

GEO 511

MASTERARBEIT

---

# Vergleich von Methoden zur Detektion von Stopps in Trajektorien

---

*Verfasserin:*

B. Sc. Alena KRÄHENBÜHL

Matrikelnummer: 08-063-604

*Betreuer und*

*Fakultätsvorsteher:*

Prof. Dr. Robert WEIBEL

*Betreuer:*

M. Sc. Christian GSCHWEND

Abteilung Geographische Informationssysteme

Geographisches Institut

UNIVERSITÄT ZÜRICH

Eingereicht am 30. Juni 2014

**Kontakt**

Alena Krähenbühl  
alena.kraehenbuehl@gmail.ch

Prof. Dr. Robert Weibel  
Universität Zürich  
Geographisches Institut  
Abteilung Geographische Informationssysteme  
Winterthurerstrasse 190  
CH-8057 Zürich

M. Sc. Christian Gschwend  
Universität Zürich  
Geographisches Institut  
Abteilung Geographische Informationssysteme  
Winterthurerstrasse 190  
CH-8057 Zürich

*“Es ist nichts beständig als die Unbeständigkeit.”*

Immanuel Kant

# *Zusammenfassung*

Trajektorien Daten können heute infolge technischer Fortschritte von verschiedenen Akteuren, wie beispielsweise Forschenden sowie auch der Allgemeinbevölkerung, erfasst werden. Die Motivation, diese Daten aufzuzeichnen kann dabei unterschiedlicher Natur sein, sowie auch die weitere Analyse, falls eine solche durchgeführt wird. Häufig wird in einem ersten Schritt bei der Analyse von Trajektorien Daten nach den Stopps gesucht, die das sich bewegende Objekt eingelegt hat.

Es existiert eine Vielzahl an Methoden zur Detektion von Stopps in Trajektorien. Diese Ansätze wurden für Untersuchungen aus verschiedenen Gebieten und für unterschiedliche Daten beziehungsweise unterschiedliche sich bewegende Objekte entwickelt. Oftmals passt die gewählte Methode ideal auf die verwendeten Daten und das Setting einer bestimmten Analyse. Eine Aussage zur Robustheit und Güte eines Algorithmus zur Detektion von Stopps für Anwendungen in einem anderen Rahmen wird dabei selten gemacht. In dieser Arbeit werden ausgewählte Bewegungsparameter und Algorithmen aus der Literatur auf verschiedene Datensätze angewendet. Dabei wird untersucht, wie robust diese Methoden bei deren Anwendung auf Datensätze verschiedener, sich bewegender Objekte reagieren. Weiter wird untersucht, ob komplexe Algorithmen Stopps grundsätzlich besser detektieren, als einfache Analysemethoden.

Es konnte gezeigt werden, dass die Algorithmen sensibel reagieren auf unterschiedliche Datensätze. So liefert kein Algorithmus für die Analyse aller sich bewegenden Objekte Resultate gleicher Qualität. Des Weiteren wurde ersichtlich, dass keine der hier untersuchten Methoden in jedem Fall die besten Resultate liefert sowie dass keine dieser Methoden in jedem Fall sehr schlechte Resultate liefert.

Schlussfolgernd kann gesagt werden, dass bei der Auswahl eines Verfahrens zur Detektion von Stopps in Trajektorien verschiedene Kriterien berücksichtigt werden sollten. So spielen die Eigenschaften des Objekts, welches untersucht wird, eine wichtige Rolle, wie auch die räumliche und zeitliche Dimension und Auflösung der Daten. Weiter hängt die Wahl für einen Algorithmus davon ab, wie viel Zeit und Arbeit in dessen Implementation investiert werden soll und wie wichtig die Performance ist.

Zusammenfassend kann gesagt werden, dass die Wahl der idealen Methode zur Detektion von Stopps in Trajektorien von diversen Faktoren abhängt und von Fall zu Fall differiert.

# *Summary*

Trajectory data can nowadays be collected by various actors, such as scientists or private people and companies. The motivation to collect trajectory data and, if wanted to further analysis this data, can be of a number of different reasons. In the majority of the cases where further analysis is investigated, the first step is to find the stops and moves in the trajectory of the moving object.

There exists a diversity of methods to detect the stops in trajectory data. These methods have been established in different research areas and for different purposes as well as treating a variety of datasets of distinct moving objects. In most of these cases, the algorithm chosen for an application fits well the data and the setting of this specific application. But in nearly no paper presenting an algorithm to find the stops in trajectories, a statement about the robustness and the suitability of this algorithm in another setting is made.

This fact leads to the motivation for the master thesis presented here. In this thesis, specifically chosen movement parameters and algorithms from the literature are implemented and run over different kind of datasets. Therefore the robustness of these techniques, when analyzing datasets of different characteristics, is investigated. Further a statement should be made about how big the possible benefits are when using complex instead of simple algorithms.

The results show that the here examined algorithms are sensitive when applied to datasets of different moving objects. There is no method showing similar results, in regard to the goodness of the results, for all of the analyzed datasets. Further it could be shown that none of these algorithms is best for all the different datasets, neither there is an algorithm showing the worst results on every dataset.

Therefore it can be said, that the choice of a method to detect stops in trajectory data of moving objects depends on different criteria. The properties of the moving object as well as the spatial and temporal extend and resolution of the data play a crucial role. Further the amount of investments, such as time for instance, must be considered carefully when choosing a certain technique. Also the importance of the performance of an implementation must be considered.

In summary it can be said, that the choice of the ideal method to detect stops in trajectories depends on a variety of factors and differs from case to case.

# *Danksagung*

Ich wurde im Rahmen meines Studiums und im Besonderen beim Erstellen dieser Masterarbeit von vielen Menschen hervorragend unterstützt, welche ich hier erwähnen möchte.

- An erster Stelle danke ich ganz herzlich meinen Betreuern Robert Weibel und Christian Gschwend für ihre grosse Unterstützung während dieser Masterarbeit, ihre hilfreichen Inputs, ihre Flexibilität und ihre konstruktive Kritik. Die Zusammenarbeit war stets angenehm, entspannt und ermutigend.
- Die vielen weiteren Ideen, welche ich im Rahmen meines Feedbackmeetings von Ross Purves und Kai-Florian Richter erhalten habe, haben mich inspiriert und dazu bewegt, meinen Horizont weit zu halten - vielen Dank.
- Ich danke auch Çöltekin für die Bereitstellung der Eye-Tracking-Daten sowie ihre Hilfe im Zusammenhang mit diesen Daten.
- Insbesondere Adrian Krähenbühl, Leonie Eicher und Melanie Herren gebührt ein grosser Dank für ihre finanzielle Unterstützung während der ganzen Studienzeit, ihr stets offenes Ohr und all die inspirierenden Gespräche.
- Auch Martin Ambühl gilt inniger Dank für die alltägliche Unterstützung, und die Hilfe während dieser Masterarbeit.
- Cyrill Krähenbühl hat mir zu jeder Tages- und Nachtzeit aus programmiertechnischen Notlagen geholfen und mich mit seiner Geduld und Ruhe von seinem Erfahrungsschatz profitieren lassen. Dafür danke ich ihm an dieser Stelle.
- Für das Korrekturlesen dieser Arbeit und die vielen motivierenden Gespräche danke ich Sabine Heinzle herzlich.

# Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>iii</b>
<b>Summary</b>	<b>iv</b>
<b>Danksagung</b>	<b>v</b>
<b>Abbildungen</b>	<b>viii</b>
<b>Tabellen</b>	<b>x</b>
<b>Abkürzungen</b>	<b>xi</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Kontext und Motivation . . . . .	1
1.2 Problemstellung und Zielsetzung . . . . .	2
1.3 Aufbau der Arbeit . . . . .	3
<b>2 Theoretische Grundlagen</b>	<b>4</b>
2.1 Trajektorien in der <i>Geographic Information Science</i> . . . . .	4
2.2 Die Wichtigkeit der Detektion von Stopps in Trajektorien . . . . .	5
2.3 Was ist ein Stopp? . . . . .	6
2.4 Physikalische Betrachtung des Forschungsgegenstandes . . . . .	9
2.5 Verfahren zur Detektion von Stopps in Trajektorien . . . . .	11
2.5.1 Ansätze aus GIScience und <i>Computational Geometry</i> . . . . .	12
2.5.2 Datenbankorientierte Ansätze . . . . .	13
2.5.3 Verhaltensbiologisch orientierte Ansätze . . . . .	15
2.6 Gemeinsamkeiten und Unterschiede der bestehenden Methoden . . . . .	16
2.7 Forschungslücken und Forschungsfragen . . . . .	20
<b>3 Datenlage und Vorverarbeitungsschritte</b>	<b>22</b>
3.1 Verwendete Daten . . . . .	22
3.1.1 Wandernde Person in der Schweiz . . . . .	22
3.1.2 Betonliefernde Lastwagen in Athen . . . . .	23
3.1.3 Ausgewählte Austernfischer auf <i>Schiermonnikoog</i> . . . . .	24
3.1.4 Zugvogel Leo in Amerika . . . . .	25
3.1.5 Eye-Tracking-Daten . . . . .	26
3.2 Datenverarbeitung . . . . .	27

---

<b>4</b>	<b>Methodisches Vorgehen</b>	<b>29</b>
4.1	Rolle der Geschwindigkeit . . . . .	29
4.1.1	Was ist Geschwindigkeit? . . . . .	29
4.1.2	Implementation der Geschwindigkeitsberechnung . . . . .	31
4.2	Bewegungsparameter bei der Detektion von Stopps . . . . .	35
4.3	Verwendete Bewegungsparameter . . . . .	38
4.4	Algorithmengruppe Stops and Moves of Trajectories . . . . .	44
4.5	Hidden Markov Models . . . . .	47
4.5.1	Anwendung bei der Detektion von Stopps in Trajektorien . . . . .	50
<b>5</b>	<b>Resultate</b>	<b>54</b>
5.1	Resultate der Bewegungsparameter . . . . .	55
5.2	Resultate der Algorithmengruppe Stops and Moves of Trajectories . . . . .	64
5.3	Resultate des Hidden Markov Model . . . . .	70
5.4	Resultate zur Analyse der ausgewählten Trajektorien . . . . .	73
<b>6</b>	<b>Diskussion</b>	<b>85</b>
6.1	Bewegungsparameter . . . . .	87
6.1.1	Geschwindigkeit . . . . .	87
6.1.2	Drehwinkel . . . . .	90
6.1.3	Dauer . . . . .	91
6.1.4	Ausdehnung . . . . .	94
6.2	Algorithmengruppe Stops and Moves of Trajectories . . . . .	96
6.2.1	Stops and Moves of Trajectories . . . . .	96
6.2.2	Clustering-Based Stops and Moves of Trajectories . . . . .	97
6.2.3	Direction-Based Stops and Moves of Trajectories . . . . .	100
6.3	Hidden Markov Model . . . . .	103
6.4	Beantwortung der Forschungsfragen . . . . .	104
<b>7</b>	<b>Schlussfolgerungen und Ausblick</b>	<b>106</b>
7.1	Fazit . . . . .	106
7.2	Ausblick . . . . .	108
	<b>Literaturverzeichnis</b>	<b>110</b>
	<b>A Übersichtstabellen der Resultate</b>	<b>115</b>
	<b>B Eigenschaften der verwendeten Datensätze</b>	<b>118</b>
	<b>C Genauigkeit und Trefferquote</b>	<b>120</b>
	<b>Persönliche Erklärung</b>	<b>121</b>

# Abbildungen

2.1	Mögliche Verläufe einer Trajektorie in einem begrenzten Raum . . . . .	11
2.2	Semantischer Anreicherungsprozess einer Trajektorie . . . . .	14
2.3	Segmentierung der Trajektorien zweier Blässgänse . . . . .	16
2.4	Taxonomie der Algorithmen zur Identifikation von Fixationen und Beispiele ausgewählter Algorithmen . . . . .	17
4.1	Implementation des räumlichen Schwellenwertes für die Berechnung der Geschwindigkeit . . . . .	34
4.2	Auswahl möglicher Ausdrucksformen einer Trajektorie . . . . .	41
4.3	Drehwinkel eines Punktes einer Trajektorie . . . . .	42
4.4	Drehwinkel in verschiedenen zeitlichen Auflösungen . . . . .	43
4.5	Beispiel einer Trajektorie und drei dazugehörigen Kandidatenregionen . .	45
4.6	Beispiel einer Trajektorie mit zwei echten und zwei unbekanntem Stopps .	46
4.7	Beispiel eines Hidden Markov Models . . . . .	50
4.8	Ein Hidden Markov Model bei der Detektion von Stopps in Trajektorien .	51
5.1	Geschwindigkeitsschwellenwert angewendet auf Wanderdaten . . . . .	56
5.2	Geschwindigkeitsschwellenwert angewendet auf Lastwagendaten . . . . .	56
5.3	Geschwindigkeitsschwellenwert angewendet auf Austernfischerdaten . . . .	57
5.4	Geschwindigkeitsschwellenwert angewendet auf Truthahngeierdaten . . . .	57
5.5	Geschwindigkeitsschwellenwert angewendet auf Eye-Tracking-Daten . . . .	57
5.6	Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Wanderdaten . . . . .	58
5.7	Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Lastwagendaten . . . . .	58
5.8	Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Austernfischerdaten . . . . .	59
5.9	Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Truthahngeierdaten . . . . .	59
5.10	Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Eye-Tracking-Daten . . . . .	59
5.11	Geschwindigkeits- und Mindestdauerschwellenwert angewendet auf Wanderdaten . . . . .	60
5.12	Geschwindigkeits- und Mindestdauerschwellenwert angewendet auf Lastwagendaten . . . . .	61
5.13	Geschwindigkeits- und Mindestdauerschwellenwert angewendet auf Austernfischerdaten . . . . .	61
5.14	Geschwindigkeits- und Mindestdauerschwellenwert angewendet auf Truthahngeierdaten . . . . .	61

---

5.15	Geschwindigkeits- und Mindestdauerschwellenwert angewendet auf Eye-Tracking-Daten . . . . .	62
5.16	Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Wanderdaten . . . . .	63
5.17	Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Lastwagendaten . . . . .	63
5.18	Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Austernfischerdaten . . . . .	63
5.19	Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Truthahngeierdaten . . . . .	64
5.20	Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Eye-Tracking-Daten . . . . .	64
5.21	CB-SMoT angewendet auf Wanderdaten . . . . .	66
5.22	CB-SMoT angewendet auf Lastwagendaten ohne Kandidatenregionen . .	66
5.23	CB-SMoT angewendet auf Austernfischerdaten . . . . .	66
5.24	CB-SMoT angewendet auf Truthahngeierdaten . . . . .	67
5.25	CB-SMoT angewendet auf Eye-Tracking-Daten . . . . .	67
5.26	DB-SMoT angewendet auf Wanderdaten . . . . .	69
5.27	DB-SMoT angewendet auf Lastwagendaten . . . . .	69
5.28	DB-SMoT angewendet auf Austernfischerdaten . . . . .	69
5.29	DB-SMoT angewendet auf Truthahngeierdaten . . . . .	70
5.30	DB-SMoT angewendet auf Eye-Tracking-Daten . . . . .	70
5.31	HMM angewendet auf Wanderdaten . . . . .	71
5.32	HMM angewendet auf Lastwagendaten . . . . .	71
5.33	HMM angewendet auf Austernfischerdaten . . . . .	72
5.34	HMM angewendet auf Truthahngeierdaten . . . . .	72
5.35	HMM angewendet auf Eye-Tracking-Daten . . . . .	72
5.36	Übersicht über die Trajektorie eines Wanderers und die dabei gemachten Stopps laut verschiedenen Bewegungsparametern . . . . .	76
5.37	Übersicht über die Trajektorie eines Wanderers und die dabei gemachten Stopps laut den Algorithmen aus der Literatur . . . . .	77
5.38	Übersicht über die Trajektorie eines Lastwagens und die dabei gemachten Stopps laut verschiedenen Bewegungsparametern . . . . .	79
5.39	Übersicht über die Trajektorie eines Lastwagens und die dabei gemachten Stopps laut den Algorithmen aus der Literatur . . . . .	81
5.40	Übersicht über die Trajektorie eines Austernfischers und die dabei gemachten Stopps laut verschiedenen Bewegungsparametern . . . . .	83
5.41	Übersicht über die Trajektorie eines Austernfischers und die dabei gemachten Stopps laut den Algorithmen aus der Literatur . . . . .	84
6.1	Verschiedene Aggregationsstufen einzelner Datenpunkten . . . . .	86
A.1	Übersicht über die Resultate der Bewegungsparameter . . . . .	116
A.2	Übersicht über die Resultate der Algorithmen aus der Literatur und des HMM . . . . .	117

# Tabellen

2.1	Beispiel für die Unterscheidung verschiedener Bewegungszustände von Menschen anhand dreier Bewegungsparameter . . . . .	19
4.1	Zeitliche Schwellenwerte zur Berechnung der Geschwindigkeit . . . . .	33
4.2	Räumliche Schwellenwerte zur Berechnung der Geschwindigkeit . . . . .	34
4.3	Verwendung der einzelnen Bewegungsparameter in der Literatur . . . . .	36
4.4	Übersicht über verschiedene Methoden zur Detektion von Fixationen . . . . .	48
5.1	Übersicht über die Werte der Bewegungsparameter in der Analyse der ausgewählten Trajektorien . . . . .	74
B.1	Übersicht über die hier verwendeten Datensätze . . . . .	119
C.1	Genauigkeit und Trefferquote der Resultate der ausgewählten Trajektorien	120

# Abkürzungen

<b>CB-SMoT</b>	<b>Clustering-Based Stops and Moves of Trajectories</b>
<b>DB-SMoT</b>	<b>Direction-Based Stops and Moves of Trajectories</b>
<b>DBSCAN</b>	<b>Density Based Spatial Clustering of Applications with Noise</b>
<b>GIS</b>	<b>Geographic Information System</b>
<b>GIScience</b>	<b>Geographic Information Science</b>
<b>GMT</b>	<b>Greenwich Mean Time</b>
<b>GPS</b>	<b>Global Positioning System</b>
<b>HMM</b>	<b>Hidden Markov Model</b>
<b>OSM</b>	<b>OpenStreetMap</b>
<b>SMoT</b>	<b>Stops and Moves of Trajectories</b>
<b>T-OPTICS</b>	<b>Temporal Ordering Points To Identify the Clustering Structure</b>
<b>UvA-BiTS</b>	<b>University of Amsterdam Bird Tracking System</b>
<b>WGS84</b>	<b>World Geodetic System 1984</b>

# Kapitel 1

## Einleitung

### 1.1 Kontext und Motivation

In dieser Arbeit geht es um die Analyse von Trajektorien­daten, wie beispielsweise GPS-Daten oder allgemeiner, Daten von sich bewegend­en Objekten. Laut Rinzivillo et al. (2008) ist eine Trajektorie die Abfolge von zeitlich annotierten Positionen, die ein sich bewegend­es Objekt frequentiert hat. Trajektorien­daten sind heutzutage in vielfältiger Weise vorhanden, da es um einiges einfacher geworden ist, Daten dieser Art zu generieren. Dies gilt sowohl für die Allgemeinbevölkerung als auch für die Forschung. Dieser Umstand ist zurückzuführen auf technische Fortschritte, beispielsweise im Bereich der GPS-Empfänger, der Navigationssysteme und der Mobiltelefone (Laube et al., 2005). So ist es möglich, knapp 14 g schwere, solarbetriebene Geräte herzustellen, die nebst einem GPS-Empfänger auch einen Beschleunigungsmesser beinhalten und die Höhe über Meer messen können. Mit Hilfe solcher Geräte können die Position und die Bewegungen kleiner, leichter Tiere wie beispielsweise Vögel sehr genau untersucht werden (Shamoun-Baranes et al., 2012). Es kann mit diesen Technologien eine Vielzahl an Trajektorien aus verschiedenen Gebieten, wie der Ökologie, der Freizeit und der Transportsysteme generiert werden. Eye-Tracking-Daten, welche sich in einem sehr feinen räumlichen und zeitlichen Rahmen bewegen, sind ebenfalls Trajektorien­daten (Salvucci und Goldberg, 2000).

Bei der Analyse von Trajektorien­daten kann ein erster Schritt sein, nach den Stopps zu suchen, die das sich bewegend­e Objekt eingelegt hat. Ein Stopp kann verschiedene Ausprägungen und Charakteristika aufweisen, je nach Anwendungsgebiet und untersuchtem

Objekt. Da es verschiedene Bereiche gibt, die sich für unterschiedliche Zwecke Trajektorien- und Daten bedienen, wird ein Stopp nicht in jedem Fall gleich definiert. Untersuchungen von Zugvögeln beispielsweise können die Rastplätze auf der Route dieser Tiere als Stopp definieren. Bei der Planung einer Veränderung des Strassennetzes einer Stadt, können Orte, an denen Autos häufig stoppen von Interesse sein, da sie Knotenpunkte darstellen. Wie die Menge an Daten zu unterschiedlichen Trajektorien hat auch die Menge an Methoden zur Analyse ebendieser Daten und im Speziellen zur Detektion von Stopps in Trajektorien in jüngerer Zeit zugenommen. Es bestehen Herangehensweisen, die nach Stopps in Trajektorien suchen, jedoch von unterschiedlicher Komplexität sind und sich verschiedener Ausgangslagen bedienen. Die Vielfalt an verfügbaren Verfahren stellt einerseits einen Mehrwert dar, da im Rahmen einer Analyse auf verschiedene Ansätze zurückgegriffen werden kann. Andererseits kann eine unüberschaubar grosse Menge an Methoden auch verwirrend sein, da bei einer Entscheidung nicht alle Alternativen restlos abgeklärt werden können und so das ideale Verfahren möglicherweise übersehen wird. Ein Vergleich der verfügbaren Algorithmen zur Detektion von Stopps in Trajektorien steht bis anhin noch aus und soll hier gemacht werden.

## 1.2 Problemstellung und Zielsetzung

Die meisten Ansätze zur Detektion von Stopps werden anhand eines einzigen Datensatzes erläutert und ausgeführt, wobei offen bleibt, wie robust die gewählten Verfahren gegenüber anderen Datensätzen und Datenstrukturen sind. Eine Methode aus der Literatur eignet sich oft ideal für die Charakteristiken der zu untersuchenden Daten und kann deren Eigenheiten in Betracht ziehen. Es bleibt im Rahmen dieser Analysen jedoch meistens offen, wie die verwendete Methode in einem anderen Setting reagieren würde. So stellt sich beispielsweise die Frage, wie ein Algorithmus zur Detektion von Rastplätzen von Zugvögeln reagiert, wenn er auf die Trajektorie einer Wandererin oder eines Autos angewendet wird. Somit macht die Diversität dieser verschiedenen Techniken zur Erkennung von Stopps in Trajektorien den Versuch eines Vergleichs interessant und notwendig.

Im Rahmen dieser Arbeit werden Methoden verglichen, die Stopps aus Trajektorien extrahieren. Dabei handelt es sich einerseits um Algorithmen aus der Literatur und andererseits um Bewegungsparameter, die in verschiedenen bestehenden Untersuchungen

verwendet werden. Hier werden somit nicht Techniken, welche darauf ausgerichtet sind, Muster in grossen Datenmengen zu finden (wie beispielsweise in Orellana und Wachowicz (2011)) oder mehrere Bewegungszustände von Objekten zu unterscheiden (wie beispielsweise in Waga et al. (2012)), untersucht. Der Methodenvergleich wird gemacht indem die einzelnen Verfahren auf Datensätze diverser sich bewegender Objekte angewendet werden. Bei diesen Objekten handelt es sich im Rahmen dieser Arbeit um eine Wandern- de Person, einen Lastwagen, einen Austernfischer, einen Truthahngerier und das Auge (Eye-Tracking-Daten). Ziel dabei ist es, herauszufinden, wie robust die bestehenden Al- gorithmen bei der Anwendung auf diese Datensätze unterschiedlicher, sich bewegender Objekte sind. Weiter soll eruiert werden, ob die Güte eines Resultats von der Kom- plexität des Algorithmus abhängt und falls ja, inwiefern dies so ist. Die spezifischen Forschungsfragen werden im Kapitel 2.7 erläutert.

### 1.3 Aufbau der Arbeit

In Kapitel 2 wird auf die Literatur eingegangen, die im Zusammenhang mit Stopps in Bewegungsdaten steht. Dabei wird der Hintergrund und das Forschungsfeld rund um die Detektion von Stopps in Trajektorien thematisiert. Daraus werden Forschungslücken herausgefiltert, um daraus die Forschungsfragen abzuleiten, welche im Abschnitt 2.7 auf- geführt sind. Kapitel 3 stellt die verwendeten Daten und deren spezifische Charak- teristika vor, wobei auch wichtige Vorverarbeitungsschritte der Daten thematisiert werden. In Kapitel 4 wird das methodische Vorgehen erläutert. Dabei werden die verwendeten Algorithmen und Bewegungsparameter sowie deren Funktionsweise vorgestellt. Die Re- sultate sind in Kapitel 5 enthalten. Kapitel 6 diskutiert diese Resultate und versucht die untersuchten Bewegungsparameter und Algorithmen zu charakterisieren sowie deren Vor- und Nachteile zu thematisieren. Die aus den Resultaten gezogenen Schlussfolge- rungen sind in Kapitel 7 dargestellt, wobei hier zudem ein Ausblick auf weiterführende Ideen gegeben wird.

Auf der CD, welche mit dieser Arbeit mitgeliefert wird, befinden sich die kommentierten Quellcodes der Methoden, die implementiert wurden. Die Implementationen sowohl der Bewegungsparameter als auch der Algorithmen aus der Literatur, die ich nach bestem Wissen und Gewissen verfasst habe, sind auf dieser CD enthalten.

## Kapitel 2

# Theoretische Grundlagen

### 2.1 Trajektorien­daten in der *Geographic Information Science*

Die *Geographic Information Science* (GIScience) ist in den letzten Jahren stark durch technische Fortschritte geprägt worden. Dies zeigt sich einerseits in der Datenverarbeitung, andererseits in der Erfassung von Daten. Dadurch sind Trajektorien­daten, seien dies GPS-Daten oder auch Eye-Tracking-Daten oder sonstige Aufzeichnungen von sich bewegenden Objekten, in den letzten Jahren unter anderem im Bereich GIScience zunehmend präsen­ter geworden (Brilhante et al., 2011).

Mit der Entwicklung des GPS in den 1990er Jahren wurde es nicht nur für Reisende einfacher, ihre Routen aufzuzeichnen, auch die Arbeit von Forschenden konnte durch nun genauere, objektiver und zuverlässiger generierte Informationen verbessert werden (Schuessler und Axhausen, 2009). Gudmundsson et al. (2008) sprechen von der Entwicklung von einem notorischen Datendefizit zu einem Übermass an Daten, wobei adäquate Methoden zu deren Analyse noch ausstehend sind. In der grossen Menge an vorhandenen Trajektorien­daten sehen einige Mitglieder der GIScience die Notwendigkeit effiziente Analyseverfahren zu erarbeiten (beispielsweise Palma et al. (2008)). Die Möglichkeit sehr vieler Menschen, fast überall GPS-Daten produzieren zu können, macht es unabdingbar, automatisierte und schnelle Methoden zu entwickeln, welche die vorhandenen Daten prozessieren können (Schuessler und Axhausen, 2009). All diese Entwicklungen in der

GIScience führten und führen schliesslich dazu, dass ältere Forschungsfragen neu aufgegriffen werden können, und, dass neue Fragestellungen auf Grund neuer Möglichkeiten erarbeitet werden können (Gudmundsson et al., 2008).

## 2.2 Die Wichtigkeit der Detektion von Stopps in Trajektorien

Bevor eingehender auf die Literatur eingegangen wird, soll hier zur Einführung des Themas erläutert werden, weshalb nach Stopps in Trajektorien gesucht wird. Dazu gibt es aus verschiedensten Forschungsgebieten unterschiedliche Motivationen, wobei hier einige Anwendungsgebiete vorgestellt werden. Diese Auswahl thematisiert einerseits Anwendungsgebiete, die in dieser Arbeit eingehender untersucht werden sowie Anwendungsgebiete, auf die später nicht weiter eingegangen wird.

**Verhaltensbiologie von migrierenden Vögeln** Wie in Abschnitt 2.5.3 ausführlicher besprochen wird, bedient sich auch die Ökologie und die Verhaltensbiologie Trajektorien zur Analyse des Bewegungsverhaltens von Tieren. Buchin et al. (2010) haben ein Konzept entwickelt, um Trajektorien nach raum-zeitlichen Kriterien zu segmentieren. Dieses Konzept haben Buchin et al. (2013) am Beispiel der Migration der Blässgans, *anser albifrons*, evaluiert. Dabei haben sie zwischen Flug und Zwischenstopp unterschieden, wobei die Winteraufenthalte, die Aufzucht und die Mauser ebenfalls zu den Zwischenstopps zählen. Ein Flug ist in ihrer Arbeit dadurch charakterisiert, dass die Richtung, in die ein Vogel fliegt, sich in einem Bereich von 120 Grad bewegt und somit nicht stark variiert. Ein Zwischenstopp wird registriert, wenn sich ein Vogel während mindestens 48 Stunden in einem Radius von 30 Kilometern befindet. In dem Sinne können Trajektorien von Vögeln genutzt werden, um deren Rastorte zu finden.

**Identifikation von Fixationen in Eye-Tracking-Daten** Bei der Analyse von Eye-Tracking-Daten wird nicht von Stopps, sondern von Fixationen gesprochen. Zusätzlich kann zwischen Sakkaden (Sakkaden sind schnelle Bewegungen des Auges<sup>1</sup>), *smooth pursuit*, Blinzeln und Rauschen unterschieden werden (Nyström und Holmqvist, 2010). Es

<sup>1</sup>Quelle (Stand 06.02.2014): <http://en.wikipedia.org/wiki/Saccade>

kann jedoch auch lediglich zwischen Fixationen und Sakkaden unterschieden werden, wie beispielsweise bei Salvucci und Goldberg (2000). Ziel solcher Analysen kann sein, herauszufinden, wie verschiedene Kartenprodukte funktionieren. Dabei wird untersucht, wo ein Leser oder eine Leserin einer bestimmten Anwendung beziehungsweise einer Karte vermehrt hinschaut, also wo Fixationen stattfinden, und welche Bildelemente die Aufmerksamkeit anziehen (Coltekin et al., 2009).

**Untersuchung und Verbesserung des Verkehrsmanagements** Im Automobilverkehr und beim Verkehr in Luftraum kann die Analyse von Trajektoriendaten hilfreich sein, um schwierige oder gefährliche Verkehrssituationen zu erkennen und entsprechend darauf zu reagieren (Gudmundsson et al., 2008). Auch Spaccapietra et al. (2008) erwähnen die Wichtigkeit von Trajektoriendaten, unter anderem, weil diese beispielsweise als Grundlage für Entscheidungsprozesse im Verkehrsmanagement dienen können. Im Rahmen der Verkehrsplanung geht es jedoch nicht nur um motorisierte Fortbewegungsmittel, so kann die Modellierung von Bewegungen der Fußgänger in dieser Hinsicht auch interessant sein (Hoogendoorn und Bovy, 2005).

**Überwachung und Sicherheit** In diesem Bereich der Anwendungen werden Trajektoriendaten unter anderem dazu verwendet, abnormales beziehungsweise unerwartetes Verhalten von Objekten, wie Personen, Kreditkarten oder Mobiltelefonsignalen, zu eruieren (Gudmundsson et al., 2008). Die Detektion von Stopps in Reiserouten könnte hier unter anderem für reisende Personen interessant sein. So könnte beispielsweise ein unerwartet langer Stopp, im Vergleich zur Dauer des Stopps im Reiseplan, gemeldet werden und vermisste Personen könnten schneller aufgefunden werden.

### 2.3 Was ist ein Stopp?

Anhand der Ausführungen in Abschnitt 2.2 wird ersichtlich, dass Stopps nicht einheitlich definiert werden. Im Folgenden soll untersucht werden, wie ein Stopp zu definieren sein kann. Collins et al. (1995) haben Videos von Hausfliegen (*musca domestica*) analysiert, um herauszufinden, wo die Fliegen einen Halt einlegen. Dabei haben sie einen Stopp Situation definiert, in der es einem Tier nicht möglich ist, eine bestimmte Distanz in einer gewissen Zeit zurück zu legen. Wäre diese einfache Definition eines Stopps, also

eine bestimmte Geschwindigkeit, absolut und allgemeingültig, wäre diese Arbeit nicht nötig, da sich jede Methode dieser Definition bedienen könnte. Dies ist jedoch nicht der Fall, weshalb verschiedene Untersuchungen auch unterschiedliche Definitionen eines Stopps anwenden. Das Stoppen eines Objektes kann somit durch eine gewisse Geschwindigkeit (Collins et al., 1995) über die Ansammlung von Trajektorienpunkten im Raum (Palma et al., 2008) bis zum Aufenthalt des sich bewegenden Objekts innerhalb eines bestimmten Gebietes für eine bestimmte Zeit (Alvares et al., 2007) definiert sein. Unterschiedliche Definitionen bestehen, weil Stopps von unterschiedlichen Objekten und für unterschiedliche Anwendungen untersucht werden (Bak et al., 2012). Weiter spielt die räumliche sowie die zeitliche Skala der Daten und der darauf basierenden Untersuchung eine wichtige Rolle. Dies zeigen Laube und Purves (2011) in einer Untersuchung von grasenden Kühen, wobei dabei primär die zeitliche Auflösung intensiver betrachtet wird.

Eine formelle Definition eines Stopps in einer Trajektorie liefern Spaccapietra et al. (2008). In ihrer Arbeit wird ein Teil einer Trajektorie als Stopp definiert, wenn folgende drei Kriterien erfüllt sind:

- der Benutzer hat den Bereich als Stopp definiert
- die zeitliche Ausdehnung ist nicht null
- das Objekt bewegt sich nicht beziehungsweise bleibt am gleichen Ort

Da eine Trajektorie nach dieser Definition vollständig durch Stopps und Bewegungen abgedeckt ist, wird ein Bewegung folgendermassen definiert. Eine Bewegung (*move*) ist ein Teil der Trajektorie, welcher durch zwei Stopps oder einen Stopp und den Start- oder Endpunkt der Trajektorie begrenzt ist, wobei die zeitliche Ausdehnung ebenfalls nicht null sein darf und die räumliche Ausdehnung eine Linie ist (Spaccapietra et al., 2008). Die obige Definition eines Stopps ist ein hilfreicher Ansatz, wenn lediglich Trajektorien-daten ohne weitere Informationen vorliegen. Ist dem Forschenden jedoch bekannt, unter welchen Bedingungen ein Objekt mit grösserer Wahrscheinlichkeit stoppt, so kann diese Kontextinformation miteinbezogen werden. Dies geschieht häufig unter Einbezug von sogenannten Kandidatenregionen (Alvares et al., 2007). In diesem Falle wird das Segment einer Trajektorie als Stopp definiert, wenn es sowohl bestimmte Kriterien erfüllt als auch innerhalb einer Kandidatenregion liegt. Es werden jedoch häufig nicht nur

räumliche, sondern auch zeitliche Grenzen für das Vorhandensein eines Stopps gesetzt. Hierbei muss eine Trajektoriensequenz über eine bestimmte Zeitdauer bestehen, um als gewissen Stopp anerkannt zu werden (Moreno et al., 2010). Die zeitliche Grenze kann jedoch auch die Dauer innerhalb einer Kandidatenregion (Alvares et al., 2007) sein.

Ein Stopp in einer Trajektorie ist somit nicht universal und eindeutig abgrenzbar. Ein wichtiges Kriterium bei dieser Diskussion ist die Anwendung wofür eine Trajektorie untersucht wird (Spaccapietra et al., 2008). Trajektoriensegmente, welche für eine bestimmte Anwendung als Stopp definiert werden können, sind in einer anderen Anwendung nicht von Bedeutung (Orellana und Wachowicz, 2011). Bei einem Datensatz zu Vogelzügen beispielsweise könnten den Forschenden nur längere Stopps interessieren, da diese für das Detektieren von Rastplätzen wichtig sind. Was hierbei nicht als Stopp gesehen wird, kann für einen Datensatz mit kleineren, nicht migrierenden Vögeln hingegen als wichtigen Stopp klassifiziert werden, da hier nur solche Stopps vorkommen. Somit muss nicht jeder Stopp in einem Datensatz für die Anwendung interessant und relevant sein.

Anhand der obigen Ausführungen habe ich Stopps in drei Kategorien eingeteilt. Diese Einteilung ist nicht universell und deckt nicht alle Arten von Stopps ab. Folgende drei Kategorien können unterschieden werden:

- Durch Parameter definierter Stopp
- Durch die räumliche Dichte definierter Stopp
- Durch Kontextinformationen definierter Stopp

Stopps der ersten Kategorie sind durch verschiedene Parameter der Trajektorie definiert, wie die Richtungsänderungen der Punkte oder der Aufenthalt der Punkte innerhalb eines Raumausschnitts (Buchin et al., 2013). In der zweiten Kategorie der Einteilung sind Stopps enthalten, die sich durch ihre räumliche Ausprägung definieren, das heisst ein Stopp äussert sich durch eine räumliche Ansammlung von Punkten (Palma et al., 2008). Die dritte Kategorie beinhaltet Stopps, die unter Einbezug von Kontextinformationen, dies können beispielsweise Kandidatenregionen sein, definiert werden (Alvares et al., 2007).

## 2.4 Physikalische Betrachtung des Forschungsgegenstandes

Bewegungsdaten können grundsätzlich reduziert werden auf eine räumliche Komponente, den Standort des sich bewegenden Objekts, und auf eine zeitliche Komponente, den Zeitpunkt zu dem sich das Objekt an dem jeweiligen Ort befindet. Somit verfügen Bewegungsdaten über einen räumlichen, einen zeitlichen und einen raum-zeitlichen Aspekt. Zu deren Analyse können diese drei Komponenten verwendet werden. Von einem rein physikalischen Gesichtspunkt aus gesehen gibt es jedoch nur diese drei Komponenten, in unterschiedlichen Formen und Auslegungen. Werden nun Methoden verglichen, die sich derselben dieser drei Komponenten bedienen, so sind auch ähnliche Resultate zu erwarten.

**Zeitliche Komponente** Wird alleinig die zeitliche Komponente berücksichtigt, so ist das nicht sehr aufschlussreich, da das zeitliche Aufnahmeintervall in der Regel unabhängig von der räumlichen oder thematischen Gegebenheit des Forschungsgegenstandes ist. Es verhält sich normalerweise nicht so, dass das Aufnahmeintervall kleiner wird, wenn das Objekt beispielsweise sehr schnell den Standort wechselt. Es kann sogar sein, dass das Aufnahmeintervall regelmässig ist, weshalb daraus keine weiteren Schlüsse gezogen werden können. Aus diesem Grund wird eine Analyse rein auf der zeitlichen Komponente hier nicht in Betracht gezogen.

**Räumliche Komponente** Für eine Analyse von Stopps und Bewegungen kann man sich auf die räumliche Komponente stützen. In diesem Sinne könnten Stopps durch eine Clusteranalyse gesucht werden, ähnlich einem DBSCAN (Ester et al., 1996), wobei eine rein räumliche Häufung von Datenpunkten ein Hinweis für das Vorkommen eines Stopps sein kann. Nun stellt sich die Frage, ob bei einem solchen Vorgehen der Untersuchung von Trajektorien alle Datenpunkte der Trajektorie oder nur die vorherigen und nachfolgenden miteinbezogen werden sollen. Im zweiten Falle wäre die Analyse nicht mehr rein räumlich, da die Reihenfolge der Punkte, also die zeitliche Komponente miteinbezogen würde. Im ersten Fall stellt sich jedoch die Frage, ob ein solches Vorgehen sinnvoll ist. Es ist bei Daten zu Autos zu erwarten, dass häufig frequentierte Strassen so als Cluster definiert werden. Diese Cluster würden sich dann jedoch nicht daraus ergeben, dass sich

ein Auto lange dort aufgehalten hat, sondern dass es sich immer wieder dort aufgehalten hat. Ob dieses Cluster dann als Stopp bezeichnet werden kann, kann von Fall zu Fall unterschiedlich sein. Auch bei Zugvögeln könnte diese Problematik auf der frequentierten Route auftreten, in diesem Fall sind die zeitlichen Unterschiede sogar noch grösser (gleicher Standort im Frühling und Herbst desselben Jahres).

**Raum-zeitliche Komponente** Eine alleinige Analyse räumlicher beziehungsweise zeitlicher Aspekte scheint also bei der Arbeit mit Trajektorien nicht in allen Fällen aufschlussreich zu sein. Die raum-zeitliche Komponente von Bewegungsdaten scheint besser geeignet zu sein für die Detektion von Stopps, da sich Stopps auf einem relativ engen Raum befinden und zeitlich begrenzt sind. Die Auslegung der raum-zeitlichen Aspekte kann auf viele verschiedene Arten gemacht werden. So kann eine Geschwindigkeit verwendet werden, oder die Mindestaufenthaltsdauer in einer bestimmten Region, wobei auch hier die Geschwindigkeit implizit beinhaltet ist.

Um diesen Sachverhalt zu verdeutlichen habe ich Abbildung 2.1 erstellt. Betrachtet man eine Trajektorie in einem definierten räumlichen Ausschnitt, so können grundsätzlich zwei Situationen unterschieden werden. Für dieses Beispiel wird davon ausgegangen, dass das zeitliche Aufnahmeintervall regelmässig ist. Entweder die Trajektorie verläuft relativ gerade durch den Ausschnitt und verbringt somit nur eine kurze Zeit innerhalb eines bestimmten Ausschnitts (Abbildung 2.1 a)). Oder aber die Trajektorie bleibt für eine länger andauernde und grösser ausfallende Bewegung in dem gleichen Ausschnitt (Abbildung 2.1 b)). Ist die räumliche Auflösung nun gröber als der räumliche Ausschnitt in Abbildung 2.1, so besteht die Gefahr, dem sich bewegenden Objekt in Abbildung 2.1 b) eine kleinere Geschwindigkeit zuzuweisen als dem sich bewegenden Objekt in Abbildung 2.1 a). Dies muss jedoch nicht unbedingt der Realität entsprechen. So kann sein, dass sich das sich bewegende Objekt in Abbildung 2.1 b) während der Zeit in diesem selben räumlichen Abschnitt sehr viel bewegt hat und somit auch viel Distanz zurückgelegt hat. Dann wäre die Geschwindigkeit sogar höher als die des Objekts in 2.1 a).

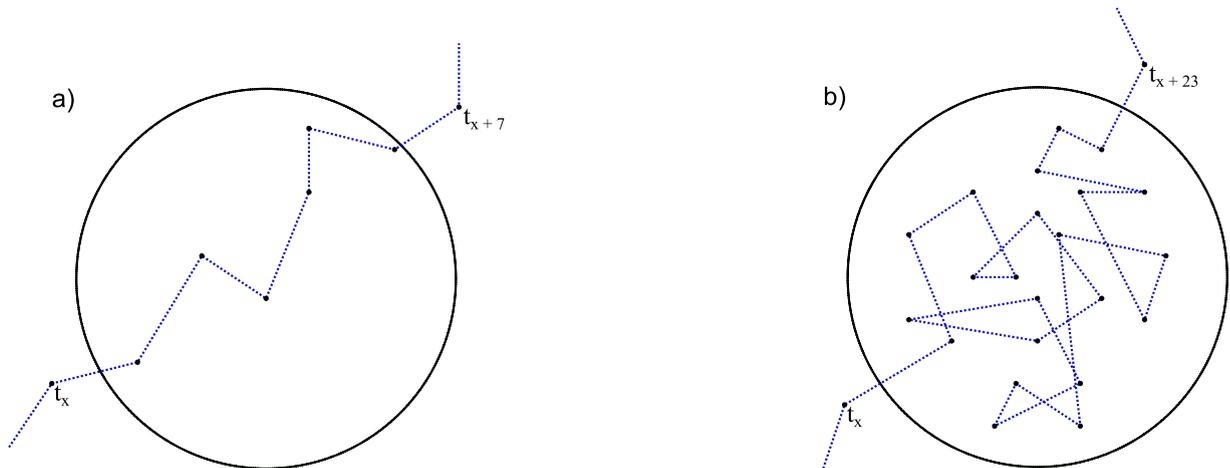


ABBILDUNG 2.1: Mögliche Verläufe einer Trajektorie in einem begrenzten Raum. Abbildung a) zeigt den Verlauf im Falle einer Bewegung, während Abbildung b) einen möglichen Stopp verdeutlicht.

Physikalisch gesehen enthalten Methoden, welche sich der Geschwindigkeit in Form bedienen, den gleichen Informationsgehalt. Die Physik geht von einer kontinuierlichen Bewegung aus, welche zeitlich in Echtzeit aufgelöst ist. Trajektorien- und Bewegungsdaten hingegen stellen immer eine Zusammenfassung beziehungsweise eine Diskretisierung dieser Bewegungen dar. Nichtsdestotrotz sollte bei der Arbeit mit Bewegungsdaten stets beachtet werden, was miteinander verglichen wird und welchen Methoden grundsätzlich dieselbe Physik zugrunde liegt.

## 2.5 Verfahren zur Detektion von Stopps in Trajektorien- daten

Im Folgenden wird eine Einteilung von Verfahren für die Detektion von Stopps in Trajektorien gemacht. Diese Einteilung richtet sich nach den Forschungsgebieten, die die entsprechenden, hier behandelten Methoden hervorgebracht haben. Die Einteilung kann auch auf eine andere Weise gemacht werden, wobei andere Akzente gesetzt werden können. Hinzu kommt, dass mit dem Einbezug weiterer Forschungsgebiete, mit technischen Fortschritten und neuen Anwendungsideen diese Einteilung stetig angepasst und verfeinert werden kann.

### 2.5.1 Ansätze aus GIScience und *Computational Geometry*

Eine Arbeit im Bereich GIScience haben Bak et al. (2012) gemacht. In Anlehnung an Situationen aus dem Autoverkehr definieren sie einen Stopp als Teil der Trajektorie während dem die Geschwindigkeit kleiner als ein gewisser Schwellenwert ist. Der Schwellenwert wird je nach Anwendung gesetzt und kann daher variieren. Auch Orellana und Wachowicz (2011) gehen grundsätzlich von einer tieferen Geschwindigkeit beim Auftreten eines Stopps aus. Sie arbeiten jedoch nicht direkt auf einer Trajektorie, sondern auf sogenannten *movement vectors*, welche aus der Trajektorie abgeleitet sind. Diese Vektoren enthalten zusätzliche Informationen, wie beispielsweise die räumliche und die zeitliche Distanz zum letzten Punkt, Geschwindigkeit und Bewegungsrichtung.

Moreno et al. (2010) gehen einen Schritt weiter als die bloße Detektion von Stopps und schauen in das Innere eines Stopps. In Anlehnung an die Methoden SMOt (Alvares et al., 2007) und CB-SMOt (Palma et al., 2008), welche in Kapitel 4 besprochen werden, verwenden die Autoren und Autorinnen drei Bewegungsparameter, um verschiedene Bewegungszustände zu unterscheiden. Diese Bewegungsparameter sind die Dauer eines Stopps, die maximale Geschwindigkeit und die Richtungsänderung während des Stopps. Aus einem Datensatz mit Fussgängerdaten können so sieben Bewegungszustände unterschieden werden. Auch Waga et al. (2012) untersuchen Bewegungszustände, dies anhand von Segmenten von Trajektorien. In ihrer Untersuchung unterscheiden sie anhand der Geschwindigkeit zwischen Stoppen, Gehen, Rennen, Fahrradfahren und Autofahren.

Zimmermann et al. (2009) liefern einen weiteren Ansatz, T-OPTICS genannt, welcher ebenfalls die Erkennung von Stopps untersucht. Dabei gehen sie von fehlerbehafteten Trajektorien aus, welche Artefakte (Bewegungen, die nicht stattgefunden haben, sondern auf Aufnahmefehler zurück zu führen sind) beinhalten. Als Lösung für solche Trajektorien liefern Zimmermann et al. (2009) eine interaktive, dichte-basierte Clusteranalyse, wobei sowohl die räumlichen als auch die zeitlichen Eigenschaften einer Trajektorie dafür miteinbezogen werden.

Die Arbeit von Laube und Purves (2011) untersucht Trajektorien aus einem anderen Gesichtspunkt, wobei sie zeitlich sehr hoch aufgelöste Daten von Trajektorien grasender Kühe untersuchen. Dabei untersuchen sie die Daten auf verschiedenen zeitlichen Auflösungsstufen, wobei als Bewegungsparameter die Geschwindigkeit, die Sinuosität der Trajektorie und der Drehwinkel der Trajektorie einer Kuh dienen.

Salvucci und Goldberg (2000) arbeiten ebenfalls mit Trajektorien, wobei es sich

um Eye-Tracking-Daten handelt. Diese Daten bewegen sich auf einer kleineren zeitlichen als auch räumlichen Skala. In ihrer Arbeit stellen die Autoren verschiedene Algorithmen zur Detektion von Stopps vor, wobei sie zwischen räumlichen und zeitlichen Kriterien unterscheiden. Einer der Algorithmen bedient sich eines Hidden Markov Model, welches auch Kiefer und Giannopoulos (2012) bei der Arbeit mit Eye-Tracking-Daten verwenden.

### 2.5.2 Datenbankorientierte Ansätze

Spaccapietra et al. (2008) haben eine häufig zitierte (beispielsweise von Alvares et al. (2007), Baglioni et al. (2009), Bak et al. (2012), etc.) Arbeit zu Trajektorien geschrieben. Motivation für ihre Arbeit war der Umstand, dass heutige Datenbankmanagementsysteme nicht zufriedenstellend mit Rohdaten von Trajektorien umgehen können. In ihrer Arbeit liefern die Autoren und Autorinnen ein konzeptuelles Modell für Trajektorien, wodurch Trajektorien ökonomischer in Datenbanken gespeichert werden können. Sie sehen Trajektorien als eine Sequenz von Bewegungen, die von einem Stopp zum nächsten gehen. Eine Trajektorie ist somit völlig durch Stopps und Bewegungen abgedeckt. Auch Baglioni et al. (2009) sehen die Notwendigkeit, Trajektorien möglichst effizient nutzen zu können. Dabei wollen sie Rohdaten von Trajektorien nicht nur konzeptuell, sondern auch semantisch anreichern, um sie sinnvoll nutzen zu können. Sie sprechen von einem stufenweisen semantischen Anreicherungsprozess, welcher in Abbildung 2.2 veranschaulicht ist. Dabei werden in einem ersten Schritt die Rohdaten zu Trajektorien umgewandelt, welche in Stopps und Bewegungen unterteilt sind. In einer nächsten, dritten Stufe können Muster erkannt werden und auf der vierten und letzten Stufe kann auf das Bewegungsverhalten geschlossen werden. Um diesen vierten Schritt zu machen muss eine gute formale Grundstruktur (*ontology*) vorliegen, welche das Bewegungsverhalten des Untersuchungsobjekts, in diesem Fall des Menschen, ideal beschreiben kann (Baglioni et al., 2009).

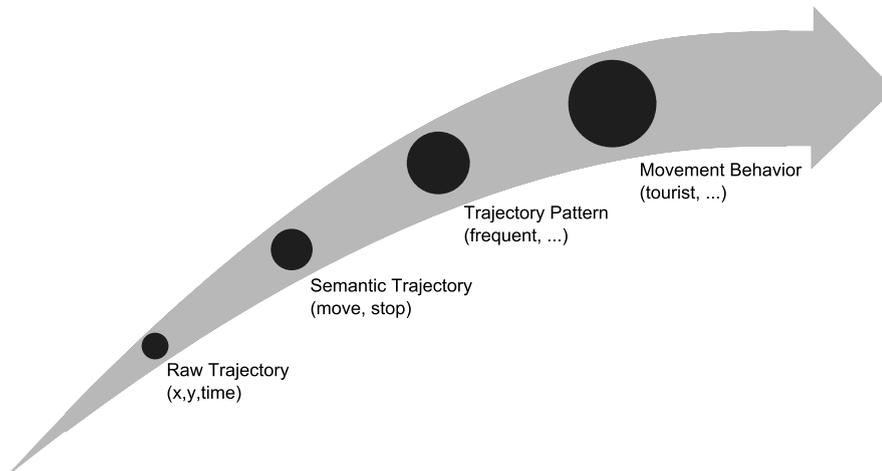


ABBILDUNG 2.2: Semantischer Anreicherungsprozess einer Trajektorie. Eigene Darstellung nach Baglioni et al. (2009): 275.

Alvares et al. (2007) gehen auch von einem Datenbankorientierten Ansatz aus. In ihrem Algorithmus SMO<sub>T</sub> spielen neben der eigentlichen Trajektorie auch sogenannte Kandidatenregionen (*candidate stops*) eine Rolle. Dies sind Regionen, in denen, ausgehend von Kontextinformationen mehr Stopps zu erwarten sind. Die Kandidatenregionen sind Flächen, deren Geometrie in einer Datenbank gespeichert ist, wobei auch die Geometrie der Trajektorien in derselben Datenbank gespeichert ist. So ist es mittels geometrischen Abfragen möglich, Trajektorienpunkte innerhalb dieser Regionen zu finden (Alvares et al., 2007). Auch Stenneth et al. (2011) und Baglioni et al. (2009) analysieren Bewegungsdaten mit dem Einbezug von Kontextinformation. Die Wichtigkeit von Kontextinformationen nennen auch Guc et al. (2008), wobei sie in ihrer Arbeit ein erweiterbares Modell zur Annotation von GPS-Trajektorien vorstellen. Auch Bogorny et al. (2009) sehen grosse Vorteile im Einbezug von Informationen zum geographischen Hintergrund der untersuchten Daten. Die Autoren und Autorinnen haben dazu eine Methodologie entwickelt, die sowohl Trajektorien als auch dazugehörige Hintergrundinformationen berücksichtigt.

Palma et al. (2008) beziehen sich auf die Arbeit von Alvares et al. (2007) und haben den Algorithmus CB-SMO<sub>T</sub> erarbeitet. Dieser Algorithmus stellt einen clusterbasierten Ansatz dar, der versucht, interessante Orte beziehungsweise Stopps in Trajektorien zu finden. Darin gehen die Autoren und Autorinnen erst in einem zweiten Schritt von Kandidatenregionen aus, wobei zuvor anhand der Geschwindigkeit und mit Hilfe einer Variation des DBSCAN (siehe dazu Ester et al. (1996)) sogenannte potentielle Stopps (Cluster) gefunden werden. Rinzivillo et al. (2008) bezeichnen Algorithmen, die sich am

DBSCAN (Ester et al., 1996) orientieren als geeignet für die dichte-basierte Analyse von Trajektorien. Dies einerseits, weil sie gut an verschiedenes Datenmaterial angepasst werden können und andererseits tolerant gegenüber Rauschen sind. Auch Rocha et al. (2010) haben eine Alternative zum SMoT von Alvares et al. (2007) entwickelt, den DB-SMoT, wobei sie zur Detektion von Stopps die Bewegungsrichtung berücksichtigen.

### 2.5.3 Verhaltensbiologisch orientierte Ansätze

Dank Fortschritten in der Technologie können immer kleinere Sensoren hergestellt werden und somit auch immer kleinere Tiere damit ausgerüstet werden. Aus diesem und anderen Gründen werden vermehrt Trajektorien von Tieren untersucht, um deren Verhalten zu studieren. Die Arbeit von Collins et al. (1995) ist ein frühes Beispiel für die Verwendung von Trajektorien zur Untersuchung des Tierverhaltens. Im Rahmen dieser Untersuchung wird sowohl ein räumlicher als auch ein zeitlicher Aspekt für die Detektion von Stopps berücksichtigt. Als Untersuchungsobjekt diente in diesem Fall die Hausfliege, deren Bewegungen auf Video aufgezeichnet wurden.

Buchin et al. (2013) verwenden Daten der Blässgans, welche im Winter in den Süden fliegen, wobei sie herausfinden wollen, wo die Gänse auf ihrer Reise einen Zwischenstopp einlegen. Als Bewegungsparameter verwenden sie einen Kriterienkatalog und unterteilen eine Trajektorie in Segmente. Ziel dieser Unterteilung ist es, möglichst wenige Segmente zu haben, deren einzelne Segmentteile alle dem definierten Kriterienkatalog entsprechen. Vier Kriterien werden berücksichtigt, diese sind Nachbarschaft (*location disk criteria*), Richtung, Geschwindigkeit und Zeitdauer. Buchin et al. (2013) können so verschiedene Verhaltensmuster der migrierenden Gänse unterscheiden und sehen unter anderem einen grossen Vorteil darin, Bewegungszustände manuell beschreiben zu können. Abbildung 2.3 zeigt die Visualisierung einer solchen Trajektorie, wobei die roten Punkte einen Zwischenstopp darstellen.

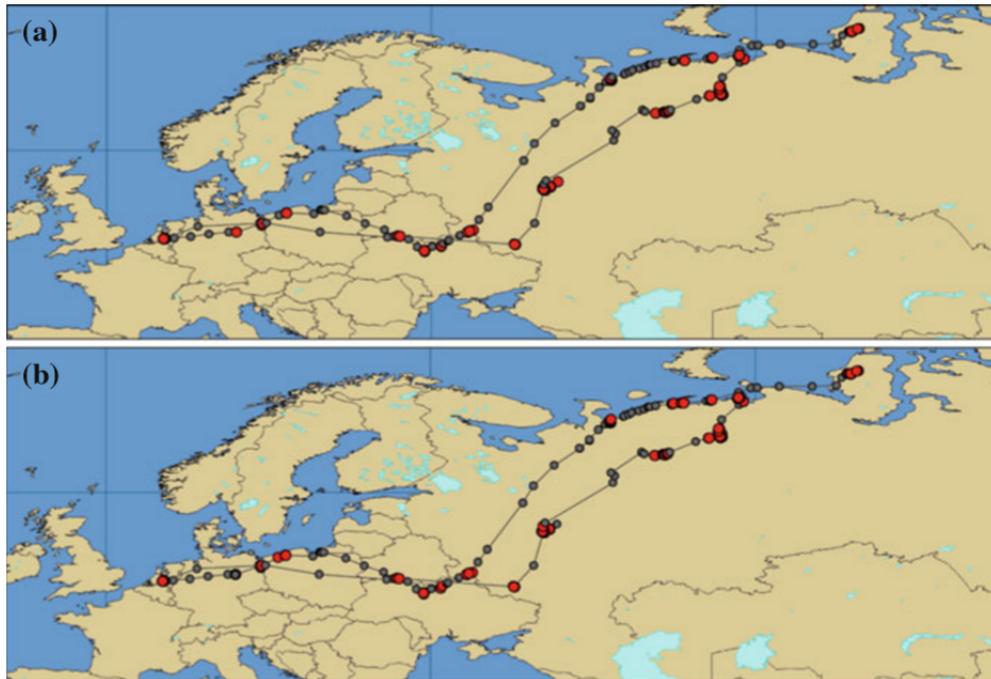


ABBILDUNG 2.3: Segmentierung der Trajektorien zweier Blässgänse, a) zeigt die manuelle Segmentierung und b) zeigt eine berechnete Segmentierung. Quelle: Buchin et al. (2013): 22.

Shamoun-Baranes et al. (2012) unterscheiden verschiedene Verhaltensweisen des Austernfischers anhand von GPS-Daten und Beschleunigungsmessern. Eine weitere, verhaltensbiologisch motivierte Untersuchung von Tieren liefern van Moorter et al. (2010). Dabei werden vier verschiedene Bewegungszustände von Elchen unterschieden. Dies wird mit Hilfe von Richtungsänderungen in der Trajektorie der Tiere, der Schrittlänge sowie den Kopfbewegungen der Tiere gemacht.

## 2.6 Gemeinsamkeiten und Unterschiede der bestehenden Methoden

Aus den obigen Ausführungen wird klar, dass das Thema der Detektion von Stopps in Trajektorien umfassend ist und Raum für weitere Analysen und Untersuchungen birgt. Im folgenden Abschnitt wird die Literatur eingehender diskutiert und Gemeinsamkeiten sowie Unterschiede zwischen den verschiedenen Forschungsgebieten und Arbeiten herausgefiltert.

**Gemeinsamkeiten der bestehenden Methoden** Salvucci und Goldberg (2000) liefern eine Zusammenstellung verschiedener Algorithmen zur Analyse von Eye-Tracking-Daten. Die spezifischen Algorithmen teilen sie in zwei Grobkategorien, räumlich und zeitlich, und weiter in fünf Untergruppen ein. Abbildung 2.4 gibt einen Überblick über diese Einteilung. Der Algorithmus I-VT arbeitet mit einem Geschwindigkeitsschwellenwert, wie Erkelens und Vogels (1995) oder auch Sen und Megaw (1984) es in ihrer Analyse tun. Der Algorithmus I-HMM, basiert auf dem Hidden Markov Model, welches in Abschnitt 4.5 besprochen wird. Beispielsweise Salvucci und Anderson (1998) arbeiten mit dem HMM, wobei sie es dazu verwenden, die Analyse von Augenbewegungen zu automatisieren. I-DT stellt einen Algorithmus dar, der auf der räumlichen Aggregation von Punkten basiert. Diese Art von Algorithmus wird beispielsweise in Widdel (1984) verwendet. Der I-MST arbeitet mit *minimum spanning trees*, wobei Punkte in bestimmter Art und Weise miteinander verbunden werden. Der Aufbau eines solchen *minimum spanning tree* wird in Camerini et al. (1988) genauer beschrieben. Der Algorithmus I-AOI verwendet als einziger der in Abbildung 2.4 aufgelisteten Algorithmen Kandidatenregionen zur Detektion von Stopps. Dabei werden lediglich Fixationen, die innerhalb dieser Kandidatenregionen liegen, identifiziert (beispielsweise Den Buurman et al. (1981)).

Criteria		Representative Algorithms				
		I-VT	I-HMM	I-DT	I-MST	I-AOI
Spatial	Velocity-based	X	X			
	Dispersion-based			X	X	
	Area-based					X
Temporal	Duration sensitive			X		X
	Locally adaptive		X	X	X	

ABBILDUNG 2.4: Taxonomie der Algorithmen zur Identifikation von Fixationen und Beispiele ausgewählter Algorithmen. Quelle: Salvucci und Goldberg (2000): 72.

Auf Grund der Rechenleistung und durch komplexe und vielseitige Anforderungen an Resultate, wird heute selten ein einzelnes der Kriterien in Abbildung 2.4 alleinig angewendet. Der grösste Teil der verwendeten Algorithmen bedient sich einer Kombination der Kriterien (beispielsweise bei Buchin et al. (2013), van Moorter et al. (2010) oder Shamoun-Baranes et al. (2012)). Eine solche Kombination von Kriterien liefert ein diversifizierteres Abbild der Daten und passende Parameter können gewählt werden. Die meisten bestehenden Algorithmen bedienen sich der räumlichen sowie zeitlichen Attribute von Bewegungsdaten. Es existieren wenige Methoden, welche sich bloss auf die räumlichen Gegebenheiten, wie beispielsweise die Dichte, beziehen. So ist die Geschwindigkeit, als raum-zeitliches Attribut ein Parameter, welcher fast bei jeder Methode berücksichtigt wird (beispielsweise bei Collins et al. (1995) und Bak et al. (2012)). Neben der Geschwindigkeit als häufig verwendeten Parameter wird auch sowohl die räumliche als auch die zeitliche Dichte in verschiedenen Varianten immer wieder herbeigezogen. Häufig wird auch eine Kombination der zeitlichen und räumlichen Dichte verwendet, so in Zimmermann et al. (2009) oder in Palma et al. (2008). Eine hohe Dichte von Datenpunkten scheint ein Indiz für einen Stopp zu sein, da sich ein Objekt an dieser Stelle vermehrt aufgehalten hat. Bei der Betrachtung der Dichte gilt es zu beachten, wie die Datenpunkte im Raum verteilt sind. Eine geringe Dichte kann unter anderem durch das Fehlen von Messpunkten bedingt sein, beispielsweise auf Grund schlechten GPS-Empfangs oder durch den Ausfall der Batterie des GPS-Empfängers, wenn sich das zu untersuchende Tier in einer Höhle befindet und der GPS-Empfänger durch eine Solarzelle angetrieben wird.

Eine weitere Gemeinsamkeit vieler Untersuchungen ist nicht die blosser Detektion von Stopps, sondern die Unterscheidung mehrerer Bewegungszustände. So detektieren Moreno et al. (2010) und Waga et al. (2012) anhand der Trajektoriendaten, wie sich Menschen fortbewegt haben. Moreno et al. (2010) bedienen sich für die Detektion der Bewegungszustände verschiedener Attribute mit spezifischen Schwellenwerten in unterschiedlichen Kombinationen. Waga et al. (2012) hingegen geben für alle Kombinationen von Attributen eine Wahrscheinlichkeit an, in welchem Bewegungszustand sich das Objekt befindet. Tabelle 2.1 ist ein Beispiel aus Moreno et al. (2010) für die Unterscheidung mehrerer Bewegungszustände von Menschen.

TABELLE 2.1: Beispiel für die Unterscheidung verschiedener Bewegungszustände von Menschen anhand dreier Bewegungsparameter. Eigene Darstellung nach Moreno et al. (2010): 17.

<b>minTime(min)</b>	<b>maxSpeed(km/h)</b>	<b>maxDirection</b>	<b>Goal</b>
15	7	50	walking
15	36	80	cycling
5	15	90	dog letting
5	4	45	photo
60	2	20	picnic
30	2	20	relaxing
30	20	50	running

**Unterschiede zwischen den bestehenden Methoden** Dem Attribut der Richtungsänderung zwischen zwei Punkten beziehungsweise des Drehwinkels eines Punktes als Indikator für das Auftreten eines Stopps wird in der Literatur weniger Aufmerksamkeit geschenkt, lediglich der DB-SMoT (Rocha et al., 2010) bedient sich ausschliesslich der Richtungsänderung, Buchin et al. (2013) verwendet die Richtungsänderung in Kombination mit anderen Parametern. Dies könnte unter anderem daran liegen, dass die Richtungsänderung für kleinere Tiere, im Gegensatz zu beispielsweise Zugvögeln, nicht direkt mit dem Bewegungsverhalten zusammenhängt. Ein kleinerer Vogel kann während eines Stopps zur Nahrungssuche sowie während eines Fluges entlang der Küste starke Richtungsänderungen aufweisen. Somit stellte der Einbezug der Richtungsänderung in diesem Fall einen kleinen Mehrwert dar, da die Richtungsänderung möglicherweise nicht direkt mit dem Auftreten von Stopps zusammenhängt.

Einige Methoden stützen sich nicht nur auf die vorhandenen Bewegungsdaten, sondern ziehen zusätzlich Kontextinformationen, wie teilweise Kandidatenregionen, mit ein (Alvares et al. (2007), Baglioni et al. (2009)). Dies scheint sinnvoll, wenn das Bewegungsverhalten, auf Grund von Eigenschaften des sich bewegenden Objekts, vorhergesehen werden kann. Wie Alvares et al. (2007) zeigen, eignet sich dieser Ansatz beispielsweise bei Touristenbewegungen in einer Stadt. Nachteil einer solchen Methode ist, dass die Daten nicht vollständig unvoreingenommen untersucht werden. Durch die alleinige Suche in den Kandidatenregionen werden allfällige Stopps ausserhalb dieser nicht wahrgenommen (Palma et al., 2008). Diese Problematik umgehen Palma et al. (2008) mit dem CB-SMoT, indem sie zuerst nach Clustern suchen und die gefundenen Cluster erst dann

auf eine Überschneidung mit Kandidatenregionen testen. Somit findet diese Methode sowohl 'echte' als auch unerwartete, sogenannte 'unbekannte' Stopps.

## 2.7 Forschungslücken und Forschungsfragen

Meines Erachtens besteht eine Forschungslücke in einem Vergleich der bestehenden Methoden. Da die vorhandenen Methoden meist nur auf einen spezifischen Datensatz angewendet werden, besteht keine Aussage zur Robustheit der Methode und der darin verwendeten Bewegungsparameter bei deren Anwendung auf andere Datensätze. Nur zu gewissen Algorithmen wird eine Aussage zu Vor- und Nachteilen gemacht. Beispielsweise für die drei miteinander verwandten Algorithmen SMoT (Alvares et al., 2007), CB-SMoT (Palma et al., 2008) und DB-SMoT (Rocha et al., 2010). Hierbei stellt der CB-SMoT eine Erweiterung des SMoT dar und der DB-SMoT richtet sich an Datensätze, welche mit dem CB-SMoT nur unzulänglich untersucht werden können. Für die meisten der genannten Algorithmen jedoch werden lediglich anwendungsspezifische Vorteile genannt, eine Aussage zur Robustheit wird selten gemacht.

Weiter sind die vorhandenen Methoden von unterschiedlicher Komplexität. Beispielsweise der Algorithmus von Collins et al. (1995), welcher einzig die Geschwindigkeit verwendet, um Stopps zu detektieren, ist im Bezug auf dessen Komplexität stark unterschiedlich vom T-OPTICS von Zimmermann et al. (2009).

Somit besteht ein relativ grosses Spektrum an unterschiedlichen Methoden, ohne dass bis heute eine vergleichende Aussage zu deren Robustheit und Eignung für unterschiedliche Daten und unterschiedliche sich bewegende Objekte gemacht worden wäre.

Im Rahmen dieser Arbeit soll untersucht werden, wie verschiedene Algorithmen reagieren, wenn diverse Datensätze damit prozessiert werden. Durch dieses systematische Vorgehen soll herausgefunden werden, wie robust die Methoden sind und ob komplexe Methoden einen signifikanten Mehrwert bieten gegenüber einfacheren Methoden. Die Vielfalt an Algorithmen unterschiedlicher Komplexität wurde entwickelt für diverse Anwendungen, jedoch mit demselben Ziel, nämlich der Detektion von Stopps in Trajektorien. Dies macht meines Erachtens einen Vergleich interessant und von Notwendigkeit. Aus diesen Beobachtungen leitet sich folgende Forschungsfrage ab:

- Wie reagieren existierende Methoden zur Detektion von Stopps in Trajektorien bei deren Anwendung auf verschiedene Datensätze, beziehungsweise wie robust sind sie?

Da in der Literatur immer wieder ähnliche Bewegungsparameter verwendet werden, jedoch mit anwendungsspezifischen Schwellenwerten und Implementierungen, ist meines Erachtens auch ein Vergleich einzelner Bewegungsparameter interessant und sinnvoll. Damit soll herausgefunden werden, ob eine Kombination berücksichtigter Bewegungsparameter zuverlässiger Stopps detektiert als ein einzelner dieser Bewegungsparameter. Daraus folgt die zweite Forschungsfrage:

- Bietet die Berücksichtigung einer Kombination von Bewegungsparametern beziehungsweise die Verwendung von komplexen Algorithmen bei der Erkennung von Stopps in Trajektorien einen Mehrwert gegenüber einfacheren Methoden?

Falls die obige Forschungsfrage mit ja beantwortet werden kann, stellt sich die unten stehende dritte Forschungsfrage, welche Bewegungsparameter oder welche Kombination von Bewegungsparametern und welche Algorithmen viel beziehungsweise wenig zum Finden von Stopps beitragen. Hierbei könnte als Ergebnis eine möglichst ideale Kombination von Bewegungsparametern hervorgehen, insofern eine solche existiert.

- Welche Algorithmen tragen wenig/viel zum Erkennen von Stopps bei?

## Kapitel 3

# Datenlage und Vorverarbeitungsschritte

### 3.1 Verwendete Daten

In einem frühen Schritt dieser Arbeit mussten geeignete Datensätze gefunden werden, anhand derer die Methoden verglichen werden können. Ideal wäre es, Datensätze zu finden, die bezüglich der Genauigkeit ähnlich sind. Werden Datensätze verwendet, welche stark unterschiedliche Genauigkeiten aufweisen, so kann das Resultat des Methodenvergleichs stark von dieser Genauigkeit abhängig sein und weniger von den verwendeten Methoden selber. Weiter wäre es für eine Analyse der hier genannten Art und Weise wünschenswert, Datensätze zur Verfügung zu haben, zu welchen die *ground truth* vorhanden ist. So könnte zuverlässig verifiziert werden, ob die gefundenen Stopps auch tatsächlich stattgefundene Stopps sind. Im Folgenden wird auf die verwendeten Datensätze eingegangen, wobei diese beschrieben sowie deren Besonderheiten erläutert werden. Eine Zusammenfassung der Datensätze und ihrer Eigenschaften ist in Anhang B zu finden.

#### 3.1.1 Wandernde Person in der Schweiz

Die Wanderdatensätze wurden im Rahmen von kleineren Ausflügen von mir und meinem Partner aufgenommen, somit sind bei diesen Datensätzen die Stopps bekannt. Bei dem sich bewegenden Objekt in diesen Datensätzen handelt es sich um Wanderer, die zu Fuss

unterwegs sind. Die Wanderdatensätze enthalten die Koordinaten im WGS84, eine Angabe zur Momentangeschwindigkeit, eine Höhenangabe und eine in Metern angegebene Genauigkeit (*accuracy*). Die Genauigkeit variiert zwischen zwei und 48 Metern und ist im Schnitt über alle verwendeten Dateien 6.74 Meter. Das Aufnahmeintervall ist rund 10 Sekunden, variiert jedoch zwischen den einzelnen Datensätzen. Die Daten wurden mit einem Mobiltelefon aufgenommen. Um dieses Telefon genügend lange gebrauchen zu können und um nicht allzu grosse Dateien zu generieren wurde die Aufnahme von Messpunkten beschränkt. So wurde kein neuer Punkt aufgezeichnet, wenn das GPS-Signal sich in einem Radius von 10 Metern des letzten Punktes befand. Bei der Analyse und der Auswertung der Resultate zum Wanderdatensatz gilt es diesem Umstand Rechnung zu tragen. Die hier verwendeten Datensätze wurden grösstenteils im Wald aufgenommen, wobei das GPS-Signal hier gestört werden kann, wie Sigrist et al. (1999) argumentieren. Sie zeigen, dass mit zunehmender Verschliessung des Baumkronendachs die Effizienz der Signalaufzeichnung sowie die Genauigkeit des Signals abnehmen.

### 3.1.2 Betonliefernde Lastwagen in Athen

Der Datensatz mit Trajektorien von Betonliefernden Lastwagen in Athen wurde unter Chorochronos<sup>1</sup> heruntergeladen, wobei die Version *Trucks revised* gewählt wurde. Pelekis et al. (2010) haben mit diesem Datensatz gearbeitet und den Effekt der Unsicherheit bei der Bildung von Clustern aus Trajektorien studiert. Der Datensatz umfasst 276 Trajektorien von insgesamt 50 Lastwagen an 33 verschiedenen Tagen. Das zeitliche Intervall ist konstant 30 Sekunden, wobei es sein kann, dass zwei verschiedene Trajektorien von einem Lastwagen an einem Tag stammen. Hierbei besteht lediglich ein längeres Zeitintervall zwischen zwei Punkten desselben Tages. Zur Analyse wurde dieser Datensatz in einzelne Dateien unterteilt (eine Datei enthält die Trajektorie eines Lastwagens an einem Tag). Der Datensatz enthält Punkte im WGS84 und im lokalen, kartesischen Koordinatensystem von Griechenland (*GGRS87*), wobei keine Angaben zur Momentangeschwindigkeit oder zur Richtung vorhanden sind. Bei diesem Datensatz gilt es zu beachten, dass die Lastwagen sich auf einem Netzwerk bewegen, sie bewegen sich demnach nicht frei und müssen dem Strassennetzwerk folgen. Weiter ist es wichtig, sich bewusst zu sein, dass das GPS-Signal in einer Stadt stärker gestört wird als im Freien<sup>2</sup>.

<sup>1</sup> Link zu den Lastwagen-Daten (Stand 06.02.2014): <http://www.chorochronos.org/?q=node/10>

<sup>2</sup> Quelle (Stand 11.06.2014): <http://www.montana.edu/gps/understd.html>

**Kandidatenregionen zu den Lastwagen in Athen** Da einer der verwendeten Algorithmen (SMoT von Alvares et al. (2007)) mit Kandidatenregionen arbeitet, ist es von Vorteil, Informationen zu diesen Regionen zur Verfügung zu haben. Dies konnte bei den Lastwagendaten teilweise realisiert werden. Die Lastwagen bewegen sich in der Region Athen und zu den Interessenregionen gehören unter anderem Ampeln, da diese zu einem Stopp auffordern können. Als Interessenregionen wurden demnach die Ampeln aus einem Datensatz von OSM extrahiert. Dazu wurden die Daten zu Griechenland<sup>3</sup> heruntergeladen, in *ArcGIS* visualisiert und auf den Bereich reduziert, in dem sich die Lastwagen bewegen. Der Datensatz enthält Informationen zu verschiedenen Bereichen und es musste entschieden werden, welche Objekte als Kandidatenregionen verwendet werden sollen. Dazu wurden alle im Datensatz vorkommenden *tags* konsultiert, wobei lediglich Punkte mit dem tag *traffic\_signals* berücksichtigt wurden. Nach der Filterung waren ausschliesslich Punktdaten vorhanden.

Die Auswahl der Interessenregionen wurde aus verschiedenen Gründen auf die Ampeln begrenzt. Einerseits konnte so die Anzahl an Interessenregionen stark reduziert werden, was eine grosse Zeitersparnis bei der Analyse bedeutet hat. Das Lesen aller Interessenregionen sowie der Algorithmus, der testet, ob sich Punkte in den Interessenregionen befinden, hat mit Abstand am meisten Rechenzeit von allen Algorithmen in Anspruch genommen. Weiter existierten beispielsweise zu dem tag *crossing* sehr viele Punkte, die teilweise sehr nahe beieinander sind. Es stellt sich die Frage, inwiefern sehr kleine Kreuzungen für Lastwagen wichtig sind. Es ist davon auszugehen, dass die Lastwagen vor allem auf Hauptstrassen unterwegs sind, da sie enge Kurven nur schlecht passieren können und da die Hauptstrassen oftmals den direktesten Weg zu einer Baustelle darstellen.

### 3.1.3 Ausgewählte Austernfischer auf *Schiermonnikoog*

Der Austernfischer, *haematopus ostralegus*, ist ein Stelzvogel, der unter anderem auf der holländischen Watteninsel *Schiermonnikoog* vorkommt. Er ernährt sich von hartschaligen Muscheln und grossen marinen Würmern, welche in den Zonen der Gezeiten vorkommen (Shamoun-Baranes et al., 2012). Der Bewegungsradius des Vogels ist im Vergleich zu den Bewegungsradien der anderen hier verwendeten Daten nicht besonders

---

<sup>3</sup> Link zu den *OpenStreetMap*-Daten (Stand 06.02.2014):  
<http://download.geofabrik.de/europe/greece.html>

gross. Die Tiere bewegen sich frei und folgen keinem Netzwerk. Auf Grund des flachen Geländes kann der Satellit, also der Sender des GPS-Signals, eine relativ direkte Verbindung zum Empfänger des GPS-Signals machen, wobei das Signal dabei weniger gestört wird als beispielsweise in einem Bergtal<sup>4</sup>.

Die in dieser Arbeit verwendeten Daten des Austernfischers stammen aus einer Feldstudie (Shamoun-Baranes et al., 2012) der Universität Amsterdam aus den Monaten Mai bis Juli 2009. Die Originaldaten wurden mit Hilfe der Arbeit von Schreutelkamp und Strang van Hees (2001) in das Koordinatensystem WGS84 transformiert. Zur Untersuchung wurden drei Austernfischer mit dem Sensorgerät UvA-BiTS ausgerüstet, wobei das zeitliche Intervall der Standortmessungen der Vögel zwischen 30 Minuten und 10 Sekunden variiert. Für die Analyse habe ich mich auf drei Trajektorien beschränkt. Die gewählten Trajektorien wurden anderen vorgezogen, da sie zeitlich sehr hoch aufgelöst sind und die *ground truth* dazu teilweise bekannt ist.

Ob es sinnvoll ist, Trajektorien mit einer hohen zeitlichen Auflösung zu nehmen, lässt sich diskutieren. Eine Arbeit von Laube und Purves (2011) hat Trajektorien von unterschiedlicher zeitlicher Auflösung untersucht, wobei sie zeigen konnten, dass eine sehr feine zeitliche Auflösung nicht immer ideal ist. Hier, im Rahmen dieser Arbeit, wurde jedoch diese Vorgehensweise gewählt, da die Geschwindigkeit wie in Abschnitt 4.1.1 berechnet wurde. Für Bewegungsdaten, welche sich auf einem engen räumlichen Ausschnitt bewegen, jedoch ein grobes zeitliches Aufnahmeintervall haben, macht die gewählte Vorgehensweise aus Abschnitt 4.1.1 wenig Sinn. Insbesondere bei einem kleinen Vogel kann davon ausgegangen werden, dass er sich während 30 Minuten bewegt, wobei all diese Bewegungen bei einer groben zeitlichen Auflösung nicht registriert werden. Ist dies der Fall, so werden geringere Geschwindigkeitswerte berechnet, als tatsächlich vorhanden waren, da nicht alle Bewegungen abgebildet werden. Aus diesem Grund wurden lediglich drei zeitlich fein aufgelöste Trajektorien ausgewählt.

### 3.1.4 Zugvogel Leo in Amerika

Die Zugvogeldata<sup>5</sup> wurden von *movebank* heruntergeladen. Es wurde der Datensatz *Turkey Vulture Acopian Center USA GPS* gewählt. Bei dieser Vogelart handelt es sich um den Truthahngerier, *cathartes aura*, der das grösste Aas-fressende Beutetier darstellt.

<sup>4</sup> Quelle (Stand 11.06.2014): <http://www.montana.edu/gps/understd.html>

<sup>5</sup> Link zu den Zugvogel-Daten (Stand 06.02.2014): [www.movebank.org](http://www.movebank.org)

Die Daten wurden von Mandel et al. (2011) verwendet und analysiert, wobei untersucht wurde, wie die Vögel während ihres Fluges in den Süden auf ihre Umgebung reagieren. Die Daten werden als eine grosse Datei zur Verfügung gestellt. Darin enthalten sind Trajektorien von 38 Vögeln. Das zeitliche Intervall der Aufnahme variiert zwischen 1 und rund 3 Stunden. Die Aufnahmedauer beträgt bei den grösseren Datensätzen bis zu sieben Jahre. Eine Angabe zur momentanen Flugrichtung ist ebenfalls in der Datei enthalten. Die Geschwindigkeit ist nur bei einigen Datenpunkten angegeben, wobei anzunehmen ist, dass es sich dabei um die Momentangeschwindigkeit, welche durch das GPS-Gerät aufgenommen wurde, handelt. Für die Analyse im Rahmen dieser Arbeit wurde die Trajektorie eines Vogels gewählt, obwohl mehrere Trajektorien zur Verfügung gestanden haben. Einerseits ist davon auszugehen, dass der Mehrwert bei der Integration einer Vielzahl ähnlicher Trajektorien in die Analyse nicht sehr gross ist, da keine neuen oder weiterführenden Informationen hinzukommen. Andererseits sind die einzelnen Dateien sehr gross, eine Analyse mit allen Dateien durchzurechnen wäre sehr aufwändig und zeitintensiv gewesen. Die verwendete Datei, die es Vogels Leo, ist die grösste aller einzelnen Dateien.

### **3.1.5 Eye-Tracking-Daten**

Die Eye-Tracking-Daten wurden von Dr. Arzu Çöltekin vom Geographischen Institut der Universität Zürich zur Verfügung gestellt. Die genannten Daten stammen aus einer Untersuchung von zwei unterschiedlich gestalteten, interaktiven Onlinekarten mit äquivalentem Informationsgehalt (Coltekin et al., 2009). Die Daten bestehen aus Dateien von 30 Versuchspersonen. Das Messintervall ist rund 17 Millisekunden und eine Messung dauert zwischen zehn und zwanzig Minuten. Da sich diese Daten auf einem kartesischen Koordinatensystem bewegen, ist die Berechnung der Geschwindigkeit und des Drehwinkels einfacher als bei Trajektorien auf der Erdoberfläche. Für die späteren Analysen habe ich mich auf eine Datei beschränkt, wobei für den Zweck dieser Arbeit sekundär ist, welche Onlinekarte dabei bearbeitet wurde.

Bei den Eye-Tracking-Daten handelt es sich nicht um Bewegungsdaten im Sinne der anderen verwendeten Datensätze. Das sich bewegende Objekt bei dieser Art von Bewegungsdaten, also die Blickbewegung des Auges, hat so gut wie keine Masse. Dies führt dazu, dass sehr hohe Geschwindigkeiten erreicht werden können. Rein physikalisch gesehen kann somit der Vergleich zwischen Eye-Tracking-Daten und den anderen

hier genannten Bewegungsdaten als fragwürdig angesehen werden. Diese Daten wurden trotzdem bei der Analyse verwendet, um genau diese Eigenart auszunutzen und die Robustheit der Methoden bei der Anwendung auf Daten anderer Charakteristika zu testen. Weiter findet die gemessene Bewegung, im Gegensatz zu den Bewegungen in den anderen Datensätzen, auf einem sehr kleinen Ausschnitt statt. Dieser Ausschnitt ist nicht Teil der Erdoberfläche, sondern stellt einen Bildschirm dar, wobei Bewegungen ausserhalb des Bildschirms möglich sind.

## 3.2 Datenverarbeitung

Da die Daten, wie oben erläutert, alle in unterschiedlichen Formaten und Grössen zur Verfügung standen, galt es in einem ersten Schritt die einzelnen Datensätze in eine sinnvolle Anzahl Dateien zu unterteilen und diese Dateien zur späteren Analyse möglichst homogen zu gestalten. Aus diesem Grund wurden die Kopfzeile und jeweils die ersten sechs Einträge jeder Zeile jeder Datei gleich gestaltet. Jede Zeile enthält folgende Angaben eines Punktes: ID, Zeitstempel, Längengrad, Breitengrad, Drehwinkel, Geschwindigkeit. Wenn eine Angabe nicht vorhanden war, wurde der Wert 0.0 eingefügt, wobei davon ausgegangen wird, dass dieser Wert sonst nicht vorkommt und er somit nicht mit einem anderen Eintrag verwechselt werden kann. Waren mehr Informationen vorhanden als die genannten sechs Attribute, wurden diese am Ende einer Zeile eingefügt. Trotz der zeitaufwändigen Umstrukturierung der Daten macht diese Vorgehensweise Sinn, da eine klare und einheitliche Struktur für die weitere Prozessierung und die Visualisierung einen grossen Vorteil darstellt.

Die Datenverarbeitung hat einige Zeit in Anspruch genommen, unter anderem, weil einige leere Zeilen oder ungültige Zeichen gefunden und speziell behandelt werden mussten. Zudem hat das Lesen, Umformatieren und Schreiben der Daten einiges an Prozessierungszeit benötigt. Beispielsweise bei der Datei zu den Zugvögeln wurden insgesamt 38 neue Dateien aus der ursprünglichen Datei abgeleitet, wobei die grössten um 6000 KB gross sind und über 35'000 Zeilen beziehungsweise Datenpunkte enthalten. Bei den Daten zum Austernfischer wurden hingegen eher kleine Dateien generiert. So ist die kleinste Datei 1 KB gross und enthält nur drei Messpunkte. Da die Eye-Tracking-Daten keine Angabe zum Datum enthalten und lediglich die Uhrzeit vorhanden ist, wurde ein

willkürliches Datum eingefügt, um die einheitliche Struktur zu erhalten. Bei den Wanderdaten wurde das Datum geändert, um die Daten zu anonymisieren.

# Kapitel 4

## Methodisches Vorgehen

### 4.1 Rolle der Geschwindigkeit

#### 4.1.1 Was ist Geschwindigkeit?

Die Geschwindigkeit scheint bei der Arbeit mit Trajektorien und im Besonderen bei der Detektion von Stopps ein allgegenwärtiger und sehr häufig erwähnter Bewegungsparameter zu sein (siehe Kapitel 3). Was genau unter Geschwindigkeit zu verstehen ist, wird jedoch selten klar definiert. Gewisse GPS-Empfänger registrieren eine Momentangeschwindigkeit, welche für jeden einzelnen Punkt gemessen wird. Wurde diese Momentangeschwindigkeit nicht aufgenommen, so kann die Geschwindigkeit aus der zurückgelegten Distanz und der vergangenen Zeit berechnet werden. Wird eine Trajektorie mit einer sehr feinen zeitlichen Auflösung aufgenommen, so ist der Unterschied zwischen der Momentangeschwindigkeit und der berechneten Geschwindigkeit klein. Wird eine Trajektorie jedoch in einem sehr groben zeitlichen Intervall aufgenommen, so ist dieser Unterschied zwischen der Momentangeschwindigkeit und der berechneten Geschwindigkeit grösser. Dies ist so, weil eine grobe zeitliche Auflösung weniger Bewegungen registrieren kann als eine sehr feine zeitliche Auflösung. Der Datensatz zu Zugvogel Leo (siehe Abschnitt 3.1.4) beispielsweise erstreckt sich über einen sehr grossen räumlichen Ausschnitt, von Kanada bis Venezuela, als auch über eine grosse Zeitdauer. Um solche Daten unter anderem sinnvoll abspeichern und prozessieren zu können, macht ein sehr kurzes Messintervall keinen Sinn, weshalb es nachvollziehbar ist, dass grosse Intervalle gewählt wurden. Wird bei diesem Datensatz nun die Momentangeschwindigkeit verglichen mit einer über

mehrere Punkte berechnete Geschwindigkeit, so resultieren unterschiedlichere Werte als bei einem Datensatz mit einer feinen zeitlichen Auflösung. Bei den hier verwendeten Datensätzen wurde in einigen Fällen keine Momentangeschwindigkeit registriert, weiter weisen die Datensätze sehr unterschiedliche zeitliche Aufnahmeintervalle auf.

**Geschwindigkeit im Rahmen dieser Arbeit** Es kann somit gesagt werden, dass es keine einheitliche Definition der Geschwindigkeit gibt. Aus diesem Grund wird in dieser Arbeit in einem ersten Schritt die Berechnung der Geschwindigkeit genauer angeschaut. Da viele der untersuchten Methoden auf einer bestimmten Definition der Geschwindigkeit beruhen, ist es von Wichtigkeit, dass dieser Geschwindigkeitswert gewissen Kriterien genügt. In einem ersten Schritt wird somit für jeden Datenpunkt aller Datensätze, mit Ausnahme der Eye-Tracking-Daten, über unterschiedliche zeitliche als auch räumliche Intervalle ein Geschwindigkeitswert berechnet. Aus all den resultierenden Geschwindigkeitswerten wird der Mittelwert für jeden Punkt berechnet. Der so erhaltene gemittelte Wert für die Geschwindigkeit wird für jeden einzelnen Punkt separat berechnet und wird für das weitere Vorgehen verwendet. Für die Berechnung einer Distanz auf der Erdoberfläche kann nicht die kartesische Geometrie verwendet werden. Aus diesem Grund wird ein Paket<sup>1</sup> verwendet, welches verschiedene Operationen im Zusammenhang mit Daten im WGS84 zur Verfügung stellt.

Bei den Eye-Tracking-Daten wird auf das obige Vorgehen verzichtet, da sich diese Daten in einem sehr kleinen, beschränkten Raum bewegen und die bewegte Masse verschwindend klein ist. Hierbei wäre bei einer Berechnung der Geschwindigkeit über mehrere Punkte die Gefahr einer Verfälschung des Geschwindigkeitswerts verhältnismässig gross. Zudem sind die zeitlichen Abstände zwischen den einzelnen Datenpunkten so klein, dass durch eine Berechnung über mehrere Punkte kein Mehrwert zu erwarten ist. Weiter sind die zeitlichen und die räumlichen Dimensionen zwischen den Eye-Tracking-Daten und den restlichen Datensätzen fundamental unterschiedlich, weshalb eine identische Prozessierung zur Berechnung der Geschwindigkeit nicht zwingend ist.

---

<sup>1</sup> Link zum Paket *geodesy* (Stand 08.11.2013):  
<http://www.gavaghan.org/blog/free-source-code/geodesy-library-vincentys-formula-java/>

**Zweck des in dieser Arbeit gewählten Vorgehens** Zweck der beschriebenen Berechnung der Geschwindigkeit ist es, die Vergleichbarkeit der zu untersuchenden Methoden zu gewährleisten. Im hypothetischen Fall, dass alle zu analysierenden Daten bereits Geschwindigkeitswerte aufweisen würden und mit einer ähnlichen Genauigkeit registriert worden wären, könnte auf die hier gemachten Berechnungen verzichtet werden. Im Rahmen dieser Arbeit werden jedoch Datensätze verwendet, die entweder zeitlich sehr hoch oder sehr tief aufgelöste oder gar keine Angaben zur Geschwindigkeit enthalten. Ein Vergleich auf einer solch uneinheitlichen Basis soll hiermit vermieden werden.

Die Aufnahmeintervalle der hier verwendeten Datensätze, ausser der Eye-Tracking-Daten, reichen von zehn Sekunden bis zu knapp vier Stunden. Wird nun die Geschwindigkeit über lediglich drei Punkte berechnet, so ist davon auszugehen, dass die resultierenden Werte unterschiedliche Gegebenheiten darstellen. Bei einem feinen zeitlichen Intervall werden die meisten Bewegungen in die Berechnung miteinbezogen, bei einem groben zeitlichen Intervall werden viele Bewegungen nicht miteinbezogen, da sie durch das Aufnahmeintervall nicht registriert worden sind.

Speziell bei der Detektion von Stopps in Trajektorien mittels unterschiedlichen Methoden, wobei die Geschwindigkeit ein elementarer Bewegungsparameter ist, ist es von Interesse, einheitlich ermittelte Geschwindigkeitswerte zu haben, damit der Methodenvergleich aussagekräftig ist. Ansonsten kann nicht eindeutig gesagt werden, ob eine auf der Geschwindigkeit basierende Methode erfolgreich Stopps detektiert hat oder ob lediglich die unterschiedlichen Geschwindigkeitsangaben zum jeweiligen Resultat geführt haben. Diese Vorgehensweise macht ebenfalls Sinn, da auch Datensätze ohne Geschwindigkeitsangabe in einen Vergleich einbezogen werden können. Auch mit der hier beschriebenen Berechnung der Geschwindigkeit bleiben die Charakteristika der einzelnen zu untersuchenden Objekte der verschiedenen Datensätze bestehen. Das Verhalten der sich bewegenden Objekte ändert sich nicht, auch deren Bewegungsradius bleibt gleich sowie deren allgemeine Eigenschaften.

#### 4.1.2 Implementation der Geschwindigkeitsberechnung

Für die Berechnung der Geschwindigkeit muss entschieden werden, über welche zeitlichen sowie räumlichen Intervalle die Geschwindigkeit ermittelt wird. Diese Intervalle werden für jeden Datensatz individuell gewählt, da auch jeder Datensatz spezifische Charakteristika aufweist. So macht es beispielsweise keinen Sinn, die Berechnung über

ein zeitliches Intervall zu machen, welches die Daten gar nicht abdecken. Bei den Vogelzugdaten ist eine Aggregation über 30 Sekunden beispielsweise nicht sinnvoll, bei den Austernfischerdaten hingegen schon.

Es ist nicht davon auszugehen, dass bei der Anwendung der verschiedenen zeitlichen und räumlichen Intervalle für die Geschwindigkeitsberechnung Punkte gefunden werden, welche genau dem Intervall entsprechen. Meistens wird der Fall vorliegen, dass ein Punkt innerhalb und der nächste ausserhalb des Intervalls zur Berechnung der Geschwindigkeit liegt und es muss entschieden werden, wie die Berechnung für diese undefinierte Zone gemacht wird. Dazu gibt es verschiedene Möglichkeiten. Eine davon ist, jeweils den letzten Punkt innerhalb oder den ersten Punkt ausserhalb des Intervalls für die Berechnung zu verwenden. Es besteht auch die Möglichkeit, zwischen den beiden Grenzpunkten zu interpolieren und so eine Approximation des auf der Grenze liegenden Wertes zu erhalten. Eine weitere Möglichkeit ist, den Punkt zu wählen, der näher an der Grenze liegt und der somit das Intervall adäquater abbildet. Letztere Variante wird bei dieser Arbeit gewählt. Im Folgenden wird diese Vorgehensweise für die zeitlichen sowie die räumlichen Intervalle erläutert.

**Zeitliche Aggregation für Berechnung der Geschwindigkeit** Tabelle 4.1 gibt Auskunft über die zeitlichen Aggregationsintervalle für die einzelnen Datensätze. Bei der Wahl dieser Werte wird darauf geachtet, dass der kleinste Wert möglichst dem Messintervall des jeweiligen Datensatzes entspricht. Der höchste Wert entspricht in den meisten Fällen dem zehnfachen des kleinsten Wertes. Dieses Vorgehen wird gewählt, da bei den meisten Datensätzen mit einer noch höheren Berechnungsdauer als der verwendeten die Geschwindigkeit stabil bleibt. Einzig bei den Wanderdaten ist dies nicht der Fall; bei diesen Daten hat sich die Geschwindigkeit erst ab einer noch höheren Berechnungsdauer stabilisiert. Daher wird die Geschwindigkeit hier über insgesamt eine längere Zeitspanne berechnet. Dieser Umstand kann daher kommen, dass bei diesen Daten kein Signal registriert wird, wenn sich der Punkt in einem Radius von 10 Metern des letzten registrierten Punktes befindet. Weiter wird bei den Wanderdaten nur alle 10 Sekunden nach dem GPS-Signal gesucht. Befindet sich ein Punkt nach 10 Sekunden noch immer innerhalb eines Radius von 10 Metern des letzten, so wird kein neuer Punkt aufgezeichnet und die nächste Registrierung eines GPS-Signals erfolgt 10 Sekunden später.

TABELLE 4.1: Zeitliche Schwellenwerte zur Berechnung der Geschwindigkeit.

	Berechnungsdauer [s]
Wanderer	15, 30, 60, 90, 120, 180, 240, 300, 360, 420
Lastwagen	30, 60, 90, 120, 150, 180, 210, 240, 270, 300
Austernfischer	20, 40, 60, 80, 100, 120, 140, 160, 180, 200
Truthahngerier	60, 120, 180, 240, 300, 360, 420, 480, 540, 600 [min]

Von jedem Datenpunkt aus wird jeweils in zwei Richtungen entlang der Trajektorie nach dem zeitlich am nächsten am Intervall gelegenen Datenpunkt gesucht. Angenommen der Schwellenwert für die Geschwindigkeitsberechnung ist 30 Sekunden. Befindet sich nun der letzte Datenpunkt innerhalb des Intervalls 20 Sekunden von dem zu analysierenden Punkt entfernt und der erste Datenpunkt ausserhalb des Intervalls 35 Sekunden entfernt, so wird der erste Datenpunkt ausserhalb des Intervalls für die Berechnung gewählt.

**Räumliche Aggregation für Berechnung der Geschwindigkeit** Bei der Berechnung der Geschwindigkeit über räumliche Intervalle wird noch mehr auf die Individualität der einzelnen Datensätze eingegangen. Die verwendeten Werte sind in Tabelle 4.2 zu finden. Mit Ausnahme der Daten zum Truthahngerier stellt das kleinste Intervall die ungefähre Grösse des untersuchten Objekts dar. Sowohl bei den Wanderdaten als auch bei den Daten zum Austernfischer ist die maximale Ausdehnung für die Berechnung das zwanzigfache des ersten Wertes. Dies macht Sinn, da so kleine und grosse Perimeter miteinbezogen werden. Bei den Lastwagendaten stellt der grösste Wert das zehnfache des kleinsten dar. Da sich Lastwagen auf einem Netzwerk bewegen und ihre Trajektorie nicht völlig frei wählen können, wird kein grösseres Intervall gewählt. Bei dem Datensatz zum Truthahngerier wird eine Spannweite von einem sehr grossen bis zu einem sehr kleinen Intervall gewählt. Buchin et al. (2013) definieren in ihrer Arbeit einen Stopp von Zugvögeln als Zustand, in dem die Vögel während mindestens 48 Stunden in einem Radius von 30 km bleiben. Es ist also gut möglich, dass ein Stopp einen Radius von 30 km aufweist; deshalb wird hier die Geschwindigkeitsberechnung über diese Grösse berechnet.

TABELLE 4.2: Räumliche Schwellenwerte zur Berechnung der Geschwindigkeit.

	Berechnungsperimeter [m]
Wanderer	2, 5, 10, 15, 20, 25, 30, 35, 40, 45
Lastwagen	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Austernfischer	2, 5, 10, 15, 20, 25, 30, 35, 40, 45
Truthahngerier	50, 100, 200, 400, 800, 1'600, 3'200, 6'400, 12'800, 25'600

Im Falle der räumlichen Berechnung gibt es ebenfalls mehrere Optionen, wie ein Schwellenwert umgesetzt werden kann, wobei folgendermassen vorgegangen wird. Für den zu analysierenden Punkt wird die direkte Distanz zu den zwei Punkten nahe des Intervalls berechnet. Der Punkt, dessen berechnete Distanz die kleinere Differenz zum Wert des Intervalls aufweist, wird für die Berechnung verwendet. Abbildung 4.1 zeigt den Sachverhalt schematisch auf. Die vier roten Punkte in dieser Abbildung kommen alle als Berechnungspunkte in Frage. Die nach dem beschriebenen Vorgehen ausgewählten Punkte sind als solche gekennzeichnet.

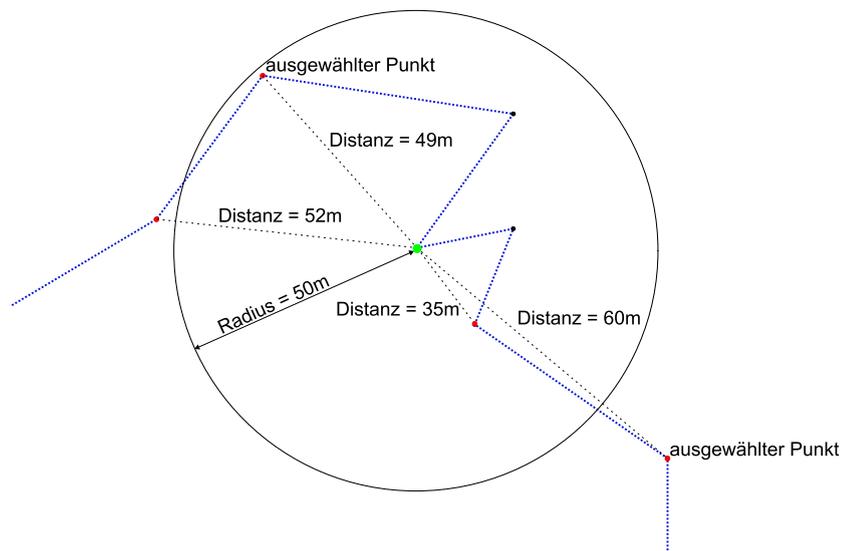


ABBILDUNG 4.1: Implementation des räumlichen Schwellenwertes für die Berechnung der Geschwindigkeit des grün eingefärbten Punktes.

## 4.2 Bewegungsparameter bei der Detektion von Stopps

Wie in Kapitel 2 erläutert gibt es einige Bewegungsparameter, die zur Detektion von Stopps in Trajektorien verwendet werden. Die Auswahl der hier untersuchten Bewegungsparameter wird einerseits auf Grund derer Vorkommen in der Literatur gemacht. Es gibt einige Bewegungsparameter, die immer wieder genannt werden und in den meisten Algorithmen in einer gewissen Form verwendet werden. Tabelle 4.3 gibt einen Überblick über die Arbeiten, in denen die für diese Arbeit ausgewählten Bewegungsparameter ebenfalls verwendet wurden. Andererseits wird bei der Auswahl der Bewegungsparameter darauf geachtet, dass diese aus den Angaben aller Datensätze berechenbar sind. Es werden somit Bewegungsparameter verwendet, die weder vom Kontext noch von der Thematik einzelner Datensätze abhängig sind.

TABELLE 4.3: Verwendung der einzelnen Bewegungsparameter in der Literatur.

Literatur	Geschwindigkeit	Dauer	Räumliche Ausdehnung	Richtungsänderung
Ankerst et al. (1999)			X	
Bak et al. (2012)	X		X	
Buchin et al. (2010)	X		X	X
Buchin et al. (2013)	X	X	X	X
Collins et al. (1995)		X	X	
Laube und Purves (2011)	X			X
Moreno et al. (2010)	X	X		X
Orellana und Wachowicz (2011)	X		X	
Salvucci und Goldberg (2000)	X	X	X	
Schuessler und Axhausen (2009)	X	X	X	
Shamoun-Baranes et al. (2012)	X			
Spaccapietra et al. (2008)		X	X	
van Moorter et al. (2010)			X	
Waga et al. (2012)	X	X	X	X
Zimmermann et al. (2009)	X		X	

Einige Beispiele weiterer, jedoch hier nicht verwendeter Parameter sollen dennoch genannt werden. Bei Trajektoriendaten zu Tieren werden häufig Daten aus Beschleunigungsmessern verwendet. Daraus können beispielsweise Kopfbewegungen oder andere Bewegungsmuster herausgelesen und für die Analyse verwendet werden (van Moorter et al., 2010). Werden beispielsweise im Rahmen einer Untersuchung von Fussgängern in einem Park GPS-Daten aufgenommen, so können Angaben zum Alter oder zur Nutzung des Parks erfasst werden. Neben diesen zusätzlichen Informationen gibt es auch einige Parameter, welche, mit Ausnahme der Eye-Tracking-Daten, bei allen hier verwendeten

Datensätzen vorhanden sind und dennoch nicht verwendet werden. Diese Parameter sind im Folgenden aufgeführt.

**Mögliche, nicht verwendete Parameter** Die Tageszeit ist bei allen Datensätzen vorhanden, wird jedoch nicht verwendet. Um die Tageszeit sinnvoll und adäquat miteinander zu können sind genaue Angaben zum Verhalten des sich bewegenden Objekts zu allen Tageszeiten nötig. Beispielsweise bei den Lastwagendaten wäre der Einbezug der Tageszeit relativ offensichtlich, da die Hauptverkehrszeiten einfach definiert werden können. Bei den Zugvögeln wird es jedoch bereits schwieriger, da sie sowohl am Tag als auch in der Nacht fliegen und ein Stopp meistens mehr als einen Tag dauert (Schaub et al., 2001). Bei den Austernfischern kann davon ausgegangen werden, dass die Zeiten der Nahrungsaufnahme und die Schlafzeiten bekannt sind. Jedoch ist davon auszugehen, dass nicht alle Verhaltensweisen zeitlich klar abgegrenzt werden können und von weiteren externen Angaben abhängen als alleinig von der Tageszeit. Bei den Wanderdaten, die teilweise einen ganzen Tag dauern, kann davon ausgegangen werden, dass mindestens ein längerer Stopp um den Mittag stattfindet. Der mögliche Zeitraum für diesen Stopp besteht über einen gewissen Zeitraum. Auf Wanderungen wird eine Mittagspause meistens nicht genau um zwölf Uhr mittags gemacht, sondern dann, wenn ein passender Rastplatz gefunden wird oder eine bestimmte Strecke zurückgelegt wurde. Da also die Angaben zur Tageszeit im Rahmen dieser Masterarbeit nicht bei allen Datensätzen mit gleicher Genauigkeit und Zuverlässigkeit erarbeitet werden können, wird auf deren Einbezug verzichtet. Für eine weiterführende Arbeit kann die Tageszeit jedoch von Interesse sein.

Die Höhe über Meer kann für gewisse Thematiken ebenfalls interessant sein, so zum Beispiel für Wanderdaten. Es ist davon auszugehen, dass die Geschwindigkeit bei der Überwindung eines grossen Höhenunterschiedes oder in unwegsamem Gelände, wie dies im Gebirge häufig vorkommt, grundsätzlich tiefer ist. Somit können Geschwindigkeits-schwellenwerte bei Wanderdaten unter Einbezug der Höhe und des Geländes angepasst werden. In einer Stadt ist diese Angabe in vielen Fällen hingegen nicht sehr aussagekräftig, da der Höhenunterschied zwischen einzelnen Standorten nicht sehr hoch zu erwarten ist. Für kleinräumige Daten wie beispielsweise die Daten zum Austernfischer scheint die Verwendung der Höhe ebenfalls nicht aussagekräftig zu sein, aus den selbigen Gründen wie oben und weil Vögel eine Höhendifferenz meistens nicht zu Fuss, sondern

im Flug überwinden. Auch bei Zugvögeln ist davon auszugehen, dass Höhendifferenzen fliegend zurückgelegt werden. Weiter legen Zugvögel sehr grosse Strecken zurück, so dass hierbei die Höhe im Verhältnis zur geflogenen Strecke einen geringeren Anteil ausmacht. Im Folgenden sind die Bewegungsparameter beschrieben, die in dieser Arbeit untersucht werden.

### 4.3 Verwendete Bewegungsparameter

**Geschwindigkeit** Die Berücksichtigung der Geschwindigkeit zur Detektion von Stopps in Trajektorien ist intuitiv, da davon ausgegangen werden, dass ein Stopp bei einer tiefen Geschwindigkeit wahrscheinlicher ist als bei einer sehr hohen Geschwindigkeit. Aus diesem Grund macht es Sinn, die Geschwindigkeit bei der Detektion von Stopps miteinzubeziehen. In Situationen mit extremen Geschwindigkeiten kann sogar fast mit Sicherheit gesagt werden, ob ein Stopp oder eine Bewegung stattfindet. So wird äusserst selten eine Bewegung stattgefunden haben, wenn die Geschwindigkeit null ist und es wird äusserst selten ein Stopp gemacht worden sein, wenn die Geschwindigkeit der Maximalgeschwindigkeit des Objekts entspricht. Alle dazwischenliegenden Geschwindigkeiten sind unklar im Bezug darauf, ob ein Stopp in einer Trajektorie stattgefunden hat.

Der Algorithmus der Mindestgeschwindigkeit eliminiert alle Punkte einer Trajektorie, deren Geschwindigkeit tiefer ist als ein spezifischer Schwellenwert. In diesem ersten Schritt werden somit verschiedene Geschwindigkeitsschwellenwerte auf die Datensätze angewendet. Alle Datenpunkte mit einer Geschwindigkeit unter dem jeweiligen Schwellenwert werden als Stopp definiert, die restlichen Datenpunkte werden als Bewegung definiert. Die als Stopps definierten Datenpunkte werden dann jeweils zu Clustern aggregiert. Dabei werden alle zeitlich direkt nacheinander folgenden Punkte als ein Cluster definiert. Es kann somit sein, dass zwischen zwei Clustern lediglich ein einziger Datenpunkt liegt, welcher eine grosse Geschwindigkeit hat, oder dass ein Cluster nur aus einem Datenpunkt besteht. Dies scheint für das Bewegungsverhalten eines Objekts nicht sehr realistisch zu sein, ist jedoch unter einer alleinigen Anwendung eines Geschwindigkeitsschwellenwertes möglich. Da es Ziel dieser Arbeit ist, auch simple Bewegungsparameter zu untersuchen, wird diese Vorgehensweise gewählt.

**Dauer** Ausgehend von den Stopps, welche mit dem Geschwindigkeitsschwellenwert gefunden werden, wird in einem weiteren Schritt die Dauer der gefundenen Stopps mit einbezogen. Dies geschieht mit Hilfe der *Java*-Klasse *Date* (*java.util.Date*). Mit dieser Klasse kann für einen beliebigen Zeitpunkt die Dauer zwischen diesem Zeitpunkt und einem Referenzzeitpunkt, dem 1. Januar 1970 00:00:00 GMT, berechnet werden. Die Zeitspanne wird in Millisekunden seit dem Referenzzeitpunkt angegeben. Es gilt hierbei zu beachten, dass die genannte *Java*-Klasse die Schaltsekunde nicht miteinbezieht. Dies ist für diese Arbeit jedoch sekundär, da lediglich die Zugvögel Trajektorien aufweisen, welche über mehrere Jahre aufgenommen wurden und hierbei das zeitliche Aufnahmeintervall grob genug ist, um die Schaltsekunde zu ignorieren. Bei den Eye-Tracking-Daten ist eine Sekunde von Wichtigkeit, da eine Sekunde im Rahmen von Eye-Tracking-Daten eine relevante Zeitspanne ist. Da diese Daten sich jedoch nicht über mehrere Jahre hinziehen, kann diese Schaltsekunde auch hier ignoriert werden.

Der Algorithmus der Mindestdauer eines Stopps wird auf die durch die Geschwindigkeit definierten Stopps angewandt. Besteht ein solcher Stopp nicht lange genug, das heisst die Zeitspanne zwischen den zwei zeitlich am weitesten entfernten Punkten ist kleiner als der gesetzte Schwellenwert der Mindestdauer, so wird er entfernt. Stopps, die nur aus einem Punkt bestehen, wird eine Dauer von 0 Sekunden hinzugefügt. Dies entspricht nicht ganz der Realität. Beispielsweise bei den Lastwagendaten mit einer zeitlichen Auflösung von 30 Sekunden, müsste ein Stopp aus nur einem Datenpunkt bestehend eine Dauer von 30 Sekunden haben und ein Stopp aus zwei Datenpunkten eine Dauer von 60 Sekunden. Der Einfachheit halber und da davon ausgegangen werden kann, dass Stopps, die aus einem einzigen Datenpunkt bestehen, bereits bei sehr tiefen Schwellenwerten für die Mindestdauer entfernt werden, wird diese Anpassung gemacht.

**Ausdehnung** Auch bei dem Algorithmus der Mindestausdehnung wird von den bereits durch die Geschwindigkeit definierten Stopps ausgegangen. Ähnlich wie bei der Dauer wird die räumliche Distanz zwischen dem ersten und dem letzten Datenpunkt des Stopps, definiert nach zeitlicher Reihenfolge, berechnet. Die Ausdehnung des Stopps wird mit dem Schwellenwert für die Ausdehnung verglichen und Stopps, welche eine zu geringe Ausdehnung aufweisen, werden eliminiert.

Es kann argumentiert werden, dass eine Berechnung der Ausdehnung nicht zwischen

den zeitlich am weitesten voneinander entfernten Punkten erfolgen sollte, sondern zwischen den räumlich am weitesten voneinander entfernten Punkten eines Stopps. Diese Vorgehensweise ist ebenfalls möglich. Es muss jedoch beachtet werden, dass sich eine Trajektorie durch den zeitlichen Ordnungsaspekt erst definiert und sich dadurch von rein räumlichen, also nicht raum-zeitlichen Daten abzeichnet. Unter anderem aus diesem Grund wird in dieser Arbeit jeweils mit den zeitlich am weitesten voneinander entfernten Punkten gearbeitet.

Weiter kommt es im Falle von Trajektorien Daten häufig vor, dass die zeitlich weit voneinander entfernten Datenpunkte auch räumlich weit voneinander entfernt sind, da Trajektorien meistens Linien gestreckter Natur sind. Ist dies nicht der Fall, so kann es sich um eine Ansammlung von Punkten handeln. In diesem Fall sind alle Punkte relativ nahe beieinander, weshalb eine Unterscheidung zwischen dem zeitlich entferntesten und dem räumlich entferntesten Punkt keinen grossen Unterschied ausmacht. Es kann auch sein, dass die Trajektorie die Form einer Haarnadelkurve hat. In dem Fall wäre die alternative Vorgehensweise, also der Einbezug der räumlich grössten Distanz, repräsentativer für die Form der Trajektorie. Das Resultat hingegen würde einen Stopp vermuten, obwohl nicht unbedingt ein Stopp vorhanden ist, da es sich lediglich auf Grund der gemachten Kurve um eine Ansammlung von Punkten handelt.

Abbildung 4.2 verdeutlicht diese verschiedenen Varianten. Die Rot gekennzeichneten Punkte stellen den Start und das Ende des durch die Geschwindigkeit definierten Stopps dar, wobei für dieses Beispiel sekundär ist, in welche Richtung die Trajektorie verläuft. Abbildung 4.2 a) zeigt den einfachsten Fall; hier sind die zeitlich und die räumlich am weitesten voneinander entfernten Punkte dieselben. In diesem Fall führt die Berechnung der Ausdehnung über die zeitliche und die räumliche Distanz zum gleichen Resultat. Abbildung 4.2 b) zeigt einen typischen Fall von einem Cluster, wobei hier alle Punkte nahe beieinander liegen. In diesem Fall sind die räumlich am weitesten voneinander entfernten Punkte des Clusters Grün gekennzeichnet. Es ist ersichtlich, dass zwar ein Unterschied besteht zwischen den zwei in Frage kommenden Distanzen, dieser jedoch im Verhältnis zum ganzen Cluster nicht sehr gross ist. Abbildung 4.2 c) zeigt den Fall einer engen Kurve; auch hier befinden sich alle Punkte des Stopps auf einem engen Raum, wobei jedoch der Anfangs- und der Endpunkt des Stopps nahe beieinander liegen. In diesem Fall gibt es einen Unterschied zwischen den zwei Herangehensweisen, die zwei verschiedenen Distanzen unterscheiden sich. Es scheint jedoch klar, dass hier die direkte Distanz zwischen den zeitlich am weitesten voneinander entfernten Punkten die Situation besser

widerspiegelt. Die räumlich grösste Distanz würde auf einen grossen und ausgedehnten Stopp hindeuten, was nicht realitätsnah ist.

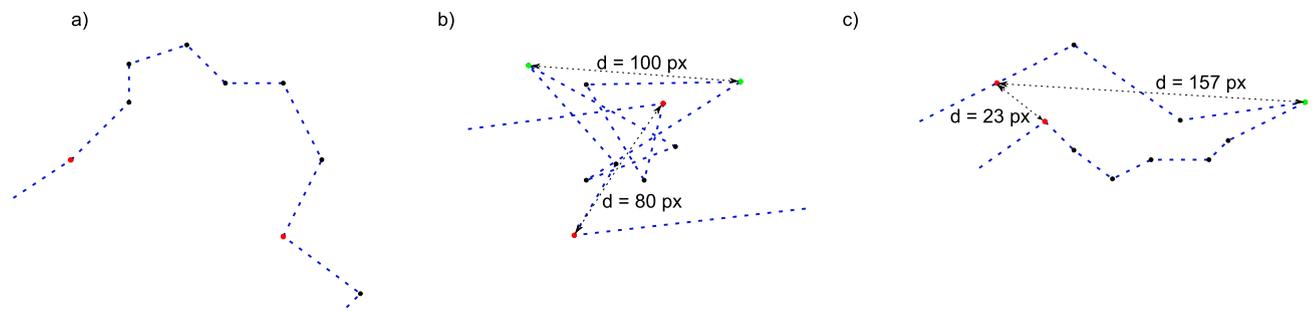


ABBILDUNG 4.2: Auswahl möglicher Ausdrucksformen einer Trajektorie.

**Drehwinkel (Zusatzparameter)** Der Drehwinkel, beziehungsweise die Variation in der Richtungsänderung einer Trajektorie, kann hier für alle Datensätze verwendet werden. Dabei wird davon ausgegangen, dass sich während einer Bewegung die Richtung nicht stark verändert, beziehungsweise der Drehwinkel nicht stark variiert. Während eines Stopps hingegen kann der Drehwinkel innerhalb einer kurzen Zeit stark variieren, weil sich das sich bewegende Objekt auf einem kleineren Raum aufhält und sich nicht grundsätzlich in eine Richtung fortbewegt.

In der allgemeinen Analyse wird der Drehwinkel auf die bereits durch die Geschwindigkeit definierten Stopps angewendet. Es werden alle Punkte, die einen Drehwinkel unter dem Schwellenwert aufweisen eliminiert. Danach werden die zeitlich direkt nacheinander folgenden Punkte neu zu Stopps aggregiert. So ist es möglich, dass neue Stopps entstehen oder auch, dass ganze Stopps eliminiert werden.

Anhand dieser Analysen der allgemeinen Resultate (Abbildungen 5.6 bis 5.10) wird ersichtlich, dass die Resultate stark schwanken. Aus diesem Grund wird bei den Analysen der ausgewählten Trajektorien der Drehwinkel nicht mit einem Geschwindigkeitsschwellenwert, sondern einerseits mit dem Algorithmus der Mindestdauer und andererseits mit dem Algorithmus der Mindestausdehnung kombiniert. Somit stellt der Drehwinkel bei der Analyse der ausgewählten Trajektorien primär einen Hilfsparameter dar, als dass er alleinig verwendet wird. Die Vorgehensweise ist ebenfalls etwas anders, so wird für jeden durch die Geschwindigkeit und die Mindestdauer oder die Mindestausdehnung definierten Stopp der Anteil an Punkten eruiert, die einen Drehwinkel unter dem gesetzten Schwellenwert für den Drehwinkel haben. Ist dieser Anteil grösser, als ein gewisser Schwellenwert (beispielsweise 30 %), so wird der Stopp eliminiert. Für diesen

Algorithmus müssen somit zwei Schwellenwerte gesetzt werden. Einerseits der minimale Drehwinkel und andererseits der Anteil an Punkten eines Stopps, die einen Drehwinkel aufweisen dürfen, der tiefer ist als dieser Schwellenwert.

Abbildung 4.3 zeigt, wie der Drehwinkel aufzufassen ist. Die Nullrichtung wird definiert durch den den zu untersuchenden Punkt  $p_i$  und den vorangegangenen Punkt  $p_{i-1}$ . Der Drehwinkel des Punktes  $p_i$  wird dann eruiert anhand des nächsten Punktes  $p_{i+1}$ . Der Drehwinkel ist definiert durch den Winkel zwischen der Nullrichtung und der Verbindungslinie von  $p_i$  zu  $p_{i+1}$ . Hierbei spielt die Richtung, in welche die Trajektorie weiter verläuft keine Rolle, da lediglich der Drehwinkel interessiert und nicht die Richtung (vereinfacht gesagt rechts oder links), in die sich das Objekt gedreht hat. Aus diesem Grund ist der maximale Drehwinkel 180 Grad und nicht 360 Grad. Der Schwellenwert ist als Winkel ausgehend von der Nullrichtung zu verstehen und nicht als gesamter Winkel, der über die Nullrichtung läuft. Wird beispielsweise von einem Schwellenwert von 45 Grad gesprochen, so sind damit 45 Grad je Seite gemeint. Mit einem Schwellenwert von 45 Grad sind somit Punkte in einem Bereich von 90 Grad betroffen.

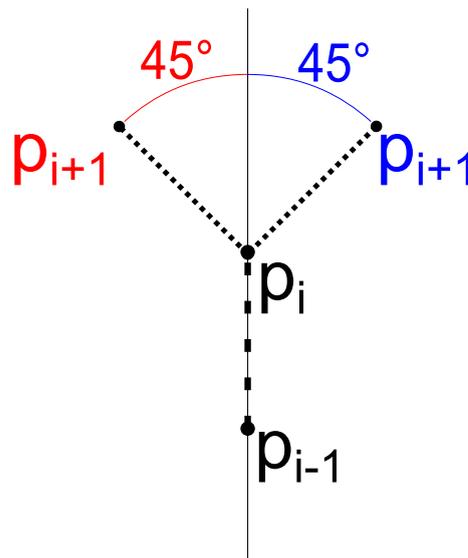


ABBILDUNG 4.3: Drehwinkel des Punktes  $p_i$  einer Trajektorie. Der Drehwinkel beträgt 45 Grad sowohl mit dem roten als auch mit dem blauen Punkt als  $p_{i+1}$ .

Ein Beispiel, um den Algorithmus des Drehwinkels zu verdeutlichen, sind Zugvögel. Abbildung 4.4 zeigt, wie eine mögliche Trajektorie eines Zugvogels aussehen könnte. Es ist ersichtlich, dass während der ganzen Trajektorie stets ungefähr die gleiche Richtung beibehalten wird, entweder fliegt der Vogel nach Süden oder nach Norden (Abbildung 4.4 a)). Die Richtungen Osten und Westen werden selten eingeschlagen. Wird nun eine Stelle, an der ein Stopp angenommen wird, genauer betrachtet, so wird ersichtlich, dass

die Richtungsänderungen zwischen den einzelnen Punkten stark variieren (Abbildung 4.4 b)). Die spitzen Winkel, das heisst die Punkte mit einem grossen Drehwinkel, sind markiert, um den Sachverhalt zu verdeutlichen.



ABBILDUNG 4.4: Drehwinkel über a) die gesamte Trajektorie eines Zugvogels und b) über einen kürzeren Zeitraum und kleineren Raumausschnitt. Quelle Basiskarte: Google Earth, Image Landsat, Data SIO, NOAA, U.S. Navy, NGA, GEBCO.

Die Richtungsänderung beziehungsweise der Drehwinkel als alleiniger Parameter wird mehrheitlich bei Objekten verwendet, welche die Möglichkeit haben, stets einer direkten Linie zu folgen und nur wenn nötig Kurven zu machen. Dies erläutern auch Rocha

et al. (2010); sie verwenden die Richtungsänderung im Zusammenhang mit Schiffen, bei welchen die Richtung ein wichtiges Bewegungsmerkmal ist. Bei den meisten Flugobjekten ist dies auch der Fall, da diese sehr wenige Hindernisse haben, welche sie zu Kurven zwingen könnten. Wandernde Personen bewegen sich zwar ebenfalls frei, wobei natürliche Objekte wie Pflanzen, Gesteine oder die Topographie eine grosse Rolle bei der Wahl des Wegs spielen. Im Rahmen dieser Arbeit sind die Lastwagendaten in ihren möglichen Drehwinkeln am stärksten eingeschränkt, da sie einem Netzwerk folgen müssen, durch andere Fahrzeuge limitiert werden und Verkehrsschilder beachten müssen.

#### 4.4 Algorithmengruppe Stops and Moves of Trajectories

Die drei Algorithmen der Gruppe SMoT, der SMoT (Alvares et al., 2007), der CB-SMoT (Palma et al., 2008) und der DB-SMoT (Rocha et al., 2010), werden ausgewählt, weil sie untereinander vergleichbar sind und deren spezifische Stärken und Schwächen bekannt sind. Da das primäre Ziel aller drei Algorithmen die Detektion von Stopps in Trajektorien ist, eignen sie sich alle im Kontext dieser Arbeit. Weiter macht der Einbezug der Kandidatenregionen den Algorithmus SMoT interessant, da Kontextinformationen weder bei den Bewegungsparametern noch bei dem HMM in Betracht gezogen werden. Somit stellen diese Algorithmen eine sinnvolle Ergänzung zu den anderen Methoden dar. Die drei Algorithmen stellen jeweils eine Weiterentwicklung oder Ergänzung eines bestehenden, vorangegangenen Algorithmus dar. Die detaillierten *pseudocodes* dieser Algorithmen sind in den jeweiligen Arbeiten enthalten.

**Stops and Moves of Trajectories (SMoT)** Der Algorithmus SMoT (Alvares et al., 2007) arbeitet mit Kandidatenregionen und der Verschneidung der Trajektorie mit diesen Regionen. Kandidatenregionen sind geographisch abgegrenzte Polygone, innerhalb jener Stopps zu erwarten sind. Somit werden bei diesem Algorithmus nicht nur die Trajektorien selbst, sondern auch Daten zu Kandidatenregionen verwendet.

Der Algorithmus funktioniert folgendermassen. Es wird konsekutiv durch die Punkte einer Trajektorie iteriert. Dabei wird für jeden Punkt kontrolliert, ob dieser in einer Kandidatenregion liegt. Ist dies der Fall, so wird der Zeitstempel dieses Punktes als Startzeit zwischengespeichert. Nun wird durch die nachfolgenden Punkte iteriert, wobei für jeden einzelnen davon ebenfalls kontrolliert wird, ob er in der Kandidatenregion liegt.

Sobald ein Punkt nicht mehr in der Kandidatenregion liegt, wird der Zeitstempel des unmittelbar vorherigen Punktes, also des letzten Punktes innerhalb der Kandidatenregion, als Endzeit zwischengespeichert. Nun wird die Differenz zwischen der Start- und Endzeit berechnet und falls diese Zeitspanne grösser als ein Schwellenwert ist, wird ein Stopp registriert. Falls die Zeitspanne zu klein ist, wird weiter durch die Punkte iteriert (Alvares et al., 2007). Abbildung 4.5 zeigt die Funktionsweise des Algorithmus SMoT auf. Die grauen Vierecke stellen Kandidatenregionen dar, die Trajektorie ist durch die Linie und die darauf enthaltenen Punkte dargestellt. Die Überschneidung der Trajektorie mit dem ersten Viereck beginnt mit dem Zeitstempel des Punktes  $p_1$  und endet mit dem Zeitstempel des Punktes  $p_6$ . Ist die Zeitdifferenz zwischen diesen zwei Punkten gross genug, so wird ein Stopp der Trajektorie in diesem ersten Viereck registriert.

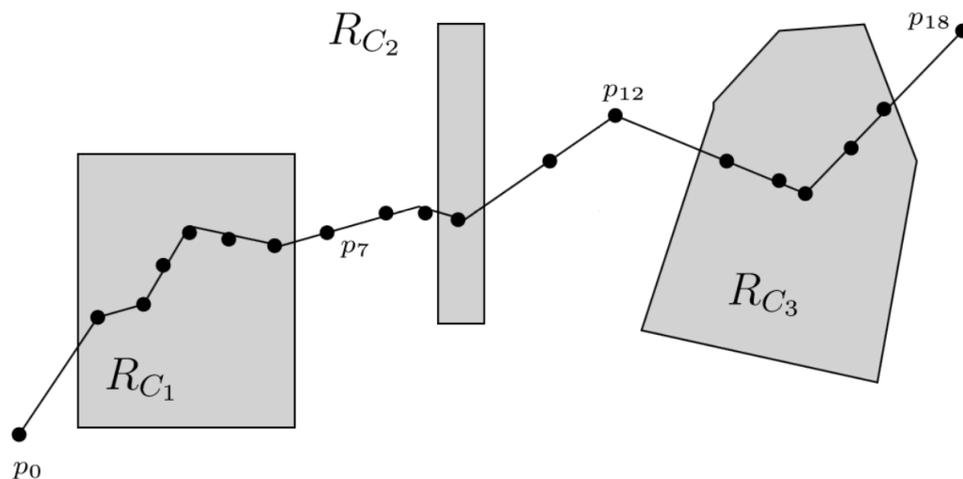


ABBILDUNG 4.5: Beispiel einer Trajektorie und drei dazugehörigen Kandidatenregionen ( $R_{C_1}$  bis  $R_{C_3}$ ). Quelle: Alvares et al. (2007): 4.

**Clustering-Based Stops and Moves of Trajectories (CB-SMoT)** Der CB-SMoT (Palma et al., 2008) stellt insofern eine Weiterentwicklung des SMoT dar, als dass nicht nur Stopps innerhalb der vorgegebenen Kandidatenregionen gesucht werden, sondern auch Stopps ausserhalb dieser Regionen, sogenannte unbekannte Stopps. Der Algorithmus orientiert sich an DBSCAN (siehe dazu Ester et al. (1996)), einem dichte-basierten Algorithmus zur Detektion von Clustern. Neben der Nachbarschaft arbeitet dieser Algorithmus mit einer minimalen Aufenthaltsdauer, wobei diese Kombination aus Nachbarschaft und Aufenthaltsdauer eine Geschwindigkeit impliziert. Somit ist die Grundlage des CB-SMoT die Geschwindigkeit (Palma et al., 2008). In einem ersten Schritt werden alle Regionen mit einer tiefen Geschwindigkeit gesucht, erst in einem zweiten

Schritt wird getestet, welche dieser gefundenen Cluster sich in Kandidatenregionen befinden. Die Cluster werden dann in echte, sich in den Kandidatenregionen befindliche, und unbekannte, sich ausserhalb der Kandidatenregionen befindliche, Stopps unterteilt. Abbildung 4.6 illustriert die Vorgehensweise des CB-SMoT.  $R_{C1}$  bis  $R_{C4}$  stellen die Kandidatenregionen dar,  $G_1$  bis  $G_4$  stellen die gefundenen Cluster dar. In diesem Beispiel sind zwei echte Stopps  $G_1$  und  $G_2$  gefunden worden, sowie zwei unbekannte Stopps  $G_3$  und  $G_4$ .

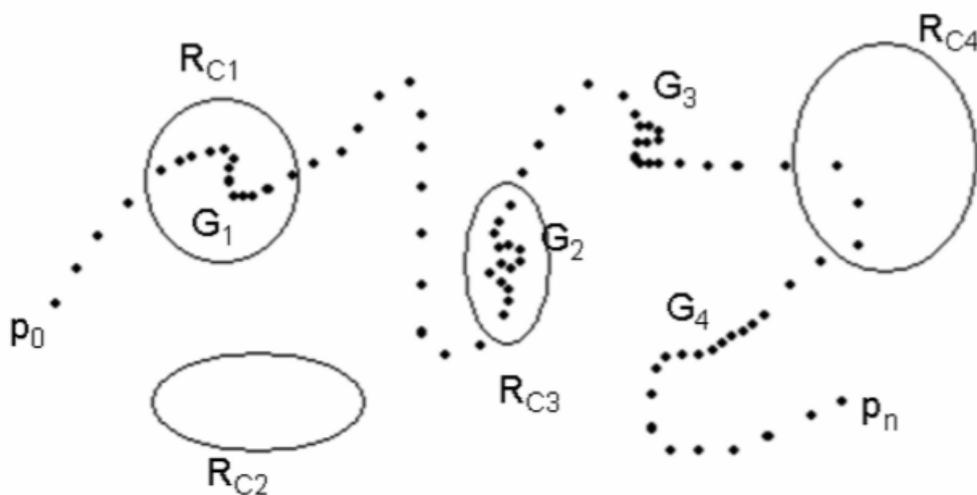


ABBILDUNG 4.6: Beispiel einer Trajektorie mit zwei echten und zwei unbekannten Stopps laut CB-SMoT. Quelle: Palma et al. (2008): 866.

**Direction-Based Stops and Moves of Trajectories (DB-SMoT)** Der DB-SMoT (Rocha et al., 2010) stellt einen weiteren Algorithmus der Gruppe SMoT dar. Der DB-SMoT basiert auf der Variation der Richtungsangabe der einzelnen Datenpunkte und berücksichtigt die Geschwindigkeit nicht. Grundsätzlich wird davon ausgegangen, dass Stopps bei Punkten vorkommen, bei denen eine grosse Richtungsänderung stattfindet. Dieser Algorithmus definiert eine minimale Richtungsänderung beziehungsweise einen minimalen Drehwinkel, damit ein Punkt zu einem potentiellen Cluster gehört. Es müssen jedoch nicht alle Punkte eines Clusters einen grossen Drehwinkel aufweisen. Es wird spezifiziert, wie viele nacheinander folgende Punkte in einem Cluster maximal eine kleine Richtungsänderung aufweisen dürfen. Weiter muss ein Cluster über eine gewisse Zeit bestehen um als Stopp angesehen zu werden (Rocha et al., 2010).

## 4.5 Hidden Markov Models

Die Idee, die Theorie der HMM im Rahmen dieser Arbeit für Bewegungsdaten zu verwenden, stammt aus Salvucci (1999) und Salvucci und Goldberg (2000), welche diese Theorie für die Analyse von Eye-Tracking-Daten verwendet haben. Kiefer und Giannopoulos (2012) verwenden HMM auch im Zusammenhang mit Eye-Tracking-Daten, wobei sie versuchen, bei der Interpretation dieser Daten die geographischen Gegebenheiten der unterliegenden Karte miteinzubeziehen. Auch Waga et al. (2012) verwenden ein HMM, jedoch zur Unterscheidung mehrerer Arten der Fortbewegung.

Ein HMM arbeitet lediglich mit verschiedenen Wahrscheinlichkeitswerten, wobei die räumlichen Informationen zur Analyse der Trajektorie nicht direkt berücksichtigt werden, die zeitliche Komponente spielt jedoch eine wichtige Rolle. HMM können für eine Vielzahl von Anwendungen nützlich sein, so wurden sie beispielsweise erfolgreich für die Sprach- und Handschrifterkennung eingesetzt (Salvucci und Goldberg, 2000). Die Entscheidung, HMM im Rahmen dieser Arbeit miteinzubeziehen wird aus folgenden Gründen gefällt. Zum einen stellt diese Methode einen sehr unterschiedlichen Ansatz im Vergleich zu den anderen hier untersuchten Verfahren dar. Zum anderen hat sich diese Methode in der Arbeit von Salvucci und Goldberg (2000) bei der Detektion von Fixationen in Eye-Tracking-Daten bewährt, wie Tabelle 4.4 zeigt. Diese Tabelle zeigt ebenfalls, dass die Implementation komplex und schwierig sein kann. Ich habe mich eines Pakets<sup>2</sup> bedient, welches HMM und all die dazugehörigen Algorithmen vollständig implementiert hat, weshalb das Argument der komplexen Implementation wegfällt. Da HMM die Räumlichkeit der Daten nicht miteinbeziehen, spielen räumliche Messungenauigkeiten in den Datensätzen hierbei keine Rolle. Dies ist ein weiteres Argument für die Wahl dieser Methode.

---

<sup>2</sup> Link zur Implementation der Hidden Markov Model in *java* (Stand 18.03.2014): <https://code.google.com/p/jahmm/downloads/list>

TABELLE 4.4: Übersicht über verschiedene Methoden zur Detektion von Fixationen.  
Eigene Darstellung nach Salvucci und Goldberg (2000): 76.

Method	Accuracy	Speed	Robustness	Impl. Ease	Parameters
Velocity Threshold (I-VT)	good	very good	not as good	very good	1
Hidden Markov Model (I-HMM)	very good	good	very good	good	8/0
Dispersion Threshold (I-DT)	very good	good	very good	good	2
Minimum Spanning Tree (I-MST)	good	not as good	very good	not as good	2
Area-of-interest (I-AOI)	not as good	good	good	good	1

**Theorie des Hidden Markov Models** Rabiner (1989) hat umfassende theoretische Grundlagen zum HMM sowie dessen Anwendung in der Spracherkennung bereitgestellt. An dieser Stelle soll die Idee hinter diesen Modellen erläutert werden. In der Natur kommt es häufig vor, dass gewisse Prozesse beobachtbare Signale produzieren. So ist beispielsweise die Umgebungstemperatur ein Signal aus verschiedenen zusammenwirkenden Prozessen in der Atmosphäre. Oft ist der Prozess nicht genau bekannt und es muss rein aus einer Abfolge von Beobachtungen auf den dahinterliegenden, versteckten Prozess geschlossen werden. Hierzu eignet sich ein HMM, da mit Hilfe dieses Modells anhand von Wahrscheinlichkeiten und einer Abfolge von Beobachtung auf die versteckten, dahinter liegenden Zustände geschlossen werden kann. Das Modell besteht aus folgenden Elementen (Rabiner, 1989):

- Menge an Zuständen im Modell
- Menge an Beobachtungen je Zustand
- Übergangsmatrix (*state transition probability distribution*)
- Beobachtungsmatrix (*observation symbol probability distribution*)
- Anfangsverteilung der Zustände

Ein HMM selbst beinhaltet drei verschiedene Algorithmen, welche alle zusammen die Mächtigkeit dieses Modells ausmachen. Ein Algorithmus heisst *Viterbi* und findet anhand einer Beobachtungssequenz die wahrscheinlichste dahinterliegende Zustandssequenz. Dieser Vorgang wird Decodierung genannt. In einem zweiten Schritt geschieht die Evaluierung anhand des Algorithmus namens *forward-backward*. Hierbei wird das durch den

Algorithmus *Viterbi* erstellte HMM und dessen Richtigkeit evaluiert. Als letzter wird der Algorithmus *Baum-Welch* zur erneuten Einschätzung der Parameter des HMM herbeigezogen, wobei versucht wird, die Wahrscheinlichkeit der Abfolge der Zustände zu erhöhen (Rabiner, 1989).

**Beispiel eines Hidden Markov Models** Ein einfaches Beispiel dazu ist folgendes: Jemand sitzt den ganzen Morgen in einem Büro ohne Fenster und würde gerne wissen, ob es draussen regnet oder nicht. Er erhält jeden Morgen die Zeitung vom Postboten und sieht, ob dessen Jacke nass ist oder nicht. Das aktuelle Wetter ist der versteckte Zustand und die Beobachtung wird anhand der Jacke gemacht. Nun weiss die betreffende Person, dass die Wahrscheinlichkeit für Regen 80% ist bei einer nassen Jacke und 20% bei einer trockenen Jacke. Weiter weiss sie, dass die Wahrscheinlichkeit für trockenes Wetter 60% ist bei einer trockenen Jacke und 40% bei einer nassen Jacke (bei starkem Regen kommt der Postbote mit dem Auto und wird daher nicht nass). Die betreffende Person weiss ebenfalls, dass nach einem Regentag mit einer Wahrscheinlichkeit von 70% nochmals ein Regentag folgt und mit 30% ein trockener Tag. Nach einem trockenen Tag folgt mit einer Wahrscheinlichkeit von 50% nochmals ein trockener Tag und mit einer Wahrscheinlichkeit von 50% ein Regentag. Der erste Tag der Woche, also die erste Beobachtung ist zu 65% ein Regentag und zu 35% ein trockener Tag. Anhand dieser Wahrscheinlichkeiten und der gemachten Beobachtungen kann die betreffende Person nun mit Hilfe eines HMM eruieren, an welchen Wochentagen es am Morgen wahrscheinlich regnet und an welchen nicht. Dazu dienen die verschiedenen Algorithmen, welche im HMM integriert sind und eine ideal zu der Beobachtungssequenz passende Abfolge von Zuständen suchen. Abbildung 4.7 stellt das Beispiel schematisch dar.

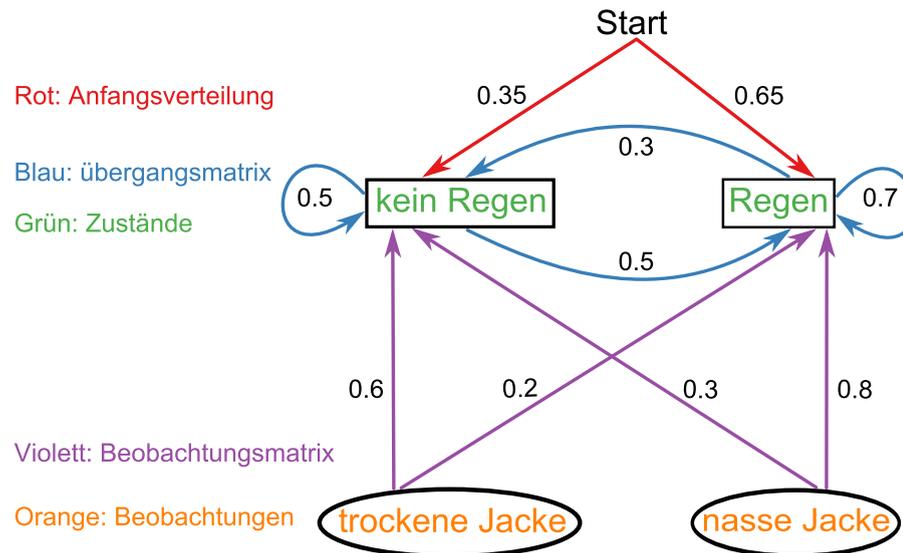


ABBILDUNG 4.7: Beispiel eines Hidden Markov Models.

#### 4.5.1 Anwendung bei der Detektion von Stopps in Trajektorien

Trajektorien- und Bewegungsdaten können in Stopps und Bewegungen unterteilt werden (Spaccapetra et al., 2008). Es kann bei der Analyse von Trajektorien- und Bewegungsdaten davon ausgegangen werden, dass eine Abfolge von häufig und regelmäßig abwechselnden Stopps und Bewegungen unwahrscheinlich ist (siehe auch Salvucci und Goldberg (2000), wobei Übergangswahrscheinlichkeiten von 0.05 verwendet werden). Befindet sich ein Objekt in Bewegung, so wird es wahrscheinlich nicht sofort wieder stoppen. Und umgekehrt, legt das Objekt einen Stopp ein, so wird es für einen gewissen Moment gestoppt bleiben und sich nicht sofort weiter fortbewegen.

Diese Idee kann am Beispiel eines Lastwagens verdeutlicht werden. So beginnt beispielsweise die Trajektorie des Lastwagens mit einer Bewegung, da das Aufnahmegerät des GPS-Signals eingeschaltet wird, wenn der Lastwagen losfährt. Nun befindet sich der Lastwagen in Bewegung, bis er seinen Zielort erreicht hat. Es kann davon ausgegangen werden, dass dieser Transport eine gewisse Zeit in Anspruch nimmt, während der der Lastwagen in Bewegung ist. Hat der Lastwagen die Zielbaustelle erreicht, so findet ein Stopp statt. Dieser Stopp dauert nicht nur einen kurzen Moment, da das Manövrieren zu der Abladestelle, das Abladen und die Vorbereitungen für die Weiterfahrt Zeit in Anspruch nehmen. Wird die Weiterfahrt angetreten, so dauert diese eine gewisse Zeit, nämlich bis der Lastwagen das Depot wieder erreicht hat. Hier wiederum kann mit einer gewissen Zeit gerechnet werden bis neuer Beton aufgeladen ist. Es kann also gesagt

werden, dass die Wahrscheinlichkeit für einen Zustandswechsel gering ist und dass die Wahrscheinlichkeit, über längere Zeit in einem Zustand zu bleiben gross ist. Diese Informationen werden in das HMM miteinbezogen.

Weiter gilt es, die Beobachtungen und die Beobachtungswahrscheinlichkeiten zu definieren. Im Rahmen von Bewegungsdaten können zwei Zustände unterschieden werden, die Stopps und die dazwischenliegenden Bewegungen. Als Beobachtung kann die gemessene Geschwindigkeit miteinbezogen werden. Somit werden die Datensätze, um für ein HMM benutzbar zu sein, in eine Abfolge von hohen und tiefen Geschwindigkeiten unterteilt. Die Beobachtungswahrscheinlichkeiten ergeben sich intuitiv, so ist die Beobachtungswahrscheinlichkeit für einen Stopp bei tieferer Geschwindigkeit grösser und die Beobachtungswahrscheinlichkeit für eine Bewegung bei höherer Geschwindigkeit grösser (siehe auch Salvucci und Goldberg (2000)). Basierend auf den verschiedenen Wahrscheinlichkeiten und der Beobachtungssequenz kann nun eruiert werden, welche Abfolge von Zuständen am wahrscheinlichsten ist.

**Auslegung des Hidden Markov Model im Rahmen dieser Arbeit** Die Parameter eines HMM bei der Detektion von Stopps in Trajektorien sind in Abbildung 4.8 dargestellt. Die genauen Werte für die Übergangs- und die Beobachtungsmatrix werden anhand eines Datensatzes gelernt und verbessert. Aus diesem Grund können in Abbildung 4.8 keine genauen Werte eingefügt werden. Eine Einschätzung dieser ist abgebildet.

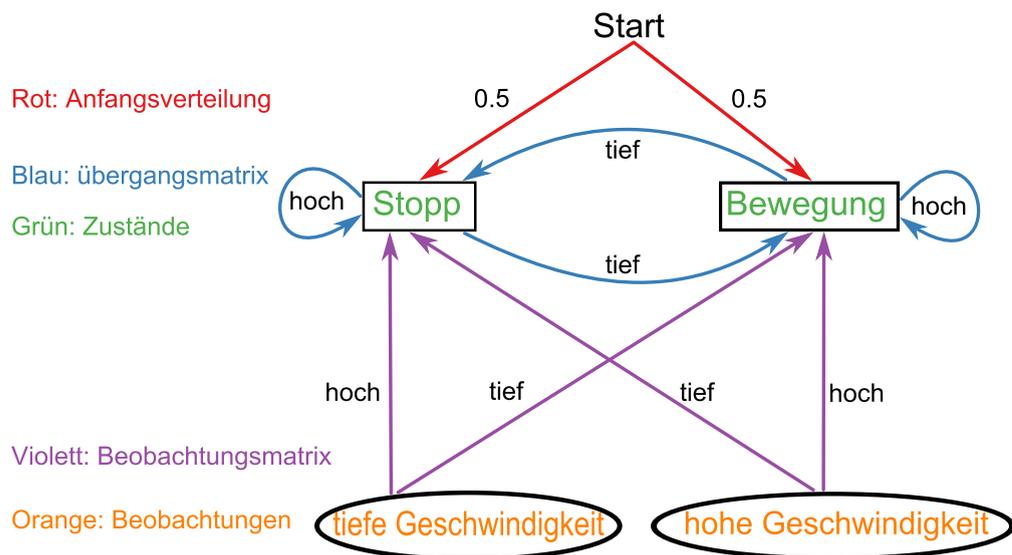


ABBILDUNG 4.8: Ein Hidden Markov Model bei der Detektion von Stopps in Trajektorien. Die Werte für die Übergangs- und Beobachtungsmatrix werden anhand eines Datensatzes erlernt, sind demnach variabel und deshalb nicht eingefügt. Für eine alternative Darstellung siehe Salvucci und Goldberg (2000): 73.

Die Geschwindigkeit stellt im Gegensatz zum obigen Beispiel mit dem Wetter eine kontinuierliche Grösse dar und muss nicht, kann aber, in zwei diskrete Beobachtungen, beispielsweise schnell und langsam, unterteilt werden (streng genommen ist auch die Jacke des Postboten im Beispiel aus Abbildung 4.7 nicht immer klar nass oder trocken, so kann sie beispielsweise auch feucht sein). Das Modell muss für die Analyse von sowohl kontinuierlichen als auch diskreten Daten jeweilig angepasst werden (Rabiner, 1989).

Ich habe mich entschieden, die Geschwindigkeitswerte in zwei Gruppen zu unterteilen, obwohl dies einen Informationsverlust bedeutet. Die Reduktion von vielen Zuständen, wie dies in den kontinuierlichen Daten der Geschwindigkeitswerte der Fall ist, auf einige wenige, reduziert den Informationsgehalt der Daten, vereinfacht jedoch auch das Verständnis für das Modell. Um die kontinuierlichen Daten miteinbeziehen zu können, hätte für jeden Datensatz die Verteilung berücksichtigt werden müssen und somit hätten auch alle Datensätze individuell interpretiert werden müssen. Weiter war es nicht von Interesse die gesamte Komplexität und Wirkungsweise der HMM auszuschöpfen. Bereits der hier verwendete Ansatz, wenn auch mit vereinfachten Daten, stellt im Rahmen dieser Arbeit einen grossen Mehrwert dar. Da sich eine solche Herangehensweise fundamental unterscheidet von den anderen hier verwendeten Methoden, ist das Resultat auch in dieser Form wichtig.

**Auswahl der Implementation des Hidden Markov Models** Eine vollständige Implementation eines HMM zu realisieren bedeutet einen grossen Aufwand und das Verständnis für alle darin ablaufenden Prozesse, die sehr komplex sein können, zu erlangen ebenfalls. Im Rahmen dieser Arbeit wird eine bereits existierende Implementation<sup>3</sup> verwendet. Die genannte Implementation stellt einige Funktionen zur Verfügung, so können ein HMM und dessen Parameter anhand eines Datensatzes eruiert werden. Dabei wird ein leeres HMM sowie ein *Baum-Welch-Learner* erstellt. Der *Baum-Welch-Learner* iteriert mehrere Male über das Modell, wobei das Modell mit jeder Iteration verbessert wird. Nach einigen Versuchen mit jeweils einer unterschiedlichen Anzahl Iterationen, habe ich festgestellt, dass das Ergebnis ab zehn Iterationen nicht mehr stark ändert, weshalb in meinen Versuchen zehn Mal iteriert wird. Ich habe mich entschieden, das HMM vor der Anwendung nicht zu definieren, sondern mit Hilfe des *Baum-Welch-Learners* die idealen Parameter für das Modell ausfindig zu machen. Diese Variante wird

---

<sup>3</sup> Link zur Implementation der Hidden Markov Model in *java* (Stand 18.03.2014): <https://code.google.com/p/jahmm/downloads/list>

bevorzugt, da alle Datensätze unterschiedliche Charakteristika aufweisen. So kann das Lernen des Modells anhand der spezifischen Datensätze ausgenutzt werden. Wird das HMM für alle Datensätze neu definiert, so kann davon ausgegangen werden, dass die verwendeten Parameter ideal auf den jeweiligen Datensatz passen.

# Kapitel 5

## Resultate

Einerseits wurden mit jeder Methode allgemeine und teilweise über mehrere Datensätze aggregierte Resultate generiert, die in Diagrammen aufgeführt sind. Andererseits konnte für die Wander-, Lastwagen- und Austernfischerdaten eine qualitative und visuell verifizierbare Analyse einer ausgewählten Trajektorie gemacht werden. Die dabei erhaltenen Resultate sind nachfolgend in Form von Karten präsentiert. Eine Zusammenfassung der Resultate befindet sich im Anhang A.

Die Resultate zeigen gut, dass es durchaus Unterschiede zwischen den Methoden und auch deren Anwendung auf verschiedene Datensätze gibt. Somit sind die Methoden sensibel auf verschiedene Datensätze und scheinen nicht in jedem Fall Resultate gleicher Charakteristika zu liefern. Bei der Analyse der erhaltenen Resultate sollte berücksichtigt werden, dass die einzelnen Datensätze einige wichtige unterschiedliche Eigenschaften aufweisen.

So sind die Wander- und Austernfischerdaten zeitlich sehr hoch aufgelöst, zudem können sich dabei der Wanderer als auch der Austernfischer frei bewegen. Die Wanderdaten sind auch räumlich sehr genau und weisen einen kleinen Fehler des GPS-Signals auf. Die Lastwagendaten hingegen bewegen sich auf einem Netzwerk und zudem in einer Stadt, was beides zu Einschränkungen in der Bewegungsfreiheit der Lastwagen führt. Die Eye-Tracking-Daten dürfen nur mit Vorsicht mit den anderen Datensätzen verglichen werden, da es sich bei den hier gemachten Bewegungen nicht um Bewegungen im physikalischen Sinne handelt, auf Grund einer verschwindend kleinen sich bewegenden Masse.

Im Rahmen einiger Datensätze ist bekannt, wie viele Stopps tatsächlich stattgefunden

haben. In der Analyse der ausgewählten Trajektorien sind die tatsächlichen Stopps visualisiert. In der allgemeinen Analyse wurde die Anzahl an Stopps über mehrere Datensätze gesucht und auch über mehrere Datensätze gemittelt. In den Trajektorien des Wanderers wurden demnach im Schnitt zwei Stopps gemacht. Für die Analyse des Austernfischers wurden drei Datensätze verwendet, wobei lediglich für zwei davon die *ground truth* vorhanden ist. Anhand dieser Angaben konnte die Anzahl an Stopps je Datensatz abgeschätzt werden, wobei die Grösse des Datensatzes miteinbezogen wurde. Somit sind für die Austernfischerdaten rund neun Stopps zu erwarten. Zu den Eye-Tracking-Daten bestehen Angaben zur Auswertung eines Programms, welches die Fixationen in diesen Trajektorien erkennt. Laut diesen Angaben sollten in den Eye-Tracking-Daten 1160 Fixationen zu finden sein. Für die Lastwagendaten und die Truthahngeierdaten kann auf Grund fehlender *ground truth* keine Angabe zu der erwarteten Anzahl an Stopps gemacht werden.

## 5.1 Resultate der Bewegungsparameter

Für die allgemeine Analyse wurden sehr viele Resultate generiert, da für jede Kombination von Datensätzen und Algorithmen eine separate Tabelle mit den Resultaten erstellt wurde. Insgesamt wurden fast fünfzig Experimente durchgeführt. Die Resultate sind zur einfacheren Veranschaulichung in Form von Diagrammen aufgeführt.

**Geschwindigkeit** Das alleinige Setzen eines Geschwindigkeitsschwellenwertes ohne Einbezug weiterer Bewegungsparameter, wie in Abschnitt 4.3 erläutert, ist in den folgenden Diagrammen abgebildet. Anhand der Diagramme 5.1 bis 5.5 ist ersichtlich, dass alle Datensätze ähnlich auf diesen Algorithmus reagieren. So weisen alle Kurven ein Maximum auf, danach nimmt die Anzahl an Stopps wieder ab. Dieses Maximum wird bei allen Datensätzen mit Ausnahme der Lastwagendaten relativ schnell erreicht. Anhand der prozentualen Anteile ist ersichtlich, dass bei diesem Maximum in allen Fällen über 80% der Punkte als Stopp klassifiziert sind. Die Kurven der grob aufgelösten Truthahngeierdaten und der sehr fein aufgelösten Eye-Tracking-Daten weisen um einiges glattere Kurven auf, als die anderen Diagramme. Interessant ist, dass die blaue Kurve (Anzahl Stopps) zwar in allen Diagrammen einen ähnlichen Verlauf hat, die rote Kurve (prozentualer Anteil an Stopppunkten) hingegen verschiedene Ausprägungen zeigt. Die rote

Kurve erreicht nie 100 %, da es immer noch Punkte gibt, die über dem höchsten abgebildeten Schwellenwert für die Geschwindigkeit liegen. Die Diagramme hätten ausgedehnter dargestellt werden können, dann hätte die rote Kurve die 100 % erreicht. Darauf wurde verzichtet, damit die Bereiche mit tieferen Geschwindigkeitswerten grösser dargestellt werden konnten, da vor allem diese Bereiche im Rahmen dieser Arbeit wichtig und interessant sind.

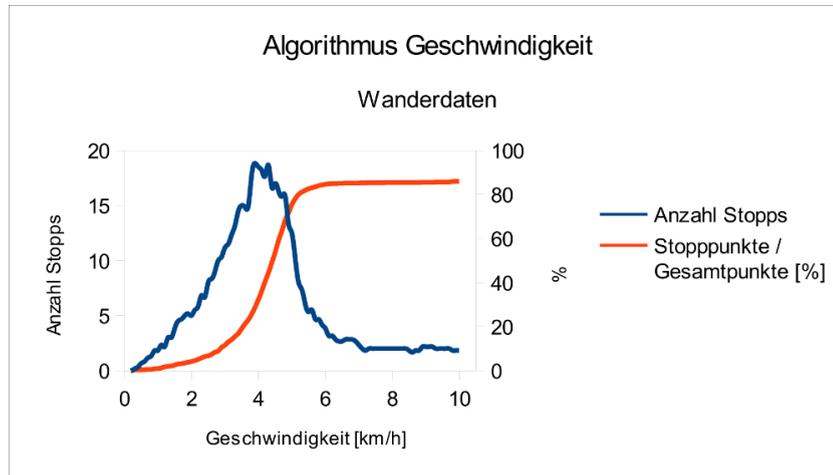


ABBILDUNG 5.1: Geschwindigkeitsschwellenwert angewendet auf Wanderdaten.

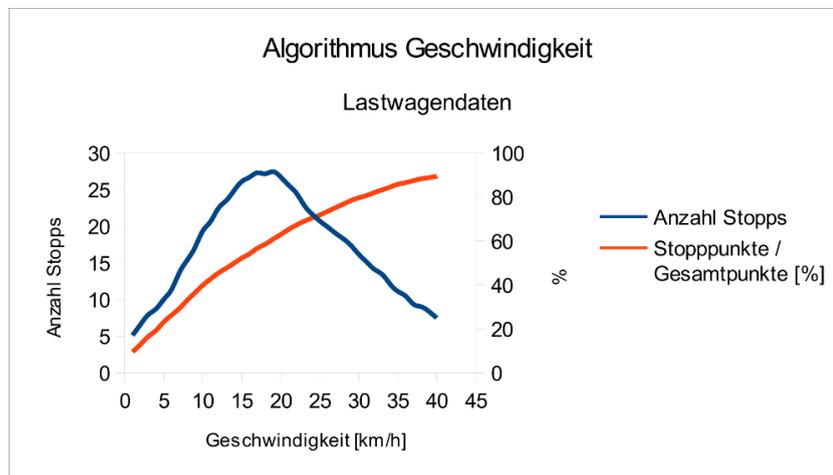


ABBILDUNG 5.2: Geschwindigkeitsschwellenwert angewendet auf Lastwagendaten.

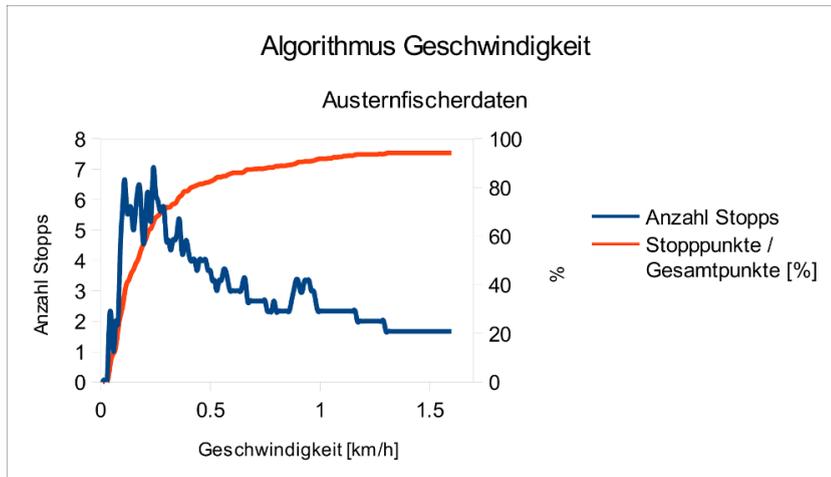


ABBILDUNG 5.3: Geschwindigkeitsschwellenwert angewendet auf Austernfischerdaten.

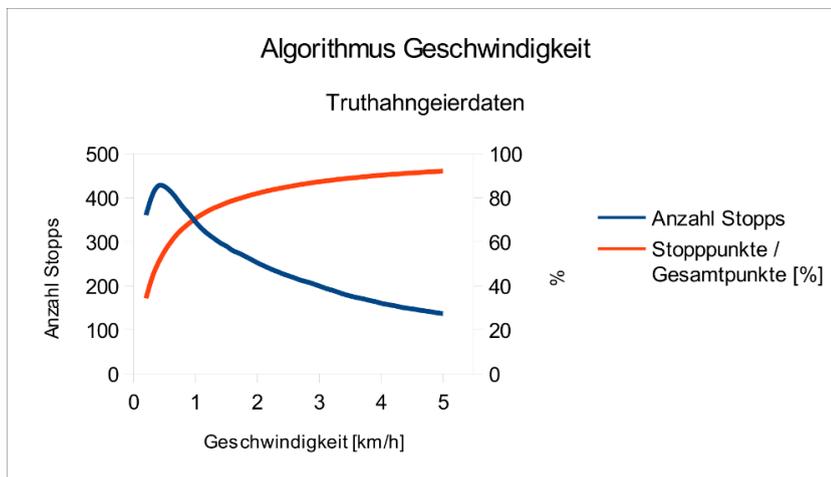


ABBILDUNG 5.4: Geschwindigkeitsschwellenwert angewendet auf Truthahngeierdaten.

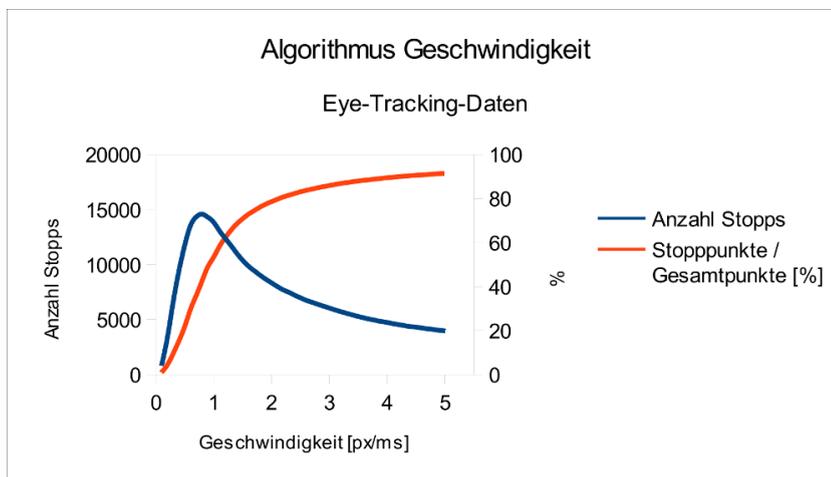


ABBILDUNG 5.5: Geschwindigkeitsschwellenwert angewendet auf Eye-Tracking-Daten.

**Drehwinkel** Im Folgenden sind die Resultate dargestellt, die unter Anwendung eines Geschwindigkeitsschwellenwertes und eines minimalen Drehwinkels der einzelnen Punkte resultieren. Dabei wurde als Schwellenwert für den Drehwinkel der Wert  $+ / - 20$  Grad verwendet. Somit wurden Punkte mit einem Drehwinkel innerhalb eines Bereichs von 40 Grad (siehe dazu Abbildung 4.3 in Abschnitt 4.3) und mit einer Geschwindigkeit über dem Schwellenwert eliminiert.

Es ist ersichtlich, dass alle Kurven starke Schwankungen aufweisen, wobei keine davon einen ähnlichen Verlauf zeigt wie die Kurven der Diagramme des Algorithmus der Geschwindigkeit. Zudem ist die Anzahl an Stopps in allen Diagrammen beachtlich tiefer als in den vorherigen Diagrammen.

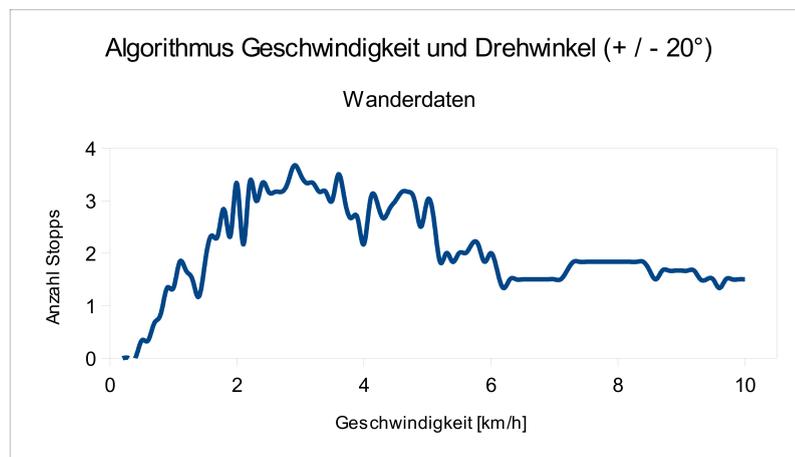


ABBILDUNG 5.6: Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Wanderdaten.

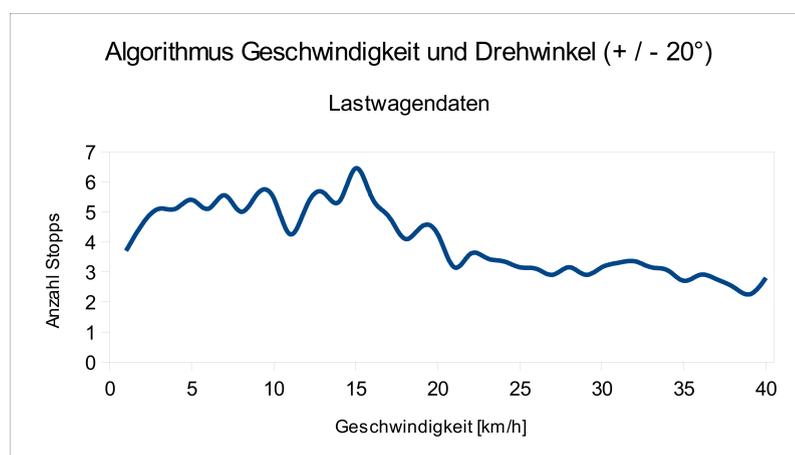


ABBILDUNG 5.7: Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Lastwagendaten.

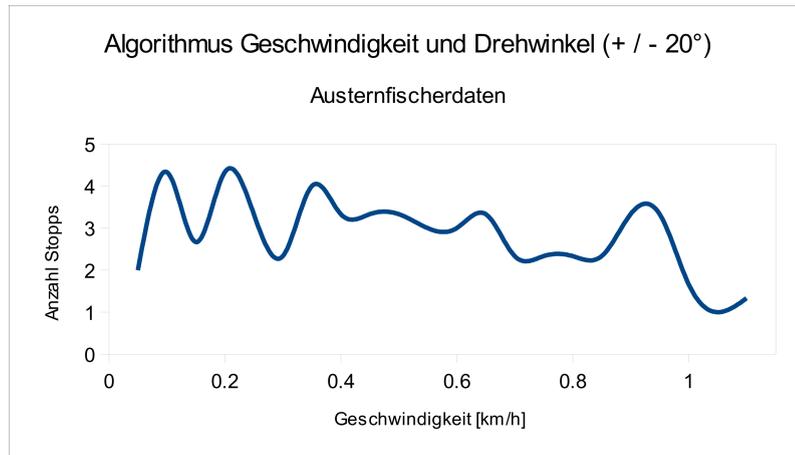


ABBILDUNG 5.8: Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Austernfischerdaten.

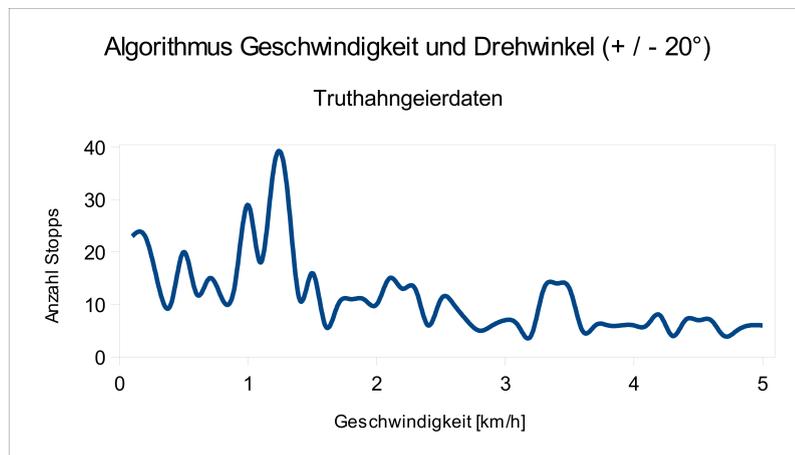


ABBILDUNG 5.9: Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Truthahngeriederdaten.

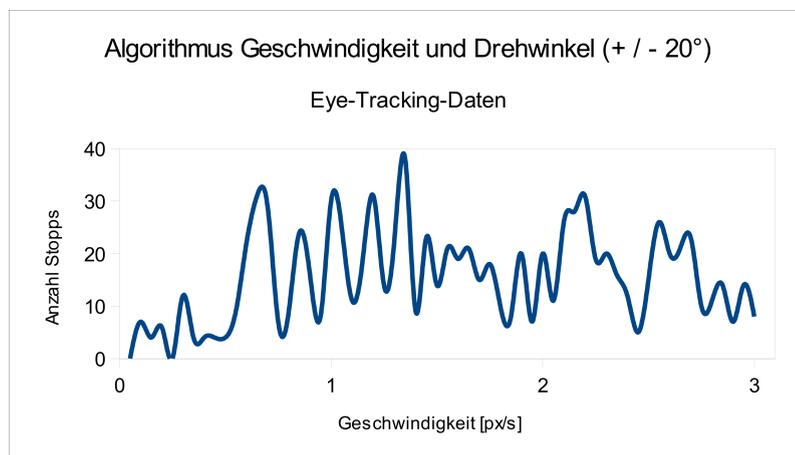


ABBILDUNG 5.10: Geschwindigkeitsschwellenwert und minimaler Drehwinkel angewendet auf Eye-Tracking-Daten.

**Dauer eines Stopps** Der Einfluss einer Mindestdauer eines Stopps, wie in Abschnitt 4.3 beschrieben, welcher bereits durch eine gewisse Geschwindigkeit definiert ist, wird anhand folgender Diagramme erläutert. Dabei stellt der Wert der Dauer jeweils einen oberen Schwellenwert dar. So sind alle durch die Geschwindigkeit definierten Stopps, die mit einer Mindestdauer von beispielsweise einer Minute als Stopp definiert werden, auch bei einer Mindestdauer von zwei Minuten als Stopp definiert. Bei einer Geschwindigkeit von ungefähr 0.3 km/h im Diagramm 5.13 scheint es, als wäre die Linie für eine Dauer grösser 30s oberhalb der Linie für eine Dauer grösser 10s, was nicht möglich ist. Dieser Effekt kommt durch die Kantenglättung, die verwendet wurde, um die Diagramme ansprechender zu gestalten, zustande.

In den Diagrammen zu den Wander-, den Lastwagen- und den Eye-Tracking-Daten (Abbildungen 5.11, 5.12, 5.15) ist jeweils bei einer gewissen Dauer eine markante Reduktion der Anzahl an Stopps zu erkennen. Diese grosse Reduktion ist im Verhältnis zu der Reduktion an Stopps zwischen den anderen Kurven um einiges grösser. Auch bei den Truthahngerdaten (Abbildung 5.14) werden die Unterschiede zwischen den einzelnen Kurven mit einer steigenden Mindestdauer kleiner. So ist der Unterschied an gefundenen Stopps zwischen einer Mindestdauer von 40 Stunden und 80 Stunden um einiges kleiner als zwischen einer Mindestdauer von 5 Stunden und 40 Stunden. Es gibt jedoch in diesem Diagramm keine starke Reduktion an einer bestimmten Stelle. Die Kurven im Diagramm 5.13 zu den Austernfischerdaten weisen auch hier die grössten Schwankungen auf, wobei mit steigender Mindestdauer auch die Anzahl an Stopps relativ gleichmässig abnimmt. Durch diese gleichmässige Abnahme an Stopps ist auch hier kein kritischer Bereich für die Mindestdauer ersichtlich.

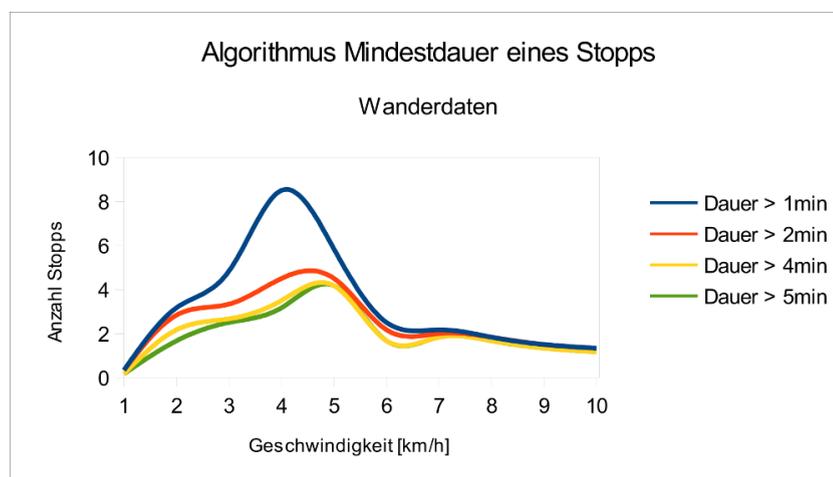


ABBILDUNG 5.11: Geschwindigkeits- und Mindestdauerschwellenwert angewendet auf Wanderdaten.

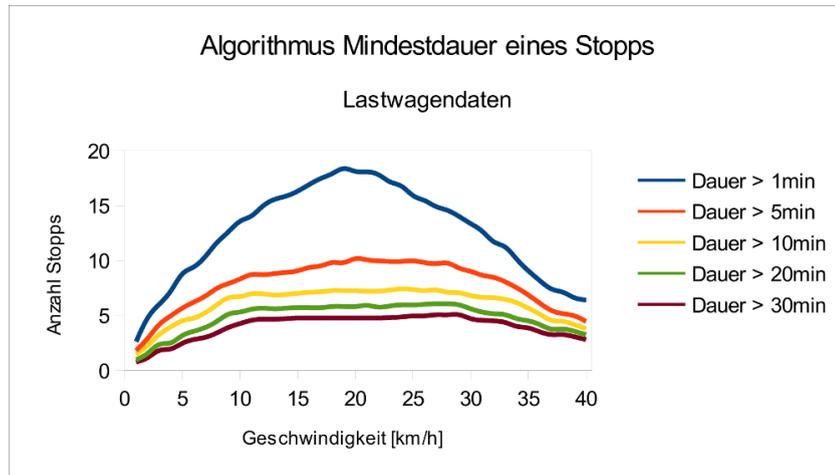


ABBILDUNG 5.12: Geschwindigkeits- und Mindestdauerschwel­lenwert angewendet auf Lastwagendaten.

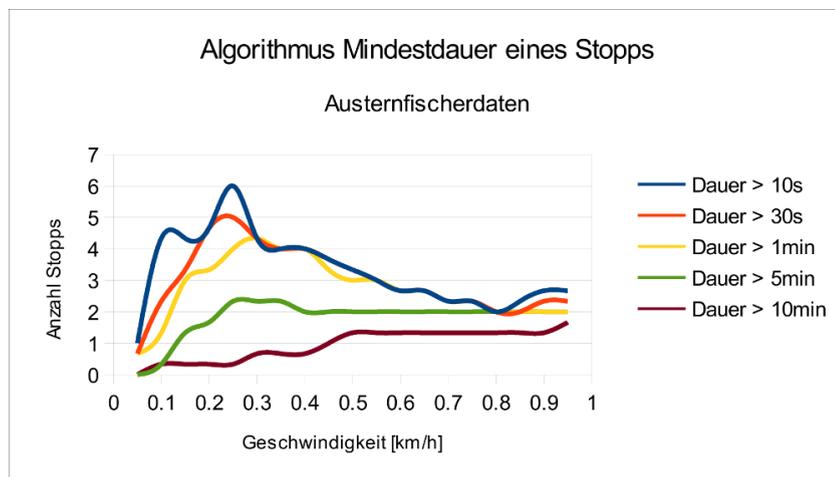


ABBILDUNG 5.13: Geschwindigkeits- und Mindestdauerschwel­lenwert angewendet auf Austernfischerdaten.

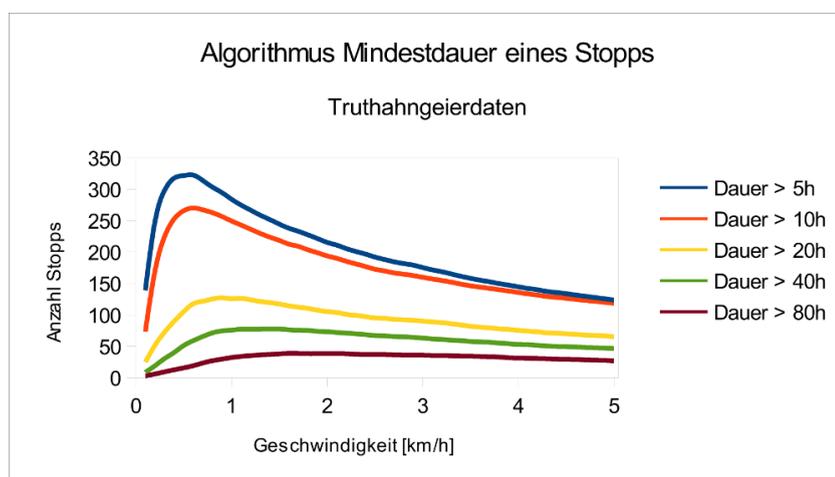


ABBILDUNG 5.14: Geschwindigkeits- und Mindestdauerschwel­lenwert angewendet auf Truthahngeierdaten.

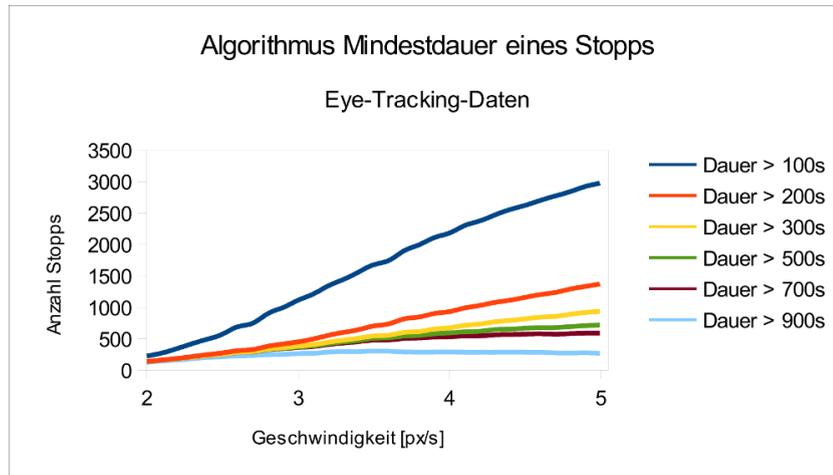


ABBILDUNG 5.15: Geschwindigkeits- und Mindestdauerschwel­lenwert angewendet auf Eye-Tracking-Daten.

**Ausdehnung eines Stopps** Der Einfluss einer Mindestausdehnung für einen Stopp ist in den Diagrammen 5.16 bis 5.20 dargestellt. Genauere Angaben zu diesem Algorithmus sind in Abschnitt 4.3 erläutert.

Es ist ersichtlich, dass die Kurven der meisten Diagramme eine ähnliche Form aufweisen wie unter Anwendung des Algorithmus der Mindestausdehnung, die maximale Anzahl gefundener Stopps hingegen differiert stärker. Der Geschwindigkeitsschwellenwert, bei welchem das Maximum an Stopps gefunden wird ist bei den Wander-, den Lastwagen- und den Austernfischerdaten (Abbildungen 5.16 bis 5.18) der gleiche wie unter Verwendung eines Schwellenwertes für die Mindestdauer. Bei den Daten zum Truthahngerier in Abbildung 5.19 wird unter Anwendung eines Schwellenwertes für die Mindestausdehnung die maximale Anzahl an Stopps bei einem höheren Geschwindigkeitsschwellenwert gefunden als bei der Anwendung einer Mindestdauer. Der Verlauf der Kurve in Diagramm 5.19 ist jedoch ähnlich wie unter Anwendung des Algorithmus einer Mindestdauer (Abbildung 5.14). Einzig die Kurven im Diagramm 5.20 zu den Eye-Tracking-Daten reagieren auf diesen Algorithmus nicht sehr ähnlich wie auf den Algorithmus der Mindestdauer. Auch die maximale Anzahl an Stopps wird bei den Eye-Tracking-Daten bei einer unterschiedlichen Geschwindigkeit gefunden als mit dem Algorithmus der Mindestdauer.

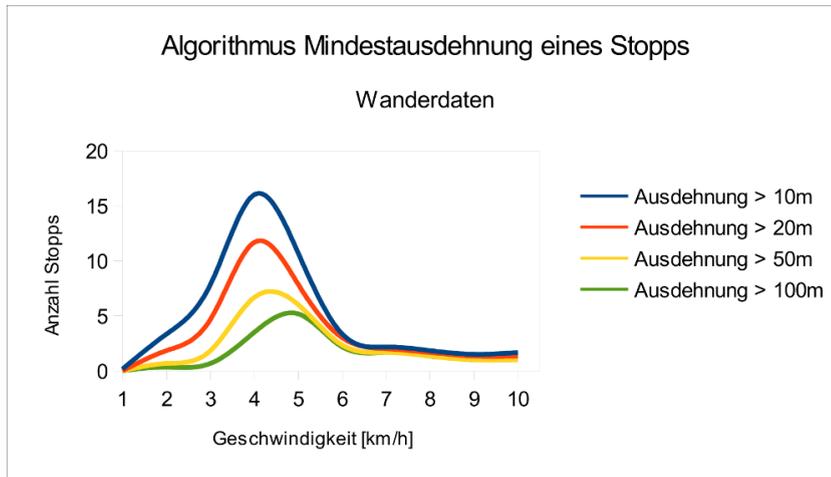


ABBILDUNG 5.16: Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Wanderdaten.

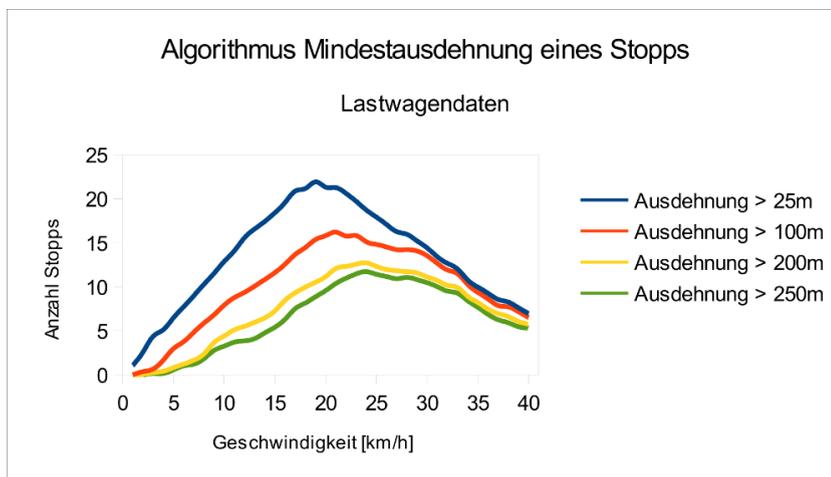


ABBILDUNG 5.17: Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Lastwagendaten.

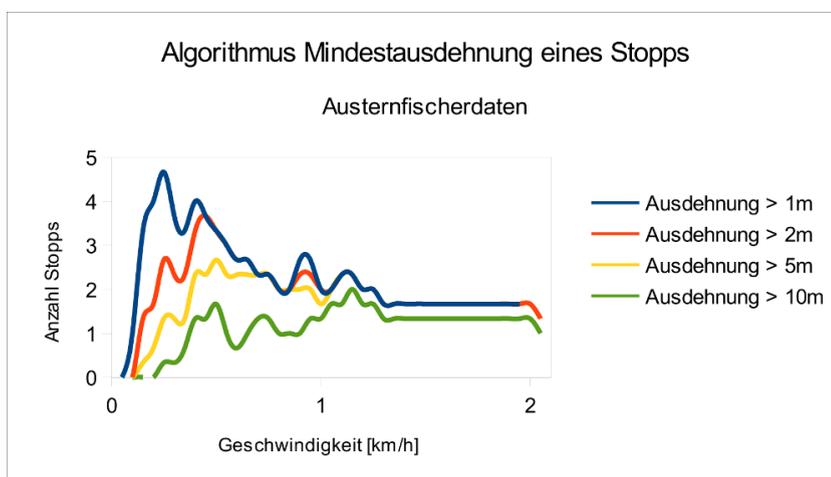


ABBILDUNG 5.18: Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Austernfischerdaten.

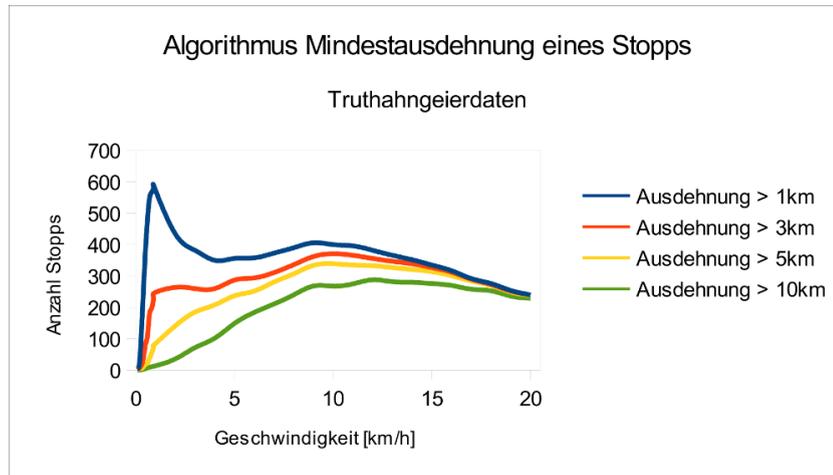


ABBILDUNG 5.19: Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Truthahngeriederdaten.

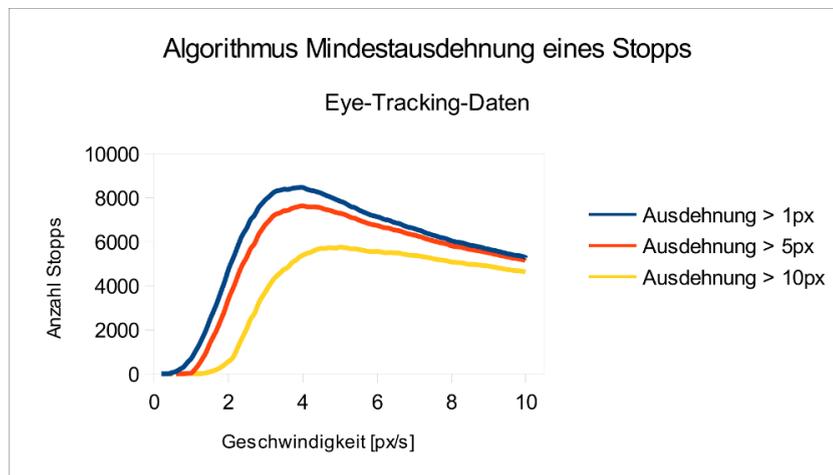


ABBILDUNG 5.20: Geschwindigkeits- und Mindestausdehnungsschwellenwert angewendet auf Eye-Tracking-Daten.

## 5.2 Resultate der Algorithmengruppe Stops and Moves of Trajectories

**Clustering-Based Stops and Moves of Trajectories** Der Algorithmus CB-SMoT orientiert sich einerseits am DBSCAN und stellt andererseits eine Erweiterung des SMoT dar (siehe Palma et al. (2008) sowie Abschnitt 4.4). Um diese Erweiterung des SMoT durchführen zu können müssen Informationen zu Kandidatenregionen vorhanden sein, was bei den hier verwendeten Datensätzen meistens nicht zutrifft. Aus diesem Grund wurde nach den unbekanntem Stopps gesucht, welche ebenfalls mit diesem Algorithmus

detektiert werden können, jedoch ohne Einbezug der Kandidatenregionen. Die Resultate sind in den Diagrammen 5.21 bis 5.25 dargestellt. Dabei wurde die Distanz zur Berechnung der Nachbarschaft als Schwellenwert variiert.

In den Diagrammen ist ersichtlich, mit Ausnahme der Eye-Tracking-Daten in Diagramm 5.25, dass die Anzahl maximaler Stopps im Gegensatz zu den bisherigen Algorithmen hoch ist. Das Diagramm 5.21 der Wanderdaten zeigt für tiefere Stoppdauerwerte eine ähnliche Kurve wie in den Diagrammen der Bewegungsparameter. Ab einer Stoppdauer von fünf Minuten stabilisieren sich die Kurven und die Anzahl an Stopps variiert nicht mehr stark. Die Kurven im Diagramm 5.22 der Lastwagendaten weisen einen ähnlichen Verlauf auf wie in den bisherigen Diagrammen, wobei das Maximum an Stopps höher ist. Die Austernfischerdaten (Diagramm 5.23) scheinen sich ab einer gewissen Distanz zur Berechnung der Nachbarschaft zu stabilisieren. Dabei nimmt sowohl die Variation zwischen den Kurven als auch innerhalb einer einzelnen Kurve ab. Dies konnte bei keinem der bisherigen Algorithmen angewendet auf die Austernfischerdaten beobachtet werden. Die Anzahl maximal gefundener Stopps ist um einiges grösser als in den bisherigen Diagrammen zum Austernfischer. Ab einer Distanz von 5 Metern zur Berechnung der Nachbarschaft sind die Kurven stabil. Diagramm 5.24 zu den Truthahngerierdaten weist zwischen einer Stoppdauer von 10 Stunden und 15 Stunden eine markante Abnahme der Anzahl an Stopps auf. Für eine Mindeststoppdauer von 15 Stunden werden ähnlich viele Stopp entdeckt wie von den bisher gezeigten Algorithmen. Eine Stoppdauer von mindestens 20 Stunden weist im Vergleich zu einer Stoppdauer von 15 Stunden ebenfalls eine Reduktion an Stopps auf, wobei der Unterschied nicht so markant ist wie zwischen einer Dauer von 10 und 15 Stunden. Diese starke Reduktion an Stopps mit einer Stoppdauer von mindestens 15 Stunden bei den Truthahngerierdaten weist darauf hin, dass diese Daten sensibel reagieren auf die Parameter des CB-SMoT.

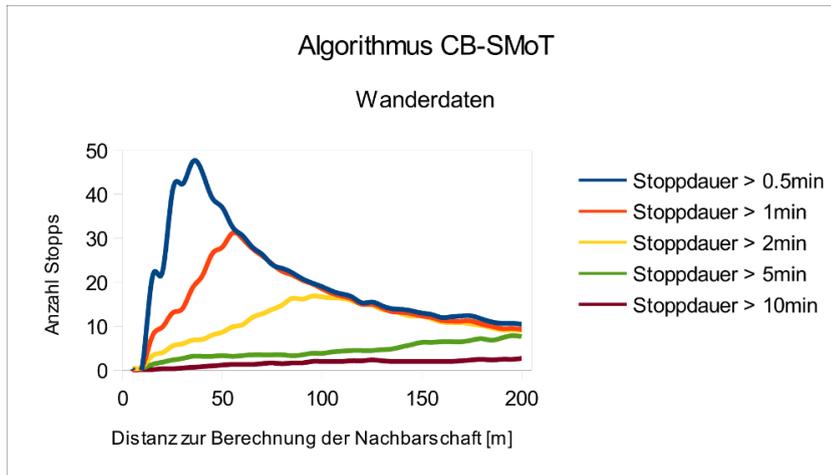


ABBILDUNG 5.21: CB-SMoT angewendet auf Wanderdaten.

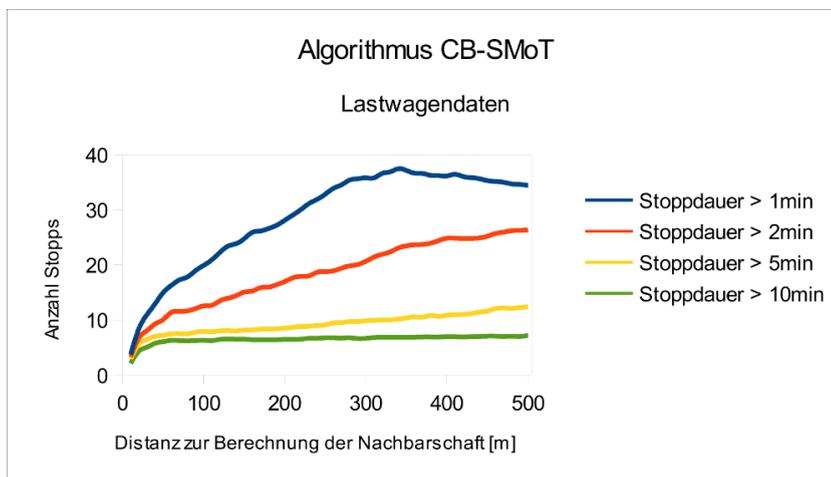


ABBILDUNG 5.22: CB-SMoT angewendet auf Lastwagendaten ohne Kandidatenregionen.

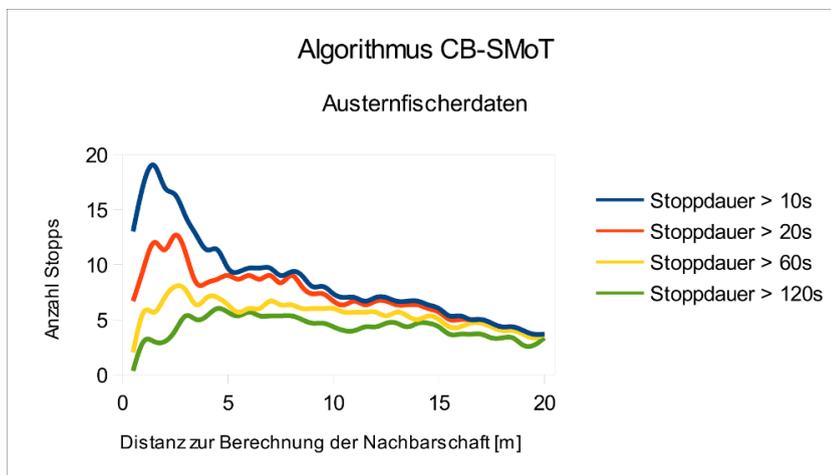


ABBILDUNG 5.23: CB-SMoT angewendet auf Austernfischerdaten.

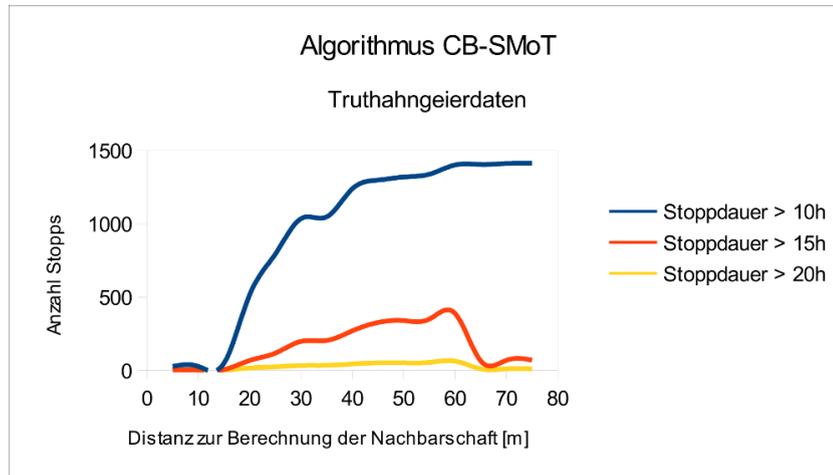


ABBILDUNG 5.24: CB-SMoT angewendet auf Truthahngeierdaten.

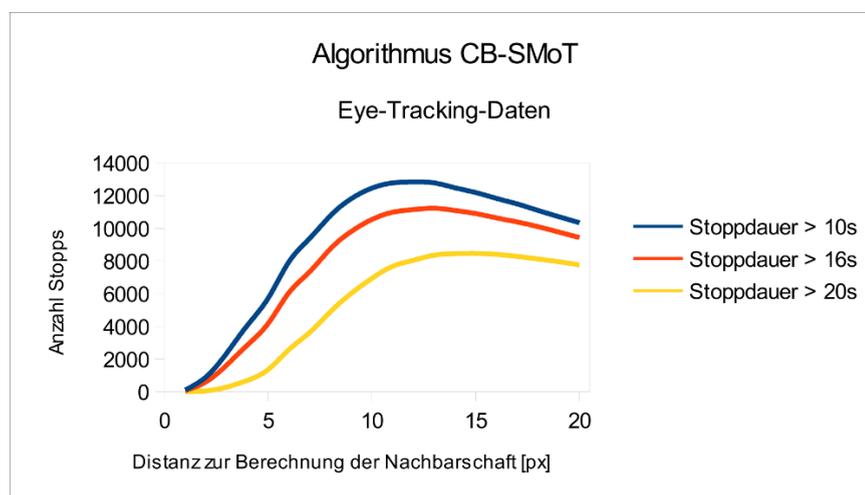


ABBILDUNG 5.25: CB-SMoT angewendet auf Eye-Tracking-Daten.

**Direction-Based Stops and Moves of Trajectories** Der DB-SMoT (siehe Rocha et al. (2010) sowie Abschnitt 4.4) liefert Resultate, die in den Abbildungen 5.26 bis 5.30 dargestellt sind. Für alle Analysen wurde der Wert 2 für den Parameter *maxTol* eingesetzt. Dieser Parameter definiert, wie viele aufeinanderfolgende Punkte in einem Stopp eine geringe Richtungsänderung aufweisen dürfen (Rocha et al., 2010). Besteht im Fall dieser Arbeit beispielsweise ein Stopp aus zehn Punkten, so kann es sein, dass nur acht davon einen grossen und zwei einen geringen Drehwinkel aufweisen. Es ist davon auszugehen, dass vereinzelte Punkte in einem Stopp einen geringen Drehwinkel aufweisen können. Wird der Wert *maxTol* sehr tief gesetzt, so eliminiert man allfällige Stopps auf Grund vereinzelter Punkte mit einem geringen Drehwinkel. Wird der Wert sehr hoch gesetzt, so besteht die Gefahr, dass Stopps definiert werden, die keine sind. Ich habe

mich entschieden, den Wert 2 für den Parameter *maxTol* zu nehmen, da so die Gefahr für ungewollte Stopps gering ist und doch eine gewisse Einschränkung für die Definition eines Stopps gemacht wird.

Die Anzahl an Stopps bewegt sich für alle Datensätze in einem mittleren Bereich im Vergleich zu den Resultate der bisherigen Algorithmen, wobei in keinem Fall besonders viel oder besonders wenige Stopps gefunden werden.

Die Diagramme zu den Wander- und den Lastwagendaten (Diagramm 5.26 und 5.27) weisen einen ähnlichen Verlauf auf. Die Anzahl an Stopps nimmt grundsätzlich mit einem steigenden minimalen Drehwinkel ab. Die Anzahl an Stopps im Diagramm 5.26 der Wanderdaten stabilisiert sich bereits bei einem minimalen Drehwinkel von ungefähr 80 Grad. Die Lastwagen reagieren relativ gleichmässig auf den Schwellenwert des Drehwinkels als auch auf den Wert für die Minstdauer.

Die Kurven im Diagramm 5.28 des Austernfischers weisen stärkere Schwankungen auf als die Kurven der anderen Diagramme des DB-SMoT. Interessant ist hierbei, dass die maximale Anzahl an Stopps für sehr kurze Minstdauerwerte erst bei einem minimalen Drehwinkel von 160 Grad gefunden wird. Für höhere Minstdauerwerte hingegen stabilisiert sich die Anzahl an Stopps bei einem tieferen Wert.

Bei den Truthahngeierdaten in Diagramm 5.29 ist zwischen einer Minstdauer von 10 und 20 Stunden eine markante Abnahme an Stopps zu erkennen. Dieses Verhalten konnte bereits in Diagramm 5.24 des CB-SMoT beobachtet werden, dort ist eine starke Reduktion der Anzahl an Stopps zwischen einer Minstdauer von 10 Stunden und 15 Stunden zu sehen.

Die Eye-Tracking-Daten in Diagramm 5.30 reagieren unterschiedlich als die anderen Datensätze bei der Anwendung des DB-SMoT. So variiert die Anzahl an Stopps unter unterschiedlichen Werten für die Stoppdauer bis zu einem minimalen Drehwinkel von 90 Grad nicht. Das Maximum an Stopps wird ähnlich wie bei den Austernfischerdaten erst bei einem hohen minimalen Drehwinkel von 150 Grad erreicht. Zwischen den einzelnen Werten für die mindeste Stoppdauer gibt es keine grossen Sprünge, so nimmt die Anzahl an Stopps mit zunehmender Mindeststoppdauer relativ gleichmässig ab (Diagramm 5.30).

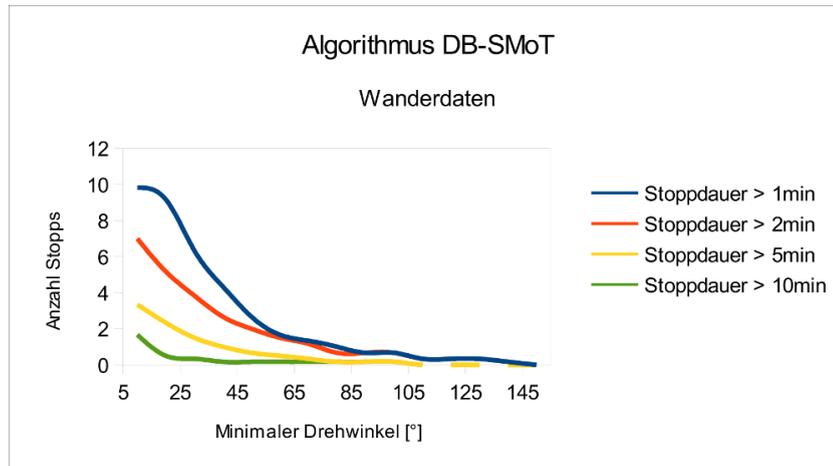


ABBILDUNG 5.26: DB-SMoT angewendet auf Wanderdaten.

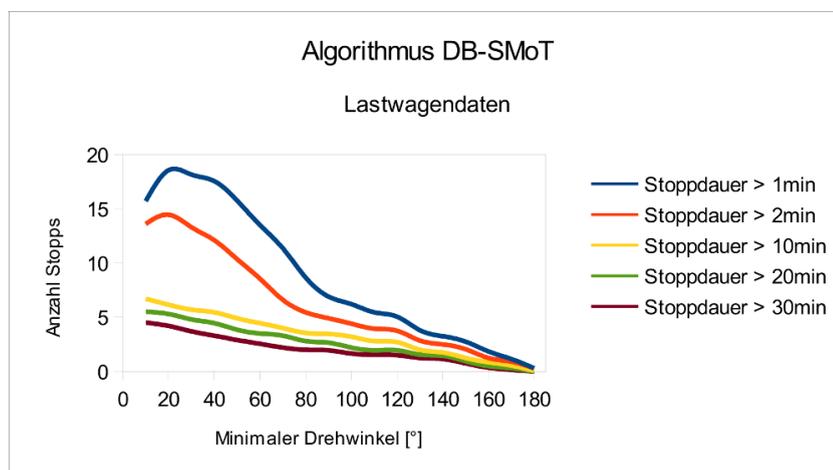


ABBILDUNG 5.27: DB-SMoT angewendet auf Lastwagendaten.

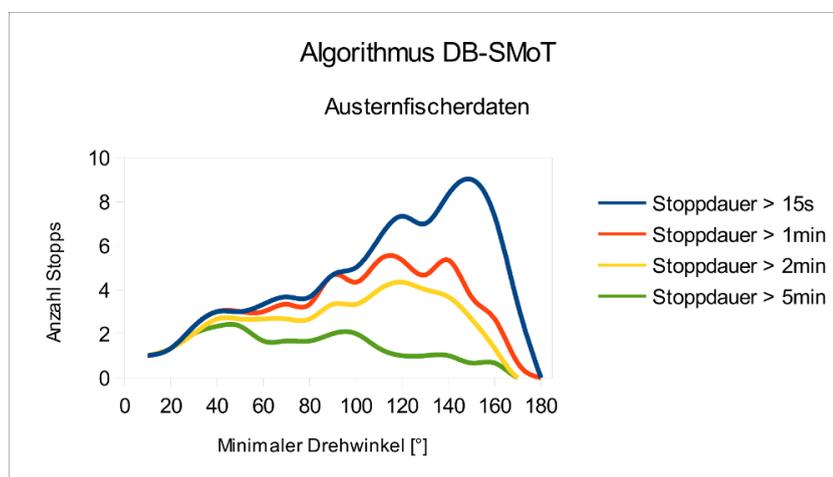


ABBILDUNG 5.28: DB-SMoT angewendet auf Austernfischerdaten.

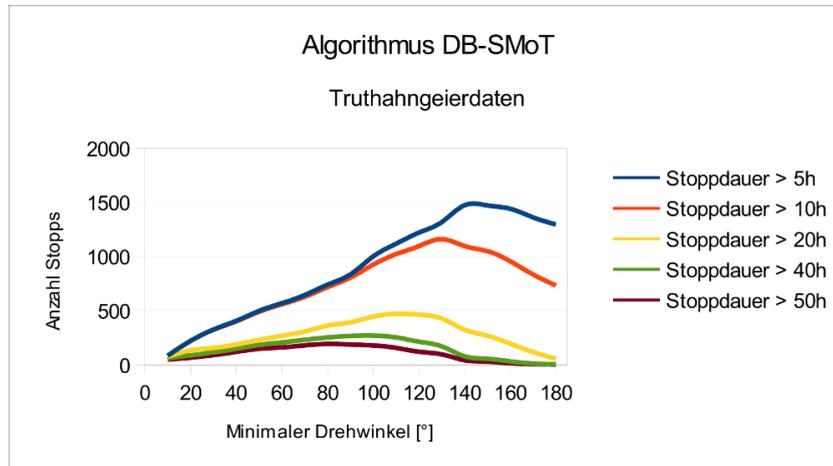


ABBILDUNG 5.29: DB-SMoT angewendet auf Truthahngeierdaten.

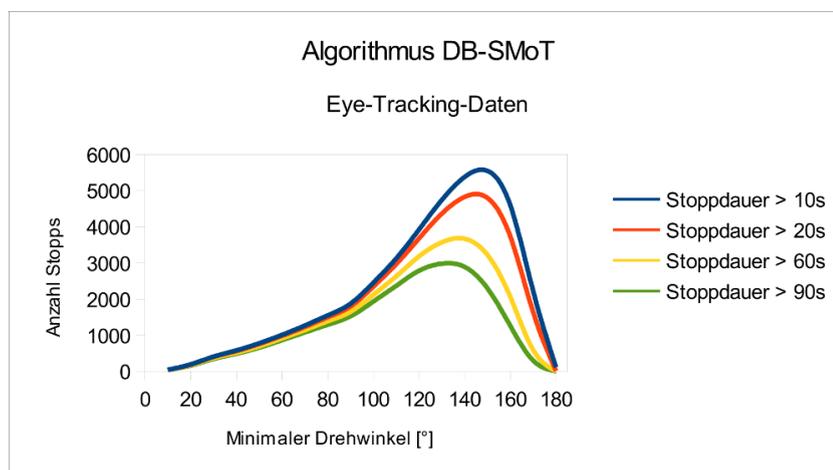


ABBILDUNG 5.30: DB-SMoT angewendet auf Eye-Tracking-Daten.

### 5.3 Resultate des Hidden Markov Model

Wie in Abschnitt 4.5 erläutert, kann auch ein Hidden Markov Model (Rabiner, 1989) zur Detektion von Stopps in Trajektorien verwendet werden. Interessant ist bei diesem Algorithmus der Umstand, dass die Räumlichkeit der Trajektorie nur indirekt im Rahmen der Geschwindigkeit miteinbezogen wird. Die Werte für die Übergangswahrscheinlichkeiten sind in allen Fällen relativ hoch ausgefallen, oftmals sogar über 0.9. In Abschnitt 4.5.1 ist ersichtlich, dass hohe Werte für die Übergangswahrscheinlichkeiten durchaus sinnvoll sind. Die Abbildungen 5.31 bis 5.35 zeigen die Resultate der Analysen des HMM.

Für die Wander- und die Lastwagendaten in den Diagrammen 5.31 und Diagramm

5.32 befindet sich die maximale Anzahl an Stopps in einem mittleren Bereich im Vergleich zu den anderen Diagrammen. Die Diagramme zu diesen Datensätzen sehen auch sehr ähnlich aus wie die Diagramme 5.1 und 5.2 des Geschwindigkeitsschwellenwertes. Bei den restlichen Datensätzen werden maximal nur wenige Stopps gefunden, bei den Eye-Tracking-Daten in Diagramm 5.35 werden gut 40 Stopps gefunden. Diagramm 5.33 zu den Austernfischerdaten weist auch hier relativ starke Schwankungen auf. Ein ähnliches, ebenfalls etwas unregelmässiges und unstetes Verhalten zeigt die Kurve der Eye-Tracking-Daten (Diagramm 5.35). Die maximale Anzahl an Stopps wird bei den Truthahngerdaten in Diagramm 5.34 bei sehr tiefen Geschwindigkeitswerten aufgezeichnet, danach nimmt die Anzahl an Stopps gleichmässig ab.

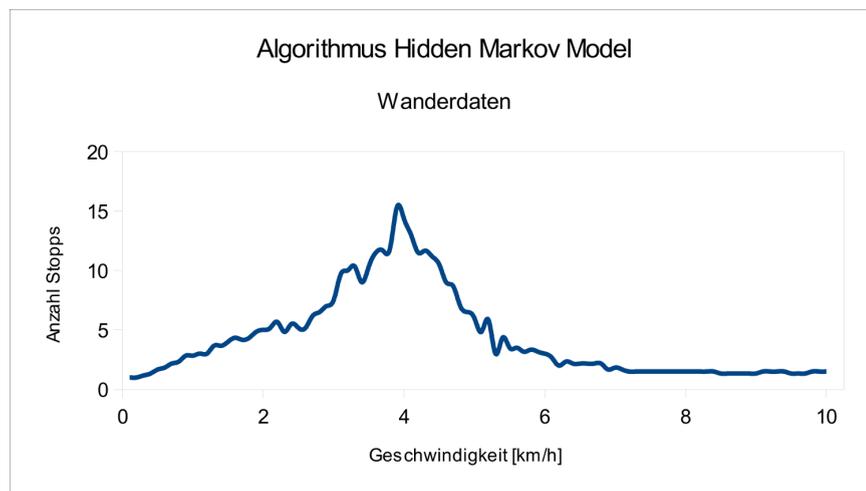


ABBILDUNG 5.31: HMM angewendet auf Wanderdaten.

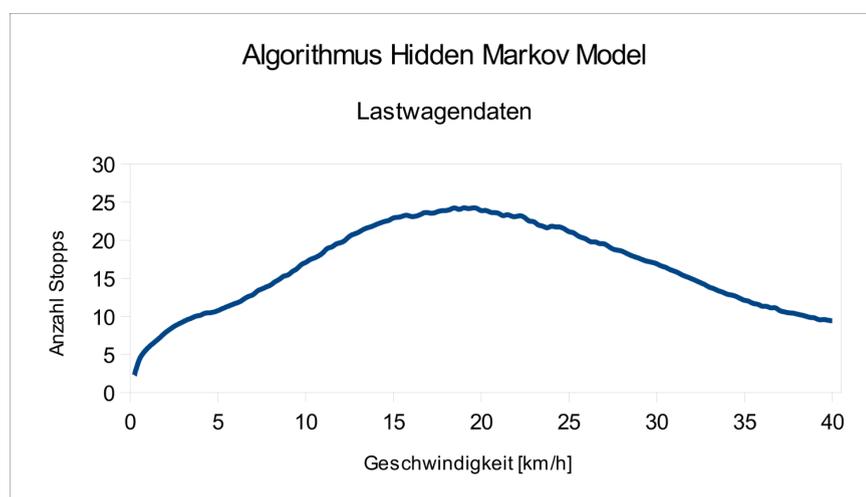


ABBILDUNG 5.32: HMM angewendet auf Lastwagendaten.

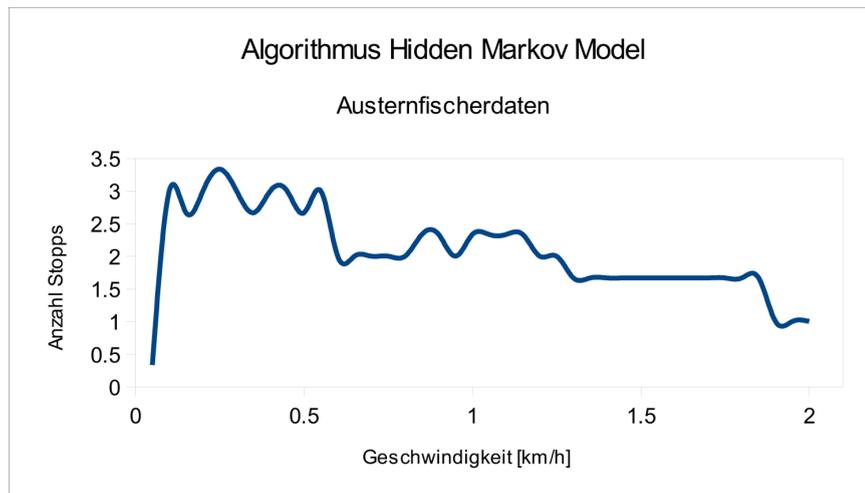


ABBILDUNG 5.33: HMM angewendet auf Austernfischerdaten.

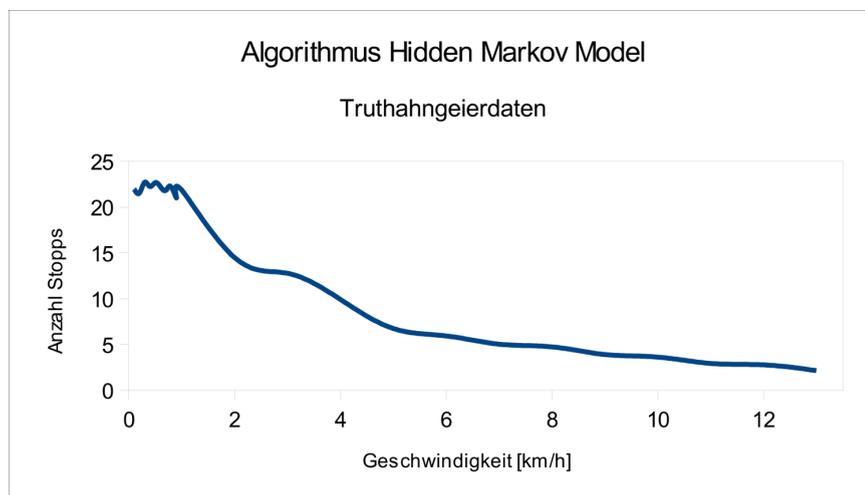


ABBILDUNG 5.34: HMM angewendet auf Truthahngeierdaten.

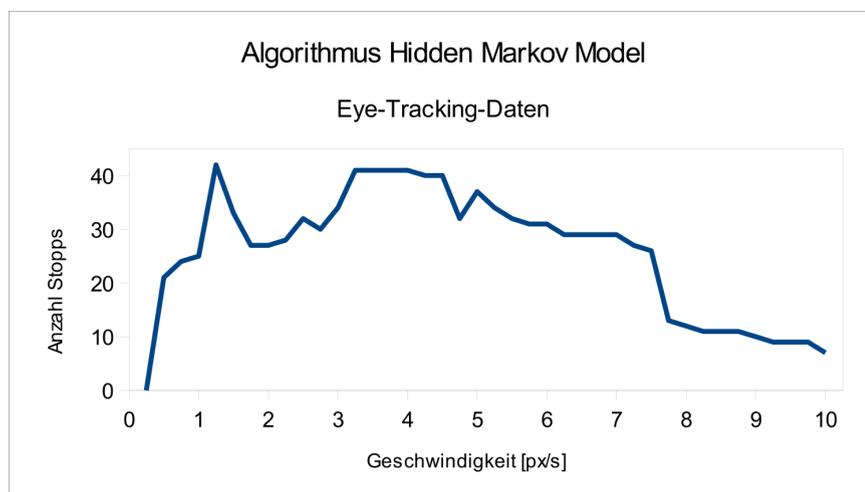


ABBILDUNG 5.35: HMM angewendet auf Eye-Tracking-Daten.

## 5.4 Resultate zur Analyse der ausgewählten Trajektorien

Im Folgenden sind die Resultate der ausgewählten Trajektorien ersichtlich. Es gilt zu beachten, dass die folgenden Abbildungen jeweils die beste Lösung aller gewählten Methoden zeigen. Dabei wurde anhand einer visuellen Analyse entschieden, welche Lösung den tatsächlich gemachten Stopps am besten entspricht. Oftmals verhielt es sich dabei so, dass mit einer Erhöhung der Schwellenwerte zwar mehr tatsächliche Stopps, aber auch mehr falsche Stopps gefunden wurden. Somit war es jeweils ein Abwägen, ob ein Resultat durch zusätzliche echte als auch falsche Stopps verbessert wird oder nicht. Die Entscheidung, welche Variante schliesslich als beste definiert wurde, ist eine individuelle und wäre von anderen Personen möglicherweise anders gefällt worden. Die gewählten Werte für die einzelnen Bewegungsparameter sind in Tabelle 5.1 aufgeführt.

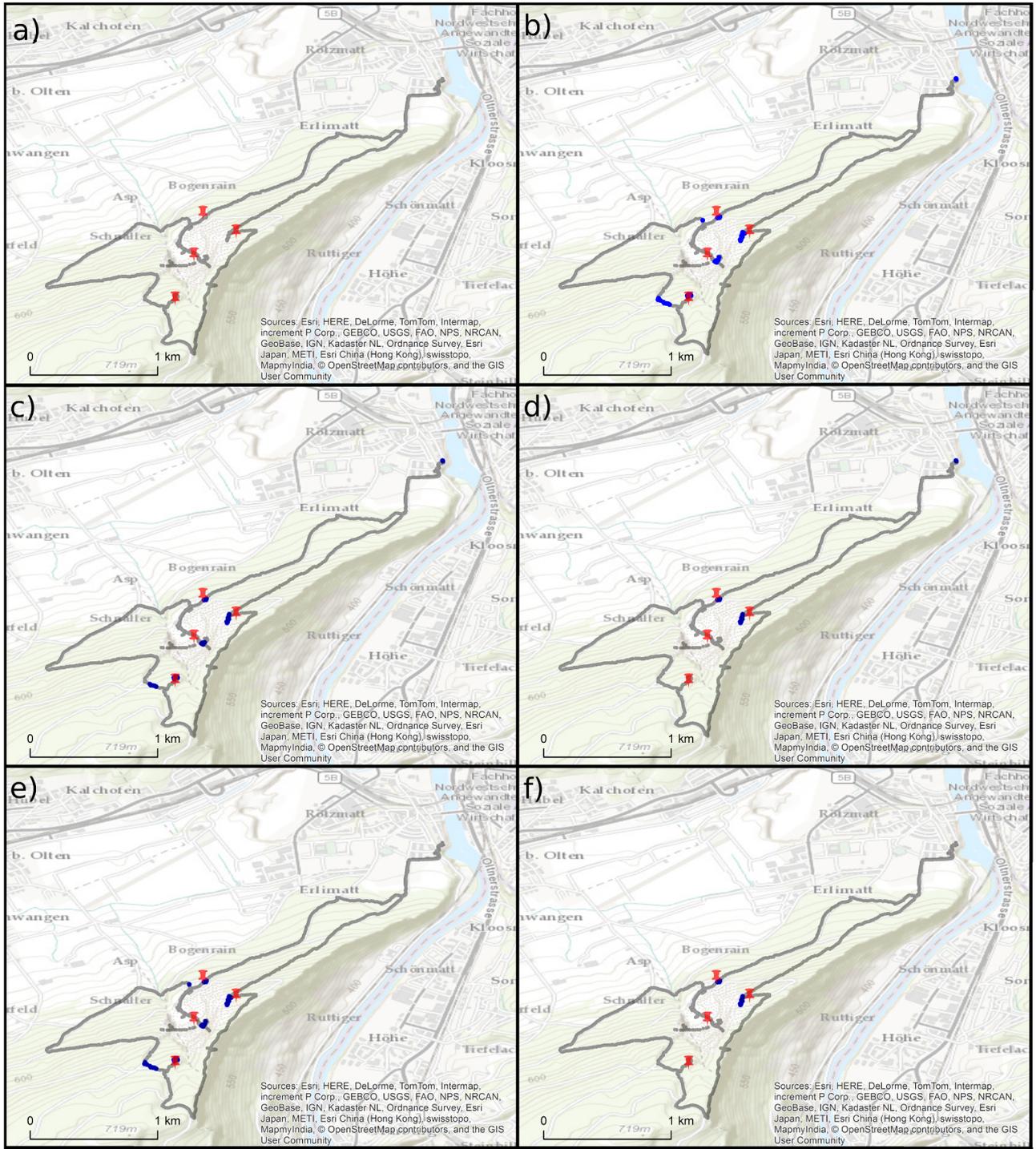
TABELLE 5.1: Übersicht über die Werte der Bewegungsparameter in der Analyse der ausgewählten Trajektorien.

	<b>Wanderer</b>	<b>Lastwagen</b>	<b>Austernfischer</b>
<b>Geschwindigkeitsschwellenwert:</b>	2 km/h	3 km/h	0.3 km/h
<b>Minstdauer:</b>	150 s	17 min	200 s
<b>Minstdauer:</b>	150 s	5 min	250 s
<b>Punkteanteil mit minimalem Drehwinkel:</b>	35 %	35 %	35 %
<b>minimaler Drehwinkel:</b>	90 Grad	90 Grad	90 Grad
<b>Mindestausdehnung:</b>	10 m	150 m	2 m
<b>Mindestausdehnung:</b>	10 m	150 m	2 m
<b>Punkteanteil mit minimalem Drehwinkel:</b>	35 %	35 %	35 %
<b>minimaler Drehwinkel:</b>	90 Grad	90 Grad	90 Grad
<b>SMoT</b>	keine Angabe		keine Angabe
<b>Radius um Kandidatenregion:</b>		360 m	
<b>Mindestaufenthaltsdauer in Kandidatenregion</b>		15 min	
<b>CB-SMoT</b>			
<b>Distanz zur Berechnung der Nachbarschaft:</b>	50 m	200 m	3 m
<b>Minstdauer:</b>	120 s	15 min	180 s
<b>DB-SMoT</b>			
<b>Minimaler Drehwinkel:</b>	50 Grad	100 Grad	100 Grad
<b>Minstdauer:</b>	120 s	15 min	180 s
<b>maxTol</b>	2	2	2
<b>HMM</b>			
<b>Geschwindigkeitsschwellenwert:</b>	1.5 km/h	3 km/h	0.21 km/h

**Trajektorie eines Wanderers** Die Resultate zu den Wanderdaten konnten validiert werden, da bekannt war, wo gestoppt wurde. Im Folgenden werden die Resultate in Form von verschiedenen Karten aufgeführt. Die Abbildungen 5.36 a) und 5.37 a) zeigen jeweils

einen Überblick über die Trajektorie und die gemachten Stopps, welche als rote Stecknadelköpfe visualisiert sind. Einige Punkte der Trajektorie in der Mitte der Karte wirken etwas lose zerstückelt, wobei nicht klar ist, wie die einzelnen Punkte zusammenhängen. Dies ist darauf zurückzuführen, dass der Peilsender sich teilweise unter überhängenden Stellen befunden hat und somit das Signal ausgefallen ist. Abbildung 5.36 zeigt die Resultate der Bewegungsparameter und Abbildung 5.37 zeigt die Lösungen der ausgewählten Algorithmen aus der Literatur.

In der Abbildung 5.36 b) ist ersichtlich, dass mit einem Geschwindigkeitsschwellenwert drei falsche Stopps und alle richtigen Stopps gefunden werden. Abbildung 5.36 c) zeigt die Lösung einer zusätzlichen Mindestdauer, hierbei wird lediglich ein falscher Stopp eliminiert. Unter Einbezug einer Mindestdauer und eines Drehwinkels in Abbildung 5.36 d) ist zu sehen, dass noch ein falscher Stopp, jedoch auch nur noch zwei richtige Stopps gefunden werden. Abbildung 5.36 e) zeigt das Resultat eines Geschwindigkeitsschwellenwertes und einer Mindestausdehnung, wobei alle richtigen und zwei falsche Stopps gefunden werden. Auch unter zusätzlichem Einbezug des Drehwinkels zu der Mindestausdehnung in Abbildung 5.36 f) werden zwar alle falschen Stopps, jedoch auch zwei richtige Stopps eliminiert.



Kartenelemente	
	Echter Stopp
	Gefundener Stopp
	Trajektorienpunkt

ABBILDUNG 5.36: Übersicht über die Trajektorie eines Wanderers und die dabei gemachten Stopps laut verschiedenen Bewegungsparametern. a) stellt einen Überblick über die Trajektorie und die gemachten Stopps dar, b) zeigt die Lösung des Bewegungsparameters Geschwindigkeit, c) zeigt die Lösung der Bewegungsparameter Geschwindigkeit und Mindestdauer, d) zeigt die Lösung der Bewegungsparameter Geschwindigkeit, Mindestdauer und Drehwinkel, e) zeigt die Lösung der Bewegungsparameter Geschwindigkeit und Mindestausdehnung, f) zeigt die Lösung der Bewegungsparameter Geschwindigkeit, Mindestausdehnung und Drehwinkel. Quelle Basiskarte: ESRI, HERE, DeLorme, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community.

Abbildung 5.37 b) zeigt das Resultat des HMM, dabei werden alle richtigen Stopps mit einer kleinen räumlichen Ungenauigkeit gefunden, sowie zwei falsche Stopps. Der CB-SMoT in Abbildung 5.37 c) findet sehr viele Stopps, darunter mit einer räumlichen Abweichung auch die vier echten Stopps, jedoch auch elf falsche Stopps. Der DB-SMoT in Abbildung 5.37 d) findet die echten Stopps alle, teilweise jedoch mit einer räumlichen Ungenauigkeit, sowie drei falsche Stopps.

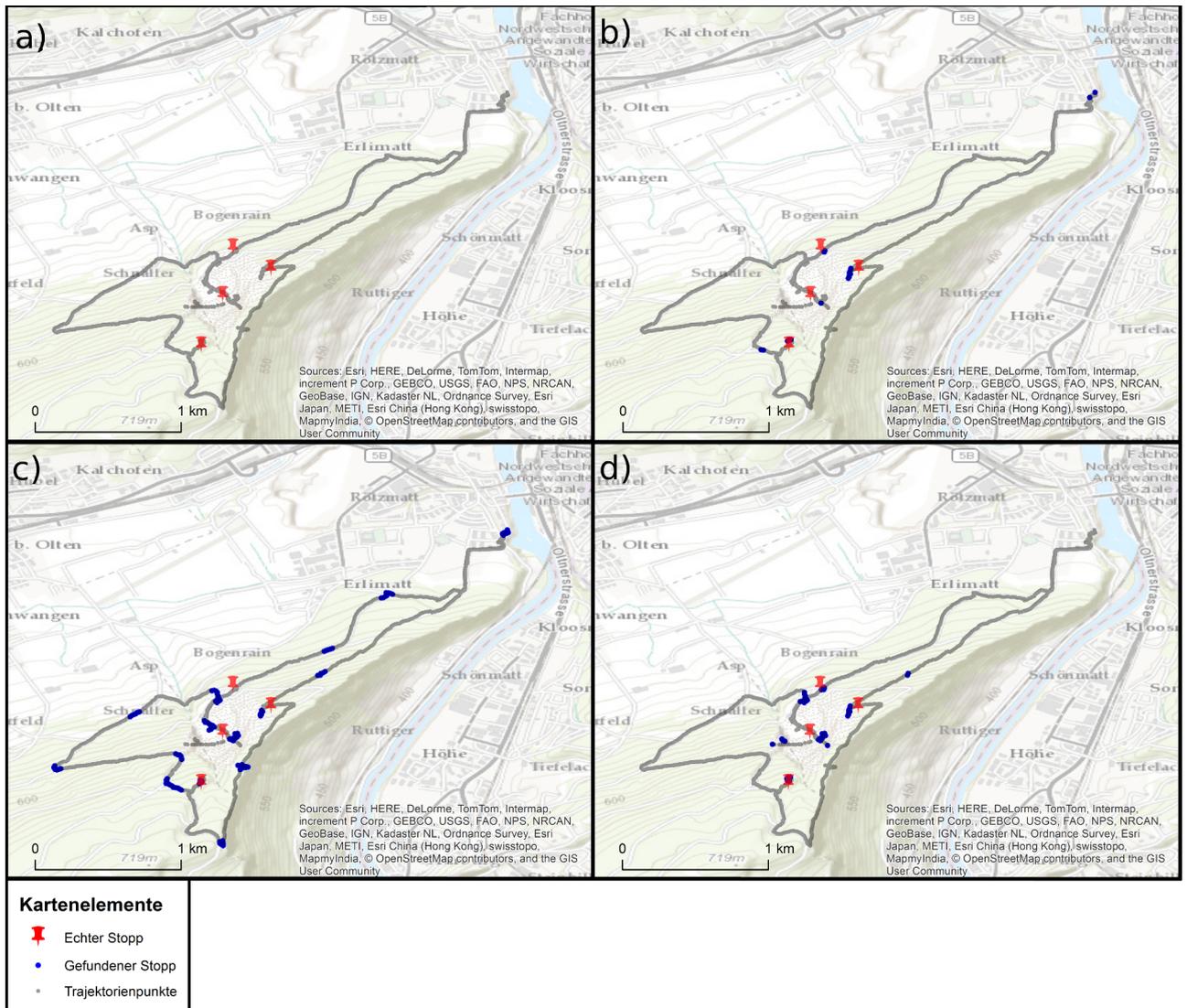
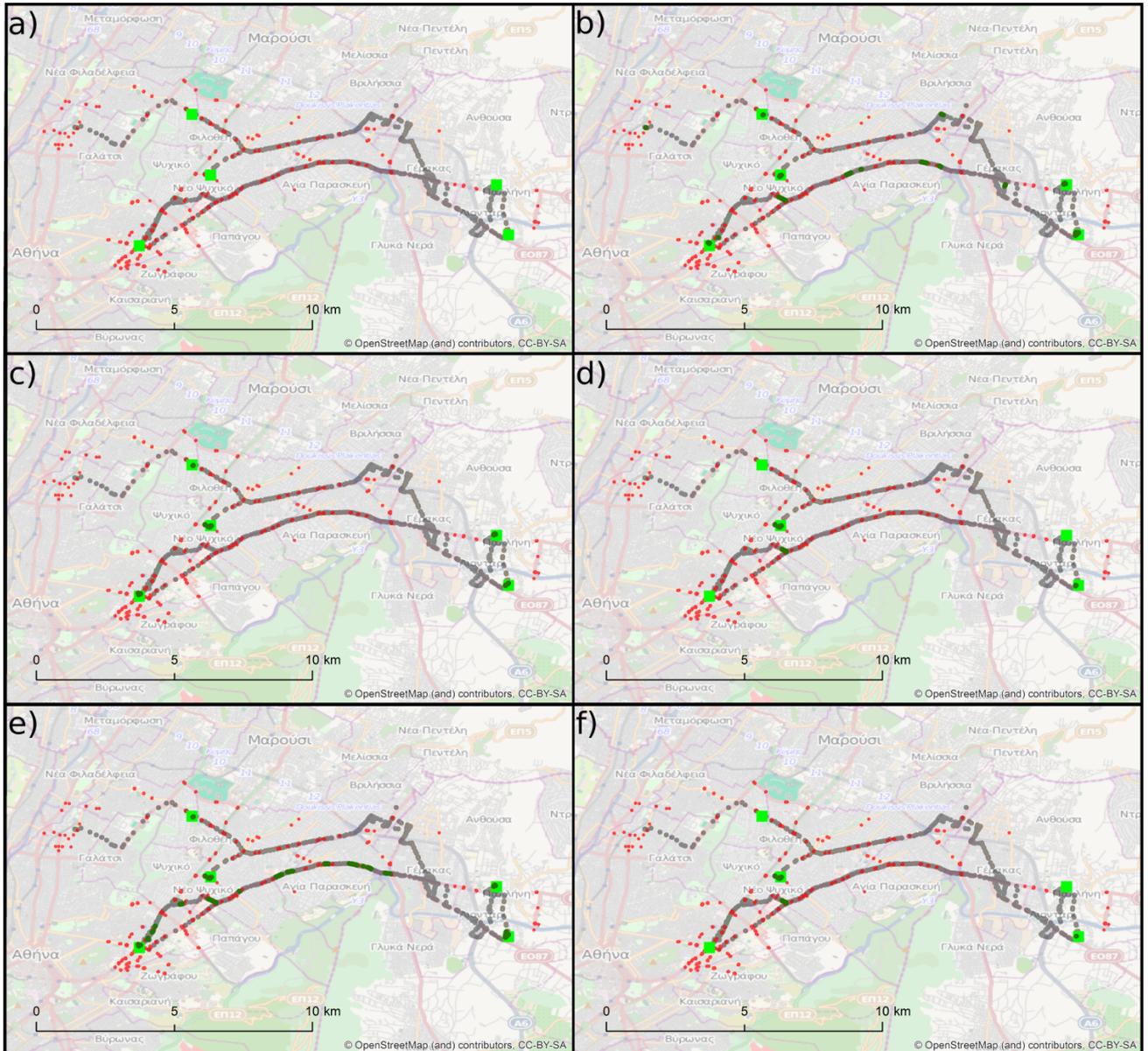


ABBILDUNG 5.37: Übersicht über die Trajektorie eines Wanderers und die dabei gemachten Stopps laut den Algorithmen aus der Literatur. a) stellt einen Überblick über die Trajektorie und die gemachten Stopps dar, b) zeigt die Lösung des HMM, c) zeigt die unbekanntenen Stopps des Algorithmus CB-SMoT, d) zeigt die Lösung des DB-SMoT. Quelle Basiskarte: ESRI, HERE, DeLorme, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community.

**Trajektorie eines betonliefernden Lastwagens über einen Arbeitstag** Zu den Lastwagendaten existiert keine *ground truth*, es konnte jedoch mit Kandidatenregionen gearbeitet werden. Als Kandidatenregionen wurden aus dem Datensatz von OSM die Ampeln extrahiert. Wie in Abbildung 5.38 ersichtlich, gibt es viele Lichtsignale entlang der ausgewählten Trajektorie.

Mit Hilfe von ArcGIS konnte die Trajektorie verhältnismässig realitätsgetreu visualisiert werden. Bei diesem Vorgehen wurde ersichtlich, wo sich wahrscheinlich Stopps befinden müssen, obwohl keine Lichtsignale vorhanden sind und wo Stopps, trotz vorhandener Lichtsignale, unwahrscheinlicher sind. Anhand dieser visuellen Analyse wurden anschliessend fünf Regionen definiert, in denen ein Stopp angenommen werden kann. Die Entscheidung wird gefestigt und nachvollziehbar durch den Umstand, dass die Trajektorie in all diesen Regionen, im Gegensatz zu den restlichen Regionen, für mindestens 25 Minuten auf einem engen Raum verbleibt, teilweise sogar bis zu einer knappen Stunde. Auch wenn der Lastwagen in den restlichen Teilen der Trajektorie teilweise langsam ist, so bewegt er sich meistens in eine Richtung und zeitlich weit entfernte Punkte sind auch räumlich weit entfernt, was an diesen Stellen gegen einen Stopp spricht.

Die Resultate der Bewegungsparameter sind in Abbildung 5.38 dargestellt, die Resultate der Algorithmen aus der Literatur sind in Abbildung 5.39 visualisiert. Es ist ersichtlich, dass mit einem einfachen Geschwindigkeitsschwellenwert wie in Abbildung 5.38 b) alle tatsächlichen Stopps gefunden werden, jedoch auch noch sieben andere, falsche Stopps. Mit dem Einbezug einer Mindestdauer in Abbildung 5.38 c) werden alle tatsächlichen und keine falschen Stopps gefunden. Der Einbezug des Drehwinkels in Abbildung 5.38 d) stellt in diesem Fall eine Verschlechterung des Ergebnisses dar, so werden drei richtige Stopps nicht mehr gefunden und ein zusätzlicher, falscher Stopp wird gefunden. Der Algorithmus der Mindestausdehnung in Abbildung 5.38 e) findet ebenfalls alle tatsächlichen Stopps, jedoch auch einige falsche. Durch die Kombination einer Mindestausdehnung und eines minimalen Drehwinkels wie in Abbildung 5.38 f) werden zwar einige falsche Stopps eliminiert, es werden jedoch auch drei tatsächliche Stopps eliminiert.



Kartenelemente	
•	Gefundener Stopp
■	Annahme eines Stopps
•	Ampel
•	Lastwagentrajektorie

ABBILDUNG 5.38: Übersicht über die Trajektorie eines Lastwagens und die dabei gemachten Stopps laut verschiedenen Bewegungsparametern. a) stellt einen Überblick über die Trajektorie und die gemachten Stopps dar, b) zeigt die Lösung des Bewegungsparameters Geschwindigkeit, c) zeigt die Lösung der Bewegungsparameter Geschwindigkeit und Mindestdauer, d) zeigt die Lösung der Bewegungsparameter Geschwindigkeit, Mindestdauer und Drehwinkel, e) zeigt die Lösung der Bewegungsparameter Geschwindigkeit und Mindestausdehnung, f) zeigt die Lösung der Bewegungsparameter Geschwindigkeit, Mindestausdehnung und Drehwinkel. Quelle Basiskarte: ©OpenStreetMap (and) contributors, CC-BY-SA.

In Abbildung 5.39 sind die Resultate der Algorithmen aus der Literatur visualisiert. Abbildung 5.39 b) zeigt die Lösung des HMM, wobei ein ähnliches Resultat zu sehen ist wie unter Anwendung des Geschwindigkeitsschwellenwertes in Abbildung 5.38 b). Das HMM findet zwar die richtigen jedoch auch einige falsche Stopps. Wie in Abbildung 5.39 c) zu sehen ist, findet der CB-SMoT zuverlässig zwei echte Stopps. Die drei anderen Stopps werden ebenfalls durch den CB-SMoT gefunden, jedoch als unbekannte Stopps. Ein Stopp wird sowohl in Abbildung 5.39 c) als auch in Abbildung 5.39 d) gefunden. Die Lösung des SMoT in Abbildung 5.39 e) findet nur zwei der fünf tatsächlichen Stopps, jedoch auch keine falschen Stopps. Der DB-SMoT in Abbildung 5.39 f) findet vier der fünf tatsächlichen Stopps, wobei einer davon räumlich ungenau lokalisiert ist.

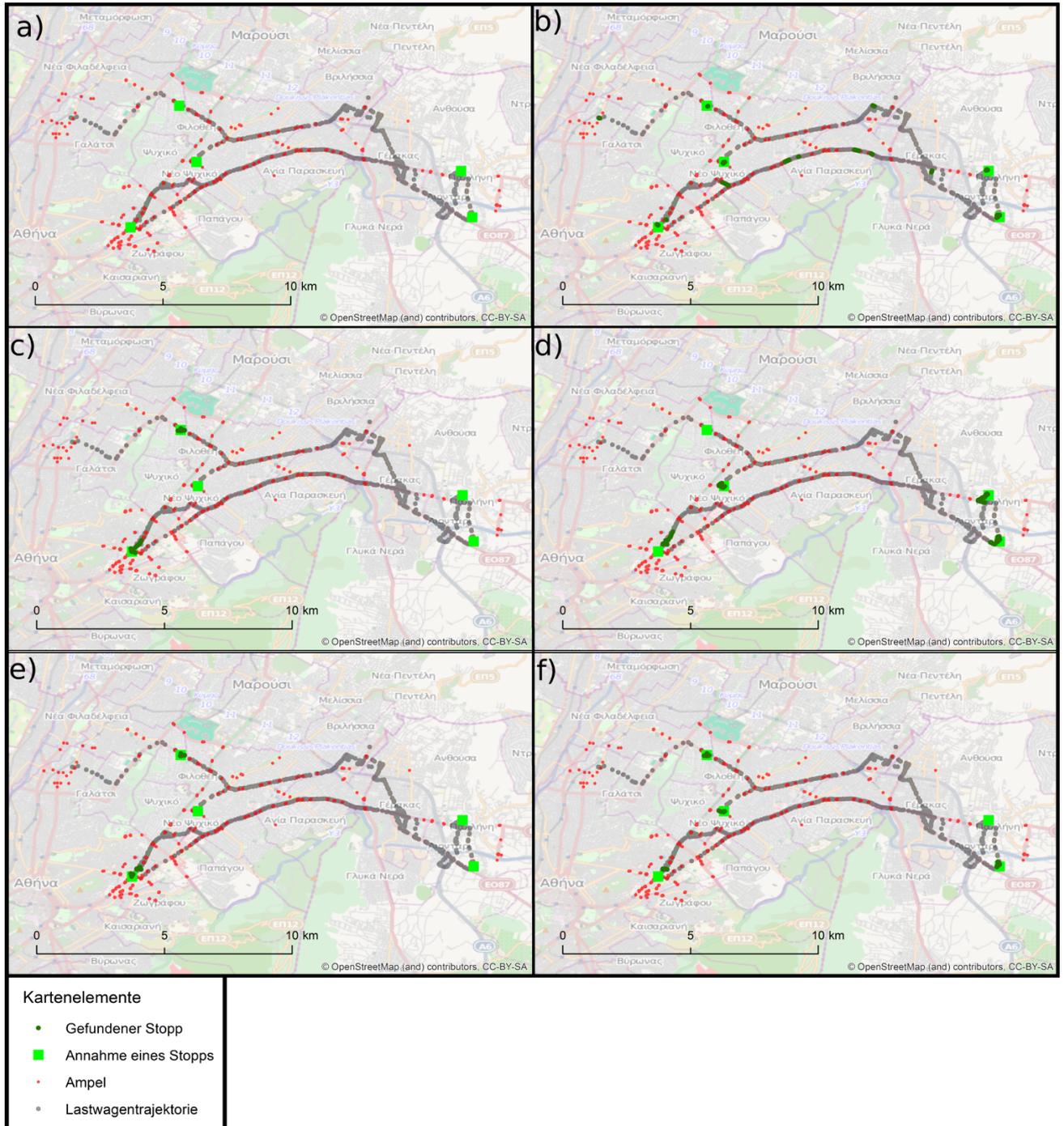


ABBILDUNG 5.39: Übersicht über die Trajektorie eines Lastwagens und die dabei gemachten Stopps laut den Algorithmen aus der Literatur. a) stellt einen Überblick über die Trajektorie und die gemachten Stopps dar, b) zeigt die Lösung des HMM, c) zeigt die echten Stopps des Algorithmus CB-SMoT, d) zeigt die unbekanntenen Stopps des Algorithmus CB-SMoT e) zeigt die Lösung des SMoT, f) zeigt die Lösung des DB-SMoT. Quelle Basiskarte: ©OpenStreetMap (and) contributors, CC-BY-SA.

**Trajektorie eines ausgewählten Austernfischers** Die Resultate zur detaillierten Analyse des Austernfischers sind in den Abbildungen 5.40 und 5.41 dargestellt. Es gilt

zu beachten, dass der Stopp, der sich am weitesten östlich auf der Karte befindet, lediglich aus einem Punkt besteht. Somit erstaunt es nicht, dass dieser Stopp von keiner der verwendeten Methoden gefunden wird.

Ein alleiniger Geschwindigkeitsschwellenwert in Abbildung 5.40 b) findet zwei richtige sowie auch drei falsche Stopps. Einer dieser falschen Stopps beinhaltet relativ viele Datenpunkte. Unter Einbezug einer Mindestdauer in Abbildung 5.40 c) wird ein falscher Stopp eliminiert, jedoch auch ein kleinerer, falscher Stopp dazu gefunden. Der Einbezug des Drehwinkels wie in Abbildung 5.40 d) eliminiert einen falschen Stopp, findet jedoch keinen echten Stopp mehr. Der Einbezug einer Mindestausdehnung wie in Abbildung 5.40 e) findet zwei richtige, jedoch auch einige falsche Stopps. Wird der Drehwinkel zu der+ Mindestausdehnung miteinbezogen, so werden wie in Abbildung Abbildung 5.40 f) zwei falsche Stopps weniger gefunden.

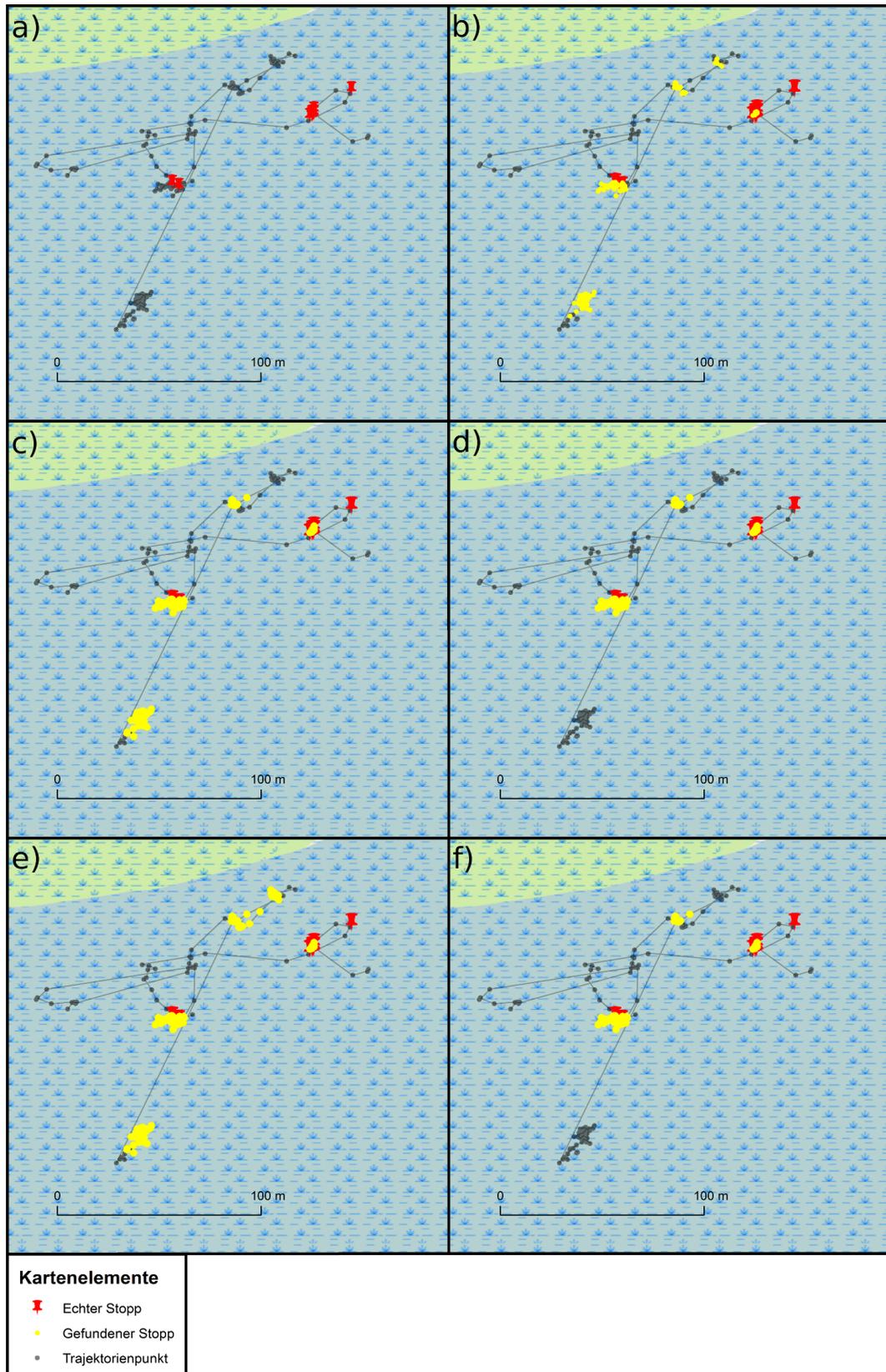


ABBILDUNG 5.40: Übersicht über die Trajektorie eines Austernfischers und die dabei gemachten Stopps laut verschiedenen Bewegungsparametern. a) stellt einen Überblick über die Trajektorie und die gemachten Stopps dar, b) zeigt die Lösung des Bewegungsparameters Geschwindigkeit, c) zeigt die Lösung der Bewegungsparameter Geschwindigkeit und Mindestdauer, d) zeigt die Lösung der Bewegungsparameter Geschwindigkeit, Mindestdauer und Drehwinkel, e) zeigt die Lösung der Bewegungsparameter Geschwindigkeit und Mindestausdehnung, f) zeigt die Lösung der Bewegungsparameter Geschwindigkeit, Mindestausdehnung und Drehwinkel. Quelle Basiskarte: ©OpenStreetMap (and) contributors, CC-BY-SA.

Die Resultate der Algorithmen aus der Literatur sind in Abbildung 5.41 dargestellt. Das Resultat des HMM in Abbildung 5.41 b) zeigt zwei richtige sowie zwei falsche Stopps, wobei einer davon relativ kleinräumig ist. Abbildung 5.41 c), welche das Resultat des CB-SMoT zeigt, findet ebenfalls zwei tatsächliche sowie zwei falsche Stopps. Der auf der Richtung basierende DB-SMoT findet dieselben richtigen Stopps wie die anderen Algorithmen sowie viele weitere Datenpunkte, die keine Stopps sind (Abbildung 5.41 d)).

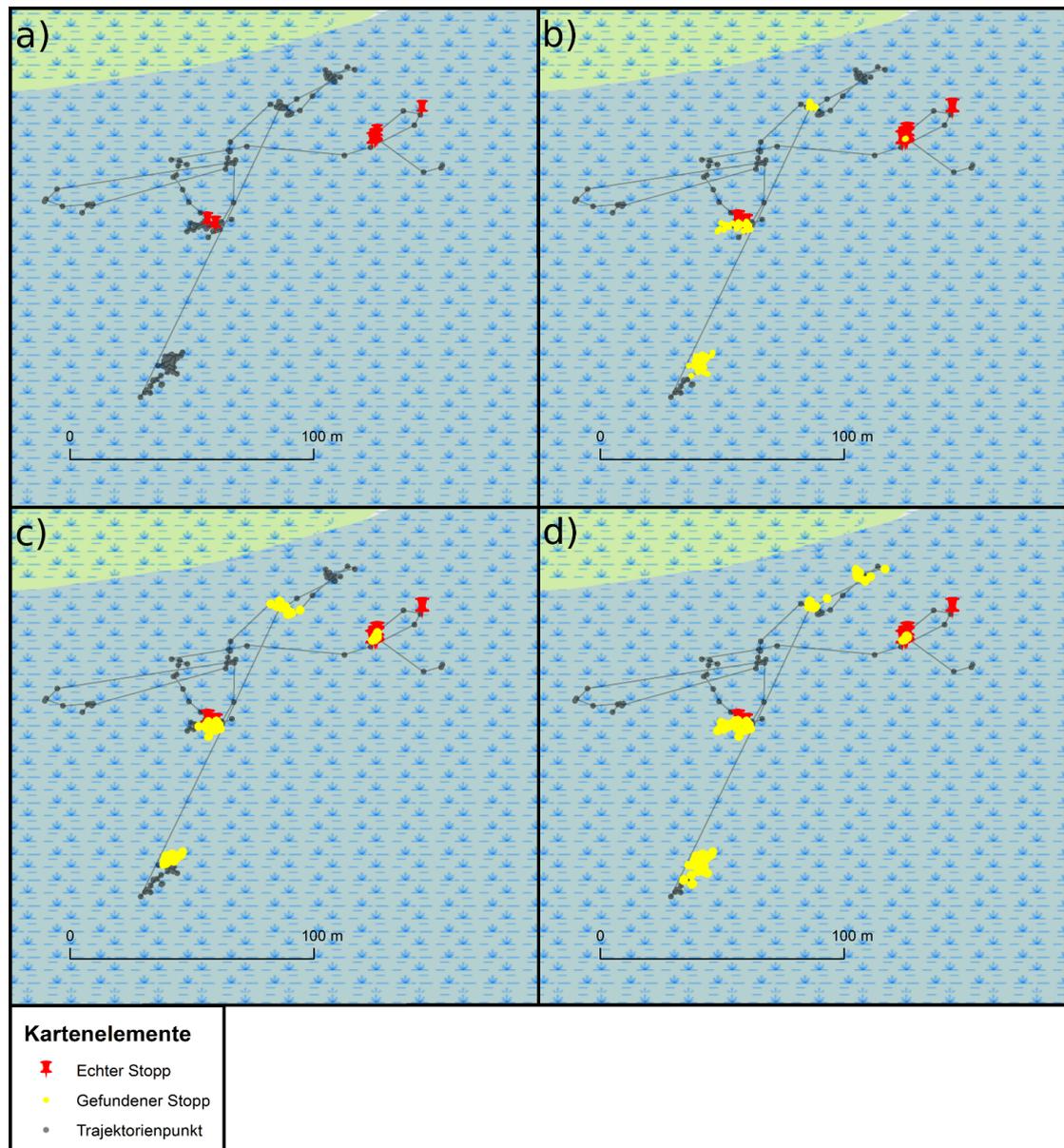


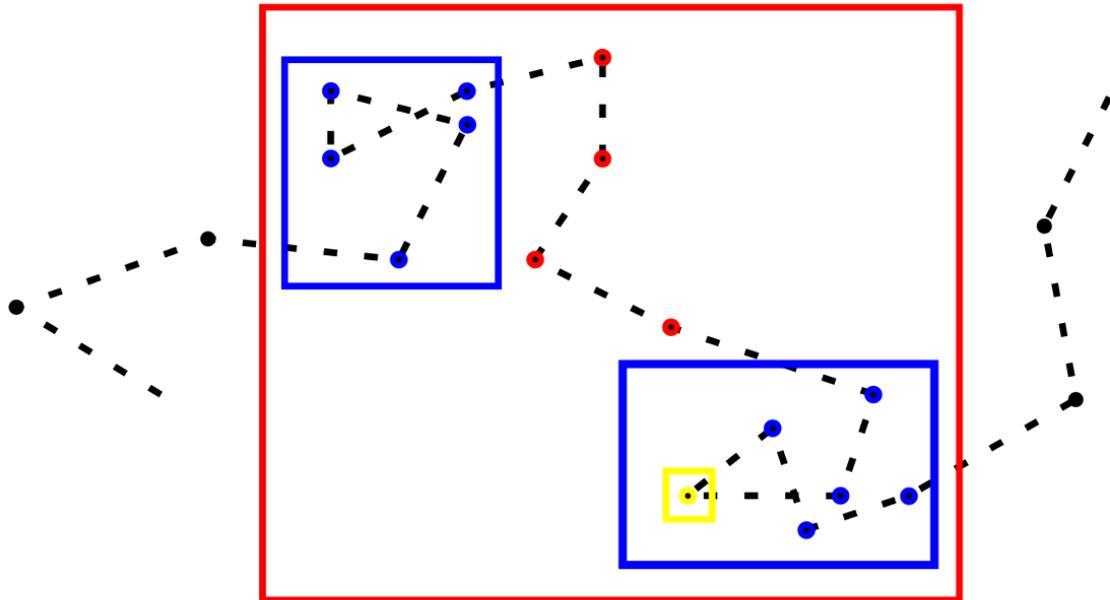
ABBILDUNG 5.41: Übersicht über die Trajektorie eines Austernfischers und die dabei gemachten Stopps laut den Algorithmen aus der Literatur. a) stellt einen Überblick über die Trajektorie und die gemachten Stopps dar, b) zeigt die Lösung des HMM, c) zeigt die unbekanntenen Stopps des Algorithmus CB-SMoT, d) zeigt die Lösung des DB-SMoT. Quelle Basiskarte: ©OpenStreetMap (and) contributors, CC-BY-SA.

# Kapitel 6

## Diskussion

Die in Kapitel 5 erhaltenen Resultate werden im Folgenden diskutiert. Wird von der allgemeinen Analyse gesprochen, so sind damit die Diagramme 5.1 bis 5.35 gemeint. Wird von der Analyse der ausgewählten Trajektorien gesprochen, so sind die Kartenprodukte der Abbildungen 5.36 bis 5.41 gemeint.

Bei der Betrachtung der Diagramme der allgemeinen Analyse fällt auf, dass die Anzahl an Stopps mit steigenden Schwellenwerten sowohl zu- als auch abnimmt. Dies ist so, weil sich mehrere Stopps ab einem gewissen Schwellenwert zu einem Stopp aggregieren können. Abbildung 6.1 zeigt diesen Sachverhalt auf. So ist ersichtlich, dass in dieser Abbildung mit einem Geschwindigkeitsschwellenwert von 1 km/h und mit einem Geschwindigkeitsschwellenwert von 10 km/h dieselbe Anzahl Stopps resultiert. Bei einem mittleren Geschwindigkeitsschwellenwert von 5 km/h resultiert die grösste Anzahl Stopps. Es kann davon ausgegangen werden, dass in dem Bereich, in dem die meisten Stopps gefunden werden, die tatsächlichen Stopps wahrscheinlich enthalten sind, jedoch auch einige falsche Stopps.



- Punktgeschwindigkeit < 1km/h
- Punktgeschwindigkeit < 5km/h
- Punktgeschwindigkeit < 10km/h
- Stopp mit Geschwindigkeitsschwellenwert = 1km/h
- Stopp mit Geschwindigkeitsschwellenwert = 5km/h
- Stopp mit Geschwindigkeitsschwellenwert = 10km/h

ABBILDUNG 6.1: Verschiedene Aggregationsstufen von einzelnen Datenpunkten zu einem oder mehreren Stopps.

Weiter kann im Allgemeinen über alle Methoden gesagt werden, dass dieselben Algorithmen bei unterschiedlichen Datensätzen auch charakteristisch unterschiedliche Resultate liefern. Somit hängt ein Resultat nicht alleinig von dem verwendeten Algorithmus ab, sondern auch von den Daten, auf die er angewendet wird sowie deren Struktur. Die Daten wiederum werden unter anderem charakterisiert durch das sich bewegende Objekt, welches sie wiedergeben. Somit reagieren die hier untersuchten Algorithmen sensibel auf das sich bewegende Objekt, welches untersucht wird. Dies zeigt sich insbesondere bei den Daten des Austernfischers.

Die Resultate zum Austernfischer zeigen unter der Analyse mit allen Methoden Schwankungen, im Gegensatz zu anderen Diagrammen, in denen die Kurven gleichmässiger

sind. Diese starken Schwankungen bei den Austernfischerdaten könnten darauf zurück zu führen sein, dass es sich hierbei um zeitlich sehr fein aufgelöste Daten handelt, die sich in einem sehr kleinen räumlichen Bereich bewegen. Somit stellt sich die Frage, ob diese sehr feine zeitliche Auflösung einen Mehrwert darstellt oder dadurch das Resultat vermehrt Unsicherheiten beinhaltet. Diese Thematik wird von Laube und Purves (2011) thematisiert, wobei die Autoren herausgefunden haben, dass Messungen mit einem feinen zeitlichen Massstab durch Unsicherheiten maskiert werden. Dabei gilt dieser Umstand sowohl für die Bewegungsparameter Geschwindigkeit, Sinuosität einer Trajektorie und Drehwinkel. Im Rahmen dieser Arbeit wurden lediglich die Geschwindigkeit und der Drehwinkel verwendet. Bei der Analyse der zeitlich sehr fein aufgelösten Daten sind auch hier grosse Schwankungen charakteristisch und könnten somit auf Unsicherheiten zurückzuführen sein.

In den folgenden Abschnitten werden die einzelnen Algorithmen und die erhaltenen Resultate einzeln betrachtet. Dabei werden sowohl allgemeine Charakteristika der einzelnen Methoden erläutert als auch wie sie im Bezug auf die verschiedenen Datensätze reagiert haben.

## 6.1 Bewegungsparameter

### 6.1.1 Geschwindigkeit

Die Geschwindigkeit als alleinigen Parameter zu verwenden führt meistens nicht zu sehr aufschlussreichen Resultaten, dieses Vorgehen wird auch in der Literatur äusserst selten gewählt (siehe dazu Tabelle 4.3). Meistens steigt die Anzahl Stopps mit einem steigenden Geschwindigkeitsschwellenwert an bis zu einem Maximum und nimmt danach wieder ab (Diagramme 5.1 bis 5.5). Es kann zwar ein Maximum an Stopps abgelesen werden, jedoch kann nicht davon ausgegangen werden, dass dieses Maximum an Stopps auch der Anzahl tatsächlich gemachter Stopps entspricht. Wie in Abbildung 6.1 zu sehen, ist davon auszugehen, dass die einzelnen Stopps im Bereich dieses Maximums relativ kleinräumig und kurzzeitig sind. Weiter ist die Anzahl an Stopps bei den meisten Datensätzen grundsätzlich hoch, insbesondere bei den Eye-Tracking-Daten in Diagramm

5.5, was aufgrund der Einfachheit dieses Algorithmus und des Einbezugs nur eines Parameters zu erwarten ist. Anhand der Analyse der ausgewählten Trajektorien ist ersichtlich, dass bei den Wander- und den Lastwagendaten (Abbildungen 5.36 und 5.38) der ideale Geschwindigkeitsschwellenwert die gemachten Stopps zuverlässig findet. Es werden in beiden Fällen weitere Regionen als Stopp definiert, welche keine sind. Dieser Umstand spricht dafür, dass auch die Anzahl der gefundenen Stopps in der allgemeinen Analyse zu hoch ist. Im Falle der Austernfischerdaten (Abbildung 5.40 b)) werden die Stopps mit diesem einfachen Algorithmus nur teilweise gefunden, wobei auch einige andere Regionen als Stopp definiert werden. Shamoun-Baranes et al. (2012) analysieren in ihrer Arbeit Daten von Austernfischern anhand verschiedenener Modelle, wobei eines davon auf Daten aus Beschleunigungsmessern basiert und ein zweites alleinig auf Angaben zur Geschwindigkeit. Auch in ihrer Analyse ist ersichtlich, dass das Modell, welches alleinig auf der Geschwindigkeit basiert weniger akkurat ist und einen grösseren Fehlerwert in der Validierung aufweist (*cross-validation error* mehr als doppelt so hoch als bei der Analyse der Daten aus dem Beschleunigungsmesser).

Es kann davon ausgegangen werden, dass mit einem Geschwindigkeitsschwellenwert zwar die tatsächlichen Stopps, mehr oder weniger genau abgegrenzt, gefunden werden, sowie auch falsche Stopps. Kennt man den idealen Schwellenwert, so können die tatsächlichen Stopps relativ gut gefunden werden. Im Falle dieser Arbeit konnte dieser ideale Schwellenwert jedoch erst durch die Visualisierung der Resultate und mit Hilfe der *ground truth* eruiert werden. Falls die *ground truth* vorhanden ist und die Stopps bekannt sind, stellt sich die Frage, wieso nach den Stopps gesucht werden soll. Ohne *ground truth* wiederum stellt sich die Detektion der Stopps mit einem einfachen Geschwindigkeitsschwellenwert als schwierig heraus. Die Geschwindigkeit liefert somit zwar einen ersten Anhaltspunkt für die Detektion von Stopps, führt jedoch als alleiniger Parameter ohne Wissen über die *ground truth* zu keinen genauen Resultaten. Dies könnte dadurch zu Stande kommen, dass die Geschwindigkeit nicht nur vom Bewegungsverhalten, sondern auch von sonstigen Faktoren abhängig ist.

**Wanderdaten** Eine wandernde Person ist langsamer, wenn sie aufwärts läuft oder eine heikle Passage, wie beispielsweise ein Flussbett, überwinden muss, ohne dass hier gestoppt wird. Insbesondere in steilem Gelände oder in dichten Wäldern, in denen das GPS-Signal stärker gestört ist, scheint die Geschwindigkeit alleine nicht zuverlässig zu

sein, um Stopps zu detektieren. Es gilt zu beachten, dass es, insbesondere bei Daten zu Fussgängern, schwierig ist, sehr langsame Bewegungen von Ungenauigkeiten des GPS-Signal zu unterscheiden, wie Orellana und Wachowicz (2011) argumentieren. Im Diagramm 5.1 ist ersichtlich, dass sich die Anzahl an Stopps bei hohen Schwellenwerten zwar bei ungefähr zwei stabilisiert, hierbei sind jedoch bereits über 80 % der Punkte als Stopp definiert, was wenig aufschlussreich ist.

**Lastwagendaten** Bei motorisierten Objekten in einer Stadt, wie dies bei den Lastwagen der Fall ist, kann eine tiefe Geschwindigkeit ebenfalls durch viele Faktoren bedingt sein. So spielen Verkehrsschilder, Baustellen, Unfälle und die anderen Verkehrsteilnehmer und Verkehrsteilnehmerinnen eine wichtige Rolle. Deshalb bietet auch hier die Geschwindigkeit alleine zu wenig Potential bei der Detektion von Stopps.

**Austernfischerdaten** Bei den Austernfischerdaten entspricht die Anzahl an Stopps in Diagramm 5.3 bei einem Geschwindigkeitsschwellenwert von 0.2 km/h zwar relativ gut der Anzahl angenommener Stopps, die bei neun liegt, jedoch sind ähnlich wie bei den Wanderdaten hier auch sehr viele Punkte als Stopp definiert. Somit ist unwahrscheinlich, dass hierbei die richtigen Stopps gefunden wurden. Weiter schwanken die Resultate relativ stark, was auf Unsicherheiten hindeutet.

**Truthahngeierdaten** Es ist offensichtlich, dass ein Vogel eine minimale Geschwindigkeit haben muss, um überhaupt in der Luft zu bleiben, weshalb in einer ersten Überlegung die Geschwindigkeit als passender Bewegungsparameter angesehen werden kann. In einer Arbeit von Bruderer und Boldt (2001) wurden die Fluggeschwindigkeiten von einigen Vögeln untersucht, wobei sich die meisten davon in einem Rahmen von 10 - 15 m/s bewegen. Im Fall der hier verwendeten Truthahngeierdaten, wobei teilweise nur alle vier Stunden ein Messpunkt vorhanden ist, weicht eine berechnete Geschwindigkeit stärker von der tatsächlichen Geschwindigkeit ab als bei einer feineren zeitlichen Auflösung. Weiter verhält es sich so, dass die Zugvögel auch während eines Stopps fliegen, jedoch kürzere Distanzen. Dies kann beispielsweise der Fall sein, wenn die Vögel Futter suchen oder einen geeigneten, geschützten Platz zur Erholung.

**Eye-Tracking-Daten** Im Rahmen der Eye-Tracking-Daten in Diagramm 5.5 werden sehr viel mehr Stopps gefunden, als tatsächlich vorhanden sind. Interessant ist, dass bei der maximalen Anzahl Stopps erst 40% der Punkte als Stopp definiert sind. Dies steht im Gegensatz zu den anderen Datensätzen, bei denen beim Maximum an Stopps bereits 60% der Punkte als Stopp definiert sind. Dieses unterschiedliche Verhalten der Eye-Tracking-Daten könnte darauf zurück zu führen sein, dass bei diesen Daten der relative Unterschied in der Geschwindigkeit zwischen Fixationen und Sakkaden im Verhältnis zu den anderen Datensätzen grösser ist. Da eine Bewegung des Auges praktisch keine Masse verschieben muss, sind viel höhere Geschwindigkeiten möglich als beispielsweise bei einem Lastwagen von mehreren Tonnen Gewicht.

### 6.1.2 Drehwinkel

Der Einbezug des Drehwinkels reduziert in allen Fällen die maximale Anzahl an Stopps. Die mengen- und verhältnismässig grösste Reduktion findet bei den Eye-Tracking-Daten statt (Diagramm 5.10). Dieser Umstand kann darauf zurückzuführen sein, dass sehr schnell spitze Drehwinkel entstehen können, da sich ein Auge sehr schnell bewegen kann. Die Resultate aller Datensätze sind zudem sehr unruhig, was Unsicherheiten in den Daten darstellen könnte. So weisen die Diagramme 5.6 bis 5.10 grössere Schwankungen auf im Vergleich zu den Diagrammen 5.1 bis 5.5. Es kann davon ausgegangen werden, dass durch eine Kombination der Bewegungsparameter Geschwindigkeit und Drehwinkel grundsätzlich zu wenige Stopps gefunden werden, wobei die Resultate bei keinem der Datensätze überzeugend sind. Laube und Purves (2011) haben in ihrer Arbeit festgestellt, dass Analysen mit einer sehr feinen zeitlichen Auflösung mehrheitlich Unsicherheiten wiedergeben können. Es kann sein, dass im Fall der hier verwendeten Daten und einer Kombination eines Schwellenwertes für die Geschwindigkeit und für den Drehwinkel diese Problematik auch bei den zeitlich weniger fein aufgelösten Datensätzen auftritt.

Der Drehwinkel scheint als Hilfsparameter, wie in den Analysen der ausgewählten Trajektorien ersichtlich, idealer zu sein als als alleiniger Parameter beziehungsweise in alleiniger Kombination mit dem Algorithmus der Geschwindigkeit. Der Drehwinkel wird in der Literatur meistens in Kombination mit anderen Bewegungsparametern verwendet, beispielsweise in Moreno et al. (2010) oder auch in Waga et al. (2012). Dabei werden entweder die Stopps selber genauer angeschaut (Moreno et al., 2010) oder es werden

mehrere Bewegungszustände unterschieden (Waga et al., 2012). Somit scheint der Bewegungsparameter des Drehwinkels, wie auch hier ersichtlich, für diese Anwendungen besser geeignet zu sein als für die bloße Detektion von Stopps.

**Allgemeine Analyse** Interessanterweise scheinen die Daten zu den Zugvögeln (Diagramm 5.9) nicht auf den Algorithmus des Drehwinkels anzusprechen, obwohl die Richtungsänderung hierbei eine essentielle Rolle spielt. Durch das grobe zeitliche Aufnahmeintervall sowohl während der Stopps als auch während des Fluges könnte die Datenstruktur für diese Besonderheit verantwortlich sein. Grundsätzlich, so argumentieren Rocha et al. (2010), eignen sich auf der Richtungsänderung basierende Methoden ideal für Anwendungen, bei denen Variationen in der Richtung die wesentliche Rolle spielen. Auch Buchin et al. (2013) beziehen sich auf die Richtungsänderungen für die Analyse von Daten zu Zugvögeln. Es ist möglich, dass die Kombination der Geschwindigkeit und des Drehwinkel suboptimal ist und die Resultate dieses Parameters deshalb hier nicht überzeugend sind.

**Analyse der ausgewählten Trajektorien** Der Vorteil des Drehwinkels in der Analyse der ausgewählten Trajektorien ist von Fall zu Fall unterschiedlich. Der Einbezug des Drehwinkels stellt hierbei somit nur teilweise einen Mehrwert dar, da sowohl gesuchte als auch falsche Stopps eliminiert werden. Nimmt man in Kauf, dass auch falsche Stopps gefunden werden, solange die wichtigen beziehungsweise gesuchten Stopps höchstwahrscheinlich auch gefunden werden, eignet sich der Einbezug des Drehwinkels nicht. Möchte man jedoch möglichst wenige falsche Stopps finden, mit dem Risiko, auch einige echte Stopps zu eliminieren, so stellt der Einbezug des Drehwinkels eine Option dar.

### 6.1.3 Dauer

Unter Einbezug einer Mindestdauer findet im Vergleich zur Verwendung eines reinen Geschwindigkeitsschwellenwertes ebenfalls eine Reduktion der Anzahl an Stopps statt (Abbildungen 5.11 bis 5.15). Der Wertebereich der Stopps befindet sich, mit Ausnahme der Eye-Tracking-Daten, in einem ähnlichen Bereich wie unter Anwendung des Algorithmus des Drehwinkels. Die Werte schwanken jedoch merklich weniger und es zeichnen sich

glattere und regelmässige Kurven ab. Dies könnte darauf hindeuten, dass die Resultate nicht Unsicherheiten und Rauschen wiedergeben, sondern tatsächliche Stopps. Auch in der Literatur wird die Dauer in einigen Analysen verwendet, so bereits bei Collins et al. (1995). Auch Spaccapietra et al. (2008) nennen als Voraussetzung eines Stopps, dass dieser eine gewisse zeitliche Ausdehnung aufweisen muss.

Bei allen Datensätzen ist zudem ersichtlich, dass sich die Anzahl der Stopps mit zunehmender Mindestgeschwindigkeit stabilisiert. Auch die Austernfischerdaten in Diagramm 5.13, welche bei den Algorithmen Geschwindigkeit und Drehwinkel unruhige und unregelmässige Resultate lieferten, stabilisieren sich unter Einbezug der Mindestdauer für einen Stopp. Durch diese Stabilisierung und anhand der Resultate der ausgewählten Trajektorien kann davon ausgegangen werden, dass dieser Algorithmus der Realität näher kommt als ein reiner Geschwindigkeitsschwellenwert und einen Überblick über die Stopps einer Trajektorie gibt.

**Wander- und Lastwagendaten** Insbesondere bei der allgemeinen Analyse der Wanderdaten sowie der Lastwagendaten zeigen sich sehr deutliche Resultate. So ist bei einer gewissen Mindestdauer eine klare Reduktion der Anzahl an Stopps sichtbar (Diagramme 5.11 und 5.12). Die allgemeine Analyse der Wanderdaten zeigt zudem, dass bereits eine Mindestdauer von zwei Minuten die Anzahl an Stopps deutlich reduziert. Insgesamt kommen die meisten Kurven bei einem Geschwindigkeitsschwellenwert von 2 km/h den zwei echten Stopps relativ nahe. Bei der Analyse der ausgewählten Trajektorie des Wanderers in Abbildung 5.36 c) nähert sich das Resultat den tatsächlich gemachten Stopps gut an, hat aber noch einen weiteren, 'falschen' Stopp zu Beginn der Trajektorie. Man könnte jedoch argumentieren, dass bei der Analyse einer Trajektorie ein Stopp zu Beginn oder zu Ende der Trajektorie individuell berücksichtigt werden sollte. Es scheint in den meisten Fällen so, dass ein Stopp an diesen Stellen unwahrscheinlich oder im Bezug auf die Thematik der Datensätze uninteressant ist.

Auch bei der Analyse des ausgewählten Lastwagens in Abbildung 5.38 c) liefert der Algorithmus der Mindestdauer die tatsächlichen Stopps ohne falsche Stopps miteinzubeziehen. Einen interessanten Unterschied gibt es dennoch zwischen den Wander- und den Lastwagendaten. Der Algorithmus der Mindestdauer funktioniert bei den Lastwagendaten besser als bei den Wanderdaten. Dies könnte dadurch zustande kommen, dass

Stopps eines Lastwagens durch deren Zweck eine gewisse Mindestdauer haben. Die Lastwagen müssen mindestens so lange stoppen, bis der Beton abgeladen ist. Eine gewisse Stoppdauer kann somit gar nicht unterschritten werden. Wohingegen ein Stopp auf einer Wanderung nur sehr kurz sein kann, auf Grund von schlechtem Wetter oder weil die Wandernden in Zeitnot sind. Die Stoppdauer ist in diesem Fall frei wählbar und somit nicht zwingender Indikator für einen Stopp.

**Austernfischerdaten** Bei der Analyse der ausgewählten Trajektorie des Austernfischers in Abbildung 5.40 c) ist sichtbar, dass zwei Stopps gefunden werden und der Mehrwert gegenüber einem Geschwindigkeitsschwellenwert in der Elimination eines einzelnen, falschen Stopps liegt. Im Gegensatz zu den Wander- und den Lastwagendaten funktioniert der Algorithmus bei den Austernfischerdaten möglicherweise schlecht, weil hier Stopps tatsächlich nur sehr kurz sein können. In der *ground truth* der Trajektorie kommt es beispielsweise vor, dass lediglich ein einziger Punkt als Stopp definiert wird und die vorangegangenen und nachfolgenden Punkte als Bewegung am Boden.

**Truthahngeierdaten** Die Truthahngeierdaten reagieren bei niedrigen Werten für die Mindestdauer relativ konstant und einheitlich auf diesen Algorithmus (Diagramm 5.14). Je höher die Mindestdauer gesetzt wird, umso flacher wird die Kurve und umso kleiner werden die relativen Unterschiede in der Anzahl an Stopps zwischen den einzelnen Kurven. So ist die Reduktion an Stopps zwischen einer Mindestdauer von fünf Stunden und zehn Stunden ähnlich gross wie zwischen einer Mindestdauer von 40 Stunden und 80 Stunden. Dies spricht dafür, dass sich das Resultat bei höheren Mindestdauerwerten stabilisiert und in diesem Bereich somit die Stopps robuster werden gegenüber variierenden Schwellenwerten. Dieser Umstand wird gestützt durch eine Analyse von Buchin et al. (2013), wobei die Autoren und Autorinnen Stopps in Trajektorien von Zugvögeln suchen. Sie verwenden das Kriterium der Mindestdauer, wobei sie eine Mindestdauer von 48 Stunden voraussetzen für das Vorkommen eines Stopps. Dieser scheinbar kritische Bereich der Mindestdauer konnte auch hier herausgelesen werden und stellt somit ein wichtiges Kriterium bei der Analyse von Zugvögeln dar.

**Eye-Tracking-Daten** Auch die Eye-Tracking-Daten in Diagramm 5.15 zeigen eine relativ gleichmässige Reaktion auf diesen Algorithmus, wobei zwischen einer Mindestdauer von 100 Sekunden und 200 Sekunden ein grosser Abstand besteht. Die Anzahl tatsächlicher Stopps wird mit einer Mindestdauer von 200 Sekunden und einer Geschwindigkeit unter 5 px/s dann auch relativ gut gefunden. Das deutet darauf hin, dass eine Dauer von 200 Sekunden ein kritischer Wertebereich ist. Mit den idealen Werten für den Geschwindigkeitsschwellenwert und die Stoppdauer können die tatsächlichen Stopps in Eye-Tracking-Daten somit detektiert werden.

#### 6.1.4 Ausdehnung

In Abschnitt 2.4 wurde angesprochen, dass verschiedene hier verwendete Methoden auf ein- und derselben Physik beruhen. Vergleicht man die Algorithmen der Mindestdauer eines Stopps und der Mindestausdehnung eines Stopps, so sollten auf Grund dieser Überlegung ähnliche Resultate resultieren. Die Daten zum Austernfischer zeigen bei der Anwendung von Ausdehnungsschwellenwerten in Diagramm 5.18 grössere Schwankungen als bei der Anwendung eines Schwellenwertes für die Stoppdauer in Diagramm 5.13. Dieser Umstand zeigt, dass den beiden Algorithmen zwar die gleiche Physik zugrunde liegen kann, sie deshalb aber nicht zwingend gleiche Resultate liefern. Dies kommt von daher, dass die Resultate nicht nur von der Physik einer Bewegung abhängen, sondern auch vom Stichprobenintervall und der Messunsicherheit, mit denen die Bewegung aufgezeichnet wird. So geht die physikalische Theorie von einer kontinuierlichen Bewegung aus. Die hier verwendeten Daten jedoch stellen eine Diskretisierung dieser Bewegung dar, da die Bewegung nur in einem gewissen Intervall abgebildet ist.

In der Analyse der ausgewählten Trajektorie des Wanderers (Abbildung 5.36 e)) findet der Algorithmus der Mindestausdehnung die Stopps zuverlässiger als die Mindestdauer. Dies ist bei den Lastwagendaten gerade umgekehrt. Hierbei findet die Mindestdauer die Stopps sehr zuverlässig, die Mindestausdehnung jedoch definiert einige falsche Punkte als Stopp.

**Lastwagendaten** Der Bewegungsparameter der Mindestausdehnung könnte für die Lastwagendaten schlechtere Resultate liefern als die Mindestdauer, da ein Lastwagen möglicherweise mehrere Orte auf ein- und derselben Baustelle bedient. In diesem Fall

bewegt er sich für eine längere Zeit langsam, da er auf der Baustelle manövriert, hat aber dennoch einen relativ grossen Bewegungsumfang, da die Baustelle auch eine gewisse Grösse hat. Dies spricht dafür, dass der zeitliche Aspekt für die Lastwagendaten wichtiger ist als beispielsweise für die Wanderdaten. Die Lastwagen sind in der Ausdehnung ihrer Bewegung stärker eingeschränkt, insbesondere auch auf grossen Strassen oder in Gebieten mit vielen Verkehrssymbolen. Bak et al. (2012) haben auch konsekutive Punkte mit einer tiefen Geschwindigkeit innerhalb einer Distanz als Stopps definiert werden. In ihrer Arbeit haben die Autoren und Autorinnen jedoch nicht nur eine Trajektorie berücksichtigt. So konnten Regionen gefunden werden, die von mehreren Trajektorien als Stopp frequentiert werden. Diese Vorgehensweise hätte auch bei den Lastwagendaten gewählt werden können, da hier mehrere Trajektorien vorhanden sind. Es ist möglich, dass der Algorithmus der Mindestausdehnung in diesem Fall robustere Resultate geliefert hätte, da mehr Informationen berücksichtigt worden wären.

**Wanderdaten** Für wandernde Personen spielt die Umgebung eine kleinere Rolle als für einen Lastwagen. In der Regel versperren einer wandernden Person keine grossen Hindernisse den Weg und die Ausdehnung der Bewegung kann frei gewählt werden. Das kann ein Grund dafür sein, dass die Mindestausdehnung gute Resultate liefert. Ein weiterer Grund für diese Besonderheit könnte sein, dass sich eine wandernde Person während eines Stopps tatsächlich für eine gewisse Zeit nicht bewegen kann, da sie beispielsweise seinen Rucksack öffnet, Proviant herausnimmt oder die Karte konsultiert. Diese Verhaltensweisen sind während einer Bewegung nur mühsam machbar, weshalb der Standort für dessen Ausführung, wenn auch nur für eine kurze Zeit, exakt gleich bleibt.

**Truthahngeierdaten** Die positive Reaktion der Truthahngeierdaten auf diesen Algorithmus weist darauf hin, dass der Algorithmus sich für diese Daten eignet, was nicht erstaunt. Vergleicht man die Distanzen, die ein Vogel während des Fluges zurücklegt, mit dem Radius, in dem er sich während eines Stopps befindet, so resultieren sehr verschiedene Grössenordnungen. Es ist also nachvollziehbar, dass dieser Bewegungsparameter aussagekräftig ist. Auch Buchin et al. (2013) verwenden unter anderem eine Form der Ausdehnung (*location disk criterion*), um bei Zugvögeln die Stopps zu detektieren, was, wie hier, für diesen Bewegungsparameter spricht.

**Austernfischerdaten** Mit Hilfe einer Mindestausdehnung können die Austernfischerdaten nur unzureichend analysiert werden. Im Vergleich zum Algorithmus der Mindestdauer werden in der Analyse der ausgewählten Trajektorie (Abbildungen 5.40 c) und e)) sogar noch mehr falsche Punkte als Stopp bezeichnet. Dieser Umstand kann darauf zurückzuführen sein, dass die Daten zum Austernfischer sehr kleinräumig sind. Somit wird relativ schnell eine hohe Ausdehnung erreicht, ohne dass ein Stopp eingelegt wurde.

**Eye-Tracking-Daten** In den Eye-Tracking-Daten (Diagramm 5.20) werden mit einer Mindestausdehnung, bei gleichem Geschwindigkeitsschwellenwert, im Vergleich zu einer Mindestdauer sehr viel mehr Stopps gefunden. Auf der Höhe der Anzahl tatsächlicher Stopps (1160 Stopps) scheint keine Mindestausdehnung sensibel zu reagieren, dieser Bereich wird nicht erkannt. Da sich das Auge sehr schnell bewegen kann, können auch sehr schnell grosse Gebiete angeschaut werden, weshalb mit einer Mindestausdehnung viele Stopps resultieren. Die Dauer scheint hierbei der idealere Parameter zur Detektion von Stopps zu sein. Das ist intuitiv, da nur durch eine genügend lange Zeitspanne einer Fixation auch ein Inhalt aus einer Karte gelesen werden kann. Eine grosse Ausdehnung eines Stopps könnte hingegen auch darauf hinweisen, dass das Auge an dieser Stelle etwas sucht, ohne konkret einen spezifischen Punkt zu fixieren.

## 6.2 Algorithmengruppe Stops and Moves of Trajectories

### 6.2.1 Stops and Moves of Trajectories

Eine genaue Beschreibung des Algorithmus SMoT (Alvares et al., 2007) ist in Abschnitt 4.4 zu finden. Anhand der visuellen Analyse in Abbildung 5.39 e) ist ersichtlich, dass bei der Verwendung von Kandidatenregionen je nach Informationslage Vorsicht geboten ist. Der gewählte Lastwagen macht auch Stopps an Orten, an denen keine Ampeln vorkommen. Dies kann darauf zurückzuführen sein, dass auf Baustellen oder Betonladestationen keine Lichtsignale vorhanden sind, weil keine Lichtsignale nötig sind. Eine Erweiterung der Kandidatenregionen könnte hierbei nützlich sein. Eine solche Anpassung ist jedoch nicht immer möglich, da nicht immer detaillierte Informationen über alle möglichen Kandidatenregionen vorhanden sind. Sind die Kandidatenregionen relativ offensichtlich und einfach detektierbar, so ist der SMoT eine gute Variante, um Stopps zu finden.

Weiter ist es möglich, dass je nach Applikation alle Kandidatenregionen oder aber nur eine Auswahl davon aufgesucht werden. Wird nur eine gewisse Anzahl an Kandidatenregionen frequentiert und es ist weder bekannt welche Regionen das sind, noch wie viele es sind, so scheint der SMoT eine gute Möglichkeit zu sein, diese Regionen zu eruieren. Somit ist der Algorithmus SMoT passend, um herauszufinden, welche der potentiellen Kandidatenregionen besucht werden und welche ausgelassen werden. Im Rahmen der Lastwagendaten gibt es nicht eine grosse Auswahl an Kandidatenregionen. Dazu gehören die Ladestation für den Beton und die zu beliefernden Baustellen. Falls diese Regionen bereits bekannt sind, stellt sich die Frage welcher Mehrwert besteht, wenn die Stopps darin gesucht und gefunden werden.

Es kann davon ausgegangen werden, dass bereits im Voraus, beispielsweise aufgrund der Arbeitszeit des Lastwagenfahrenden, definiert wird, welcher Lastwagen welche Baustelle bedient. Somit stellt sich die Frage nach dem Zweck einer Analyse mit dem SMoT für Objekte, welche wenige Kandidatenregionen haben oder diese nicht nach freier Wahl besuchen oder auslassen können. Weiter können mit dieser Methode Informationen über eine Trajektorie und die bekannten Kandidatenregionen gewonnen werden, nicht jedoch über allfällige andere interessante Regionen. Diese Kritik äussern auch Palma et al. (2008) und liefern den CB-SMoT als Alternative zum SMoT.

Es gilt zu beachten, dass der SMoT eine Implementation der Geschwindigkeit ist, da von einem Radius und einer Zeitspanne innerhalb dieses Radius ausgegangen wird. Dies führt zu der in Abschnitt 2.4 diskutierten Problematik. Somit stellt sich die Frage, welchen Mehrwert dieser Algorithmus gegenüber einem einfachen Geschwindigkeitsschwellenwert bietet. Insbesondere, wenn die Erfassung der Kandidatenregionen noch aussteht, gilt es meines Erachtens, den Aufwand und den Ertrag pragmatisch abzuwägen.

### **6.2.2 Clustering-Based Stops and Moves of Trajectories**

Moreno et al. (2010) argumentieren, ein Problem des CB-SMoT von Palma et al. (2008) sei, dass die unbekanntes Stopps nicht weiter untersucht werden. So könne bei der Detektion eines unbekanntes Stopps eines Tieres nicht gesagt werden, ob das Tier dabei frisst oder sich ausruht. Die Autoren und Autorinnen liefern als Alternative dazu eine Methode, die die Stopps genauer untersucht und so versucht, herauszufinden, was das Ziel des gemachten Stopps ist. Die grosse Anzahl an gefundenen Stopps des CB-SMoT

im Rahmen der hier gemachten Untersuchungen sprechen ebenfalls dafür, dass eine genauere Analyse dieser Stopps aufschlussreich sein könnte.

Die allgemeine Analyse zeigt, dass Stopps mit einer idealen Kombination aus der Berechnung der Nachbarschaft und der Mindestdauer für einen Stopp bei vielen Datensätzen gefunden werden. Einzig die Eye-Tracking-Daten in Diagramm 5.25 können mit diesem Algorithmus nicht ideal analysiert werden. Der CB-SMoT scheint somit für Eye-Tracking-Daten nicht geeignet zu sein.

Für die restlichen Datensätze scheint die Verwendung eines Cluster-Algorithmus bei der Detektion von Stopps jedoch eine sinnvolle Methode zu sein. Ein Vorteil dieser Methode ist, dass sie nur implizit auf der Geschwindigkeit beruht, also über eine Ausdehnung pro Mindestdauer, und nicht von den Geschwindigkeitsangaben der Datensätze abhängig ist. Diese Methode kann auf jegliche Daten angewendet werden, ohne dass dazu eine Geschwindigkeitsangabe nötig ist oder berechnet werden muss.

Ein weiterer Vorteil dieser Methode ist, dass auch Datensätze untersucht werden können, zu denen keine Kandidatenregionen vorhanden sind. Die grosse Anzahl Stopps zeigt gut, dass verschiedene Stopps gefunden werden. Es kann davon ausgegangen werden, dass beispielsweise im Rahmen der Lastwagendaten die Stopps an den Auf- und Abladestellen mit dem CB-SMoT gefunden werden, daneben jedoch auch weitere Stopps, beispielsweise in engen Passagen, in denen ein Lastwagen manövrieren muss. Grundsätzlich findet der CB-SMoT viele Stopps, wobei davon auszugehen ist, dass einige davon für eine bestimmte Applikation nicht von Interesse sind.

**Wanderdaten** Die allgemeine Analyse der Wanderdaten zeigt bei einer Mindestdauer von 10 Minuten unabhängig von der Distanz zur Berechnung der Nachbarschaft die tatsächliche, richtige Anzahl Stopps. Dies spricht dafür, dass der CB-SMoT für diese Daten geeignet ist. Auch im Rahmen der Analyse der ausgewählten Trajektorie des Wanderers werden die Regionen, in denen ein Stopp stattgefunden hat, zwar gefunden, räumlich jedoch nicht besonders genau lokalisiert. Zudem werden noch einige andere Regionen mit Stopps gefunden, wobei diese meistens plausibel sind (beispielsweise in Kurven), jedoch nicht unbedingt einen Stopp, sondern vielmehr eine reduzierte Geschwindigkeit oder einen Umweg wiedergeben.

**Lastwagendaten** Auch bei den Lastwagendaten stabilisiert sich die Anzahl an Stopps bereits ab einer Stoppdauer von mindestens 5 Minuten. Bei der Analyse der ausgewählten Trajektorie eines Lastwagens kommen die von Rocha et al. (2010) genannten Vorteile des CB-SMoT zur Geltung. So werden die 'echten' Stopps gefunden, welche sich innerhalb der Kandidatenregionen bewegen, jedoch auch die 'unbekannten' Stopps, die sich ausserhalb dieser Regionen befinden. Somit findet dieser Algorithmus in den Lastwagendaten insgesamt alle Stopps, durch eine separate Erfassung der 'echten' und der 'unbekannten' Stopps.

**Austernfischerdaten** In der allgemeinen Analyse der Austernfischerdaten werden rund neun Stopps als echte Stopps angenommen, wobei sowohl die rote als auch die blaue Kurve im Diagramm 5.23 (Stoppdauer grösser 10 Sekunden und Stoppdauer grösser 20 Sekunden) bei dieser Anzahl an Stopps über eine gewisse Distanz zur Berechnung der Nachbarschaft stabil bleibt. Mit diesem Wissen kann der kritische Bereich detektiert werden. Ist jedoch kein Wissen über die tatsächliche Anzahl an Stopps vorhanden, so ist es schwieriger, diesen Bereich zu erkennen. Das Resultat des CB-SMoT bei den Austernfischerdaten in der Analyse der ausgewählten Trajektorie ist wenig überzeugend. Dies könnte wiederum auf die Kleinräumigkeit dieser Daten zurück zu führen sein. Es bilden sich bei diesen Daten auf Grund des kleinen Bewegungsradius sehr schnell Cluster, ohne dass ein Stopp stattgefunden hat. So kann sich beispielsweise die Futtersuche ebenfalls auf einem sehr kleinen Raum bewegen, ohne dass hier ein tatsächlicher Stopp stattgefunden hat.

**Truthahngeierdaten** Bei der allgemeinen Analyse der Truthahngeierdaten in Diagramm 5.24 fällt der grosse Unterschied zwischen den Kurven mit einer Mindestdauer von 10 Stunden und einer Mindestdauer von 15 Stunden auf. Diese Beobachtung spricht dafür, dass dies den kritischen Bereich darstellt und davon ausgegangen werden kann, dass eine Stoppdauer von mindestens 15 Stunden falsche Stopps eliminiert. Somit scheint dieser Algorithmus auch die Truthahngeierdaten gut abbilden zu können. Eine kritische Stoppdauer im ähnlichen Bereich wie bei diesem Algorithmus wurde bei den Truthahngeierdaten auch bei den Algorithmen der Mindestdauer (Abbildung 5.14) und des DB-SMoT (Abbildung 5.29) gefunden. Dies spricht dafür, dass der Einbezug einer Mindestdauer bei der Analyse von Daten zu Zugvögeln essentiell ist.

### 6.2.3 Direction-Based Stops and Moves of Trajectories

Der DB-SMoT wurde mehrheitlich zur Analyse von Objekten entwickelt, bei denen die Richtungsänderungen beziehungsweise die Drehwinkel grundlegende Charakteristika sind (Palma et al., 2008). Es ist zu erwarten, dass im Rahmen dieser Arbeit der DB-SMoT bei den Truthahngeriederdaten die besten Resultate liefert.

Anhand der Resultate der allgemeinen Analyse ist ersichtlich, dass die Wander- und die Lastwagendaten ähnlich auf diesen Algorithmus reagieren. Die Anzahl an Stopps nimmt mit steigendem Schwellenwert der Richtungsänderung ab. Dies erscheint einleuchtend, da die Auswahl möglicher Stopppunkte immer stärker eingeschränkt wird. Der Einfluss der Stoppdauer scheint hier wenig einflussreich zu sein, die Abnahme der Stopps mit einer erhöhten Stoppdauer ist kontinuierlich.

Einige Kurven der Abbildung 5.26 zur allgemeinen Analyse der Wanderdaten geben zwar bei gewissen Werten des Drehwinkels die richtige Anzahl Stopps an, zeigen an diesen Stellen jedoch keine aussergewöhnlichen Reaktionen. Somit ist es ohne *ground truth* schwierig, zu entscheiden, bei welchen Schwellenwerten die tatsächlichen Stopps gefunden werden.

**Wander- und Lastwagendaten** Die guten Resultate des DB-SMoT in der Analyse der ausgewählten Trajektorien des Wanderers und des Lastwagens widersprechen den Resultaten des Bewegungsparameters des Drehwinkels. Dieser Umstand kann darauf zurück zu führen sein, dass der DB-SMoT mit mehreren Parametern, wie der *maxTol* und der Mindestdauer, arbeitet als die Kombination der Algorithmen Geschwindigkeit und Drehwinkel. Weiter kann argumentiert werden, dass im Falle der Lastwagendaten der Drehwinkel durchaus charakteristisch für das Bewegungsverhalten des sich bewegenden Objekts sein kann. Strassen, insbesondere grössere Strassen weisen oft wenig enge Kurven auf, und falls doch, kommen sie vereinzelt vor. Auch Wanderer bewegen sich meistens entlang von Wegen, wobei auch Wege oftmals wenige Kurven und nur selten scharfe Winkel aufweisen. Bei den Wanderdaten ist jedoch Vorsicht geboten. Würde man beispielsweise die Trajektorie eines Hochgebirgswanderers an einem sehr steilen Hang entlang einer Serpentine analysieren, so könnte die Form dieses Weges bei einer Analyse mit dem DB-SMoT irreführend sein.

Insgesamt scheint der DB-SMoT jedoch auch für diese sich bewegenden Objekte ein geeigneter Algorithmus zu sein. Moreno et al. (2010) argumentieren, dass der DB-SMoT

in spezifischen Fällen interessant ist. Diese spezifischen Fälle stellen Trajektorien dar, in denen die Richtungsänderung einen grösseren Einfluss hat als die Geschwindigkeit. Nun stellt sich meines Erachtens die Frage, mit welchen Mitteln und wie gut dieser Einfluss definiert werden kann. Für die Wander- als auch die Lastwagendaten spielt die Geschwindigkeit meines Erachtens eine wichtigere Rolle als die Richtungsänderung. Dies widerspricht der Aussage von Moreno et al. (2010) in Anbetracht der hier erhaltenen Resultate. Um diesen Sachverhalt zu klären könnten weitere Trajektorien von motorisierten Fahrzeugen und wandernden Personen miteinbezogen werden. Dabei müsste darauf geachtet werden, dass bei diesen Trajektorien sowohl die Geschwindigkeit als auch die Richtungsänderungen charakteristisch sind. So könnte herausgefunden werden, welcher Parameter den grösseren Einfluss hat.

**Austernfischerdaten** Das unzulängliche Finden von Stopps bei der Anwendung auf die Austernfischerdaten könnte darauf zurückzuführen sein, dass der Austernfischer, als relativ kleiner und wendiger Vogel, sowohl während eines Stopps als auch während einer Bewegung die Richtung stark wechselt (beispielsweise bei der Futtersuche). Im Gegensatz zu grösseren Tieren, wie beispielsweise dem Truthahngeier, ist davon auszugehen, dass es für den Austernfischer einfacher ist, sich schnell zu bewegen und dennoch Bewegungen mit hohen Drehwinkeln zu machen. Bei den Austernfischerdaten kann davon ausgegangen werden, dass der DB-SMoT auf Grund des Bewegungsverhaltens dieser Vögel nicht ideal ist für die Detektion der Stopps.

**Truthahngeierdaten** Die Truthahngeierdaten weisen ebenfalls keinen kritischen Wertebereich des Drehwinkels auf, die Zunahme an Stopps steigt kontinuierlich an bis zu einem relativ flachen Maximum. Dies erstaunt etwas, da Buchin et al. (2013) für die Detektion von Flugphasen von Zugvögeln als Kriterium einen *bounded heading angular range* von 120 Grad verwendet und dieser Bewegungsparameter somit charakteristisch für diese Klasse von sich bewegendem Objekt sein sollte. Dieser kritische Wertebereich ist in Diagramm 5.29 jedoch nicht zu erkennen. Dieser Umstand könnte auf die zeitliche Auflösung der Daten darauf zurückzuführen sein. In der Arbeit von Buchin et al. (2013) wurde eine Trajektorie von 126 Tagen untersucht. Dies ist eine viel kürzere Zeitspanne als die des hier untersuchten Zugvogels. Es ist somit möglich, dass die zeitliche Auflösung

feiner ist als bei den hier verwendeten Daten und die Drehwinkel somit adäquater mit-  
einbezogen werden können.

Weiter könnte es idealer sein, den Drehwinkel nicht als Schwellenwert miteinzubeziehen,  
sondern als Bereich, *range*, wie Buchin et al. (2013) es tun. Weiter nehme ich an, dass  
Buchin et al. (2013) den Bereich des Drehwinkels für ein ganzes Segment definieren und  
nicht für einzelne Punkte. Diese Herangehensweise mit dem Einbezug der Richtung über  
ein grösseres räumliches Fenster ist möglicherweise auch für den hier verwendeten Algo-  
rithmus des Drehwinkels adäquater.

Eine weitere Eigenheit der abgebildeten Resultate im Diagramm 5.29 ist, dass die An-  
zahl Stopps bei einem Drehwinkel von 180 Grad nicht immer null ist. Dies ist darauf  
zurückzuführen, dass der Wert *maxTol* auf 2 gesetzt wurde. Weiter gilt es zu beachten,  
dass eine Mindestdauer von 5 Stunden bei einem Aufnahmeintervall von gut 3 Stunden  
Stopps erlaubt, die nur aus zwei Punkten bestehen.

Der markante Sprung zwischen einer Mindeststoppdauer von 10 und 20 Stunden ist  
auch bei dem Algorithmus der Mindestdauer und dem CB-SMoT zu erkennen. Dies  
zeigt, dass die Mindestdauerwerte bei den Daten zum Truthahngeier essentiell sind für  
eine plausible Analyse.

**Eye-Tracking-Daten** Im Rahmen der Eye-Tracking-Daten (Abbildung 5.30) werden  
grundsätzlich zu viele Stopps gefunden. Interessant ist, dass bis zu einem minimalen  
Drehwinkel von knapp 100 Grad die Stoppdauer fast keinen Einfluss hat. Die tatsächliche  
Anzahl Stopps wird mit einem Drehwinkel von knapp 70 Grad gefunden, unabhängig  
von der Stoppdauer. Diese Resultate sind meines Erachtens kontraintuitiv. Betrachtet  
man Diagramm 5.30, so scheinen die wichtigen Wertebereiche dort zu liegen, wo die  
unterschiedlichen Stoppdauerwerte auch unterschiedliche Resultate liefern. Andererseits  
kann auch argumentiert werden, dass das beste Resultat in den Bereichen erzeugt wird,  
in denen der Datensatz robust ist gegenüber unterschiedlichen Schwellenwerten für die  
Stoppdauer.

### 6.3 Hidden Markov Model

Die ähnlichen Ergebnisse des HMM und des Geschwindigkeitsschwellenwertes bei den Wander- und Lastwagendaten erstaunen nicht, da das HMM lediglich auf der Geschwindigkeit beruht. Die starken Schwankungen bei den Austernfischerdaten in Diagramm 5.33 könnten auf die Kleinräumigkeit der Daten zurückzuführen sein.

Salvucci und Goldberg (2000) nennen einige Vorteile des HMM gegenüber einem allein auf der Geschwindigkeit beruhenden Algorithmus. So ist nach ihnen das HMM robuster als ein Schwellenwert für die Geschwindigkeit. Diese Aussage kann im Rahmen der hier erhaltenen Resultate teilweise unterstützt werden. So weisen zwar die Austernfischerdaten unter der Analyse mit dem HMM weniger Schwankungen auf und auch die Analyse der ausgewählten Trajektorie liefert bessere Resultate. Die anderen Datensätze hingegen reagieren eher unruhiger unter der Analyse des HMM. Dieses Verhalten könnte auf Unsicherheiten hindeuten oder darauf, dass die kritischen Bereiche besser detektiert werden. Die Kurven in Diagramm 5.31 des HMM angewendet auf die Wanderdaten schwanken zwar mehr als die Kurven in Diagramm 5.1 des Geschwindigkeitsschwellenwertes, die tatsächliche Anzahl an Stopps wird jedoch nicht klarer abgegrenzt. Dies spricht dagegen, dass das HMM im Rahmen dieser Daten robustere Resultate liefert als ein Geschwindigkeitsschwellenwert.

Weiter werden bei Salvucci und Goldberg (2000) die Fixationen mittels HMM besser gruppiert als mit einem Geschwindigkeitsschwellenwert. Im Rahmen der hier gemachten Analysen wird dieser Vorteil bei den Austernfischern ersichtlich, wobei keine losen Punkte als Stopp definiert werden. Bei den Lastwagendaten hingegen werden einige einzelne Punkte mit dem HMM als Stopp definiert, somit scheint dieser Vorteil nicht bei allen Datensätzen vorzuliegen.

Eine Analyse mit dem HMM unter der Verwendung von kontinuierlichen Werten für die Geschwindigkeit würde wahrscheinlich idealere Resultate liefern, da so mehr Informationen miteinbezogen werden könnten.

**Truthahngeierdaten** Bei den Truthahngeierdaten in Diagramm 5.34 findet das HMM nur wenige Stopps. Die Übergangswahrscheinlichkeiten für das HMM, welches hierbei benutzt wurde, sind sehr niedrig. Dies könnte Grund für das Finden nur weniger Stopps sein. Bei der sowohl zeitlich als auch räumlich groben Auflösung dieser Daten könnte

es sein, dass die Unterschiede zwischen den verschiedenen Geschwindigkeitswerten, wie sie im Rahmen dieser Arbeit berechnet wurden, zu gering sind und die Unterteilung der Beobachtungen in zwei Kategorien die Daten nicht ideal abbilden kann. Im Fall der Truthahngerdaten könnte eine Analyse mit kontinuierlichen Daten für die Beobachtungen des HMM einen grossen Mehrwert bieten.

## 6.4 Beantwortung der Forschungsfragen

Die gemachten Analysen und die daraus erhaltenen Resultate geben Auskunft über die Reaktion der einzelnen Bewegungsparameter und Algorithmen auf die verschiedenen Datensätze. Im Folgenden werden die Forschungsfragen (Abschnitt 2.7) diskutiert und beantwortet.

- Wie reagieren existierende Methoden zur Detektion von Stopps in Trajektorien bei deren Anwendung auf verschiedene Datensätze, beziehungsweise wie robust sind sie?

Die Resultate zeigen, dass sowohl die Bewegungsparameter als auch die Algorithmen aus der Literatur gegenüber verschiedenen Datensätzen nicht sehr robust sind. So liefern alle Methoden je nach Datensatz zuverlässige oder auch falsche Resultate. Die Robustheit der Verfahren hängt also zu einem gewissen Teil von den Datensätzen, dem sich bewegenden Objekt, das damit dargestellt ist, dem Abtastintervall und der Genauigkeit der Positionsmessungen ab. Zum anderen spielt eine ideale Kombination aus Variablen und Wertebereichen der gesetzten Schwellenwerte eine essentielle Rolle. Mit den ideal passend gesetzten Schwellenwerten können die Stopps teilweise fehlerlos gefunden werden. Mit suboptimalen Variablen und Schwellenwerten ist eine Detektion der Stopps bereits schwieriger. Insbesondere die Austernfischerdaten konnten mit keinem der hier behandelten Algorithmen adäquat analysiert werden. Grundsätzlich finden die meisten Algorithmen eher eine hohe Anzahl Stopps, wobei davon auszugehen ist, dass nicht alle gefundenen Stopps von Interesse sind. Insofern kann gefolgert werden, dass alle hier behandelten Bewegungsparameter und Algorithmen mit zu vielen Stopps als Resultate auf die verschiedenen Datensätzen reagieren. Die zweite Forschungsfrage lautet:

- Bietet die Berücksichtigung einer Kombination von Bewegungsparametern beziehungsweise die Verwendung von komplexen Algorithmen bei der Erkennung von Stopps in Trajektorien einen Mehrwert gegenüber einfacheren Methoden?

Im Falle der Bewegungsparameter ist ersichtlich, dass der Einbezug einer Mindestdauer oder einer Mindestausdehnung die Stopps besser findet als ein reiner Schwellenwert der Geschwindigkeit. Der Einbezug des Drehwinkels der einzelnen Punkte hingegen bietet keinen Mehrwert. Der DB-SMoT, welcher ein relativ einfacher Algorithmus ist, liefert ebenfalls gute Resultate, wobei hier jedoch bereits mehrere Variablen miteinander kombiniert werden müssen. Somit liefert dieser Algorithmus von mittlerer Komplexität ähnlich adäquate Resultate wie eine Kombination aus einem Geschwindigkeitsschwellenwert und einer Mindestdauer beziehungsweise einer Mindestausdehnung. Der CB-SMoT, der im Vergleich zu den anderen Algorithmen komplexer ist, liefert grundsätzlich gute Resultate. Für die Implementation und die Rechenzeit dieses Algorithmus musste jedoch sehr viel mehr Zeit aufgewendet werden. Weiter müssen relativ viele Variablen und Schwellenwerte definiert werden, was zeitaufwändig ist und viele verschiedene Kombinationen zulässt. Insofern kann gesagt werden, dass ein komplexer Algorithmus die Stopps in gewissen Fällen zuverlässiger detektiert als einfache Algorithmen, dieser Mehrwert jedoch durch einen grösseren Aufwand bei der Implementation und durch längere Rechenzeiten relativiert wird. Die dritte Forschungsfrage lautet folgendermassen:

- Welche Algorithmen tragen wenig/viel zum Erkennen von Stopps bei?

Der Bewegungsparameter der Geschwindigkeit trägt viel zum Finden von Stopps bei. Die Geschwindigkeit ist mit Ausnahme des DB-SMoT in allen Algorithmen und Kombinationen von Bewegungsparametern enthalten, die teilweise sehr gute Resultate liefern. Dabei ist die Geschwindigkeit jedoch nicht überall gleichermassen definiert. Diese Problematik ist in Abschnitt 4.1.1 erläutert. Bei den Truthahngerätedaten scheint eine Mindestdauer einen besonders grossen Einfluss zu haben. Im Rahmen der Algorithmen aus der Literatur kann gesagt werden, dass ein Algorithmus mit möglichst vielen Variablen und Parametern nicht zwingend die besten Resultate liefert. Insofern gibt es keinen ultimativen Parameter, der in jedem Fall ideal passend ist.

# Kapitel 7

## Schlussfolgerungen und Ausblick

### 7.1 Fazit

In dieser Arbeit konnte gezeigt werden, dass verschiedene Methoden zur Detektion von Stopps in Trajektorien sensibel reagieren auf die Datensätze, auf die sie angewendet werden. Dabei konnte kein Algorithmus oder Bewegungsparameter ermittelt werden, der für alle Datensätze oder Anwendungen ideal ist. Auch Rinzivillo et al. (2008) thematisieren diesen Sachverhalt. So müssen nach Rinzivillo et al. (2008) verschiedene Methoden erarbeitet werden, um auch verschiedene Trajektorien analysieren zu können.

Weiter konnte im Rahmen dieser Arbeit gezeigt werden, dass jeder Algorithmus, angewendet auf einen passenden Datensatz und mit ideal gesetzten Parametern, gute Resultate liefert. Es gibt also kein in dieser Arbeit untersuchtes Verfahren, welches nie angewendet werden sollte. Diese Arbeit hat zudem Hinweise dazu geliefert, dass die Auswahl einer Methode zur Detektion von Stopps von einigen Faktoren abhängig ist. Solche Faktoren können die gewünschte Genauigkeit der Resultate sein, die Eigenschaften der zu analysierenden Daten oder des sich bewegenden Objekts sowie die vorhandenen Ressourcen zur Ausführung eines Verfahrens.

In der Literatur bestehen zu einigen der hier verwendeten Algorithmen bereits Alternativen oder Erweiterungen; beispielsweise baut die Arbeit von Moreno et al. (2010) auf dem Algorithmus von Palma et al. (2008) auf, welcher eine Erweiterung des Verfahrens von Alvares et al. (2007) darstellt. Die Vorteile dieser Erweiterungen wurden auch in dieser Arbeit ersichtlich.

Anhand dieser Arbeit kann gesagt werden, dass die Detektion von Stopps in Trajektorien von Anwendung zu Anwendung mit unterschiedlichen Methoden angegangen werden sollte. Diese Thematik wird auch von Orellana und Wachowicz (2011) erläutert. Sie bemerken, dass sowohl die Bedeutung eines Stopps als auch dessen räumliche und zeitliche Schwellenwerte von Fall zu Fall variieren. Die Wichtigkeit der zeitlichen Auflösung, welche je nach verfügbaren Trajektorien Daten differiert, thematisieren auch Laube und Purves (2011). Es konnte in der hier vorliegenden Arbeit somit bestätigt werden, dass bei der Wahl eines Analyseverfahrens auf verschiedene Aspekte eingegangen werden sollte. Dies ist unter anderem darauf zurückzuführen, dass Eigenschaften des sich bewegenden Objekts, unterschiedliche Messverfahren und -intervalle sowie die räumliche Auflösung der Daten und die Zielsetzung in jeder Situation unterschiedlich sind. Diese Parameter spielen bei der Detektion von Stopps in Trajektorien eine essentielle Rolle. Somit erstaunt es nicht, dass auch die Verfahren dazu spezifisch gewählt werden sollten, um ideale Resultate zu erhalten.

Bei den hier untersuchten Methoden stellte es sich häufig als Gratwanderung heraus, wie viele falsche Stopps in Kauf genommen werden, um möglichst viele richtige Stopps zu finden. Die Entscheidung, wie viele falsche Stopps toleriert beziehungsweise wie viele richtige Stopps mit grosser Sicherheit detektiert werden wollen, hängt vom Ziel einer Analyse ab. So können ein ideales Verfahren sowie passende Parameter nur im Hinblick auf Ansprüche an eine gewisse Zielerreichung definiert werden.

In dieser Arbeit war die *ground truth* nicht zu allen Datensätzen vorhanden, somit konnten einige sich bewegende Objekte nur im Rahmen einer allgemeinen und wenig detaillierten Analyse behandelt werden. Dies stellt eine Limitierung dieser Arbeit dar, da tatsächliche Stopps für bestimmte Datensätze nicht von Artefakten unterschieden werden konnten. Diese Problematik von Artefakten diskutieren auch Zimmermann et al. (2009). Meines Erachtens spielt diese Thematik eine essentielle Rolle, da mit ungenügendem Wissen über die Daten auch die Resultate nur unzureichend interpretiert werden können.

Die hier vorliegende Arbeit macht lediglich eine Aussage zur Eignung der untersuchten Algorithmen für die hier ausgewählten Datensätze. Somit gelten die hier gemachten Aussagen streng genommen lediglich für die in dieser Arbeit behandelten sich bewegenden Objekte. Es kann jedoch gesagt werden, dass die verwendeten fünf Datensätze auch fünf verschiedene, stark individuelle sich bewegende Objekte darstellen. Somit können

anhand der hier erhaltenen Resultate Schlüsse auf ähnliche sich bewegende Objekte gezogen werden.

## 7.2 Ausblick

Die strukturellen Gemeinsamkeiten sowie Gemeinsamkeiten der räumlicher Ausprägung der hier verwendeten Datensätze könnten in einer weiterführenden Arbeit tiefer thematisiert werden. Das Wissen darüber und über allfällige Unterschiede wäre für die Verallgemeinerung der hier erhaltenen Resultate essentiell. Zudem könnte versucht werden, die Bewegungsparameter und Algorithmen sowie daraus resultierenden Kombinationen nach deren Eignung für verschiedene sich bewegende Objekte zu klassifizieren. So könnte eine Basis gelegt werden, die zukünftigen Anwendern und Anwenderinnen einen Überblick gibt über mögliche Methoden und deren Eigenschaften.

Insbesondere die Austernfischerdaten konnten mit keinem der hier untersuchten Algorithmen adäquat analysiert werden. Somit könnte in einer weiterführenden Arbeit nach Algorithmen und Parametern gesucht werden, die Daten dieser Art sinnvoll analysieren können. Dabei wäre es von Vorteil, einen Datensatz zur Verfügung zu haben, der sich über einen längeren Zeitraum erstreckt.

Eine weitere Idee ist, dieselbe Strecke mit verschiedenen Verkehrsmitteln, wie beispielsweise dem Auto, dem Fahrrad oder zu Fuss, zurückzulegen und dabei die Besonderheiten und Charakteristika der einzelnen Varianten der Fortbewegung zu studieren. Dabei könnte grösstenteils ausgeschlossen werden, dass verschiedene Gegebenheiten durch eine veränderte Umgebung definiert sind. So kann man erwarten, dass Personen zu Fuss nicht unbedingt die gleichen Routen wählen würden wie Personen, die mit dem Auto unterwegs sind. Durch eine Festlegung der Strecke könnten diese Einflüsse reduziert werden. Weiterführend könnten bei den verschiedenen Algorithmen und Bewegungsparametern die einzelnen Variablen noch genauer untersucht werden. So wäre es interessant, den Einfluss der verschiedenen Variablen sowie dessen Ausmass detaillierter zu untersuchen. Dazu könnten Informationen über die Datensätze miteinbezogen werden, um spezifischer auf die Eigenheiten dieser einzugehen.

Meines Erachtens stellte der Einbezug der geometrischen Ausprägung der Trajektorie einen Mehrwert dar. Im Rahmen der Analyse der ausgewählten Trajektorie des Lastwagens beispielsweise wurde ersichtlich, dass die Lastwagen grundsätzlich immer die gleiche

Strecke hin- und zurückfahren und bei den Auf- und Abladestellen des Betons wenden. Diese Form der Trajektorie könnte in die Analyse miteinbezogen werden. Dabei könnten die Enden der Linien der Trajektorie (also die Wendepunkte) eines solchen Lastwagens ein Indikator für einen Stopp sein.

Abschliessend kann gesagt werden, dass im Rahmen dieser Arbeit und der hier verwendeten Methodik die Forschungsfragen beantwortet werden konnten. Die hier gewonnenen Erkenntnisse können, und werden hoffentlich, in diversen Forschungs- und Anwendungsgebieten für weitere Analysen oder als Referenz verwendet werden.

# Literaturverzeichnis

- Alvares, L., Bogorny, V., Kuijpers, B., de Macedo, J., Moelans, B. und Vaisman, A. Model for Enriching Trajectories with Semantic Geographical Information. In: *Proceedings of the 15th International Symposium on Advances in Geographic Information Systems*. ACM (2007).
- Ankerst, M., Breunig, M., Kriegel, H.P. und Sander, J. OPTICS: Ordering Points To Identify the Clustering Structure. In: *ACM SIGMOD*, S. 49–60. ACM (1999).
- Baglioni, M., de Macedo, J., Renso, C., Trasarti, R. und Wachowicz, M. Towards Semantic Interpretation of Movement Behavior. In: *Advances in GIScience*, S. 271–288. Springer (2009).
- Bak, P., Packer, E., Ship, H. und Dotan, D. Algorithmic and Visual Analysis of Spatio-temporal Stops in Movement Data. In: *ACM SIGSPATIAL GIS*, S. 462–465 (2012).
- Bogorny, V., Kuijpers, B. und Alvares, L. ST-DMQL: A Semantic Trajectory Data Mining Query Language. *International Journal of Geographical Information Science*, 23(10):1245–1276 (2009).
- Brilhante, I., de Macedo, J., Renso, C. und Casanova, M. Trajectory Data Analysis using Complex Networks. *Proceedings of the 15th Symposium on International Database Engineering & Applications - IDEAS*, 17:17 – 25 (2011).
- Bruderer, B. und Boldt, A. Flight characteristics of birds. *Ibis*, 143(2):178–204 (2001).
- Buchin, M., Driemel, A., van Kreveld, M. und Sacristán, V. An Algorithmic Framework for Segmenting Trajectories based on Spatio-Temporal Criteria. In: *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, S. 202–211. ACM (2010).

- Buchin, M., Kruckenberg, H. und Kölzsch, A. Segmenting Trajectories by Movement States. In: *Advances in Spatial Data Handling, Advances in Geographic Information Science*, S. 15–25. Springer (2013).
- Camerini, P., Galbiati, G. und Maffioli, F. Algorithms for finding optimum trees: Description, use and evaluation. *Annals of Operations Research*, 13(1):263–397 (1988).
- Collins, R., Tourtellot, M. und Bell, W. Defining stops in search pathways. *Journal of Neuroscience Methods*, 60:95 – 98 (1995).
- Coltekin, A., Heil, B., Garlandini, S. und Fabrikant, S. Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-movement Analysis. *Cartography and Geographic Information Science*, 36(1):5 – 17 (2009).
- Den Buurman, R., Roersema, T. und Gerrissen, J. Eye movements and the perceptual span in reading. *Reading Research Quarterly*, 16(2):227–235 (1981).
- Erkelens, C. und Vogels, I. The initial direction and landing position of saccades. *Studies in Visual Information Processing*, 6:133–144 (1995).
- Ester, M., Kriegel, H.P., Sander, J. und Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: *KDD*, Bd. 96, S. 226–231 (1996).
- Guc, B., May, M., Saygin, Y. und Körner, C. Semantic Annotation of GPS Trajectories. In: *11th AGILE International Conference on Geographic Information Science*, S. 1–9 (2008).
- Gudmundsson, J., Laube, P. und Wolle, T. Movement Patterns in Spatio-temporal Data. In: *Encyclopedia of GIS*, S. 726–732. Springer US (2008).
- Hoogendoorn, S. und Bovy, P. Pedestrian Travel Behavior Modeling. *Networks and Spatial Economics*, 5(2):193–216 (2005).
- Kiefer, P. und Giannopoulos, I. Gaze map matching: Mapping eye tracking data to geographic vector features. In: *Proceedings of the 20th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM (2012).

- Laube, P., Imfeld, S. und Weibel, R. Discovering relative motion patterns in groups of moving objects. *International Journal of Geographical Information Science*, 19(6):639 – 668 (2005).
- Laube, P. und Purves, R. How fast is a cow? Cross-Scale Analysis of Movement Data. *Transactions in GIS*, 15(3):401 – 418 (2011).
- Mandel, J., Bohrer, G., Winkler, D., Barber, D., Houston, S. und Bildstein, K. Migration path annotation: cross-continental study of migration-flight response to environmental conditions. *Ecological Applications*, 21(6):2258 – 2268 (2011).
- van Moorter, B., Visscher, D., Jerde, C., Frair, J. und Merrill, E. Identifying Movement States From Location Data Using Cluster Analysis. *Journal of Wildlife Management*, 74(3):588 – 594 (2010).
- Moreno, B., Times, V., Renso, C. und Bogorny, V. Looking Inside the Stops of Trajectories of Moving Objects. In: *XI Brazilian Symposium on Geoinformatics*, S. 9–20 (2010).
- Nyström, M. und Holmqvist, K. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, 42(1):188–204 (2010).
- Orellana, D. und Wachowicz, M. Exploring Patterns of Movement Suspension in Pedestrian Mobility. *Geographical Analysis*, 43:241 – 260 (2011).
- Palma, A., Bogorny, V., Kuijpers, B. und Alvares, L. A Clustering-based Approach for Discovering Interesting Places in Trajectories. In: *Proceedings of the 2008 ACM symposium on Applied computing*, S. 863–868. ACM (2008).
- Pelekis, N., Kopanakis, I., Kotsifakos, E., Frentzos, E. und Theodoridis, Y. Clustering uncertain trajectories. *Knowledge and Information Systems*, 28(1):117–147 (2010).
- Rabiner, L. A Tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286 (1989).
- Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N. und Andrienko, G. Visually driven analysis of movement data by progressive clustering. *Information Visualization*, 7:225–239 (2008).

- Rocha, J., Oliveira, G., Alvares, L. und Bogorny, V. DB-SMoT: A Direction-Based Spatio-Temporal Clustering Method. In: *Proceedings of the IEEE International Conference on Intelligent Systems*, S. 114–119. IEEE (2010).
- Salvucci, D. *Mapping Eye Movements to Cognitive Processes*. Dissertation, Carnegie Mellon University Pittsburg (1999).
- Salvucci, D. und Anderson, J. Tracing eye movement protocols with cognitive process models. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, S. 923–928 (1998).
- Salvucci, D. und Goldberg, J. Identifying Fixations and Saccades in Eye-Tracking Protocols. In: *Proceedings of the 2000 symposium on Eye tracking research & applications*, S. 71–78. ACM (2000).
- Schaub, M., Pradel, R., Jenni, L. und Lebreton, J.D. MIGRATING BIRDS STOP OVER LONGER THAN USUALLY THOUGHT: AN IMPROVED CAPTURE-RECAPTURE ANALYSIS. *Ecology*, 82(3):852 – 859 (2001).
- Schreutelkamp, F. und Strang van Hees, G. Benaderingsformules voor de transformatie tussen RD-en WGS84-kaartcoördinaten. *Geodesia: tijd-schrift voor geodesie en geoinformatie*, 43(2):64–69 (2001).
- Schuessler, N. und Axhausen, K. Processing Raw Data from Global Positioning Systems Without Additional Information. *Journal of the Transportation Research Board*, 2105:28 – 36 (2009).
- Sen, T. und Megaw, T. The effects of task variables and prolonged performance on saccadic eye movement parameters. *Advances in Psychology*, 22:103–111 (1984).
- Shamoun-Baranes, J., Bom, R., van Loon, E., Ens, B., Oosterbeek, K. und Bouten, W. From Sensor Data to Animal Behaviour: An Oystercatcher Example. *PloS one*, 7(5):1–13 (2012).
- Sigrist, P., Coppin, P. und Hermy, M. Impact of forest canopy on quality and accuracy of GPS measurements. *International Journal of Remote Sensing*, 20(18):3595–3610 (1999).

- Spaccapietra, S., Parent, C., Damiani, M., de Macedo, J., Porto, F. und Vagenot, C. A conceptual view on trajectories. *DATA & KNOWLEDGE ENGINEERING*, 65:126 – 146 (2008).
- Stenneth, L., Wolfson, O., Yu, P. und Xu, B. Transportation Mode Detection using Mobile Phones and GIS Information. In: *ACM SIGSPATIAL GIS 11*, S. 54–63. ACM (2011).
- Waga, K., Tabarcea, A., Chen, M. und Fränti, P. Detecting movement type by route segmentation and classification. In: *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2012 8th International Conference on*, S. 508–513. IEEE (2012).
- Widdel, H. Operational problems in analysing eye movements. *Advances in Psychology*, 22:21–29 (1984).
- Zimmermann, M., Kirste, T. und Spiliopoulou, M. Finding Stops in Error-Prone Trajectories of Moving Objects with Time-Based Clustering. In: *Intelligent Interactive Assistance and Mobile Multimedia Computing*, Bd. 53, S. 275–286. Springer (2009).

## Anhang A

# Übersichtstabellen der Resultate

Die Abbildungen A.1 und A.2 geben einen Überblick über die Resultate der untersuchten Methoden für die verschiedenen Datensätze. Dabei sind sowohl die Resultate der allgemeinen Analyse als auch die Resultate der Analyse der ausgewählten Trajektorien zusammengefasst. Diese Informationen sind ebenfalls in Kapitel 5 in Form von Diagrammen und Karten dargestellt.

	Algorithmus Geschwindigkeit	Algorithmus Dauer	Algorithmus Ausdehnung	Algorithmus Drehwinkel
	Maximal gefundene Stopps: 19	Maximal gefundene Stopps: 9	Maximal gefundene Stopps: 17	Maximal gefundene Stopps: 4
<b>Wanderer</b>	Mit dem idealen Schwellenwert, welcher in der allgemeinen Analyse nicht auffällig und ohne <i>ground truth</i> schwierig zu detektieren ist, werden alle tatsächlichen als auch drei falsche Stopps gefunden.	In der allgemeinen Analyse markante Reduktion an Stopps mit einer Dauer grösser 2 Minuten, in der Analyse der ausgewählten Trajektorie werden alle tatsächlichen als auch zwei falsche Stopps gefunden.	In der allgemeinen Analyse stetige Abnahme an Stopps mit steigendem Schwellenwert, kein markanter Sprung, in der Analyse der ausgewählten Trajektorie werden alle tatsächlichen als auch zwei falsche Stopps gefunden.	Führt in der allgemeinen Analyse zu sehr unruhigen Resultaten und einer Reduktion an Stopps, in der Analyse der ausgewählten Trajektorie sowohl eine Reduktion an falschen als auch an tatsächlichen Stopps.
	Maximal gefundene Stopps: 28	Maximal gefundene Stopps: 18	Maximal gefundene Stopps: 22	Maximal gefundene Stopps: 6.5
<b>Lastwagen</b>	Mit dem idealen Schwellenwert, welcher in der allgemeinen Analyse nicht auffällig und ohne <i>ground truth</i> schwierig zu detektieren ist, werden alle tatsächlichen als auch einige falsche Stopps gefunden.	In der allgemeinen Analyse etwas markante Abnahme an Stopps mit einer Mindestdauer von 5 Minuten, ansonsten ähnliche Verläufe der Kurven wie bei einem Geschwindigkeitsschwellenwert, in der Analyse der ausgewählten Trajektorie werden mit dem idealen Schwellenwert alle tatsächlichen ohne falsche Stopps gefunden.	In der allgemeinen Analyse markante Abnahme an Stopps mit einem Schwellenwert von 100 Metern, ansonsten ähnliche Verläufe der Kurven wie bei einem Geschwindigkeitsschwellenwert, in der Analyse der ausgewählten Trajektorie werden alle richtigen und einige falsche Stopps gefunden.	Führt in der allgemeinen Analyse zu sehr unruhigen Resultaten und einer Reduktion an Stopps, in der Analyse der ausgewählten Trajektorie wird eine Reduktion an falschen als auch an tatsächlichen Stopps.
	Maximal gefundene Stopps: 7	Maximal gefundene Stopps: 6	Maximal gefundene Stopps: 5	Maximale gefundene Stopps: 4.5
<b>Austernfischer</b>	In der allgemeinen Analyse unruhige Kurve, bereits bei einem tiefen Schwellenwert für die Geschwindigkeit ist ein grosser Anteil an Punkten als Stopp definiert, in der Analyse der ausgewählten Trajektorie werden zwei von drei tatsächlichen sowie drei falsche Stopps gefunden.	In der allgemeinen Analyse relativ stetige Abnahme an Stopps mit steigendem Schwellenwert, ab einer Mindestdauer von 5 Minuten konstantere Anzahl an Stopps, in der Analyse der ausgewählten Trajektorie werden zwei tatsächliche sowie zwei falsche Stopps gefunden.	In der allgemeinen Analyse starke Schwankungen der Kurven, wobei die Anzahl an Stopps mit steigendem Schwellenwert stetig abnimmt, in der Analyse der ausgewählten Trajektorie werden zwei tatsächliche sowie drei falsche Stopps gefunden.	Führt in der allgemeinen Analyse zu sehr unruhigen Resultaten und einer Reduktion an Stopps, in der Analyse der ausgewählten Trajektorie wird die Anzahl an falschen Stopps reduziert, zwei tatsächliche Stopps werden weiterhin gefunden.
	Maximal gefundene Stopps: 440	Maximal gefundene Stopps: 330	Maximal gefundene Stopps: 600	Maximal gefundene Stopps: 40
<b>Trutthahngerler</b>	In der allgemeinen Analyse relativ rascher Anstieg des prozentualen Anteils an Stopps, bei dem Maximum an Stopps sind über 60% der Punkte als Stopp definiert, insgesamt glatte Kurven.	In der allgemeinen Analyse eine grosse Reduktion an Stopps zwischen einer einer Mindestdauer von 10 und 20 Stunden, ab einer Mindestdauer von 20 Stunden flachere Kurven.	In der allgemeinen Analyse markante Abnahme an Stopps mit einer Mindestausdehnung von 3 km bei tiefen Schwellenwerten für die Geschwindigkeit, das Maximum an Stopps wird für alle Kurven ab einer Mindestausdehnung von 3 km erst ungefähr bei einem Geschwindigkeitsschwellenwert von 10 km/h erreicht.	Führt in der allgemeinen Analyse zu unruhigen Resultaten und reduziert die Anzahl an Stopps stark.
	Maximal gefundene Stopps: 15000	Maximal gefundene Stopps: 3000	Maximal gefundene Stopps: 8050	Maximal gefundene Stopps: 40
<b>Auge</b>	In der allgemeinen Analyse relativ rascher Anstieg des prozentualen Anteils an Stopps, bei dem Maximum an Stopps sind über 60% der Punkte als Stopp definiert, insgesamt glatte Kurven.	In der allgemeinen Analyse eine grosse Reduktion an Stopps zwischen einer einer Mindestdauer von 100 und 200 Sekunden, ab einer Mindestdauer von 200 Sekunden sehr ähnliche Kurven für die weiteren Schwellenwerte der Mindestdauer.	In der allgemeinen Analyse sind keine markanten Sprünge bei einer gewissen Mindestausdehnung zu erkennen, die maximale Anzahl an Stopps wird für alle Kurven bei einem Geschwindigkeitsschwellenwert von 4 px/s erreicht.	Führt in der allgemeinen Analyse zu extrem unruhigen Resultaten und reduziert die Anzahl an Stopps sehr stark.

ABBILDUNG A.1: Übersicht über die Resultate der Bewegungsparameter.

	Algorithmus IB-SMoT	Algorithmus CB-SMoT	Algorithmus DB-SMoT	Algorithmus HMM
Wanderer	keine Angaben	Maximal gefundene Stopps: 48	Maximal gefundene Stopps: 10	Maximal gefundene Stopps: 16
		In der allgemeinen Analyse sehr stabile Resultate ab einer Stoppdauer grösser 5 Minuten, in der Analyse der ausgewählten Trajektorie werden drei von vier tatsächlichen Stopps gefunden sowie viele falsche Stopps, wobei diese sich an 'plausiblen' Orten befinden.	In der allgemeinen Analyse kontinuierliche Abnahme an Stopps mit steigendem Wert für den Mindestdrehwinkel, in der Analyse der ausgewählten Trajektorie werden mit kleinen räumlichen Ungenauigkeiten alle tatsächlichen und drei falsche Stopps gefunden.	In der allgemeinen Analyse ähnliche Resultate wie der Algorithmus der Geschwindigkeit, in der Analyse der ausgewählten Trajektorie werden alle tatsächlichen sowie zwei falsche Stopps gefunden.
Lastwagen	keine Angaben	Maximal gefundene Stopps: 38	Maximal gefundene Stopps: 19	Maximal gefundene Stopps: 24
	In der Analyse der ausgewählten Trajektorie werden zwei von fünf Stopps gefunden, die Ampeln als Kandidatenregionen sind möglicherweise nicht ideal beziehungsweise reichen nicht aus.	In der allgemeinen Analyse ist eine markante Reduktion an Stopps mit einer Mindstdauer von 5 Minuten zu erkennen, in der Analyse der ausgewählten Trajektorie werden alle Stopps ohne falsche Stopps gefunden.	In der allgemeinen Analyse ausser bei sehr tiefen Schwellenwerten eine kontinuierliche Abnahme an Stopps mit steigendem Wert für den Mindestdrehwinkel, in der Analyse der ausgewählten Trajektorie werden vier von fünf tatsächlichen Stopp ohne falsche Stopps gefunden.	In der allgemeinen Analyse ähnliche Resultate wie der Algorithmus der Geschwindigkeit, grundsätzlich jedoch etwas flachere Kurve, in der Analyse der ausgewählten Trajektorie werden alle tatsächlichen sowie einige falsche Stopps gefunden.
Austernfischer	keine Angaben	Maximal gefundene Stopps: 19	Maximal gefundene Stopps: 9	Maximal gefundene Stopps: 3.5
		In der allgemeinen Analyse mit einer Stoppdauer von über einer Minute konstante Anzahl an Stopps, ab einer Distanz von 10 Metern zur Berechnung der Nachbarschaft gleiche Anzahl an Stopps für alle alle Schwellenwerte für die Stoppdauer. In der Analyse der ausgewählten Trajektorie werden zwei tatsächliche sowie zwei falsche Stopps gefunden.	In der allgemeinen Analyse markante Reduktion an Stopps für eine Mindstdauer von einer Minute, die maximale Anzahl an Stopps wird bei erst einem minimalen Drehwinkel von 160 Grad gefunden, in der Analyse der ausgewählten Trajektorie werden zwei tatsächliche sowie zwei grosse, falsche Stopps gefunden.	In der allgemeinen Analyse werden weniger Stopps gefunden als mit dem Algorithmus der Geschwindigkeit, die Kurve schwankt relativ stark, in der Analyse der ausgewählten Trajektorie werden zwei tatsächliche sowie zwei falsche Stopps gefunden, wobei einer davon sehr klein ist.
Truthahngerier	keine Angaben	Maximale gefundene Stopps: 1400	Maximale gefundene Stopps: 1500	Maximale gefundene Stopps: 23
		In der allgemeinen Analyse ist eine sehr markante Reduktion an Stopps bei einer Stoppdauer von mindestens 20 Stunden zu erkennen. Die Anzahl an Stopps steigt grundsätzlich mit einer zunehmenden Distanz zur Berechnung der Nachbarschaft, wobei bei einer Mindeststoppdauer von 15 Stunden weniger steil als bei einer Mindeststoppdauer von 10 Stunden.	In der allgemeinen Analyse ist ein markanter Sprung bei einer Mindeststoppdauer von 20 Stunden zu erkennen, die Kurven mit einer minimalen Stoppdauer von über 20 Stunden sind relativ flach, die Kurven mit einer minimalen Stoppdauer von unter 20 Stunden steigen mit zunehmendem Schwellenwert für den minimalen Drehwinkel bis zu einem Maximum an.	In der allgemeinen Analyse wird bei einem tiefen Geschwindigkeitsschwellenwert die maximale Anzahl an Stopps gefunden, danach sinkt die Anzahl an Stopps zuerst steiler, dann flacher.
Auge	keine Angaben	Maximal gefundene Stopps: 13000	Maximal gefundene Stopps: 5500	Maximal gefundene Stopps: 41
		In der allgemeinen Analyse sind keine grossen Sprünge ersichtlich, so nimmt die Anzahl an Stopps mit zunehmender Mindeststoppdauer stetig ab, das Maximum an Stopps ist bei einer Distanz zur Berechnung der Nachbarschaft von 11 px zu erkennen.	In der allgemeinen Analyse verhalten sich alle Kurven für die unterschiedlichen Werte der Mindeststoppdauer bis zu einem minimalen Drehwinkel von 100 Grad gleich, danach steigt in allen Kurven die Anzahl gefundener Stopps bis zu einem Maximum, wobei keine grossen Sprünge ersichtlich sind.	Die Kurve der allgemeinen Analyse ist unruhiger, weist jedoch ein klares Maximum an gefundenen Stopps auf, die Kurve zeigt jedoch zwei Bereiche, die eine hohe Anzahl an Stopps zeigen.

ABBILDUNG A.2: Übersicht über die Resultate der Algorithmen aus der Literatur und des HMM.

## Anhang B

# Eigenschaften der verwendeten Datensätze

In der folgenden Tabelle B.1 sind die Eigenschaften der verwendeten Datensätze zusammengefasst. Es gilt zu beachten, dass sich die Anzahl an Punkten auf die Datensätze bezieht, die für die allgemeine Analyse verwendet wurden. Bei den Lastwagendaten wurden insgesamt 20 Datensätze verwendet, bei den Wanderdaten wurden 6 Datensätze verwendet. Weiter stellen sowohl die räumliche Skala als auch das zeitliche Aufnahmeintervall eine Größenordnung dar und variieren innerhalb der Datensätze. Die Einschränkungen können ergänzt werden, hier wurden lediglich einige, für diese Arbeit relevante Einschränkungen aufgeführt.

TABELLE B.1: Übersicht über die hier verwendeten Datensätze sowie deren Eigenschaften.

	Wanderdaten	Lastwagen- daten	Austernfischer- daten	Truthahngerier- daten	Eye-Tracking- Daten
Anzahl Datenpunkte in der allgemeinen Analyse	2889	7232	394	35256	85708
Räumliche Skala (Größenordnung)	5 km	15 km	50 m	10000 km	1 mm
Zeitliches Aufnahmeintervall (Größenordnung)	10 s	30 s	16 s	1 h - 3 h	0.017 s
Aufnahmedauer	9 h 8 min	ungefähr 7 Tage	1 h 54 min	2103 Tage	23 min 50 s
Genauigkeit	6.7 m	keine Angabe	keine Angabe	keine Angabe	keine Angabe
Einschränkung	Topographie, Netzwerk der Wanderwege	Strassennetzwerk, Verkehrsschilder, Verkehrs- teilnehmende	Umgebung	Umgebung	Bildschirmrand
<i>Ground truth</i>	vorhanden	abgeleitet mittels visueller Analyse	vorhanden	nicht vorhanden	durch Tracking- Programm eruiert

## Anhang C

# Genauigkeit und Trefferquote

In der folgende Tabelle sind die Werte für die Genauigkeit (*Precision*) und die Trefferquote (*recall*) der Analysen der ausgewählten Trajektorie enthalten. Es gilt zu beachten, dass sich die Werte bei den Wander- und Lastwagendaten auf die Anzahl an Stopps beziehen. Bei den Austernfischerdaten beziehen sich die Werte auf die Anzahl an Punkten. Das wurde so gemacht, da für den Austernfischer ein Stopp lediglich aus einem Punkt besteht. Würde nun von der Anzahl an Stopps ausgegangen werden, so würde dieser eine Stopp das Resultat verhältnismässig stark beeinflussen. Die Abkürzung p steht für *precision*, also Genauigkeit und die Abkürzung r steht für *recall*, also Trefferquote.

TABELLE C.1: Genauigkeit und Trefferquote der Resultate der ausgewählten Trajektorien.

	<b>Wanderdaten</b>	<b>Lastwagendaten</b>	<b>Austernfischerdaten</b>
Geschwindigkeit	p = 0.571; r = 1	p = 0.357; r = 1	p = 0.101; r = 0.867
Mindestdauer	p = 0.667; r = 1	p = 1; r = 1	p = 0.111; r = 0.933
Mindestdauer und Drehwinkel	p = 0.0.667; r = 0.5	p = 0.667; r = 0.4	p = 0.152; r = 0.933
Mindestausdehnung	p = 0.667; r = 1	p = 0.294; r = 1	p = 0.095; r = 0.933
Mindestausdehnung und Drehwinkel	p = 0.1; r = 0.5	p = 0.667; r = 0.4	p = 0.152; r = 0.933
IB-SMoT	keine Angabe	p = 0.667; r = 0.4	keine Angabe
CB-SMoT	p = 0.267; r = 1	p = 0.833; r = 1	p = 0.111; r = 0.8
DB-SMoT	p = 0.571; r = 1	p = 0.75; r = 0.6	p = 0.098; r = 0.8
HMM	p = 0.667; r = 1	p = 0.357; r = 1	p = 0.001; r = 0.067

## Persönliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und die den verwendeten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Alena Krähenbühl:

---

Datum:

---