oft um multivariate Datensätze handelt. Bis zu drei Dimensionen können problemlos mithilfe von Choroplethenkarten dargestellt werden (typischerweise geographische Länge, geographische Breite und eine weitere Variable). Da Sensordatensätze in einigen Fällen jedoch mehr als nur ein Qualitätsmaß haben, gestaltet sich die Visualisierung schwierig. Röger (2019) extrahiert aus einem mobil aufgenommenen Sensordatensatz neun Qualitätsfaktoren, die im Zuge einer einzigen Visualisierung dargestellt werden sollen. Es handelt sich dabei um die Punktedichte im jeweiligen Teilgebiet der Messung sowie die Anzahl an erreichbaren Satelliten und horizontale Abweichung der Präzision (HDOP) der Positionierungsdaten. Des Weiteren sollen die Standardabweichungen der Messwerte für die Helligkeit, die Temperatur, die Luftfeuchtigkeit, den Kohlenstoffdioxidwert und die Feinstaubwerte PM_{1} , $PM_{2.5}$ und PM_{10} dargestellt werden. In vielen Studien werden multivariate Datensätze durch *Wettersymbole*, *parallele Koordinaten* und *Star Plot Maps* visualisiert (Chambers 2018; Hauser et al. 2002). *Star Plot Maps* haben den Vorteil, dass die Daten im Gegensatz zu der Visualisierung mit parallelen Koordinaten auf einer zweidimensionalen Karte dargestellt werden können, sodass auch räumliche Einordnungen möglich sind. *Star Plots* sind sternförmige Symbole, die auf einer Karte angeordnet werden. Die Anzahl dargestellter Werte entspricht der Anzahl der Zacken in jedem Stern. Ist ein Wert höher, dann ist der jeweilige Zacken größer. Bei einem niedrigeren Wert stellt sich der Sternzacken kleiner dar. Ein Beispiel für eine *Star Plot Map* mit Qualitätsfaktoren mobiler Sensordatensätze wird auf Abbildung 13 gezeigt.

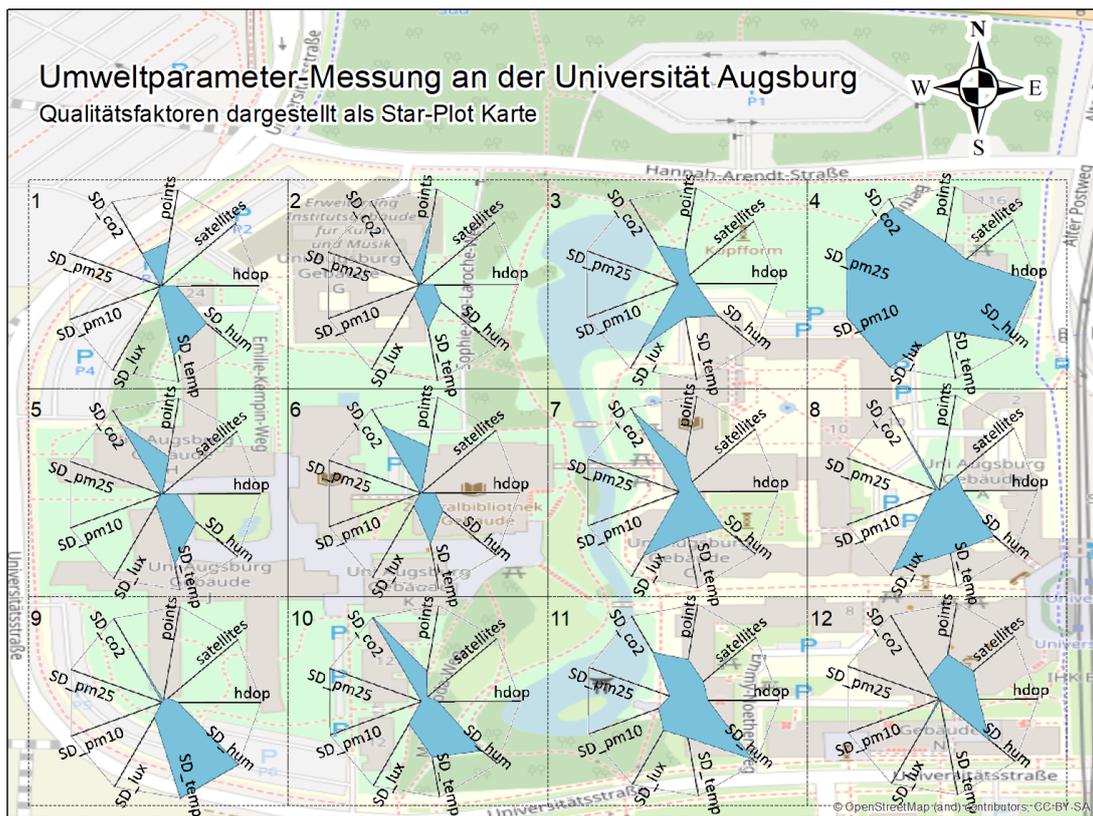


Abbildung 13: Qualitätsfaktoren einer mobilen Sensormessung, dargestellt als *Star Plot Map*.
Quelle: Röger (2019)

Abbildung 13 zeigt eine *Star Plot Map*, die Qualitätsfaktoren einer mobiler Sensordatenmessung darstellt. Das Testgebiet ist in zwölf Teilgebiete untergliedert, wobei jede Zelle mit einem Star Plot versehen ist. Durch die Darstellung lassen sich alle neun Qualitätsfaktoren übersichtlich und zeitgleich ablesen. Die Kartendarstellung lässt darüber hinaus auch im Gegensatz zu Diagrammen eine räumliche Zuordnung der Werte zu (Röger 2019).

3 Stand der Forschung zur Qualität von Sensordaten

Die Erfassung und Analyse von Sensordaten ist ein wissenschaftlich hochaktuelles Thema. Dabei geraten nicht nur Datensätze statischer Sensoren in den Fokus, sondern auch mobil erfasste Daten. Neben der Entwicklung und Anwendung von Methoden zur Analyse dieser Daten ist die Datenqualität ein viel diskutierter Faktor. Im Endeffekt entscheidet die Qualität der erfassten Geodaten über die Aussagekraft der durchgeführten Analysen. Standards der Datenqualität helfen dabei, die Güte eines Sensordatensatzes zu beschreiben. Sie definieren Faktoren, die Einfluss auf qualitative Facetten eines Datensatzes nehmen. Bei der mobilen Erfassung von Feinstaubsensordaten existieren allerdings noch weitere Faktoren, die über die bekannten Qualitätsstandards hinausgehen.

3.1 Standards zur Beschreibung der Qualität von Sensordaten

Die Beschreibung der Qualität von Geodaten findet in den Metadaten statt. Diese enthalten weiterführende Informationen über den jeweiligen Datensatz. In der Richtlinie *ISO 19115-2:2019* werden Standards für die Metadaten von Geodaten festgelegt (ISO Central Secretary 2019). Typischerweise beinhalten diese unter anderem Informationen über die Urheber eines Datensatzes, die Art und das Format der Daten, die Qualität der Daten, die räumliche Auflösung und das Koordinatensystem. Die Datenqualität setzt sich aus verschiedenen Bestandteilen zusammen, die in der Richtlinie *ISO 19157:2013* adressiert werden (ISO Central Secretary 2013). Diese sollen dabei helfen, verschiedene Datensätze bezüglich ihrer Qualität vergleichbar zu machen. Das ist eine wichtige Hilfestellung, wenn für einen bestimmten Anwendungszweck der passendste Datensatz ausgewählt werden soll. Die Richtlinie zur Beschreibung der Datenqualität definiert unterschiedliche Ebenen. Diese beinhalten:

- die Vollständigkeit,
- die logische Konsistenz,
- die Positionsgenauigkeit,
- die temporale Genauigkeit,
- die thematische Genauigkeit,

- sowie die Benutzerfreundlichkeit.

Die Vollständigkeit (*Completeness*) setzt sich zusammen aus dem Datenüberschuss (*Commission*) und dem Datendefizit (*Omission*). Dabei beschreibt der Datenüberschuss die Präsenz von Elementen in einem Datensatz, die als überschüssig gelten. Das Datendefizit adressiert fehlende Werte in einem Datensatz, die normalerweise vorhanden sein sollten (Brodeur et al. 2019).

Die logische Konsistenz (*logical consistency*) besteht aus den Sub-Elementen der konzeptuellen Konsistenz (*conceptual consistency*), der Konsistenz des Wertebereiches (*domain consistency*) und der Formatkonsistenz (*format consistency*). Sie ist ein Maß für die Einhaltung von logischen Regeln der Attributierung, der Topologie und der Datenstruktur. Unter der konzeptuellen Konsistenz versteht man dabei die Einhaltung des konzeptuellen Schemas eines Datensatzes. Die Konsistenz des Wertebereiches bezieht sich darauf, dass die Werte eines Datensatzes in einem definierten Wertebereich liegen. Die Formatkonsistenz legt Regeln für die Nutzung verschiedener Datenformate fest (Brodeur et al. 2019).

Die Positionsgenauigkeit (*positional consistency*) gliedert sich in die absolute Positionsgenauigkeit (*absolute positional consistency*), die relative Positionsgenauigkeit (*relative positional consistency*) und die Rasterpositionsgenauigkeit (*gridded data positional consistency*). Die absolute Positionsgenauigkeit beschreibt dabei die Abweichung der gemessenen und realen Positionen. Die relative Positionsgenauigkeit bezieht sich auf die Abweichung einer Messung zu einer gewählten Referenzposition. Die Rasterpositionsgenauigkeit ist ein Maß für die Genauigkeit von Positionsdaten in einem Gitternetz (Brodeur et al. 2019).

Die temporale Genauigkeit (*temporal accuracy*) ist ein Maß für die Qualität aller zeitbezogenen Parameter eines Datensatzes. Sie besteht aus der Genauigkeit der Zeitmessung (*accuracy of a time measurement*), der temporalen Konsistenz (*temporal consistency*) und der temporalen Validität (*temporal validity*). Die allgemeine Korrektheit der im Datensatz vorkommenden zeitlichen Informationen wird von der Genauigkeit der Zeitmessung beschrieben. Die temporale Konsistenz bezieht sich auf die richtige zeitliche Einordnung von gemessenen Ereignissen. Die temporale Validität beschreibt den zeitlichen Gültigkeitsbereich von erfassten Geodaten (Brodeur et al. 2019).

Die thematische Genauigkeit besteht aus der Klassifizierungskorrektheit (*classification correctness*), der Genauigkeit nicht-quantitativer Attribute (*non quantitative attribute correctness*) und der Genauigkeit quantitativer Attribute (*quantitative attribute correctness*). Die Klassifizierungskorrektheit beschreibt, ob die

im Datensatz vorkommenden Klassen die vorgesehenen Wertebereiche abdecken. Quantitative Attribute liegen vor, wenn deren Werte miteinander vergleichbar sind. Dementsprechend wird deren Validität durch die Genauigkeit quantitativer Attribute beschrieben. Andere Attribute fallen in die Kategorie der Genauigkeit nicht-quantitativer Attribute (Brodeur et al. 2019).

Die Benutzerfreundlichkeit (*usability*) beschäftigt sich mit der Eignung eines Datensatzes für einen bestimmten Anwendungsbereich (Brodeur et al. 2019). Sensordaten verstehen sich als eine Art von Geodaten. Sie sind Datenströme, die aus m Tupeln und n Attributen bestehen. Die in der Richtlinie *ISO 19157:2013* genannten Dimensionen der Datenqualität Q sind demnach ebenfalls für Sensordaten gültig. Einige Studien wie Klein et al. (2009) nennen jedoch weitere wichtige Elemente der Datenqualität, die über die offiziellen Standards hinaus gehen. Zu diesen zählen:

- die Genauigkeit a ,
- das Vertrauen ϵ ,
- die Vollständigkeit c ,
- das Volumen d ,
- sowie die Aktualität u .

Die Genauigkeit a (*Accuracy*) beschreibt Messfehler, die durch statische Einflüsse auf die Datenqualität entstehen. Dazu gehören sensorspezifikationsbedingte Ungenauigkeiten, fehlerhafte Kalibrierungen von Sensoren und Umwelteinflüsse. Die Einheit und die Größe des Einflusses der Genauigkeit sind während einer Messung stets gleichbleibend (Klein et al. 2009).

Das Vertrauen ϵ (*Confidence*) beinhaltet Messfehler, die durch zufällige Ereignisse während einer Messung entstehen. Diese dynamische Faktoren können ihren Einfluss während einer Messung ändern. Dazu gehören fehlerhafte Messungen durch Vibrationen im Umfeld des Sensors und auftretende Störquellen in der Messumgebung (Klein et al. 2009).

Die Vollständigkeit c (*Completeness*) beschreibt das Auftreten von Messfehlern durch Funktionsstörungen eines Sensors. Fehlende Werte werden in der Regel interpoliert oder geschätzt. Durch die Analyse der Vollständigkeit kann abgeschätzt werden, zu welchem Grad ein Datensatz aus tatsächlich gemessenen Werten besteht (Klein et al. 2009).

Die Menge der erfassten Rohdaten einer Sensormessung wird mit dem Faktor des Volumens d (*Data Volume*) beschrieben. Dabei ist es entscheidend, welche

Menge an Rohdaten für eine einzelne Analyse verwendet wird. Je höher das Volumen der Rohdaten für eine Analyse ist, desto höher ist die Datenqualität (Klein et al. 2009).

Die Aktualität u (*Timeliness*) beschreibt das Alter der verwendeten Geodaten für eine Analyse. Sie wird definiert als der temporale Unterschied zwischen den Zeitstempeln der verwendeten Datenpunkte und der durchgeführten Analyse der Daten (Klein et al. 2009).

Die Datenqualität Q gilt als Kombination der Dimensionen a , ϵ , c , d und u . Die Genauigkeit a und das Vertrauen ϵ bilden in der Summe den numerischen Gesamtfehler der Messungen ab. Die Vollständigkeit c und das Volumen d geben Informationen über quantitative Qualitätsmerkmale eines Datensatzes. Die Aktualität beschreibt den Datensatz im temporalen Kontext (Klein et al. 2009).

Die Studie fokussiert sich auf die Datenqualitätsfaktoren, die den numerischen Gesamtfehler einer Messung beschreiben. Dazu gehören die statischen Einflüsse der Genauigkeit und die dynamischen Einflüsse des Vertrauens. Eine Reihe von Teilfaktoren beeinflussen beide Größen bei der mobilen Erfassung von Feinstaubdaten. Diese werden im Folgenden diskutiert.

3.2 Teilfaktoren der Qualität von mobil erfassten Feinstaub-sensordaten

Nach Röger (2021) wirken unterschiedliche Einflussfaktoren auf die Qualität von mobil erfassten Geodaten mit Feinstaubsensoren:

- Die Positionierungsgenauigkeit,
- die Sensorspezifikationen,
- baulich bedingte Messunterschiede identischer Sensoren,
- Witterungsverhältnisse,
- die Bewegungsgeschwindigkeit,
- die Sensorposition und -ausrichtung,
- sowie der Umgebungsverkehr während einer Messung.

Einige der genannten Faktoren gelten dabei auch für statische Messungen. Dazu gehören die Spezifikationen der verwendeten Sensoren, die Messunterschiede identischer Sensoren, die Witterungsverhältnisse sowie bedingt der

Umgebungsverkehr, sollte ein Sensor in unmittelbarer Straßennähe installiert sein.

Röger (2021) nimmt ebenfalls eine Unterteilung in statische und dynamische Faktoren vor. Einige der Einflussgrößen sind während einer Messung gleichbleibend und vorhersehbar, während andere häufigen und unvorhersehbaren Änderungen unterworfen sind. Die Positionierungsgenauigkeit lässt sich beiden Gruppen zuordnen. Zu den statischen Facetten gehören die Genauigkeit des verwendeten Positionierungssystems und die Art der Positionsbestimmung. Einfachere Ansätze sehen lediglich eine Lokalisierung auf Basis der Triangulation von Signallaufzeiten der Satelliten vor. Genauere Methoden nutzen weitere erdgebundene Referenzstationen (differenzielles GPS) oder die Kombination verschiedener Lokalisierungstechniken (assistiertes GPS) zur Optimierung der Positionierung. Zu den dynamischen Facetten der Positionierungsgenauigkeit gehören unter anderem die Anzahl der erreichbaren Satelliten, sowie die Unterbrechung und Abfälschung der Satellitensignale durch Wolken oder hohe Bebauung am Straßenrand. Sensorspezifikationen sind statischer Natur. Die Genauigkeit von Messungen mit verschiedenen Sensoren ist maßgeblich von der Qualität des genutzten Messgerätes beeinflusst. Neben der Messgenauigkeit selbst spielt dabei auch die Messverzögerung, der mögliche Messbereich und der Temperatur- und Luftfeuchtigkeitsbereich eine Rolle. Auch baulich bedingte Unterschiede identischer Sensoren sind ein statischer Faktor. Im Fall von lichtstreuungs-basierten Feinstaubsensoren können sich minimale Unterschiede der Abmessungen der Messkammern, der Leistung des Ventilators oder der Ausrichtung der Lichtquelle auf die Messergebnisse auswirken. Diverse Vergleichsmessungen wissenschaftlicher Studien belegen dieses Phänomen (Zheng et al. 2018; Zusman et al. 2020). Witterungsbedingungen wie der Verlauf der Luftfeuchtigkeit oder der Temperatur sind ein hochdynamischer Faktor. Wasserpartikel, die sich an Feinstaubpartikel anheften, verändern deren Reflektionseigenschaften. Das kann dazu führen, dass ein deutlich höherer Feinstaubgehalt in der Umgebungsluft gemessen wird, als er in Realität vorherrscht. Das Wetter ist hochfrequenten kleinräumigen Änderungen unterworfen, die für eine Messung nicht immer vorhersehbar sind. Die Bewegungsgeschwindigkeit eines Sensors ist ebenfalls dynamisch. Sie ist abhängig davon, mithilfe welchem System sich ein Sensor durch den Raum bewegt. Zwischen der Bewegung zu Fuß, per Fahrrad oder per Fahrzeug kann es große Geschwindigkeitsunterschiede geben, die sich auf die Qualität der Daten auswirken können. Feinstaubsensoren können sehr anfällig bezüglich des Luftstromes in ihrer direkten Umgebung sein

(Ahn et al. 2020; Goyal et al. 2018). Der Luftstrom ändert sich mit der Bewegungsgeschwindigkeit. Auch davon betroffen ist die Ausrichtung und Position eines Sensors. Je nach Ausrichtung und Positionierung eines Messgerätes kann der Einfluss des sich ändernden Luftstromes größer oder kleiner ausfallen. Dieser Faktor ist jedoch statisch und ändert sich in der Regel während einer Messung nicht. Letztendlich nimmt mit dem Verkehrsvolumen in der direkten Umgebung einer Messung ein weiterer hochdynamischer Faktor Einfluss auf die Qualität von Sensordaten. Fahrzeuge emittieren unabhängig ihrer Antriebstechnik (Verbrennungs- oder Elektromotoren) Feinstaub. Dabei stammen Emissionen nicht nur aus Verbrennungsprozessen, sondern auch aus Brems- und Reifenabrieb. Darüber hinaus wird auf dem Straßenbelag abgesetzter Feinstaub durch das Überfahren von Fahrzeugen wieder aufgewirbelt und in die Umgebungsluft gebracht (Gehrig et al. 2004; Röger 2021). In den anschließenden Unterkapiteln wird relevante Literatur vorgestellt, die sich mit der Erforschung der genannten Einflussgrößen auf die Qualität von mobil erfassten Sensordaten beschäftigt.

3.3 Erfassung und Eliminierung von Positionierungsungenauigkeiten

Die Analyse von raum-zeitlichen Daten gewinnt immer mehr an Bedeutung, nicht zuletzt durch die steigende Beliebtheit von mobilen Erfassungsansätzen von Sensordaten oder durch die Präsenz von vernetzten Fahrzeugen im Straßenverkehr. Durch Orts- und Zeitstempel in Sensordaten wie zum Beispiel erweiterten Floating Car Daten können Observations genau im Raum verortet werden. Ein Problem bei der Nutzung von Positionsdaten aus GNSS-Sensoren ist, dass diese nie die wahre Position wiedergeben. Ausgelesene Ortsdaten verstehen sich stets nur als Annäherung an die Realität mit gewissen Fehlerwerten, die von verschiedenen Faktoren wie dem Stand der Satelliten oder der Wolkenbedeckung abhängen. Mapmatching ist eine Möglichkeit, etwaige Positionierungsfehler durch Verschiebung der Punkte auf ein Straßennetz auszugleichen. Verschobene Punkte repräsentieren dann zwar immer noch nicht ihre tatsächliche Position, allerdings ist die Anordnung der Punkte innerhalb einer Trajektorie dann standardisiert. Das Mapmatching kann somit als Basis weiterer Analysen der Sensorwerte fungieren. Man unterscheidet zwischen drei grundlegenden Ansätzen im Mapmatching: *lokal*, *inkrementell* und *global* (Dogramadzi et al. 2021).

Lokales Mapmatching sieht vor, einzelne Punkte einer Trajektorie auf die

nächstgelegene Netzwerkkante zu verschieben. Dabei spielen Positionen vor und nach dem verschobenen Punkt keine Rolle. Distanzbasiertes lokales Mapmatching funktioniert dann gut, wenn sich Punkte entlang einer einzelnen Straße ohne Kreuzungen oder Nebenstraßen befinden. Problematisch wird dieser Ansatz dann, wenn sich mehrere Netzwerkkanten in der direkten Umgebung des jeweiligen Datenpunktes befinden. In der Umgebung von Kreuzungen kann es darüber hinaus passieren, dass einzelne Stützpunkte einer Trajektorie zwischen mehreren Straßen hin- und herspringen. Datenrauschen ist eine große Limitation von distanzbasierten Mapmatching-Algorithmen. Folglich wird der lokale Ansatz trotz geringem Rechenaufwand in wissenschaftlichen Studien nur wenig genutzt (Dogramadzi et al. 2021). Beispiele in der Literatur finden sich in Yin et al. (2004) und Blazquez et al. (2005).

Der inkrementelle Ansatz ist ein weiterer Typ von Mapmatching Lösungen. Dabei werden Punkte auf Basis der Position ihres Vorgängerpunktes auf Netzwerkkanten verschoben. Ein großer Vorteil dabei ist, dass sowohl topologische Informationen als auch die Distanz zwischen Punkten gewahrt werden können. Die Qualität der Ausgaben gilt als höher als beim lokalen Mapmatching, bei einem etwas höherem Rechenaufwand. Als ein häufig genutzter Vertreter inkrementeller Mapmatching-Ansätze gilt der Algorithmus nach Newson et al. (2009). Dabei werden die Methoden des *Hidden Markov Modells* und des *Viterbi Algorithmus* genutzt. Die Metriken der Messwahrscheinlichkeit (die Wahrscheinlichkeit, dass ein GPS Punkt einer bestimmten Netzwerkkante zugehörig ist) und der Übergangswahrscheinlichkeit zwischen Netzwerkkanten sind die Basis des Ansatzes. Eine Route, bei der das Fahrzeug in einer hohen Frequenz von Straße zu Straße springt, hat zum Beispiel eine niedrige Übergangswahrscheinlichkeit. Mithilfe eines *Hidden Markov Modells* wird die Zugehörigkeit der GPS Punkte zu den Netzwerkkanten festgestellt. Um die wahrscheinlichste Route durch das Netzwerk zu bestimmen, kommt ein *Viterbi Algorithmus* zum Einsatz. Diejenige Route, die die beste Kombination aus Messwahrscheinlichkeiten und Übergangswahrscheinlichkeiten aufweist, wird als die wahrscheinlichste Route durch das Netzwerk deklariert (Newson et al. 2009).

Der Rechenaufwand zur Implementierung des *Newson Algorithmus* nach Newson et al. (2009) gilt als sehr hoch. Dogramadzi et al. (2021) lösen dieses Problem, indem sie eine Optimierung des Ansatzes vornehmen. Zunächst detektieren die Autoren Elemente des Algorithmus, die besonders viel Zeit in Anspruch nehmen. Sie kommen zum Schluss, dass die Kalkulation der Distanzen von GPS Punkten zu Netzwerkelementen den größten Rechenaufwand

erzeugt. Um das Problem zu umgehen, wird ein Vorprozessierungs-Schritt für das Straßennetzwerk vorgeschlagen. Netzwerksegmente, die vom Punkt X_a bis zum Punkt X_z verlaufen, werden so lange mit weiteren Knoten verdichtet, bis jeder Punkt maximal 60 Meter vom nächsten Punkt entfernt liegt. Dabei wird die Einzelsegmentlänge in jeder Iteration halbiert. Experimente mit zwei Datensätzen zeigen signifikante Verbesserungen des Rechenaufwandes, der um bis zu 90 Prozent reduziert wird (Dogramadzi et al. 2021).

Globale Mapmatching Algorithmen bauen auf inkrementellen und lokalen Ansätzen auf. Zusätzlich wird der Verlauf der kompletten Trajektorie von Anfang an in die Berechnungen mit einbezogen. Lokale und inkrementelle Ansätze gelten als Anfällig hinsichtlich der Fehlerausbreitung. Ein Datenpunkt mit besonders hohem Rauschen kann dafür sorgen, dass alle darauf folgenden Punkte dessen Fehlkalkulation mittragen. Die Berechnungsgenauigkeit von globalen Mapmatching Ansätzen gilt als die höchste aller drei Gruppen. Andererseits ist dafür in der Regel der höchste Rechenaufwand nötig. Darüber hinaus kommen viele Algorithmen mit hohem Aufwand in der Vorprozessierung der Daten mit einher. Da globale Mapmatching-Ansätze den Verlauf der kompletten Trajektorie mit einbeziehen, ist es möglich, Faktoren wie die ursprüngliche Form der Routen oder individuelle Fahrerprofile in die Kalkulation mit einzubeziehen. Ein Vertreter globaler Mapmatching Ansätze ist der von Gong et al. (2017) entwickelte *AntMapper* Algorithmus. Die Autoren bauen die resultierende Route auf Basis geometrischer und topologischer Eigenschaften der Ursprungstrajektorie auf. Des Weiteren spielen Ähnlichkeitsmaße der Eingabe- und Ausgabetrajektorie eine Rolle. Um die Ergebnisse des Ansatzes zu optimieren, kommt darüber hinaus ein auf natürlichen Beobachtungen einer Ameisenkolonie beim Transport von Essen inspirierter Algorithmus zum Einsatz. Ameisen verfügen nur über eine eingeschränkte Sicht. Trotzdem lässt sich beobachten, dass die Tierart besonders talentiert im Finden der kürzesten Route zwischen zwei Punkten ist. Sie sondern beim Laufen *Pheromone* aus. Die Agglomeration der Duftstoffe am Boden hilft nachfolgenden Ameisen, die perfekte Route zu finden. In einem ersten Schritt berechnen Gong et al. (2017) einen Informationsgraphen aus gegebenen GPS Punkten und einem Netzwerk. Der Informationsgraph besteht aus mehreren Sets C_n . Jedes Set C_n verfügt über m verschiedene Kandidatenpunkte. Die Kandidatenpunkte aus C_n sind mit allen Kandidatenpunkten aus C_{n+1} verbunden. Allen Verbindungslinien wird ein *Pheromon*-Wert zugewiesen. Im Anschluss wird eine Ameisenkolonie u definiert. Jede Ameise der Kolonie wird für den ersten GPS Punkt auf einen zufälligen Kandidatenpunkt im ersten Set des Informati-

onsgraphen gesetzt. Für die nächsten n Schritte (die Anzahl der GPS-Punkte) wird abwechselnd die Ameise auf den nächst wahrscheinlichen Punkt gesetzt und der *Pheromon*-Wert der genutzten Verbindungslinie aktualisiert. Der beste Pfad, gemessen an den *Pheromon*-Werten, wird schlussendlich ausgegeben (Gong et al. 2017).

3.4 Messunterschiede identischer und verschiedener Feinstaubsensoren

Für Messkampagnen sind Feinstaubsensoren unterschiedlicher Genauigkeit verfügbar. Dabei spiegelt sich die Qualität der Messungen in der Regel im bezahlten Preis für die Geräte wieder. Während für den Betrieb offizieller Messstationen hochpräzise Sensoren verwendet werden, wird ebenfalls viel Forschung mit kostengünstigen Sensoren betrieben. Alleine durch das Studieren der technischen Datenblätter kann allerdings keine präzise Aussage darüber getroffen werden, wie unterschiedlich genau bestimmte Sensoren gegenüber Referenzgeräten arbeiten. Aus diesem Grund gibt es einige Forschungsarbeiten, die sich mit Messunterschieden unterschiedlicher Sensoren beschäftigen.

3.4.1 Kalibrierung von Feinstaubsensoren

Ein Forschungsfeld der Evaluierung von Messunterschieden unterschiedlich präziser Feinstaubsensoren ist die Kalibrierung kostengünstiger Messgeräte. Durch Kalibrierungsansätze können kostengünstige Sensoren auch vertrauenswürdige Werte liefern. Im Folgenden werden unterschiedliche aktuelle Forschungsarbeiten dazu vorgestellt.

Lee et al. (2020) nutzen verschiedene Kalibrierungstechniken, um die Genauigkeit eines kostengünstigen Sensoraufbaus zu verbessern. Als Referenzgerät nutzen die Autoren einen *TE-PM711*-Sensor. Die zu kalibrierende Sensorbox besteht aus einem Mikrocontroller, einem Temperatur- und Luftfeuchtigkeitsmessgerät, einem Sensor zur Erfassung der Strahlungsintensität und drei *Plantower PMS7003*-Messgeräten. Für die Kalibrierung der Sensoren werden drei Kalibrierungstechniken angewandt. Die erste davon ist die *lineare Kalibrierung*, die auf der *multivariaten linearen Regression* basiert. Dafür verwenden die Autoren die Luftfeuchtigkeit, die Temperatur und die erfassten Feinstaubwerte der *PMS7003*-Sensoren als Prediktoren. Die selben Variablen nutzen die Autoren für die *nichtlineare Kalibrierung*, die auf einem *mehrlagigen Perzeptron* basiert. Als dritten Kalibrierungsansatz stellen Lee et al. (2020) die *kombinierte*

Kalibrierung vor. Dabei werden sowohl die *lineare*, als auch die *nichtlineare Kalibrierung* durchgeführt. Auf Basis des Vorhersagefehlers wird im Anschluss automatisch die genauer arbeitende Kalibrierungstechnik ausgewählt (Lee et al. 2020).

Zheng et al. (2019) nutzen ein Gaußprozess Regressionsmodell (GPR) zur dynamischen Kalibrierung eines Feinstaubsensornetzwerkes in Delhi (Indien). Die Autoren führen an, dass die Nutzung kostengünstiger Feinstaubmessgeräte in Sensornetzwerken immer relevanter wird. Dabei ist jedoch bei steigender Netzwerkgröße die Instandhaltung und Kalibrierung der einzelnen Messknoten eine Herausforderung. Für ihr Sensornetzwerk in Delhi nutzen Zheng et al. (2019) Sensorboxen, die auf dem Plantower PMS7003 Sensor basieren. Des weiteren verfügen die Boxen über einen Temperatur- und Luftfeuchtigkeitssensor, einen Positionierungssensor, einen SD-Kartenleser sowie ein GSM-Übertragungsmodul. Neben den kostengünstigen Sensorboxen befinden sich ebenfalls hochpräzise Referenzmessknoten im Netzwerk. Zur Kalibrierung der günstigen Module Nutzen die Autoren ein Gaußprozess Regressionsmodell, das folgendermaßen arbeitet:

1. Ausschluss einer Station aus der Gesamtheit der Referenzmessstationen
2. Berechnung einer linearen Regression (zur jeweils nächsten Referenzstation) der kostengünstigen Messpunkte
3. Optimierung der Hyperparameter des GPR
4. Kalibrierung aller kostengünstigen Messpunkte zu ihrem bedingten Erwartungswert mithilfe von linearer Regression
5. Aktualisierung der GPR Hyperparameter
6. Aktualisierte Kalibrierung der kostengünstigen Messpunkte
7. Vorhersage der ausgelassenen Referenzstation (Konvergenz)

Durch den letzten Schritt (Vorhersage der Werte der ausgelassenen Messstation) können Genauigkeitswerte wie der *Root Mean Square Error* (RMSE) sowie der Fehlerwert in Prozent berechnet werden. Die Autoren nutzen das GPR Modell, um ihre Messpunkte in Delhi zu kalibrieren und instand zu halten. Eine Fallstudie wird vom ersten Januar 2018 bis zum 31. März 2018 durchgeführt. Die Ergebnisse zeigen eine gute Stabilität und vergleichsweise gute Genauigkeit der Messungen. Der Fehlerwert liegt für 22 Referenzmessgeräte und zehn kostengünstige Messpunkte bei 30 Prozent (RMSE: $33 \mu\text{g}/\text{m}^3$)

). Durch die Anwendung des GPR Modells können zwei mit 50 Prozent Fehlerwert überdurchschnittlich ungenau messende Sensoren detektiert und aus dem Netzwerk ausgeschlossen werden (Zheng et al. 2019).

3.4.2 Gewichtete Interpolation unterschiedlicher Messwerte

Sensoren unterschiedlicher Spezifikationen liefern unterschiedlich qualitative Messwerte. Daher ist es nicht sinnvoll, Messwerte unterschiedlicher Datenquellen in Analysen gleichzustellen. Gerade bei der Interpolation von Feinstaubwerten macht es Sinn, genauere Sensoren höher zu gewichten als kostengünstige Messgeräte. Die Analysemethode der *gewichteten Interpolation* ist eine häufig angewandte Lösung bei der Verschneidung von Sensordaten aus unterschiedlich qualitativen Quellen (Shukla et al. 2020; Liang et al. 2021). Taichung City (Taiwan) verfügt über drei große Sensornetzwerke zur Erfassung von Feinstaub: die offiziellen Messstationen (17 Messpunkte), Messstationen der *Taiwan Environmental Protection Administration* (EPA, 506 Messpunkte) sowie ein Netzwerk privat betriebener Sensoren (420 Messpunkte). Dabei sind die offiziellen Messstationen mit hochpräzisen Sensoren ausgestattet, während die anderen beiden Messnetzwerke auf kostengünstigen Lösungen zur Feinstaubdatenerfassung basieren. Liang et al. (2021) untersuchen, welche Qualitätsmerkmale der unterschiedlichen Sensoren für eine gewichtete Interpolation eine Rolle spielen könnten. Die Autoren identifizieren die Linearität, die Sensitivität, die konstante Messabweichung, die Präzision, die Genauigkeit sowie die Verzerrung der zwei kostengünstigen Messnetzwerke als Qualitätsmerkmale. Eine Analyse der Daten der drei Messnetzwerke für das Jahr 2019 zeigt, dass die Linearität der Daten der offiziellen Messstationen durchgängig höher ist als die der beiden anderen Messnetzwerke. Dagegen ist die Sensitivität, die Standardabweichung, der relative Fehlerwert, die Messabweichung und die Verzerrung der kostengünstigen Sensoren höher als bei den hochpräzisen Messgeräten. Aus den Erkenntnissen können die Autoren eine gewichtete Interpolation berechnen, die eine Vielzahl unterschiedlicher Sensoren mit abweichenden Messgenauigkeitswerten mit einschließt. Die Autoren schließen, dass zur Nutzung von kostengünstigen Sensoren für eine akkurate Messung der Feinstaubkonzentration Schritte zur Kalibrierung und Korrektur der Messwerte nötig sind (Liang et al. 2021). Shukla et al. (2020) nutzen Interpolationsmethoden, um die $PM_{2.5}$ Feinstaubkonzentration in ausgewählten Gebieten in Delhi (Indien) zu beurteilen. Innerhalb des Untersuchungsgebietes existieren zwar 36 Messstationen, diese decken nach Ansicht der Autoren jedoch die Hotspots der Stadt nur unzu-

reichend ab. Aus diesem Grund unterstreichen Shukla et al. (2020) die Notwendigkeit, in der Datenanalyse auf modellbasierte Techniken zu setzen. In einer Fallstudie werden die Interpolationsmethoden *Inverse Distance Weighted* (IDW) und *Kriging* eingesetzt, um die räumliche Verteilung von $PM_{2.5}$ Feinstaub zu modellieren. Dafür werden die Daten von Januar 2015 bis Dezember 2016 verwendet. Eine Kreuzvalidierung zeigt Fehlerwerte von 19 Prozent (2015) und 27 Prozent (2016) für *IDW*, sowie 17 Prozent (2015) und 26 Prozent (2016) für *Kriging*. Die Autoren schließen, dass es keine wesentlichen Unterschiede bezüglich der Fehlerwerte gibt zwischen *IDW* und *Kriging*. Allerdings kann der *Kriging* Algorithmus bei ungleich verteilten Messstationen Lücken in der Interpolation produzieren, die bei der Anwendung von *IDW* nicht auftreten (Shukla et al. 2020).

3.5 Witterungsbedingte Einflüsse auf die Datenqualität

Ein Nebeneffekt, der bei vielen Feinstaubsensoren auftritt, sind Messfehler aufgrund von Wasserpartikeln in der Luft. Bei einer Messung mit Feinstaubsensoren fließt Luft durch eine Messkammer. Eine Laserdiode emittiert ein Lichtsignal. Dieses Signal wird durch die sich in der Luft befindlichen Feinstaubpartikel gestreut und reflektiert. Je nach dem, wie sehr dabei das Lichtsignal abgelenkt wird, können Rückschlüsse über die Größe und Menge der Feinstaubpartikel gezogen werden. Heften sich Wasserpartikel in der Messumgebung an die Feinstaubpartikel an, können diese Ergebnisse abgefälscht werden. Je höher die Luftfeuchtigkeit in der Umgebung des Sensors ist, desto höher fällt der Messfehler aus. Durch die Anwendung einer Luftfeuchtigkeitskorrektur kann dieser Messfehler aus dem Ergebnis eliminiert werden. Man unterscheidet hier zwischen dem trockenen Feinstaubwert PM_{dry} sowie dem nassen Feinstaubwert PM_{wet} . Letzterer beschreibt den Messwert vor der Luftfeuchtigkeitskompensation, während PM_{dry} den korrigierten Wert, der die Feinstaubkonzentration auf trockene Bedingungen projiziert, darstellt. Für die Luftfeuchtigkeitskorrektur ist ein Luftfeuchtigkeitswert der Umgebungsluft nötig. Dieser lässt sich über die Hinzunahme eines Luftfeuchtigkeitsensors bestimmen (Streibl 2017).

Der Zusammenhang zwischen PM_{dry} und PM_{wet} wird anhand der Funktion

$$PM_{dry} = \frac{PM_{wet}}{gf(rh)} \quad (1)$$

mit der Wachstumsfunktion $gf(rh)$ dargestellt. Nach Streibl (2017) existieren unterschiedliche Wachstumsfunktionen $gf(rh)$, die in Tabelle 6 gezeigt

werden.

Wachstumsfunktion	Definition
gf_{Haenel}	$gf_{Haenel} = \frac{1}{(1-rh)^\beta}$
gf_{Soneja}	$gf_{Soneja} = 1 + \frac{\alpha * rh^2}{1-rh}$
gf_{Combo}	$gf_{Combo} = 1 + \alpha * \frac{rh^2}{(1-rh)^\beta}$
gf_{Skupin}	$gf_{Skupin} = \begin{cases} \frac{\alpha}{(1-rh)^\beta} & rh \geq 0.7 \\ \frac{1}{(1-rh)^{\beta - \frac{\log(\alpha)}{\log(0.3)}}} & rh < 0.7 \end{cases}$

Tabelle 6: Wachstumsfunktionen für die Luftfeuchtigkeitskorrektur von Feinstaubsensordaten nach Streibl (2017)

Tabelle 6 zeigt unterschiedliche Wachstumsfunktionen, die für die Luftfeuchtigkeitskorrektur von Feinstaubsensoren anwendbar sind. Nach (Chakrabarti et al. 2004) sind für die Funktionen ein α -Wert von 1 sowie ein β -Wert von 0.25 zu verwenden.

3.6 Einfluss der Fahrgeschwindigkeit und der Sensorausrichtung auf die Datenqualität

Der Einfluss der Fahrgeschwindigkeit und der Sensorausrichtung ist ein bisher wenig beforschtes Feld. In der Literatur finden sich jedoch Hinweise für die Präsenz der Faktoren. In ihrer Studie kommen Goyal et al. (2018) zum Schluss, dass die Datenqualität von Feinstaubsensoren abnimmt, sobald diese bewegt werden. Des weiteren weisen Ahn et al. (2020) darauf hin, dass Feinstaubsensoren einen konstanten Luftstrom erfordern, was bei Messungen in Bewegung nicht immer möglich ist. Wang et al. (2020a) erforschen in ihrer Studie die Auswirkungen des Luftflusses auf die Qualität von Drohnen-basierten Feinstaubmessungen. Sensoren auf unbemannten Luftfahrzeugen (UAVs) stehen unter einem besonderen Einfluss des Luftflusses, da die Bewegung der Rotoren des Fluggerätes direkt auf das jeweilige Messgerät wirkt. Um deren Einfluss zu bestimmen, befestigen die Autoren Feinstaubsensoren statisch in unterschiedlichen Höhen. Darauf hin fliegen sie diese Messpunkte mit ihrem UAV ab und vergleichen die Messwerte der statischen Sensoren mit denen des bewegten Messgerätes. Wang et al. (2020a) stellen deutliche Messunterschiede der Datenreihen fest. Diese betragen 6,2 Prozent für $PM_{2.5}$ -Feinstaub und 6,6 Prozent für PM_{10} -Feinstaub. Die Autoren arbeiten an einem Korrekturalgorithmus, mithilfe dessen der Einfluss des durch die Rotorbewegung

zustande kommenden instabilen Luftflusses minimiert werden kann (Wang et al. 2020a).

3.7 Modellierung von Umwelt-Verkehr Beziehungen

Der Straßenverkehr sorgt weltweit für 27 Prozent der Luftverschmutzung. Eine Studie in 127 Städten, durchgeführt von Heydari et al. (2020), gibt Aufschluss über den Einfluss des Verkehrs auf die Feinstaubkonzentration in urbanen Räumen. Dafür werden mithilfe einer *bayesschen Meta-Regression* 155 bereits existierende Studien analysiert. Die Ergebnisse zeigen, dass der Einfluss des Verkehrs in Europa, Nordamerika und Ozeanien um 37 Prozent niedriger ist als im Rest der Welt. Außerdem zeigt die Analyse, dass Studien, die nach dem Jahr 2005 publiziert wurden, einen im Schnitt um 28 Prozent niedrigeren Einfluss berechnen. Auch hinsichtlich der Population einer Stadt gibt es deutliche Unterschiede. So ist der Einfluss des Verkehrs auf die Luftverschmutzung in Städten unter 500.000 Einwohnern um 24 Prozent geringer als in größeren Agglomerationen (Heydari et al. 2020).

Verkehr-Umwelt Beziehungen werden in der Literatur unterschiedlich modelliert. Zunächst einmal gibt es verschiedene Einflussgrößen auf die Luftverschmutzung, die untersucht werden. Laut der *European Environmental Agency* (EAA) sind Feinstaub, Kohlenstoffoxide, flüchtige organische Verbindungen (NMVOC), Stickoxide und Schwefeloxide primär für die Luftverschmutzung durch den Transportsektor verantwortlich¹⁰. Feinstaub in den Größen PM_1 , $PM_{2.5}$ und PM_{10} wird durch Verbrennungsprozesse, Reifenabrieb, Bremsabrieb und Aufwirbelungseffekte durch den Verkehr emittiert. Eine weitere Quelle von Feinstaub sind Schwefeloxide (SO_x). Sie treten bei der Verbrennung schwefelhaltiger Brennstoffe wie Kohle und Öl auf. SO_x können sich in der Atmosphäre zu Sulfatpartikeln umwandeln, die wiederum zur Gruppe des PM_{10} Feinstaubes gehören. Röger (2020), Röger et al. (2020a) und Habre et al. (2021) untersuchen Feinstaubemissionen von Fahrzeugen hinsichtlich ihres Einflusses auf die Luftqualität. Ein weiterer Einfluss des Verkehrs auf die Luftqualität ist die Emission von Kohlenstoffoxiden wie CO und CO_2 . Das Treibhausgas wird primär durch die Verbrennung von fossilen Stoffen wie Benzin ausgestoßen. Beispiele für Studien, die sich mit dem Einfluss des Verkehrs auf den Kohlenstoffoxidgehalt der Luft beschäftigen, finden sich in Röger et al. (2018), Li et al. (2019) und Zheng et al. (2020a). Stickoxide (NO_x)

¹⁰<https://www.eea.europa.eu/data-and-maps/indicators/transport-emissions-of-air-pollutants-8/transport-emissions-of-air-pollutants-8,03/21>

entstehen durch chemische Reaktionen von Stickstoff und Sauerstoff. Im Straßenverkehr entstehen diese Reaktionen vorrangig durch die Verbrennung von Dieselkraftstoff. Lawson et al. (2011) und Schiavon et al. (2015) untersuchen die verkehrsbedingte Emission von Stickoxiden. Flüchtige organische Verbindungen (NMVOC) entstehen ebenfalls durch Verbrennungsprozesse. Ihre Emission durch den Verkehr wird von Xie et al. (2020) und Freire et al. (2020) untersucht.

Es existieren unterschiedliche Ansätze, Verkehrsparameter mit Luftschadstoffkonzentrationen in Verbindung zu bringen. Einige Studien sehen vor, diverse Luftschadstoffe sowohl bei starkem, als auch bei schwachen Verkehrsaufkommen zu messen und zu vergleichen. Ein Beispiel dafür sind Messungen während der COVID-19 Pandemie. Rossi et al. (2020) und Chen et al. (2021) nutzen die Ausgangsbeschränkungen und deren Effekt auf das Verkehrsaufkommen mithilfe von Vergleichsmessungen, um Rückschlüsse des Verkehrsaufkommens auf die Luftqualität zu ziehen. Röger et al. (2020a) nutzen einen alternativen Ansatz. Ihr auf Random Forest basierendes Modell setzt gemessene Feinstaubwerte und die Präsenz/Absenz von Umgebungsverkehr eines Fahrzeuges in den Zusammenhang. Den Autoren gelingen dabei Vorhersagegenauigkeiten von bis zu 99 Prozent (Röger et al. 2020a). Auch zur Emission von Kohlenstoffoxiden gibt es unterschiedliche Modellierungsansätze. Eine Möglichkeit ist, wie bei der Arbeit mit Feinstaubwerten, Messungen bei unterschiedlichen Verkehrsvolumina zu tätigen. Eine Studie dazu findet sich in (Li et al. 2019). Röger et al. (2018) nutzen einen anderen Ansatz. Über OBDII-Adapter lassen sich die individuellen CO₂ Emissionen von Fahrzeugen während dem Fahren auslesen. Die Autoren untersuchen Korrelationen zwischen der Steigung von Straßen und dem Kohlenstoffdioxid-Ausstoß eines Testfahrzeuges. Eine Fallstudie zeigt moderat positive Korrelationswerte von 0,37 und 0,41 für zwei Streckenabschnitte in Augsburg, Deutschland (Röger et al. 2018).

3.8 Floating Car Observer zur Bestimmung des Umgebungsverkehrs

Gurczik (2016) modellieren die Nutzung von auf der Bluetooth Technologie basierenden Floating Car Observern. Im Gegensatz zu herkömmlichen FCO-Systemen sind diese in der Lage, auch Objekte aufzuspüren, die sich nicht direkt in Sichtweite des ausgestatteten Fahrzeuges bewegen. Das vom *deutschen Zentrum für Luft-und Raumfahrt* (DLR) entwickelte DYNAMIC-System

scannt dafür mithilfe eines Bluetooth-Empfängers in einem maximalen Radius von 100 Metern die Umgebung nach weiteren Bluetooth-Signalen ab. Um mehrfache Observationen desselben Gerätes zu vermeiden, werden die Ergebnisse nach der MAC-Adresse des Bluetooth-Signals gruppiert. Zur Modellierung verwenden Gurczik (2016) die *SUMO*-Simulationsumgebung. Die Umgebung besteht sowohl aus Fahrzeugen mit FCO Technologie, die als Empfänger fungieren, Fahrzeugen mit Bluetooth Sendern die von den Empfängern detektiert werden können, als auch aus nicht detektierbaren Fahrzeugen ohne Bluetooth-Sender oder Empfänger. Mithilfe einer Messkampagne in Berlin werden die modellierten Ergebnisse validiert. Dafür werden im August 2013 täglich von 9 bis 10 Uhr und von 13 bis 15 Uhr mit Floating Car Observern Daten aufgezeichnet. Die Messkampagne produziert ähnliche Ergebnisse wie die Simulation, wobei die Wahrscheinlichkeitsverteilung noch optimierbar ist und Gegenstand zukünftiger Forschung sein wird (Gurczik 2016).

Kühnel (2012) implementiert in seiner Arbeit einen Floating Car Observer zur Verkehrsdatenerfassung auf zweispurigen Landstraßen. Der Autor nennt das steigende Verkehrsvolumen auf den deutschen Straßen durch einen wachsenden Güter- sowie Individualverkehr eine große Herausforderung für die bestehenden Straßennetze. Stationäre Messsysteme zur Abschätzung des Verkehrsaufkommens sind gerade in städtischen Räumen abseits von Autobahnen unzureichend vorhanden, was die Zählung von Fahrzeugen zur weiteren Verkehrsplanung aufwändig und teuer macht. Mobile Ansätze wie Floating Car Daten sind zwar in der Lage, einzelne ausgestattete Fahrzeuge zu detektieren, ein Rückschluss auf das Gesamtverkehrsaufkommen ist hingegen schwierig zu ziehen. Aus diesem Grund nennt Kühnel (2012) die Weiterentwicklung der Floating Car Observer Technologie als wegweisend.

Wolf et al. (2008c) sehen Floating Car Observer als eine Möglichkeit, Daten über die Verkehrsdichte und Verkehrsgeschwindigkeiten für Straßennetze zu generieren. Die Autoren vergleichen unterschiedliche Technologien zur Detektion von Fahrzeugen im Straßenverkehr. Dazu gehören Infrarotsensoren, Laserscanner, Ultraschallsensoren sowie Videokameras. Wolf et al. (2008c) kommen zum Schluss, dass Laserscanner aufgrund der Anschaffungskosten nicht dazu geeignet sind, großflächige FCO-Kampagnen durchzuführen. Ultraschallsensoren sind hingegen sehr günstig in der Anschaffung, liefern jedoch nur unzureichend genaue Daten. Infrarotelemente können gut als Ergänzung für Kamerasysteme genutzt werden, um unter schwierigen Lichtbedingungen die Messungen zu unterstützen. Folglich implementieren die Autoren einen Floating Car Observer, der mit Infrarotsensoren und einer Vi-

deokamera ausgestattet ist. Die von den Autoren vorgestellte Software, die mit einem Mikrocontroller die angeschlossenen Sensoren steuert und auslöst, soll in zukünftigen Arbeiten evaluiert und optimiert werden (Wolf et al. 2008c).

Ein besonderer Fokus wird auf Feinstaub-basierte und Bild-basierte Floating Car Observer gelegt, die im Folgenden diskutiert werden.

3.8.1 Feinstaub-basierte Floating Car Observer

Erfasste Feinstaubwerte korrelieren positiv mit der Anzahl an Fahrzeugen im Umgebungsverkehr während einer Messung (Röger 2020). Aufgrund dieses festgestellten Zusammenhangs implementieren Röger et al. (2020a) einen Floating Car Observer auf der Basis von Feinstaubdaten. Dafür messen die Autoren die Emission von Feinstaub bei Verbrennungsprozessen, beim Bremsen, durch Reifenabrieb sowie durch Aufwirbelungseffekte auf der Fahrbahn. Ein vor dem Lüftungsgitter eines Testfahrzeuges angebrachter Sensor misst Feinstaubwerte. Dazu werden über eine Videokamera vor der Frontscheibe sowie durch ein GPS-Modul in der Fahrerkabine weitere Daten aufgezeichnet. Mithilfe der produzierten Videodatei wird die Präsenz und Absenz von vorausfahrenden Fahrzeugen während Testfahrten bestimmt. Die Sensor- und Verkehrsdaten werden mit den GPS-Daten zur zeitlichen und räumlichen Einordnung verschnitten. Die aus den sieben Testfahrten gewonnenen Daten werden mithilfe eines Random Forest Modells analysiert. Dabei fungieren die Feinstaubwerte als Prediktoren und die Boole'schen Verkehrsdaten (mit *TRUE* für Präsenz des Umgebungsverkehrs und *FALSE* für Absenz von vorausfahrenden Fahrzeugen) als vorhergesagte Variablen. Bei einem 70/30 Train/Test Split prognostiziert das Modell den Verkehrsstatus mit einer Genauigkeit von bis zu 99%. Die Autoren sehen Potential für die Nutzung von Feinstaubsensoren zur Implementierung eines Floating Car Observers, weisen jedoch auch auf die vergleichsweise kleine Datengrundlage sowie auf die Möglichkeit ungenauerer Ergebnisse durch die Heterogenität des Verkehrs und der Straßen hin (Röger et al. 2020a).

3.8.2 Objektdetektionsalgorithmen für bildbasierte Floating Car Observer

Mit der Veröffentlichung des *YOLOv2* (auch bekannt als *YOLO9000*) Algorithmus im Jahr 2017 werden mit *RetinaNet* (Lin et al. 2017) und dem Single Shot Detector (Liu et al. 2016) weitere Objektdetektionsalgorithmen publiziert, die die Präzision von *YOLOv2* übertreffen. Folglich veröffentlichten Redmon

et al. (2018) mit *YOLOv3* eine optimierte Version des *YOLO9000* Algorithmus. Der Fokus von *YOLOv3* liegt vorrangig auf der Präzision der Detektionen, während die Berechnungszeiten länger als beim Vorgänger ausfallen. Auf technischer Seite unterscheidet sich primär die Architektur. Während *YOLO9000* auf dem *darknet-19* Framework basiert mit 19 Ebenen und 11 zusätzlichen Ebenen für die Objektdetektion, setzt *YOLOv3* auf das *darknet-53* Framework. Dieses besteht aus 53 Ebenen, wobei *YOLOv3* zusätzlich noch auf 53 weitere Ebenen zugreift. Somit besteht die Architektur des *YOLOv3* Algorithmus aus 106 vollständig gefalteten Ebenen. Darüber hinaus macht *YOLOv3* drei Prognosen, jeweils beim Reduzieren der Eingabeebene um die Faktoren 32, 16 und 8. Diese Eigenschaften machen *YOLOv3* zwar langsamer als *YOLO9000*, jedoch auf der anderen Seite deutlich genauer als sein Vorgänger. Ein weiterer Vorteil, der gegenüber der zweiten Generation des *YOLO* Algorithmus geschaffen wird, ist die Detektionsgenauigkeit bei besonders kleinen Objekten. Das kommt unter anderem davon, dass *YOLOv3* zehn mal so viele Begrenzungsrahmen wie *YOLO9000* zur Objektdetektion aufspannt.

Im Gegensatz zu *YOLOv1*, *YOLO9000* und *YOLOv3* stammt die Weiterentwicklung *YOLOv4* nicht von den ursprünglichen Autoren *Joseph Redmon* und *Ali Farhadi*. Im Februar 2020 verkündet *Joe Redmon* über den Kurznachrichtendienst Twitter, dass er aufgrund der neuerlichen militärischen Anwendung des Algorithmus aus der Forschung im Bereich des computerbasierten Sehens aussteigt¹¹. Tatsächlich wird die Zieldetektion heute als eines der Kernthemen der Militärtechnik gehandelt, wobei auch die Objektdetektion aus Bildern und Videos eine große Rolle spielt (Wang et al. 2020b). So sind an der Entwicklung von *YOLOv4* *Alexey Bochkovskiy*, *Chien-Yao Wang* und *Hong-Yuan Mark Liao* beteiligt mit dem Ziel, einen schnellen und für parallele Berechnungen geeigneten Objektdetektionsalgorithmus zu implementieren. *YOLOv4* basiert wie *YOLOv3* auf dem *darknet-53* Framework. Im Gegensatz zur Vorgängerversion werden die Parameter aus den unterschiedlichen *Framework* und *Detektor* Ebenen nicht mit der *Feature Pyramid Network* (FPN) Methode aggregiert, sondern mit dem schnelleren *Path Aggregation Network* (PANet). Um das Training eines Detektors zu optimieren, kann auf ein *Bag of Freebies* zugegriffen werden. Dieses Feature enthält Funktionalitäten zur Optimierung des Trainingsvorganges, bei deren Anwendung kein vergrößerter Rechenaufwand zu erwarten ist. Des Weiteren präsentieren die Autoren ein *Bag of Specials*, das zusätzliche Funktionalitäten enthält, die bei leicht gesteigertem Rechenaufwand die Präzision der Detektionen signifikant erhöhen können. Zusammenfassend stellen Bo-

¹¹<https://twitter.com/pjreddie/status/1230524770350817280>, 06/21

chkovskiy et al. (2020) einen schnellen und präzisen Algorithmus bereit, der durch den mitgelieferten *Bag of Freebies* sowie *Bag of Specials* eine Optimierung für spezielle Anwendungsgebiete zulässt (Bochkovskiy et al. 2020).

Mahto et al. (2020) präsentieren eine auf Fahrzeugdetektion optimierte Implementation des *YOLOv4* Algorithmus. Jährlich erscheinen neue Algorithmen zur Objektidentifizierung in Bildern. Diese sind in der Regel auf Standard-Datensätzen wie *MS COCO*¹² trainiert. *MS COCO* ist ein Datensatz, der 91 verschiedene Objekte enthält. Diese sind in 2,5 Millionen Annotationen auf 328.000 verschiedenen Bildern enthalten. Der Datensatz enthält verschiedene Objektklassen, wie Automobile, Fahrräder oder Personen (Lin et al. 2014). Mahto et al. (2020) wollen in ihrer Arbeit den bereits existierenden *YOLOv4* Algorithmus so konfigurieren, dass er besser zur Fahrzeugidentifizierung geeignet ist als andere herkömmliche, unkonfigurierte Objektdetektoren. Dafür wenden die Autoren vier Techniken aus dem *Bag of Freebies/Specials* an, die der *YOLOv4* Detektor zur Konfiguration bereitstellt:

- die Optimierung der Ankerboxen mithilfe von *k-means Clustering*,
- die Nutzung der *distance-IoU* für die Non-Maximum Suppression (DIoU-NMS),
- die Nutzung des *Spatial Attention* Moduls (SAM),
- und eine Implementierung des *Self-Adversarial Trainings* (SAT)

Ankerboxen beinhalten die initialen Dimensionen der Begrenzungsrahmen um Objekte. Diese werden zur Laufzeit eines Detektionsalgorithmus angepasst, sodass sie als Ergebnis die wahren Dimensionen der detektierten Objekte zeigen. Diese Anpassung wird durch Feedback aus dem Objektdetektionsalgorithmus vorgenommen. Die *K-Means Clustering*-Methode kann dafür angewandt werden. Dafür werden *K* zufällige Boxen ausgewählt, repräsentiert durch ihre Zentroiden. Um den jeweils nächsten Zentroiden werden im Anschluss Cluster gebildet, die auf Basis eines Grenzwertes aktualisiert und optimiert werden. Der Grenzwert wird durch den *Intersection over Union*-Wert bestimmt, der die Überlappung einer Ankerbox mit den wahren Abmessungen eines Objektes, dividiert durch die nicht-überlappende Fläche, repräsentiert. Bei einem IoU-Wert von über 50% kann davon ausgegangen werden, dass eine Ankerbox die Abmessungen eines Objektes repräsentiert. Die Auswahl der Menge an genutzten Zentroiden (*K*-Wert) für das *K-Means*

¹²<https://cocodataset.org/>, aufgerufen 06/21

Clustering ist eine Entscheidung zwischen höheren IoU-Werten (höherer K-Wert) und schnelleren Berechnungszeiten (niedrigerer K-Wert). Mahto et al. (2020) definieren einen idealen Wert bei $K = 12$, bei dem sie einen mittleren IoU-Wert von 78.3% herausbekommen. Die *Non-Maximum Suppression* eliminiert Begrenzungsrahmen, die sich gegenseitig überlappen und dabei das selbe detektierte Objekt repräsentieren. In der Standardkonfiguration der *YOLOv4*-Implementation wird dafür das *Generalised Intersection over Union* (GIoU) Verfahren angewandt. Zheng et al. (2020b) schlagen vor, stattdessen die *Distance-based Intersection over Union* (DIoU) zu verwenden, da diese Methode bei zusammenhängenden Objekten bessere Ergebnisse erzielt als die herkömmliche Vorgehensweise. Zur Dekodierung und Enkodierung der Bildpunkte verwenden Mahto et al. (2020) das *Spatial Attention* Modul (SAM). Um die Überanpassung (*Overfitting*) des verwendeten Modells zu reduzieren, nutzen die Autoren darüber hinaus das *Self-Adversarial Training* (SAT).

Das konfigurierte *YOLOv4* Modell wird von Mahto et al. (2020) mit dem *UA-DETRAC Benchmark* Datensatz¹³ trainiert. Der Datensatz besteht aus 100 Videosequenzen, die in einer Auflösung von 960×540 aufgenommen sind und eine Bildwiederholungsrate von 25 *Frames per Second* (FPS) aufweisen (Wen et al. 2015). In einem Experiment werden die Ergebnisse von *YOLOv3*, *YOLOv4* sowie den genannten Modifikationen miteinander verglichen. Die Resultate zeigen, dass das modifizierte *YOLOv4*-Modell mit 67,7% durchschnittlicher Präzision um 10% genauer arbeitet als das unmodifizierte Modell. Des Weiteren unterstreichen die Autoren, dass *YOLOv4* signifikant schneller arbeitet als die vorherigen Implementationen des Algorithmus und dazu in der Lage ist, den untersuchten Datensatz in Echtzeit auf einer einzigen *Graphics Processing Unit* (GPU) zu analysieren (Mahto et al. 2020).

¹³<https://detrac-db.rit.albany.edu/>, 06/21

4 Ausgewählte Methoden zur Analyse der Qualitätsparameter von Sensordaten

Mit welchen Methoden werden Qualitätsparameter von Sensordaten verarbeitet und analysiert? Mehrere Fachbereiche stellen Lösungen dazu bereit. Das Spektrum reicht dabei von Methoden der angewandten Geoinformatik über statistische Modelle bis hin zum computerbasiertem Sehen.

Die Arbeit mit Sensordaten findet in der Geoinformatik häufig im *Monitoring* (Überwachung) statt. Hinter dem Term *Monitoring* verbirgt sich die systematische Observation von Parametern, die einem bestimmten Problem zugeordnet sind sowie deren Änderung mit der Zeit (SCEP 1970). Spellerberg (2005) definiert die Überwachung als *surveying* und *surveillance*, wobei *surveying* die Kollektion von qualitativen sowie quantitativen Daten adressiert und *surveying* diesem Prozess eine zeitliche Komponente verleiht. Nach Spellerberg (2005) lässt sich der Prozess des *Monitoring* in vier Gruppen unterteilen:

- das einfache Monitoring,
- das überwachte Monitoring,
- das stellvertretende Monitoring,
- und das integrierte Monitoring.

Das *einfache Monitoring* nimmt hierbei eine Variable an einem definierten Punkt über eine festgelegte Zeit auf. Das *überwachte Monitoring* sieht vor, umwelttechnische Gegebenheiten in betroffenen sowie nicht-betroffenen Regionen gleichermaßen zu überwachen. Das *stellvertretende Monitoring* sieht vor, den Datenbestand vorher nicht überwachter Gebiete durch neue Messungen zu komplettieren. *Integriertes Monitoring* nutzt bei der Messung von Variablen in Gebieten bereits vorhandene ökologische Informationen (Spellerberg 2005). Downes et al. (2002) wählen eine alternative Kategorisierung des *Monitorings*, die sich mehr auf die Zielstellung des Prozesses bezieht. Sie unterteilen die Überwachung in:

- das Umweltmonitoring,
- das Langzeitmonitoring und Referenzstandort-Monitoring,
- das Compliance-Monitoring,
- und das Impact-Monitoring.

Das *Umweltmonitoring* bezeichnet Vorgänge, in denen umwelttechnische Fragestellungen durch Messungen behandelt werden. *Langzeitmonitoring* beschreibt Messungen über längere Zeiträume, um die Dynamik natürlicher Prozesse abzubilden. Unter *Compliance-Monitoring* versteht man Messprozesse, die an Orten durchgeführt werden, an denen bestimmte Grenzwerte eingehalten werden müssen. Das betrifft beispielsweise Feinstaubwerte entlang definierter Straßen. *Impact-Monitoring* wird dann angewandt, wenn der menschliche Einfluss auf natürliche Prozesse bestimmt werden soll (Downes et al. 2002).

Bei der Nutzung von Methoden aus der Geoinformatik zum Umweltmonitoring werden in der Regel Technologien verschiedener Forschungsbereiche wie der Bodenkunde oder der Atmosphärenphysik integriert. Als Grundlage ist die Geoinformatik jedoch bestens geeignet, denn ihre Methoden bilden die Basis zur Messung und Dokumentation räumlicher Veränderungen in der Umwelt auf lokaler, regionaler und globaler Ebene. Die Geoinformatik bietet einige Anwendungsbeispiele zur Messung und Dokumentation umwelttechnischer Aspekte. Abbildung 14 zeigt eine Auswahl.

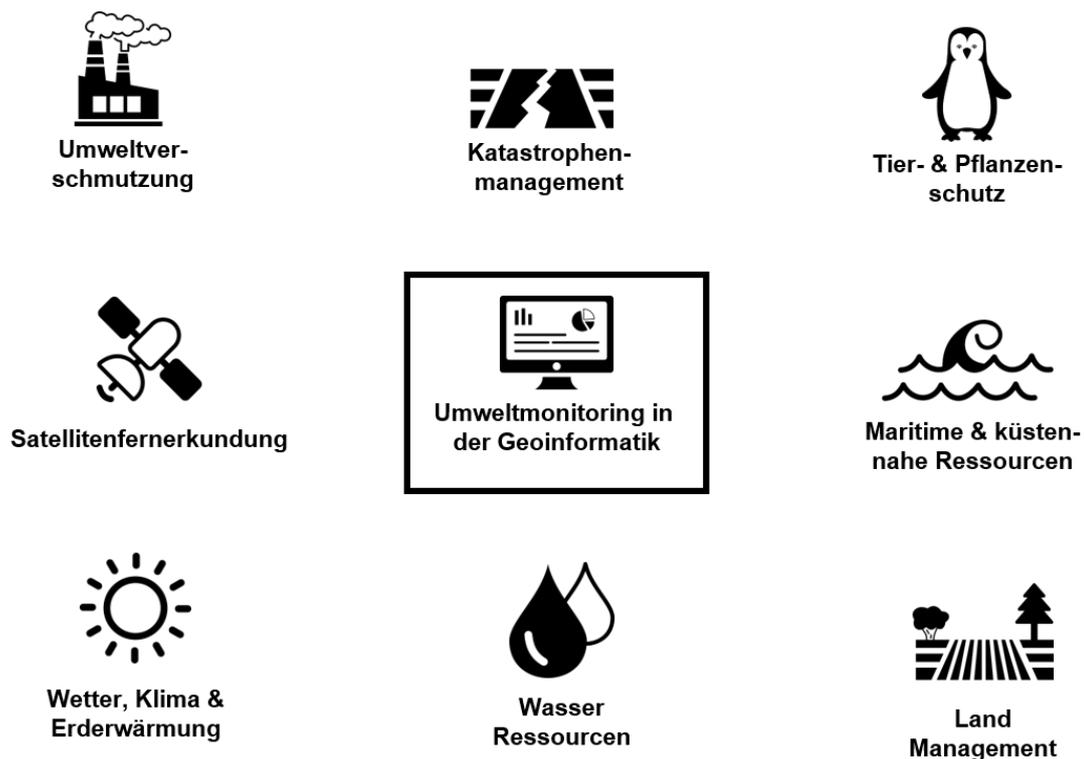


Abbildung 14: Anwendungsgebiete des Umweltmonitoring in der Geoinformatik

Abbildung 14 zeigt eine Auswahl an Anwendungsgebieten zum Monitoring umwelttechnischer Aspekte in der Geoinformatik. Als eins der wichtigen Elemente umwelttechnischer Dokumentationen und Messungen ist die Fer-

nerkundung zu sehen. Diese deckt eine große Bandbreite an Anwendungsfällen ab, in etwa durch Thermalbilder, Laservermessungen oder die Bereitstellung von Luftbildern. Messungen und Analysen zum Thema Wetter, Klima und Erderwärmung sind ebenfalls ein Anwendungsgebiet der Geoinformatik. Hier kommen in der Regel Sensoren zum Einsatz, die wetter- und klimatechnische Phänomene aufzeichnen. Durch Analysen lässt sich dann herausfinden, ob und inwiefern sich Wetter und Klima in definierten Regionen ändern. Das Anwendungsgebiet der Wasser-Ressourcen ist der Analyse von Wasserkörpern wie Flüssen oder Seen zugeordnet. Beispiele hierfür sind die Analyse der Veränderung von Flussläufen oder die Flächenveränderung von Seen. Das Landmanagement ist der Bodenpolitik zugeordnet. Es setzt sich zum Ziel, Bodengüte zu dokumentieren und landwirtschaftliche Prozesse sowie Planungen ökonomisch und nachhaltig zu gestalten. Im Management maritimer und küstennaher Ressourcen geht es um die Dokumentation sowie den Schutz von Flora, Fauna und weiteren natürlichen Ressourcen, die in Meeren oder in Küstennähe vorkommen. Der Schutz von Tieren und Fauna ist darüber hinaus auch auf dem Festland ein Anwendungsgebiet geoinformatischer Methoden. Im Katastrophenmanagement trägt die Geoinformatik durch die Prognose sowie die Dokumentation von Schäden durch Naturereignisse wie Erdbeben dazu bei, potentielle Gefahren schon früher zu erkennen und Hilfskräften im Fall einer Katastrophe die Arbeit durch Karten des betroffenen Gebietes zu erleichtern (De Lange 2013).

4.1 Methoden der Geostatistik zur Analyse von Sensordaten

Statistische Auswertungen sind ein wichtiger Bestandteil der Analyse von Geodaten. Dazu zählen auch Verkehrs- und Umweltdaten. Das Spektrum reicht von einfacher deskriptiver Statistik bis hin zu Korrelationsanalysen und komplexen Modellen des maschinellen Lernens. In diesem Subkapitel wird eine Auswahl der in der Sensordatenanalyse angewandten Methoden vorgestellt.

Die Statistik ist allgemein in drei Teilbereiche zu untergliedern: die *erkundende* (explorative) Statistik, die *schließende* (mathematische) Statistik und die *beschreibende* (deskriptive) Statistik. Die Suche nach bisher unbekanntem Mustern und Zusammenhängen in Datensätzen und die darauf basierende Schließung neuer Hypothesen gehören zur *explorativen* Statistik. Die *schließende* Statistik hat das Ziel, diese Hypothesen auf ihre allgemeine Gültigkeit hin zu untersuchen. Basis vieler Methoden und Algorithmen in der Geostatistik ist

die *deskriptive* Statistik. Diese sieht vor, empirische Daten (wie zum Beispiel Sensordaten) durch Kennzahlen, Tabellen und graphische Darstellungen zu beschreiben. Damit lässt sich leicht ein Überblick verschaffen, vor allem auch über größere Datensätze. Verwendete Grafiken in der *deskriptiven* Statistik sind *Linien-, Kreis- und Balkendiagramme*. Außerdem bieten *Boxplots* und *Streudiagramme* Möglichkeiten, Daten in der *beschreibenden* Statistik zu visualisieren. Zu den Kennzahlen in der deskriptiven Statistik gehören *Häufigkeits- und Lagemaße*. Diese beinhalten unter anderem den *Mittelwert*, den *Median*, die *Varianz*, die *Summe*, das *Maximum*, das *Minimum* und den *Modus* einer Datenreihe (Assenmacher 2013).

In vielen Fällen ist bei der Analyse empirischer Daten die Nutzung einfacher, beschreibender Methoden der Statistik nicht genug. Komplexere Ansätze helfen dabei, Zusammenhänge in Datensätzen besser zu verstehen und darüber hinaus Vorhersagen zu treffen. Teil der komplexen Ansätze in der Statistik sind unter anderem Korrelationsanalysen und Methoden des maschinellen Lernens.

4.1.1 Korrelationsanalyse zur Untersuchung von Zusammenhängen verschiedener Datenreihen

Korrelationsanalysen untersuchen den Zusammenhang zweier oder mehrerer Datenreihen. In der Geostatistik werden diese Verfahren angewandt, wenn der Einfluss einer Variable auf eine andere kalkuliert werden soll. Es existieren unterschiedliche Ansätze zur Untersuchung von Korrelationen. Die *Pearson-Korrelationsanalyse* ist einer davon. Dabei können zwei metrisch skalierte Variablen auf Zusammenhänge überprüft werden. Ist die Datenbasis ordinalskaliert, kann die *Spearman-Rangkorrelation* verwendet werden. Für die Nutzung einer *Pearson-Korrelationsanalyse* müssen die Daten normalverteilt sein, während man bei nicht-normalverteilten Daten die *Spearman-Rangkorrelation* berechnet (Artusi et al. 2002; Fahrmeir et al. 2016). Die *Pearson-Korrelationsanalyse* basiert auf der Kalkulation des Korrelationskoeffizienten r :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)\left(\sum_{i=1}^n (y_i - \bar{y})^2\right)}} \quad (2)$$

mit der Stichprobengröße n , den Werten der Variablen x und y am i -ten Punkt x_i y_i und den Mittelwerten der Variablen x und y am i -ten Punkt \bar{x}_i \bar{y}_i .

Der Korrelationskoeffizient r kann Werte zwischen -1 und $+1$ annehmen. Ist der r -Wert positiv, besteht eine positiv lineare Korrelation. Ist er kleiner als 0 , besteht ein negativer Zusammenhang. Je näher sich der Wert r an -1 oder $+1$ annähert, desto größer ist die Korrelation der untersuchten Datenreihen. Bei einem Wert von 0 kann nicht von einem Zusammenhang ausgegangen werden (Artusi et al. 2002; Fahrmeir et al. 2016).

Bei der Rangkorrelation nach *Spearman* wird ein Korrelationskoeffizient ρ berechnet:

$$\rho = 1 - \frac{6 * \sum_{i=1}^n (r_i - s_i)^2}{n^3 - n} \quad (3)$$

mit der Stichprobengröße n , dem Rangplatz des i -ten Punktes innerhalb der Variable X (r_i) und dem Rangplatz des i -ten Punktes innerhalb der Variable Y (s_i). Die Interpretation von ρ geschieht analog zur Interpretation des Pearson Korrelationskoeffizienten r . Das Ergebnis liegt zwischen -1 und $+1$. Bei einem Koeffizienten größer als 0 liegt ein positiver Zusammenhang vor, während eine negative Korrelation bei einem ρ -Wert von kleiner als 0 besteht (Artusi et al. 2002; Fahrmeir et al. 2016).

Um zu überprüfen, ob der berechnete Zusammenhang (*Pearson* oder *Spearman*) zweier Variablen signifikant ist, kann der p -Wert untersucht werden. Dabei geht man zunächst von einem Signifikanzniveau α aus. $\alpha = 0,05$ ist ein häufig verwendeter Wert dafür. Ist der resultierende p – Wert kleiner oder gleich α , kann von einer statistisch signifikanten Korrelation ausgegangen werden. Als statistisch nicht signifikant gilt die Korrelationsanalyse, wenn der p – Wert größer als das Signifikanzniveau α ist (Fahrmeir et al. 2016).

4.1.2 Methoden des maschinellen Lernens

Wie können Klassifikations- und Regressionsprobleme im Zusammenhang mit der Analyse von Sensordaten gelöst werden? Das maschinelle Lernen (ML) ist Teil des Gebietes der künstlichen Intelligenz (KI). Durch den Einsatz von ML sind Computer dazu in der Lage, komplexe Zusammenhänge zu verstehen und eigenständig Lösungen zu Problemen zu finden. Methoden des maschinellen Lernens liegen in der Form von Algorithmen vor. Datenbestände haben gewisse Gesetzmäßigkeiten und Muster, die von ML Algorithmen erkannt werden. Durch den Aufbau einer künstlichen Intelligenz, die durch Lernprozesse anhand der Analyse von Datensätzen entsteht, können neue Problemstellungen gelöst werden. Außerdem sind Algorithmen des maschinellen Lernens dazu in der Lage, auf Basis des Lernzuwachses beim Betracht-

ten von Datensätzen auch neue Datensätze zu verstehen und zu analysieren (Mohri et al. 2018).

Bei Methoden des maschinellen Lernens spielt auch die menschliche Komponente eine tragende Rolle. Algorithmen müssen angelernt werden. Dabei ist es entscheidend, Modelle mit einer Datenbasis zu versorgen. Außerdem müssen bei der Implementierung von ML-basierenden Modellen, je nach gewähltem Algorithmus, Einstellungen getätigt werden. Sinnvoll gewählte Modellparameter sind die Basis präziser Modelle (Mohri et al. 2018).

Es existieren unterschiedliche Arten von ML Algorithmen. Diese umfassen:

- das überwachte Lernen,
- das unüberwachte Lernen,
- das teilüberwachte Lernen,
- das bestärkte Lernen,
- sowie das aktive Lernen.

Beim *überwachten Lernen* (*supervised learning*) werden im Vorfeld Beispielmuster definiert. Ein Beispiel eines Algorithmus des maschinellen Lernens, der *überwachtes Lernen* vorsieht, ist die *lineare Regression*. Das *unüberwachte Lernen* (*unsupervised learning*) sieht keine vorher definierten Beispielmuster vor. Der Algorithmus selbst findet wiederkehrende Muster in den Datensätzen. Ein Beispiel für *unüberwachtes Lernen* sind Cluster-Algorithmen wie *DB-Scan* oder *K-Means*. Das *teilüberwachte Lernen* (*semi-supervised learning*) stellt eine Kombination aus *überwachtem* und *unüberwachtem* Lernen dar. Dabei werden zwar Basisdaten zum Lernen bereitgestellt, die Zielvariable ist allerdings noch nicht im Vorfeld bekannt. Ein Beispiel für *teilüberwachtes Lernen* ist die Nutzung von Objektidentifikationsalgorithmen bei der Analyse von Bilddateien. Das *bestärkte Lernen* (*reinforcement learning*) basiert auf dem Konzept von Bestrafungen und Belohnungen. Dabei findet das Modell selbst eine Strategie, ein Problem zu lösen. Je nach dem, ob der Algorithmus einer Problemlösung näher kommt oder sich weiter davon entfernt, erhält es entweder positive (Lob) oder negative (Bestrafung) Rückmeldungen. *Monte Carlo Simulationen* sind ein Beispiel für das *bestärkte Lernen*. *Aktives Lernen* (*active learning*) basiert auf Feedback von Nutzern der Algorithmen. Dabei gelingt es den Algorithmen zu lernen, welche Datenpunkte sie am optimalsten zum Anlernen verwenden sollten (Mohri et al. 2018).

Algorithmen des maschinellen Lernens sind entweder Klassifikatoren, Regressoren oder können beides. Bei der Regression wird eine Formel $y = \beta * x + a$ durch eine gegebene Punktwolke gesucht, die die Entwicklung der Werte der Observationen möglichst gut beschreibt. Auf Basis dieser Formel können dann Vorhersagen für bisher unbekannte Datenpunkte gemacht werden. Bei Klassifikationsproblemen werden Gruppenzugehörigkeiten bestimmt. Neue Punkte werden im Anschluss der Gruppe zugeordnet, die anhand der Vorhersage am wahrscheinlichsten ist. Manche Algorithmen, wie zum Beispiel *Random Forest*, beherrschen sowohl die Klassifikation als auch die Regression (Mohri et al. 2018).

Aus der Gesamtmenge aller Algorithmen des maschinellen Lernens werden einige Methoden besonders häufig für mit Umwelt- und Verkehrsdaten zusammenhängenden Problemstellungen genutzt. Dazu gehören die *lineare Regression*, *Gradient Boost* und *Random Forest*.

Lineare Regression

Die *lineare Regression* ist ein Verfahren, bei dem Werte einer Variable mithilfe von einer oder mehreren anderen Variablen vorhergesagt werden. Dabei unterscheidet man zwischen der abhängigen Variable (auch *Kriterium* genannt) und der/den unabhängigen Variablen (auch *Prädikatoren* genannt). Das *Kriterium* ist der Wert, der vorhergesagt werden soll. Die Variablen, die zur Vorhersage genutzt werden, nennen sich *Prädikatoren*. Bei einer Vorhersage mithilfe einer *linearen Regression* werden die Zusammenhänge der abhängigen und unabhängigen Variablen untersucht (Montgomery et al. 2021). Ihr Zusammenhang wird mit einer Regressionsgleichung beschrieben:

$$\hat{y}_k = bx_k + a \quad (4)$$

mit \hat{y}_k für das vorhergesagte *Kriterium*, y für den k -ten Wert, x_k für den *Prädikator* k an x -ter Stelle, b für die Steigung der Regressionsgeraden und a für den y -Achsen Abschnitt der Regressionsgeraden. Die Regressionsgleichung lässt sich auch graphisch darstellen. Das hilft dabei, optisch die Zusammenhänge der abhängigen und unabhängigen Variablen zu verstehen. Dabei klassifiziert man die Art des jeweiligen Zusammenhangs in zwei Gruppen. Ist die Regressionsgerade ansteigend, spricht man von einem positiven Zusammenhang der *Prädikatoren* mit dem *Kriterium*. Bei absteigender Regressionsgerade wird der Zusammenhang als negativ bezeichnet. Ferner unterscheidet man zwischen der einfachen und der multiplen linearen Regression. Die Nomen-

klatur kommt von der Anzahl der *Prädikatoren*. Bei einer unabhängigen Variable spricht man von einer einfachen linearen Regression, während bei zwei oder mehr *Prädikatoren* von einer multiplen linearen Regression die Rede ist (Montgomery et al. 2021).

Gradient Boost

Der *Gradient Boost*-Algorithmus (GB) basiert auf der Idee von Breiman (1997). Der Ansatz ist nutzbar für Klassifizierungs- und Regressionsprobleme. Er ist verwandt mit dem *AdaBoost*-Algorithmus. Im *AdaBoost*-Verfahren wird ein Entscheidungsbaum trainiert, innerhalb dessen jeder Observation initial die gleiche Gewichtung zugewiesen wird. Nach der Evaluierung des ersten Baumes, wird die Gewichtung der schwer klassifizierbaren Observationen erhöht und die Gewichtung der einfach klassifizierbaren Observationen vermindert. Im Anschluss wird der zweite Entscheidungsbaum unter Berücksichtigung der angepassten Gewichtungen gebaut. Das neue Modell besteht nun aus beiden Entscheidungsbäumen. Auf Basis des Klassifizierungsfehlers wird im Anschluss ein dritter Entscheidungsbaum aufgebaut. In mehreren Iterationen wird das Modell dann optimiert. Der Unterschied zwischen *Gradient Boost* und *AdaBoost* liegt in der Kalkulation der Fehlerwerte schwer klassifizierbarer Observationen. Der *Gradient Boost*-Algorithmus verwendet dafür die Verlustfunktion $y = ax + b + e$ mit dem Fehlerwert e (Friedman 2001). Vereinfacht lässt sich die Implementierung eines *Gradient Boost*-Modells wie folgt erklären:

1. Initialisierung eines vereinfachten Vorhersagemodells
2. Implementierung eines weiteren Modells auf Basis der Fehler der vorherigen Modells
3. Kombination der schwachen Modelle zu einem starken Modell
4. Wiederholung von Schritt 2 und 3, bis die gewünschte Genauigkeit erreicht ist

Der Vorteil in der Nutzung von *Gradient Boost* als Klassifikations- oder Regressionsalgorithmus liegt darin, dass die Daten in der Regel nicht vorprozessiert werden müssen. Außerdem zeichnen GB-Modelle eine hohe Flexibilität aus, da Kosten- und Verlustfunktionen durch den Nutzer selbst definierbar sind (Friedman 2001).

Random Forest

Eine weitere häufig bei der Analyse von Verkehrs- und Umweltdaten verwendete Methode ist *Random Forest*. Der von Breiman (2001) entwickelte Algorithmus ist sowohl für Klassifikations- als auch für Regressionsprobleme nutzbar. Er gehört zur Gruppe des überwachten Lernens. RF basiert auf einer Vielzahl von Entscheidungsbäumen, auch Wald (*forest*) oder *Ensemble* genannt. Jeder einzelne Baum liefert dem Algorithmus eine Lösung. Aus den verschiedenen Lösungen berechnet RF das wahrscheinlichste Ergebnis der Regressions- oder Klassifikationsaufgabe. Jeder einzelne Baum des Algorithmus besteht aus mehreren Verzweigungen. Diese entstehen, indem auf Grundlage der Datenbasis mit *ja* oder *nein* beantwortbare Fragen gestellt werden. Diese Entscheidungen erfolgen randomisiert. Das bedeutet, dass jedes RF-Modell auf der gleichen Datengrundlage zu unterschiedlichen Ergebnissen kommen kann. Das sogenannte *Bagging* kombiniert Einzelentscheidungen der Entscheidungsbäume. Einige Entscheidungen liegen dem Nutzer des Algorithmus offen, so in etwa die Anzahl der Entscheidungsbäume (*n_tree*) oder die Anzahl der zufällig ausgewählten Prädiktoren pro Baum (*mtry*). RF-Modelle bestehen in der Praxis häufig aus einem Trainingsdatensatz, einem Testdatensatz und in manchen Fällen aus einem Validierungsdatensatz. Die Ausgabe eines RF-Algorithmus ist stets die Klassifizierung beziehungsweise die Regression (Breiman 2001; Pavlov 2019).

Es stehen des Weiteren auch noch einige Parameter zur Überprüfung der Genauigkeit eines RF-Modells zur Verfügung:

- die Genauigkeit (*Accuracy*),
- die Sensitivität (*Sensitivity*),
- die Spezifität (*Specificity*),
- die ROC-Kurve (*Receiver Operating Characteristic Curve*),
- sowie die Fläche unter der Kurve (*Area Under the Curve*, *AUC*).

Die *Genauigkeit* gibt an, wie viel Prozent aller Punkte richtig klassifiziert sind. Die *Sensitivität* ist die *true positive rate*, gibt also an, wie viel Prozent der mit *TRUE* deklarierten Punkte auch als *TRUE* klassifiziert werden. Umgekehrt ist das bei der *Spezifität*. Diese gibt an, wie viel Prozent der mit *FALSE* deklarierten Datenpunkte korrekt erkannt werden. Die *ROC-Kurve* ist eine

Visualisierung der *Sensitivität* gegen die *Spezifität*. Die resultierende Linie verläuft bei einem akkuraten Modell in der Regel so weit links und oben wie möglich. Die *Fläche unter der Kurve* wird mit dem *AUC-Wert* beschrieben. Je höher dieser ist, desto akkurater ist das Modell. Hyperparameter-Tuning ist eine Möglichkeit, Modellvorhersagen mit *Random Forest* zu verbessern. Hierbei werden Modellparameter wie n_{tree} oder $mtry$ in unterschiedlichen Iterationen variiert, bis die optimalen Parameter für ein Modell gefunden werden (Breiman 2001; Pavlov 2019).

RF hat eine Reihe an Vor- und Nachteilen gegenüber anderen Algorithmen des maschinellen Lernens. Die Trainingszeit von Random Forest gilt als sehr gering. Ein weiterer Vorteil ist, dass RF auch mit großen Datensätzen und einer Vielzahl von Klassen gut zurecht kommt. Das Training eines RF-Modells ist gegenüber anderen Methoden leicht bewerkstelligbar. Des Weiteren hat RF eine Vielzahl an potentiellen Anwendungsfällen und wird in vielen Wissenschaftsbereichen gerne verwendet. Dahingegen gilt Random Forest als *Black Box*. Es ist in der Regel nicht nachzuvollziehen, warum ein Modell zu einem bestimmten Ergebnis kommt. Außerdem ist der Algorithmus nur bedingt für Echtzeitberechnungen verwendbar, denn im Gegensatz zum Training gilt die Klassifizierung neuer Datensätze als Zeitaufwendig (Breiman 2001; Pavlov 2019).

4.1.3 Hidden Markov Modelle in Mapmatching Algorithmen

Hidden Markov Modelle (HMMs) werden für das Mapmatching von Trajektorien benötigt. Sie bestehen aus *Markov-Ketten*, die als Graphen mit definierten Übergangswahrscheinlichkeiten zwischen verschiedenen Status definiert sind. Sieht man diese Status als Netzwerkknoten (beispielsweise i und j), dann wäre die Wahrscheinlichkeit, dass Knoten j auf Knoten i folgt, definiert als $p_{i,j}$. Vor der Implementierung einer *Markov-Kette* werden die jeweiligen Status $\{S_1, \dots, S_k\}$ festgelegt. Die Wahrscheinlichkeit, in einen neuen Status zu wechseln, wird allein vom aktuellen Status bestimmt und ist nicht von vorhergehenden Ereignissen beeinflusst. Ein *Hidden Markov Modell* besteht aus einer unsichtbaren *Markov-Kette*. Jeder Status einer Kette generiert eine von k Observationen, die sichtbar sind (Rabiner et al. 1986; Choudhury et al. 2007). Eine schematische Darstellung der Funktionsweise eines *Hidden Markov Modells* findet sich in Abbildung 15.

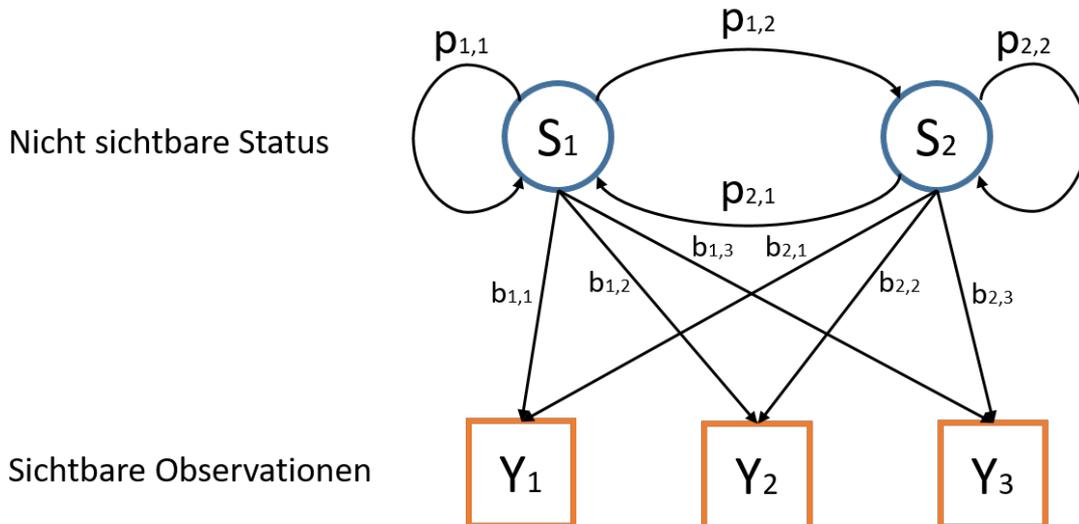


Abbildung 15: Schematische Darstellung eines *Hidden Markov Modells* für zwei Status (S_1, S_2) und drei Observationen (Y_1, Y_2, Y_3)

Abbildung 15 zeigt schematisch die Funktionsweise eines *Hidden Markov Modells* für zwei Status $\{S_1, S_2\}$ und drei mögliche Observationen $\{Y_1, Y_2, Y_3\}$. Die Status sind für den Benutzer nicht sichtbar. Während der Laufzeit kann das Modell seinen Status ändern oder den gleichen Status behalten mit der Wahrscheinlichkeit $p_{x,y}$. x steht hier für den aktuellen Status und y für den darauf folgenden Status. Jeder der beiden Status kalkuliert in jedem Schritt eine für den Nutzer sichtbare Observation. Die Wahrscheinlichkeit, dass eine Observation Y_z eintritt, wird vom Wert $b_{i,j}$ bestimmt (Rabiner et al. 1986; Choudhury et al. 2007).

Im Falle eines Mapmatching-Algorithmus liegen eine Reihe von Datenpunkten vor, für die die wahrscheinlichsten Status (in diesem Fall Positionen entlang eines definierten Netzwerkes) gesucht werden. Für etwaige Anwendungen ein einfaches *Hidden Markov Modell* zu verwenden, gilt als sehr rechenaufwendig. Der *Viterbi Algorithmus* (VA) löst dieses Problem durch einige Optimierungen eines HMM. Ein VA ist dafür konzipiert, den wahrscheinlichsten Pfad durch ein Netzwerk, beziehungsweise Gitter, zu finden. Jeder Knoten eines Pfades repräsentiert ein Status, während jede Kante des Netzwerkes eine Übergangsmöglichkeit in einen neuen Status bietet. Ein VA nutzt eine Reihe an Metriken, die von den Observationen und Übergangskanten abhängen. Darauf aufbauend wird der wahrscheinlichste Pfad durch das Netzwerk kalkuliert. Beim Iterieren durch die Status wird stets die Kante mit der größeren Metrik gewählt. Bei gleicher Gewichtung erfolgt der Übergang randomisiert. Auf diese Weise werden mehrere *survivor*-Pfade aufgebaut. Sobald der Algo-

rithmus terminiert, werden diese Pfade wieder anhand ihrer Metriken verglichen. Der Pfad sp_T mit der höchsten Metrik Γ_T wird als wahrscheinlichster Weg durch das Netzwerk definiert (Ryan et al. 1993).

4.1.4 Beschreibung und Vergleich von Trajektorien

Wie können unterschiedliche, korrigierte und nicht-korrigierte, Trajektorien miteinander verglichen werden? Datenpunkte werden durch Koordinaten lokalisiert. Bei der Analyse der Positionierungsgenauigkeit von Observationen kann auf Genauigkeitsmaße wie die Satellitenanzahl oder die Verschiebung der Präzision (DOP) zurückgegriffen werden. Diese werden in der Regel von GNSS-Modulen bei der Datenerfassung mit aufgenommen. Einige Arbeiten, wie Lyu et al. (2017), verwenden zur Beschreibung der Positionierungsgenauigkeit von Trajektorien darüber hinaus auch Maße der Kurvengleichheit. Das Prinzip sieht vor, die jeweilige aufgenommene Trajektorie mit einer Referenzkurve zu vergleichen. Diese kann zum Beispiel durch Mapmatching Algorithmen erzeugt werden und verläuft dabei entlang den Kanten eines Netzwerkdatensatzes. Methoden der Kurvengleichheit können unterschiedliche Vergleichsmaße bestimmen. Ansätze wie die *Discrete Fréchet* (DF) Distanz oder das *Partial Curve Mapping* (PCM) bestimmen die Distanz zwischen zwei Kurven. Andere Methoden sehen die Messung der Fläche zwischen zwei Kurven vor. Eine weitere Möglichkeit bietet der Vergleich der Trajektorienlängen.

Abbildung 16 zeigt das *Partial Curve Mapping*, die *Discrete Fréchet* Distanz sowie die *Fläche zwischen den Kurven* schematisch. Dabei wird eine Kurve A (blaue Linien) mit einer Kurve A' (orangene Linien) verglichen. Die jeweiligen zu den Kurven gehörigen Netzwerkknoten werden mit blauen beziehungsweise orangenen Punkten dargestellt. Die im oberen Teil der Abbildung dargestellte Methode ist das *Partial Curve Mapping*. Dabei kann der Versatz zweier Kurven festgestellt werden. In einem ersten Schritt werden zunächst beide Kurven A und A' zu den Kurven a und a' normalisiert. Dafür müssen die Begrenzungsrahmen von A und A' berechnet werden. Im Anschluss werden die Kurven- und Segmentlängen von a und a' bestimmt. Sollte die Länge von a größer ausfallen als die von a' , tauschen beide Kurven ihre Bezeichner. a muss beim *Partial Curve Mapping* immer länger sein als a' . Im nächsten Schritt wird jeder Knoten von a auf die Kurve a' gelegt. Dafür wird in einer ersten Iteration ein *offset*-Wert bestimmt, der die Distanz eines Punktes auf a zu seinem zugehörigen neuen Punkt auf a' beschreibt. Wichtig dabei ist, dass die Distanz zwischen den neuen Koordinaten der Distanz zwischen den Knoten in a entspricht. Im Anschluss wird der sogenannte *Mismatch-Error* kalkuliert.

Partial Curve Mapping (PCM)



Discrete Fréchet Distanz (DF)



Fläche zwischen den Kurven

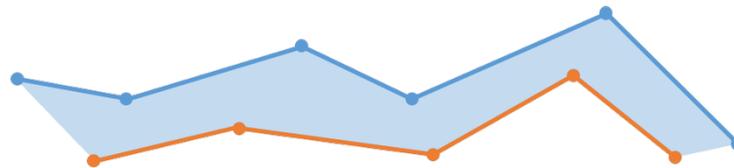


Abbildung 16: Verschiedene Ansätze zum Vergleich von Kurven: *Partial Curve Mapping* (oben), *Discrete Fréchet* (mitte) und die *Fläche zwischen den Kurven* (unten)

Dafür wird die Distanz d der Knoten auf a zu ihren jeweiligen neu erstellten Pendants auf a' berechnet. Danach wird die Volumenkomponente v_i jedes Segmentes i kalkuliert:

$$v_i = \frac{d_i + d_{i-1}}{2} * \tilde{S}_i; v_1 = 0; i = 2, 3, \dots, m \quad (5)$$

mit den Segmentlängen von a (\tilde{S}) und der Gesamtzahl der Knoten m . Die Summe aller v_i Werte ist die Diskrepanz ϵ . In weiteren Iterationen des Algorithmus wird der *offset*-Wert schrittweise variiert. Der final berechnete Versatz Wert, der die Gleichheit beider Kurven A und A' beschreibt, ist der kleinstmögliche Wert von ϵ (Witowski et al. 2011).

Die mittlere Darstellung in Abbildung 16 zeigt das Prinzip der *Discrete Fréchet* Distanz, die von Fréchet (1906) entwickelt wurde. Sie wird weitläufig folgendermaßen definiert: Eine Kurve repräsentiert die Laufstrecke eines Hundes, die andere Kurve die seines Halters. Der Hund ist an einer Leine. Beide Akteure können ihre Laufgeschwindigkeit beliebig anpassen, allerdings ist rückwärts laufen nicht erlaubt. Die *Discrete Fréchet* Distanz repräsentiert die Länge der kürzest möglichen Leine. Einen besonderen Fokus legt die Methode auf

die Verortung und die Reihenfolge der Punkte. Gegeben seien zwei Kurven $f : [a, b] \rightarrow V$ und $g : [a', b'] \rightarrow V$. Die *Discrete Fréchet* Distanz $\delta_F(f, g)$ wird demnach folgendermaßen kalkuliert:

$$\delta_F(f, g) = \inf_{\alpha, \beta} \max_{t \in [0, 1]} d(f(\alpha(t)), g(\beta(t))) \quad (6)$$

mit einer beliebigen kontinuierlichen, nicht abnehmenden Funktion von $[0, 1]$ nach $[a, b]$ (α), beziehungsweise einer beliebigen kontinuierlichen, nicht abnehmenden Funktion von $[0, 1]$ nach $[a', b']$ (β) (Eiter et al. 1994).

Bei der Berechnung der *Fläche zwischen zwei Kurven*, die schematisch im unteren Teil von Abbildung 16 dargestellt ist, wird der Raum zwischen zwei Kurven in viereckige Elemente aufgeteilt. Dafür muss zunächst sicher gestellt werden, dass beide verglichenen Kurven die selbe Anzahl an Knoten besitzen. Ist das nicht der Fall, werden der Kurve mit weniger Ecken so viele Knoten hinzugefügt, bis die Anzahl gleich ist. Im Anschluss werden die Knoten beider Kurven miteinander verbunden. Ein zur Kalkulation der Fläche generiertes Viereck besteht aus zwei Verbindungsstücken, einer Kante der ersten Kurve und einer Kante der zweiten Kurve. Der Flächeninhalt A eines unregelmäßigen Vierecks berechnet sich folgendermaßen:

$$F = \frac{1}{2} * d_1 * d_2 * \sin \alpha \quad (7)$$

mit den Diagonalen d_1 und d_2 des Vierecks und dem Winkel zwischen den Diagonalen α . Die *Fläche zwischen den Kurven* entspricht der Summe aller Flächeninhalte der Vierecke (Jekel et al. 2019).

4.2 Computerbasiertes Sehen (CV) in der Verkehrsdatenanalyse

Computerbasiertes Sehen kommt in der Dissertation zum Einsatz, um Verkehrsinformationen aus dem erfassten Bildmaterial zu extrahieren. Bilder bestehen aus Gitterzellen, die Farbinformationen beinhalten. Das menschliche Auge ist dazu in der Lage, die in einem Bild gespeicherten Informationen zu verarbeiten und zu interpretieren. Dabei werden bestimmte Farbmuster und die daraus entstehenden Formen (Linien, Punkte, Grenzen) mit bereits bekannten Informationen verglichen. Dieser Ablauf macht es dem menschlichen Gehirn möglich, Objekte auf Bildern richtig zuzuordnen (Klette 2014; Learned-Miller 2011; Zhang 2010).

Das Ziel des computerbasierenden Sehens ist es, Algorithmen zu befähigen,

ebenfalls Bilder interpretieren zu können. Dafür müssen Computer lernen, bestimmte Muster in Gitterzellen und deren Nachbarschaft bestimmten Merkmalen eines Bildes zuzuordnen. Einige populäre Ansätze, wie zum Beispiel *gefaltete neuronale Netze* (CNN) orientieren sich dabei sogar an der Architektur des menschlichen Gehirnes (Khan et al. 2018; Zhang 2010).

Computerbasiertes Sehen kommt auch in der Verkehrsdatengenerierung zum Einsatz. Verkehrsvideos und -bilder sind dazu in der Lage, ein vollumfängliches Bild des aktuellen Verkehrsgeschehens aufzuzeichnen. Andere Ansätze wie Radarmessungen oder induktive Schleifen können dagegen nur eine limitierte Anzahl an Parametern aufzeichnen. Da die manuelle Auswertung als zeitaufwändig gilt, beschäftigen sich einige Arbeiten mit der automatischen Extraktion von Informationen aus Verkehrsbildern und -videos mithilfe von Objektdetektionsalgorithmen aus dem computerbasierenden Sehen (Kun et al. 2009).

4.2.1 Gefaltete Neuronale Netzwerke (CNN) in CV Anwendungen

Gefaltete neuronale Netzwerke (CNN) orientieren sich an biologischen Prozessen. Sie finden Anwendung in der Analyse von Audio- und Bilddateien. So ist es mithilfe von CNNs beispielsweise möglich, Objekte innerhalb von Bildern rechnerisch zu bestimmen und zu verorten. CNNs gelten als ein Element des maschinellen Lernens (ML) und zählen als ein Beispiel für künstliche Intelligenz (KI). CNNs akzeptieren unterschiedliche Eingabedaten. Diese können in Form einer Matrix (zum Beispiel Bilder mit Breite, Höhe und Farbkäna len) oder in Form eines Vektors vorliegen. Wird ein Bild als Eingabedatei in Form eines Vektors gewählt, müssen die einzelnen Bildpunkte in Form einer Kette angereiht werden (*Flattening*). Die Art des CNN-Algorithmus gibt an, in welcher Form Eingabedateien vorliegen müssen. Eine Datei durchläuft innerhalb eines CNN mehrere *Convolutional Layer* (Filter) sowie *Pooling Layer* (Aggregations-Schichten), die abwechselnd durchlaufen werden. Die Anzahl an Filtern und Aggregations-Schichten wird durch den Algorithmus festgelegt. Am Ende des Algorithmus steht ein *Dense-* beziehungsweise *Fully Connected-Layer*, der aus vollständig verbundenen Neuronen besteht. Der schematische Aufbau eines gefalteten neuronalen Netzwerkes wird in Abbildung 17 gezeigt (Albawi et al. 2017).

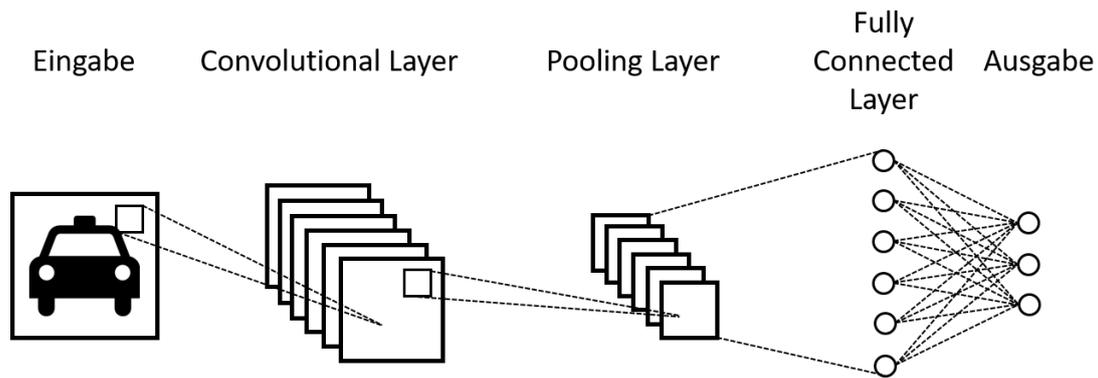


Abbildung 17: Schematischer Aufbau eines gefalteten neuronalen Netzwerkes. Einer Eingabeebene werden im *Convolutional Layer* Filter angewandt. Im *Pooling Layer* werden die Ergebnisse aggregiert. Nach beliebig vielen *Convolutional*- und *Pooling*-Ebenen folgt ein *Fully Connected*-Layer. Die Ausgabebene legt die finalen Klassen fest.

Abbildung 17 zeigt den schematischen Aufbau eines CNN mit einer Eingabeebene, einem *Convolutional Layer*, einem *Pooling Layer*, einem *Fully Connected Layer* und einer Ausgabebene. Die *Convolutional*-Schicht faltet Eingabedateien mithilfe von Filtern. Diese erkennen in Bildern einfache Merkmale wie Formen, Linien und Kanten. Filter haben eine vom Algorithmus festgelegte Größe, in etwa 2×2 Pixel. Die Filter springen als bewegliche Fenster von links nach rechts und anschließend eine Ebene tiefer auf der Eingabeebene. Aus dem Fensterinhalt wird so eine Ergebnismatrix berechnet. Die vier Pixel innerhalb eines Fensters werden damit auf einen Wert reduziert. Da beispielsweise ein 2×2 Fenster jeweils vier Pixel einer Eingabe mit dem Filter verbindet, spricht man von *gefalteten* neuronalen Netzwerken (Albawi et al. 2017). *Pooling Layer* geben nun die stärksten Signale der Filter weiter und aggregieren somit die Ergebnisse der *Convolutional Layer*. Damit werden nur die relevantesten Informationen eines Eingabebildes an die nächste Ebene innerhalb des CNN weitergegeben. Möglich ist das *Maximal-Pooling* sowie das *Mittelwert-Pooling*. Da in diesem Schritt redundante Informationen eines Bildes entfernt werden, erhöht jeder Schritt des *Pooling* die Berechnungsgeschwindigkeit des neuronalen Netzes. Ein CNN besteht je nach Konfiguration aus mehreren Filtern sowie *Pooling Layern*. Ist die Größe der ursprünglichen Eingabe maximal reduziert bei größtmöglichstem Informationsgewinn, folgen mehrere oder ein *Fully Connected Layer*, bei denen alle Neuronen der vorgelegerten Schichten mit jedem Ausgabemerkmale verbunden sind. Der darauf folgende Ausgabebereich besitzt genau die gleiche Anzahl an Neuronen, wie es Ausgabeklassen gibt (Albawi et al. 2017). Die Objektidentifikation ist ein weit verbreitetes Problem im Feld des computerbasierten Sehens. Als eine Anwendung der gefalteten neuronalen

Netzwerke sind Objektidentifikationsmethoden dazu in der Lage, Bilder zu klassifizieren und Objekte darin zu erkennen sowie zu lokalisieren. Man unterteilt die dazugehörigen Algorithmen in zwei Klassen, *Klassifikationsalgorithmen* und *Regressionsalgorithmen*. Klassifikationsalgorithmen selektieren zunächst Bildregionen mit einer hohen Wahrscheinlichkeit, dass sich in dieser ein Objekt befindet. Danach werden diese Regionen mit einem CNN klassifiziert. Beispiele für Klassifikationsalgorithmen sind *Region-based Convolutional Neural Networks* (RCNN), *Fast-CNN*, *Faster-CNN* sowie *Mask-CNN*. Regressionsalgorithmen prognostizieren Objektklassen und deren Begrenzungsrahmen innerhalb eines ganzen Bildes in einem Durchlauf des Algorithmus. Beispiele hierfür sind *You Only Look Once* (YOLO) sowie der *Single Shot Multi-box Detector* (SSD). Regressionsalgorithmen gelten als schneller, verglichen mit klassifikationsbasierten Objektidentifikationsalgorithmen (Morera et al. 2020).

4.2.2 You Only Look Once (YOLO) zur Objektdetektion in Bildern

Der *You Only Look Once* Algorithmus ist dazu in der Lage, in einem Durchlauf mehrere Objekte auf einem Eingabebild zu identifizieren. Das unterscheidet ihn von anderen, traditionellen, CNN Ansätzen. Der *YOLO*-Algorithmus prognostiziert Objekte sowie deren Begrenzungsrahmen. Dabei wird ein Modell mit vorher annotierten Bildern antrainiert, dass im Anschluss in Echtzeit Objekte auf neuen, unbekanntem, Bildern identifizieren kann. Die Objektdetektion mithilfe des *YOLO*-Algorithmus folgt einem definierten Schema. Die dargestellten Dimensionsgrößen und Parameter können sich je nach Wahl der Hyperparameter oder der gewählten *YOLO*-Version unterscheiden. Für die Identifizierung der Objekte erwartet der Algorithmus ein Eingabebild mit der Dimension $416,416,3$. Dieses Bild wird im Anschluss an ein gefaltetes neuronales Netzwerk übergeben. Dabei wird die Eingabeebene auf ein Raster der Dimension $19,19,425$ reduziert. Dann werden jeder Zelle einer Ebene der Dimension $19,19,425$ Werte zugewiesen. Jede Zelle besteht aus fünf Ankerboxen, die jeweils 80 Ausgabeklassen bestimmen können. Das Ergebnis dieses Ablaufes sind Begrenzungsrahmen y , die folgende fünf Parameter besitzen (Redmon et al. 2018; Bochkovskiy et al. 2020):

1. der Mittelpunkt des Begrenzungsrahmens,
2. die Breite des Begrenzungsrahmens,
3. die Höhe des Begrenzungsrahmens,

4. die Objektklasse c ,
5. und die Wahrscheinlichkeit pc .

Dabei verorten der *Mittelpunkt*, die *Breite* und die *Höhe* das Objekt innerhalb eines Bildes. Die Objektklasse c gibt an, um welche Art von Objekt es sich bei der Prognose handelt. Der Wert pc zeigt, mit welcher Wahrscheinlichkeit es sich dabei um ein Objekt c handelt. Aufgrund der Menge an generierten Ankerboxen kann es zu Überlappungen und Mehrfachdetektionen von identischen Objekten kommen. Deshalb werden im letzten Schritt das *Intersection over Union* (IoU) und das *Non-Max Suppression* Verfahren angewandt. Für die IoU wird der generierte Begrenzungsrahmen sowie die antrainierte Position des Begrenzungsrahmens benötigt. Im Anschluss wird die überlappende Fläche beider Rahmen durch die Vereinigungsfläche (*Union*) geteilt. Das *Non-Max Suppression* Verfahren sieht hingegen vor, das Problem überlappender Begrenzungsrahmen durch den Vergleich ihrer pc -Werte zu lösen. Dabei wird zunächst der Rahmen mit der höchsten Wahrscheinlichkeit pc identifiziert. Alle anderen Begrenzungsrahmen, die sowohl (1) einen niedrigeren pc -Wert vorweisen und (2) sich mit dem wahrscheinlichsten Rahmen überlappen, werden verworfen (Redmon et al. 2018; Bochkovskiy et al. 2020).

5 Auswahl der Qualitätsparameter

Welche der vorgestellten Qualitätsparameter werden im Zuge der Fallstudie dieser Arbeit analysiert?

Die Auswahl der Faktoren, deren Einfluss experimentell untersucht wird, ist eine zentrale Fragestellung der Arbeit. Die Selektion basiert auf den Erkenntnissen vorangegangener Forschung (Röger 2020; Röger et al. 2020a; Röger 2021). In Abbildung 18 ist eine Auswahl verschiedener Faktoren dargestellt, die mithilfe einer experimentellen Analyse untersucht werden.

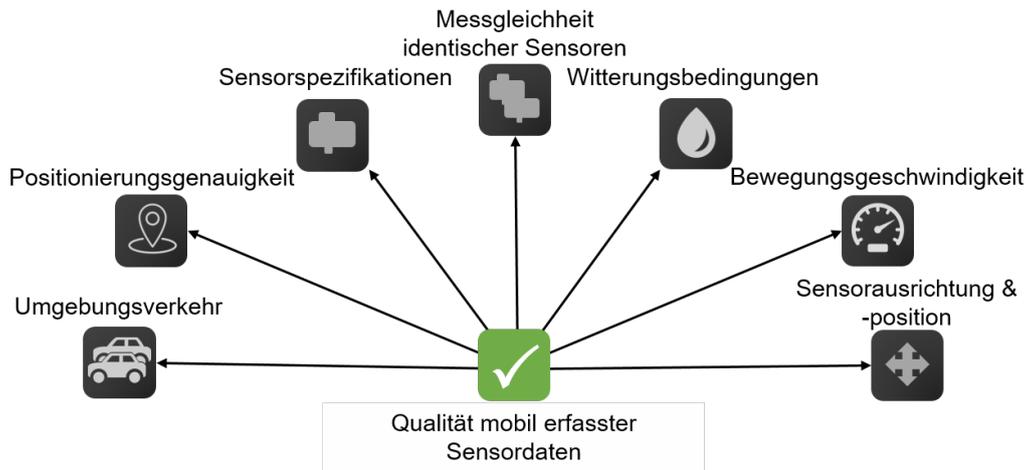


Abbildung 18: Einflussfaktoren auf die Datenqualität mobil erfasster Feinstaubsensordaten

Abbildung 18 zeigt die Faktoren, die im Zuge der Fallstudie untersucht werden. Zunächst werden dabei die Messunterschiede unterschiedlicher und identischer Sensoren analysiert. In beiden Fällen soll dabei beurteilt werden, inwiefern verschiedene Messeinheiten unter ähnlichen oder gleichen Umgebungsbedingungen vergleichbare Messwerte liefern. Da sich die Analyse beider Faktoren ähnelt, werden diese zu einem Faktor *Messunterschiede* zusammengefasst. Sensoren zur Positionsbestimmung liefern einen weiteren Unsicherheitsfaktor, der die allgemeine Messqualität beeinflussen kann. Deshalb wird evaluiert, wie stark der Einfluss der *Positionierungsgenauigkeit* ausfällt. Da auch äußere Faktoren wie die Luftfeuchtigkeit einen Einfluss auf die Messdatenqualität nehmen können, werden ebenfalls die *Witterungsbedingungen* während einer Messung analysiert. Ein weiterer untersuchter Faktor ist die *Sensorausrichtung*, bei dem analysiert wird, ob sich die Qualität der Messdaten bei unterschiedlich ausgerichteten Sensoren unterscheidet. Bei der Analyse der *Bewegungsgeschwindigkeit* wird untersucht, ob sich die Messdatenqualität bei verschiedenen Fahrgeschwindigkeiten unterscheidet. Bei Messungen auf Fahrzeugen ist nicht nur die Bewegungsgeschwindigkeit sehr variabel, son-

dern auch der Einfluss weiterer Feinstaubquellen. Diese treten beispielsweise in Form anderer Fahrzeuge auf, die sich ebenfalls auf den Straßen fortbewegen. Deshalb wird ebenfalls der Einfluss des *Umgebungsverkehrs* während einer mobilen Feinstaubmessung untersucht.

6 Experimentelle Analyse von Einflüssen auf die Messgenauigkeit bei der mobilen Erfassung von Feinstaubsensordaten

Eine Fallstudie soll über statische und dynamische Einflüsse auf mobil erfasste Feinstaubsensordaten Aufschluss geben. Dafür werden zunächst ein Studiengebiet definiert und verwendete Sensoraufbauten beschrieben. Es folgt die Erfassung mehrerer Datensätze, die im Anschluss analysiert werden. Zuletzt werden die Erkenntnisse aus der Fallstudie zusammengefasst.

6.1 Sensoren und Testgebiet der Fallstudie

Die Einflussfaktoren auf die Qualität mobiler Feinstaubsensordaten werden durch mehrere Experimente untersucht. Dafür stehen diverse Feinstaubmessgeräte zur Verfügung, die im Folgenden vorgestellt werden. Darüber hinaus wird ein Überblick über das Testgebiet der Fallstudie gegeben.

6.1.1 Verwendete Sensoren zur Datenerfassung

Es kommen im Experiment mehrere Sensoren zum Einsatz. Diese werden zur Bestimmung des Feinstaubgehaltes in der Luft, zur Positionierung und zur Temperatur- und Luftfeuchtigkeitsmessung eingesetzt. Zur Verortung der erfassten Datenpunkte wird der *NEO-6M* Sensor genutzt, der in Abbildung 19 gezeigt wird.

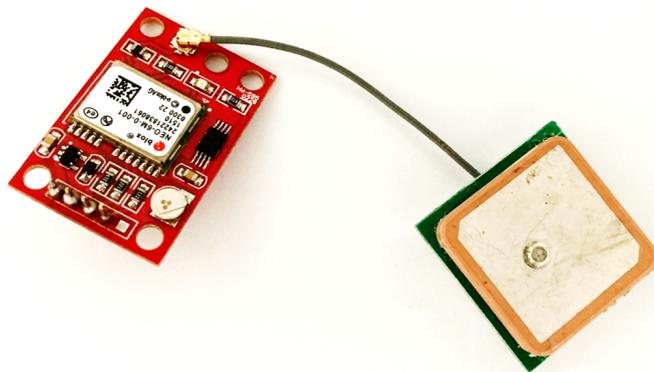


Abbildung 19: *NEO-6M*-Sensor zur Positionsbestimmung

Abbildung 19 zeigt das GNSS-Modul *NEO-6M* zur Positionsbestimmung mit der verwendeten Antenne. Der Sensor hat eine Positionierungsgenauigkeit von bis zu 2,5 Metern und erfordert zur Inbetriebnahme eine Stromquelle

mit fünf Volt Gleichstrom¹⁴. Die Position wird vom Sensor mit einer Frequenz von einem Hertz selbst kalkuliert, und kann anschließend über eine serielle Verbindung mit einem Mikrocontroller wie dem *Arduino Uno* ausgelesen werden. Zur Messung der Temperatur und der Luftfeuchtigkeit in der Umgebungsluft kommt der *DHT22*-Sensor zum Einsatz. Dieser wird in Abbildung 20 gezeigt.

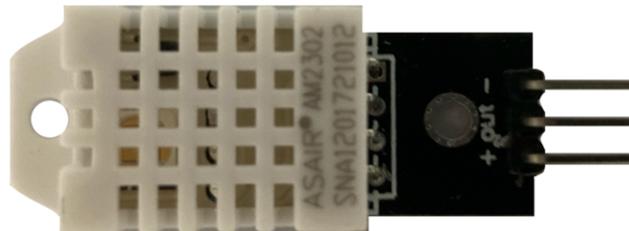


Abbildung 20: *DHT22*-Sensor zur Messung der Luftfeuchtigkeit und der Temperatur

Abbildung 20 zeigt den verwendeten *DHT22*-Sensor¹⁵ zur Bestimmung von Temperatur und Luftfeuchtigkeit in der Umgebungsluft. Wie in der Abbildung zu sehen ist, hat der Sensor je einen Anschluss für Plus (*VCC*) und Minus (*GND*) für die Stromversorgung. Außerdem verfügt er über einen digitalen Pin, um die Daten auszulesen (*DAT*). Die Spezifikationen des Sensors werden in Tabelle 7 gezeigt.

Bezeichnung	<i>DHT22</i>
Spannung	3.3 - 6 Volt Gleichstrom
Ausgabesignal	Digital (Single Bus)
Messspannweite Luftfeuchtigkeit	0-100%
Messspannweite Temperatur	-40 - 80 Grad Celsius
Genauigkeit Luftfeuchtigkeit	+ - 2 %
Genauigkeit Temperatur	+ - 0,5 Grad Celsius
Erfassungszeitraum	2 Sekunden

Tabelle 7: Spezifikationen des *DHT22*-Sensors

Tabelle 7 zeigt die Spezifikationen des *DHT22*-Sensors. Durch die variabel anlegbare Spannung von 3,3 bis 6 Volt ist der Sensor gut für die Nutzung mit einem Mikrocontroller wie dem *Arduino Uno* geeignet. Die Luftfeuchtigkeit wird in einer Spanne von 0 bis 100 Prozent angegeben, was das gesamte mögliche Spektrum an Werten abdeckt. Die möglichen Ungenauigkeiten von bis zu 2 Prozent (Luftfeuchtigkeit) sowie 0,5 Grad Celsius (Temperatur) müssen bei der Analyse und Auswertung der Daten beachtet werden.

¹⁴https://www.u-blox.com/sites/default/files/products/documents/NEO-6_DataSheet_%28GPS.G6-HW-09005%29.pdf, 05/21

¹⁵<https://www.electroschematics.com/am2302-dht22-datasheet/>, 06/21

Zur Messung von Feinstaubwerten kommen die zwei Sensoren *Plantower PMS3003* sowie *Alphasense OPC-N3* zum Einsatz.



Abbildung 21: *Plantower PMS3003* Sensor (links) sowie *Alphasense OPC-N3* Sensor (rechts)

Abbildung 21 zeigt die verwendeten Sensoren *Plantower PMS3003* (links im Bild, rund zehn Euro pro Einheit) und *Alphasense OPC-N3* (rechts in der Abbildung, rund 450 Euro pro Einheit). Beide Messgeräte gehören der kostengünstigen Kategorie an, wobei der *Alphasense OPC-N3* Sensor deutlich genauer arbeitet als der *Plantower PMS3003*. Die Sensoren unterscheiden sich anhand ihrer Spezifikationen.

Tabelle 8 zeigt die Spezifikationen der Sensoren *Plantower PMS3003*¹⁶ sowie *Alphasense OPC-N3*¹⁷. Die Sensoren haben grundsätzlich unterschiedliche Anschaffungspreise. Dabei kostet der *Alphasense OPC-N3* etwa 13-mal so viel wie der *Plantower PMS3003*. Anhand des Preisunterschiedes lässt sich erahnen, dass die Spezifikationen des *Alphasense OPC-N3* denen des *Plantower PMS3003* überlegen sind. Während der *Plantower PMS3003* Sensor die Werte PM_1 , $PM_{2.5}$ sowie PM_{10} aufzeichnet, lässt sich mit dem *Alphasense OPC-N3* optional $PM_{4.25}$ detektieren. Darüber hinaus speichert der *Alphasense OPC-N3*

¹⁶<https://aqicn.org/sensor/pms3003/de/,06/21>

¹⁷<https://www.alphasense.com/wp-content/uploads/2019/03/OPC-N3.pdf,06/21>

Bezeichnung	Plantower PMS3003	Alphasense OPC-N3
Preis	30 Euro	400 Euro
Spannung	5 Volt	4.8 - 5.2 Volt
Ausgabesignal	SPI	SPI, Mikro SD
Messwerte	$PM_1, PM_{2.5}, PM_{10}$	$PM_1, PM_{2.5}, PM_{4.25}, PM_{10}$
Messspannweite	0-1000 $\mu\text{g}/\text{m}^3$	0.01-2000 $\mu\text{g}/\text{m}^3$
Temperaturspanne	-10 bis 60 °C	-10 bis 50 °C
Luftfeuchtigkeitsspanne	0 bis 99 Prozent	0 bis 95 Prozent

Tabelle 8: Spezifikationen des *Plantower PMS3003* Sensors und des *Alphasense OPC-N3* Sensors

für jede einzelne Messung ein Histogramm ab, das Informationen über die absolute Anzahl von Feinstaubpartikeln in verschiedenen Größenordnungen enthält (bin_0, \dots, bin_{23}). Unterschiede gibt es ebenfalls bei der Messspannweite, bei der der *Alphasense OPC-N3* mit $0,1 - 2000 \mu\text{g}/\text{m}^3$ eine doppelt so große Spannweite aufweist als der *Plantower PMS3003*. Die Temperatur sowie die Luftfeuchtigkeit, bei denen die Sensoren arbeiten können, unterscheidet sich marginal. Allerdings ist festzuhalten, dass die genannten Luftfeuchtigkeiten nur gültig sind bei nicht-kondensierender Luft, das heißt wenn es nicht regnet.

6.1.2 Räumliche Abgrenzung des Testgebietes

Die Region Augsburg fungiert als Testgebiet für die Untersuchung der Einflussfaktoren auf die Qualität mobil erfasster Feinstaubsensordaten. Das Gebiet befindet sich im Süden Deutschlands. Die Stadt Augsburg umfasst 296.582 Einwohner über eine Fläche von 146 Quadratkilometern. Zum Ballungsraum Augsburg, zu dem des Weiteren der Landkreis Augsburg sowie der Landkreis Aichach-Friedberg gehören, zählen 685.000 Einwohner. Abbildung 22 zeigt die räumliche Ausdehnung des Testgebietes.

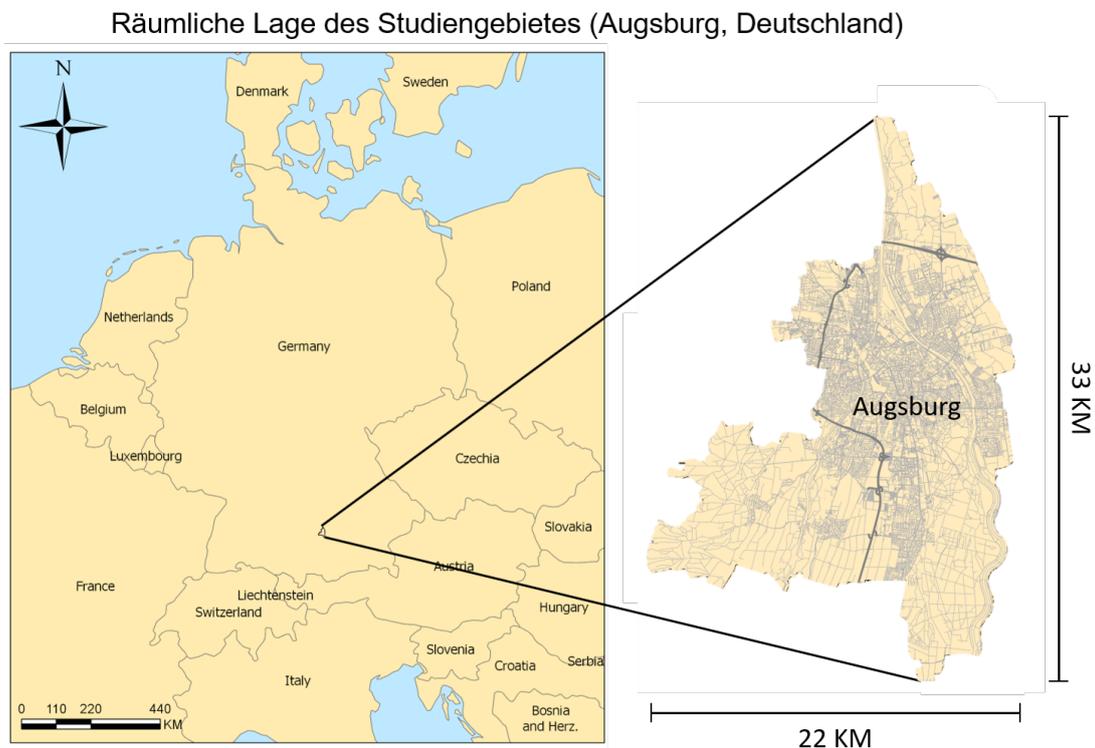


Abbildung 22: Karte mit der räumlichen Abgrenzung des Studiengebietes

Abbildung 22 zeigt die Lage und die räumliche Ausdehnung des Testge-

bietet. Links in der Abbildung ist Zentral- und Nordeuropa zu sehen. Der rechte Ausschnitt zeigt die Stadt Augsburg vergrößert. Die Messungen finden primär, aber nicht ausschließlich, in diesem Gebiet statt.

6.2 Versuchsaufbauten für die Datenerfassung

Für die Überprüfung der Einflüsse der unterschiedlichen vorgestellten Faktoren auf die Genauigkeiten von mobilen Feinstaubmessungen werden unterschiedliche Versuchsaufbauten implementiert. Für Vergleichsmessungen der verwendeten Feinstaubsensoren ist ein statischer Versuchsaufbau vorgesehen, bei dem alle Sensoren zeitgleich und unter gleichen Umgebungsbedingungen laufen (Versuchsaufbau 1). Versuchsaufbau 2 soll den Einfluss von Geschwindigkeit und Sensorausrichtung auf die Messergebnisse untersuchen. Dabei werden die Feinstaubmessgeräte in unterschiedlichen Ausrichtungen auf einem Fahrzeug montiert. Versuchsaufbau 3 stellt eine für Alltagsfahrten praktikable Lösung zur Installation von Feinstaubsensoren auf Fahrzeugen dar, bei der die Messgeräte so auf einem KFZ installiert werden, dass diese für andere Verkehrsteilnehmer kaum sichtbar sind. Mit Versuchsaufbau 3 werden über fünf Monate hinweg größere Datenmengen gesammelt, mit denen der Einfluss des Umgebungsverkehrs eines Fahrzeuges untersucht wird.

6.2.1 Versuchsaufbau 1: statische Messungen mit Feinstaubsensoren

Der beschriebene Versuch sieht vor, drei Feinstaubparameter (PM_{01} , $PM_{2.5}$, PM_{10}) aus sechs Sensoren (fünf *Alphasense OPC-N3*, ein *Plantower PMS3003*) über den Zeitraum von 14 Tagen aufzuzeichnen und miteinander zu vergleichen. Um für gleiche Messvoraussetzungen zu sorgen, werden die Sensoren nebeneinander auf einer Holzleiste platziert. Um witterungsbedingte Einflüsse wie die Sonneneinstrahlung oder Feuchtigkeit in Form von Regen zu minimieren, wird jeder einzelne Sensor in einem Winkelrohr befestigt. Abbildung 23 zeigt ein solches Rohr mit einem Feinstaubsensor.

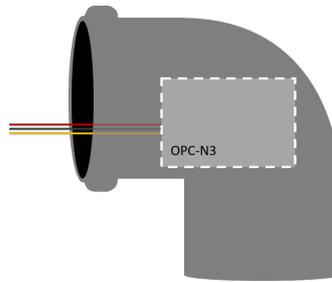


Abbildung 23: Feinstaubsensor im Winkelrohr zur Reduzierung von witterungsbedingten Einflüssen

Abbildung 23 zeigt einen Feinstaubsensor, der in einem 90 Grad gebogenen Rohr befestigt ist. Der Winkel des Rohres verhindert Sonneneinstrahlung in beide Öffnungen. Darüber hinaus wird so verhindert, dass Regen auf die Sensoroberfläche oder in den Lufteingang des Sensors gelangt. Feuchtigkeit kann sich darüber hinaus durch die Biegung des Materials nicht im Inneren des Rohres sammeln.

Die bestückten Rohre werden im Anschluss nebeneinander auf einer Leiste befestigt, wie in Abbildung 24 zu sehen ist.

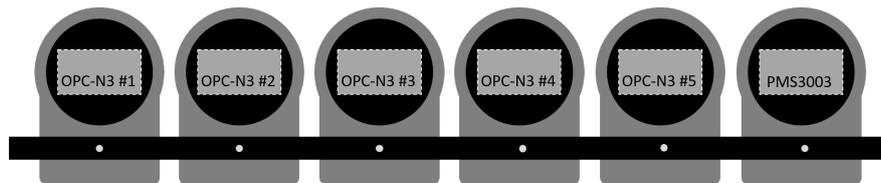


Abbildung 24: Anordnung von fünf *Alphasense OPC-N₃* Sensoren sowie von einem *PMS3003* Sensor in Winkelrohren

Abbildung 24 zeigt sechs Sensor-Rohre, die nebeneinander angeordnet sind. Die enthaltenen Sensoren sind nummeriert (OPC-N₃ #1, OPC-N₃ #2, OPC-N₃ #3, OPC-N₃ #4, OPC-N₃ #5, PMS3003). Alle Geräte werden zur Stromversorgung und zum Auslesen der Daten miteinander verkabelt. Die Leiste mit den Sensoren wird auf einem Sägebock befestigt und im Freien platziert, wie in Abbildung 25 zu sehen ist.



Abbildung 25: Vergleichsmessung von fünf *OPC-N₃* sowie einem *PMS₃₀₀₃* Feinstaubsensor

6.2.2 Versuchsaufbau 2: statische und mobile Vergleichsmessungen

Versuchsaufbau 2 soll zeigen, inwiefern die Ausrichtung der Sensoren (Lufteinlass entgegen der Fahrtrichtung, in Richtung der Fahrtrichtung oder seitlich) als auch die Fahrgeschwindigkeit Einfluss auf die Genauigkeit der Messwerte bei mobilen Einsätzen nehmen. Dafür werden drei *Alphasense OPC-N₃* Sensoren unterschiedlich ausgerichtet auf einem Testfahrzeug montiert. Zwei weitere Sensoren befinden sich an festen Positionen auf einer Teststrecke. Diese Strecke wird mehrfach mit unterschiedlichen Geschwindigkeiten abgefahren. Die Werte der beweglichen Sensoren werden mit den stationären Sensoren verglichen, wenn das Testfahrzeug deren Positionen passiert. Die getesteten Positionen der Sensoren im Kontext der Fahrtrichtung sowie des Fahrtwindes werden auf Abbildung 26 dargestellt.

Abbildung 26 zeigt die drei getesteten Sensorausrichtungen mit Angaben zur Fahrtrichtung sowie des Fahrtwindes. Der Fahrtwind wirkt entgegen der Fahrtrichtung. Die verwendeten *Alphasense OPC-N₃* Sensoren haben jeweils einen Lufteinlass und einen Luftauslass, die sich auf zwei gegenüberliegenden Seiten der Sensoren befinden. Während der Lufteinlass durch eine einfache Öffnung im Sensorgehäuse bewerkstelligt wird, befindet sich am Luftauslass ein Ventilator, der einen Luftstrom durch die Messkammern des Sensors erzeugt. Auf Abbildung 26 werden drei Sensorpositionierungen gezeigt. Sensorposition 1 sieht vor, dass sich der Lufteinlass in Fahrtrichtung befindet. Somit ist der Sensor dem entgegengesetzten Fahrtwind ausgesetzt, der zusätzlich zum Ventilator am Lufteinlass für Luftstrom sorgt. Sensorausrich-



Abbildung 26: Ausrichtung der Feinstaubsensoren entgegen, mit und seitlich der Fahrtrichtung

Abbildung 26 zeigt die drei verschiedenen Ausrichtungen der Feinstaubsensoren. Sensorposition 1 ist so ausgerichtet, dass der Lufteinlass nach oben und der Luftauslass nach unten zeigt. Sensorposition 2 ist so ausgerichtet, dass der Lufteinlass an der Seite zeigt. Sensorposition 3 ist so ausgerichtet, dass der Lufteinlass nach unten zeigt. Die Fahrtrichtung ist durch einen vertikalen Pfeil nach oben und der Fahrtwind durch einen vertikalen Pfeil nach unten dargestellt.

Sensorposition 2 sieht vor, den Sensor um 90 Grad gedreht gegenüber Sensorposition 1 auszurichten. Dadurch sind weder Luftein- noch Luftauslass direkt vom Fahrtwind beeinflusst. Sensorposition 3 sieht vor, den Lufteinlass des Sensors entgegengesetzt zur Fahrtrichtung zu positionieren. Dadurch wirkt der Fahrtwind auf den Ventilator am Lufteinlass und erzeugt somit einen Gegendruck zur eigentlichen Strömungsrichtung innerhalb des Sensors.

Die verschiedenen ausgerichteten Sensoren sind an unterschiedlichen Positionen eines Testfahrzeuges montiert. Diese werden auf Abbildung 27 dargestellt.

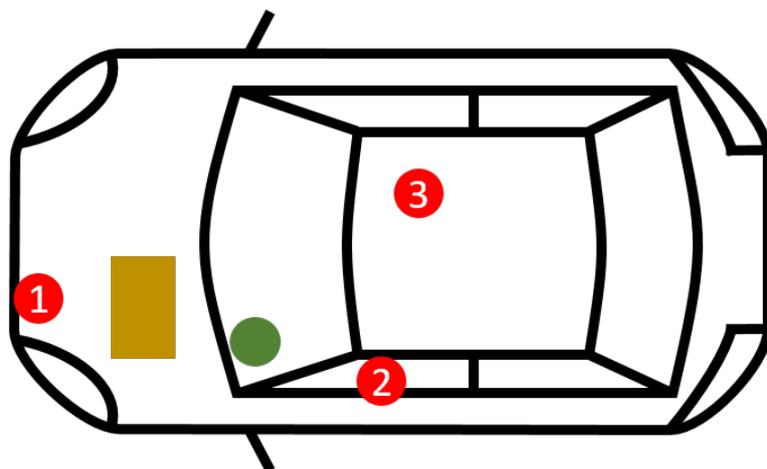


Abbildung 27: Schematische Zeichnung mit den Positionen der Feinstaubsensoren (rote Kreise), des GNSS-Moduls (grüner Kreis) sowie der zentralen Stromquelle (gelbes Rechteck)

Abbildung 27 zeigt das Testfahrzeug mit den Positionen der Feinstaubsensoren (rote Kreise, nummeriert von 1 bis 3), der Position des GNSS-Moduls zur Lokalisierung (grüner Kreis) sowie der Autobatterie als zentrale Stromquelle (gelbes Rechteck). Alle Module werden über die Autobatterie gespeist.

Dafür muss die Betriebsspannung der KFZ Batterie zunächst von 12 auf 5 Volt gewandelt werden, wofür entsprechende Spannungswandler genutzt werden. Diese werden per USB-Kabel mit den Sensoren verbunden. Die Feinstaubsensoren speichern die Messwerte automatisch auf ihrer internen SD-Karte ab. Das GNSS-Modul zur Positionierung wird über einen *Arduino Uno* Mikrokontroller ausgelesen. Die Werte werden mithilfe eines Datenloggers auf einer SD Karte gespeichert. Über den Zeitstempel der Feinstaubsensoren sowie des GNSS-Moduls können die Werte verschiedener Datenquellen am Ende zusammengeführt werden. *Feinstaubsensor 1* befindet sich im Motorraum. Am Lufteinlass ist ein Schlauch befestigt, der über den Kühlergrill des Fahrzeuges nach außen führt. Die Ausrichtung des Schlauchendes ist mit der Fahrtrichtung, sodass der Fahrtwind entgegen der internen Strömungsrichtung des Sensors wirkt. *Feinstaubsensor 2* befindet sich auf der Fahrerseite des Fahrzeuges. Ein damit verbundener Schlauch führt durch die Scheibe ins Freie, sodass die Luft von der linken Seite des Kraftfahrzeuges in den Sensor gelangt. *Feinstaubsensor 3* befindet sich über einen Magneten befestigt auf dem Dach des Autos. Der Lufteinlass ist nach hinten, entgegen der Fahrtrichtung, ausgerichtet. Das System kann über einen Funkschalter an- und ausgeschaltet werden. Die Architektur und Verkabelung der Elemente des Messsystems wird in Abbildung 28 gezeigt.

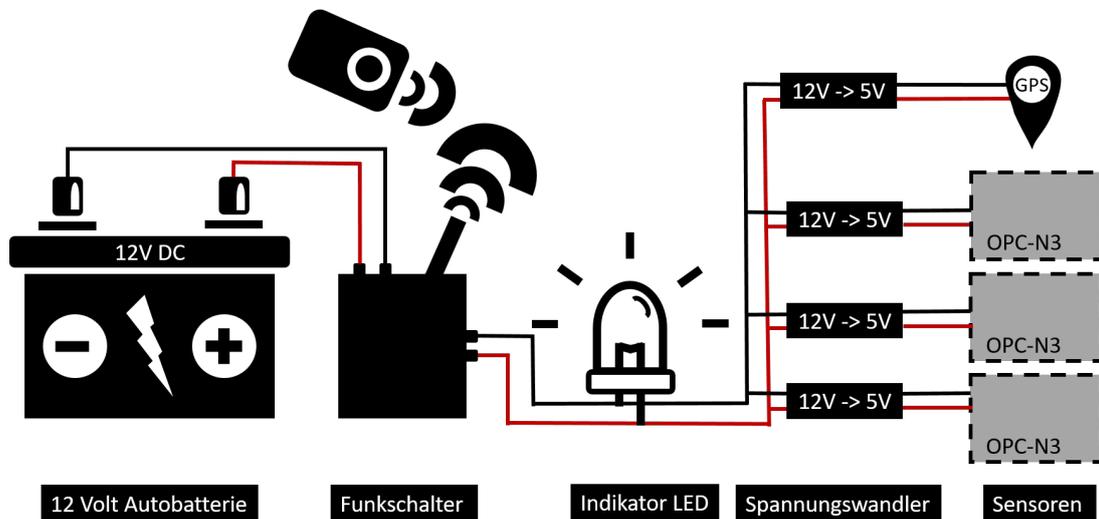


Abbildung 28: Architektur des Messsystems zu Versuchsaufbau 2

Abbildung 28 zeigt die Architektur der Messelemente zu Versuchsaufbau 2. Zu sehen auf der Abbildung sind die Autobatterie, der Funkschalter, eine Indikator LED, Spannungswandler sowie vier Sensoren (drei Feinstaubsensoren und ein GNSS-Positionierungsmodul). Das System hat eine Betriebsspannung von 12 Volt, die durch die Autobatterie bereit gestellt wird. Der Plus-

sowie Minuspol der Batterie sind mit einem Funkschalter verbunden. Dieser ermöglicht es, den Stromfluss innerhalb des Messsystems entweder zu unterbrechen (*Schalterstellung 0*) oder zu aktivieren (*Schalterstellung 1*). Der Status des Schalters kann durch eine ebenfalls im System inkludierte Indikator LED nachvollzogen werden. Ist das Messsystem aktiv, leuchtet diese. Wird die Messung unterbrochen durch die Fernbedienung des Funkschalters, leuchtet sie nicht. Während das Messsystem mit einer Betriebsspannung von 12 Volt läuft, benötigen die verwendeten Sensoren jedoch 5 Volt. Deshalb sind den drei *OPC-N₃* Feinstaubsensoren und dem GNSS-Positionierungsmodul Spannungswandler vorgeschaltet.

Das Experiment sieht nur ein Teilgebiet des in Abbildung 22 definierten Messgebietes vor. Die Messstrecke befindet sich im Zentrum des befahrenen Gebietes. Innerhalb der Teststrecke gibt es zwei stationäre Messpunkte, die durchgehend Feinstaubwerte aufzeichnen, während das Testfahrzeug an diesen vorbeifährt und mit ihren mobilen Sensoren Werte aufzeichnet. Die gewählte Messstrecke ist in Abbildung 29 zu sehen.

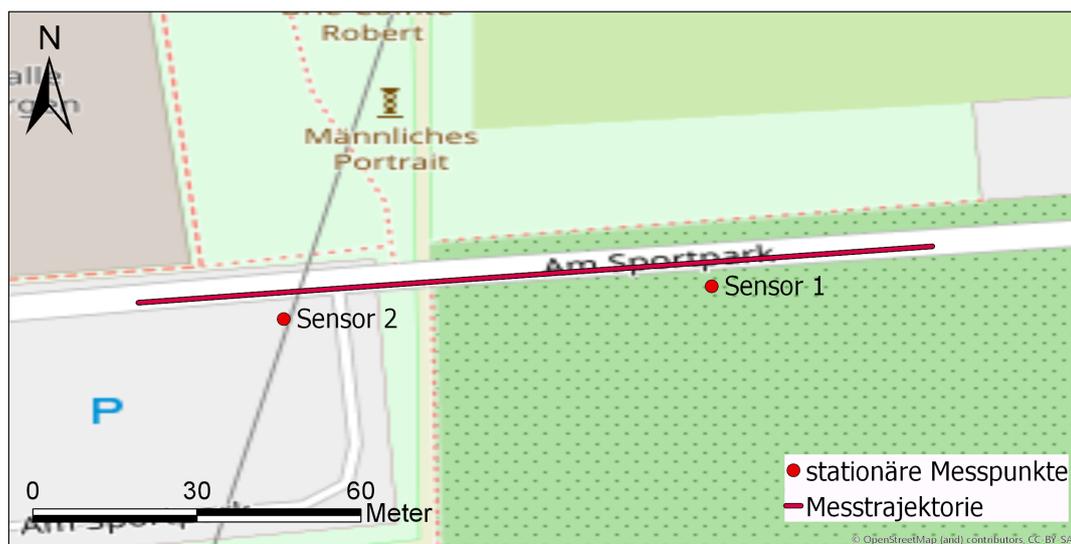


Abbildung 29: Testgebiet zur Überprüfung des Einflusses von Sensorausrichtung und Fahrgeschwindigkeit auf Messwerte von Feinstaubsensoren

Abbildung 29 zeigt das gewählte Testgebiet für Versuchsaufbau 2. Die Strecke befindet sich in der Stadt Stadtbergen, zentral des Messgebietes. Innerhalb des Gebietes wird eine 100 Meter lange Messstrecke definiert, auf der mit drei unterschiedlich ausgerichteten Sensoren bei unterschiedlichen Geschwindigkeiten Feinstaub gemessen wird. Die Strecke ist mit einer roten Linie auf der Karte dargestellt. Nach 20 sowie 80 Metern befindet sich neben der Messstrecke jeweils ein stationärer Messpunkt, der durchgehend Feinstaubwerte

aufzeichnet. Diese Messpunkte sind mit roten Punkten dargestellt. Für das Experiment wird die Messstrecke jeweils mit den folgenden Geschwindigkeiten abgefahren:

- 5 km/h
- 10 km/h
- 20 km/h
- 30 km/h
- 40 km/h
- 50 km/h

Damit werden typische Fahrgeschwindigkeiten abgedeckt, die innerorts gefahren werden. Während der Fahrt werden die Geschwindigkeiten durch den Tempomat des Testfahrzeuges geregelt, damit diese über die Messstrecke hinweg gleich bleiben.

6.2.3 Versuchsaufbau 3: mobile Erfassung von Feinstaubsensordaten

Versuchsaufbau 3 ähnelt Versuchsaufbau 2, mit dem Unterschied dass sich keine Messmodule außerhalb des Fahrzeuges befinden. Des Weiteren wird zur Observation des Umgebungsverkehrs eine Videokamera mit in das Messsystem integriert. Abbildung 30 zeigt eine schematische Zeichnung des Messaufbaus.

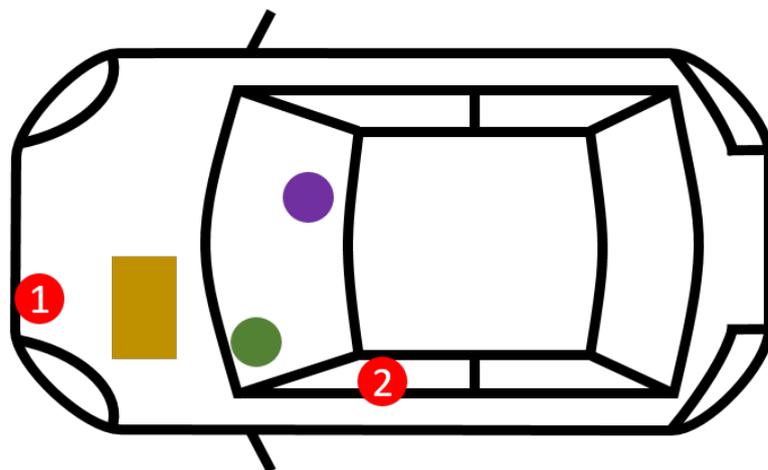


Abbildung 30: Schematische Zeichnung mit den Positionen der Feinstaubsensoren (rote Kreise), des GNSS-Moduls (grüner Kreis), der zentralen Stromquelle (gelbes Rechteck) sowie der Videokamera zur Aufzeichnung des Umgebungsverkehrs (violetter Kreis)

Abbildung 30 zeigt die Positionen der Feinstaubsensoren (nummerierte rote Kreise), des GNSS-Moduls (grüner Kreis), der Videokamera (violetter Kreis) sowie der Stromquelle (gelbes Rechteck) schematisch. Wie in Versuchsaufbau 2 (Abbildung 27) befindet sich ein Feinstaubsensor im Motorraum des Fahrzeuges mit einem durch den Kühlergrill führenden Schlauch (*Feinstaubsensor 1*). *Feinstaubsensor 2* befindet sich ebenfalls analog zu Abbildung 27 an der Fahrerseite des Kraftfahrzeuges mit einem Schlauch, der seitlich aus dem Fenster des Autos führt. Das GNSS-Modul befindet sich auf der linken Seite unter der Windschutzscheibe des Fahrzeuges. An der Windschutzscheibe befestigt ist eine Videokamera des Typs *Apeman C450* mit einer Auflösung von 1080P und einem Weitwinkel von 170 Grad. Um den Umgebungsverkehr unabhängig von der Umgebungshelligkeit aufzuzeichnen, hat die Videokamera ebenfalls eine Nachtsichtfunktion. Die Stromquelle der Module ist die Autobatterie. Abbildung 31 zeigt die Architektur des Messsystems.

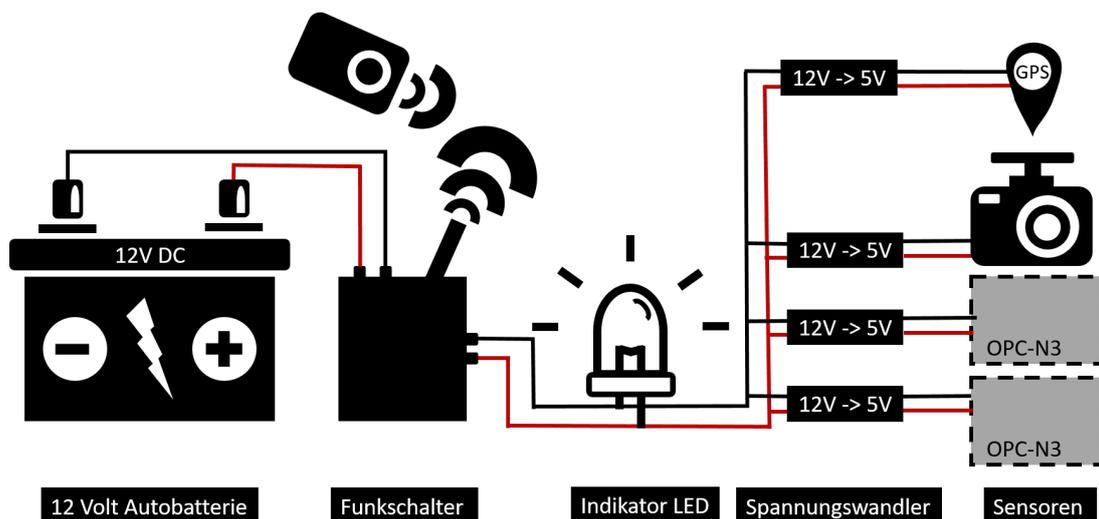


Abbildung 31: Architektur des Messsystems zu Versuchsaufbau 3

Abbildung 31 zeigt die Architektur des Messsystems zu Versuchsaufbau 3 mit der 12 Volt Autobatterie als Stromquelle, einem Funkschalter, einer Indikator LED, Spannungswandlern sowie den Messmodulen (Feinstaubsensoren, Videokamera und Positionierung). Das System ist im Motorraum des Testfahrzeuges an der 12 Volt Autobatterie angeklemmt. Das erste Modul innerhalb des Schaltkreises ist ein Funkschalter. Dieser kann per Fernbedienung aus dem Fahrzeug heraus die Stromversorgung des Messsystems starten und unterbrechen. Ob das System gerade läuft, zeigt eine LED, die bei *Schaltstatus 1* (Stromversorgung läuft) leuchtet und bei *Schaltstatus 0* (Stromversorgung unterbrochen) kein Licht abgibt. Auf die LED folgen im Schaltkreis Spannungswandler, die die Betriebsspannung von 12 Volt auf 5 Volt

herunterregeln. Das ist nötig, da die im Messsystem verwendeten Messmodule eine Spannung von 5 Volt benötigen. Diese Messmodule umfassen zwei Feinstaubsensoren (einer im Motorraum, einer innerhalb der Fahrerkabine), die Videokamera sowie ein GNSS-Positionierungsmodul. Alle Messkomponenten speichern ihre Werte autark auf ihrer internen SD-Speicherkarte ab. Der Zeitstempel aller Messungen lässt es zu, die unterschiedlichen Datensätze miteinander zu kombinieren.

6.3 Datenaufbereitung und Überblick über die resultierenden Datensätze

Im Zuge der Datenaufbereitung werden die mit den jeweiligen Versuchsaufbauten aufgenommenen Daten für weitere Analysen vorbereitet. Insgesamt werden drei Datentypen generiert. Sensordaten beinhalten alle Feinstaub-, Temperatur- und Luftfeuchtigkeitsmessungen. Raum-zeitliche Daten werden in den Positionsdaten gespeichert. Die Videodaten beinhalten die Aufnahmen des Umgebungsverkehrs bei Messungen. Jeder Versuchsaufbau sieht die Messung unterschiedlicher Datentypen vor, wie in Tabelle 9 zu sehen ist.

Versuchsaufbau	Sensordaten	Positionsdaten	Videodaten
1	+	-	-
2	+	+	-
3	+	+	+

Tabelle 9: Präsenz (+) und Absenz (-) von Datentypen bei Versuchsaufbau 1, 2 und 3

Tabelle 9 zeigt die drei Versuchsaufbauten mit ihren gemessenen Datentypen. Versuchsaufbau 1 beinhaltet statische Vergleichsmessungen mit Feinstaubsensoren. Dabei werden lediglich Feinstaub-, Temperatur- und Luftfeuchtigkeitsdaten erfasst. Versuchsaufbau 2 untersucht den Einfluss von Sensorausrichtung und Fahrtgeschwindigkeit auf die Messergebnisse mit Feinstaubsensoren. Dazu werden neben den Sensordaten auch Positionsdaten benötigt. Versuchsaufbau 3 sieht sowohl Sensordaten, als auch Positionsdaten und Videodaten vor.

Jegliche Datentypen werden zur weiteren Analyse mithilfe eines Python Skriptes aufbereitet und miteinander verbunden. Abbildung 32 zeigt die Aufbereitung der erfassten Daten in einem Ablaufdiagramm. Der Datensatz ist so abgespeichert, dass sich jede Messung in einem individuellen Ordner befindet. Der Name des Ordners entspricht der Mess-ID. Jeder Ordner enthält die jeweiligen, zur Messung passenden, Datensätze. Somit beinhalten die Ordner

für die statischen Messungen (Versuchsaufbau 1) lediglich die Messungen der Sensoren, während die Ordner für Versuchsaufbau 2 und 3 zusätzlich noch Positionsdaten beziehungsweise Videodaten enthalten.

Zunächst wird über die Ordner mit dem Namen der jeweiligen Mess-ID iteriert. Danach werden die Datensätze aus jedem Ordner (Sensordaten, Positionsdaten, Videodaten) identifiziert. Videodaten werden mit der Funktion *objektdetektion()* aufbereitet. Das Ziel ist es, aus dem Videostream für jede Sekunde die Anzahl an Fahrzeugen zu extrahieren und in eine Komma-separierte Datei zu schreiben. Dafür wird zunächst das jeweilige Video in Einzelframes konvertiert. Die Framerate eines Videos zeigt, wie viele Frames das Video pro Sekunde hat. Um die Verkehrssituation sekundlich zu bestimmen, wird also nur jeder x -te Frame für die weitere Analyse verwendet (wobei x für die Framerate steht). Die extrahierten Frames werden im Anschluss mithilfe eines auf dem *YOLOv3*-Algorithmus basierenden Modells analysiert. Dabei wird die Häufigkeit der Fahrzeugklassen *truck*, *car*, *motorbike* und *bus* in den Bildern gezählt. Diese Ergebnisse werden im Anschluss in eine komma-separierte Datei geschrieben mit der Datenstruktur:

- sekunde (Frame-ID)
- cars (Anzahl der Autos)
- motorbikes (Anzahl der Motorräder)
- trucks (Anzahl der Lastkraftwagen)
- busses (Anzahl der Busse)
- total (Gesamtanzahl Fahrzeuge)
- ambient (Boole'scher Wert mit FALSE für *kein Fahrzeug* und TRUE für *ein oder mehr Fahrzeuge*)

Nach der Analyse der Videos werden die Positionsdaten aufbereitet. Dafür werden die Positionierungsgenauigkeiten untersucht und auf dieser Basis Ausreißerwerte eliminiert. Die Aufbereitung der Sensordaten sieht vor, zunächst mehrfach vorhandene Sensordaten zu kombinieren. Wurden zum Beispiel für eine Messung zwei Sensoren verwendet, werden aus den zwei resultierenden Datenreihen eine Datei gemacht. Dafür werden die Daten auf Basis des Zeitstempels kombiniert. Die zusammengefügte Werte werden im Anschluss wiederum in eine komma-separierte Datei geschrieben.

Aufbereitung der Datensätze 1, 2 und 3

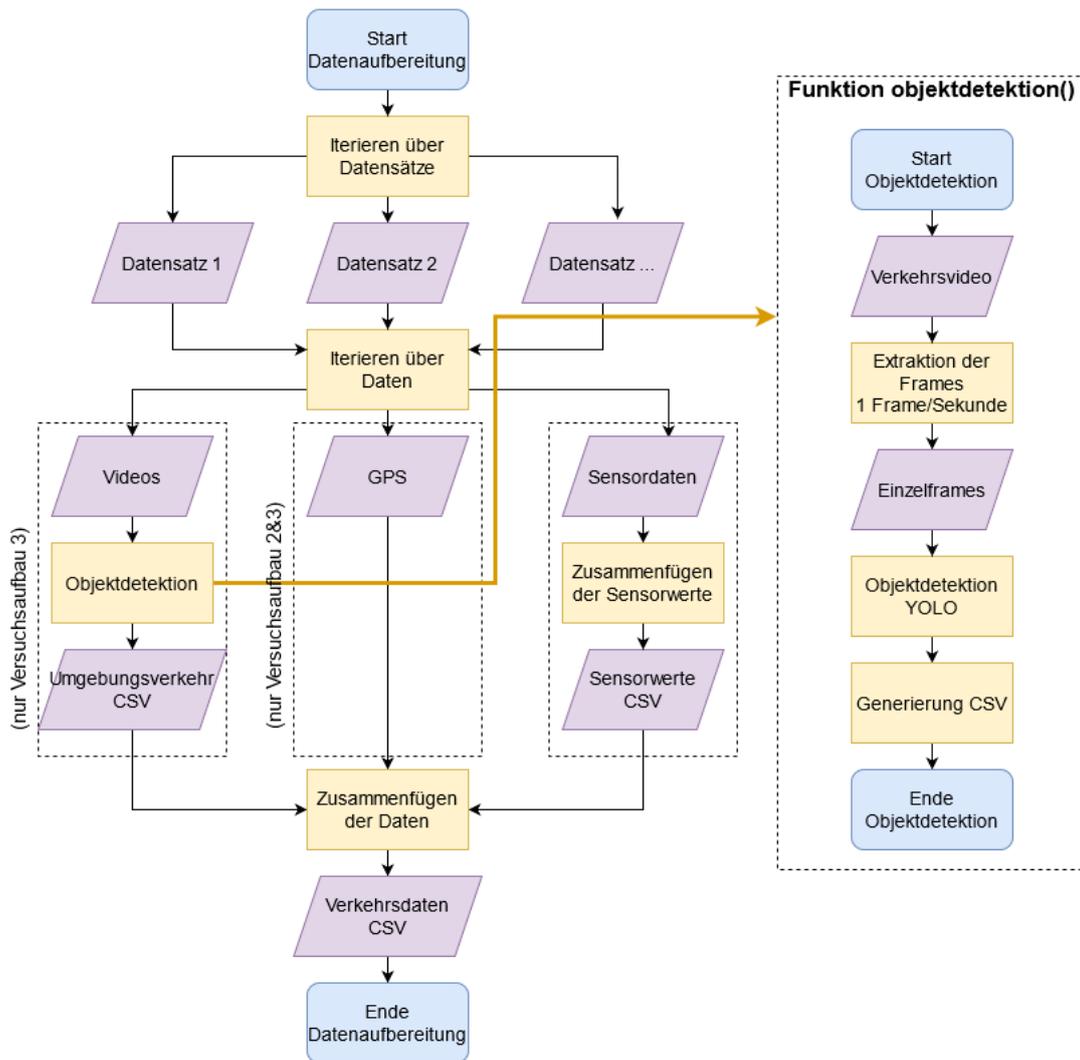


Abbildung 32: Flowchart zur Datenaufbereitung und Detektion der Fahrzeuge aus den Verkehrsvideos

Sobald sowohl die Videodaten als auch die Positions- sowie Sensordaten aufbereitet sind, werden diese zu einer Datei zusammengefügt. Dafür verwendet das Skript den Zeitstempel der jeweiligen Messungen. Die Ergebnisse werden wieder in eine neue komma-separierte Datei geschrieben. Die Datei hat unterschiedliche Spalten, die in Tabelle 10 gezeigt werden.

Spalte	Datentyp	Beispiel
ID	Long	1
Zeitstempel	Datetime	2020-01-28T11:14:45
Latitude*	Float	48.366512
Longitude*	Float	10.894446
HDOP*	Double	2.6
VDOP*	Double	2.9
Satelliten*	Integer	5
PM_1	Float	8.53
$PM_{2.5}$	Float	10.43
PM_{10}	Float	14.12
bin_0, \dots, bin_{23}	Integer	53
Temperatur	Float	8.5
Luftfeuchtigkeit	Float	88.5
cars**	Integer	2
motorbikes**	Integer	0
trucks**	Integer	1
busses**	Integer	0
total**	Integer	3
ambient**	Boolean	TRUE

Tabelle 10: Übersicht über die Spaltennamen der aus der Datenaufbereitung resultierenden Datei mit Datentypen und Beispielen (* = nur Versuchsaufbau 2 und 3, ** = nur Versuchsaufbau 3)

Wie in der Tabelle 10 zu sehen ist, werden raum-zeitliche Informationen (Zeitstempel, geographische Länge, geographische Breite, HDOP, VDOP, Anzahl der Satelliten), Feinstaubwerte (PM_1 , $PM_{2.5}$, PM_{10} , bin_0, \dots, bin_{23}), Temperatur, Luftfeuchtigkeit und Informationen über den Umgebungsverkehr (KFZ, Motorräder, LKW, Busse, Umgebungsverkehr als Boole'sche Variable) abgespeichert.

Mithilfe der Sensoraufbauten werden Datensätze erfasst. Datensatz 1 besteht aus Messwerten von fünf OPC-N3-Sensoren und einem PMS3003-Messgerät, die vom 23.09.2020 bis zum 07.10.2020 nebeneinander Daten aufgezeichnet haben. Die Parameter Datumzeit, PM_1 , $PM_{2.5}$, PM_{10} , Temperatur und Luftfeuchtigkeit werden nach Ablauf der Messzeit ausgelesen. Jeder Sensor trägt 856.945 Observationen zu Datensatz 1 bei, was eine Gesamtanzahl von 5.141.670 Datenpunkten zur Folge hat. Abbildung 33 zeigt die Verläufe der Temperatur sowie der Luftfeuchtigkeit.

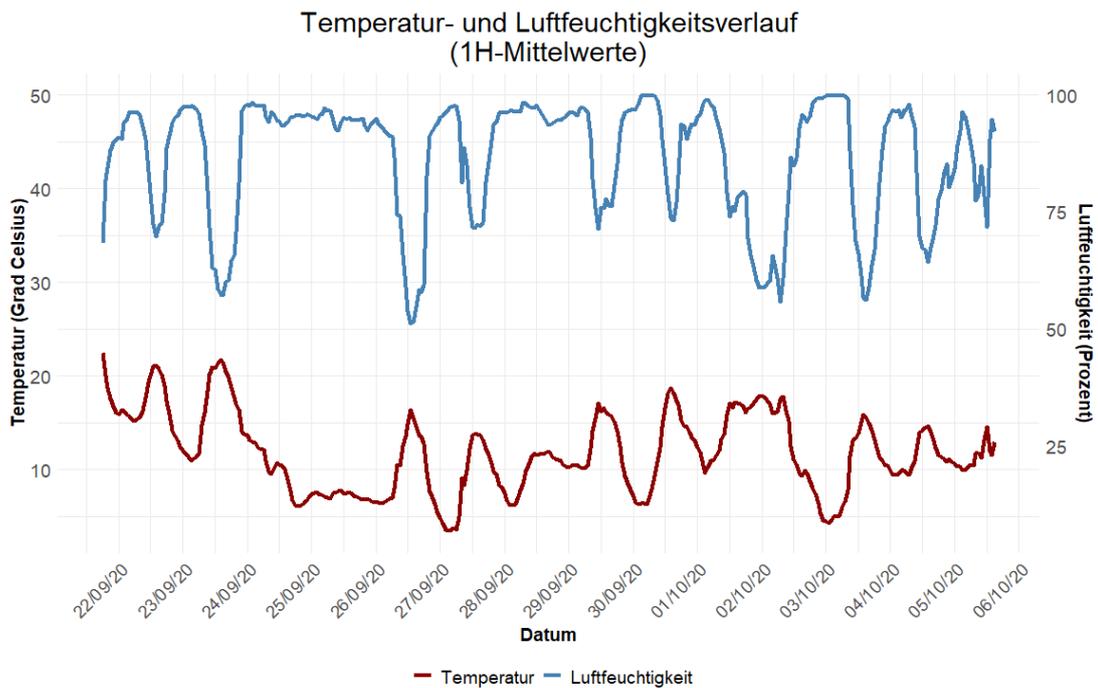


Abbildung 33: Temperatur- und Luftfeuchtigkeitsverlauf während der Erfassung von Datensatz 1

Abbildung 33 zeigt den Verlauf von Temperatur und Luftfeuchtigkeit über den Messzeitraum. Auf der X-Achse wird der zeitliche Verlauf der Messung dargestellt. Die Skala reicht vom 22.09.2020 bis zum 06.10.2020. Die linke Y-Achse zeigt die Temperatur während den Messungen mithilfe einer roten Linie. Die Temperatur reicht von *drei* Grad Celsius bis 23 Grad Celsius. Die zweite Y-Achse stellt die Luftfeuchtigkeit mithilfe einer blauen Linie dar. Die Werte reichen von 51 % bis 100 %. Alle Temperaturwerte liegen im möglichen Arbeitsbereich der Feinstaubsensoren. Die Luftfeuchtigkeitswerte erreichen jedoch in der Nacht vom 30.09.2020 bis zum 01.10.2020 sowie am 03.10.2020 den Maximalwert von 100 %.

Datensatz 2 besteht aus den Daten zweier Messereignisse am 16.11.2020 und am 28.02.2021. Während mithilfe von Sensoraufbau 2 am 16.11.2020 die Strecke einmal mit verschiedenen Geschwindigkeiten $v = \{5, 10, 20, 30, 40, 50\} km/h$ abgefahren wurde, fanden am 28.02.2021 zwei Durchläufe statt. Der Datensatz besteht aus insgesamt 382 Observationen. Am 16.11.2020 beträgt die mittlere Luftfeuchtigkeit 44 Prozent und die durchschnittliche Temperatur 17 Grad. Dahingegen liegt am 28.02.2021 die Luftfeuchtigkeit bei 31 Prozent und die Temperatur bei 15 Grad Celsius.

Die Datenaufnahme mithilfe von Versuchsaufbau 3 lief vom 01.11.2020 bis zum 01.03.2021. Der resultierende Datensatz besteht aus 23.725 Observationen, zusammengesetzt aus 47 Einzeltrajektorien. Die Punkte sind über das

gesamte Testgebiet verteilt. Abbildung 34 gibt einen Überblick über die räumliche Verteilung der Datenpunkte.



Abbildung 34: Überblick über die mit Versuchsaufbau 3 erfassten Daten mit 23.725 Observationen

Abbildung 34 zeigt das Testgebiet mit den aufgenommenen Datenpunkten, dargestellt durch gelbe Punkt-Features. Die Observationen befinden sich größtenteils im Stadtgebiet von Augsburg. Ein kleiner Teil der Datenpunkte ist im westlich angrenzenden Landkreis Augsburg sowie im östlich angrenzenden Landkreis Aichach-Friedberg lokalisiert.

Der Datensatz enthält verschiedene Variablen, die einzelne Observationen bezüglich der aktuellen Position, dem Umgebungsverkehr sowie Umweltparametern beschreiben. Histogramme für die Geschwindigkeit, sowie über aufgenommene Feinstaubwerte sind in Abbildung 35 zu sehen.

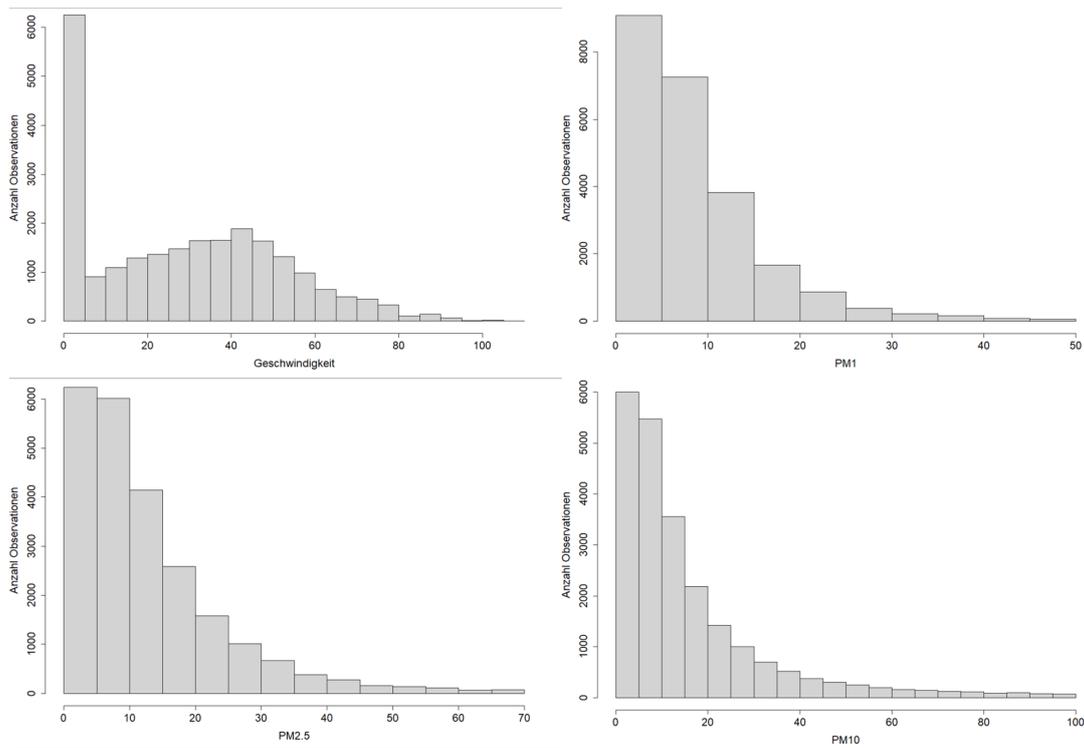


Abbildung 35: Histogramme für die Geschwindigkeit und die Feinstaubwerte (PM_1 , $PM_{2.5}$, PM_{10}) der einzelnen Observationen

Abbildung 35 zeigt Histogramme für die Variablen *Geschwindigkeit*, PM_1 , $PM_{2.5}$ und PM_{10} der Observationen. Während die X-Achse der Histogramme den jeweiligen Wert der untersuchten Variable angibt, zeigt die Y-Achse die Anzahl der betroffenen Datenpunkte. Bezüglich der Geschwindigkeit hat ein Großteil der Observationen den Wert 0 km/h. Das bedeutet, dass das Testfahrzeug steht. Die restlichen Datenpunkte zeigen Fahrgeschwindigkeiten von bis zu 106 Kilometer pro Stunde. Der Mittelwert liegt bei 28,78 km/h. Die Werte für PM_1 , $PM_{2.5}$ und PM_{10} stellen den jeweiligen Mittelwert der zwei im Testfahrzeug installierten Sensoren dar. Für die Darstellung als Histogramm werden die feuchten (nicht luftfeuchtigkeits-korrigierten) Werte verwendet. Für PM_1 Feinstaub liegen die meisten Werte im Bereich um $12 \mu\text{g}/\text{m}^3$. Einzelne Punkte nehmen Werte von bis zu $390 \mu\text{g}/\text{m}^3$ an, Feinstaubwerte von über $80 \mu\text{g}/\text{m}^3$ sind jedoch sehr selten. Der Mittelwert für PM_1 liegt bei $8,80 \mu\text{g}/\text{m}^3$. Für $PM_{2.5}$ Feinstaub zeigt sich ein ähnliches Bild. Die meisten Werte finden sich im niedrigeren Bereich von 0 bis $100 \mu\text{g}/\text{m}^3$. Ausreißer nehmen allerdings Werte von bis zu $900 \mu\text{g}/\text{m}^3$ an. Der Mittelwert für $PM_{2.5}$ liegt bei $13,67 \mu\text{g}/\text{m}^3$. Für PM_{10} Feinstaub ergibt sich ein Mittelwert von $23,34 \mu\text{g}/\text{m}^3$. Ausreißerwerte zeigen bis zu $1000 \mu\text{g}/\text{m}^3$. Die Temperatur und die Luftfeuchtigkeit sind weitere aufgenommene Parameter. Diese geben Aufschluss über Witterungsbedingungen während einer

Fahrt und werden für die Korrektur der Feinstaubwerte benötigt. Histogramme, die die Verteilung der Werte für die genannten Variablen zeigen, sind in Abbildung 36 zu sehen.

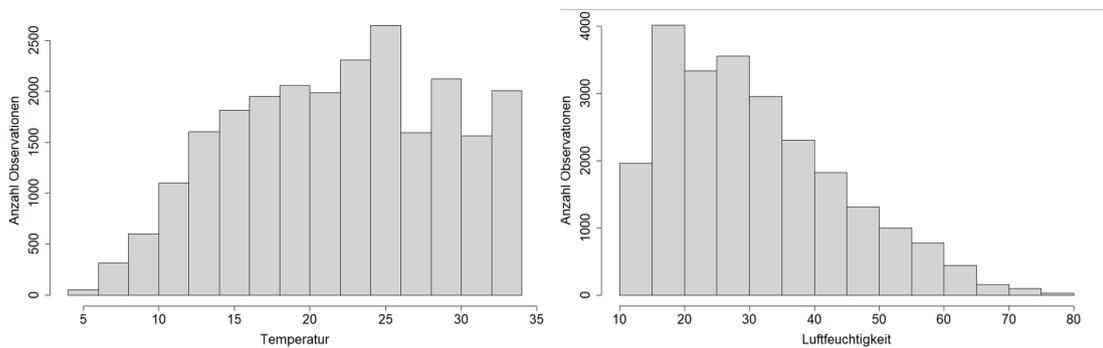


Abbildung 36: Histogramme für die Temperatur und Luftfeuchtigkeit der einzelnen Observationen

Abbildung 36 zeigt Histogramme für die Temperatur und Luftfeuchtigkeit der Observationen. Die Temperatur, dargestellt im linken Histogramm der Abbildung, hat einen Wertebereich von 4,26 Grad Celsius bis 34 Grad Celsius. Die Werteverteilung ist linkschief, was bedeutet, dass höhere Temperaturen häufiger vorkommen als niedrigere. Der Mittelwert liegt bei 22 Grad Celsius. Da sich die Temperatursensoren in der Fahrerkabine beziehungsweise im Motorraum befinden, kann nicht davon ausgegangen werden, dass die im Datensatz aufgenommenen Werte der Außentemperatur entsprechen. Für die Luftfeuchtigkeit ergibt sich eine rechtsschiefe Werteverteilung. Der analysierte Datensatz verfügt über mehr Observationen mit niedriger Luftfeuchtigkeit, als über Punkte mit höheren Feuchtigkeitswerten. Aufgenommene Werte reichen von 8,6 Prozent bis zu 75,81 Prozent Luftfeuchtigkeit. Der Mittelwert des Datensatzes liegt bei 30,97 Prozent.

Informationen über den Umgebungsverkehr des Testfahrzeuges stammen von den Videoaufnahmen, die per Objektdetektionsalgorithmus analysiert werden. Ein resultierender Parameter daraus ist die Gesamtanzahl der Fahrzeuge im Umgebungsverkehr, deren Histogramm in Abbildung 37 gezeigt wird.

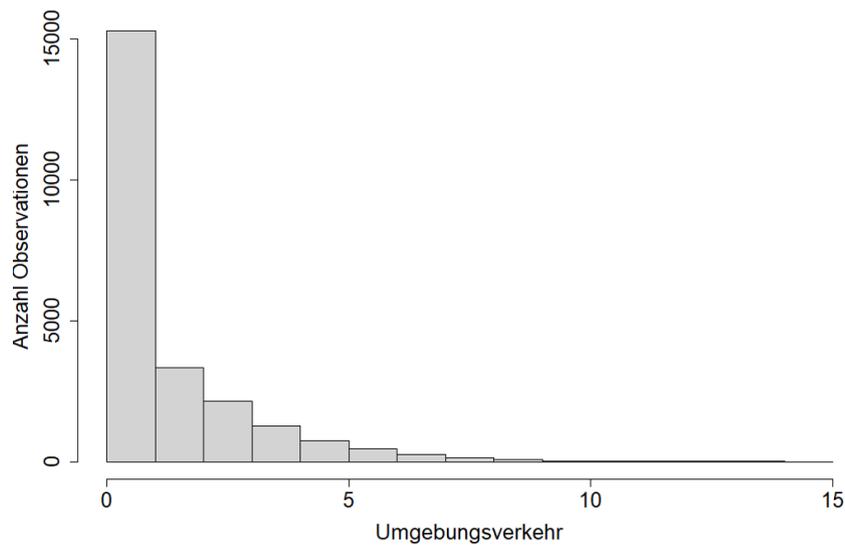


Abbildung 37: Histogramm für die Anzahl der Fahrzeuge im Umgebungsverkehr

Abbildung 37 zeigt ein Histogramm für die Anzahl von Fahrzeugen im Umgebungsverkehr während dem Messen. Den größten Anteil machen Datenpunkte ohne detektierte Fahrzeuge aus. Je höher die Anzahl der Fahrzeuge im Umgebungsverkehr, desto weniger Observationen gibt es. Die maximal detektierte Fahrzeugmenge liegt bei 15. Durchschnittlich zeigen die Datenpunkte eine Anzahl von 1,42 Fahrzeugen pro Observation im Umgebungsverkehr.

6.4 Analyse der Positionierungsgenauigkeit

Alle Observationen verfügen über einen Orts- und Zeitstempel. Die Qualität der raum-zeitlichen Verortung der Datenpunkte hängt essentiell von der Genauigkeit des verwendeten Positionierungssensors ab. Wie bereits ausgeführt, haben die Anzahl der verwendeten Satelliten und die Verwendung externer Ortungstechniken einen Einfluss auf die Positionierungsgenauigkeit. Dazu kommen dynamische Faktoren wie Witterungsverhältnisse und der Bebauungsgrad in der direkten Umgebung einer Messung. In den folgenden Subkapiteln wird der generierte Sensordatensatz auf die Positionierungsgenauigkeit hin untersucht.

6.4.1 Analyse der horizontalen und vertikalen Positionierungsabweichung

Die Positionsgenauigkeit (*DOP*) einer Lokalisierung kann vertikal (*VDOP*) und horizontal (*HDOP*) abweichen. Viele *GNSS*-Module sind dazu in der Lage, diese Abweichungen selbst zu bestimmen. Histogramme mit der Verteilung von *HDOP* und *VDOP* der Observationen des aufgenommenen Daten-

satzes werden in Abbildung 38 gezeigt.

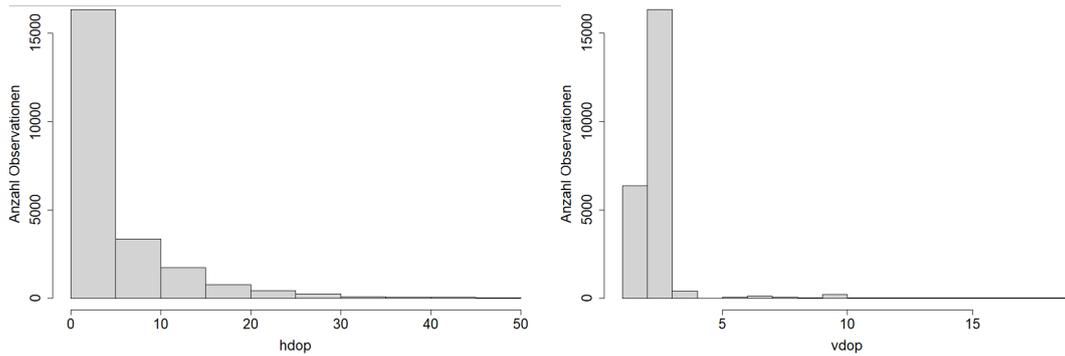


Abbildung 38: Histogramme für die vertikalen (*VDOP*) und horizontalen (*HDOP*) Abweichungen der Positionierungsgenauigkeit im aufgenommenen Datensatz

Abbildung 38 zeigt zwei Histogramme. Links ist die Verteilung der horizontalen Positionsabweichung zu sehen. Der Wertebereich geht von 4 bis 50, wobei ein großer Teil der Observationen Abweichungen unter 30 aufweist. Werte darüber stellen zumeist einzelne Ausreißer dar. Der Mittelwert liegt bei 12,1. Folglich sind bei den meisten Observationen horizontale Abweichungen auszumachen, die auf eine ausreichend genaue Positionierung hindeuten. Die vertikalen Positionsabweichungen, rechts in Abbildung 38 dargestellt, zeigen einen deutlich kleineren Wertebereich. Dieser reicht von 1 bis 15. Im Durchschnitt liegt die vertikale Abweichung der Positionierungen bei 2,84. Analog zu den *HDOP*-Werten sind zwar Ausreißer festzustellen, diese nehmen jedoch bei *VDOP* deutlich niedrigere Werte an. Zusammenfassend zeigt die Analyse der horizontalen Abweichungen einige Punkte, die potentiell ungenau verortet sind.

6.4.2 Anwendung von Mapmatching-Algorithmen

Zum Mapmatching der Datenpunkte werden unterschiedliche Ansätze getestet. Dazu gehören einfache distanzbasierte Fang (*Snapping*) Algorithmen und die Verwendung eines *Hidden Markov Modells*. Für die Ausführung der Mapmatching-Optionen werden *ArcGIS Pro* (*Snapping*) sowie die Python-Bibliothek *noiseplanet* (*nearest* und *hidden markov model*) verwendet. Ein Liniendatensatz aus *OpenStreetMap*, heruntergeladen von *Geofabrik*¹⁸, fungiert als Netzwerk. Die Datenpunkte werden über die Anwendung der verschiedenen Mapmatching-Ansätze auf eine Position innerhalb des Netzwerkes verschoben. Eine Auswahl der Ergebnisse für ein Teilgebiet der Trajektorie mit der ID 47 wird in Abbildung 39 gezeigt.

¹⁸<https://www.geofabrik.de/>, 06/21

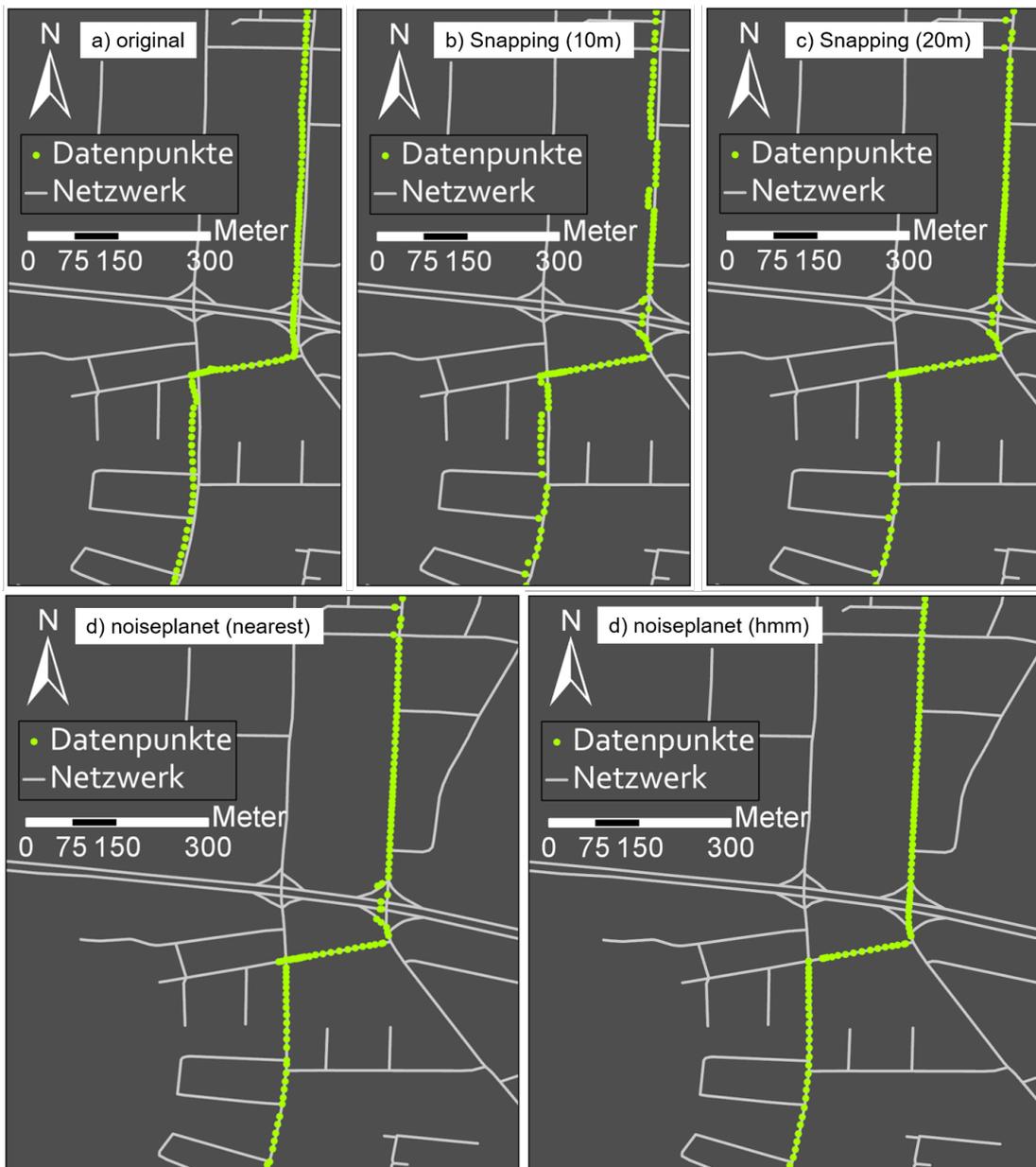


Abbildung 39: Verschiedene Ansätze zum Mapmatching, angewandt auf die Datenpunkte der Fallstudie: originale Punkte (a), Nutzung des *Snapping*-Tools mit zehn Metern Suchradius (b), Nutzung des *Snapping*-Tools mit 20 Metern Suchradius (c), *nearest*-Option der *noiseplanet*-Umgebung (d) und *hidden markov model*-Option der *noiseplanet* Umgebung (e)

Abbildung 39 zeigt Kartenausschnitte mit Ergebnissen der unterschiedlichen Mapmatching-Ansätze für die Trajektorie 47. Dabei werden die Positionen der Observations als hellgrüne Punkte dargestellt. Das Netzwerk, das als Basis für das Mapmatching verwendet wird, ist mit hellgrauen Linien dargestellt. Teil *a*) der Abbildung zeigt die unveränderten Datenpunkte, bevor ein Algorithmus angewandt wird. Es ist zu sehen, dass die Punkte entlang der Netzwerkkanten verlaufen. An einigen Stellen ist der Versatz jedoch deutlich zu erkennen. Teil *b*) der Abbildung zeigt die Datenpunkte nach der Anwen-

dung des *Snapping-Tools* von *ArcGIS Pro* mit einer maximalen Versatz-Distanz von zehn Metern. Im Umkreis von zehn Metern um jeden Punkt wird dabei nach Kanten des zugrundeliegenden Netzwerkes gesucht. Gibt es einen Treffer, wird der Punkt zur Position der nächstgelegenen Kante verschoben. Es ist zu sehen, dass der Ansatz für Punkte, die sich nicht neben Straßenkreuzungen oder -abzweigungen befinden, gut funktioniert. An einigen Stellen ist jedoch zu erkennen, dass Punkte zwischen Straßen hin- und herspringen, was im Sinne einer Fahrt mit einem Fahrzeug nicht als realistisch scheint. Außerdem befinden sich manche Punkte noch an ihrer herkömmlichen Position. Das indiziert, dass die Suchdistanz in diesen Fällen nicht groß genug war. Konsequenterweise sind die betroffenen Punkte weiter als zehn Meter von ihrer nächstgelegenen Netzwerkkante entfernt. Teil *c*) der Abbildung zeigt den gleichen Ansatz, nur mit einem Suchradius von 20 Metern. Nicht-verschobene Punkte sind im Gegensatz zu Teil *b*) nicht mehr in der Abbildung erkennbar. Das lässt auf einen sinnvoll gewählten Suchradius schließen. Punkte finden sich jedoch immer noch auf sich oft ändernden Straßen. Das wird vor allem deutlich beim Betrachten der Straßenkreuzung östlich des Teilgebietes. Teil *d*) zeigt die Ergebnisse für die *nearest*-Option der *noiseplanet* Python-Bibliothek. Im Gegensatz zu den Versuchen mithilfe des *Snapping-Tools* ergeben sich hier weniger kurzfristige Positionssprünge. Diese sind an der komplexen Kreuzung jedoch immer noch deutlich zu erkennen. Teil *e*) zeigt die Datenpunkte nach der Nutzung der *hidden markov model* Option der *noiseplanet*-Bibliothek. Es ist zu erkennen, dass alle Punkte sinnvoll verschoben werden. Jeder Datenpunkt des Teilgebietes findet sich auf einer Kante des Netzwerkes wieder. Auch die Straßenkreuzung zeigt keine fehlerhaft verschobenen Punkte.

6.4.3 Vergleich von Trajektorien

Die Anwendung von Mapmatching Algorithmen ändert die Position der Datenpunkte. Dabei sind sowohl die unveränderten Messungen von GNSS-Modulen, als auch die auf ein Netzwerk verschobenen Punkte nur Annäherungen an die tatsächlichen Positionen der Observationen. Unter der Annahme, dass die mit dem *hidden markov model* (Abbildung 39, Teil *e*) verschobenen Punkte die genaueste Repräsentation darstellen, soll nun anhand von diversen Vergleichsparametern der Unterschied zu den unveränderten Originalpunkten festgestellt werden. Dadurch kann abgeschätzt werden, welchen Einfluss das Mapmatching auf die verschiedenen Trajektorien hat. Die Python-Bibliothek *similaritymeasures* liefert einige Methoden zum Vergleich von Kurven (Jekel et al. 2019). Dazu zählen das *Partial Curve Mapping*, die *Discrete*

Fréchet-Distanz, die Kalkulation der Fläche zwischen zwei Kurven und der Vergleich von Kurvenlängen. Dafür werden die originalen sowie die veränderten Punkte jeweils zu einer Trajektorie verbunden. Die Ergebnisse für die Parameter *PCM* und *DF* finden sich in Abbildung 40.

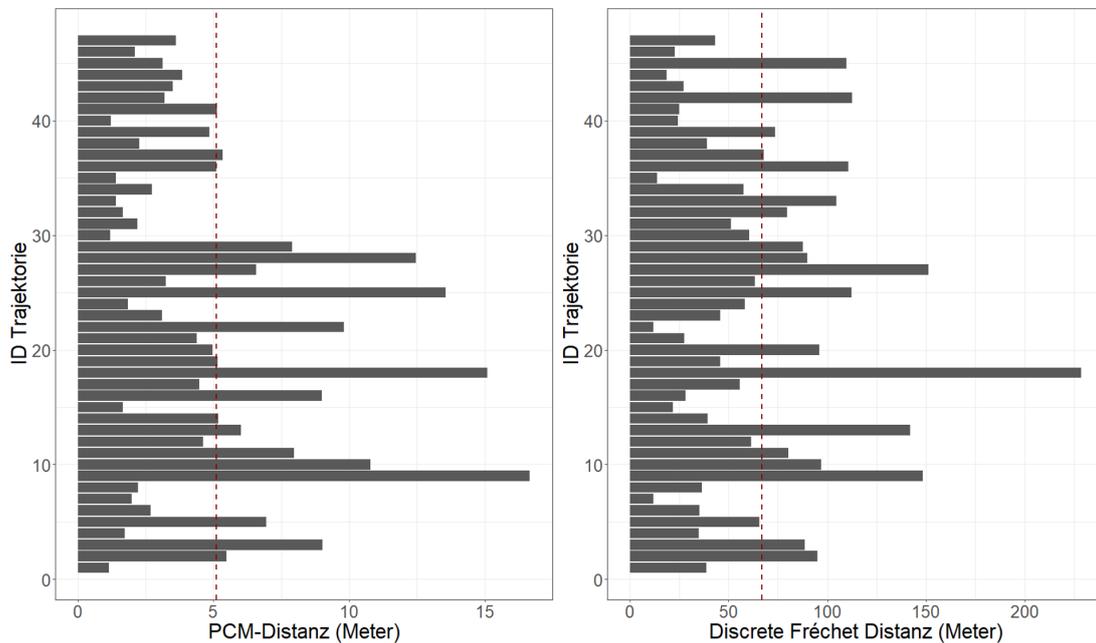


Abbildung 40: Balkendiagramme für die Parameter *PCM* (links) und *DF* (rechts) der einzelnen Trajektorien beim Vergleich der originalen Kurve mit der korrigierten Kurve

Abbildung 40 zeigt zwei Balkendiagramme. Im linken Teil ist die *PCM*-Distanz dargestellt, während das rechte Diagramm die *DF*-Distanz zeigt. Auf der jeweiligen x-Achse sind die unterschiedlichen Trajektorien-IDs zu sehen. Der Mittelwert ist mit einer dunkelrot gestrichelten Linie dargestellt. Dieser liegt für die *PCM*-Distanz bei 5,1 Metern und für die *DF*-Distanz bei 70 Metern. Es kann abgeschätzt werden, dass bei Trajektorien mit hoher *PCM*-Distanz sich auch die *DF*-Distanz im höheren Bereich befindet. Sind die Werte beider Parameter hoch, hat sich der Verlauf der Kurve beim Mapmatching stärker geändert. Die Trajektorien 1, 15, 24, 30 und 40 weisen sehr kleine *PCM*-Distanzen von unter zwei Metern auf, was darauf hinweist, dass deren Observierungen eine Position sehr nahe an den Netzwerkkanten haben. Die *PCM*- und *DF*-Werte haben Aussagekraft hinsichtlich von Ausreiserwerten. Gibt es einzelne Punkte, die sich sehr weit von ihren zugeordneten Netzwerkkanten entfernt befinden, nehmen beide Parameter hohe Werte an. Weitere Informationen über die Gleichheit zweier Kurven lassen sich über die Analyse der Längenunterschiede und der Fläche zwischen den Trajektorien herausfinden. Ergebnisse beider Analysen sind in Abbildung 41 zu sehen.

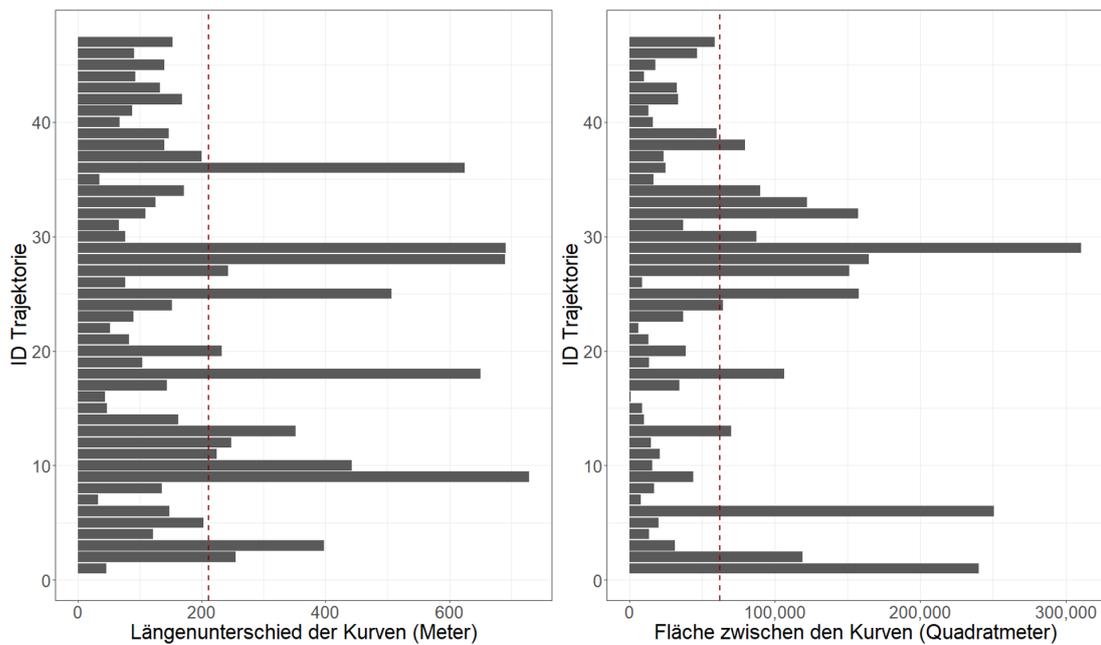


Abbildung 41: Balkendiagramme für den Unterschied der *Kurvenlängen* (links) und die *Fläche zwischen den Kurven* (rechts) der einzelnen Trajektorien beim Vergleich der originalen Kurve mit der korrigierten Kurve

Abbildung 41 zeigt ein Balkendiagramm für die Längenunterschiede der Trajektorien und ein Balkendiagramm für die Fläche zwischen den Kurven. Längenunterschiede existieren von 30 Metern (Trajektorie 7) bis 730 Metern (Trajektorie 9). Diese kommen dadurch zu Stande, dass Trajektorien durch das Mapmatching geglättet werden. Unbearbeitet können Punkte durch ungenaue Lokalisierungen zwischen Positionen hin- und herspringen, während Netzwerkanten in der Regel Geraden sind. Der Mittelwert aller Längenabweichungen beträgt 211 Meter. Die Fläche zwischen den Kurven reicht von 983 Quadratmetern (Trajektorie 16) bis hin zu 310.769 Quadratmetern. Dieser Wert wird maßgeblich von der Länge der Kurven beeinflusst. Eine längere Trajektorie hat eine größere Fläche zwischen den Kurven als eine kürzere Trajektorie mit selbem Abstand zur Vergleichskurve. Der Mittelwert aller Trajektorien-Paare beträgt 62.089 Quadratmeter.

Die Analysen zeigen, dass durch die Berechnung von Vergleichsmaßen Unterschiede zwischen Trajektorien berechnet werden können. Somit kann die Genauigkeit der Positionsbestimmungen eines Sensordatensatzes näherungsweise bestimmt werden. Dabei wird jedoch angenommen, dass die zum Vergleich herangezogene Trajektorie die realen Positionen besser abbildet als die unbearbeitete Kurve. Für den Fall eines mit Mapmatching bearbeiteten Datensatzes kann davon im Regelfall nicht ausgegangen werden.

6.5 Statische Analyse von Messunterschieden

Die verwendeten *OPC-N₃* Feinstaubsensoren zeichnen sich durch eine relativ gute Genauigkeit, verglichen mit hochpräzisen Referenzgeräten, aus. Allerdings muss vor der Arbeit mit den Sensoren sicher gestellt werden, dass die aus unterschiedlichen Geräten abgeleiteten Werte miteinander vergleichbar sind. Während der Produktion von Feinstaubsensoren jeglicher Art kann es zu minimalen Unterschieden der Messkammern oder der Laser-Ausrichtung innerhalb eines Gerätes kommen, weshalb ein Gerät unter gleichen Messvoraussetzungen unterschiedliche Werte wie ein Sensor des selben Typs liefern kann. Deshalb wird im Folgenden evaluiert, ob die fünf verfügbaren *Alpha-sense OPC-N₃*-Sensoren unter gleichen Voraussetzungen gleiche Werte liefern. Darüber hinaus sollen die Messungen mit dem vergleichsweise günstigen *Plantower PMS3003*-Sensor verglichen werden. Der Verlauf der *PM₁*-Werte der sechs Sensoren wird in Abbildung 42 gezeigt.

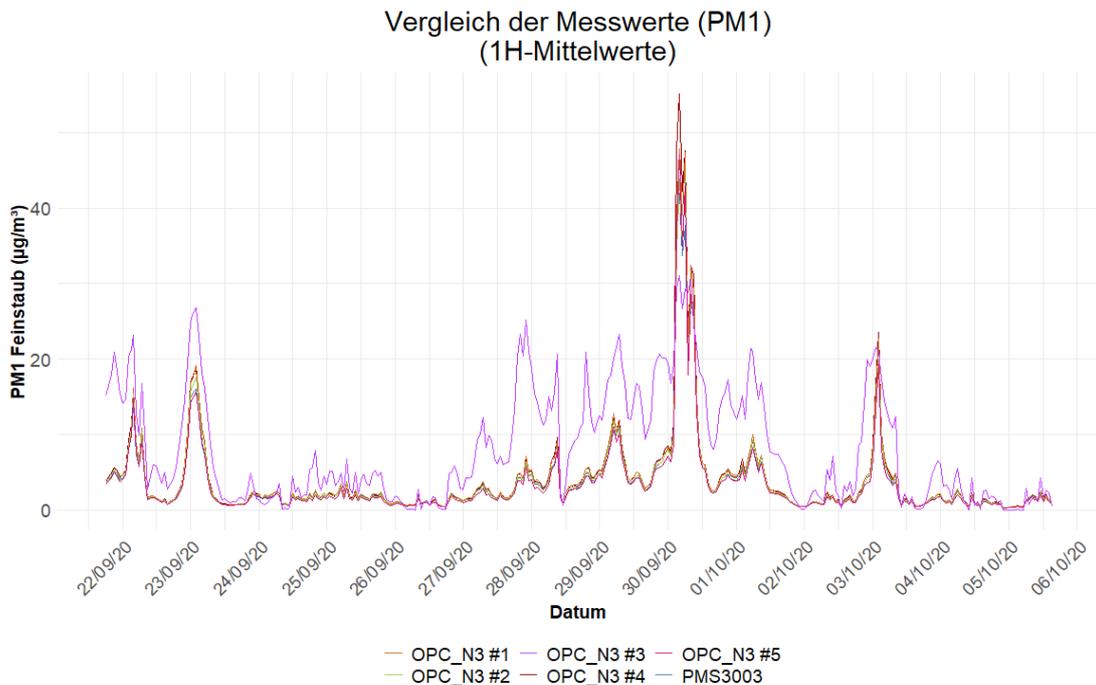


Abbildung 42: Ergebnis des Sensorenvergleichs für *PM₁* Feinstaub (stündliche Mittelwerte)

Abbildung 42 zeigt den Verlauf der Feinstaubwerte für *PM₁*-Partikel. Die X-Achse zeigt den zeitlichen Verlauf der Messungen. Auf der Y-Achse werden die Feinstaubwerte für die Sensoren *OPC-N₃* #1 (dunkelrot), *OPC-N₃* #2 (blau), *OPC-N₃* #3 (braun), *OPC-N₃* #4 (grün), *OPC-N₃* #5 (pink) und *PMS3003* (violett) mithilfe von Linien dargestellt. Die angezeigten Werte sind stündliche Mittelwerte. In der Nacht von 30.09.2020 auf den 01.10.2020 lässt sich ein starker Anstieg der Werte feststellen. Auch im zweiten Zeitraum mit

erhöhter Luftfeuchtigkeit findet sich ein lokales Maximum der PM_1 Feinstaubwerte. Allgemein verlaufen die Kurven der mit $OPC-N_3$ Sensoren gemessenen Werte sehr analog mit nur leicht abweichenden Messungen. Der PMS_{3003} -Sensor gibt deutlich unterschiedliche Werte aus, verglichen mit den $OPC-N_3$ Sensoren. Allerdings befinden sich die lokalen Maxima aller Kurven in der Regel an den gleichen Zeitpunkten, weshalb davon auszugehen ist, dass der PMS_{3003} -Sensor relative Veränderungen des Feinstaubgehaltes in der Luft gut erkennt.

Inwiefern die Sensoren gleiche Werte ausgeben, soll eine Korrelationsanalyse zeigen. Korrelationsmatrizen für die PM_1 -, $PM_{2.5}$ - und PM_{10} -Werte sind in Abbildung 43 dargestellt.

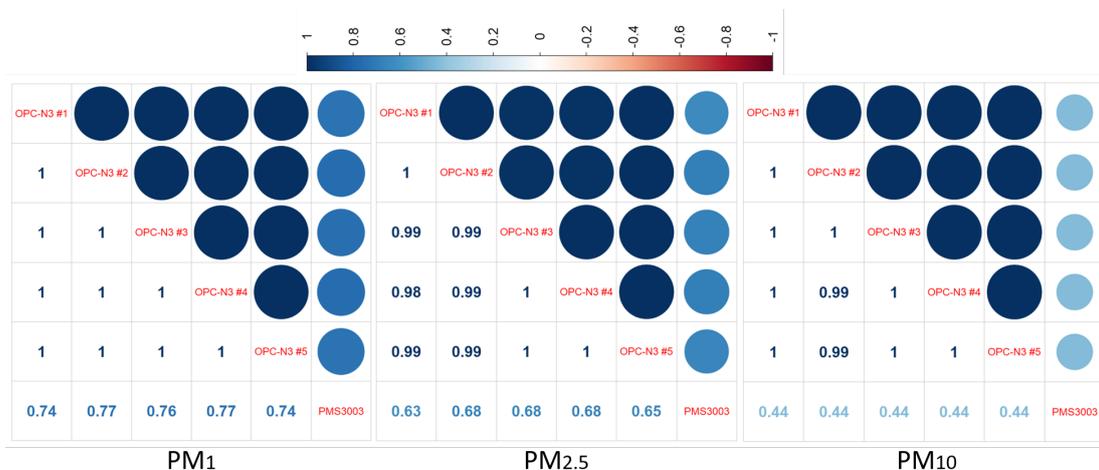


Abbildung 43: Korrelationsmatrizen für fünf $OPC-N_3$ sowie einen PMS_{3003} Sensor für PM_1 , $PM_{2.5}$ und PM_{10} Feinstaub (stündliche Mittelwerte)

Abbildung 43 zeigt drei Korrelationsmatrizen für die PM_1 , $PM_{2.5}$ und PM_{10} -Werte der Feinstaubsensoren. Die Analyse ist statistisch signifikant, da alle p -Werte unter $2,2e^{-16}$ liegen bei einem Signifikanzniveau von 0,05. Die linke Matrix der Abbildung zeigt Korrelationen der mit den Sensoren gemessenen PM_1 -Werten zueinander. Die Werte links der Diagonalen zeigen die *Pearson*-Korrelationsfaktoren. Die Kreise rechts der Diagonalen visualisieren die Faktoren. Alle Darstellungen sind farbkodiert. Dunkelrot dargestellte Werte und Kreise zeigen eine hohe negative Korrelation. Dunkelblaue Werte und Kreise zeigen eine hohe positive Korrelation. Je heller die Farbe in beide Richtungen wird, desto niedriger ist der Korrelationsfaktor. Es ist auffällig, dass alle PM_1 -Werte der $OPC-N_3$ Sensoren zueinander größtmögliche Korrelationsfaktoren vom Wert 1 zeigen. Das bedeutet, dass alle Sensoren zur Zeit der Messung gleiche Feinstaubverläufe aufgezeichnet haben. Der PMS_{3003} -Sensor zeigt mit hohen Korrelationsfaktoren von 0,74 bis 0,77 ebenfalls sehr

analoge Werteverläufe für PM_1 -Feinstaub. Der Sensorenvergleich für $PM_{2.5}$ -Feinstaub wird in Abbildung 44 dargestellt.

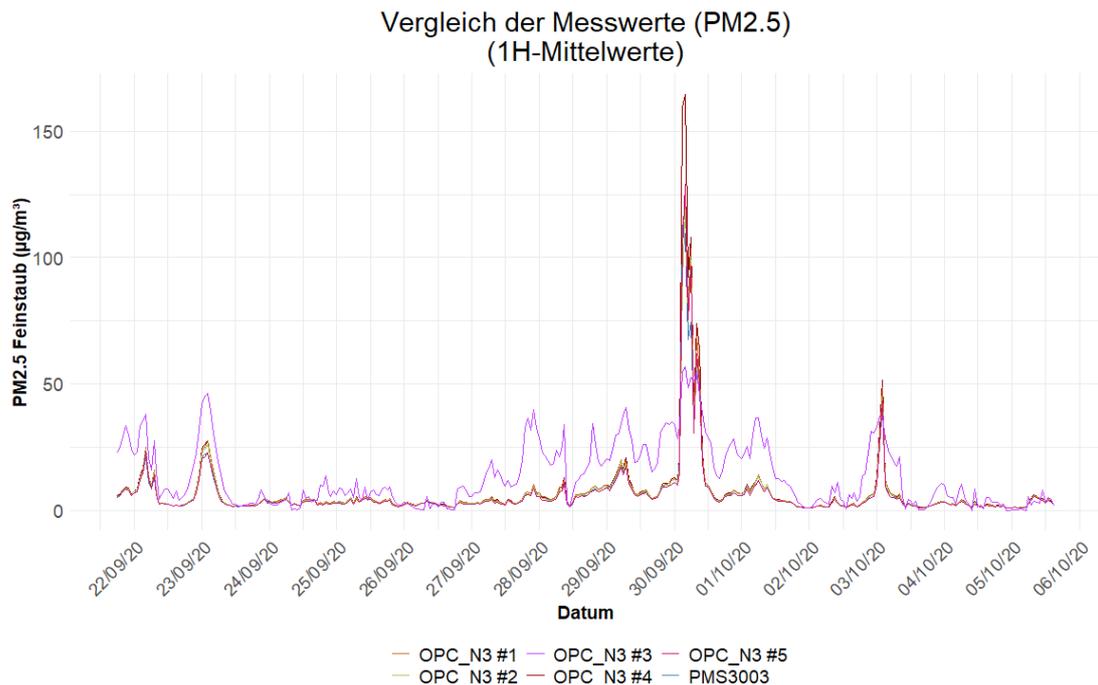


Abbildung 44: Ergebnis des Sensorenvergleichs für $PM_{2.5}$ Feinstaub (stündliche Mittelwerte)

Abbildung 44 zeigt den Verlauf der stündlich gemittelten $PM_{2.5}$ Feinstaubwerte während der Messung. Die X-Achse zeigt den zeitlichen Verlauf der Messungen, während auf der Y-Achse die Höhe der Feinstaubwerte dargestellt wird. Es werden die selben Linienfarben für die jeweiligen Sensoren wie in Abbildung 42 verwendet. Die Kurvenverläufe der Feinstaubwerte zeigen ähnliche Charakteristika wie die der PM_1 -Werte. Alle fünf OPC-N₃-Sensoren liefern analoge Werte, die hoch zu korrelieren scheinen. Der PMS3003-Sensor gibt die lokalen Maxima und Minima der OPC-N₃-Werte gut wieder, zeigt aber abweichende absolute Werte. Der gemessene Maximalwert zwischen dem 30.09.2020 und dem 01.10.2020 fällt auf Abbildung 44 noch deutlicher aus als bei der Analyse des PM_1 -Feinstaubs.

Die Korrelationen der Werte der verschiedenen Sensoren werden in der mittleren Matrix in Abbildung 43 dargestellt. Die Korrelationsindizes links der Diagonalen der Matrix indizieren sehr hohe positive Zusammenhänge zwischen den einzelnen OPC-N₃-Sensoren. Diese liegen jedoch im Gegensatz zu PM_1 nicht mehr alle bei 1, sondern haben einen Wertebereich von 0,98 bis 1. Es lässt sich jedoch festhalten, dass die OPC-N₃-Sensoren auch für $PM_{2.5}$ Feinstaub sehr analoge Werte ausgeben bei einer gleichzeitigen Messung. Die Korrelationsindizes für den PMS3003-Sensor im Zusammenhang

mit den fünf $OPC-N_3$ -Sensoren zeigen niedrigere Zusammenhänge als für den PM_{10} -Datensatz. Die Werte indizieren moderate positive Korrelationen. Ein Wertevergleich für PM_{10} -Feinstaub wird in Abbildung 45 dargestellt.

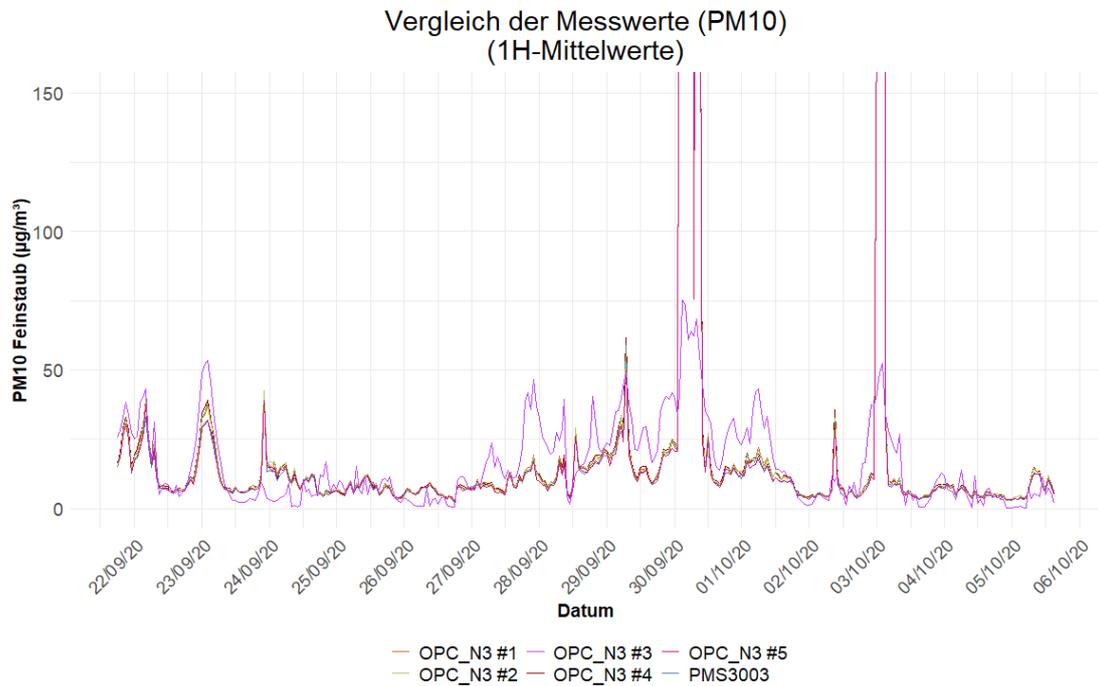


Abbildung 45: Ergebnis des Sensorenvergleichs für PM_{10} Feinstaub (stündliche Mittelwerte)

Abbildung 45 zeigt den Verlauf der Feinstaubwerte für größere PM_{10} -Partikel (stündliche Mittelwerte). Die X-Achse zeigt die zeitliche Skala, während die Y-Achse die Menge der Feinstaubpartikel angibt mit einer Skala von $0 \mu g/m^3$ bis $150 \mu g/m^3$. Werte über $150 \mu g/m^3$ werden nicht dargestellt. Die Ausschläge beider in der Abbildung sichtbaren Maxima reichen bis $1.500 \mu g/m^3$. Beide lokalen Maxima scheinen von der gleichzeitig auftretenden Luftfeuchtigkeit von 100 % beeinflusst zu sein. Abgesehen von diesen beiden Ausschlägen zeigt auch Abbildung 45 analoge Werteverläufe der Feinstaubmenge. Der $PMS3003$ -Sensor gibt die gemessenen lokalen Maxima und Minima der $OPC-N_3$ -Sensoren gut wieder.

Die rechte Matrix in Abbildung 43 zeigt Korrelationen der PM_{10} -Werte der sechs Sensoren. Werte der $OPC-N_3$ -Geräte zueinander indizieren sehr hohe Zusammenhänge mit Korrelationsindizes von 0,99 bis 1. Die Indizes des Zusammenhangs der $PMS3003$ -Werte und der $OPC-N_3$ -Werte von 0,44 indizieren jedoch nur geringe Zusammenhänge. Diese niedrigen Werte könnten den zwei lokalen Maxima der $OPC-N_3$ Sensoren geschuldet sein, denn der $PMS3003$ -Sensor hat lediglich eine relative Änderung der Feinstaubwerte erkannt.

6.6 Analyse der Fahrgeschwindigkeit und Sensorausrichtung

Der Einfluss der Sensorausrichtung und der Fahrgeschwindigkeit soll mithilfe von zwei Experimenten überprüft werden. Zunächst werden die statisch und mobil erfassten Messwerte von Datensatz 2 miteinander verglichen. Im zweiten Experiment werden Messunterschiede der unterschiedlich ausgerichteten Sensoren in Datensatz 3 untersucht.

6.6.1 Vergleich von statisch und mobil erfassten Daten

Um zu überprüfen, ob die Bewegungsgeschwindigkeit und die Sensorausrichtung bei bewegten Feinstaubsensoren einen Einfluss auf die Höhe der gemessenen Werte haben, werden Vergleichsmessungen durchgeführt. Dabei werden die statisch erfassten Messwerte und die mobil aufgenommenen Daten aus Datensatz 2 verglichen. Diese Analyse wird mithilfe eines *R*-Skriptes durchgeführt. Ein Ablaufdiagramm dafür wird in Abbildung 46 gezeigt.

Vergleich von statisch und mobil erfassten Feinstaubsensormesswerten

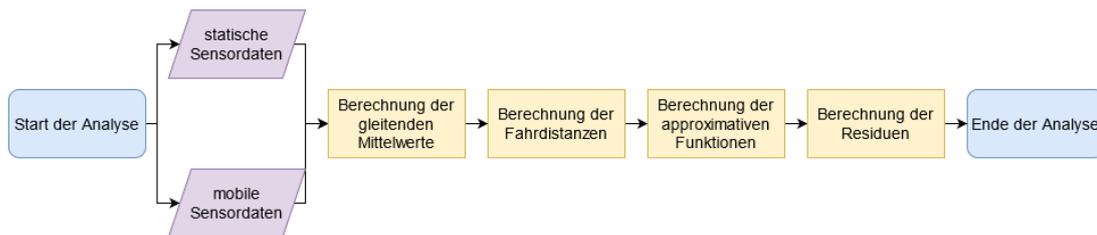


Abbildung 46: Ablaufdiagramm zur Analyse der Messunterschiede statischer und bewegter Sensoren unter Berücksichtigung der Fahrgeschwindigkeit und der Sensorausrichtungen

Abbildung 46 zeigt ein Ablaufdiagramm für die Analyse der Vergleichsmessungen in Datensatz 2. Teil des Datensatzes sind Messwerte zwei statischer und drei mobiler Feinstaubsensoren. Die statischen Messstationen (Sensor 1 und Sensor 2) befinden sich jeweils am Anfang und am Ende der Messstrecke. Auf dem Testfahrzeug befindet sich ein frontal zur Fahrtrichtung, ein seitlich zur Fahrtrichtung und ein entgegen der Fahrtrichtung ausgerichteter Sensor. Zunächst werden gleitende Mittelwerte der Feinstaubkurven berechnet, um den Einfluss plötzlicher Ausschläge der Werte nach oben zu vermindern. Dafür wird eine Berechnungsbreite von vier Werten gewählt. Anschließend wird für jede Datenreihe die jeweilige Fahrdistanz der Observations bestimmt. Bei einer Messstrecken-Länge von 100 Metern bekommt somit der erste Punkt den Wert 0 und der letzte Punkt den Wert 100 zugewiesen. Das ist wichtig, denn so kann beispielsweise auf den mobil erfassten Feinstaub-

wert am Punkt einer statischen Messstation zugegriffen werden. Um das zu bewerkstelligen, müssen außerdem approximative Funktionen AF der Feinstaubkurven bestimmt werden. Mit dem Befehl $AF(15)$ kann somit auf den Feinstaubwert nach 15 Metern Fahrt zugegriffen werden, sollte sich die jeweilige Messstation an diesem Punkt befinden. Im Anschluss werden die Residuen berechnet, die den Messunterschied der statischen Werte zu den mobilen Werten beschreiben. Als Residuen gelten die vertikalen Distanzen von Observationspunkten zu Kurven. Die Observations sind die Messwerte der statischen Sensoren, die mit dem Verlauf der Kurven der mobil erfassten Werte verglichen werden. Die Analyse wird für die Daten der unterschiedlichen Geschwindigkeiten $v = \{5, 10, 20, 30, 40, 50\} km/h$, der unterschiedlichen Sensorausrichtungen $a = \{frontal, seitlich, entgegen\}$ zur Fahrtrichtung und der unterschiedlichen Partikeldurchmesser $PM_x = \{PM_1, PM_{2.5}, PM_{10}\}$ durchgeführt. Ein Beispiel für die Geschwindigkeit $v = 20 km/h$ und den Partikeldurchmesser PM_{10} wird in Abbildung 47 gezeigt.

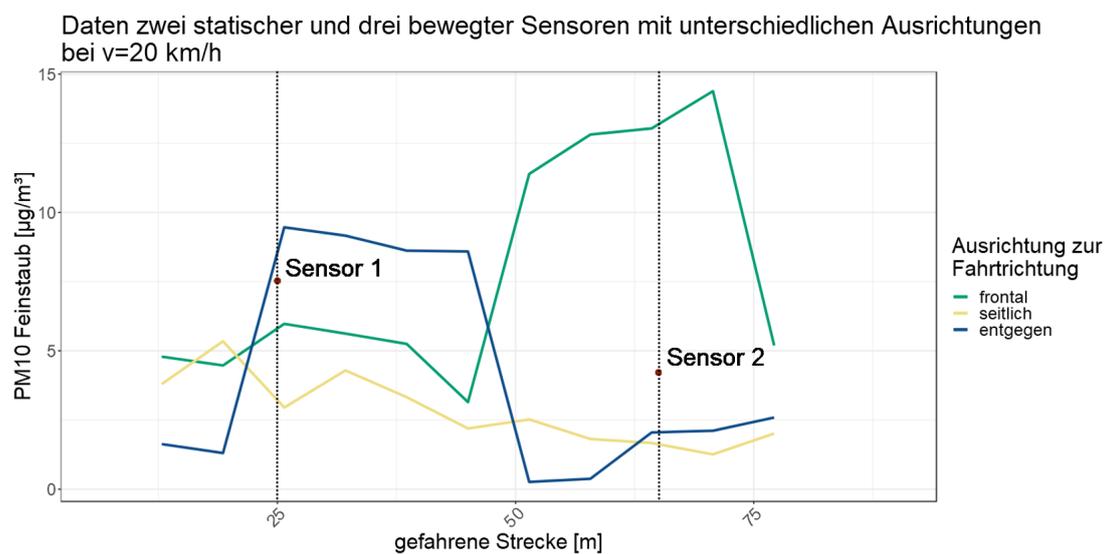


Abbildung 47: Vergleichsmessung (PM_{10}) von zwei statischen Sensoren (Sensor 1 und Sensor 2, rote Punkte) und drei bewegten Sensoren (frontal, seitlich und entgegen der Fahrtrichtung) bei einer Geschwindigkeit von 20 km/h

Abbildung 47 zeigt ein Liniendiagramm mit den Feinstaubkurven der mobilen Sensoren (*frontal* als grüne Linie, *seitlich* als gelbe Linie und *entgegen* als blaue Linie). Ebenfalls dargestellt sind die Werte der Sensoren *Sensor 1* und *Sensor 2* als rote Punkte. Diese befinden sich an der Stelle, an der das Testfahrzeug die Messstationen passiert. Die Werte der mobilen Sensoren, die frontal und entgegen der Fahrtrichtung ausgerichtet sind, zeigen große Messunterschiede im Laufe der Fahrt. Der seitlich ausgerichtete Sensor zeigt eine leichte und stetige Abnahme der Messwerte während dem Befahren der Teststrecke.

Die gestrichelten vertikalen Linien in Abbildung 47 helfen dabei, den Messunterschied zwischen den statischen und mobilen Sensoren festzustellen. Am Schnittpunkt der Hilfslinie mit der jeweiligen Feinstaubkurve befindet sich der Vergleichswert zur Kalkulation des Messunterschiedes. Dieser stellt sich für das Beispiel in der Abbildung als hoch dar. Die Unterschiede für *Sensor 1* betragen zwischen 2 und 5 $\mu\text{g}/\text{m}^3$. *Sensor 2* zeigt noch höhere Diskrepanzen mit Werten zwischen 4 und 16 $\mu\text{g}/\text{m}^3$. Das Beispiel indiziert, dass die Bewegungsgeschwindigkeit und die Sensorausrichtung einen Einfluss auf die Qualität der Messergebnisse nehmen. Eine Übersicht der Residuen für den kompletten Datensatz wird in Tabelle 11 gegeben.

a & PM_x \ v [km/h]	5	10	20	30	40	50	\emptyset
frontal & PM_1	0,33	0,27	1,45	0,53	1,27	0,82	0,78
seitlich & PM_1	0,72	0,33	0,71	1,28	0,82	1,64	0,92
entgegen & PM_1	1,07	1,04	0,49	0,29	0,25	0,84	0,66
frontal & $PM_{2.5}$	2,27	0,73	2,53	4,47	6,96	2,82	3,25
seitlich & $PM_{2.5}$	1,97	1,84	1,82	2,15	1,96	7,60	2,89
entgegen & $PM_{2.5}$	4,26	4,02	2,84	0,97	3,89	5,19	3,53
frontal & PM_{10}	13,93	14,94	5,20	37,83	15,46	4,39	15,29
seitlich & PM_{10}	3,56	8,68	2,94	4,11	3,30	258,61	46,86
entgegen & PM_{10}	16,11	20,32	2,15	11,81	14,55	8,31	12,21
\emptyset	4,91	5,79	2,24	7,05	5,38	32,25	

Tabelle 11: Messunterschiede bei zeitgleicher statischer und mobiler Datenerfassung bei unterschiedlichen Geschwindigkeiten. Aufschlüsselung nach Sensorausrichtung a (frontal, seitlich und entgegen der Fahrtrichtung) und Partikeldurchmesser PM_x (PM_1 , $PM_{2.5}$, PM_{10}). Farbliche Hervorhebung des jeweils **niedrigsten** und **höchsten** Wertes einer Kombination.

Tabelle 11 zeigt die Messunterschiede unter der Berücksichtigung aller Kombinationen aus Bewegungsgeschwindigkeit v , Sensorausrichtung a und Partikeldurchmesser PM_x . Die jeweils niedrigsten und höchsten Unterschiede einer Kombination aus a und PM_x werden in **grüner** und **roter** Farbe hervorgehoben. Tendenziell zeigt ein Großteil der Kombinationen die höchsten Abweichungen bei 40 und 50 km/h und die niedrigsten Abweichungen bei 10 und 20 km/h. Das ist allerdings nicht immer der Fall, denn beispielsweise hat die Kombination *frontal & PM_{10}* die geringste Abweichung bei der höchsten Bewegungsgeschwindigkeit. Ein Trend lässt sich bei Aufschlüsselung der Werte nach dem Partikeldurchmesser ablesen. Hier zeigen die Kombinationen mit PM_1 die geringsten Abweichungen und die PM_{10} -Werte die höchsten. Allerdings haben jedoch die Feinstaubkurven ein unterschiedliches Wertenniveau, was einen Schluss über die absoluten Abweichungen nicht zulässt. Hin-

sichtlich der Bewegungsgeschwindigkeit zeigen die Werte bei $v = 50\text{km/h}$ im Schnitt die höchsten Abweichungen. Die Messunterschiede sind, aufgeschlüsselt nach der Sensorausrichtung, im Durchschnitt am höchsten, wenn der Sensor seitlich ausgerichtet ist. Die Aussagen hinsichtlich der Bewegungsgeschwindigkeit und der Sensorausrichtung sind stark beeinflusst von dem extrem hohen Wert bei *seitlich & PM₁₀* und $v = 50\text{km/h}$. In Folge dessen kann man aus der Analyse keinen allgemeinen Schluss hinsichtlich des genauen Einflusses von Geschwindigkeit und Ausrichtung ziehen. Tabelle 11 zeigt allerdings klare Indizien dafür, dass beide Größen die Qualität der Messungen beeinflussen.

6.6.2 Vergleich verschieden ausgerichteter Sensoren

Der aus Versuchsaufbau 3 entstandene Datensatz enthält unter anderem Feinstaubmessungen, die von zwei unterschiedlich ausgerichteten Sensoren stammen. Der erste Sensor ist nach vorne ausgerichtet, wobei der Lufteinlass entgegen dem Fahrtwind steht. Der zweite Sensor ist zur Seite ausgerichtet. Wie in Kapitel 6.5 dargestellt, geben die Sensoren unter identischen Messbedingungen nahezu die gleichen Werte ab. Korrelieren ebenfalls die Werte des seitlich ausgerichteten Sensors mit den Werten des frontal ausgerichteten Sensors in Datensatz 3, kann davon ausgegangen werden, dass die Ausrichtung von Sensoren während dem Fahren keinen Einfluss auf die Qualität der Daten nimmt. Für die Überprüfung dieser Aussage wird ein R-Skript genutzt. Der Ablauf der Analyse sieht vor, für jeden Feinstaubwert zunächst ein Tupel aus dem jeweiligen Wert des frontal beziehungsweise seitlich ausgerichteten Sensors zu bilden. Berücksichtigt werden die Feinstaubwerte PM_1 , $PM_{2.5}$, PM_{10} , die Feinstaubbins bin_0, \dots, bin_{23} sowie die Temperatur und die Luftfeuchtigkeit. Anschließend wird für die beiden Werte jedes Tupels die Korrelation mithilfe einer *Pearson*-Korrelationsanalyse berechnet.

Abbildung 48 zeigt ein Punktdiagramm mit den Korrelationswerten für jedes Wertetupel des Datensatzes. Ein Tupel besteht aus dem jeweiligen Wert des frontal ausgerichteten Feinstaubsensors und dem jeweiligen Wert des seitlich ausgerichteten Feinstaubsensors. Die X-Achse zeigt die verschiedenen Variablen, während auf der Y-Achse der dazugehörige Korrelationswert dargestellt wird. Die Werte der X-Achse sind so sortiert, dass die höchste Korrelation zuerst und die niedrigste Korrelation zuletzt gezeigt wird. Die p -Werte der Korrelationsanalyse liegen im Schnitt bei weniger als $2,2e^{-16}$. Bei der Annahme eines Signifikanzniveaus von 0,05 gelten die Analysen somit als statistisch signifikant. Beim Betrachten der Korrelationswerte fällt primär auf,

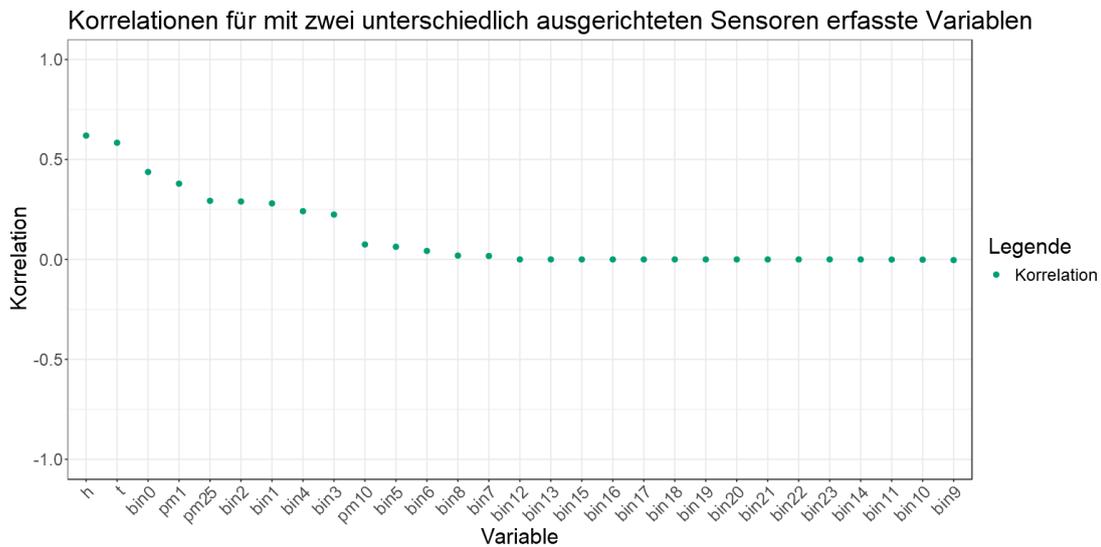


Abbildung 48: Korrelationen der Messwerte aus dem nach vorne ausgerichteten Sensor zu den dazugehörigen Werten aus dem seitlich ausgerichteten Sensor

dass es höchstens zu moderat positiven Korrelationen kommt. Der höchstmögliche Wert liegt bei 0,6 für die Luftfeuchtigkeit. Jegliche Feinstaubwerte zeigen Ergebnisse von unter 0,5, der Großteil der untersuchten Werte weist darüber hinaus überhaupt keine Zusammenhänge auf. Verglichen mit den Ergebnissen aus Kapitel 6.5 muss festgehalten werden, dass trotz gleicher Umgebungsbedingungen bei verschiedenen Sensorausrichtungen in Bewegung keine Wertegleichheit erwartet werden darf.

Um den Einfluss der Fahrgeschwindigkeit noch weiter zu untersuchen, zeigt Abbildung 49 die Korrelationen nur für Punkte mit einer hohen Bewe-

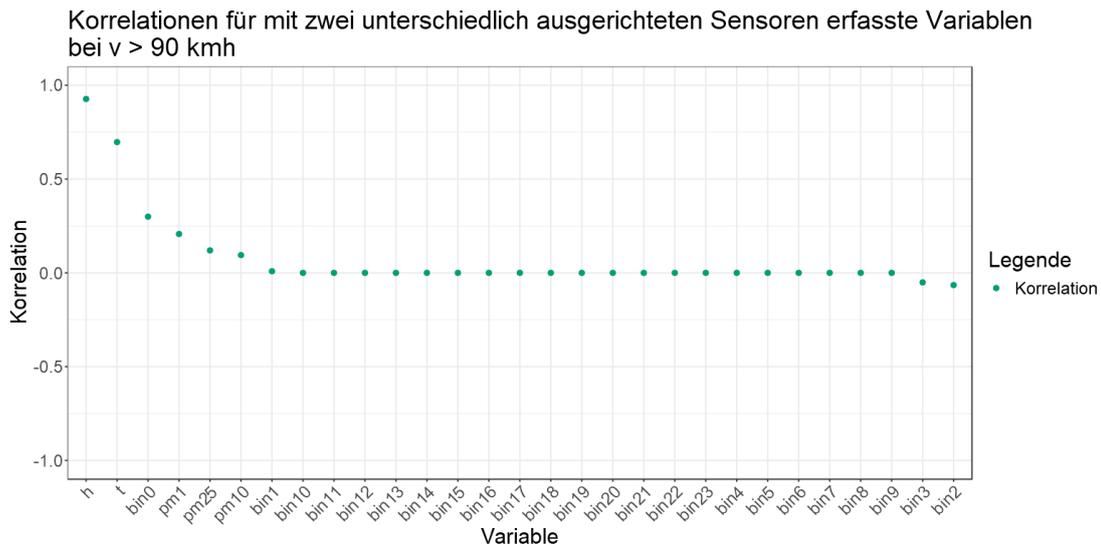


Abbildung 49: Korrelationen der Messwerte aus dem nach vorne ausgerichteten Sensor zu den dazugehörigen Werten aus dem seitlich ausgerichteten Sensor bei $v > 90\text{km/h}$

gungsgeschwindigkeit > 90 Kilometer pro Stunde. Die Korrelationsanalysen sind erneut statistisch signifikant bei einem durchschnittlichen p -Wert von unter $2,2e^{-16}$ und einem Signifikanzniveau von $0,05$. Es werden nur Observa-tionen mit einer Bewegungsgeschwindigkeit von über 90 Stundenkilometern berücksichtigt. Zunächst fällt auf, dass die Temperatur und die Luftfeuch-tigkeit eine höhere Korrelation aufweisen als bei der Analyse des gesamten Datensatzes. Jegliche Feinstaubwerte korrelieren jedoch weniger und in mehr als der Hälfte der Fälle überhaupt nicht. Das ist ein Indiz dafür, dass die Rolle der Sensorausrichtung bei höheren Geschwindigkeiten zunimmt.

6.7 Analyse witterungsbedingter Einflüsse

Die Luftfeuchtigkeit in der Umgebung einer Feinstaubmessung kann für abgefälschte Werte sorgen. Wasserpartikel heften sich an Feinstaubpartikel an und lassen die Werte durch veränderte Reflektionseigenschaften in der Messkammer höher erscheinen, als sie in Wahrheit sind. Die Rolle der Witte-rungsbedingungen auf gemessene Feinstaubwerte soll in zwei Schritten un-tersucht werden. Zunächst wird Datensatz 1 auf Zusammenhänge zwischen dem Luftfeuchtigkeitsverlauf und der Höhe der gemessenen Feinstaubwer-te untersucht. Danach werden unterschiedliche Korrekturalgorithmen auf die Feinstaubwerte von Datensatz 3 angewandt, um die Auswirkung der in der Literatur vorgeschlagenen Faktoren zu untersuchen. Ein Vergleich des Verlau-fes der Feinstaubwerte mit der Entwicklung der Luftfeuchtigkeit in Datensatz 1 wird in Abbildung 50 dargestellt.

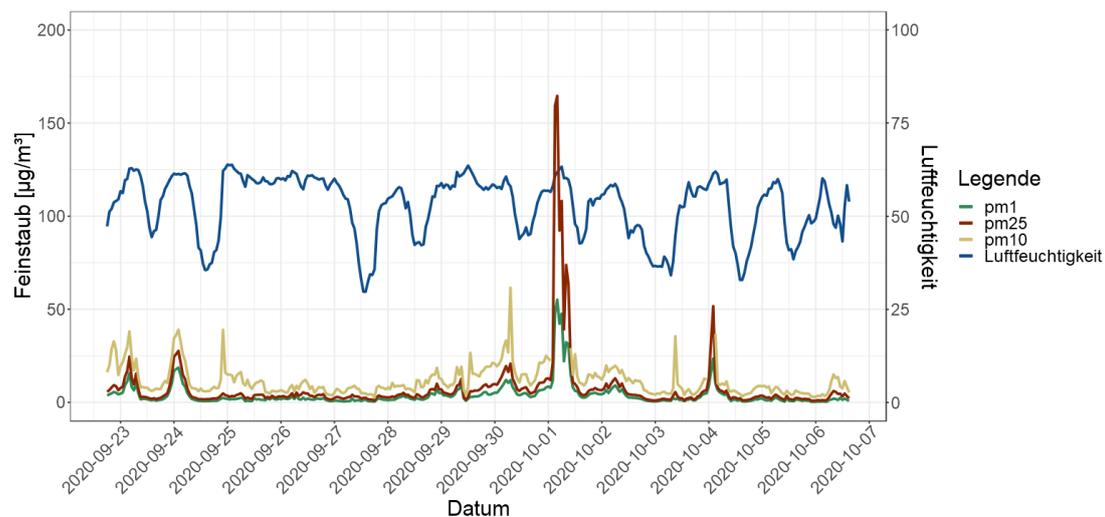


Abbildung 50: Vergleich der PM_1 , $PM_{2.5}$ und PM_{10} Feinstaubwerte mit dem Luftfeuchtigkeitsverlauf von Datensatz 1

Abbildung 50 zeigt ein Liniendiagramm mit den Feinstaubwerten PM_1 (grüne Linie), $PM_{2.5}$ (dunkelrote Linie), PM_{10} (dunkelgelbe Linie) und der Luftfeuchtigkeit (blaue Linie) von Datensatz 1. Die X-Achse zeigt den zeitlichen Verlauf der Messungen. Auf der linken Y-Achse sind die Feinstaubwerte in $\mu\text{g}/\text{m}^3$ dargestellt und auf der rechten Y-Achse die relative Luftfeuchtigkeit in Prozent. Die Feinstaubwerte PM_1 , $PM_{2.5}$ und PM_{10} zeigen zwar unterschiedlich extreme Ausschläge nach oben und unten, haben jedoch die lokalen Maxima und Minima an identischen Stellen. Über weite Strecken hinweg zeigt die Luftfeuchtigkeitskurve einen ähnlichen Verlauf. Bei lokalen Maxima der Feinstaubwerte indiziert das Diagramm ebenfalls höhere Feuchtigkeitswerte. Besonders auffällig ist dieses Phänomen am 23.09.2020, am 24.09.2020, am 04.10.2020 und zwischen dem 01.10.2020 und dem 02.10.2020. Die Gemeinsamkeiten sind für die PM_{10} -Werte deutlicher sichtbar als für PM_1 und $PM_{2.5}$. Das ergibt Sinn, denn wenn sich Wasserpartikel an Feinstaubpartikel anheften, erscheint der Partikeldurchmesser größer als er in Wahrheit ist. Nicht alle Ausschläge lassen sich jedoch mit einem Anstieg oder einem Abstieg der Luftfeuchtigkeit in Verbindung bringen. Das gilt insbesondere für die Feinstaubwerte zwischen dem 25.09.2020 und dem 29.09.2020. Auch der besonders ausgeprägte Anstieg der Feinstaubwerte zwischen dem 01.10.2020 und dem 02.10.2020 lässt sich nicht über einen besonders hohen Luftfeuchtigkeitswert erklären. Zusammengefasst zeigt die Darstellung in Abbildung 50 starke Indizien eines Zusammenhangs zwischen der Höhe der gemessenen Feinstaubwerte und der relativen Luftfeuchtigkeit. Allerdings lassen sich nicht alle Wertentwicklungen durch die Luftfeuchtigkeit erklären, was entweder für eine natürliche Entwicklung der Feinstaubwerte oder für die Präsenz weiterer Einflussfaktoren spricht.

Um Unstimmigkeiten hinsichtlich des Einflusses der Luftfeuchtigkeit zu korrigieren, lässt sich der trockene Feinstaubwert PM_{dry} berechnen. Dafür wird der feuchte Feinstaubwert PM_{wet} durch einen Wachstumsfaktor gf geteilt, der maßgeblich vom zum Zeitpunkt der Messung aufgenommenen Luftfeuchtigkeitswert beeinflusst ist. Streibl (2017) nennen vier mögliche Funktionen zur Berechnung des Wachstumsfaktors: $gf_{Hänel}$, gf_{Soneja} , gf_{Combo} und gf_{Skupin} . Um deren Einfluss auf die Feinstaubwerte des Datensatzes zu untersuchen, wird anhand der Wachstumsfunktionen der jeweilige trockene Feinstaubwert PM_{dry} bestimmt. Die Analyse wird mit R durchgeführt. Der Ablauf der Analyse wird in Abbildung 51 gezeigt.

Ablauf der Luftfeuchtigkeitskorrektur

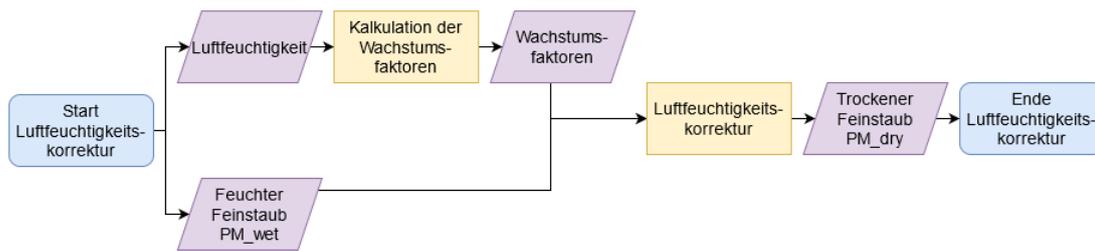


Abbildung 51: Ablaufdiagramm für die Luftfeuchtigkeitskorrektur der Feinstaubwerte

Abbildung 51 zeigt ein Ablaufdiagramm für die Korrektur der Feinstaubwerte hinsichtlich gemessener Luftfeuchtigkeitswerte. Dafür werden die Luftfeuchtigkeit h sowie die jeweiligen Feinstaubwerte PM_1^{wet} , $PM_{2.5}^{wet}$ und PM_{10}^{wet} benötigt. Mithilfe der Luftfeuchtigkeitswerte werden anhand der in Kapitel 3.5 vorgestellten Formeln die Wachstumsfaktoren $gf_{Hänel}$, gf_{Soneja} , gf_{Combo} und gf_{Skupin} eines jeden Datenpunktes berechnet. Aus den Wachstumsfaktoren und den PM_{wet} Werten werden im Anschluss die trockenen Feinstaubwerte PM_1^{dry} , $PM_{2.5}^{dry}$ und PM_{10}^{dry} kalkuliert. Die Ergebnisse für eine Trajektorie (ID 17) mit relativ hohen Luftfeuchtigkeitswerten (durchschnittlich 61%) werden in Abbildung 52 gezeigt.

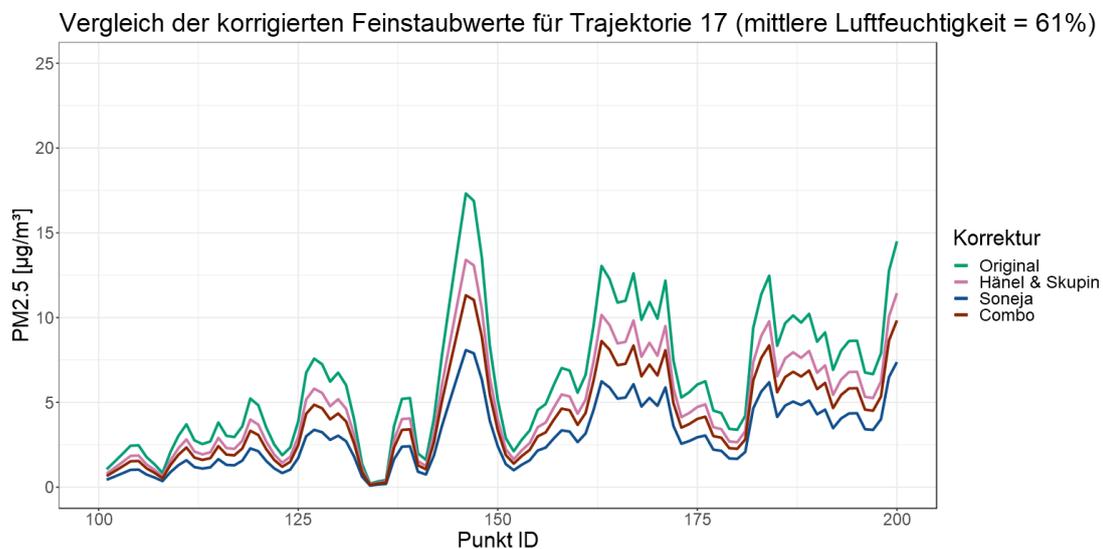


Abbildung 52: Vergleich der Feinstaubwerte ($PM_{2.5}$) nach Anwendung unterschiedlicher Wachstumsfunktionen (Hänel, Soneja, Combo, Skupin) mit den originalen Werten für Trajektorie 17 bei einer durchschnittlichen Luftfeuchtigkeit von 61%

Abbildung 52 zeigt ein Liniendiagramm für originale und korrigierte $PM_{2.5}$ Feinstaubwerte eines Ausschnitts von Trajektorie 17. Die unbearbeiteten Feinstaubwerte (PM_{wet}) sind mit einer grünen Linie dargestellt. Die korrigierten Feinstaubwerte PM_{dry} sind dargestellt mit einer magentafarbenen (Hänel und

Skupin), blauen (Soneja) und dunkelroten (Combo) Linie. Die unkorrigierten Werte sind stets höher als die korrigierten. Folglich ist der Wachstumsfaktor immer höher als der Wert 1. Den größten Einfluss auf die Werte hat die Wachstumsfunktion gf_{Soneja} , die die Feinstaubwerte in einigen Fällen mehr als halbiert. $gf_{Hänel}$ und gf_{Skupin} produzieren identische Feinstaubwerte. Besonders bei lokalen Maxima (zum Beispiel Punkt ID 140 und 185) fällt auf, dass die Korrekturen bei höheren Feinstaubwerten schwergewichtiger sind als bei niedrigen. Abbildung 52 zeigt Ergebnisse für eine Trajektorie mit einer relativ hohen Luftfeuchtigkeit von 61 Prozent. Als Gegenbeispiel sind in Abbildung 53 die Ergebnisse für eine Trajektorie mit einer Luftfeuchtigkeit von lediglich 20 Prozent zu sehen.

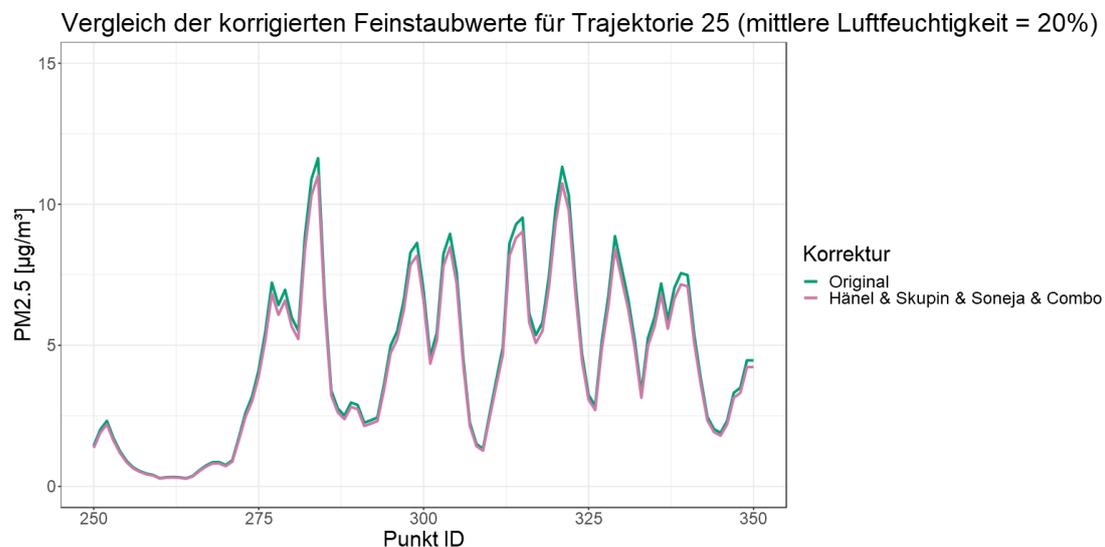


Abbildung 53: Vergleich der Feinstaubwerte ($PM_{2.5}$) nach Anwendung unterschiedlicher Wachstumsfunktionen (Hänel, Soneja, Combo, Skupin) mit den originalen Werten für Trajektorie 25 bei einer durchschnittlichen Luftfeuchtigkeit von 20%

Abbildung 53 zeigt ein Liniendiagramm mit den originalen und korrigierten Feinstaubwerten für Trajektorie 25. Die mittlere Luftfeuchtigkeit liegt bei 20%, folglich sind kaum Korrekturen durch die Wachstumsfunktionen erkennbar. Bei den lokalen Maxima der originalen Feinstaubkurve (Punkt IDs 288, 298, 305, 317, 321, 335) zeigen die mit den Wachstumsfaktoren bearbeiteten Kurven leicht geringere Werte. Abgesehen davon haben alle Kurven nahezu identische Verläufe. Unterschiede sind zwischen den verschiedenen Wachstumsfunktionen $gf_{Hänel}$, gf_{Soneja} , gf_{Combo} und gf_{Skupin} , dargestellt mit einer magentafarbenen Linie, nicht auszumachen.

Die Verläufe der Kurven in den Abbildungen 52 und 53 zeigen, dass die Luftfeuchtigkeitskorrekturalgorithmen bei höheren Feuchtigkeitswerten einen deutlich größeren Einfluss auf die Feinstaubwerte nehmen als bei relativ niedrigen

Feuchtigkeitswerten. Außerdem ist zu erkennen, dass der Einfluss auf hohe Feinstaubwerte in absoluten Zahlen deutlicher sichtbar ist als auf niedrige Feinstaubwerte. Die Wachstumsfunktion gf_{Soneja} nimmt den größten Einfluss auf die Feinstaubwerte. $gf_{Hänel}$ und gf_{Skupin} produzieren für einen α -Wert von 1 nach Chakrabarti et al. (2004) identische Ergebnisse.

6.8 Analyse des Umgebungsverkehrs

Zu Störfaktoren innerhalb des Umgebungsverkehrs gehören Feinstaub emittierende Fahrzeuge, die kleinräumig einen direkten Einfluss auf Messergebnisse nehmen. Wenn ein Fahrzeug im Verkehrsfluss anderen Fahrzeugen hinterher fährt, treten einige Effekte auf, die die Messergebnisse von mobil eingesetzten Feinstaubsensoren beeinflussen. Diese Effekte verfälschen die Messergebnisse zwar nicht maßgeblich, jedoch treten hierdurch lokale, kleinräumige, Unterschiede der Messungen auf. Somit misst ein Fahrzeug im Verkehrsfluss durch den Einfluss des Umgebungsverkehrs höhere Feinstaubwerte als ein Fahrzeug, das sich ohne vorausfahrende KFZ auf einer Straße bewegt. Vorangegangene Forschung (Röger et al. (2020a), Röger (2020)) zeigt, dass auch Feinstaubmessungen eine mögliche Datenquelle für die Implementierung eines Floating Car Observers sein können. Da die Korrelationsanalyse in Röger (2020) indiziert, dass die Feinstaubmessungen eines Testfahrzeuges mit der Anzahl der sich vor diesem Automobil befindlichen Fahrzeuge in Relation steht, scheint eine Modellierung dieses Zusammenhangs möglich. Zur Implementierung dieses Modells wird der auf maschinellem Lernen basierende Random Forest Algorithmus verwendet. Die Feinstaubwerte PM_1 , $PM_{2.5}$ und PM_{10} sowie die Werte der einzelnen Feinstaubbins bin_0, \dots, bin_{23} fungieren als Prediktoren innerhalb des Modells. Sie prognostizieren den Boole'schen Parameter *Umgebungsverkehr* mit dem Wert 0 für *keine vorausfahrenden Fahrzeuge* sowie dem Wert 1 für *ein oder mehrere vorausfahrende Fahrzeuge*. Die Analyse wird mithilfe eines R-Skriptes durchgeführt. Ein Ablaufdiagramm für die Analyse wird in Abbildung 54 gezeigt.

Implementierung eines Floating Car Observers auf Basis von Feinstaubwerten

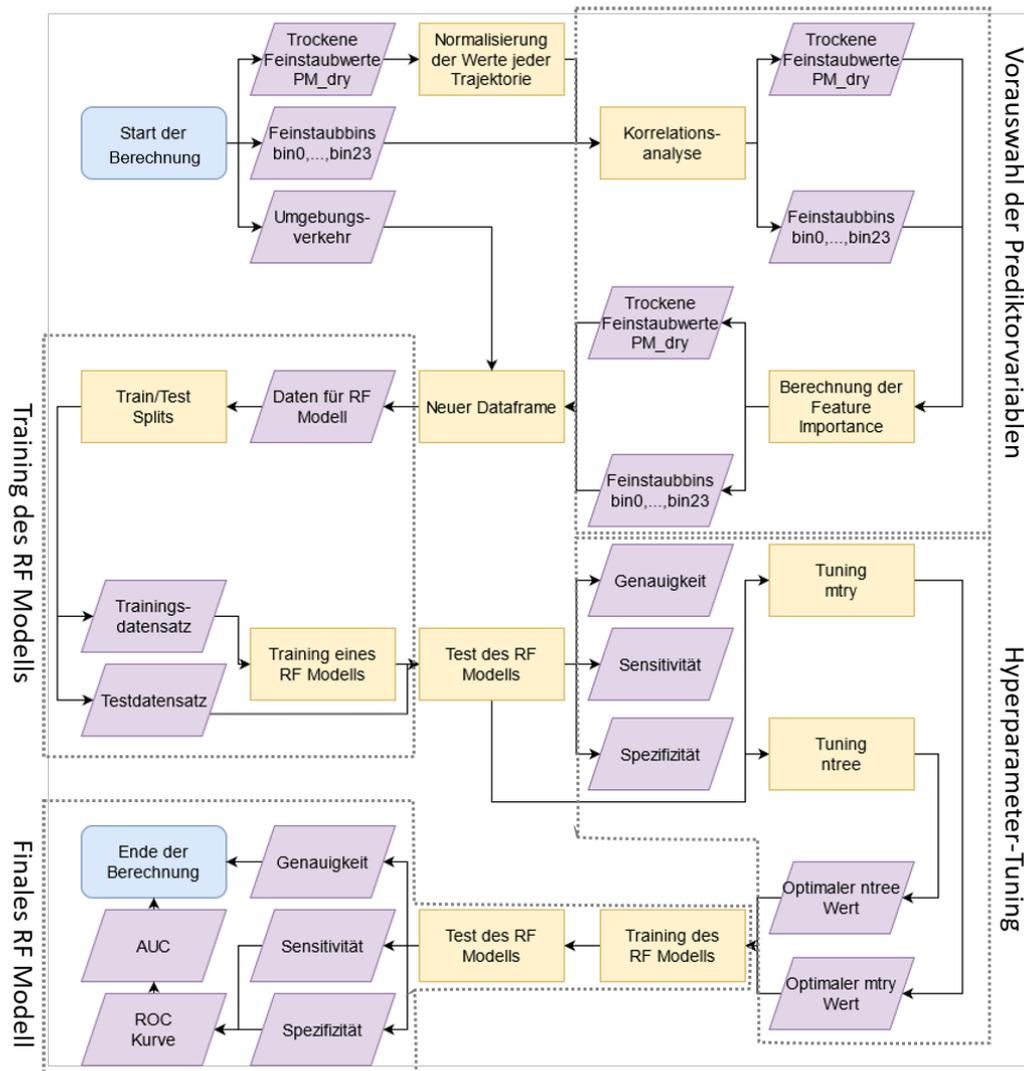


Abbildung 54: Ablaufdiagramm für die Implementierung eines FCO auf Basis von Feinstaubwerten

Abbildung 54 zeigt den Ablauf der Analyse. Zunächst werden die Feinstaubwerte PM_1 , $PM_{2.5}$ sowie PM_{10} vorprozessiert. Dafür werden die mit der Hänel Wachstumsfunktion korrigierten Werte verwendet. Diese werden normalisiert, um die Feinstaubwerte aller Trajektorien auf den gleichen Wertebereich von 0 bis 1 zu bringen. Das ist nötig, denn die mittlere Feinstaubkonzentration in der Umgebungsluft jeder Trajektorie ist unterschiedlich groß. Eine Trajektorie mit niedriger Feinstaubkonzentration könnte somit bei existierendem Umgebungsverkehr höhere Werte aufweisen als eine Trajektorie mit hoher Feinstaubkonzentration bei nicht existierendem Umgebungsverkehr. Nach der Normalisierung werden die Feinstaubwerte mit den Partikel-

zählungen aus bin_0, \dots, bin_{23} auf Korrelationen untersucht. Liegen starke Zusammenhänge vor, ist das wichtig für die Interpretation der Ergebnisse. Bei perfekter Korrelation zweier Datenreihen kann keine der beiden Variablen als wichtiger als die andere identifiziert werden. Nach der Korrelationsanalyse folgt die Berechnung der *Feature Importance* der Prediktoren. Diese gibt an, welche Variablen einen höheren Einfluss auf die Ergebnisse eines Random Forest Modells haben und welche einen niedrigeren. Variablen mit einer besonders niedrigen *Feature Importance* werden nicht weiter verwendet, um den Rechenaufwand des Modells zu reduzieren. Mit den Prediktoren der Feinstaubwerte $PM_1, PM_{2.5}, PM_{10}$ und den Partikelzählungen bin_0, \dots, bin_{23} sowie den Daten des Umgebungsverkehrs wird ein neuer Dataframe erstellt. Dieser wird in ein Trainings- und einen Testdatensatz aufgeteilt. Getestet werden die Split Optionen 50/50, 60/40 und 70/30. Der Trainingsdatensatz wird zum Anlernen eines RF Modells verwendet. Für die Hyperparameter n_{tree} und m_{try} werden zunächst die initialen Werte 500 und \sqrt{p} (mit der Anzahl der Prediktoren p) gewählt, die in weiteren Iterationen optimiert werden sollen. Das Modell wird im Anschluss mit dem Testdatensatz validiert. Parameter zur Überprüfung der Leistungsfähigkeit des Modells sind die *Genauigkeit*, die *Sensitivität* und die *Spezifität*. Danach findet ein Hyperparameter-Tuning statt für die Werte n_{tree} und m_{try} . Die daraus resultierenden optimierten Werte werden für ein weiteres Training des Modells verwendet. Nach dem erneuten Test des Modells werden wieder die *Genauigkeit*, die *Sensitivität* und die *Spezifität* ausgegeben. Dazu wird aus der *Sensitivität* und der *Spezifität* eine ROC-Kurve generiert, die im Anschluss auch den Parameter *AUC* (Fläche unter der Kurve) liefert. Abschließend werden die Ergebnisse des Modells interpretiert. Zunächst wird eine Auswahl unter den Prediktorvariablen $PM_1, PM_{2.5}, PM_{10}$ und bin_0, \dots, bin_{23} getroffen. Dafür kommt eine Korrelations-, sowie eine *Feature Importance* Analyse zum Einsatz. Untersucht werden dabei lediglich die Feinstaubwerte $PM_1, PM_{2.5}, PM_{10}$ und die Partikelzählungen bin_0, \dots, bin_{14} sowie bin_{18} und bin_{19} . Die verbliebenen bin_x -Werte weisen über alle Messungen hinweg einen Durchschnittswert von 0 auf und werden deshalb für das RF-Modell nicht weiterverwendet. Die Korrelationen zwischen den Prediktorvariablen werden aufgrund der Menge an zu visualisierenden Werten mithilfe eines Korrelationsnetzwerkes dargestellt, das in Abbildung 55 gezeigt wird.

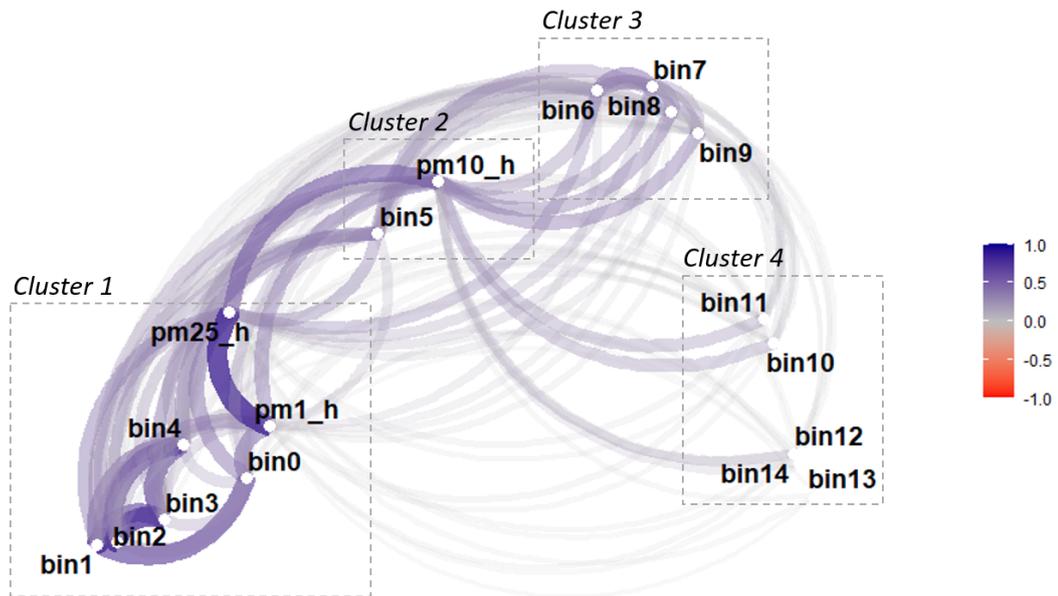


Abbildung 55: Korrelationsnetzwerk für die Prediktoren des Random Forest Modells

Abbildung 55 zeigt ein Korrelationsnetzwerk für die Prediktorvariablen des Random Forest Modells. Dargestellt sind *Pearson*-Korrelationswerte für die Feinstaubwerte PM_1 , $PM_{2.5}$, PM_{10} und die Partikelzählungen bin_0, \dots, bin_{14} . Die Werte bin_{18} und bin_{19} haben eine Korrelation von 0 und werden deshalb nicht mit aufgezählt. Mithilfe eines Korrelationsnetzwerkes lassen sich Zusammenhänge zwischen Werten auf zwei Arten herauslesen. Zum einen sind alle dargestellten Werte über farblich hervorgehobene Linien verbunden. In Abbildung 55 zeigen dicke, violette Linien eine stark positive Korrelation und dicke, rote Linien eine stark negative Korrelation. Dünne graue Linien stehen für niedrige Korrelationswerte. Zusammenhänge lassen sich ebenfalls über die Cluster-Bildungen erkennen. Bilden zwei oder mehrere Werte in einem Korrelationsnetzwerk eine Gruppe, weisen diese ähnlich hohe Korrelationen auf. Im Fall von Abbildung 55 lassen sich die höchsten Korrelationen zwischen $PM_{2.5}$ und PM_1 sowie zwischen bin_2 und bin_3 herauslesen. Des Weiteren existieren vier Cluster:

- $bin_0, bin_1, bin_2, bin_3, bin_4, PM_1, PM_{2.5}$ (Cluster 1)
- bin_5, PM_{10} (Cluster 2)
- $bin_6, bin_7, bin_8, bin_9$ (Cluster 3)
- $bin_{10}, bin_{11}, bin_{12}, bin_{13}, bin_{14}$ (Cluster 4)

Im nächsten Schritt wird zur Auswahl der im Modell verwendeten Prediktorvariablen eine *Feature Importance* Analyse durchgeführt. Die Ergebnisse sind in Abbildung 56 zu sehen.

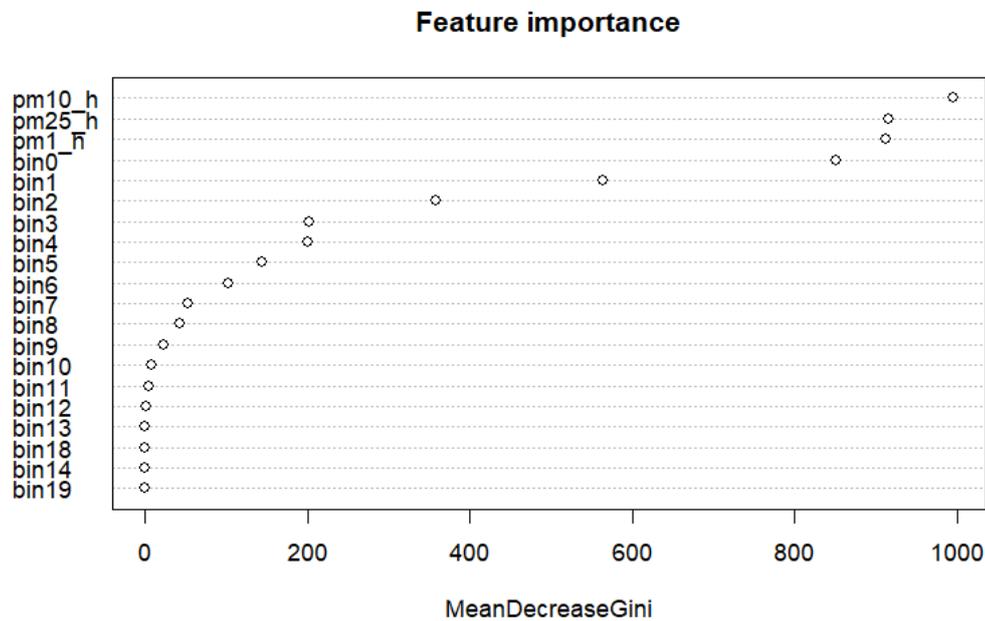


Abbildung 56: *Feature Importance* Analyse für die Prediktoren des Random Forest Modells

Abbildung 56 zeigt die *Variable Importance* Werte der Prediktorvariablen, gemessen am *Mean Decrease Gini* Index. Dieser ist ein Maß, das angibt, wie viel die jeweilige Variable zur Homogenität eines Modells beiträgt. Besonders niedrige *Mean Decrease Gini* Werte weisen darauf hin, dass die jeweiligen Prediktorvariablen nur einen geringen Beitrag zum Ergebnis des Modells erbringen. Gemessen daran ist der Parameter PM_{10} am wichtigsten für das RF Modell, gefolgt von den Werten $PM_{2.5}$, PM_1 und bin_0 . bin_{12} , bin_{13} , bin_{14} , bin_{18} und bin_{19} haben einen *Mean Decrease Gini* Index von 0. Relativ niedrige Werte sind darüber hinaus für bin_7 , bin_8 , bin_9 , bin_{10} und bin_{11} auszumachen. Für das Training und den Test des Modells werden somit folgende Prediktorvariablen ausgewählt:

- PM_1
- $PM_{2.5}$
- PM_{10}
- $bin_0, bin_1, bin_2, bin_3, bin_4, bin_5, bin_6$

Mithilfe der Prediktorvariablen wird ein Random Forest Modell erstellt, dass aus den Feinstaubwerten die Präsenz und Absenz von Umgebungsverkehr vorhersagt. Für die Splitooptionen 50/50, 60/40 und 70/30 werden die Ergebnisse in Tabelle 12 gezeigt.

Splitooption	Genauigkeit	Sensitivität	Spezifität
50/50	0,68	0,55	0,78
60/40	0,68	0,55	0,78
70/30	0,68	0,56	0,77

Tabelle 12: Genauigkeit, Sensitivität und Spezifität für die Splitooptionen 50/50, 60/40 und 70/30 eines RF Modells zur Vorhersage von Umgebungsverkehr mithilfe von Feinstaubwerten als Prediktoren

Tabelle 12 zeigt die Resultate des Random Forest Modells für die Splitooptionen 50/50, 60/40 und 70/30. Alle drei Splits produzieren ähnliche Ergebnisse. Die Genauigkeit (also der Wert, der angibt, wie viele Punkte des Testdatensatzes richtig erkannt werden) liegt bei 68 Prozent. Für die Sensitivität (Anteil der richtig klassifizierten Punkte mit vorhandenem Umgebungsverkehr) liegt der Wert nur bei 55 Prozent beziehungsweise 56 Prozent. Das bedeutet, dass nur knapp über die Hälfte der Punkte mit präsentem Umgebungsverkehr korrekt detektiert werden. Die Spezifität (Anteil der richtig klassifizierten Punkte ohne Umgebungsverkehr) zeigt bessere Ergebnisse mit 77 bis 78 Prozent. Die Ergebnisse für den ersten Durchlauf des Modells implizieren keinen Unterschied hinsichtlich der Splitooptionen. Um das Modell zu optimieren, wird ein Hyperparameter-Tuning für die Modellparameter m_{try} und n_{tree} durchgeführt.

Abbildung 57 zeigt ein Liniendiagramm mit der Entwicklung der Genauigkeitswerte für unterschiedliche m_{try} Konfigurationen des RF Modells. Dabei werden verschiedene Werte von 1 bis 15 getestet. Die Genauigkeit bei $m_{try} = 1$ ist am geringsten. Ab $m_{try} = 2$ befindet sich der Genauigkeitswert fast konstant auf dem gleichen Niveau von knapp unter 69 Prozent. Von allen überprüften Werten ergibt sich für $m_{try} = 4$ die höchste Genauigkeit. Die Abbildung gibt kein Indiz dafür, dass die Optimierung des m_{try} -Wertes für bessere Ergebnisse sorgt. Trotzdem wird der Wert von 4 für zukünftige Modellausführungen verwendet. Ein weiterer Hyperparameter neben m_{try} ist die Anzahl der Entscheidungsbäume n_{tree} .

Abbildung 58 zeigt ein Liniendiagramm für Genauigkeitswerte bei der Verwendung unterschiedlicher n_{tree} -Parameter. Überprüft werden zehn Werte in Hunderter-schritten zwischen 100 und 1000. Es ist zu sehen, dass die Genauigkeit bei den Werten von 100 bis 300 am abnehmen ist, und danach

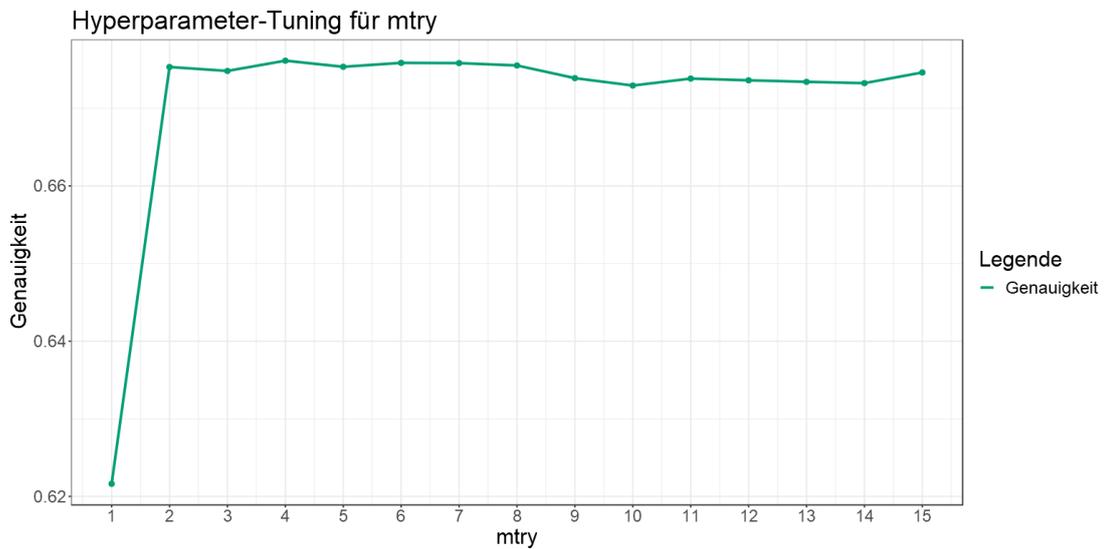


Abbildung 57: Hyperparameter-Tuning für den m_{try} -Wert: Genauigkeitswerte für $m_{try} = 1, \dots, m_{try} = 15$

wieder zunimmt bis zum Wert 1000. Die höchstmögliche Genauigkeit in der Abbildung ist für $n_{tree} = 100$ mit 70 Prozent auszumachen. Folglich wird in zukünftigen Modellausführungen ein n_{tree} -Wert von 100 gewählt.

Die zwei Schritte des Hyperparameter-Tunings zeigen, dass die Genauigkeitswerte bei $m_{try} = 4$ und $n_{tree} = 100$ am höchsten erscheinen. Deshalb wird das Random Forest Modell noch einmal mit diesen Einstellungen antrainiert. Die Ergebnisse des zweiten Modells für die Genauigkeit, die Sensitivität und die Spezifität für die drei Splitoptionen 50/50, 60/40 und 70/30 sind in Tabelle 13 dargestellt.

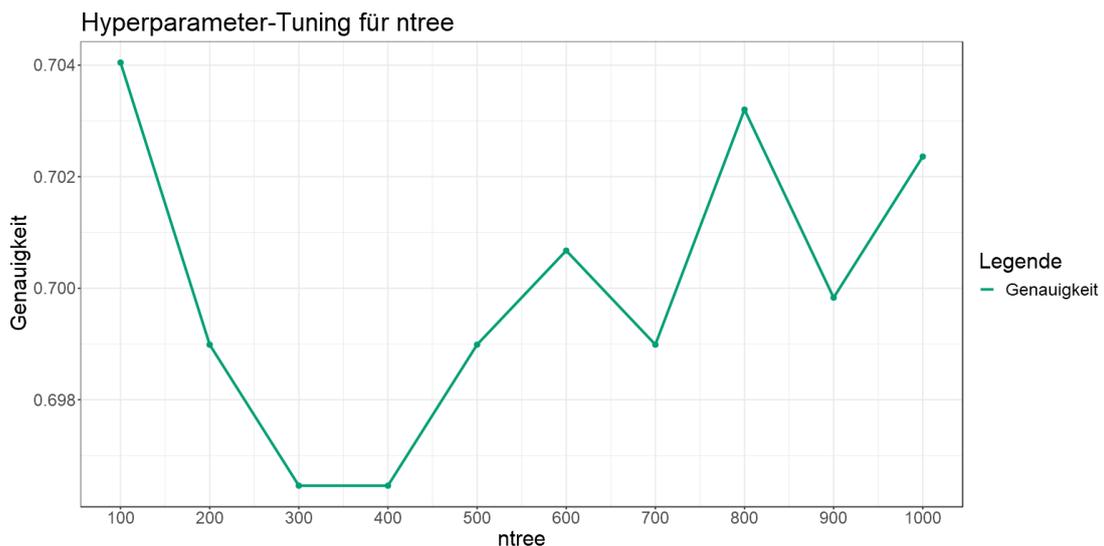


Abbildung 58: Hyperparameter-Tuning für den n_{tree} -Wert: Genauigkeitswerte für $n_{tree} = 100, \dots, n_{tree} = 1000$

Splitoption	Genauigkeit	Sensitivität	Spezifität
50/50	0,68	0,55	0,79
60/40	0,69	0,56	0,78
70/30	0,69	0,55	0,80

Tabelle 13: Genauigkeit, Sensitivität und Spezifität für die Splitoptionen 50/50, 60/40 und 70/30 eines RF Modells zur Vorhersage von Umgebungsverkehr mithilfe von Feinstaubwerten als Prediktoren mit $m_{try} = 4$ und $n_{tree} = 200$

Tabelle 13 zeigt die Genauigkeit, die Sensitivität und die Spezifität der Splitoptionen 50/50, 60/40 und 70/30 mit den optimierten Hyperparametern $m_{try} = 4$ und $n_{tree} = 200$. Es ist zu erkennen, dass sich durch die Optimierung die Werte nur geringfügig geändert haben. Für die Optionen 60/40 und 70/30 ist die Genauigkeit um ein Prozent auf 0,69 gestiegen. Die Sensitivität bleibt nach wie vor niedrig bei 55 bis 56 Prozent. Dafür steigt die Spezifität auf bis zu 80 an Prozent (70/30 Split). Die ROC-Kurve (Abbildung 59) und der AUC-Wert sind weitere Parameter zur Überprüfung der Genauigkeit eines RF Modells. Beide Maße werden im Folgenden diskutiert.

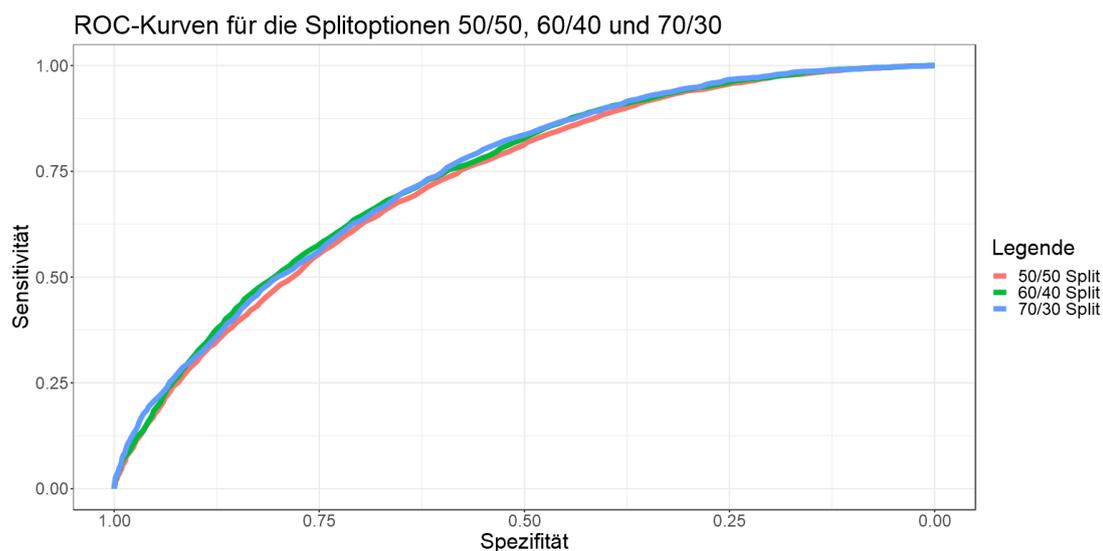


Abbildung 59: ROC-Kurven für die Splitoptionen 50/50, 60/40 und 70/30

Abbildung 59 zeigt die ROC-Kurven der Splitoptionen 50/50 (rote Linie), 60/40 (grüne Linie) und 70/30 (blaue Linie). Je näher die Kurven entlang der linken oberen Ecke des Diagramms verlaufen, desto genauer arbeitet das jeweilige Modell. Die drei dargestellten Kurven verlaufen alle ähnlich. Die blaue Kurve indiziert das beste Modell. Die dazugehörigen AUC-Werte, die den Anteil der Fläche rechts einer Kurve von der Gesamtfläche darstellen, lauten 0,73 (50/50), 0,73 (60/40) und 0,74 (70/30). Auch hier indiziert ein höherer Wert eine bessere Modellperformance. Dementsprechend sorgt der

70/30 Split für das bestmögliche Ergebnis des untersuchten Modells. Was bedeutet die Genauigkeit des Floating Car Observer Modells hinsichtlich dem Zusammenhang des Umgebungsverkehrs eines Fahrzeuges mit der messbaren Feinstaubkonzentration? Zunächst müssen die Ergebnisse des Modells eingeordnet werden. Die Genauigkeitswerte von 68 Prozent beziehungsweise 69 Prozent indizieren, dass das Modell für viele Observationen in der Lage ist, aus den gegebenen Prediktoren den korrekten Status des Umgebungsverkehrs zu bestimmen. Allerdings können nicht alle Verkehrsstatus mithilfe der Feinstaubmessungen erklärt werden. Die verbleibenden 31 beziehungsweise 32 Prozent sind Fehlerwerte. Bemerkenswert sind die Ergebnisse für die Sensitivität und Spezifität. Während bis zu 80 Prozent der Observationen mit nicht vorhandenem Umgebungsverkehr korrekt klassifiziert werden, ist das nur für knapp über 50 Prozent der Punkte mit Fahrzeugen in der direkten Umgebung der Fall. Das gibt Grund zur Annahme, dass für viele Punkte mit vergleichsweise niedrig gemessenen Feinstaubwerten ganz klar konstatiert werden kann, dass diese nicht im direkten Einfluss von Emissionen des Umgebungsverkehrs stehen. Ein Einfluss des Umgebungsverkehrs eines Fahrzeuges auf gemessene Feinstaubwerte kann mithilfe des durchgeführten Experiments nicht klar belegt werden. Allerdings bleiben Indizien, vor allem hinsichtlich der Genauigkeit des Modells im niedrigen Feinstaubwertebereich, dass es durchaus Verbindungen der zwei Phänomene gibt. Ob die Fehlerwerte primär durch eine schlechte Datenlage, ein fehlerhaftes Modell oder eine falsche Grundannahme zu erklären sind, bleibt diskutabel.

6.9 Vorhersage der Feinstaubkonzentration mithilfe von Einflussfaktoren der Datenqualität

Hinsichtlich der Datenqualität von mobil erfassten Sensordaten werden unterschiedliche potentielle Einflussfaktoren genannt. Um die Präsenz des Einflusses der unterschiedlichen Einflüsse zu untermauern, soll mithilfe eines Modells geprüft werden, ob sich Feinstaubkonzentrationen mithilfe von ausgewählten numerisch beschreibbaren Einflussfaktoren näherungsweise vorhersagen lassen. Es geht hierbei nicht um eine exakte Vorhersage der jeweiligen Feinstaubwerte, sondern um die Frage, ob An- und Abstiege der Feinstaubkonzentrationen anhand der Faktoren abgebildet werden können. Als Prediktorvariablen werden die Fahrtgeschwindigkeit, die Luftfeuchtigkeit, sowie die Anzahl an vorausfahrenden Bussen, Kraftfahrzeugen, Motorrädern und Lastkraftwagen genutzt. Als Methode fungieren ausgewählte Algorithmen.

men des maschinellen Lernens zur Regressionsanalyse (*Random Forest*, *Gradient Boost*, *Lineare Regression*). Die Modelle werden in *R* implementiert. Der Ablauf der Analyse wird in Abbildung 60 gezeigt.

Vorhersage der Feinstaubkonzentration mithilfe von Einflussfaktoren auf die Datenqualität

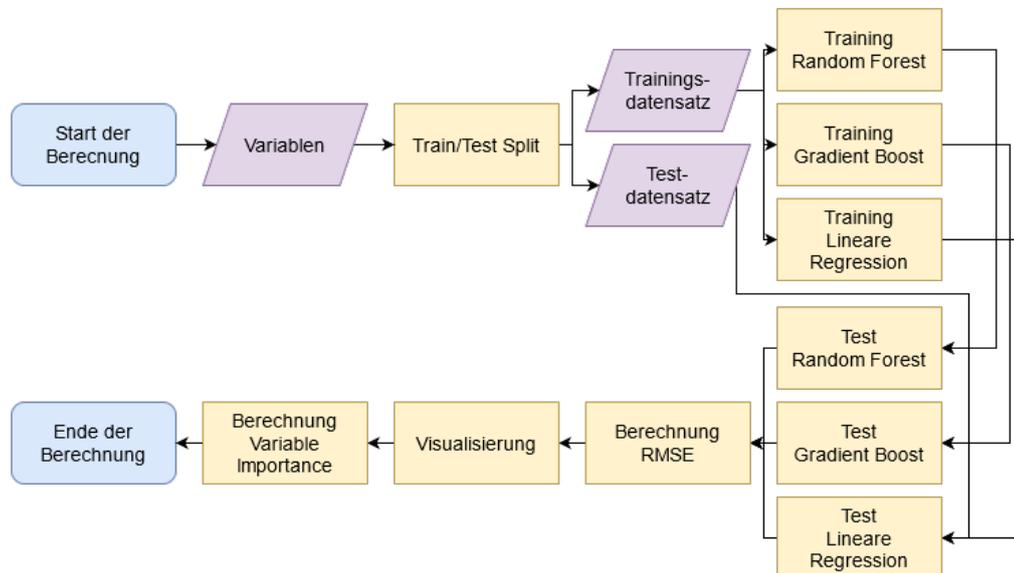


Abbildung 60: Ablaufdiagramm für die Vorhersage der Feinstaubkonzentration mithilfe von Einflussfaktoren der Datenqualität

Abbildung 60 zeigt ein Ablaufdiagramm für die näherungsweise Vorhersage der Feinstaubkonzentration mithilfe der Einflussfaktoren auf die Datenqualität als Prediktorvariablen. Dabei wird zunächst ein neuer Dataframe erstellt mit den Feinstaubwerten (PM_1 , $PM_{2.5}$, PM_{10}) und den vorgesehenen Prediktorvariablen *Fahrtgeschwindigkeit*, *Luftfeuchtigkeit*, *Busse*, *Kraftfahrzeuge*, *Motorräder* und *Lastkraftwagen*. Dieser wird im Anschluss in einen Trainings- und Testdatensatz aufgeteilt im 70/30 Split. Danach wird je ein *Random Forest*, *Gradient Boost* und *Lineare Regression* Modell antrainiert. Der Test der jeweiligen Modelle erzeugt die vorhergesagte Feinstaubkonzentration und *Root Mean Square Error* (RMSE)-Werte. Der RMSE gibt an, wie weit der vorhergesagte Wert im Schnitt vom originalen Wert abweicht. Die vorhergesagten Feinstaubwerte werden zusammen mit den originalen Werten der Testdatensätze zum Vergleich visualisiert.

Der RMSE stellt die Standardabweichung der Residuen dar. Das bedeutet, dass mithilfe dieses Wertes die mittlere Abweichung der vorhergesagten Werte zu den Originalwerten abgelesen werden kann. Die RMSE-Werte der verschiedenen Modelle sowie der Feinstaubkurven PM_1 , $PM_{2.5}$ und PM_{10} werden in Tabelle 14 gezeigt.

ML-Algorithmus	RMSE (PM_1)	RMSE ($PM_{2.5}$)	RMSE (PM_{10})
Random Forest	7,72	11,84	53,48
Gradient Boost	8,12	12,63	54,36
Lineare Regression	8,95	13,37	54,86

Tabelle 14: RMSE-Werte für *Random Forest*, *Gradient Boost* und *Lineare Regression* bei der Vorhersage von PM_1 , $PM_{2.5}$ und PM_{10}

Tabelle 14 zeigt die *Root Mean Square Error*-Werte für die Modelle *Random Forest*, *Gradient Boost* und *Lineare Regression* für die Feinstaubwerte PM_1 , $PM_{2.5}$ und PM_{10} . Es sind zwei Trends zu erkennen. Zum einen zeigt *Random Forest* die niedrigsten Abweichungen, gefolgt von *Gradient Boost* und *Lineare Regression*. Zum anderen nimmt der RMSE mit höherem Partikeldurchmesser zu. Während das *Random Forest*-Modell für PM_1 in einer Abweichung von $7,72 \mu\text{g}/\text{m}^3$ resultiert, zeigt das jeweilige Ergebnis für PM_{10} einen deutlich höheren Wert von $53,48 \mu\text{g}/\text{m}^3$. Das dürfte mit der steigenden Standardabweichung der Feinstaubwerte mit höherem Partikeldurchmesser zu erklären sein. Die Abweichungen stellen sich insgesamt als hoch dar. Da die Höhe der Feinstaubwerte nicht nur durch die Einflussfaktoren auf die Datenqualität, sondern auch mit der tatsächlichen Feinstaubkonzentration in der Umgebungsluft zu erklären ist, ist das Ergebnis nicht überraschend. Interessanter sind die Trends in der Entwicklung der vorhergesagten und originalen Feinstaubkurven. Ein Ausschnitt der Ergebnisse für PM_1 ist in Abbildung 61 zu sehen.

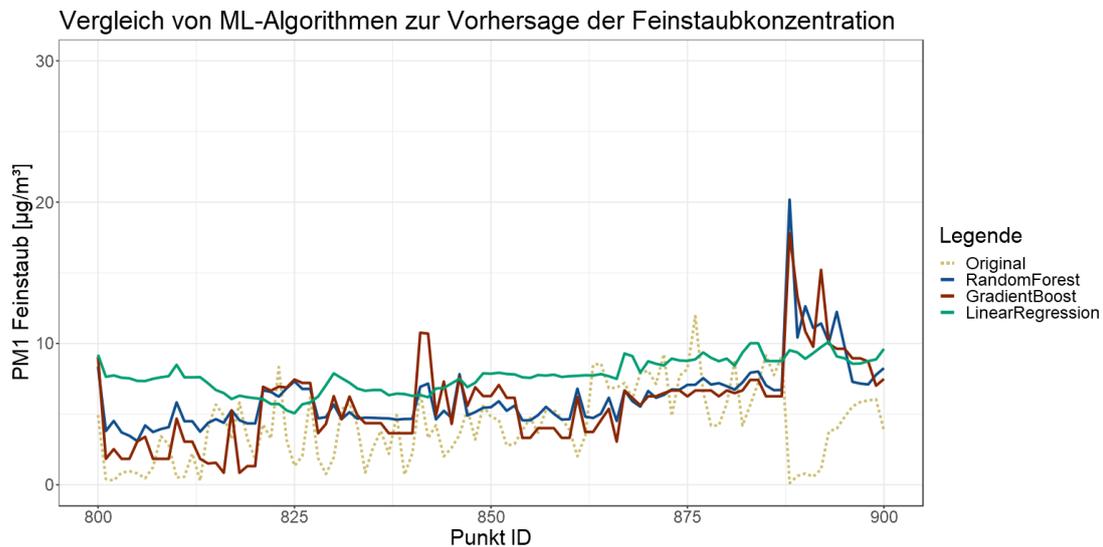


Abbildung 61: Vorhersage der PM_1 Feinstaubkonzentration mithilfe von Luftfeuchtigkeit, Bewegungsgeschwindigkeit und der Anzahl von Fahrzeugen im Umgebungsverkehr mit *Random Forest*, *Gradient Boost* und *Lineare Regression* zwischen Punkt-ID 800 und 900

Abbildung 61 zeigt ein Liniendiagramm mit den Kurven der originalen Feinstaubwerte (dunkelgelb) und der vorhergesagten Werte der Modelle *Random Forest* (blau), *Gradient Boost* (rot) und *Lineare Regression* (grün). Der Ausschnitt zeigt Werte zwischen den Punkt-IDs 800 und 900. Die originalen Werte steigen kontinuierlich bis Punkt-ID 887, wo sie deutlich abnehmen. Der lange Anstieg wird auch von den vorhergesagten Werten der Modelle *Random Forest* und *Gradient Boost* abgebildet. Die Werte der *linearen Regression* bleiben konstant auf einem ähnlichen Niveau und bilden zu keinem Zeitpunkt die Entwicklung der originalen Kurve ab. Der Abstieg ab Punkt-ID 887 wird von keinem der eingesetzten Modelle vorhergesehen. Der Ausschnitt in Abbildung 61 zeigt, dass die Feinstaubkonzentration in Teilen durch die ausgewählten Prediktorvariablen zu erklären ist. Ein Ausschnitt für $PM_{2.5}$ ist in Abbildung 62 zu sehen.

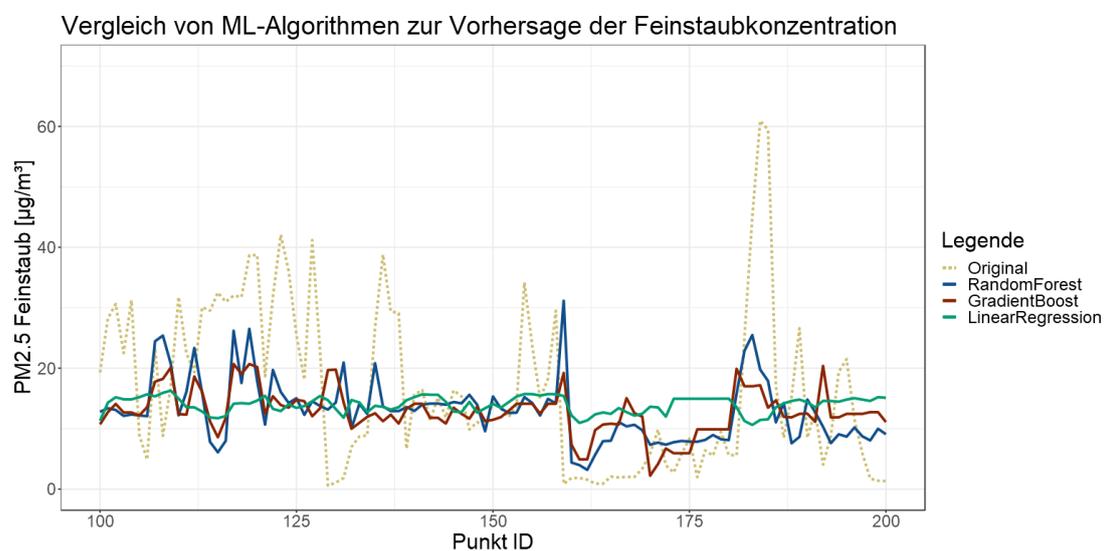


Abbildung 62: Vorhersage der $PM_{2.5}$ Feinstaubkonzentration mithilfe von Luftfeuchtigkeit, Bewegungsgeschwindigkeit und der Anzahl von Fahrzeugen im Umgebungsverkehr mit *Random Forest*, *Gradient Boost* und *Lineare Regression* zwischen Punkt-ID 100 und 200

Abbildung 62 zeigt ein Liniendiagramm mit den originalen $PM_{2.5}$ Feinstaubwerten (dunkelgelb) und den vorhergesagten Werten der Modelle *Random Forest* (blau), *Gradient Boost* (rot) und *Lineare Regression* (grün). Der Bereich reicht von Punkt-ID 100 bis Punkt-ID 200. Die originale Feinstaubkurve zeigt, anders als in Abbildung 61, relativ große Ausschläge in den höheren und niedrigeren Wertebereich. Besonders gut sichtbar ist das bei Punkt-ID 160 und Punkt-ID 185. Es fällt auf, dass die vorhergesagten Werte der *linearen Regression* keine Ähnlichkeit mit dem originalen Kurvenverlauf vorweisen. Anders sieht es aus bei den Modellen *Random Forest* und *Gradient Boost*, deren Vorhersagen in einigen Fällen den Trends der originalen Kurve folgen. Beispielhaft

dafür steht der Abstieg der Werte bei Punkt-ID 160, dem auch die rote und blaue Kurve folgen. Das Maximum der originalen Kurve bei Punkt-ID 185 wird zwar nicht mit ähnlich hohen Absolutwerten vorhergesehen, allerdings ist ein Anstieg der modellbasierten Kurven deutlich sichtbar. Einige lokale Maxima, wie bei Punkt-ID 120 oder 160, werden ebenfalls korrekt erkannt. Wie Abbildung 61 zeigt auch Abbildung 62 starke Indizien dafür, dass Feinstaubkonzentrationen näherungsweise anhand von Einflussfaktoren auf die Datenqualität vorhergesehen werden können. Ebenfalls deutlich wird, dass die *lineare Regression* im Gegensatz zu *Random Forest* und *Gradient Boost* dafür nicht geeignet zu sein scheint.

Im nächsten Schritt wird mithilfe einer *Variable Importance* Analyse überprüft, welche Prediktorvariablen für die Vorhersagen der genutzten Modelle wichtiger sind, und welche weniger wichtig. Die Ergebnisse der Analyse werden in Abbildung 63 gezeigt.

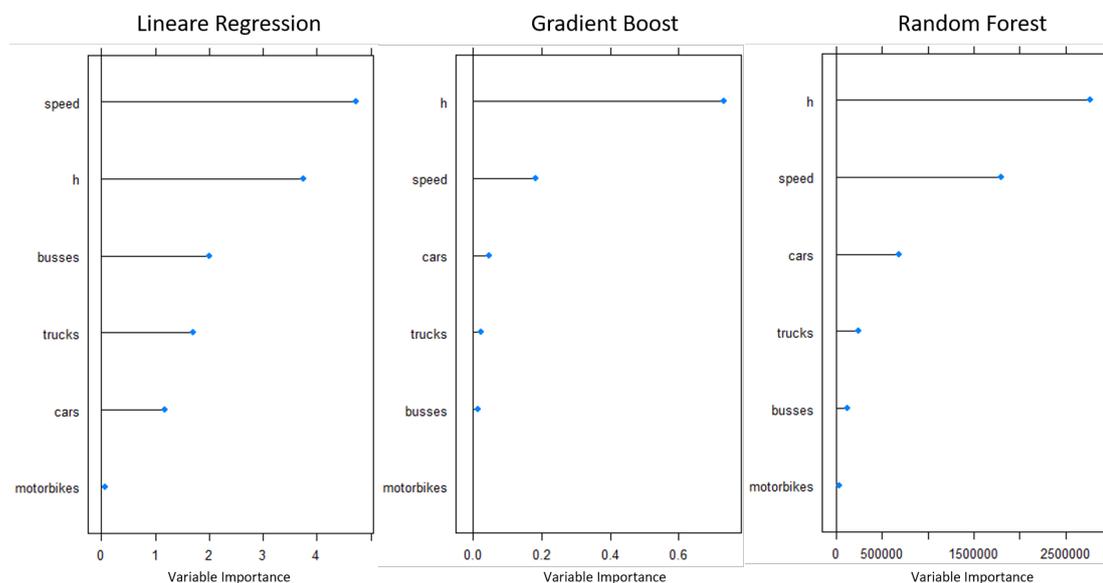


Abbildung 63: Variable Importance Analyse für die Prediktorvariablen der genutzten Modelle

Abbildung 63 zeigt drei Diagramme mit den Variable Importance Werten für die Prediktorvariablen des linearen Regressionsmodells, des Gradient Boost Modells und des Random Forest Modells. Die absoluten Werte der verschiedenen Diagramme sind nicht modellübergreifend miteinander vergleichbar, geben aber Informationen über die Wichtigkeit der Variablen innerhalb der Modelle. Für die lineare Regression ist die Geschwindigkeit die einflussreichste Variable. Die Luftfeuchtigkeit weist ebenfalls einen hohen Wert auf. Darauf folgen die Zählungen der Busse, LKW, PKW und Motorräder. Dabei beeinflusst die Anzahl der Motorräder die Ergebnisse des Modells überhaupt nicht, wäh-

rend sich die Anzahl der Busse, PKW und LKW als wichtig für den Verlauf der vorhergesagten Feinstaubkurven herausstellt. Für das *Gradient Boost* Modell ist die Luftfeuchtigkeit der mit Abstand wichtigste Prediktor. An zweiter Stelle folgt die Geschwindigkeit. Die Anzahl der Fahrzeuge im Umgebungsverkehr hat einen sehr geringen Einfluss auf die Ergebnisse des Modells. Das *Random Forest* Modell zeigt die gleiche Reihenfolge der Wichtigkeit der Variablen wie das *Gradient Boost* Modell. Hier haben jedoch die Geschwindigkeit und die Anzahl an PKW im Umgebungsverkehr einen deutlich höheren Einfluss. Die Dominanz der Geschwindigkeit und der Luftfeuchtigkeit zeigt sich in allen drei Modellen. Darüber hinaus stellt sich der Umgebungsverkehr als ein weniger wichtiger Faktor heraus für die Ergebnisse der Modelle. Die Ergebnisse der *Variable Importance* Analyse indizieren, dass die Geschwindigkeit und die Luftfeuchtigkeit einen höheren Einfluss auf die Datenqualität haben, als der Umgebungsverkehr. Allerdings setzt das voraus, dass die Ergebnisse der Modelle auch auf weitere Datensätze übertragbar sind.

6.10 Einflüsse auf die Qualität mobil erfasster Feinstaubdaten

Verschiedene Experimente überprüfen den Einfluss von ausgewählten Faktoren auf die Qualität von mobil erfassten Feinstaubdaten. Teilweise wird ein hoher, teilweise ein geringer Einfluss festgestellt. Darüber hinaus sind manche Faktoren gleichbleibend, während sich andere dynamisch während dem Messen ändern.

Abbildung 64 gibt einen Überblick über die untersuchten Einflussfaktoren mit einer Einordnung in statische und dynamische Teilaspekte. Im Folgenden sind die jeweiligen Faktoren aufgelistet. Es wird anhand von Erkenntnissen aus den Experimenten und der dazugehörigen Literatur argumentiert, welchen Einfluss der jeweilige Faktor auf die Datenqualität hat und inwiefern er sich einer statischen beziehungsweise dynamischen Gruppe zuordnen lässt.

Positionierungsgenauigkeit

Die Positionierungsgenauigkeit hängt von Faktoren wie den Sensorspezifikationen, der Art der Positionierung (dGPS, aGPS) und der Anzahl erreichbarer Satelliten ab. Darüber hinaus spielt das Sichtfeld auf und die relative Position des Empfängers zu den Satelliten eine Rolle. Im dazugehörigen Experiment wird die gemessene Abweichung der Positionierungsgenauigkeit (DOP) untersucht, die *GNSS-Empfänger* typischerweise beim Lokalisieren aus

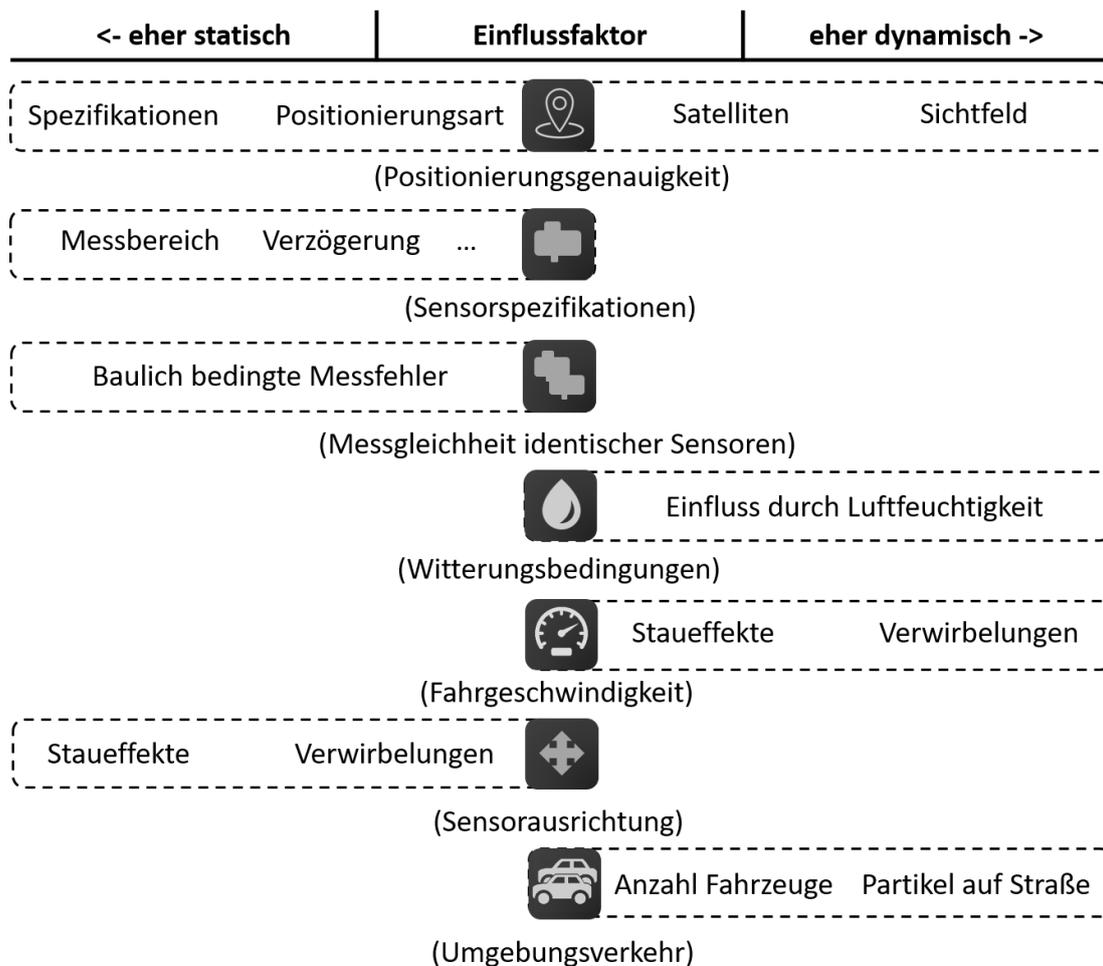


Abbildung 64: Zusammenfassende Unterteilung der Einflussfaktoren *Positionierungsgenauigkeit*, *Sensorspezifikationen*, *Messgleichheit identischer Sensoren*, *Witterung*, *Bewegungsgeschwindigkeit*, *Sensorausrichtung* und *Umgebungsverkehr* in statische und dynamische Faktoren

den Satellitenpositionen berechnen. Die resultierende Streubreite liegt für die meisten Messungen mit niedrigen Werten in einem akzeptablen Bereich. Allerdings gilt das nicht für alle untersuchten Punkte, denn bei einigen Observa-tionen liegen der *HDOP*- beziehungsweise *VDOP*-Wert über zehn. Neben der Analyse der Streubreiten werden auch unterschiedliche Mapmatching Ansätze untersucht. Dabei werden die Methoden *Snapping* mit unterschiedlichen Suchradien, die auf Routing basierende *nearest*-Methode sowie ein auf einem *Hidden Markov Modell* basierender Ansatz gewählt. Der zuletzt genannte Al-gorithmus ist dabei der Einzige, der die Punkte zufriedenstellend auf das zugrundeliegende Straßennetz verschiebt. Andere getestete Ansätze haben Schwächen in der realistischen Platzierung von Datenpunkten, denn verschobene Punkte springen in viele Fällen zwischen unterschiedlichen Straßen hin und her. Gematchte Trajektorien können nach Lyu et al. (2017) zur Berech-nung eines weiteren Parameters zur Überprüfung der Positionierungsgenau-

igkeit verwendet werden. Die Autoren nutzen Maße der Kurvengleichheit, um den Verlauf der unbearbeiteten Trajektorie mit dem Verlauf des zugeordneten Straßennetzes zu vergleichen. Für einen Versuch werden die *Discrete Fréchet* Distanz, das *Partial Curve Mapping*, die Fläche zwischen den Trajektorien und der Längenunterschied der Trajektorien verglichen. Die Ergebnisse zeigen teils deutliche Unterschiede zwischen den verglichenen Trajektorien. Aufgrund der Regel, dass weder die unbearbeiteten Trajektorien noch die auf ein Straßennetz projizierten Kurven jedoch den tatsächlichen Verlauf einer Testfahrt abbilden, sind diese Vergleiche kein aussagekräftiges Argument für den Einfluss von Positionierungsabweichungen auf die Qualität von mobilen Sensordatensätzen.

Für den Faktor der Positionierungsgenauigkeit lassen sich sowohl statische, als auch dynamische Gesichtspunkte beobachten. Die Genauigkeit des Sensors hängt maßgeblich von der Qualität des verwendeten Sensors und der dazugehörigen Antenne ab. Außerdem spielt es eine Rolle, wie der Sensor seine Position bestimmt - ob nur über die Satelliten, durch zusätzliche Referenzstationen oder mithilfe der Kombination verschiedener Lokalisierungsmethoden. Dynamische Einflüsse kommen durch die Bewegung der Satelliten und die Bewegung des Sensors hinzu. Die Genauigkeit hängt von der Anzahl an erreichbaren Satelliten und deren relative Position zueinander ab. Beide Faktoren können sich im Laufe einer Messung ändern, was eine dynamische Veränderung der Genauigkeitswerte zur Folge hat. Ebenfalls ein dynamischer Einfluss ist der Bewölkungsgrad am Himmel und der Bebauungsgrad am Straßenrand. Etwaige Elemente können die Sicht zu Satelliten verdecken oder das Signal abfälschen, was zu Fehllokalisierungen führen kann. Beide Phänomene ändern sich dynamisch während einer Messung und sind schwer vorherzusehen.

Sensorspezifikationen und Messunterschiede baugleicher Sensoren

Unterschiedliche Spezifikationen von Sensoren oder bauliche Unterschiede identischer Messgeräte können zu Abweichungen der Messgenauigkeit führen. Sensoren haben unterschiedliche Spezifikationen, die für die Messqualität mitverantwortlich sind. Dazu gehören unter anderem der jeweilige Messbereich, die Verzögerung einer Messung, der Temperatur- und Luftfeuchtigkeitsbereich, die Qualität des Lasers und die Konstanz des Luftflusses. Eine zeitgleiche Vergleichsmessung unter identischen Umweltbedingungen mit fünf *OPC-N3* Sensoren und einem *PMS3003*-Messgerät zeigt, dass die

Messwerte weder unter den gleichen Messgeräten, noch unter unterschiedlichen Sensoren identisch sind. Die Unterschiede zwischen den Messwerten der *OPC-N₃* Sensoren sind zwar minimal, aber dennoch sichtbar. Der hingegen kostengünstigere *PMS3003*-Sensor erkennt zwar dieselben lokalen Maxima und Minima, zeigt aber hinsichtlich der absoluten Werte deutliche Messunterschiede. Das Experiment gibt ein starkes Indiz für einen schwergewichtigen Einfluss der Sensorspezifikationen und einen verglichen leichten Einfluss baulich bedingter Messunterschiede auf die Datenqualität von Sensordatensätzen.

Die Sensorspezifikationen und die Messunterschiede baugleicher Sensoren üben einen statischen Einfluss auf die Datenqualität aus. Sind Messunterschiede zwischen den für eine Kampagne verwendeten Sensoren bekannt, können diese durch Kalibrierung beseitigt werden. Sensorspezifikationen bleiben während einer Messung konstant und sind den Nutzern in der Regel schon vor der Datenerfassung bewusst.

Witterungsverhältnisse

Witterungsverhältnisse beeinflussen unter anderem die Temperatur und die Luftfeuchtigkeit während einer Messung. Relevant für die Datenqualität ist jedoch primär die Luftfeuchtigkeit, da sich Wasserpartikel an Feinstaubpartikel anheften können und somit Messungen beeinflussen können. In einem ersten Versuch werden die Feinstaubkurven der Vergleichsmessungen von Datensatz 1 der jeweiligen Luftfeuchtigkeit gegenübergestellt. Ausschläge der Feinstaubkonzentration nach oben und unten kommen tendenziell mit dazu analogen Entwicklungen der Luftfeuchtigkeitskurve. Der Einfluss der Luftfeuchtigkeit ist bereits hinreichend erforscht und lässt sich mithilfe von Korrekturalgorithmen minimieren (Streibl 2017). In der Literatur werden vier verschiedene Wachstumsfunktionen zur Korrektur der Luftfeuchtigkeit genannt: gf_{Soneja} , $gf_{Hänel}$, gf_{Combo} und gf_{Skupin} . Ein Versuch mit den Feinstaubwerten aus Datensatz 3 zeigt, dass gf_{Soneja} den größten Einfluss auf die Wertentwicklungen nimmt. Die unterschiedlich hohen Auswirkungen der Wachstumsfunktionen auf die Wertentwicklung indiziert, dass die vorgestellten Korrekturfaktoren nur eine Annäherung an die tatsächlichen Feinstaubwerte bereitstellen, was für die Interpretation von Ergebnissen beachtet werden muss. Zuletzt zeigt das Modell zur Vorhersage der Feinstaubkonzentration mithilfe von Qualitätsfaktoren als Prediktoren, dass die Luftfeuchtigkeit einer von mehreren Einflüssen auf die Datenqualität ist.

Die Luftfeuchtigkeit ist ein dynamischer Faktor. Auch wenn dieser kleinräumig und in kurzen Zeiträumen teils nur geringen Schwankungen unterworfen ist, kann davon ausgegangen werden, dass plötzlich auftretender Regen einen deutlichen Einfluss auf die Datenqualität nimmt. Nicht immer sind witterungsbedingte Phänomene vorhersehbar. Deshalb macht es Sinn, bei Feinstaubmessungen auch die Luftfeuchtigkeit mit aufzuzeichnen.

Bewegungsgeschwindigkeit

Die Bewegung ist der Faktor, der statische Messungen maßgeblich von mobilen Messungen unterscheidet. Sie kann den Luftfluss und Verwirbelungen in der Umgebung eines Messgerätes beeinflussen, was Auswirkungen auf die Qualität von Sensordaten zur Folge haben kann. Mehrere Versuche zeigen Indizien für den Einfluss der Bewegungsgeschwindigkeit auf die Qualität von Messwerten. Der Vergleich der Werte beider unterschiedlich ausgerichteten Sensoren führt zum Schluss, dass die Korrelationswerte beider Messgeräte bei höherer Geschwindigkeit niedriger sind, als beim Betrachten des gesamten Datensatzes. Der Vergleich statischer und mobiler Messpunkte indiziert, dass Messunterschiede statischer und bewegter Sensoren an identisch lokalisierten Messstationen bestehen. Darüber hinaus ist auch die Fahrgeschwindigkeit Teil der Prediktorvariablen des Modells zur Vorhersage der Feinstaubkonzentration. Da mithilfe des Geschwindigkeitswertes (neben anderen Variablen) Ausschläge der Feinstaubkurven nach oben und unten korrekt abgebildet werden können, kann davon ausgegangen werden, dass die Fahrgeschwindigkeit Einfluss auf die gemessenen Feinstaubwerte nimmt. Die Geschwindigkeit nimmt einen dynamischen Einfluss auf die Genauigkeit von Sensormesswerten. In Datensatz 3 nimmt diese Werte von 0 km/h bis zu 106 Kilometer pro Stunde an. Beim Fahren im Straßenverkehr lässt sich die Geschwindigkeit nur bedingt selbst beeinflussen. Sie hängt von Faktoren wie dem Umgebungsverkehr, Verkehrsregeln und der Beschaffenheit der Straßen ab. Auch das verwendete Verkehrsmittel zum Messen nimmt darauf Einfluss. Kraftfahrzeuge wie Autos oder Busse haben in der Regel einen höheren Geschwindigkeitsbereich als Fahrräder.

Sensorposition und -ausrichtung

Die Ausrichtung eines Sensors spielt bei statischen Anwendungen nur eine untergeordnete Rolle. Mit der Bewegungsgeschwindigkeit kommt aller-

dings ein Faktor hinzu, der die Ausrichtung des Lufteinlasses eines Sensors für die Qualität der Messergebnisse relevant macht. Wie für die Bewegungsgeschwindigkeit gezeigt, verändern unterschiedliche Luftströme die Messgenauigkeit. Durch die Positionierung und Ausrichtung eines Sensors kann dieser mal mehr, und mal weniger von diesem Phänomen betroffen sein. Der Einfluss der Ausrichtung ist in der experimentellen Analyse klar zu erkennen. Zwei Sensoren, die zeitgleich auf dem selben Fahrzeug messen, jedoch unterschiedlich ausgerichtet sind, produzieren nur leicht bis moderat korrelierende Messwerte. Dahingegen resultiert eine statische Messung in sehr hohen Korrelationswerten beim Vergleich verschieden ausgerichteter Sensoren. Auch der Versuch in Kapitel 6.6 zeigt Indizien für den Einfluss der Sensorausrichtung auf die Qualität der Messwerte. Beim Vergleich der statisch gemessenen Feinstaubkonzentrationen mit den mobil bestimmten Messwerten verursachen unterschiedliche Ausrichtungen des Lufteinlasses (entgegen, mit und seitlich der Fahrrichtung) variierende Ergebnisse.

Die Sensorausrichtung und -position ist ein statischer Faktor. Das Messgerät wird vor einer Messung auf dem jeweiligen Fahrzeug befestigt und ändert seine Position und Ausrichtung während der Erfassung von Daten nicht.

Umgebungsverkehr

Der Umgebungsverkehr während einer Feinstaubsensormessung spielt eine Rolle, wenn Daten in Straßennähe beziehungsweise während dem Befahren einer Straße erfasst werden. Andere Fahrzeuge gelten dabei als Störquellen, die die Werte kurzzeitig stark ansteigen lassen können. Der Einfluss des Umgebungsverkehrs auf Messwerte wird in zwei Analysen von Datensatz 3 untersucht. In einem ersten Versuch wird ein Floating Car Observer implementiert, mithilfe dessen auf Basis der gemessenen Feinstaubwerte die Präsenz und Absenz des Umgebungsverkehrs vorhergesagt werden soll. Das Modell erzielt Vorhersagegenauigkeiten von bis zu 69 Prozent. Während einige Observationen falsch interpretiert werden, gibt der Floating Car Observer für die Mehrheit der Datenpunkte den korrekten Verkehrsstatus aus. In einem weiteren Experiment werden unter anderem mithilfe der absoluten Zahlen der Autos, Lastkraftwagen, Busse und Motorräder sowie der Geschwindigkeit und der Luftfeuchtigkeit die Feinstaubkonzentration vorhergesagt. Dafür kommen Algorithmen des maschinellen Lernens (*Random Forest*, *Gradient Boost* und *Lineare Regression*) zum Einsatz. Die Ergebnisse zeigen, dass die absoluten Feinstaubwerte nur selten korrekt vorhergesagt werden. Allerdings

werden kurzzeitige Sprünge der Werte nach oben und unten oft richtig erkannt, was ein Indiz für den Einfluss der getesteten Prediktorvariablen auf die Feinstaubwerte darstellt. Anhand der Ergebnisse der Versuche zeichnet sich ab, dass der Umgebungsverkehr Einfluss auf die Höhe der gemessenen Feinstaubkonzentrationen nimmt.

Der Umgebungsverkehr während einer Messung ist ein hochdynamischer Faktor. In Datensatz 3 schwankt die Zahl der Fahrzeuge zwischen 0 und 15. Auch die Zusammensetzung des Umgebungsverkehrs kann unterschiedlich ausfallen. Das ist relevant, da ein Motorrad in der Regel für deutlich weniger Feinstaubemissionen verantwortlich ist als ein Lastkraftwagen. Vor einer Messfahrt lässt sich das Volumen und die Zusammensetzung des Umgebungsverkehrs nur schwer vorhersagen.

7 Diskussion und Limitationen

Unterschiedliche Einflussfaktoren werden in den vorangegangenen Experimenten analysiert und auf ihre statischen und dynamischen Facetten hin untersucht. Die Ergebnisse der Analysen zeigen, dass es einige Einflüsse gibt, die bei einer mobilen Messkampagne beachtet werden müssen. Daraus ergeben sich diverse Limitationen, die im Folgenden erörtert werden. Darüber hinaus werden die Hypothese sowie die gestellten Forschungsfragen kritisch diskutiert und eingeordnet.

7.1 Auswahl der Einflussfaktoren auf die Qualität von mobil erfassten Feinstaubsensordaten

Röger (2021) nennt sieben Einflussfaktoren auf die Qualität mobil erfasster Feinstaubsensordaten. Dazu gehören die *Positionierungsgenauigkeit*, die *Sensorspezifikationen*, *baulich bedingte Unterschiede identischer Messgeräte*, die *Witterung*, die *Bewegungsgeschwindigkeit*, die *Sensorausrichtung* und der *Umgebungsverkehr* während der Messung. Mögliche Einflüsse auf die Datenqualität können mit den in der Fallstudie verwendeten Methoden zum Teil identifiziert werden. Wie einflussreich sind diese Faktoren tatsächlich, und existieren noch weitere Einflussfaktoren?

Die Positionierungsgenauigkeit ist ein viel diskutiertes Problem. Durch die Präsenz des Themas in zahlreichen Studien wie Dogan et al. (2014) und die Erkenntnisse aus den Analysen der Datensätze lässt sich zweifelsfrei feststellen, dass der Faktor tatsächlich Einfluss auf die Datenqualität nimmt. Wie groß der Einfluss tatsächlich ist, lässt sich nur näherungsweise bestimmen und variiert von Datensatz zu Datensatz. Einen guten Ansatz liefern hierbei die *DOP*-Werte, die die Streubreite der Messwerte beschreiben. Beachtet werden muss, dass diese Werte keine Einheit haben und nur eine ungefähre Aussage über die Positionierungsgenauigkeit treffen lassen. Wichtig ist außerdem, dass selbst bei hohen *DOP*-Werten nicht auszuschließen ist, dass sich ein individuell betrachteter Punkt nahe der realen Position befindet. Nachdem Positionsmessungen allgemein nur eine Annäherung an die Realität darstellen, hängt die Beurteilung der Datenqualität anhand der Positionierungsgenauigkeit stark davon ab, was mit der Analyse des Datensatzes ausgesagt werden soll. Wird die Feinstaubkonzentration nur kleinräumig auf wenigen Straßen analysiert, können leichte Ungenauigkeiten bezüglich der Lokalisierung der Messpunkte durchaus für eine Abfälschung der Ergebnisse sorgen. Dahinge-

gen fällt die Positionierungsgenauigkeit bei der Analyse von Feinstaubwerten ganzer Städte eher weniger ins Gewicht.

Ebenfalls diskutiert sind der Einfluss von Sensorspezifikationen und baulich bedingten Messunterschieden identischer Geräte. Parallel dazu zeigen auch die Erkenntnisse aus der Fallstudie klare Indizien für den Einfluss dieser Faktoren. Die Qualität und Genauigkeit des Messgerätes stellen sich hier als besonders ausschlaggebend bezüglich der Datenqualität heraus. Die Messunterschiede identischer Sensoren zeigen sich als minimal und je nach Anforderung an die Datenqualität als vernachlässigbar. Ein entscheidender Faktor bezüglich der Beurteilung der Datenqualität ist jedoch die allgemeine Genauigkeit der verwendeten Messgeräte. Die Wahl der Sensoren ist kritisch für das Ziel einer Studie. Aus der Literatur und den Erkenntnissen aus der Fallstudie geht hervor, dass ein kostengünstiger Sensor durchaus dazu in der Lage ist, relative Veränderungen der Feinstaubkonzentration abzubilden. Wird allerdings ein Augenmerk auf die Absolutwerte gelegt, gibt es keine Alternative zur Verwendung von hochpräzisen Sensoren zur Messung.

Witterungsbedingungen zeichnen sich unter anderem in der Temperatur und der Luftfeuchtigkeit ab. Die Temperatur nimmt dabei keinen wahrnehmbaren Einfluss auf die Qualität von Sensormesswerten. Einzig auf den erlaubten Temperaturbereich der verwendeten Feinstaubsensoren muss geachtet werden, um valide Daten zu erfassen. Die Luftfeuchtigkeit spielt hingegen eine bedeutende Rolle hinsichtlich der Datenqualität von Feinstaubsensordaten. Wie in Studien wie Di Antonio et al. (2018) ausführlich erörtert, verändern sich an Feinstaub anhängende Wasserpartikel die Reflektionseigenschaften von Feinstaubpartikeln und fälschen so Messungen ab. Bei der Verwendung von vergleichsweise kostengünstigen Feinstaubsensoren, die auf dem Lichtbrechungs-Prinzip basieren, muss also ein besonderes Augenmerk auf die Luftfeuchtigkeit während einer Messung gelegt werden. Bei niedrigen Luftfeuchtigkeitswerten ist der Faktor jedoch unbedeutender als bei hoher Luftfeuchtigkeit. Auch bei der Nutzung hochpräziser Sensoren spielt die Luftfeuchtigkeit teilweise keine Rolle, wenn die Feinstaubpartikel innerhalb des Sensors getrocknet werden.

Die Bewegungsgeschwindigkeit und die Sensorausrichtung sind weitestgehend unerforschte Faktoren. Bei der Analyse der Datensätze zeichnen sich jedoch starke Indizien ab, dass ein Einfluss vorliegt. Im Gegensatz zu den anderen untersuchten Faktoren lassen sich jedoch keine Trends erkennen. Weder eine zunehmende Geschwindigkeit, noch eine sich ändernde Sensorausrichtung nehmen einen deutlich sichtbaren Einfluss auf den Grad der Genauigkeit

der Sensordaten. An dieser Stelle lässt sich weiterhin eine Forschungslücke ausmachen. Um allgemeingültig beurteilen zu können, wie hoch der Einfluss der Bewegungsgeschwindigkeit und der Sensorausrichtung wirklich ist, sind weitere Studien notwendig.

Dass Störquellen in der Umgebung von Feinstaubmessungen für Ungenauigkeiten in den resultierenden Datensätzen sorgen, ist unbestritten. Daher scheint es allgemein kritisch, eine Messung auf einer Straße durchzuführen. Bei der mobilen Erfassung von Feinstaubsensordaten ist es jedoch schwer zu vermeiden, dass Sensoren in die direkte Umgebung von Emittenten von Feinstaub gelangen. Folglich zeigen sich auch im Datensatz der Fallstudie klare Indizien für einen Einfluss des Umgebungsverkehrs auf die Datenqualität. Wie schon bei der Bewegungsgeschwindigkeit und der Sensorausrichtung fällt es schwer, eine Einschätzung über die Höhe des Einflusses des Umgebungsverkehrs zu treffen. Vermutlich hängt der Einfluss des Faktors auch von der Zusammensetzung der Fahrzeuge in der Umgebung der Messung ab. Die Art des Motors (Benzin, Diesel, Elektro) spielt dabei eine Rolle. Zwar sorgen jegliche Antriebstypen für Feinstaubemissionen, jedoch kann bei einem Dieselfahrzeug von einer höheren Freisetzung von Partikeln ausgegangen werden als bei einem Elektrofahrzeug. Unbestritten ist jedoch, dass auch Fahrzeuge ohne Verbrennungsmotor für Feinstaubemissionen sorgen (Timmers et al. 2016). Hinsichtlich der Aufwirbelungseffekte von Partikeln ist das Fahrzeuggewicht mitentscheidend. Lastkraftwagen können hierbei für einen größeren Störfaktor sorgen als Kleinwagen (Sehmel 1973).

Die Analyse der potentiellen Einflussfaktoren zeigt klare Indizien für eine Beeinflussung der Datenqualität. Unklar bleibt jedoch weiterhin, ob es noch weitere Faktoren gibt, die einen Anteil an der Messgenauigkeit mobil erfasster Feinstaubdaten haben. Messungen im Feld stehen unter dem Einfluss komplexer Umgebungsbedingungen. Diese umschließen neben weiteren Wetterparametern (wie zum Beispiel der Strahlungsintensität der Sonne) auch die direkte Umgebung einer Messung. Bauliche Begebenheiten, wie das Durchfahren von Tunneln oder die Präsenz von hohen Gebäuden, werden bisher nicht berücksichtigt. Es ist daher nicht auszuschließen, dass mobile Feinstaubsensormessungen unter weiteren, bisher unbekanntem oder nicht erforschten, Einflüssen stehen.

7.2 Kritische Auseinandersetzung mit der Einordnung in statische und dynamische Faktoren

Die Analyse der Faktoren ergibt, dass nicht nur statische Einflüsse auf die Qualität mobil erfasster Feinstaubdaten wirken. Dynamische Faktoren wie der Umgebungsverkehr oder die Bewegungsgeschwindigkeit können sich während der Datenerfassung mit Sensoren ändern. Wie klar kann eine Einordnung in statisch und dynamisch erfolgen?

Positionierungsungenauigkeiten zeigen sich im Experiment sowohl mit statischen, als auch mit dynamischen Gesichtspunkten. Es ist fraglich, ob beide Dimensionen dabei im Gleichgewicht stehen, oder ob es eine Tendenz in Richtung statisch oder dynamisch gibt. Dogan et al. (2014) argumentieren, dass die Genauigkeit der Positionierung komplett als dynamisch zu kategorisieren ist. Mithilfe von Langzeitmessungen mit Sensoren finden sie Indizien für einen saisonalen Wandel der Positionierungsgenauigkeit. Die Autoren schließen, dass zu bestimmten Jahreszeiten, abhängig vom Messstandort, die Genauigkeit höher zu erwarten ist als in anderen Jahreszeiten. In einem vergleichsweise kurzem Messzeitraum wie in der vorliegenden Fallstudie kann diese Beobachtung nicht reproduziert werden. Es ist jedoch davon auszugehen, dass die Positionierungsgenauigkeit mehr in Richtung dynamischer Faktor tendiert.

Die Sensorspezifikationen sowie die Messungleichheiten identischer Sensoren zeigen sich im Experiment als statischer Faktor. In Langzeitstudien wird allerdings deutlich, dass auch diese Faktoren dynamischen Änderungen unterworfen sind. Zheng et al. (2019) observieren, dass sich Messgenauigkeiten individueller Sensoren über einen längeren Zeitraum ändern können. Außerdem merken die Autoren an, dass die Abnutzung der elektronischen Komponenten eines Messgerätes zu Defekten führen kann. Folglich können die Spezifikationen eines Sensors nur über einen bestimmten Zeitraum als statischer Faktor klassifiziert werden. Ab welchem Zeitpunkt Änderungen eintreten, ist nicht vorhersehbar. Unter anderem hängt dieser Faktor auch von dem Einfluss von Umgebungsbedingungen wie der Witterung auf einen Sensor ab (Zheng et al. 2019).

Witterungsbedingungen unterliegen hochdynamischen Änderungen. Cuadrat et al. (2015) untersuchen die Dynamik der Luftfeuchtigkeit in Saragossa (Spanien). Sie observieren Zusammenhänge der Luftfeuchtigkeit mit urbanen, geographischen und meteorologischen Faktoren. Die Autoren finden heraus, dass die Luftfeuchtigkeit in urbanen Räumen um bis zu 20 Prozent niedri-

ger ist als in ländlichen Umgebungen. Außerdem werden Korrelationen mit der Temperatur und diversen Windintensitäten und -richtungen erkannt. Die Observationen decken sich mit den Ergebnissen des Experiments in der Fallstudie. Kleinräumig und kurzzeitig fällt die Dynamik der Luftfeuchtigkeit allerdings geringer aus. Bei Messungen über einen längeren Zeitraum und in unterschiedlichen Umgebungen spielt die Dynamik der Witterungsbedingungen allerdings eine nicht zu vernachlässigende Rolle.

Die Bewegungsgeschwindigkeit ist dynamisch - unklar ist, inwiefern sich ändernde Fahrgeschwindigkeiten auf die Messgenauigkeit auswirken. Unterschiedliche Fahrzeugtypen bringen allerdings auch diverse Voraussetzungen für die Geschwindigkeit mit sich. Bei der Bewegung durch den fließenden Verkehr gibt es viele Parameter, die die Fahrgeschwindigkeit beeinflussen. Dazu gehören primär das Verkehrsvolumen und der Verkehrsfluss. Sind Fahrzeuge allerdings außerhalb des motorisierten Individualverkehrs unterwegs, ist die Dynamik teilweise besser vorhersehbar. Das ist zum Beispiel der Fall bei der Nutzung von Straßenbahnen als Messträger (Rinke et al. 2010). Das Experiment unterstreicht die Dynamik der Fahrgeschwindigkeit.

Der Umgebungsverkehr während einer Messung ist ständig im Wandel. Parameter, die diesen beschreiben, gelten als hoch-dynamisch. Dazu gehören der Verkehrsfluss, das Verkehrsvolumen und die Verkehrsdichte (Theja et al. 2010). Die Verkehrsdatenerfassung im Zuge der experimentellen Analyse zeigt die Dynamik des Umgebungsverkehrs mit 0 bis 15 detektierten Fahrzeugen pro Observation. Aufgrunddessen kann der Umgebungsverkehr zweifelsfrei als dynamisch kategorisiert werden.

Die Diskussion zeigt, dass Einflussfaktoren auf die Qualität mobil erfasster Feinstaubsensordaten dynamischer sein könnten als zunächst angenommen. Faktoren wie die Positionierungsgenauigkeit und die Sensorspezifikationen sollten daher bei längerfristigen Messungen ebenfalls auf dynamische Änderungen hin überprüft werden.

7.3 Kalibrierung und Korrektur von Feinstaubsensordaten auf Basis von Qualitätsfaktoren

Einige der detektierten Einflüsse lassen sich durch Funktionen und Modelle korrigieren. Dazu zählen primär die Luftfeuchtigkeit und die spezifikationsbedingte Genauigkeit von Sensoren (Di Antonio et al. 2018; Zheng et al. 2019). Dazu stellen verschiedene Mapmatching-Verfahren eine Lösung zur Korrektur von Messungenauigkeiten dar (Dogramadzi et al. 2021). Wann ist

eine Kalibrierung oder Korrektur sinnvoll, und wann werden Daten dadurch unnötig verfälscht?

Hagler et al. (2018) sehen die Anwendung von Korrekturverfahren auf Sensormesswerte kritisch. Sie begründen ihre Annahme damit, dass in anderen Studien angewandte Modelle oft darauf basieren, Messwerte mithilfe von Vergleichsmessungen zu kalibrieren. Es ist dabei nicht auszuschließen, dass während den Vergleichsmessungen andere Bedingungen vorherrschen, als zum Zeitpunkt der Messung, in der ein Korrekturmodell angewandt wird. Wenn die Vergleichsmessung darüber hinaus noch kurzzeitiger ausfällt als die darauf folgende Feldmessung, kann sich dieses Ungleichgewicht noch weiter verstärken. Weiter merken Maag et al. (2018) an, dass viele Sensoren unterschiedlich auf sich ändernde Umweltbedingungen reagieren. Wenn ein Messgerät weniger anfällig ist für Änderungen der Luftfeuchtigkeit als ein anderes, fällt es schwer, mit diesen Sensoren Kalibrierungen durchzuführen. Die Ergebnisse des Experiments decken sich mit diesen Ausführungen.

Ebenfalls kritisch zeigen sich die Korrekturfunktionen für die Verminderung des Einflusses der Luftfeuchtigkeit. Die in Studien wie Streibl (2017) diskutierten Wachstumsfaktoren resultieren in der Fallstudie bei hohen Feuchtigkeitswerten in völlig unterschiedlichen Werten. Während der Faktor gf_{Skupin} den Verlauf der Feinstaubkurven nur minimal anpasst, halbiert der Faktor gf_{Soneja} im Beispiel dieser Studie alle Werte. Mit steigender Luftfeuchtigkeit verschärft sich dieses Ungleichgewicht. Es ist während der Analyse nicht nachzuvollziehen, welche Korrekturfunktion die realistischsten Werte liefert.

Für die Korrektur von Positionierungsungenauigkeiten stehen eine Reihe von Mapmatching-Algorithmen zur Verfügung. Die experimentelle Studie in Augsburg hat gezeigt, dass zwischen lokalen, inkrementellen und globalen Ansätzen deutliche Unterschiede auftreten können. Lokale Mapmatching-Lösungen benötigen dabei zwar die wenigste Berechnungszeit, stellen sich aber als ungenau heraus. Wird die Korrektur nur auf Basis der Entfernung eines jeden Datenpunktes zur nächstgelegenen Netzwerkkante durchgeführt, kann von Fehlpositionierungen ausgegangen werden. Die Fallstudie resultiert dabei in unrealistischen Sprüngen einer Route von einem zum anderen Straßensegment. Bessere Ergebnisse versprechen inkrementelle und globale Ansätze, die die Verschiebung der Datenpunkte nicht nur auf Basis der Position des jeweiligen betrachteten Punktes bewerkstelligen. *Hidden Markov* basierende Algorithmen geben hierbei die besten Ergebnisse aus, haben allerdings auch eine nicht zu vernachlässigende Berechnungszeit. Kritisch zum Thema Mapmatching-Algorithmen anzumerken ist, dass auch die korrigierten Posi-

tionen der Datenpunkte nur eine Annäherung an die Realität darstellen. Soll jedoch die Feinstaubkonzentration auf Basis von Straßenabschnitten analysiert werden, können Mapmatching-Ansätze die Qualität der Analyse verbessern.

Weniger bis nicht präsent sind Korrekturalgorithmen für die Bewegungsgeschwindigkeit, die Sensorausrichtung und den Umgebungsverkehr. Folglich werden die Daten im Zuge der experimentellen Analyse nicht hinsichtlich dieser Faktoren korrigiert. Im Feld der Multicopter-basierenden Feinstaubmessungen existiert jedoch ein Ansatz zum Ausgleich von Ungenauigkeiten durch den auf den Sensor wirkenden Wind (Wang et al. 2020a). Da dieser allerdings primär von den Rotoren des Fluggerätes ausgeht und somit näherungsweise bestimmbar ist, lässt sich die Technik nicht auf Fahrzeuggetragene Feinstaubsensoren anwenden. Auftretender Fahrtwind beim Messen mit einem Kraftfahrzeug ist von vielen Faktoren, wie etwa den aerodynamischen Voraussetzungen der Karosserie, abhängig (Bella et al. 2011).

Von allen überprüften Einflussfaktoren lassen sich Positionierungsungenauigkeiten, sensorspezifikationsbedingte Ungenauigkeiten und witterungsbedingte Ungenauigkeiten durch Korrekturalgorithmen vermindern. Es sollte dabei aber beachtet werden, dass auch Sensorkalibrierungen nur eine Annäherung an die realen Werte darstellen. Außerdem ist diskutabel, ob Korrekturmodelle in allen Fällen für realistischere Werte sorgen.

7.4 Qualitätsfaktoren im Zusammenhang mit Standards und Richtlinien

Eine Richtlinie für die Beurteilung der Qualität von Geodaten wird in der *ISO 19157:2013* gegeben. Sensordaten bringen jedoch weitere Anforderungen bezüglich der Beurteilung ihrer Datenqualität mit sich. Unter anderem sollen dabei nach Klein et al. (2009) die Kategorie der Genauigkeit statische Einflussparameter und die Kategorie des Vertrauens dynamische Einflussparameter beschreiben. Inwiefern leisten die in der Fallstudie untersuchten Einflussparameter einen Beitrag hinsichtlich der bereits existierenden Standards und Richtlinien?

Die Genauigkeit und das Vertrauen summieren sich zum numerischen Gesamtfehler. Dabei sind jedoch nicht alle in der experimentellen Studie untersuchten Einflussfaktoren mit numerischen Werten beschreibbar. Eine Einordnung der Einflussfaktoren in die Kategorien der Genauigkeit sowie des Vertrauens wird in Tabelle 15 gegeben.

Einflussfaktor	Genauigkeit	Vertrauen	numerisch?
Positionierungsgenauigkeit	✓	✓	✓
Sensorspezifikationen	✓	(✓)	✓
Witterungsbedingungen	x	✓	✓
Bewegungsgeschwindigkeit	x	✓	x
Sensorausrichtung	✓	x	x
Umgebungsverkehr	x	✓	x

Tabelle 15: Einordnung der (nicht-)numerisch beschreibbaren Einflussfaktoren in die Kategorien *Genauigkeit* und *Vertrauen*

Tabelle 15 ordnet die untersuchten Einflussparameter in die Kategorien der *Genauigkeit* und des *Vertrauen* ein. Darüber hinaus werden die Faktoren nach ihrer numerischen Beschreibbarkeit eingeordnet. Das Symbol *x* steht dabei für Einflüsse, die nicht in die jeweilige Kategorie fallen oder nicht numerisch sind. Faktoren, die in eine Kategorie eingeordnet werden oder numerisch beschreibbar sind werden mit dem Symbol ✓ markiert. (✓) repräsentiert eine bedingte Einordnung in eine Kategorie. Die Einflussparameter werden in die Kategorien des *Vertrauens* und der *Genauigkeit* eingeordnet. Die Positionierungsgenauigkeit fällt dabei in beide untersuchten Kategorien aufgrund ihrer statischen und dynamischen Teilaspekte. Die Fallstudie zeigt, dass sie durch den Vergleich des gemessenen und originalen Kurvenverlaufes numerisch beschreibbar ist. Die Sensorspezifikationen fallen in die Kategorie der *Genauigkeit* und bedingt in die Kategorie des *Vertrauens*. Die Ausprägung von Messunterschieden lässt sich durch Modelle berechnen und gilt deshalb als numerisch beschreibbar. Witterungsbedingungen ändern sich dynamisch und lassen sich deshalb dem *Vertrauen* zuordnen. Es existieren Korrekturfunktionen für die Luftfeuchtigkeit, mit Hilfe derer man den Einfluss numerisch beschreiben kann. Die Einflüsse der Bewegungsgeschwindigkeit und der Sensorausrichtung sind dagegen nicht numerisch beschreibbar. Anhand der Ergebnisse der Fallstudie fällt die Bewegungsgeschwindigkeit in die Gruppe der *Genauigkeit* und die Sensorausrichtung in die Gruppe des *Vertrauens*. Der Umgebungsverkehr ändert sich dynamisch und gehört somit zur Gruppe des *Vertrauens*. Er ist zwar durch das Verkehrsvolumen beschreibbar, allerdings lässt sich die Größe des Einflusses des Faktors nicht in Zahlen ausdrücken. *ISO 19157:2013* beinhaltet weitere Kategorien zur Beschreibung der Qualität von Geodaten. In der Richtlinie werden die Kategorien *Vollständigkeit*, *logische Konsistenz*, *Positionsgenauigkeit*, *temporale Genauigkeit*, *thematische Genauigkeit* und *Benutzerfreundlichkeit* definiert. Eine Einordnung der in der Studie untersuchten Einflussfaktoren in die Kategorien der *ISO 19157:2013* wird in

Abbildung 65 gegeben (ISO Central Secretary 2013).

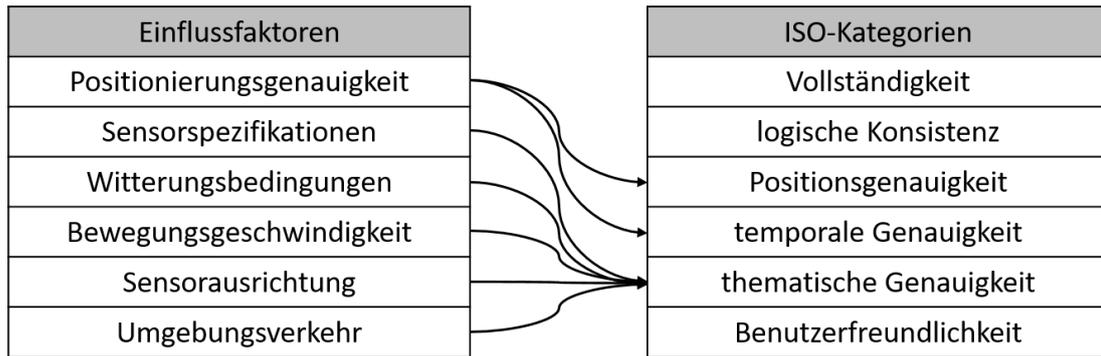


Abbildung 65: Verbindungen der in der Studie untersuchten Einflussfaktoren mit den Kategorien der Qualität von Geodaten nach *ISO 19157:2013*

Abbildung 65 ordnet die in der experimentellen Studie untersuchten Einflussparameter den jeweiligen Kategorien der Richtlinie *ISO 19157:2013* zu. Bei mobilen Messungen mit Sensoren wird die Position in der Regel mit GNSS-Empfängern bestimmt. Diese können sowohl räumliche, als auch temporale Daten aufzeichnen. Konsequenterweise wird der Einflussfaktor der Positionierungsgenauigkeit den ISO-Kategorien *temporale Genauigkeit* und *Positionsgenauigkeit* zugeordnet. Alle anderen untersuchten Faktoren beeinflussen die gemessene Feinstaubkonzentration. Diese zählt als quantitatives Attribut, weshalb die Einflussparameter der *thematischen Genauigkeit* zugeordnet werden.

Die untersuchten Einflussfaktoren finden ihren Platz in den gängigen Standards zur Beschreibung der Qualität von Geo- und Sensordaten. Dazu müssen jedoch keine neuen Hauptkategorien ergänzt werden. Vielmehr zeigt sich, dass die Faktoren bestehende Kategorien erweitern. Die *Vollständigkeit* der Daten, die sowohl von Klein et al. (2009), als auch in *ISO 19157:2013* genannt wird, wird von den untersuchten Einflussfaktoren nicht adressiert. Das gilt auch für die *logische Konsistenz*, die *Benutzerfreundlichkeit*, das *Volumen* und die *Aktualität* eines Sensordatensatzes.

7.5 Temporale und räumliche Übertragbarkeit der Fallstudie

Nicht nur einzelne Einflussfaktoren, sondern auch der komplette Bereich des Verkehrs befindet sich in einem stetigen, dynamischen Wandel. Außerdem sorgt der Klimawandel weltweit für sich verändernde Umgebungsbedingungen von Messungen. Abgesehen davon stellt sich die Frage, ob die durchgeführte experimentelle Analyse in einem anderen Testgebiet in identischen

Ergebnissen resultiert, wie in Augsburg. Sind die Ergebnisse der Fallstudie auch im Falle einer veränderten Umwelt- und Verkehrssituation noch valide? Die Stadt Augsburg befindet sich im Süden Deutschlands. Die Lage in Zentraleuropa sorgt für Messvoraussetzungen, die nicht für jeden beliebigen weiteren Ort auf der Welt gelten. Das betrifft unter anderem das Klima, die Verkehrszusammensetzung und die Art der Straßen. In einem alternativen Testgebiet kann es zu unterschiedlichen Ausprägungen der klimatologischen Faktoren wie der Luftfeuchtigkeit kommen. Folglich kann davon ausgegangen werden, dass sich der Einfluss der Witterungsbedingungen auf die Datenqualität in anderen Städten unterschiedlich darstellt. Darüber hinaus gibt es in unterschiedlichen Ländern auch unterschiedliche Zusammensetzungen der Verkehrsmittel. Effekte auf den Umgebungsverkehr können bei einem unterschiedlich ausgeprägten öffentlichen Nahverkehr oder der vornehmlichen Präsenz bestimmter Fahrzeugtypen nicht ausgeschlossen werden. Auch die Art der Straßen und das verwendete Material zum Straßenbau können die Übertragbarkeit der Fallstudie beeinträchtigen. In einigen Ländern sorgen ungepflasterte Verkehrswege für einen deutlich intensiveren Aufwirbelungseffekt von Partikeln auf dem Straßenbelag. Dieser spiegelt sich konsequenterweise in den erfassten Feinstaubwerten wieder.

Welchen Einfluss haben zukünftige Entwicklungen wie die Verkehrswende oder der Klimawandel auf die Übertragbarkeit der Fallstudie? Der modale Split beschreibt die Verteilung des Gesamtverkehrsaufkommens auf unterschiedliche Transportmittel. Im Jahr 2019 liegt dieser in Deutschland bei 79 Prozent für den motorisierten Individualverkehr, 8,4 Prozent für den Schienenpersonenverkehr, 6,8 Prozent für den öffentlichen Straßenpersonenverkehr und sechs Prozent für den Luftverkehr¹⁹. Es zeigt sich, dass es einen Trend in Richtung eines nachhaltigeren modalen Splits gibt. Nachdem der Verkehr für 18,5 Prozent der Feinstaubemission in Deutschland verantwortlich ist, kann davon ausgegangen werden, dass hier ein Effekt in zukünftigen Messkampagnen sichtbar sein könnte²⁰. Unter der Annahme, dass sich der modale Split in Zukunft weiter in Richtung öffentlicher Straßenpersonenverkehr und Schienenpersonenverkehr verschiebt, würde sich der Einfluss des Umgebungsverkehrs während einer Messung verändern. Dem ist jedoch entgegenzuhalten, dass es zwischen 1950 und 2020 keine deutlich sichtbare

¹⁹<https://de.statista.com/statistik/daten/studie/168397/umfrage/modal-split-im-personenverkehr-in-deutschland/>, 05/21

²⁰<https://de.statista.com/statistik/daten/studie/1090888/umfrage/verteilung-der-der-luftverschmutzung-nach-verursachergruppen-in-deutschland/>, 05/21

Abnahme der KFZ-Neuzulassungen in Deutschland zu verzeichnen gibt ²¹. Konsequenterweise kann festgehalten werden, dass Änderungen im modalen Split hin zu nachhaltigeren Transportmitteln die Ergebnisse der Fallstudie nicht beeinflussen (Jonkeren et al. 2011).

Ein weiterer Trend hinsichtlich des Verkehrs zeichnet sich in der Relevanz neuer Antriebstechniken ab. In Deutschland hat sich der Marktanteil von Elektrofahrzeugen von 2019 bis 2021 verdreifacht ²². Elektronisch betriebene Fahrzeuge emittieren beim Fahren deutlich weniger Schadstoffe, da Verbrennungsprozesse wegfallen. Aus mehreren Gründen limitiert diese Entwicklung jedoch nicht die Übertragbarkeit der Fallstudie. Zum einen betrifft die Reduktion der Schadstoffemission durch elektrisch betriebene Fahrzeuge primär die lokale Freisetzung von Kohlenstoffdioxid. CO₂-Messungen sind nicht Teil der Fallstudie. Zum anderen wird Feinstaub im Straßenverkehr nicht nur durch Verbrennung von fossilen Stoffen erzeugt. Durch die Rekuperationstechnik wird zwar beim Bremsen von Elektrofahrzeugen weniger Feinstaub freigesetzt als bei Fahrzeugen mit Verbrennungsmotor, allerdings sorgt auch der Reifenabrieb bei jeglichen Antriebsarten für Feinstaubemissionen. Darüber hinaus bestehen auch weiterhin Aufwirbelungseffekte. Feinstaub, der sich auf der Straße abgesetzt hat, kann durch das Überfahren mit Fahrzeugen aufgewirbelt und somit wieder in die Umwelt gebracht werden. Sowohl der Reifenabrieb, als auch Aufwirbelungseffekte sind von dem Faktor des Fahrzeuggewichtes abhängig. Viele Elektrofahrzeuge sind durch die Masse des Fahrakus schwerer als vergleichbare Verbrenner. Konsequenterweise ist auch die Zunahme des Marktanteiles von elektronisch betriebenen Fahrzeugen nur bedingt eine Limitation für die Übertragbarkeit der Studie. Es kann zwar zu einer Abnahme des Einflusses des Umgebungsverkehrs kommen, allerdings ist eine Aufhebung des Faktors nicht absehbar (Holland et al. 2019). Die Luftfeuchtigkeit gilt als ein wichtiger Einflussfaktor der Messgenauigkeit. Verschiedene Studien untersuchen den Einfluss des Klimawandels auf den Parameter der Luftfeuchtigkeit. Coffel et al. (2017) sehen Indizien für einen Anstieg dieser in den kommenden Jahrzehnten. Sie untersuchen den *Feuchtkegeleffekt*, der als eine Kombination aus Temperatur und Luftfeuchtigkeit zu verstehen ist. Konsequenterweise kann ein Effekt auf die Messgenauigkeit in zukünftigen Studien nicht ausgeschlossen werden. Das ist problematisch, denn die Ergebnisse der Fallstudie zeigen, dass sich Messfehler mit steigender

²¹<https://de.statista.com/statistik/daten/studie/74433/umfrage/neuzulassungen-von-pkw-in-deutschland/>, 05/21

²²<https://de.statista.com/statistik/daten/studie/784986/umfrage/marktanteil-von-elektrofahrzeugen-in-deutschland/>, 05/21

Luftfeuchtigkeit intensivieren. Limitiert werden dadurch die Erkenntnisse der Studie allerdings nicht. Es ist jedoch abzusehen, dass der Klimawandel eine Auseinandersetzung mit der Qualität von Feinstaubsensordatensätzen umso wichtiger macht (Coffel et al. 2017).

Ausgehend von einer Veränderung des modalen Splits, der Einführung neuer Antriebstechniken für Fahrzeuge und dem Klimawandel sind keine Limitationen der Fallstudie für die Zukunft auszumachen. Die potentiellen Auswirkungen der drei Ereignisse deuten allerdings darauf hin, dass die Relevanz des Themas der Studie weiter ansteigen wird. Dazu lässt sich beobachten, dass die Installation von Sensoren in Fahrzeugen an Popularität gewinnt. Die Datenerfassung mit mobilen Sensoren hilft bei der Bewältigung der Implementierung des autonomen Fahrens (Jo et al. 2013), bei der Einhaltung von Vorgaben der Luftschadstoffkonzentration (Re et al. 2014) und bei der Unterstützung von intelligenten Transportsystemen (Dimitrakopoulos et al. 2010). Technologien wie der 5G-Standard ermöglichen darüber hinaus die Kommunikation von mobilen Sensoren miteinander und mit zentralen Servern in Echtzeit (Sliwa et al. 2018). Konsequenterweise wird die Untersuchung von Einflussfaktoren der Datenqualität von Sensordaten in zukünftigen Arbeiten einen hohen Stellenwert haben.

7.6 Limitationen der in der Fallstudie durchgeführten Experimente

Die durchgeführte Fallstudie gibt Aufschluss über Einflussfaktoren auf die Datenqualität mobiler Feinstaubmessungen. Dabei kommen unterschiedliche Analysemethoden zum Einsatz. Jede eingesetzte Methode bringt auch potentielle Limitationen mit sich. Diese reichen von Ungenauigkeiten wegen Fehlern im Datensatz bis hin zu Fehlerwerten wegen einer nicht-optimalen Wahl von Modellparametern.

7.6.1 Limitationen der Genauigkeit bei der Fahrzeugdetektion aus Verkehrsvideos

Die Nutzung der Daten des Umgebungsverkehrs basiert auf der Annahme, dass alle Fahrzeuge vom Objektdetektionsalgorithmus korrekt erkannt werden. Störfaktoren wie eine schlechte Ausleuchtung des Videos, Objekte zwischen der Kamera und dem Fahrzeug sowie Reflexionen an der Windschutzscheibe vor der Kamera können jedoch für abgefälschte Ergebnisse sorgen. Darüber hinaus können auch Fahrzeuge, die am Straßenrand geparkt

sind und somit gar nicht am Verkehr teilnehmen als Umgebungsverkehr detektiert und klassifiziert werden. Abbildung 66 zeigt verschiedene Verkehrssituationen und deren dazugehörige Detektionen.

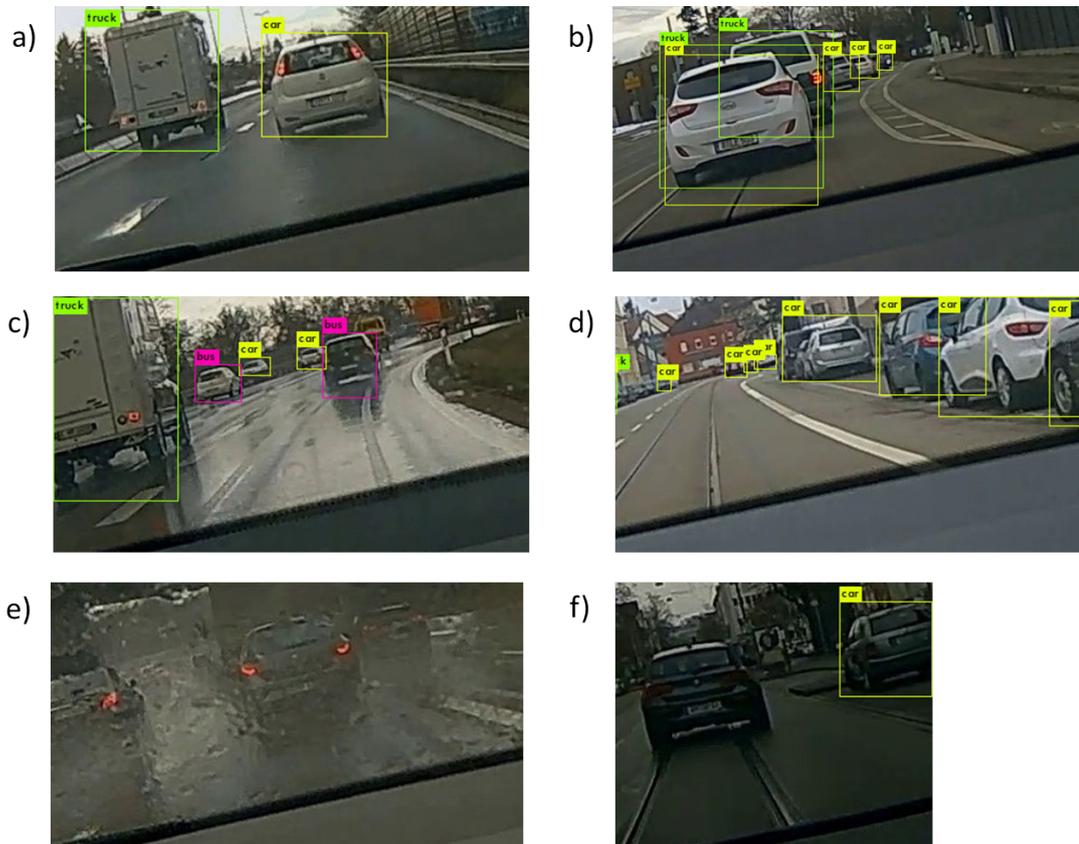


Abbildung 66: Detektion unterschiedlicher Verkehrssituationen: korrekte Objektdetektion (a), doppelte Detektion gleicher Fahrzeuge (b), korrekte Detektion mit falschen Klassifizierungen (c), Detektion von parkenden Fahrzeugen (d), nicht funktionierende Detektion bei starkem Regen (e) und nicht funktionierende Detektion bei schlechten Lichtverhältnissen (f)

Abbildung 66 zeigt Ausschnitte der Videokamera beim Aufnehmen von Verkehrsvideos. Detektierte Fahrzeuge haben einen gelben Rahmen, LKW grüne Rahmen und Busse rote Rahmen. Ausschnitt a) zeigt die korrekte Identifikation eines LKW und eines Fahrzeuges. In Ausschnitt b) ist zu sehen, dass in manchen Fällen doppelte Begrenzungsrahmen um individuelle Fahrzeuge gezogen werden. Die irrtümliche Detektion einiger Fahrzeuge als Busse ist in Ausschnitt c) zu sehen. Bild d) zeigt die Detektion parkender Fahrzeuge am Straßenrand. Der Einfluss von Witterungsbedingungen ist in Ausschnitt e) zu sehen. Hier behindert ein starker Regenfall die Identifikation eines Fahrzeuges. Auch schlechte Lichtverhältnisse können für Fehldetektionen sorgen, wie in Bild f) zu sehen ist.

Die Untersuchung der unterschiedlichen Verkehrssituationen zeigt, dass die Mehrfachdetektion gleicher Fahrzeuge, die falsche Klassifizierung von Objekten

ten (zum Beispiel Fahrzeuge als Busse) und die Detektion von am Straßenrand parkenden Fahrzeugen mögliche Fehlerquellen sind. Fehldetektionen können an der Qualität des Trainingsdatensatzes oder an nicht ausreichender Bildqualität liegen. Wurde ein bestimmter Fahrzeugtyp nicht zum Trainieren des Modells verwendet, kann es sein, dass es nicht korrekt detektiert wird. In der Fallstudie werden Frames aus den Verkehrsvideos extrahiert. Da die Frames nicht einzeln aufgenommen, sondern eigentlich Teil eines Videos sind, kann die Bildqualität beeinträchtigt sein. Die Detektion von parkenden Fahrzeugen am Straßenrand ist kritisch zu betrachten. Zwar kann durch Zuschneiden der Verkehrsvideos ein gewisser Teil des Straßenrandes entfernt werden, jedoch sind in einigen Kamerawinkeln und in der Distanz nicht-gewollte Detektionen nicht auszuschließen. Um parkende Fahrzeuge aus den Verkehrsvideos zu extrahieren, könnte man durch Identifikation der Fahrstreifen alles außerhalb der eigentlichen Fahrbahnen maskieren. Als weiteres Problem bei der Objektidentifikation aus Verkehrsvideos fungieren Witterungsverhältnisse. Es zeigt sich, dass vornehmlich starke Regenfälle für eine eingeschränkte Identifikation von Fahrzeugen sorgen können. Im Regelfall reicht es, wenn bei der Aufnahme der Videos der Scheibenwischer läuft. Ist der Regen allerdings so stark, dass im Zeitraum zwischen den Wischintervallen die Frontscheibe trotzdem nass bleibt, kann die Objektidentifizierung nur eingeschränkt funktionieren. Lichtverhältnisse sind eine weitere Limitation bei der Objektidentifikation aus Verkehrsvideos. Viele Dashcams haben eine Nachtsichtfunktion, bei der die Belichtung automatisch angepasst wird, wenn wenig natürliche Beleuchtung vorherrscht. Beim Betrachten der Verkehrsvideos zeigt sich, dass diese Belichtungskorrektur durchaus gut funktioniert. Allerdings kommt es in manchen Fällen vor, dass bei schwierigen Lichtverhältnissen unter Tags (wie in Abbildung 66, Ausschnitt f) die Objektdetektion nur eingeschränkt möglich ist. Das ist der Fall, wenn die natürliche Beleuchtung durch die Sonne niedrig ist, aber von der Kamera aus noch keine Korrektur der Belichtung stattfindet. Eine Lösung dazu ist die Nachbearbeitung der Verkehrsvideos, abhängig von den Belichtungswerten der jeweiligen Frames.

Die beschriebenen Limitationen können eine Rolle spielen hinsichtlich der Qualität der Detektionen. Eine Stichprobe soll Aufschluss über die Qualität der Fahrzeugdetektion aus den Verkehrsvideos bringen. Dafür werden aus dem Gesamtdatensatz zunächst zehn Trajektorien ausgewählt. Diese zehn Trajektorien sollen verschiedene Situationen wiedergeben: gute Bedingungen, viele Fahrzeuge, wenige Fahrzeuge, Regen, Abenddämmerung. Durch eine manuelle Zählung wird zu jeder vollen Sekunde der Zustand des Umge-

bungsverkehrs bestimmt. Die Ergebnisse der Zählung werden im Anschluss mit dem Ergebnis der automatischen Objektdetektion verglichen. Tabelle 16 zeigt die überprüften Trajektorien anhand ihrer IDs, ihrem automatisch generierten sowie ihrem händisch bestimmten Umgebungsverkehrszustand.

$ID_{\text{Trajektorie}}$	Anzahl Observationen	$[\text{ambient} == 1]$ (identifiziert)	$[\text{ambient} == 1]$ (händisch bestimmt)
1	755	425	399
5	335	15	110
10	419	298	322
15	251	86	169
20	335	156	151
25	839	77	90
30	671	485	495
35	335	297	197
40	335	305	304
45	167	117	120

Tabelle 16: Trajektorien mit ihrer Anzahl an Observationen sowie Datenpunkten mit Umgebungsverkehr (identifiziert und händisch bestimmt)

Tabelle 16 zeigt unterschiedliche Trajektorien-IDs, ihre Anzahl an Einzelobservationen, ihr automatisch bestimmter Verkehrszustand sowie ihr händisch bestimmter Verkehrszustand. Überprüft wird eine Stichprobe des Gesamtdatensatzes mit den Trajektorien-IDs 1, 5, 10, 15, 20, 25, 30, 35, 40 und 45. Nicht von Regen, Dunkelheit oder viele parkenden Fahrzeugen beeinflusst sind die Trajektorien 1, 10, 20, 25, 30, 40 und 45. Das spiegelt sich auch in den Ergebnissen des Stichprobentests wieder. Die untersuchten Trajektorien zeigen Abweichungen bis zu einem Wert von 26 Observationen. Ein stark abweichendes Ergebnis zeigt Trajektorie 5. Vom Objektdetektionsalgorithmus werden lediglich 15 Punkte mit Umgebungsverkehr erkannt, während die händische Auszählung ein Ergebnis von 110 zeigt. Dunkelheit und Regen sorgen für erschwerte Detektionsbedingungen. Dasselbe gilt auch für die Trajektorie mit der ID 15. Bei beiden Teildatensätzen werden vom Algorithmus deutlich weniger Punkte erkannt als durch eine händische Zählung. Trajektorie 35 zeigt auch eine hohe Differenz, allerdings ist dabei der händisch bestimmte Wert niedriger als der automatisch generierte Wert. Die Betrachtung des dazugehörigen Videos zeigt wenig Umgebungsverkehr bei gleichzeitig vielen parkenden Fahrzeugen am Straßenrand.

Die Analyse von einzelnen Frames und der Stichprobentests der Trajektorien-Videos zeigt, dass die automatische Detektion von Fahrzeugen aus Verkehrsvideos mithilfe des *YOLOv3*-Algorithmus nicht immer problemfrei funktio-

niert. Gerade bei schlechter Ausleuchtung der Straße, Regen oder vielen parkenden Fahrzeugen am Straßenrand schafft es der Ansatz nicht immer, genaue Ergebnisse zu liefern. Die Limitation betrifft sowohl den Objektdetektionsalgorithmus, als auch die verwendete Kamera. Bei der Nutzung des Ansatzes bei Nacht oder schwierigen Beleuchtungsverhältnissen sollte ein passendes Aufnahmegerät verwendet werden. Auch die Nachbelichtung der Verkehrsvideos kann dabei die Ergebnisse verbessern. Darüber hinaus zeigt sich die Wahl eines passenden Begrenzungsrahmens kritisch, um parkende Fahrzeuge aus dem Verkehrsvideo auszuschließen.

7.6.2 Bestimmung des Einflusses von Bewegungsgeschwindigkeit und Sensorausrichtung

In der Fallstudie kommen Vergleichsmessungen von statischen und mobilen Sensoren zum Einsatz, um den Einfluss von Bewegungsgeschwindigkeit und Sensorausrichtung zu analysieren. Dafür werden zwei fixe Messpunkte definiert, die durchgehend die Feinstaubkonzentration aufzeichnen. Ein mit verschiedenen ausgerichteten Sensoren ausgestattetes Messfahrzeug passiert diese Punkte mit unterschiedlichen Geschwindigkeiten. Die Datenanalyse resultiert in der Erkenntnis, dass die bewegten Sensoren teils deutlich abweichende Werte erfassen, verglichen mit den statischen Messpunkten. Darüber hinaus lässt sich allerdings kein Trend feststellen. Höhere Fahrgeschwindigkeiten führen nicht automatisch zu ungenaueren Werten.

Diskutable Punkte finden sich zunächst im Aufbau und in der Durchführung des Experiments. Das Testfahrzeug passiert die Messpunkte dabei nicht an exakt gleicher Stelle, sondern in einem Abstand von etwa zwei Metern. Es muss davon ausgegangen werden, dass kurzfristige Ereignisse wie beispielsweise aufkommender Wind kleinräumig für unterschiedliche Feinstaubkonzentrationen sorgen können. Darüber hinaus spiegeln die getesteten Fahrgeschwindigkeiten von fünf bis 50 Kilometer pro Stunde nicht zwangsweise real vorkommende Geschwindigkeiten im Straßenverkehr wieder. Sowohl inner- als auch außerstädtisch existieren zum Teil höhere Maximalgeschwindigkeiten. Diese werden mit dem verwendeten Versuchsaufbau nicht abgedeckt. Ebenfalls kritisch wird die Länge der definierten Teststrecke gesehen. Vor dem ersten und nach dem zweiten statischen Messpunkt liegen 20 Meter, die das Testfahrzeug mit der überprüften Geschwindigkeit befährt. Das kann dazu führen, dass das Fahrzeug die vorgesehene Geschwindigkeit erst kurz vor der Messstation erreicht und somit noch im Einfluss der Beschleunigung steht. Nachdem der Einfluss zunehmender und abnehmender Geschwindig-

keiten nicht überprüft wird, kann nicht davon ausgegangen werden, dass sich ändernde Feinstaubwerte nur mit der Geschwindigkeit im Zusammenhang stehen.

Datensatz 2 ist deutlich kleiner als die Datensätze 1 und 3. Lediglich 382 Datenpunkte werden für die Fallstudie erfasst. Das ist zwar genug zur Observierung von Indizien, dass ein Einfluss der Bewegungsgeschwindigkeit existiert. Allerdings ist das nicht genug, um eine allgemein gültige Aussage über die Wichtigkeit des Einflusses zu treffen.

7.6.3 Limitationen der Modell- und Variablenauswahl für die Vorhersage der Feinstaubkonzentration

Für die Vorhersage der Feinstaubkonzentration mithilfe von Parametern der Datenqualität werden die ML-Algorithmen *Lineare Regression*, *Gradient Boost* und *Random Forest* ausgewählt. Als Prediktorvariablen fungieren die Parameter *Fahrtgeschwindigkeit*, *Luftfeuchtigkeit*, *Busse*, *Kraftfahrzeuge*, *Motorräder* und *Lastkraftwagen*. Die Wahl der Modellierungsalgorithmen erklärt sich primär durch die Popularität der Ansätze für die Sensordatenverarbeitung. Dazu gilt die *Lineare Regression* als robust und nicht fehleranfällig für die Lösung von Regressionsproblemen. Der Vorteil in der Nutzung des *Gradient Boost* Algorithmus ist, dass Datensätze nicht vorprozessiert werden müssen. Die gewählten Prediktorvariablen stammen größtenteils aus den Rohdaten der Messungen, weshalb sich *Gradient Boost* als Modellierungsalgorithmus anbietet. *Random Forest* ist ein häufig verwendeter ML-Algorithmus für die Analyse großer Datensätze, weshalb er sich für die Vorhersage der Feinstaubkonzentration eignet (23.725 Observationen). Trotzdem existieren neben den drei verwendeten Algorithmen noch eine Reihe weiterer Modelle für Regressionsprobleme. Anstelle der *linearen Regression*, die für den verwendeten Datensatz keine guten Ergebnisse zeigt, wäre der Test weiterer Algorithmen sinnvoll gewesen.

Die zur Regression verwendeten Prediktorvariablen stammen von denjenigen Qualitätsparametern, die (1) sich dynamisch ändern und (2) numerisch beschreibbar sind. Die Ergebnisse der Analyse zeigen, dass mithilfe dieser Variablen bereits relative Veränderungen der Feinstaubkonzentration vorhersehbar sind. Es bleibt eine offene Fragestellung, ob mit einer weiteren Reduzierung oder Hinzunahme von Prediktorvariablen eine höhere Genauigkeit hätte erzielt werden können. Die *Feature Importance* Analyse zeigt unterschiedliche Wichtigkeitswerte für die verschiedenen Parameter. Es gibt Indizien dafür, dass die *Fahrtgeschwindigkeit* und die *Luftfeuchtigkeit* einen größeren Einfluss

auf die Höhe der Feinstaubwerte nehmen, als die Zahlen des Umgebungsverkehrs. Fraglich ist, ob eine Zusammenfassung der einzelnen Variablen *Busse, Kraftfahrzeuge, Motorräder* und *Lastkraftwagen* zur Variable *Umgebungsverkehr* für eine höhere Wichtigkeit des Faktors gesorgt hätten. Das Ziel des Experimentes war es jedoch lediglich, herauszufinden, ob relative Veränderungen der Feinstaubkonzentration vorhersehbar sind. Da dieses Ziel mit den implementierten Modellen erreicht wird, kann die Optimierung der Modelle als zukünftige Arbeit deklariert werden.

7.7 Umgebungsverkehr - Störfaktor oder wichtiger Bestandteil von Feinstaubmessungen?

Dass der Straßenverkehr für erhöhte Feinstaubwerte sorgt, ist unbestritten und wird von einigen Studien bestätigt. Tatsächlich sorgen die Verbrennung von fossilen Stoffen, der Abrieb der Bremsen und Reifen sowie Aufwirbelungseffekte auf dem Straßenbelag für eine messbare Mehrbelastung mit Feinstaub. Diskutabel ist, inwiefern dieser Vorgang mobile Feinstaubmessungen beeinträchtigt. Messende Fahrzeuge befinden sich im Gegensatz zu stationären Messpunkten inmitten der Emissionsquelle des Feinstaubes und stellen darüber hinaus auch selber einen Emittenten von Feinstaub dar. Findet nun eine Messung direkt neben einem Auspuff oder inmitten einer Wolke aus aufgewirbeltem Feinstaub statt, können lokal wesentlich höhere Werte festgestellt werden als am Straßenrand. Grundsätzlich besitzen sowohl die auf der Straße, als auch die neben der Straße gemessenen Werte ihre Richtigkeit. Entscheidend ist die Interpretation der Daten. Findet die Datenerhebung zur Einschätzung des Risikos der aktuellen Feinstaubkonzentration für Fußgänger statt, sind direkt an der Quelle gemessene Daten sicherlich nicht zu 100 Prozent aussagekräftig. Auch deshalb gelten für die Errichtung offizieller Messstationen Richtwerte. Folgende Regeln gelten unter anderem laut dem 39. *Bundes-Immissionsschutzgesetz*²³ für die Positionierung von Messpunkten:

- Der Lufteinlass einer Station darf nicht von potentiellen Ströquellen beeinträchtigt sein.
- In einem Umkreis von einigen Metern um den Messeinlass dürfen sich keine Objekte befinden.
- Der Messeinlass muss sich in eine Höhe von 1,5 bis 4 Metern befinden.

²³http://www.gesetze-im-internet.de/bimschv_39/index.html, 03/21

- Der Wiedereintritt von Abluft soll vermieden werden.
- Eine Station darf nicht weiter als zehn Meter von der Fahrbahn entfernt sein. Für verkehrsreiche Kreuzungen gilt jedoch ein Mindestabstand von 25 Metern.

Derartige Regelungen sind schwer, bei der mobilen Datenerfassung zu berücksichtigen. Ein Fahrzeug ist im Straßenverkehr unterschiedlichen Störquellen ausgesetzt. Fahrzeuge in der Umgebung beeinflussen den Lufteinlass durch die Emission von Feinstaub. Die eigene Geschwindigkeit eines Messfahrzeuges sorgt für Verwirbelungen, wobei nicht ausgeschlossen werden kann, dass es zum Wiedereintritt von bereits analysierter Luft kommt. Des weiteren können sich messende Fahrzeuge direkt auf hochfrequentierten Kreuzungen befinden.

Zusammenfassend muss der Umgebungsverkehr als potentieller Störfaktor in Erwägung gezogen werden. Die Studie dieser Arbeit zeigt, dass Feinstaubemissionen des Umgebungsverkehrs durchaus messbar sind. Bei mobilen Feinstaubmessungen liegt die Herausforderung daher vornehmlich in der Dateninterpretation.

8 Schlussfolgerungen und Ausblick

In diesem Kapitel werden Schlussfolgerungen gezogen, die Hypothese verifiziert und die Forschungsfragen beantwortet. Zum Abschluss wird ein Überblick über zukünftige Arbeiten gegeben.

8.1 Schlussfolgerungen

Die Analyse der untersuchten Qualitätsparameter mobil erfasster Feinstaubdaten zeigt, dass unterschiedliche Faktoren Einfluss auf die Datenqualität nehmen. Primär indiziert das Modell zur Vorhersage von Feinstaubkonzentrationen mithilfe von Qualitätsfaktoren als Prediktorvariablen die Relevanz der untersuchten Einflüsse. Die Wirkung der Einflüsse variiert je nach Umgebung einer Messung und den verwendeten Sensoren. Auch die Dynamik der Faktoren unterscheidet sich. Ein Teil der untersuchten Einflüsse zeigt sich statisch, während sich andere Faktoren dynamisch während der Datenerfassung verändern. Trotz des bestätigten Einflusses der untersuchten Faktoren und deren Dynamik zeigt der Vergleich mit den gängigen Standards und Richtlinien zur Beschreibung der Qualität von Geo- und Sensordaten nicht, dass diese um weitere Kategorien erweitert werden müssen. Die Auswirkung der untersuchten Faktoren bilden sich ab in der *Positionsgenauigkeit*, der *temporalen Genauigkeit* und der *thematischen Genauigkeit* nach ISO 19157:2013. In den Richtlinien zur Beschreibung der Qualität von Sensordaten nach Klein et al. (2009) decken die Kategorien der *Genauigkeit* (statische Einflussfaktoren) und des *Vertrauens* (dynamische Einflussfaktoren) die Erkenntnisse der Studie ab. Konsequenterweise wird die Hypothese der Arbeit verworfen. Die untersuchten Qualitätsparameter stellen vielmehr eine Ergänzung zu den bereits existierenden Kategorien zur Beschreibung der Qualität von Geo- und Sensordaten dar.

Mithilfe der Erkenntnisse der experimentellen Studie werden die Forschungsfragen beantwortet.

Mit welchen Standards und Richtlinien kann die Qualität von Sensordaten beschrieben werden?

Die Qualität von Geodaten wird im ISO-Standard ISO 19157:2013 adressiert. Bestandteile davon sind die Vollständigkeit, die logische Konsistenz, die Positionsgenauigkeit, die temporale Genauigkeit, die thematische Genauigkeit und die Benutzerfreundlichkeit. Die Ergebnisse der Fallstudie zeigen,

dass sich Messungenauigkeiten von Feinstaubsensoren primär in der Positionsgenauigkeit, der temporalen Genauigkeit und der thematischen Genauigkeit zeigen. Eine weitere Richtlinie wird von Klein et al. (2009) diskutiert. Diese beinhaltet die Genauigkeit, das Vertrauen, die Vollständigkeit, das Volumen und die Aktualität. Hier zeigen die Ergebnisse der Fallstudie vornehmlich Einflüsse auf die Genauigkeit (statische Faktoren) und das Vertrauen (dynamische Faktoren). Informationen zur Qualität von Sensordatensätzen können nach der Richtlinie *ISO 19115-2:2019* in den Metadaten gespeichert werden.

Welche Einflussgrößen wirken auf die Qualität mobil erfasster Feinstaub-sensordaten?

Als Faktoren, die die Qualität mobil erfasster Feinstaubsensordaten beeinflussen, werden die Positionierungsgenauigkeit, die Sensorspezifikationen, die Witterungsbedingungen, die Fahrgeschwindigkeit, die Sensorausrichtung und der Umgebungsverkehr ausgemacht. Die Sensorspezifikationen können sowohl für unterschiedliche Messergebnisse verschiedener, als auch identischer Sensoren sorgen. Hinsichtlich der Witterungsbedingungen lässt sich lediglich ein Einfluss der Luftfeuchtigkeit ausmachen, während die Temperatur keine Wirkung auf die Datenqualität hat. Die Fahrgeschwindigkeit und die Sensorausrichtung beeinflussen die Genauigkeit erfasster Feinstaubsensordaten, auch wenn die Höhe des Einflusses mit dem aktuellen Stand der Forschung noch nicht bestimmbar ist. Der Umgebungsverkehr während einer Messung kann durch die direkte Nähe der Messgeräte zu Emissionsquellen für einen starken Einfluss sorgen.

Welche Einflussgrößen sind statisch, und welche Einflussgrößen ändern sich dynamisch während einer Messung?

Die untersuchten Einflussfaktoren nehmen sowohl statischen, als auch dynamischen Einfluss auf die Datenqualität. Die Positionierungsgenauigkeit ist dabei sowohl statisch, als auch dynamisch. Die Art der Positionsbestimmung (aGPS, dGPS) ist während einer Messung gleichbleibend. Die Anzahl der empfangbaren Satelliten und der relative Winkel der Satelliten zum Empfänger sind Änderungen unterworfen. Außerdem lassen sich saisonale Veränderungen der Positionierungsgenauigkeit feststellen. Die Sensorspezifikationen sind ein statischer Faktor. Betrachtet über einen längeren Zeitraum kann es jedoch zu Änderungen der Genauigkeit von Messgeräten kommen. Witterungs-

bedingungen, und damit auch die Luftfeuchtigkeit, ändern sich dynamisch. Die Fahrgeschwindigkeit gilt ebenfalls als ein dynamischer Faktor. Statischen Einfluss nimmt die Sensorausrichtung, die sich in der Regel während einer Messung nicht ändert. Der Umgebungsverkehr ist dynamisch und vor einer Messung nur schwer vorhersehbar.

Welche der Einflussgrößen können mithilfe von Datenkorrektur und Kalibrierung eliminiert werden?

Modelle und Algorithmen zur Korrektur von Messungenauigkeiten existieren für die Positionierungsgenauigkeit, die Sensorspezifikationen und die Luftfeuchtigkeit. Positionierungsungenauigkeiten lassen sich durch Mapmatching Algorithmen ausgleichen. Liegen Messungenauigkeiten durch die limitierte Genauigkeit von Sensoren vor, kann man diese kalibrieren. Zur Korrektur von Witterungsbedingungen existieren Algorithmen, die den Einfluss der Luftfeuchtigkeit durch Wachstumsfaktoren ausgleichen. Zur Korrektur von Ungenauigkeiten durch die Fahrgeschwindigkeit, die Sensorausrichtung und den Umgebungsverkehr existieren noch keine Lösungen.

Ist der Umgebungsverkehr während einer mobilen Erfassung von Feinstaubdaten ein Störfaktor oder ein Teil der natürlichen Feinstaubkonzentration?

Die Experimente der Fallstudie zeigen Indizien für einen starken Einfluss des Umgebungsverkehrs auf die Genauigkeit der Messergebnisse. Dieser Einfluss lässt sich nur schwer umgehen, da sich bei mobilen Datenerfassungen auf der Straße die Messgeräte in direkter Nähe zu Emissionsquellen bewegen. Datenpunkte, die davon betroffen sind, zeigen deutlich erhöhte Feinstaubkonzentrationen. Daher kann der Umgebungsverkehr im Falle von mobilen Messungen auf Straßen als Störfaktor definiert werden.

Die vorliegende Dissertation zeigt, wie durch Geoinformationsanalysen mit Qualitätsfaktoren hochdynamischer Daten umgegangen werden kann. Dabei wird deutlich, dass die Nutzung klassischer Methoden aus der angewandten Geoinformatik zur Analyse dieser multivariaten Daten nicht ausreichend ist. Es ist vielmehr eine Kombination von Methodik aus unterschiedlichen Fachbereichen wie der Geostatistik oder dem computerbasierenden

Sehen, die dabei ans Ziel führt. Ferner zeigt sich, dass die Auseinandersetzung mit der Qualität eines Datensatzes ein bedeutender Arbeitsschritt ist. Eine Vielzahl von Einflüssen unterschiedlicher Ausprägung kann dazu führen, dass Analyseergebnisse nur limitiert gültig sind. Diese Einflüsse können dabei sowohl von internen Faktoren (Sensorqualität) stammen, als auch durch externe Ereignisse (Wetter, Umgebungsverkehr) auftreten.

Die Erkenntnisse aus der Arbeit mit Qualitätsfaktoren mobil erfasster Feinstaubsensordaten leisten einen wichtigen Beitrag zur Entwicklung vernetzter, sensorbasierter Technologien. Dazu gehören unter anderem das autonome Fahren oder Anwendungen in intelligenten Transportsystemen. Die Implementierung solcher Systeme stellt nicht nur hohe Anforderungen an das Datenvolumen, sondern auch an die Datenqualität. Es wird deutlich, dass neben der zeitlichen Komponente auch vornehmlich die räumliche Komponente einer Sensormessung für die Datenqualität entscheidend ist. Es ist dabei nicht ausreichend, nur statische Einflussfaktoren zu analysieren. Durch die Bewegung eines Sensors durch den Raum treten Messungenauigkeiten unterschiedlichster Art auf, die bei der Datenanalyse beachtet werden müssen. Die Dynamik einer Sensormessung stellt sich als besonders kritisch dar und wird in zukünftigen Arbeiten mit Feinstaubsensoren noch weiter an Bedeutung gewinnen.

8.2 Zukünftige Arbeiten

Folglich der Erkenntnisse der Studie und der Ablehnung der Hypothese bleiben einige Fragestellungen offen. Diese resultieren in weiteren potentiellen Forschungsfragen, die in zukünftigen Arbeiten bearbeitet werden.

Wie können Einflussfaktoren der Datenqualität mobil erfasster Feinstaubsensordaten in bestehende Metadatenstandards und -richtlinien integriert werden?

Neben der Überwachung von Qualitätsparametern ist auch die Dokumentation dieser ein wichtiger Vorgang bei der Arbeit mit Sensordaten. Es stellt sich die Frage, wie die untersuchten Faktoren in bestehende Richtlinien und Standards, wie beispielsweise *ISO 19115-2:2019*, integriert werden können. Die Metadatenstruktur nach *ISO 19115-2:2019* zur Erfassung und Prozessierung von Sensordaten sieht unter anderem Informationen über die Temperatur, die Luftfeuchtigkeit und meteorologische Bedingungen vor. Des weiteren

ist definiert, wie Informationen über verwendete Messgeräte und Sensoren dokumentiert werden sollen. Fraglich ist, inwiefern potentielle Störquellen wie der Umgebungsverkehr und die Fahrgeschwindigkeit in den Metadaten beschrieben werden können (Brodeur et al. 2019).

In zukünftigen Arbeiten wird untersucht, wie die in dieser Studie behandelten Qualitätsparameter in Metadatenstrukturen integriert werden können.

Welche infrastrukturellen Umgebungsfaktoren beeinflussen die Datenqualität mobil erfasster Feinstaubsensordaten?

Die Umgebung einer Messung ist komplex. Mit dem Umgebungsverkehr und den Witterungsbedingungen werden bereits zwei Faktoren untersucht, die von außen Einfluss auf die Datenqualität nehmen. Nicht berücksichtigt werden bisher infrastrukturelle Einflüsse wie Straßentunnel. Diese beeinträchtigen jedoch in der Regel die Messung mit Positionierungssensoren, da die Kommunikation mit Satelliten dort nur bedingt möglich ist. Darüber hinaus können sich innerhalb von Tunneln Feinstaubpartikel vermehrt ansammeln und absetzen, weil diese nur durch den Tunnelein- und ausgang entweichen können. Ein konstanter Luftfluss wird darüber hinaus lediglich durch Ventilationssysteme aufrecht erhalten (Hung-Lung et al. 2009).

Ein weiterer Umgebungsfaktor, der die Datenqualität beeinträchtigen könnte, sind hohe Bebauungsgrade am Straßenrand. Wenn beide Straßenseiten eines Abschnittes über eine größere Distanz von höheren Gebäuden bedeckt werden, spricht man von einer *Straßenschlucht*. Studien wie De Nicola et al. (2013) schließen, dass die Feinstaubkonzentration entlang von Straßenschluchten einigen Besonderheiten unterliegt. So beobachten die Autoren, dass gemessene Feinstaubwerte an einer Seite der Straßenschlucht stets höher sind als auf der anderen Seite (De Nicola et al. 2013).

In zukünftigen Arbeiten werden diese Umgebungsfaktoren mit in Betracht gezogen. Essentiell wird dabei die Erfassung zusätzlicher Datensätze sein. Informationen über den Verlauf der Straßen können Netzwerkdatensätzen entnommen werden. Des weiteren können Informationen über Gebäude mithilfe digitaler Oberflächenmodelle (DOM) oder dreidimensionalen Städtmodellen extrahiert werden, um Messfehler bedingt durch Straßenschluchten zu untersuchen.

Wie können Qualitätsparameter in Echtzeit überwacht werden?

Im Zuge einiger Analysen wird nicht nur auf bereits bestehende Datensätze zurückgegriffen. In manchen Fällen müssen Sensordaten in Echtzeit erfasst und analysiert werden. Das betrifft beispielsweise Anwendungen des autonomen Fahrens, die auf hochaktuelle Daten angewiesen sind. Videokameras und Laserscanner liefern dabei Informationen, die kurz nach der Datenerfassung bereits zur Entscheidungsfindung hinsichtlich der befahrenen Trajektorie genutzt werden. Auch in größeren Sensornetzwerken ist die Qualitätsüberwachung in Echtzeit ein wichtiges Thema. Dabei können Messfehler zur Laufzeit der Sensoren erkannt werden, sodass nach den Ursachen der fehlerhaften Messwerte gesucht werden kann. Ein Lösungsansatz dafür ist das *Internet of Things* (IoT), bei dem Akteure in einem Netzwerk mithilfe von Kommunikationstechniken Daten in Echtzeit übertragen können. She et al. (2019) nennen die Blockchain-Technologie als einen wichtigen Bestandteil zukünftiger Forschung im Bereich der Qualitätsüberwachung von Sensordaten. Dabei agieren die Sensoren als Knotenpunkte in einem dezentralisiertem Netzwerk. Das führt dazu, dass jeder Sensor zur Laufzeit über die Messwerte der anderen Messgeräte im Netz Bescheid weiß. Konsequenterweise kann so schnell ermittelt werden, wenn ein Netzwerkknoten fehlerhafte Werte erfasst. Ein ähnlicher Ansatz zur Qualitätsüberwachung findet sich in Zheng et al. (2019). Die Autoren verwenden ein Gaußprozess-Regressionsmodell, um Sensoren in einem Netzwerk dynamisch zu überwachen und zu kalibrieren.

Wie genau kann die Feinstaubkonzentration mithilfe von ML-Algorithmen mit Faktoren der Datenqualität vorhergesagt werden?

In einem Experiment der Fallstudie wird die Feinstaubkonzentration mithilfe von Prediktorvariablen, die aus der Analyse der Qualitätsfaktoren stammen, vorhergesagt. Dabei kommen die ML-Algorithmen *Lineare Regression*, *Gradient Boost* und *Random Forest* sowie die Prediktorvariablen *Fahrgeschwindigkeit*, *Luftfeuchtigkeit*, *Busse*, *Kraftfahrzeuge*, *Motorräder* und *Lastkraftwagen* zum Einsatz. Es stellt sich heraus, dass relative Veränderungen der Feinstaubkonzentration mithilfe der implementierten Modelle vorhersehbar sind. Gegenstand zukünftiger Forschung wird die Optimierung dieser Modelle sein. Dafür werden weitere ML-Algorithmen getestet. Des Weiteren soll durch Hyperparameter Tuning die Genauigkeit der verwendeten Modelle erhöht werden. Außerdem wird geprüft, ob sich die Vorhersagegenauigkeit durch die Hinzu-

nahme weiterer Prediktorvariablen oder die Wegnahme bestehender Parameter ändert.

Literaturverzeichnis

- Abedi, Naeim, Ashish Bhaskar und Edward Chung (2013). „Bluetooth and Wi-Fi MAC address based crowd collection and monitoring: benefits, challenges and enhancement“. In: *Australasian Transport Research Forum 2013 Proceedings*. Australasian Transport Research Forum, S. 1–17.
- Ahn, Kang-Ho, Handol Lee, Hae Dong Lee und Sang Chul Kim (2020). „Extensive evaluation and classification of low-cost dust sensors in laboratory using a newly developed test method“. In: *Indoor air* 30.1, S. 137–146.
- Albawi, Saad, Tareq Abed Mohammed und Saad Al-Zawi (2017). „Understanding of a convolutional neural network“. In: *2017 International Conference on Engineering and Technology (ICET)*. IEEE, S. 1–6.
- Alessandrini, Adriano, Francesco Filippi und Fernando Ortenzi (2012). „Consumption calculation of vehicles using OBD data“. In: *20th International Emission Inventory Conference- Emission Inventories-Meeting the Challenges Posed by Emerging Global, National, and Regional and Local Air Quality Issues*.
- Alpatov, Boris A, Pavel V Babayan und Maksim D Ershov (2018). „Vehicle detection and counting system for real-time traffic surveillance“. In: *2018 7th Mediterranean Conference on Embedded Computing (MECO)*. IEEE, S. 1–4.
- Altintasi, Oruc, Hediye Tuydes-Yaman und Kagan Tuncay (2017). „Detection of urban traffic patterns from Floating Car Data (FCD)“. In: *Transportation research procedia* 22, S. 382–391.
- Anjomshoaa, Amin, Fábio Duarte, Daniël Rennings, Thomas J Matarazzo, Priyanka deSouza und Carlo Ratti (2018). „City scanner: Building and scheduling a mobile sensing platform for smart city services“. In: *IEEE Internet of things Journal* 5.6, S. 4567–4579.
- Ariola, V, A D’Alessandro, F Lucarelli, G Marcazzan, F Mazzei, S Nava, I Garcia-Orellana, P Prati, G Valli und R Vecchi (2006). „Elemental characterization of PM₁₀, PM_{2.5} and PM₁ in the town of Genoa (Italy)“. In: *Chemosphere* 62.2, S. 226–232.
- Artusi, R, P Verderio und E Marubini (2002). „Bravais-Pearson and Spearman correlation coefficients: meaning, test of hypothesis and confidence interval“. In: *The International journal of biological markers* 17.2, S. 148–151.
- Assenmacher, Walter (2013). *Deskriptive Statistik*. Springer-Verlag.
- Bai, Lu, Lin Huang, Zhenglu Wang, Qi Ying, Jun Zheng, Xiaowen Shi und Jianlin Hu (2020). „Long-term field evaluation of low-cost particulate matter sensors in Nanjing“. In: *Aerosol and Air Quality Research* 20.2, S. 242–253.

- Báthory, Csongor und Arpad Bence Palotas (2019). „Hotspot identification with portable low-cost particulate matter sensor“. In: *International Journal of Energy Water Food Nexus* 1.1, S. 13–17.
- Bella, G und VK Krastev (2011). „On the rans modeling of turbulent airflow over a simplified car model“. In: *Fluids Engineering Division Summer Meeting*. Bd. 44403, S. 871–883.
- Benabbas, Aboubakr, Martin Geißelbrecht, Gabriel Martin Nikol, Lukas Mahr, Daniel Nähr, Simon Steuer, Gabriele Wiesemann, Thomas Müller, Daniela Nicklas und Thomas Wieland (2019). „Measure particulate matter by yourself: data-quality monitoring in a citizen science project“. In: *Journal of Sensors and Sensor Systems* 8.2, S. 317–328.
- Blazquez, Carola A und Alan P Vonderohe (2005). „Simple map-matching algorithm applied to intelligent winter maintenance vehicle data“. In: *Transportation Research Record* 1935.1, S. 68–76.
- Blogg, Miranda, Conor Semler, Manu Hingorani und Rod Troutbeck (2010). „Travel time and origin-destination data collection using Bluetooth MAC address readers“. In: *Australasian transport research forum*. Bd. 36.
- Bochkovskiy, Alexey, Chien-Yao Wang und Hong-Yuan Mark Liao (2020). „YOLOv4: Optimal Speed and Accuracy of Object Detection“. In: *arXiv preprint arXiv:2004.10934*.
- Breiman, Leo (1997). *Arcing the edge*. Techn. Ber. Berkeley, CA, USA: Technical Report 486, Statistics Department, University of California at Berkeley.
- Breiman, Leo (2001). „Random forests“. In: *Machine learning* 45.1, S. 5–32.
- Brockfeld, Elmar, Peter Wagner, Stefan Lorkowski und Peter Mieth (2007). „Benefits and limits of recent floating car data technology—an evaluation study“. In: *CDROM-WCTR2007*, S. C2–830.
- Brodeur, Jean, Serena Coetzee, David Danko, Stéphane Garcia und Jan Hjelmer (2019). „Geographic information metadata—an outlook from the international standardization perspective“. In: *ISPRS International Journal of Geo-Information* 8.6, S. 280.
- Budde, Matthias, Simon Leiner, Marcel Köpke, Johannes Riesterer, Till Riedel und Michael Beigl (2019). „FeinPhone: Low-cost smartphone camera-based 2D particulate matter sensor“. In: *Sensors* 19.3, S. 749.
- Bulot, Florentin MJ, Steven J Johnston, Philip J Basford, Natasha HC Easton, Mihaela Apetroaie-Cristea, Gavin L Foster, Andrew KR Morris, Simon J Cox und Matthew Loxham (2019). „Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment“. In: *Scientific reports* 9.1, S. 1–13.

- Campbell, John L, Lindsey E Rustad, John H Porter, Jeffrey R Taylor, Ethan W Dereszynski, James B Shanley, Corinna Gries, Donald L Henshaw, Mary E Martin und Wade M Sheldon (2013). „Quantity is nothing without quality: automated QA/QC for streaming environmental sensor data“. In: *BioScience* 63.7, S. 574–585.
- Cerqueira, Vitor, Luis Moreira-Matias, Jihed Khiari und Hans van Lint (2018). „On Evaluating Floating Car Data Quality for Knowledge Discovery“. In: *IEEE Transactions on Intelligent Transportation Systems* 19.11, S. 3749–3760.
- Chakrabarti, Bhabesh, Philip M Fine, Ralph Delfino und Constantinos Sioutas (2004). „Performance evaluation of the active-flow personal DataRAM PM2.5 mass monitor (Thermo Anderson pDR-1200) designed for continuous personal exposure measurements“. In: *Atmospheric Environment* 38.20, S. 3329–3340.
- Chambers, John M (2018). *Graphical methods for data analysis*. CRC Press.
- Chen, Zhongfei, Xinyue Hao, Xiaoyu Zhang und Fanglin Chen (2021). „Have traffic restrictions improved air quality? A shock from COVID-19“. In: *Journal of cleaner production* 279, S. 123622.
- Choudhury, Charisma, Moshe E Ben-Akiva, Tomer Toledo, Anita Rao und Gunwoo Lee (2007). *State dependence in lane changing models*.
- Coffel, Ethan D, Radley M Horton und Alex de Sherbinin (2017). „Temperature and humidity based projections of a rapid rise in global heat stress exposure during the 21st century“. In: *Environmental Research Letters* 13.1, S. 014001.
- Cuadrat, José M, Sergio Vicente-Serrano und Miguel A Saz (2015). „Influence of different factors on relative air humidity in Zaragoza, Spain“. In: *Frontiers in Earth Science* 3, S. 10.
- De Fabritiis, Corrado, Roberto Ragona und Gaetano Valenti (2008). „Traffic estimation and prediction based on real time floating car data“. In: *2008 11th international IEEE conference on intelligent transportation systems*. IEEE, S. 197–203.
- De Lange, Norbert (2013). *Geoinformatik: in Theorie und Praxis*. Springer-Verlag.
- De Nicola, Flavia, Fabio Murena, M Antonietta Costagliola, Anna Alfani, Daniela Baldantoni, M Vittoria Prati, Ludovica Sessa, Valeria Spagnuolo und Simonetta Giordano (2013). „A multi-approach monitoring of particulate matter, metals and PAHs in an urban street canyon“. In: *Environmental Science and Pollution Research* 20.7, S. 4969–4979.
- Di Antonio, Andrea, Olalekan AM Popoola, Bin Ouyang, John Saffell und Roderic L Jones (2018). „Developing a relative humidity correction for

- low-cost sensors measuring ambient particulate matter". In: *Sensors* 18.9, S. 2790.
- Dimitrakopoulos, George und Panagiotis Demestichas (2010). „Intelligent transportation systems". In: *IEEE Vehicular Technology Magazine* 5.1, S. 77–84.
- Dogan, U, M Uludag und DO Demir (2014). „Investigation of GPS positioning accuracy during the seasonal variation". In: *Measurement* 53, S. 91–100.
- Dogramadzi, Marko und Aftab Khan (2021). „Accelerated Map Matching for GPS Trajectories". In: *IEEE Transactions on Intelligent Transportation Systems*.
- Downes, Barbara J, Leon A Barmuta, Peter G Fairweather, Daniel P Faith, Michel J Keough, Phillip Sam Lake, Bruce D Mapstone und Gerry P Quinn (2002). *Monitoring ecological impacts: concepts and practice in flowing waters*. Cambridge University Press.
- Ebendt, Rüdiger, Alexander Sohr, Touko Tcheumadjeu, Louis Calvin und Peter Wagner (2010). „Utilizing historical and current travel times based on floating car data for management of an express truck fleet". In: *5th International Scientific Conference: Theoretical and Practical Issues in Transport*.
- Eiter, Thomas und Heikki Mannila (1994). *Computing discrete Fréchet distance*. Techn. Ber. TU Vienna, Austria: Christian Doppler Laboratory for Expert Systems.
- Erkan, Ilker und Hasan Hastemoglu (2016). „Bluetooth as a traffic sensor for stream travel time estimation under Bogazici Bosphorus conditions in Turkey". In: *Journal of Modern Transportation* 24.3, S. 207–214.
- Fahrmeir, Ludwig, Christian Heumann, Rita Künstler, Iris Pigeot und Gerhard Tutz (2016). *Statistik: Der weg zur datenanalyse*. Springer-Verlag.
- Fang, Jianxin, Huadong Meng, Hao Zhang und Xiqin Wang (2007). „A low-cost vehicle detection and classification system based on unmodulated continuous-wave radar". In: *2007 IEEE Intelligent Transportation Systems Conference*. IEEE, S. 715–720.
- Ford, Bonne, Jeffrey R Pierce, Eric Wendt, Marilee Long, Shantanu Jathar, John Mehaffy, Jessica Tryner, Casey Quinn, Lizette van Zyl und Christian L'Orange (2019). „A low-cost monitor for measurement of fine particulate matter and aerosol optical depth–Part 2: Citizen-science pilot campaign in northern Colorado". In: *Atmospheric Measurement Techniques* 12.12, S. 6385–6399.
- Fréchet, M Maurice (1906). „Sur quelques points du calcul fonctionnel". In: *Rendiconti del Circolo Matematico di Palermo (1884-1940)* 22.1, S. 1–72.

- Freire, Sandra M, Helder Relvas und Myriam Lopes (2020). „Impact of traffic emissions on air quality in Cabo Verde“. In: *Environmental Monitoring and Assessment* 192.11, S. 1–13.
- Friedman, Jerome H (2001). „Greedy function approximation: a gradient boosting machine“. In: *Annals of statistics*, S. 1189–1232.
- Friesen, Marcia R und Robert D McLeod (2015). „Bluetooth in intelligent transportation systems: a survey“. In: *International Journal of Intelligent Transportation Systems Research* 13.3, S. 143–153.
- Fritz, Steffen, Linda See, Tyler Carlson, Mordechai Muki Haklay, Jessie L Oliver, Dilek Fraisl, Rosy Mondardini, Martin Brocklehurst, Lea A Shanley und Sven Schade (2019). „Citizen science and the United Nations sustainable development goals“. In: *Nature Sustainability* 2.10, S. 922–930.
- Gajda, Janusz, Piotr Piwowar, Ryszard Sroka, Marek Stencel und Tadeusz Zeglen (2012). „Application of inductive loops as wheel detectors“. In: *Transportation Research Part C: Emerging Technologies* 21.1, S. 57–66.
- Gao, Yi, Wei Dong, Kai Guo, Xue Liu, Yuan Chen, Xiaojin Liu, Jiajun Bu und Chun Chen (2016). „Mosaic: A low-cost mobile sensing system for urban air quality monitoring“. In: *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, S. 1–9.
- Garcia, Fernando, Pietro Cerri, Alberto Broggi, Jose Maria Armingol und Arturo De La Escalera (2009). „Vehicle detection based on laser radar“. In: *International Conference on Computer Aided Systems Theory*. Springer, S. 391–397.
- Gehrig, Robert, Matz Hill, Brigitte Buchmann, David Imhof, Ernest Weingartner und Urs Baltensperger (2004). „Separate determination of PM₁₀ emission factors of road traffic for tailpipe emissions and emissions from abrasion and resuspension processes“. In: *International Journal of Environment and Pollution* 22.3, S. 312–325.
- Gökaşar, İlgin und Yiğit Çetinel (2017). „Analysis of the Effect of the Lane Drops on the Traffic near Bus Stops using Bus GPS Data“. In: *Procedia Computer Science* 109, S. 466–473.
- Gong, Yue-Jiao, En Chen, Xinglin Zhang, Lionel M Ni und Jun Zhang (2017). „AntMapper: An ant colony-based map matching approach for trajectory-based applications“. In: *IEEE Transactions on Intelligent Transportation Systems* 19.2, S. 390–401.
- Goyal, Tanishka, Ankita Singh, Smriti Chhaya, Aditi Vikas, Poorva Garg, Ritika Malik und Rijurekha Sen (2018). „Low cost platform design for pollution measurement in delhi-ncr using vehicle-mounted sensors“. In: *Pro-*

- ceedings of the 24th Annual International Conference on Mobile Computing and Networking*, S. 759–761.
- Gozzi, Fernando, Giancarlo Della Ventura und Augusto Marcelli (2016). „Mobile monitoring of particulate matter: State of art and perspectives“. In: *Atmospheric Pollution Research* 7.2, S. 228–234.
- Graves, Spencer (2000). „Why do we need international standards?“ In: *Quality Progress* 33.6, S. 138.
- Gurczik, Gaby (2016). „Bluetooth-Based Floating Car Observer: Model Evaluation Using Simulation and Field Measurements“. In: *German Aerospace Center (DLR) Institute of Transportation Systems*.
- Gurczik, Gaby (2017). „Performance measurement of a bluetooth-based floating car observer“. In: *Transportation research procedia* 25, S. 1834–1845.
- Habre, Rima, Mariam Girguis, Robert Urman, Scott Fruin, Fred Lurmann, Martin Shafer, Patrick Gorski, Meredith Franklin, Rob McConnell und Ed Avol (2021). „Contribution of tailpipe and non-tailpipe traffic sources to quasi-ultrafine, fine and coarse particulate matter in southern California“. In: *Journal of the Air & Waste Management Association* 71.2, S. 209–230.
- Hagler, Gayle SW, Eben D Thoma und Richard W Baldauf (2010). „High-resolution mobile monitoring of carbon monoxide and ultrafine particle concentrations in a near-road environment“. In: *Journal of the Air & Waste Management Association* 60.3, S. 328–336.
- Hagler, Gayle SW, Ronald Williams, Vasileios Papapostolou und Andrea Polidori (2018). *Air quality sensors and data adjustment algorithms: When is it no longer a measurement?*
- Hauser, Helwig, Florian Ledermann und Helmut Doleisch (2002). „Angular brushing of extended parallel coordinates.“ In: *INFOVIS*. Bd. 2, S. 127.
- Heydari, Shahram, Marko Tainio, James Woodcock und Audrey de Nazelle (2020). „Estimating traffic contribution to particulate matter concentration in urban areas using a multilevel Bayesian meta-regression approach“. In: *Environment international* 141, S. 105800.
- Holland, Stephen P, Erin T Mansur, Nicholas Z Muller und Andrew J Yates (2019). „Distributional effects of air pollution from electric vehicle adoption“. In: *Journal of the Association of Environmental and Resource Economists* 6.S1, S65–S94.
- Howerton, James Montgomery und Benjamin Leo Schenck (2020). „The deployment of a LoRaWAN-based IoT air quality sensor network for public good“. In: *2020 Systems and Information Engineering Design Symposium (SIEDS)*. IEEE, S. 1–6.

- Hu, Yidan, Guojun Dai, Jin Fan, Yifan Wu und Hua Zhang (2016). „BlueAer: A fine-grained urban PM_{2.5} 3D monitoring system using mobile sensing“. In: *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, S. 1–9.
- Hung-Lung, Chiang und Huang Yao-Sheng (2009). „Particulate matter emissions from on-road vehicles in a freeway tunnel study“. In: *Atmospheric Environment* 43.26, S. 4014–4022.
- ISO Central Secretary (Dez. 2013). *Geographic information — Data quality*. Standard ISO/TC 211 TR 19157:2013. Geneva, CH: International Organization for Standardization. URL: <https://www.iso.org/standard/32575.html>.
- ISO Central Secretary (Jan. 2019). *Geographic information — Metadata — Part 2: Extensions for acquisition and processing*. Standard ISO/TC 211 TR 19115-2:2019. Geneva, CH: International Organization for Standardization. URL: <https://www.iso.org/standard/67039.html>.
- Jekel, Charles F, Gerhard Venter, Martin P Venter, Nielen Stander und Raphael T Haftka (Mai 2019). „Similarity measures for identifying material parameters from hysteresis loops using inverse analysis“. In: *International Journal of Material Forming*. ISSN: 1960-6214.
- Jo, Kichun und Myounggho Sunwoo (2013). „Generation of a precise roadway map for autonomous cars“. In: *IEEE Transactions on intelligent transportation systems* 15.3, S. 925–937.
- Jonkeren, Olaf, Bart Jourquin und Piet Rietveld (2011). „Modal-split effects of climate change: The effect of low water levels on the competitive position of inland waterway transport in the river Rhine area“. In: *Transportation Research Part A: Policy and Practice* 45.10, S. 1007–1019.
- Keler, Andreas und Jukka M Krisp (2015). „Spatio-temporal visualization of interpolated particulate matter (PM_{2.5}) in Beijing“. In: *GI_Forum—Journal for Geographic Information Science*, S. 464–474.
- Kelly, Frank J und Julia C Fussell (2012). „Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter“. In: *Atmospheric environment* 60, S. 504–526.
- Kerner, BS, C Demir, RG Herrtwich, SL Klenov, H Rehborn, M Aleksic und A Haug (2005). „Traffic state detection with floating car data in road networks“. In: *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005*. IEEE, S. 44–49.
- Khan, Salman, Hossein Rahmani, Syed Afaq Ali Shah und Mohammed Benamoun (2018). „A guide to convolutional neural networks for computer vision“. In: *Synthesis Lectures on Computer Vision* 8.1, S. 1–207.

- Klein, Anja und Wolfgang Lehner (2009). „Representing data quality in sensor data streaming environments“. In: *Journal of Data and Information Quality (JDIQ)* 1.2, S. 1–28.
- Klette, Reinhard (2014). *Concise computer vision*. Springer.
- Kühnel, Carsten (2012). *Verkehrsdatenerfassung mittels Floating Car Observer auf zweistreifigen Landstraßen*. Bd. 23. kassel university press GmbH.
- Kun, Attila Jozsef und Zoltan Vamossy (2009). „Traffic monitoring with computer vision“. In: *2009 7th International Symposium on Applied Machine Intelligence and Informatics*. IEEE, S. 131–134.
- Lawson, Anneka Ruth, Bidisha Ghosh und Brian Broderick (2011). „Prediction of traffic-related nitrogen oxides concentrations using Structural Time-Series models“. In: *Atmospheric environment* 45.27, S. 4719–4727.
- Learned-Miller, Erik G (2011). „Introduction to computer vision“. In: *University of Massachusetts, Amherst*.
- Lee, Hoochang, Jiseock Kang, Sungjung Kim, Yunseok Im, Seungjung Yoo und Dongjun Lee (2020). „Long-term evaluation and calibration of low-cost particulate matter (pm) sensor“. In: *Sensors* 20.13, S. 3617.
- Lee, Joseph K, Andreas Christen, Rick Ketler und Zoran Nesic (2017). „A mobile sensor network to map carbon dioxide emissions in urban environments“. In: *Atmospheric Measurement Techniques* 10.2, S. 645–665.
- Li, Shuguang, Hongkai Yu, Jingru Zhang, Kaixin Yang und Ran Bin (2013). „Video-based traffic data collection system for multiple vehicle types“. In: *IET Intelligent Transport Systems* 8.2, S. 164–174.
- Li, Yu, Ji Zheng, Suocheng Dong, Xuefa Wen, Xueting Jin, Leiming Zhang und Xiao Peng (2019). „Temporal variations of local traffic CO₂ emissions and its relationship with CO₂ flux in Beijing, China“. In: *Transportation Research Part D: Transport and Environment* 67, S. 1–15.
- Liang, Chen-Jui und Pei-Rong Yu (2021). „Assessment and Improvement of Two Low-Cost Particulate Matter Sensor Systems by Using Spatial Interpolation Data from Air Quality Monitoring Stations“. In: *Atmosphere* 12.3, S. 300.
- Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He und Piotr Dollár (2017). „Focal loss for dense object detection“. In: *Proceedings of the IEEE international conference on computer vision*, S. 2980–2988.
- Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár und C Lawrence Zitnick (2014). „Microsoft coco: Common objects in context“. In: *European conference on computer vision*. Springer, S. 740–755.

- Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu und Alexander C Berg (2016). „Ssd: Single shot multibox detector“. In: *European conference on computer vision*. Springer, S. 21–37.
- Liu, Xiaofeng, Chaosheng Xiang, Bin Li und Aimin Jiang (2015). „Collaborative bicycle sensing for air pollution on roadway“. In: *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*. IEEE, S. 316–319.
- Lorkowski, Stefan, Elmar Brockfeld, Peter Mieth, Bert Passfeld, Kai-Uwe Thiesenhusen und Ralf-Peter Schäfer (2003). „Erste mobilitätsdienste auf basis von"floating car data"“. In: *Tagungsband zum 4. Aachener Kollouium"Mobilität und Stadt" 75*, S. 93–100.
- Lyu, Haiyang, Yehua Sheng, Ningning Guo, Baoqun Huang und Siyang Zhang (2017). „Geometric quality assessment of trajectory-generated VGI road networks based on the symmetric arc similarity“. In: *Transactions in GIS* 21.5, S. 984–1009.
- Maag, Balz, Zimu Zhou und Lothar Thiele (2018). „A survey on sensor calibration in air pollution monitoring deployments“. In: *IEEE Internet of Things Journal* 5.6, S. 4857–4870.
- Mahajan, Sachit, Prashant Kumar, Janaina Antonino Pinto, Agnese Riccetti, Katinka Schaaf, Guillem Camprodon, Viktor Smári, Antonella Passani und Giuseppe Forino (2020). „A citizen science approach for enhancing public understanding of air pollution“. In: *Sustainable Cities and Society* 52, S. 101800.
- Mahto, Pooja, Priyamm Garg, Pranav Seth und J Panda (2020). „Refining YoloV4 for Vehicle Detection“. In: *International Journal of Advanced Research in Engineering and Technology (IJARET)* 11.5.
- Makowski, Lukasz (2018). „Mobile Sensor Network for Suspended Particulate Matter Measurements using Kriging Interpolation“. In: *19th International Conference Computational Problems of Electrical Engineering*. IEEE, S. 1–4.
- Masic, Adnan, Boran Pikula und Dzevad Bibic (2017). „Mobile measurements of particulate matter concentrations in urban area“. In: *Proceedings of the 28th DAAAM International Symposium*, S. 0452–0456.
- Mazzoleni, Claudio, Hampden D Kuhns, Hans Moosmüller, Robert E Keislar, Peter W Barber, Norman F Robinson, John G Watson und Djordje Nikolic (2004). „On-road vehicle particulate matter and gaseous emission distribu-

- tions in Las Vegas, Nevada, compared with other areas". In: *Journal of the Air & Waste Management Association* 54.6, S. 711–726.
- Messelodi, Stefano, Carla M Modena, Michele Zanin, Francesco GB De Natale, Fabrizio Granelli, Enrico Betterle und Andrea Guarise (2009). „Intelligent extended floating car data collection". In: *Expert systems with applications* 36.3, S. 4213–4227.
- Mohri, Mehryar, Afshin Rostamizadeh und Ameet Talwalkar (2018). *Foundations of machine learning*. MIT press.
- Montgomery, Douglas C, Elizabeth A Peck und G Geoffrey Vining (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Morawska, Lidia, Phong K Thai, Xiaoting Liu, Akwasi Asumadu-Sakyi, Godwin Ayoko, Alena Bartonova, Andrea Bedini, Fahe Chai, Bryce Christensen und Matthew Dunbabin (2018). „Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?" In: *Environment international* 116, S. 286–299.
- Morera, Ángel, Ángel Sánchez, A Belén Moreno, Ángel D Sappa und José F Vélez (2020). „SSD vs. YOLO for detection of outdoor urban advertising panels under multiple variabilities". In: *Sensors* 20.16, S. 4587.
- Mukherjee, Anondo, Levi G Stanton, Ashley R Graham und Paul T Roberts (2017a). „Assessing the utility of low-cost particulate matter sensors over a 12-week period in the Cuyama valley of California". In: *Sensors* 17.8, S. 1805.
- Mukherjee, Arideep und Madhoolika Agrawal (2017b). „World air particulate matter: sources, distribution and health effects". In: *Environmental Chemistry Letters* 15.2, S. 283–309.
- Newson, Paul und John Krumm (2009). „Hidden Markov map matching through noise and sparseness". In: *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, S. 336–343.
- Oh, Seri, Stephen G Ritchie und Cheol Oh (2002). „Real-time traffic measurement from single loop inductive signatures". In: *Transportation Research Record* 1804.1, S. 98–106.
- Pan, Yuepeng, Shili Tian, Yuanhong Zhao, Lin Zhang, Xiaying Zhu, Jian Gao, Wei Huang, Yanbo Zhou, Yu Song und Qiang Zhang (2018). „Identifying ammonia hotspots in China using a national observation network". In: *Environmental science & technology* 52.7, S. 3926–3934.
- Pant, Pallavi, Raj M Lal, Sarath K Guttikunda, Armistead G Russell, Ajay S Nagpure, Anu Ramaswami und Richard E Peltier (2019). „Monitoring par-

- ticulate matter in India: recent trends and future outlook". In: *Air Quality, Atmosphere & Health* 12.1, S. 45–58.
- Parvez, Fatema, Carmen Lamancusa und Kristina Wagstrom (2017). „Primary and secondary particulate matter intake fraction from different height emission sources". In: *Atmospheric Environment* 165, S. 1–11.
- Pavlov, Yu L (2019). *Random forests*. Walter de Gruyter GmbH & Co KG.
- Prasad, BVP, Jing-Jou Tang und Sheng-Jhu Luo (2019). „Design and implementation of SAE J1939 vehicle diagnostics system". In: *2019 IEEE International Conference on Computation, Communication and Engineering (ICCCCE)*. IEEE, S. 71–74.
- Puckett, Darryl D und Michael J Vickich (2010). *Bluetooth-based travel time/speed measuring systems development*. Techn. Ber. TX, USA: Texas Transportation Institute.
- Qin, Xiaoliang, Lujian Hou, Jian Gao und Shuchun Si (2020). „The evaluation and optimization of calibration methods for low-cost particulate matter sensors: Inter-comparison between fixed and mobile methods". In: *Science of The Total Environment* 715, S. 136791.
- Rabiner, Lawrence und Biinghwang Juang (1986). „An introduction to hidden Markov models". In: *ieee assp magazine* 3.1, S. 4–16.
- Ranacher, Peter, Richard Brunauer, Stefan Van der Spek und Siegfried Reich (2016). „What is an appropriate temporal sampling rate to record floating car data with a GPS?" In: *ISPRS international journal of geo-information* 5.1, S. 1.
- Re, Giuseppe Lo, Daniele Peri und Salvatore Davide Vassallo (2014). „Urban air quality monitoring using vehicular sensor networks". In: *Advances onto the Internet of Things*. Springer, S. 311–323.
- Reddy, C Rajashekar, T Mukku, A Dwivedi, A Rout, S Chaudhari, K Vemuri, KS Rajan und AM Hussain (2020). „Improving Spatio-Temporal Understanding of Particulate Matter using Low-Cost IoT Sensors". In: *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, S. 1–7.
- Reddy, Sasank, Jeff Burke, Deborah Estrin, Mark Hansen und Mani Srivastava (2007). „A framework for data quality and feedback in participatory sensing". In: *Proceedings of the 5th international conference on Embedded networked sensor systems*, S. 417–418.
- Redmon, Joseph und Ali Farhadi (2018). „Yolov3: An incremental improvement". In: *arXiv preprint arXiv:1804.02767*.

- Rinke, R, A Wieser, B Vogel, U Corsmeier und Ch Kottmeier (2010). „Messungen der räumlichen Variabilität der Luftqualität in einem Ballungsraum mittels einer Straßenbahn“. In: *Programm Lebensgrundlage Umwelt und ihre Sicherung* "(BWPLUS) Zwischenbericht anlässlich des Statusseminars des BWPLUS" 24.
- Röger, Christian (2019). „Star-Plot-Karten für die visuelle Untersuchung von mobilen Sensordatensätzen“. In: *AGIT – Journal für Angewandte Geoinformatik*, S. 383–394.
- Röger, Christian (2020). „Quantifizierung des Umgebungsverkehrs eines Fahrzeuges durch die Messung von Straßenabrieb“. In: *AGIT – Journal für Angewandte Geoinformatik*, 6-2020, S. 290–295.
- Röger, Christian (2021). „Statische und Dynamische Unsicherheitsfaktoren bei Mobilien Messungen mit Feinstaubsensoren“. In: *AGIT – Journal für Angewandte Geoinformatik*, 6-2021.
- Röger, Christian und Irada Ismayilova (2020a). „Predicting ambient traffic of a vehicle from road abrasion measurements using random forest“. In: *Proceedings of the 13th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, S. 1–7.
- Röger, Christian, Maja Kalinic und Jukka M Krisp (2020b). „Extracting Densely Covered Areas Within Floating Car Datasets Using Inductive Loop Detector Data“. In: *KN-Journal of Cartography and Geographic Information*, S. 1–7.
- Röger, Christian, Andreas Keler und Jukka M Krisp (2018). „Examining the influence of road slope on carbon dioxide emission using extended floating car data“. In: *Adjunct Proceedings of the 14th International Conference on Location Based Services*. ETH Zurich, S. 135–140.
- Röger, Christian und Jukka M Krisp (2019). „Using Cartograms for Visualizing extended Floating Car Data (xFCD)“. In: *Proceedings of the ICA*. Bd. 2, S. 107.
- Rossi, Riccardo, Riccardo Ceccato und Massimiliano Gastaldi (2020). „Effect of Road Traffic on Air Pollution. Experimental Evidence from COVID-19 Lockdown“. In: *Sustainability* 12.21, S. 8984.
- Ryan, Matthew S und Graham R Nudd (1993). *The viterbi algorithm*. Techn. Ber. Coventry, UK: University of Warwick. Department of Computer Science.
- SCEP (1970). *Study of critical environmental problems*.
- Schäffer, Florian (2013). *OBD: Fahrzeugdiagnose in der Praxis*. Franzis Verlag.
- Schiavon, Marco, Martina Redivo, Gianluca Antonacci, Elena Cristina Rada, Marco Ragazzi, Dino Zardi und Lorenzo Giovannini (2015). „Assessing

- the air quality impact of nitrogen oxides and benzene from road traffic and domestic heating and the associated cancer risk in an urban area of Verona (Italy)". In: *Atmospheric Environment* 120, S. 234–243.
- Sehmel, GA (1973). „Particle resuspension from an asphalt road caused by car and truck traffic". In: *Atmospheric Environment* (1967) 7.3, S. 291–309.
- Setti, Leonardo, Fabrizio Passarini, Gianluigi De Gennaro, Pierluigi Barbieri, Maria Grazia Perrone, Massimo Borelli, Jolanda Palmisani, Alessia Di Gilio, Valentina Torboli und Francesco Fontana (2020). „SARS-Cov-2RNA found on particulate matter of Bergamo in Northern Italy: first evidence". In: *Environmental research* 188, S. 109754.
- Sharma, RB und Ram Bansal (2013). „CFD simulation for flow over passenger car using tail plates for aerodynamic drag reduction". In: *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)* 7.5, S. 28–35.
- She, Wei, Qi Liu, Zhao Tian, Jian-Sen Chen, Bo Wang und Wei Liu (2019). „Blockchain trust model for malicious node detection in wireless sensor networks". In: *IEEE Access* 7, S. 38947–38956.
- Shukla, Komal, Prashant Kumar, Gaurav S Mann und Mukesh Khare (2020). „Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi". In: *Sustainable cities and society* 54, S. 101997.
- Si, Minxing, Ying Xiong, Shan Du und Ke Du (2020). „Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods". In: *Atmospheric Measurement Techniques* 13.4, S. 1693–1707.
- Sliwa, Benjamin, Thomas Liebig, Robert Falkenberg, Johannes Pillmann und Christian Wietfeld (2018). „Efficient machine-type communication using multi-metric context-awareness for cars used as mobile sensors in upcoming 5G networks". In: *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, S. 1–6.
- Snik, Frans, Jeroen HH Rietjens, Arnoud Apituley, Hester Volten, Bas Mijling, Antonio Di Noia, Stephanie Heikamp, Ritse C Heinsbroek, Otto P Hasekamp und J Martijn Smit (2014). „Mapping atmospheric aerosols with a citizen science network of smartphone spectropolarimeters". In: *Geophysical Research Letters* 41.20, S. 7351–7358.
- Spellerberg, Ian F (2005). *Monitoring ecological change*. Cambridge University Press.

- Streibl, N (2017). *Influence of Humidity on the Accuracy of Low-Cost Particulate Matter Sensors*. Techn. Ber. Technical report. DOI: 10.13140/RG. 2.2. 21095.75683.
- Theja, PVV und Lelitha Vanajakshi (2010). „Short term prediction of traffic parameters using support vector machines technique“. In: *2010 3rd International Conference on Emerging Trends in Engineering and Technology*. IEEE, S. 70–75.
- Timmers, Victor RJH und Peter AJ Achten (2016). „Non-exhaust PM emissions from electric vehicles“. In: *Atmospheric Environment* 134, S. 10–17.
- Tingsheng, Shi, Ian K Duncan und Michael T Gastner (2019). „go-cart. io: a web application for generating contiguous cartograms“. In: *Abstr. Int. Cartogr. Assoc* 1, S. 333.
- Tung, Nguyen Thanh, Po-Ching Cheng, Kai-Hsien Chi, Ta-Chi Hsiao, Timothy Jones, Kelly BéruBé, Kin-Fai Ho und Hsiao-Chi Chuang (2021). „Particulate matter and SARS-CoV-2: a possible model of COVID-19 transmission“. In: *Science of The Total Environment* 750, S. 141532.
- Wang, Jackson, ER Case und Dan Manor (1992). „The road traffic microwave sensor (RTMS)“. In: *The 3rd International Conference on Vehicle Navigation and Information Systems*, IEEE, S. 83–90.
- Wang, Tonghua, Wenting Han, Mengfei Zhang, Xiaomin Yao, Liyuan Zhang, Xingshuo Peng, Chaoqun Li und Xvjia Dan (2020a). „Unmanned aerial vehicle-borne sensor system for atmosphere-particulate-matter measurements: Design and experiments“. In: *Sensors* 20.1, S. 57.
- Wang, Yang, Jiayu Li, He Jing, Qiang Zhang, Jingkun Jiang und Pratim Biswas (2015). „Laboratory evaluation and calibration of three low-cost particle sensors for particulate matter measurement“. In: *Aerosol Science and Technology* 49.11, S. 1063–1077.
- Wang, Yanwen, Yanjun Du, Jiaonan Wang und Tiantian Li (2019). „Calibration of a low-cost PM_{2.5} monitor using a random forest model“. In: *Environment international* 133, S. 105161.
- Wang, Zhengye und Yongbin Zhou (2020b). „A HED-optimized Automatic Detection and Tracking Algorithm for Marine Moving Targets based on YOLO V3“. In: *Journal of Physics: Conference Series*. Bd. 1449. 1. IOP Publishing, S. 012126.
- Weisheit, Toni, Carsten Kühnel und Robert Hoyer (2011). „Video-and sound parameter based detection of opposing traffic in Floating Car Observer applications“. In: *Proceedings of the 18th World Congress on Intelligent Transport Systems, Orlando*.

- Wen, Longyin, Dawei Du, Zhaowei Cai, Zhen Lei, Ming-Ching Chang, Honggang Qi, Jongwoo Lim, Ming-Hsuan Yang und Siwei Lyu (2015). „UADETRAC: A new benchmark and protocol for multi-object detection and tracking“. In: *arXiv preprint arXiv:1511.04136*.
- Wender, Stefan und Klaus Dietmayer (2008). „3D vehicle detection using a laser scanner and a video camera“. In: *IET Intelligent Transport Systems* 2.2, S. 105–112.
- Witowski, Katharina, Markus Feucht und Nielen Stander (2011). „An effective curve matching metric for parameter identification using partial mapping“. In: *8th European LS-DYNA, Users Conference Strasbourg, pgs*, S. 1–12.
- Wolf, Franziska, Sebastian Naumann, Chr Engel und René Schönrock (2008a). „Floating Car Observer—Approaches for Traffic Management Strategies by Analysing Oncoming Vehicles“. In: *Proceedings of the 2nd International Workshop on Intelligent Vehicle Control Systems (IVCS 2008), Madeira, Portugal*.
- Wolf, Franziska, Sebastian Naumann, Christoph Engel und Rene Schonrock (2008b). „Low-cost Floating Car Observer Implementation by a Video Camera based Method“. In: *15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting ITS America ERTICO ITS Japan TransCore*.
- Wolf, Franziska, Sebastian Naumann, Christoph Engel und René Schönrock (2008c). „Floating Car Observer—Approaches for Traffic Management Strategies by Analysing Oncoming Vehicles“. In: *Proceedings of the 2nd International Workshop on Intelligent Vehicle Control Systems (IVCS 2008), Madeira, Portugal*.
- Xie, Ming, Lifeng Wu, Bin Li und Zechen Li (2020). „A novel hybrid multivariate nonlinear grey model for forecasting the traffic-related emissions“. In: *Applied Mathematical Modelling* 77, S. 1242–1254.
- Yang, Hongxi, Shu Li, Li Sun, Xinyu Zhang, Jie Hou und Yaogang Wang (2017). „Effects of the ambient fine particulate matter on public awareness of lung cancer risk in China: evidence from the internet-based big data platform“. In: *JMIR public health and surveillance* 3.4, e64.
- Yao, Ye, Jinhua Pan, Weidong Wang, Zhixi Liu, Haidong Kan, Xia Meng und Weibing Wang (2020). „Spatial correlation of particulate matter pollution and death rate of COVID-19“. In: *MedRxiv*.
- Yin, Huabei und Ouri Wolfson (2004). „A weight-based map matching method in moving objects databases“. In: *Proceedings. 16th International Conference on Scientific and Statistical Database Management, 2004*. IEEE, S. 437–438.

- Zhang, Bo (2010). „Computer vision vs. human vision“. In: *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*. IEEE, S. 3–3.
- Zhang, Jian-bo, Guo-hua Song, Lei Yu, Ji-fu Guo und Hong-yu Lu (2018). „Identification and characteristics analysis of bottlenecks on urban expressways based on floating car data“. In: *Journal of Central South University* 25.8, S. 2014–2024.
- Zheng, Ji, Suocheng Dong, Yingjie Hu und Yu Li (2020a). „Comparative analysis of the CO₂ emissions of expressway and arterial road traffic: A case in Beijing“. In: *PloS one* 15.4, e0231536.
- Zheng, Tongshu, Michael H Bergin, Karoline K Johnson, Sachchida N Tripathi, Shilpa Shirodkar, Matthew S Landis, Ronak Sutaria und David E Carlson (2018). „Field evaluation of low-cost particulate matter sensors in high-and low-concentration environments“. In: *Atmospheric Measurement Techniques* 11.8, S. 4823–4846.
- Zheng, Tongshu, Michael H Bergin, Ronak Sutaria, Sachchida N Tripathi, Robert Caldow und David E Carlson (2019). „Gaussian process regression model for dynamically calibrating and surveilling a wireless low-cost particulate matter sensor network in Delhi“. In: *Atmospheric Measurement Techniques* 12.9, S. 5161–5181.
- Zheng, Zhaohui, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye und Dongwei Ren (2020b). „Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression.“ In: *AAAI*, S. 12993–13000.
- Zusman, Marina, Cooper S Schumacher, Amanda J Gasset, Elizabeth W Spalt, Elena Austin, Timothy V Larson, Graeme Carvlin, Edmund Seto, Joel D Kaufman und Lianne Sheppard (2020). „Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study“. In: *Environment international* 134, S. 105329.