

Routenwahl von Fahrradfahrern

Diskrete Entscheidungsmodellierung mittels GPS-Daten

Autor: Gianluca Menghini

Betreuer:

Prof. Dr. R. Weibel (Uni ZH),

Prof. Dr.-Ing. K.W. Axhausen (ETH), Dipl.-Ing. M. Löchl (ETH)

Fakultätsvertreter:

Prof. R. Weibel

Masterarbeit Universität Zürich, Mai 2008

MSc in Geographie, mit Spezialisierung in Geographischer Informationswissenschaft

Dank

Ich möchte mich besonders bei Prof. K.W. Axhausen, Prof. R. Weibel sowie Michael Löchl für die kompetente Betreuung während meiner gesamten Masterarbeit bedanken.

Ebenso ein Dankeschön gebührt Dr. Michael Balmer für die Implementierung des MATSim Moduls für die Berechnung der Routenalternativen sowie Frau Nadine Schüssler für die Hilfe beim Matching der GPS-Daten.

Für die grosse Unterstützung während des gesamten Studiums und im Speziellen während der Masterarbeit danke ich besonders meiner Familie und meinen engsten Freunden.

Inhaltsverzeichnis

1	Einführung	3
1.1	Motivation und Ausgangslage	3
1.2	Zielsetzung und Methodik.....	4
1.3	Bemerkungen	7
2	Literaturüberblick und Hintergrund	8
2.1	Räumliche Datenqualität.....	8
2.2	Identifizierung von Wegen und Verkehrsmittel aus GPS-Daten.....	13
2.3	Map Matching	14
2.4	Generierung eines Alternativensatzes.....	20
2.5	Einführung in die Entscheidungsmodellierung	23
2.6	Routenwahl und Verhalten von Fahrradfahrern	29
3	Datenlage	35
3.1	GPS-Daten am IVT	35
3.2	Strassennetz.....	36
3.3	Weitere GIS-Daten	39
4	Datenaufbereitung	41
4.1	Strassennetz.....	41
4.2	GPS-Daten	52
4.3	Ergänzung des Strassennetzes um weitere Eigenschaften.....	73
5	Ermittlung der Alternativen	79
5.1	Methodik	79
5.2	Ergebnisse.....	82
5.3	Diskussion	84
5.4	Ermittlung von Alternativen mit ArcGIS	85
6	Berechnung der Ähnlichkeit sich überlappender Routen.....	87

6.1	Methodik	87
6.2	Ergebnisse	89
6.3	Diskussion	92
7	Berechnung weiterer Routeneigenschaften.....	93
8	Deskriptive und räumliche Analyse.....	96
8.1	Eigenschaften des gesamten Strassennetzes	96
8.2	Räumliche Verteilung der gewählten Routen	97
8.3	Deskriptive Analyse des Schätzdatensatzes.....	102
9	Schätzung der Entscheidungsmodelle	118
9.1	Korrelationsanalyse	119
9.2	Modellschätzungen.....	120
9.3	Analyse der verschiedenen Modelle.....	132
9.4	Zusammenfassung der Modellergebnisse.....	148
9.5	Diskussion	149
10	Schlussfolgerungen und Ausblick.....	151
10.1	Beantwortung der Forschungsfragen	151
10.2	Ausblick.....	155
11	Literatur.....	159
12	Glossar	167

Tabellenverzeichnis

Tabelle 1	Wichtigste Faktoren bei der Routenwahl von Fahrradfahrern	30
Tabelle 2	Topologieüberprüfung: Anzahl der Fehler	42
Tabelle 3	Automatisches Matching: Optionen und Toleranzen	48
Tabelle 4	Strassentypen und Bestimmung der Buffergrösse	57
Tabelle 5	Ergebnisse nach der Filterung des GPS-Datensatzes	60
Tabelle 6	Ergebnisse nach der speziellen Filterung des GPS-Datensatzes	66
Tabelle 7	Matching-Eigenschaften	69
Tabelle 8	Rechenzeit für das Matching und Import der Daten	70
Tabelle 9	Statistische Grössen für die Festlegung der DTV-Kategorien	77
Tabelle 10	DTV-Kategorien	77
Tabelle 11	Anteil DTV-Kategorien für Route i	94
Tabelle 12	Dummy-Kodierung der DTV-Kategorien von Route i	94
Tabelle 13	Variablen des Schätzdatensatzes	95
Tabelle 14	Eigenschaften des gesamten Strassennetzes	97
Tabelle 15	Ergebnisse der Modellschätzung mit einzelnen Variablen	121
Tabelle 16	Nutzenfunktionen der linearen Modelle	123
Tabelle 17	Schätzergebnisse für das Modell (14)	123
Tabelle 18	Nutzenfunktionen der linearen Modelle mit Transformationen	125

Tabelle 19	Schätzergebnisse für die Modelle (15), (16), (17) und (18).....	125
Tabelle 20	Nutzenfunktionen der nichtlinearen Modelle	126
Tabelle 21	Schätzergebnisse für das Modell (20).....	126
Tabelle 22	Nutzenfunktionen mit Wechselwirkungstermen.....	127
Tabelle 23	Schätzergebnisse für das Modell (26).....	128
Tabelle 24	Nutzenfunktionen mit transformierten Wechselwirkungstermen	129
Tabelle 25	Nutzenfunktionen mit nichtlinearen Wechselwirkungstermen	129
Tabelle 26	Schätzergebnisse für das Modell (38).....	130
Tabelle 27	Schätzergebnisse für das Modell (39).....	131
Tabelle 28	Nutzenfunktionen mit nichtlinearen Wechselwirkungstermen	132
Tabelle 29	Beispiel Nutzendifferenz.....	141
Tabelle 30	Gewählte Routen mit grösstem Nutzen.....	143
Tabelle 31	Eigenschaftsäquivalente für das Modell (14).....	147
Tabelle 32	Attributkatalog von VECTOR25.....	A-1
Tabelle 33	Attributkatalog des Strassendatensatzes vom Kanton Zürich	A-1
Tabelle 34	Attributkatalog von MAP Zürich	A-3
Tabelle 35	Attributkatalog der Fahrradwege vom kommunalen Richtplan.....	A-4
Tabelle 36	Attributkatalog der GPS-Daten (für Import in PostGIS).....	A-4
Tabelle 37	Attributkatalog der nach dem Matching erhaltenen Routen	A-5

Tabelle 38 Statistik der Modellvariablen für den gesamten Schätzdatensatz	A-5
Tabelle 39 Statistik der Modellvariablen für die gewählten Routen.....	A-5
Tabelle 40 Statistik der Modellvariablen für die Alternativen.....	A-5
Tabelle 41 Spearman Rangkorrelationskoeffizient für ausgeschlossene Variablen ..	A-6
Tabelle 42 Spearman Rangkorrelationskoeffizient für die Modellvariablen	A-6
Tabelle 43 Elastizitäten des Modells (14).....	A-6
Tabelle 44 Elastizitäten des Modells (26).....	A-7
Tabelle 45 Elastizitäten des Modells (38).....	A-7
Tabelle 46 Elastizitäten des Modells (39).....	A-7
Tabelle 47 Beispiel für die Syntax einer BioRoute Eingabedatei	A-15
Tabelle 48 Beispiel einer BIOGEME Modelldatei.....	A-17

Abbildungsverzeichnis

Abbildung 1	Arbeitsschritte	6
Abbildung 2	MAP Zürich: Dangling nodes.....	43
Abbildung 3	MAP Zürich: Aufhebung der Dangling nodes	44
Abbildung 4	Roadmatcher Werkzeuge und Legende.....	47
Abbildung 5	Ausgangslage vor dem Matching.....	47
Abbildung 6	Manuelle Nachbearbeitung	49
Abbildung 7	Integriertes Strassennetzwerk.....	50
Abbildung 8	Matching Probleme und mögliche Fehler.....	52
Abbildung 9	Evaluation der GPS-Daten	53
Abbildung 10	Abstand der GPS-Punkte zum Strassennetz	54
Abbildung 11	Bildung des Strassenbuffers	56
Abbildung 12	Berechnung des Schwerpunktes.....	57
Abbildung 13	Dichtefilterung für einen Ausschnitt aus Etappe 1.....	58
Abbildung 14	Abstand der gefilterten GPS-Punkte zum Strassennetz	60
Abbildung 15	Mehrere Selbstverschneidungen einer Etappe	64
Abbildung 16	Maximale Neigung der GPS-Etappen	65
Abbildung 17	Geschwindigkeit verglichen mit der Steigung.....	65
Abbildung 18	Geschwindigkeit verglichen mit dem Gefälle.....	66

Abbildung 19	Matching der GPS-Punkte.....	70
Abbildung 20	Matching der GPS-Punkte: Probleme und Fehler	71
Abbildung 21	Strassenkanten-Auflösung	72
Abbildung 22	Lichtsignalanlagen an Strassenkreuzungen.....	74
Abbildung 23	Matching der Lichtsignalanlagen.....	75
Abbildung 24	Algorithmus für die Generierung des Alternativensatzes	81
Abbildung 25	Alternativen und gewählte Route fuer ein OD-Paar	82
Abbildung 26	Längendifferenz zwischen der gewählten und der kürzesten Route	83
Abbildung 27	Zwei Strecken mit gleichen Start- und Endknoten	84
Abbildung 28	Path-Size Faktor	89
Abbildung 29	Logarithmus des Path-Size Faktors	90
Abbildung 30	Path-Size für unterschiedliche Routenlängen	91
Abbildung 31	Log(Path-Size) für unterschiedliche Routenlängen.....	91
Abbildung 32	Gewählte Routen verglichen mit der Netzdichte	98
Abbildung 33	Gewählte Routen verglichen mit der Neigung.....	99
Abbildung 34	Anteil Routen pro Kreis	100
Abbildung 35	Normierter Anteil Routen pro Kreis	101
Abbildung 36	Länge der Alternativen und gewählten Routen	102
Abbildung 37	Durchschnittliche Steigung.....	103

Abbildung 38	Maximale Steigung.....	104
Abbildung 39	Kumulierte Höhenmeter aufwärts.....	105
Abbildung 40	Durchschnittliches Gefälle.....	106
Abbildung 41	Maximales Gefälle.....	107
Abbildung 42	Kumulierte Höhenmeter abwärts.....	108
Abbildung 43	Anteil markierter Fahrradwege an der Gesamtroute.....	109
Abbildung 44	Anteil Brücken an der Gesamtroute.....	110
Abbildung 45	Anteil Tunnels an der Gesamtroute.....	111
Abbildung 46	Anteil Naturbelag an der Gesamtroute.....	112
Abbildung 47	Anzahl Lichtsignalanlagen.....	113
Abbildung 48	Anteil DTV-Kategorie 1.....	114
Abbildung 49	Anteil DTV-Kategorie 2.....	115
Abbildung 50	Anteil DTV-Kategorie 3.....	116
Abbildung 51	Anteil DTV-Kategorie 4.....	117
Abbildung 52	Mittlere Nutzenzusammensetzung aller Routen.....	133
Abbildung 53	Mittlere Nutzenzusammensetzung der gewählten Routen.....	134
Abbildung 54	Mittlere Nutzenzusammensetzung der Alternativen.....	135
Abbildung 55	Einfluss der Anzahl Lichtsignalanlagen und max. Steigung.....	137
Abbildung 56	Einfluss der Ähnlichkeit und der max. Steigung.....	138

Abbildung 57	Einfluss der Länge und max. Steigung.....	139
Abbildung 58	Einfluss der Länge und des Anteils markierter Fahrradwege.....	140
Abbildung 59	Mittlere Nutzendifferenz zwischen den gewählten Routen und Alternativen	142
Abbildung 60	Elastizitäten Modell 14	145
Abbildung 61	Elastizitäten Modell 26	145
Abbildung 62	Elastizitäten Modell 38	146
Abbildung 63	Elastizitäten Modell 39	146
Abbildung 64	Box-Cox-Transformation für verschiedene λ	168
Abbildung 65	Dangling node	169
Abbildung 66	Strassendaten: Topologieprüfung und Fehlerkorrektur.....	A-1
Abbildung 67	Matching der Strassennetze.....	A-1
Abbildung 68	Allgemeine GPS-Daten Filterung	A-2
Abbildung 69	Spezielle GPS-Daten Filterung für Fahrradfahrer	A-3
Abbildung 70	Matching der Lichtsignalanlagen.....	A-4
Abbildung 71	Distanz zwischen der Strassenmitte und Mitte der Fahrspur	A-8
Abbildung 72	Distanz zwischen einem Punkt und einem Strassensegment.....	A-9
Abbildung 73	Pfaderzeugung bei Schnittstellen.....	A-12
Abbildung 74	Flussdiagramm des Map Matching Algorithmus	A-14

Abkürzungen

API	Application Programming Interface
BIOGEME	Bierlaire's Optimization Toolbox for GEV Model Estimation
CNL	Cross Nested Logit
ESRI	Environmental Systems Research Institute
GDAL	Geospatial Data Abstraction Library
GEV	Generalized Extreme Value
GML	Geography Markup Language
GNL	Generalized Nested Logit
GPS	Global Positioning System
GUI	Graphical User Interface
HTML	Hypertext Markup Language
JPEG	Joint Photographic Experts Group
JUMP	Java Unified Mapping Platform
IAP	Implicit Availability/Perception
INSPIRE	Infrastructure for Spatial Information in the European Community
IVT	Institut für Verkehrsplanung und Transportsysteme
KML	Keyhole Markup Language

KOGIS	Koordination, Geo-Information und Services
MMNL	Mixed Multinomial Logit
MNL	Multinomial Logit
NGEV	Network Generalized Extreme Value
NL	Nested Logit
NMEA	National Marine Electronics Association
ODBC	Open Database Connectivity
OD-Paar	Origin-Destination-Paar
OGC	Open Geospatial Consortium
SQL	Structured Query Language
VBA	Visual Basic for Applications
WKT	Well Known Text
XSLT	Extensible Stylesheet Language Transformation

Masterarbeit Studiengang Geographie (Fachrichtung GIScience)

Routenwahl von Fahrradfahrern

Diskrete Entscheidungsmodellierung mittels GPS-Daten

Gianluca Menghini
Andrau 9
7013 Domat/Ems

Telefon: +41 79 564 80 90
gianluca@menghini.biz

Mai 2008

Kurzfassung

In der vorliegenden Masterarbeit (erstellt am IVT der ETH Zürich) geht es um die Modellierung der Routenwahl von Fahrradfahrern. Für diese Untersuchung wird ein GPS-Datensatz verwendet, welcher von einer freiwilligen Zufallsstichprobe der Bevölkerung der Stadt Zürich erhoben wurde. Ziel der Arbeit ist es mit den vorhandenen GPS-Daten und dem zugehörigen detaillierten Strassennetz, Modelle zu schätzen, welche die Routenwahl der Beobachteten erklären. Die Verwendung von räumlich und temporal gut aufgelösten Daten eröffnet verglichen mit früheren Untersuchungen vollkommen neue Möglichkeiten.

Die Ergebnisse der verschiedenen Modelle werden eingehend analysiert und diskutiert. Das geeignetste Modell postuliert eine lineare Nutzenfunktion der Routeneigenschaften Länge, maximale Steigung, Anzahl Lichtsignalanlagen, Ähnlichkeit sowie Anteil Fahrradwege. Den grössten Entscheidungseinfluss hat die Routenlänge. Mit Ausnahme des Anteils Fahrradwege fließen alle Eigenschaften negativ in die Nutzenfunktion ein.

Schlagworte

Fahrradweg; Map Matching; Entscheidungsmodellierung; Routenwahl

Zitierungsvorschlag

Menghini, G. (2008) Routenwahl von Fahrradfahrern: Diskrete Entscheidungsmodellierung mittels GPS-Daten, *Masterarbeit*, Universität Zürich, Zürich.

1 Einführung

1.1 Motivation und Ausgangslage

„Beste Velo-Infrastrukturen mit dem PRIX VELO ausgezeichnet“, „Mit dem Velo an die EURO 08“, „Fuss- und Veloverkehr auf gemeinsamen Flächen“ (Pro Velo Schweiz 2008).

Diese drei exemplarischen Schlagzeilen offenbaren die kontinuierliche Bedeutungszunahme des Fahrrads in unserer Gesellschaft. Einerseits liegen gesundheitsfördernde Aspekte durch mehr Bewegung dahinter, andererseits spielen auch Gedanken an die Umwelt eine grosse Rolle für dieses zunehmende Interesse. Dass durch die Benutzung eines Fahrrads auch Geld für Treibstoff oder Fahrkarten gespart werden kann, ist ebenso eine schöne Nebensache. Zu guter Letzt kann Fahrradfahren aber auch einfach nur Spass machen und den persönlichen Gemütszustand verbessern.

Die genannten Vorzüge werden beispielsweise durch Aktionen wie „bike to work“ oder „bike to school“ gefördert. Erstere soll Leute animieren während eines Monats mit dem Fahrrad zur Arbeit zu fahren. Etwas jünger ist hingegen die Kampagne „bike to school“, welche Jugendliche zu vermehrtem Fahrradfahren anregen will (Pro Velo Schweiz 2008).

Nebst diversen individuellen Gegebenheiten und Umwelteinflüssen spielt die vorhandene Infrastruktur (Fahrradspuren und separate Fahrradwege) sicherlich eine entscheidende Rolle beim Erfolg eines solchen Projektes. Dill und Carr (2003) konnten beispielsweise zeigen, dass in grösseren US-Städten ein höheres Angebot an Fahrrad-Infrastruktur zu einer vermehrten Fahrradbenutzung anregt. Die genannten Gründe beeinflussen aber nicht nur den Aspekt der Radnutzung als solches, sondern auch die entsprechende Routenwahl.

Damit Stadtplaner in qualitativer und quantitativer Hinsicht die bestmöglichen Infrastrukturen zur Verfügung stellen, sowie weitere planen können, ist es wichtig die Präferenzen der Fahrradfahrer auszumachen. Ist bekannt, welche Routeneigenschaften (z.B. Steigung, Anzahl Lichtsignalanlagen etc.) für Fahrradfahrer relevant sind, kann die Infrastruktur der Nachfrage optimal angepasst werden.

Bisherige Erkenntnisse über die Routenwahl von Fahrradfahrern basieren auf *Stated Preference* Untersuchungen oder auf grob kodierte *Revealed Preference* Analysen. Entsprechende Literatur (z.B. Axhausen und Smith 1986; Bovy und Stern 1990; Hochmair 2004; Stinson und Bhat 2003 oder Tilahun 2007) ist daher kritisch zu beurteilen. Dies, weil die besagten Studien nur über eine beschränkte Verfügbarkeit von Beobachtungen verfügen oder auf hypothetischen Situationen beruhen.

Für diese Arbeit steht ein umfangreicher GPS-Datensatz über zurückgelegte Wege von einer freiwilligen Zufallstichprobe der Bevölkerung der Stadt Zürich zur Verfügung. Sowohl die räumliche als auch die zeitliche Auflösung dieser Daten eröffnen vollkommen neue Möglichkeiten bei der Modellierung der Routenwahl von Fahrradfahrern. Die bisherigen Studien kommen zum Schluss, dass separate Fahrradwege oder -spuren, kurze Distanzen, wenig Verkehrsaufkommen, gute Strassenbelagsqualität sowie geringe Steigungen zu den wichtigsten Routeneigenschaften für Fahrradfahrer gehören. Ob diese Erkenntnisse, welche wie erwähnt auf bestimmten Annahmen beruhen, gestützt und erweitert werden können, soll die hier vorliegende Arbeit untersuchen.

1.2 Zielsetzung und Methodik

Ziel der Arbeit ist mit dem vorhandenen GPS-Datensatz und dem zugehörigen detaillierten Strassennetz diskrete Entscheidungsmodelle zu schätzen, welche die Routenwahl der Beobachteten erklären.

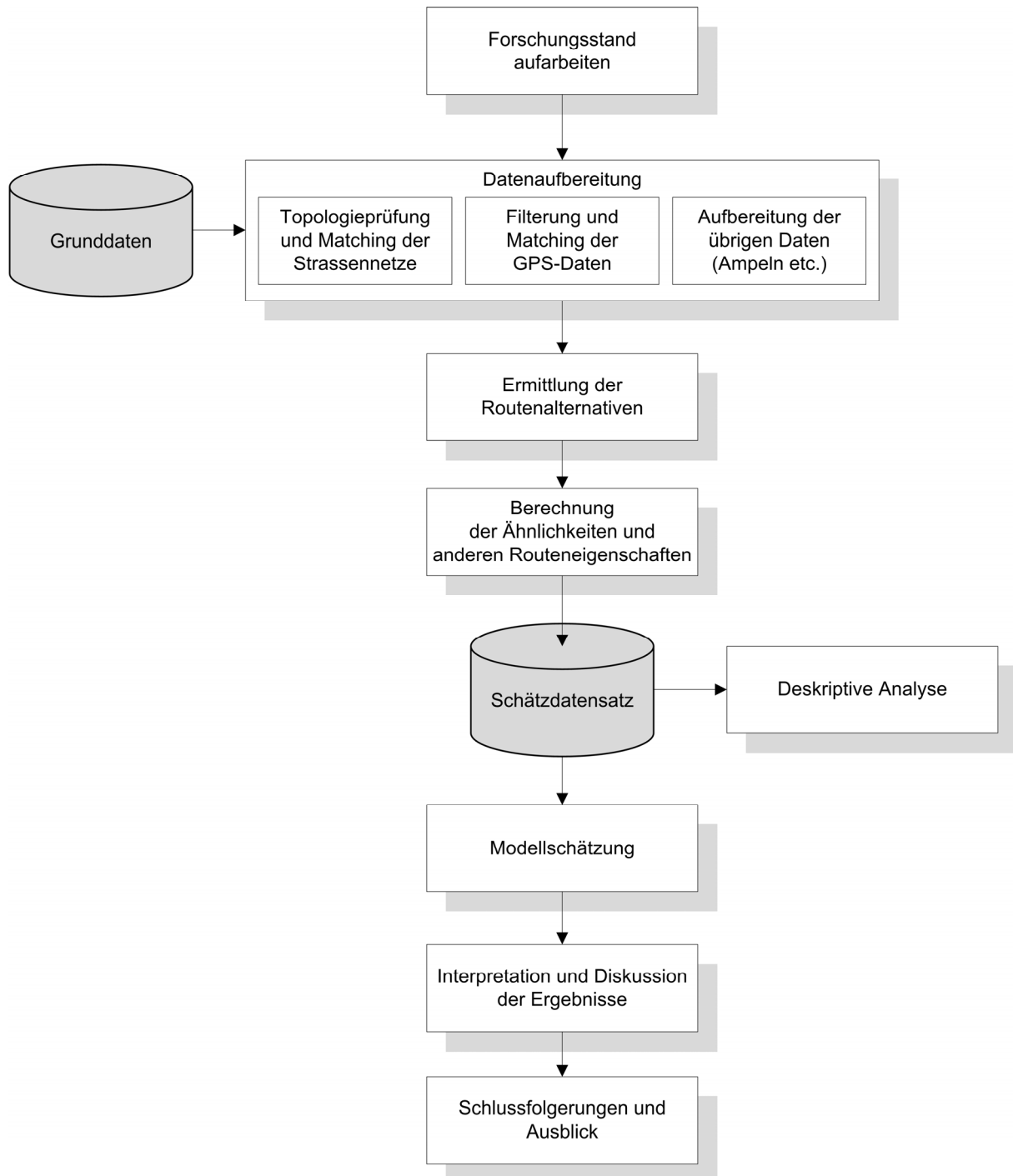
Das Treffen von Entscheidungen über die Wahl eines bestimmten Weges gehört zu den täglichen Routinen und wird heutzutage meistens in Form von diskreten Entscheidungsmodellen modelliert. Bei einem solchen Ansatz geht es immer um die Simulation der Wahloptionen eines Individuums. Die grundlegende Idee dahinter ist, dass ein Entscheidungsträger mit einem Satz von diskreten Alternativen konfrontiert wird und eine davon auswählen muss. Das Modell selbst schätzt für jede Alternative die Wahrscheinlichkeit gewählt zu werden. Dabei wird von der Annahme ausgegangen, dass der Entscheidungsträger seinen Nutzen stets maximieren will. Um das erwähnte Ziel zu erreichen sind verschiedene Arbeitsschritte notwendig, welche nachfolgend kurz behandelt werden.

Nach einem Überblick über den aktuellen Forschungsstand und der Formulierung entsprechender Forschungsfragen, erfolgt als Erstes die Aufbereitung der Daten. Dieser Arbeitsschritt beinhaltet diverse Datenbereinigungen sowie verschiedene Filterungs- und Mat-

chingschritte. Anschliessend wird für jede beobachtete Route ein Satz von Alternativen ermittelt. Weiter wird die Ähnlichkeit zwischen diesen Routen bestimmt sowie verschiedene Routeneigenschaften auf Basis des Strassennetzes berechnet. Schliesslich werden im letzten Schritt diskrete Entscheidungsmodelle geschätzt, welche erklären sollen, warum bestimmte Routen gewählt wurden. Die Ergebnisse der besten Modelle werden mit bisherigen Erkenntnissen verglichen und analysiert. Ebenso wird eine kritische Diskussion über die erzielten Resultate abgehandelt. Abgerundet wird die Arbeit mit einem Blick in die Zukunft sowie mit Hinweisen über offene Fragen oder Probleme.

Die folgende Grafik illustriert das generelle Vorgehen bei der gesamten Arbeit. Es soll als Leitfaden dienen und das Verständnis über die durchgeführten Arbeitsschritte verbessern.

Abbildung 1 Arbeitsschritte



1.3 Bemerkungen

In der folgenden Arbeit werden englische Begriffe, so weit das Verständnis nicht eingeschränkt wird, übersetzt. Im Kontext von Strassennetzen wird beispielsweise der englische Begriff *link* in das deutsche Wort Kante oder Strecke übersetzt. Existiert jedoch keine passende deutsche Übersetzung, wird das entsprechende Wort (z.B. Clip) in der ursprünglichen englischen Schreibweise belassen, um Missverständnisse zu vermeiden.

2 Literaturüberblick und Hintergrund

In den folgenden Abschnitten sollen die für diese Arbeit wesentlichen Theorien und Forschungsbereiche näher vorgestellt werden. Die Reihenfolge ist dem Ablauf der verschiedenen Arbeitsschritte angelehnt. Jeder Abschnitt wird mit einem kurzen Fazit und mit der Formulierung von ein bis zwei Forschungsfragen abgeschlossen. Die Beantwortung dieser Forschungsfragen erfolgt am Schluss der Arbeit.

2.1 Räumliche Datenqualität

Damit die verschiedenen Arbeitsschritte besser verstanden sowie hinterfragt werden können, wird hier auf die wesentlichsten Aspekte bezüglich Qualität von räumlichen Daten eingegangen. Der Abschnitt über die Qualität der GPS-Daten ist für diese Arbeit besonders relevant und wird bewusst speziell behandelt.

Eine der grössten Herausforderungen heutiger Zeit ist die grosse Anzahl vielfältiger, räumlicher Daten. Beinahe jeder und jede produziert räumliche Daten in irgendeiner Form. Dies geschieht meistens unabhängig voneinander mit dem Resultat, dass sich die Datensätze bezüglich Verfügbarkeit, Format, Genauigkeit, Metadaten etc. sehr stark unterscheiden.

Aufgrund der eben beschriebenen Problematik müssen die unterschiedlichen Daten einander angepasst werden. Dies erfordert eine Reihe von Integrationsprozessen, die gemäss Shepherd (1991) folgendermassen zusammengefasst werden können:

- Integration von Geometrie- und Sachdaten (Attribute)
 - Verknüpfung von Geometrie und Attributen über IDs oder *Spatial Join*
- Geometrische Integration verschiedener Objektklassen
 - Angleichung der Projektionen; *Edge matching*; Fusion verschiedener Datensätze der gleichen Region (*Map Conflation*); Zuordnung homologer Objekte über Massstäbe usw.
- Semantische Integration verschiedener Attributtabelle der gleichen Objektklassen
 - Homogenisierung von Attribut- und Klassendefinitionen über verschiedene Tabellen oder Datenbanken sowie Verknüpfung von Tabellen

- Integration verschiedener Datenmodelle
 - Konversion Raster > Vektor, Vektor > Raster, Raster > Quadtree usw.
- Integration von Datenbanken (DB):
 - Einbindung DB in GIS; Verbindung DB-DB (z.B. verteilte oder föderierte DBs)
- Import/Export (Datenaustausch) von Geometrie- und Sachdaten zwischen verschiedenen Systemen
 - System A zu System B; gemeinsamer Zugriff auf Datenserver usw.
- Integration der Funktionalität verschiedener Systeme
 - Verbindung verschiedener Systeme und Services (z.B. GIS und Fernerkundungs-Systeme) über gemeinsame Daten- oder Funktionalitätsschnittstellen (Interoperabilität)

Die Integration von Geodaten sollte nicht leichtsinnig durchgeführt werden. Bei jedem einzelnen Schritt können Fehler und Unsicherheiten eingeführt werden, die sich bei der weiteren Verarbeitung fortpflanzen und sich damit auf die Qualität der Daten auswirken. Damit mit diesen Unsicherheiten besser umgegangen werden kann, haben Burrough und McDonnell (1998) verschiedene Faktoren, welche die Qualität von Geodaten beeinflussen, definiert. Dabei unterscheiden sie folgende Faktoren:

- Aktualität
- Vollständigkeit
- Konsistenz
- Verfügbarkeit
- Genauigkeit und Präzision
- Quellfehler in den Daten
- Verarbeitungsfehler

Für eine detaillierte Abhandlung dieser Aspekte wird auf Burrough und McDonnell (1998) verwiesen.

Wichtig zu unterscheiden sind jedoch die beiden Begriffe Genauigkeit und Präzision. Veregin (1999) definiert die Genauigkeit wie folgt: „Beziehung (Übereinstimmung) zwischen einem Messwert und dem wahren Wert, der dadurch dargestellt werden soll. Der wahre Wert wird in

der Praxis ermittelt durch Referenzmessungen mit einer Messmethode oder Apparatur höherer Genauigkeit und Präzision.“ Die Präzision definiert er als: „Detaillierungsgrad mit der die numerische Darstellung eines Messwerts oder die rechnerische Manipulation desselben erfolgen kann.“

Als Antonym zur Genauigkeit steht der Begriff Fehler. Auf die verschiedenen Fehlerarten und Fehlermodellen wird hier nicht eingegangen, da dies den Rahmen der Arbeit sprengen würde. Eine gute Behandlung dieser Thematik ist in Burrough und McDonnell (1998) oder Veregin (1999) zu finden.

Die Gründe, weshalb räumliche Daten mit Unsicherheiten behaftet sind, können sehr vielfältig sein. So stellen Geodaten meistens Stichproben dar, welche die reale Welt repräsentieren sollen. Die Bestimmung eines „wahren“ Wertes ist häufig nur approximativ möglich. Vor allem bei metrischen Daten (Intervall-, Ratioskala) hängt dies von der Präzision des Aufnahme- und Verarbeitungsgerätes ab. Kategoriale Daten (Nominal-, Ordinalskala) oder Zähldaten können hingegen in gewissen Fällen genau sein. Die Bestimmung der Anzahl Studenten in einer Vorlesung ist beispielsweise genau möglich.

Betreffend die Problematik unterschiedlicher Datenqualität und -modelle gibt es in der Schweiz seit einigen Jahren vielversprechende Ansätze, wie beispielsweise diejenigen von KOGIS¹ bzw. e-geo.ch². In der Europäischen Union ist man mit der INSPIRE³ Richtlinie bereits weiter fortgeschritten. Mit dem neuen Geoinformationsgesetz⁴ sollte aber auch in der Schweiz diese Richtlinie vorzeitig sowie parallel und analog zur EU umgesetzt werden. Abschliessend gilt es jedoch zu erwähnen, dass man in der Praxis immer noch weit davon entfernt ist mit einheitlichen Datenmodellen und -beschreibungen, sogenannten Standards, arbeiten zu können.

¹ <http://www.swisstopo.admin.ch/internet/swisstopo/de/home/swisstopo/org/kogis.html> , Zugriff: 07.03.2008

² www.e-geo.ch , Zugriff: 07.03.2008

³ <http://inspire.jrc.it/index.cfm> , Zugriff: 07.03.2008

⁴ Bundesgesetz über die Geoinformation vom 5. Oktober 2007, SR 7155

2.1.1 Qualität von GPS-Daten

Die Genauigkeit von GPS wurde seit Mai des Jahres 2000 dank der Aufhebung der *Selective Availability* (SA) durch die US-Regierung grundlegend verbessert. Trotzdem ist die Qualität dieser Daten immer noch abhängig vom verwendeten GPS-Gerät (Antenne), von der Anzahl sichtbarer Satelliten, von der Anordnung derselben sowie von der Existenz möglicher Signalhindernisse und herrschenden Umweltbedingungen. In eng bebauten Stadtgebieten mit hohen Häusern treten Signalreflexionen ziemlich häufig auf und erzeugen fehlerhafte Punktwolken anstatt klaren Datenströmen. Weitere Probleme, welche zu Qualitätsverlusten führen, ergeben sich durch dichte Baumkronen, Datenlücken in Tunnels und Galerien sowie durch das sogenannte Kaltstartproblem. Letzteres deckt den Fall ab, dass kurz nach dem Einschalten des Gerätes die Datenaufzeichnung noch nicht funktioniert. Selbstverständlich ist die Grösse dieses Fehlers abhängig vom verwendeten Gerät. Ferner sei noch erwähnt, dass auch menschliche Fehler bei der Handhabung eines GPS-Gerätes auftreten können (Chung und Shalaby, 2004; Czerniak, 2002; Tsui und Shalaby, 2006).

Häufig werden qualitative Angaben wie die *Horizontal Dilution of Precision* (HDOP), die *Positional Dilution of Precision* (PDOP), die Anzahl sichtbarer Satelliten u.a. automatisch durch das jeweilige Gerät erhoben. Nachstehende Liste nach Czerniak (2002) fasst mögliche Fehler von GPS-Daten zusammen:

- Satellitenfehler
 - Zeitfehler der Satelliten (minimale Differenzen in den Uhren der Satelliten können grosse Fehler erzeugen)
 - Fehler durch Satellitendrift
- Atmosphärische Fehler
 - Ionosphärenfehler (Signalinterferenz in der oberen Atmosphäre)
 - Troposphärenfehler (Signalinterferenz in der unteren Atmosphäre)
- Operatorfehler (Benutzungsfehler)
 - Fahren auf der falschen Strassenspur
 - Falsche Handhabung des Gerätes
- Limiten des GPS-Geräts
 - Fehler beim Messen und Berechnen der Satellitenposition
 - Fehler durch Signalinteraktionen mit Gebäuden oder Vegetation
 - *Dilution of Precision* Fehler (GDOP, HDOP, PDOP, TDOP, VDOP)

Eine gute Abhandlung und Erklärung aller Fehlerquellen von GPS ist in Kowoma (2007) zu finden. Zusätzliche Angaben über die Techniken mobiler GPS-Geräte sowie allgemeine Erläuterungen bei der Verarbeitung von GPS-Daten mit Geographischen Informationssystemen (GIS) können in Czerniak (2002) nachgelesen werden.

In einigen Jahren sind durch das European Galileo System⁵ und den European Geostationary Overlay Service (EGNOS) zusätzliche Qualitätsverbesserungen zu erwarten.

2.1.2 Fazit und Forschungsfragen

Für die Verwendung der Geodaten dieser Arbeit ergibt sich folgende Situation: Quellfehler können leider nicht mehr beeinflusst werden. Trotzdem sollte versucht werden die richtigen Daten für den richtigen Zweck zu benutzen. Häufig wird dabei der Begriff „*fitness for use*“ für die Datenqualität gebraucht (Burrough und McDonnell, 1998). Damit wird die Eignung der Daten für die beabsichtigte Anwendung gemeint. Die bei den verschiedenen Arbeitsschritten entstehenden Verarbeitungsfehler können hingegen grösstenteils beeinflusst werden. Diesbezüglich sollte in gewissen Situationen aber auch ein wenig pragmatisch vorgegangen werden, da ein unendlicher Perfektionismus zu einer Zielverfehlung führen kann. Eine sorgfältige Überwachung sowie Dokumentierung muss in jedem Fall angestrebt werden. Ferner können Fehler bei der Verwendung der Daten entstehen. Diese liegen aber in der Verantwortung des Benutzers und dürfen nicht geduldet werden.

Auch die Qualität der GPS-Daten hängt wie weiter oben erwähnt von verschiedenen Faktoren ab. Obwohl die Genauigkeit ständig verbessert wurde und wird, sind die GPS-Daten teilweise noch fehlerhaft. In einem städtischen Gebiet wie Zürich mit engen Strassenschluchten und vielen Gebäuden ist die Problematik der Signalreflexion sicherlich eine nicht zu unterschätzende Tatsache. Weiter erschwerend kommt hinzu, dass in den Daten für diese Arbeit keine Qualitätsangaben wie z.B. die HDOP etc. vorhanden sind. Dies behindert die Eliminierung möglicher Fehler und Ungenauigkeiten bei der Vorverarbeitung der Daten.

Basierend auf diesen ersten Erkenntnissen können folgende Forschungsfragen formuliert werden:

⁵ http://ec.europa.eu/dgs/energy_transport/galileo/index_en.htm , Zugriff: 07.03.2008

Forschungsfrage 1: Wie wirken sich die qualitativ unterschiedlichen Daten auf die Integration sowie auf nachfolgende Berechnungen aus?

Forschungsfrage 2: Sind GPS-Daten für Untersuchungen über die Routenwahl von Fahrradfahrern geeignet?

2.2 Identifizierung von Wegen und Verkehrsmittel aus GPS-Daten

Die Detektion der Wege und Etappen sowie die Identifikation der Fahrradfahrer aus dem ursprünglichen GPS-Datensatz sind nicht Ziel dieser Arbeit. Dennoch sollen ein paar wichtige Punkte erwähnt werden, da die hier verwendeten GPS-Daten in einzelne Wege aufgeteilt sind und mit einer bestimmten Wahrscheinlichkeit Fahrradfahrten sind.

Die Wahl eines geeigneten Nachbearbeitungsansatzes hängt davon ab, ob die GPS-Daten Fahrzeug-basiert oder Personen-basiert aufgezeichnet wurden. Erstere Aufzeichnungsmethode führt ein GPS-Gerät im Fahrzeug mit und erhebt lediglich Daten, wenn der Motor des Gefährts läuft. Die Detektion von einzelnen Wegen erfolgt aufgrund von Zeitdifferenzen zwischen GPS-Punkten und ist ziemlich einfach. Trotzdem hat die Fahrzeug-basierte Aufzeichnung auch einige Nachteile, auf die aber nicht eingegangen werden soll.

Schüssler und Axhausen (2008) weisen darauf hin, dass in letzter Zeit Personen-basierte GPS-Studien an Popularität gewonnen haben. Bei dieser Aufzeichnungsart trägt eine Person das GPS-Gerät ständig auf sich und die Datenaufzeichnung erfolgt nicht nur während der Fahrt. Ein weiterer Vorzug dieser Methode ist, dass die fahrende Person eindeutig identifiziert werden kann und entsprechende soziodemographische Informationen für spätere Auswertungen hinzugefügt werden können. Jedoch erhöht sich der Nachbearbeitungsaufwand solcher Daten beträchtlich. Aus den Daten müssen zuerst einzelne Wege, welche aus mehreren Etappen bestehen können, ausgemacht werden. Des Weiteren wird nicht zwischen den benutzten Verkehrsmitteln unterschieden. Folglich muss anhand bestimmter Annahmen das verwendete Verkehrsmittel identifiziert werden.

Die in dieser Arbeit benutzen GPS-Daten (vgl. dazu Kapitel 3.1) wurden ebenfalls Personen-basiert aufgenommen und nach den Ansätzen von Schüssler und Axhausen (2008) nachbearbeitet. An dieser Stelle gilt es zu erwähnen, dass die Algorithmen für die Detektion einzelner Wege und Etappen sowie die Identifikation der Verkehrsmittel zum Zeitpunkt der Arbeit noch

nicht einwandfrei funktionierten. Folglich ist die Qualität der hier zur Verfügung stehenden Daten nicht immer zufriedenstellend. In den verschiedenen Datenaufbereitungskapiteln wird auf diesbezügliche Mängel hingewiesen.

Aufgrund der eben genannten Tatsache sowie der Gegebenheit, dass die Detektion einzelner Wege nicht die Aufgabe dieser Arbeit ist, wird auf eine detaillierte Erklärung dieser Algorithmen verzichtet und auf die Literatur (z.B. Schüssler und Axhausen, 2008 oder Tsui und Shalaby, 2006) verwiesen. Die Formulierung einer separaten Forschungsfrage wird ebenfalls unterlassen.

2.3 Map Matching

In diesem Kapitel soll etwas näher auf den bereits erwähnten Aspekt der Integration von Geodaten eingegangen werden. Nach Zhang *et al.* (2007) ist das *Map Matching* oder auch *Data Matching* eine der wichtigsten Methoden um verschiedene Datensätze zu integrieren.

2.3.1 Klassierung der Matching Algorithmen

Der Begriff *Map Matching* ist als der Prozess der Integration zweier räumlicher Datensätze definiert. Die dabei verarbeiteten *Matching*-Typen können Punkt-zu-Punkt, Punkt-zu-Linie, Linie-zu-Linie oder Polylinie-zu-Polylinie sein. Wird zudem die zeitliche Komponente als Kriterium einbezogen, können *Matching* Algorithmen auch eingeteilt werden in *offline Map Matching* oder *online Map Matching*. Letzteres führt die *Matching* Prozesse in sogenannter Echtzeit aus. *Offline Map Matching* versucht hingegen erst nach der Datenerhebung die einzelnen Datensätze einander anzupassen.

In jüngster Zeit wurden zahlreiche Algorithmen für verschiedene Verwendungszwecke und auf Basis unterschiedlicher Techniken entwickelt (Quddus *et al.*, 2007; Stigmar, 2005; Zhang *et al.*, 2007). Gemäss Quddus *et al.* (2007) lassen sich die in der Literatur vorgestellten *Matching Algorithmen* in vier Gruppen einteilen. Je nach angewandter Technik wird folgende Gliederung vorgenommen:

- Algorithmen auf Basis geometrischer Techniken
- Algorithmen auf Basis topologischer Techniken
- Algorithmen auf Basis probabilistischer Techniken

- Algorithmen auf Basis anderer fortgeschrittener Techniken

Geometrische *Matching* Algorithmen benutzen lediglich die Geometrie der Strassenkanten als Information. Die Art und Weise wie die Strecken miteinander verbunden sind, wird nicht berücksichtigt. Diesbezügliche Ansätze sind meistens ziemlich einfach und für die *Matching* Typen Punkt-zu-Punkt, Punkt-zu-Linie sowie Linie-zu-Linie entwickelt. Beispiele hierfür sind Bentley und Maurer (1980); Bernstein und Kornhauser (1996); Phuyal (2002); White *et al.* (2000) sowie Taylor *et al.* (2001).

In GIS verweist die Topologie auf die Beziehung zwischen den Objekten (Punkte, Linien und Polygone). Dabei unterscheidet man Nachbarschaft (bei Polygonen), Verbundenheit (bei Linien) oder Enthaltensein (bei Punkten in Polygonen) als Beziehung (Quddus *et al.* 2007). Ansätze, welche neben der geometrischen auch die hier genannten topologischen Informationen miteinbeziehen, gehören somit zur Gruppe der topologischen Algorithmen. Diesbezügliche Anwendungen sind zu finden in Chen *et al.* (2003); Greenfeld (2002); Meng (2006) sowie Quddus *et al.* (2003).

Die probabilistischen *Matching* Algorithmen definieren einen elliptischen oder rechteckigen Konfidenzbereich um die Position, die vom Navigationssensor (z.B. GPS-Gerät) gemessen wird. Ein erster entsprechender Ansatz wurde schon früh entwickelt und geht auf Honey *et al.* (1989) zurück. Weitere probabilistische Algorithmen wurden von Ochieng *et al.* (2004) und Zhao (1997) entwickelt.

Zur Gruppe der fortgeschrittenen *Matching* Algorithmen gehören Ansätze, welche raffinierte Methoden wie ein Kalman Filter, die Dempster-Shafer's Theorie (auch *belief theory* genannt), ein Fuzzy Logik Modell oder die Multiple Hypothese Techniken benutzen. Kalman Filter werden beispielsweise in den Ansätzen von Jo *et al.* (1996) sowie Kim *et al.* (2000) gebraucht. Methoden, welche die Dempster-Shafer's Theorie anwenden sind in Najjar und Bonnifait (2003) sowie Yang *et al.* (2003) vorgestellt. *Matching* Algorithmen auf Basis eines Fuzzy Logik Modells finden sich hingegen in Kim *et al.* (1998); Kim und Kim (2001); Quddus *et al.* (2006b); Syed und Cannon (2004) sowie in Zhao (1997). Zu guter Letzt sei noch auf Pyo *et al.* (2001) für ein Beispiel der Anwendung der Multiplen Hypothese Technik verwiesen.

Nicht nur die hier kurz vorgestellten Techniken unterscheiden sich grundlegend, auch deren Anwendungsbereiche sind ziemlich heterogen und Disziplinen übergreifend. Deshalb drängt sich auch eine Klassierung nach Anwendungsgebieten der Algorithmen auf. Zhang *et al.* (2007) schlagen folgende Einteilung vor:

- Erweiterung der Anwendungsmöglichkeiten existierender Daten durch einen Transfer von Attributen oder Objektklassen von einem Datensatz zum andern. Beispiele hierfür sind Devogele *et al.* (1998); Mantel und Lipeck (2004); Xiong und Sperling (2004) und Zhang *et al.* (2006).
- Evaluierung und Verbesserung der Datenqualität durch den Vergleich verschiedener Datensätze. Ein Beispiel findet man in Walter (1997).
- Unterhalt und/oder Aktualisierung von Datensätzen in Multirepräsentationsdatenbanken. Diesbezügliche Referenzen sind Anders und Bobrich (2004); Badard (1999); Gösseln und Sester (2003) sowie Volz (2006).
- Anbietet von Navigationslösungen für *Location Based Services*. Vgl. dazu Qudus *et al.* (2006) sowie Stigmar (2005).

An dieser Stelle wird darauf hingewiesen, dass die grosse Vielfalt und Anwendungsbereiche dieser Algorithmen eine vollständige Literaturabdeckung für diese Arbeit verunmöglichen. Allein zwischen 1989 und 2006 wurden mindestens 35 *Matching* Algorithmen publiziert (Qudus *et al.* 2007). Gleichwohl sollen nachfolgend einige Ansätze, ohne Rücksicht auf die genannten Klassierungsmöglichkeiten, näher vorgestellt werden.

2.3.2 Beispiele von Algorithmen

Der Ansatz von Devogele (1997) basiert auf einen dreistufigen Zuordnungsprozess für zwei Strassennetze mit leicht verschiedenen Massstäben. Im ersten Schritt werden die Strassen provisorisch einander zugeordnet. Dabei wird die Hausdorff-Distanz von den Strassen grösseren Massstabs zu den Strassen in kleinerem Massstab als Vergleichsgrösse verwendet. Der Schwellwert dieser Komponente wird schrittweise verkleinert und in jedem Iterationsschritt werden eindeutige Strassenzuordnungen bestimmt. Im zweiten Teil des Zuordnungsprozesses erfolgt die Verknüpfung von Knoten. Zwei Knoten werden einander zugeordnet, wenn im vorherigen Schritt alle zu diesen Knoten gehörenden Strassen miteinander vereinigt wurden. Treten mehrere Knoten mit nur teilweise zugeordneten Strassen als Kandidaten auf, so wird eine n:1-Knotenzuordnung definiert. Schliesslich erfolgt im letzten Teil des Integrationsprozesses die definitive Strassenzuordnung. Hierbei werden zwischen den Knoten die kürzesten Wege berechnet. Die kürzesten Wege zwischen vereinten Knoten werden dann zueinander zugeordnet.

In Walter (1997) wird ein geometrischer *Matching* Ansatz für GDF und ATKIS Daten vorgestellt. Es kombiniert verschiedene Methoden wie *Buffer Growing*, Winkel-, Längen- und Form-Vergleiche, um korrekte *Matching* Resultate zu erreichen. Diese Arbeit wurde von Volz

(2006) durch einen iterativen Ansatz erweitert. Dabei wird der *Matching* Prozess durch die Identifizierung von Strassen mit hoher Zuordnungswahrscheinlichkeit begonnen. Ein kombinierter Strecken und Knoten *Matching* Algorithmus detektiert anschliessend 1:1 Übereinstimmungen. Falls keine 1:1-Zuordnungen gefunden wurden, wird ein erweiterter Strecken *Matching* Algorithmus ausgelöst, der 1:2-Beziehungen erkennt. Der ganze Prozess verläuft in mehreren Iterationen mit schrittweise abgeschwächten Schwellwerten.

Der Ansatz von Stigmar (2005) basiert auf einem Programm, das auch in dieser Arbeit benutzt wird (vgl. Kapitel 4.2). Stigmar (2005) beschreibt, wie sich Routendaten in ein Echtzeit-Service-System importieren und mit diesen Daten matchen lassen. Hierzu werden die Inputdaten zuerst mittels XSLT (*Extensible Stylesheet Language Transformation*) zu GML (*Geography Markup Language*) transformiert, um sie dann in die Java Umgebung JUMP importieren zu können. Der Ansatz beruht auf die Verbesserung des bereits existierenden JCS (*Java Conflation Suite*) Algorithmus. Zu diesem Zweck wurden drei *Matching*-Erweiterungen (*Merge*-, *Topologie*- und *Buffer*-Erweiterung) entwickelt. Die Kombination der drei Erweiterungen mit dem ursprünglichen JCS Algorithmus zeigte dabei ansprechende Resultate. Trotzdem weist Stigmar (2005) darauf hin, dass die vorgestellte Lösung nur für die dort involvierten Daten optimal geeignet ist.

Der in Lüscher und Burghardt (2006) vorgestellte Algorithmus wurde für das *Matching* von Strassendaten stark unterschiedlicher Massstäbe konzipiert und erfolgreich für Daten des Massstabs 1:25000 und 1:200000 getestet. Der Algorithmus erzeugt zuerst einen Buffer und generiert so mögliche Kandidaten für Strassen und Knoten. Unwahrscheinliche Kandidaten werden anhand semantischer, geometrischer und topologischer Informationen herausgefiltert. Schliesslich werden die übrig gebliebenen Knoten durch geometrische Masse verglichen und 1:1 Verknüpfungen zwischen Knoten gebildet. Um Strassenverknüpfungen zu erhalten, werden die Knotenverknüpfungen durch einen modifizierten kürzesten-Pfad-Algorithmus verbunden.

Der von Waldner (2005) entwickelte Algorithmus wurde für die Integration eines Navteq-Netzes und des Strassennetzes vom Kanton Zürich entwickelt. Berechnet werden verschiedene *Matching* Kriterien, welche die Geometrie und Attribut Affinität von Strassen unterschiedlicher Netzwerke beschreiben. Diese werden dann in eine *Matching* Tabelle zusammengefasst. Die Methode wurde als Prototyp entwickelt und benutzt hierfür zwei Programme. In ArcGIS werden anhand eines VBA-Skriptes (durch Benutzung von ArcObjects) alle möglichen *Matching* Kandidaten in eine neue Feature Klasse geschrieben. Dabei sind alle Informationen über die geometrische Nähe Attribute dieser Feature-Klasse. Zusätzlich wird eine Kon-

figurations-Datenbank benutzt um die verschiedenen Netzwerke einander anzupassen. Wiederum mit Hilfe eines VBA-Skriptes werden in Access SQL-Abfragen erzeugt und die *Matching*-Kandidaten rangiert.

Für die Integration von Positions-Daten (z.B. GPS-Punkte) wurden in den letzten Jahren ebenfalls etliche Algorithmen entwickelt bei denen mehrheitlich Algorithmen auf Basis fortgeschrittener Techniken angewendet wurden.

Es kann generell gesagt werden, dass die Leistung dieser Algorithmen in jüngster Zeit verbessert wurde. Dies einerseits durch die Anwendung verbesserter Techniken beim *Map Matching* Prozess sowie andererseits aufgrund besserer Qualität der GPS-Daten und des Strassennetzwerkes. Trotz dieser Verbesserung sind die Resultate nicht bei jeder Anwendung befriedigend. Grössere Schwierigkeiten treten auf, wenn grosse Navigationsleistungen in komplexen und dichten städtischen Gebieten erforderlich sind (Quddus *et al.*, 2007).

In der Literatur wird immer wieder darauf hingewiesen, dass für ein effizientes *Map Matching* die Integration von Informationen über die Netzwerk-Topologie unerlässlich ist. Die von Pyo *et al.* (2001) vorgeschlagene Multiple Hypothese Technik (MHT) wurde beispielsweise für das *online Map Matching* beim Gebrauch eines GPS-Gerätes und eines *dead reckoning device* (Koppelungsgerät) entwickelt und greift unter anderem auf Informationen über die Netzwerk-Topologie zurück.

Neben beliebig komplexen Algorithmen gibt es auch einfachere Methoden wie die Suche des naheliegendsten Knotens oder der naheliegendsten Kante. Diese sind ziemlich schnell, doch ignorieren sie die Korrelation zwischen nachfolgenden Punkten. Folglich ist die Konsistenz der gesamten Route nicht gewährleistet.

Marchal *et al.* (2006) kritisieren, dass diese früheren Arbeiten zu sehr auf die Genauigkeit der Position auf dem Netzwerk fokussiert sind, als auf die Effizienz bezüglich der Rechenzeit. So seien viele Algorithmen für grosse Datensätze überhaupt nicht brauchbar. Zhang und Meng (2006) sind hingegen der Meinung, dass die Genauigkeit nicht zu unterschätzen sei und deshalb eine grosse Rolle spiele.

Aufgrund des oben genannten Einwandes versucht der von Marchal *et al.* (2006) vorgeschlagene Algorithmus die Rechenzeit zu verkürzen ohne jedoch die Genauigkeit der *Matching* Ergebnisse zu vernachlässigen. Die Hauptidee besteht darin einen Satz von Netzwerkpfaden zu definieren, die als Kandidaten für die GPS-Daten fungieren. Diese werden kontinuierlich

und schrittweise neu rangiert, wobei der Pfad mit der besten Rangierung am Schluss beibehalten wird. Rangiert werden die Pfade aufgrund eines Gewichtungswertes, der anhand des Abstandes der GPS-Punkte zur Strasse berechnet wird. Der Algorithmus ist abhängig von zwei Parametern, welche je nach Anwendung verschieden gewählt werden können. Im Anhang ist der Algorithmus mit all seinen Annahmen und Berechnungen detailliert beschrieben.

2.3.3 Fazit und Forschungsfrage

In dieser Arbeit wird für die Integration der verschiedenen Strassendaten der *Matching* Typ Polylinie-zu-Polylinie sowie für die nachfolgende Verarbeitung der GPS-Punkte die *Matching* Art Punkt-zu-Linie behandelt. Beim Betrachten des temporalen Aspektes kann ohne Zweifel behauptet werden, dass es sich in beiden Fällen um ein *offline Map Matching* Problem handelt.

Wie im vorherigen Kapitel dargestellt, gibt es eine Vielzahl von Algorithmen, die auf verschiedenen Techniken beruhen sowie unterschiedlich komplex sind. Meistens wurden die Algorithmen für eine spezielle Anwendung konzipiert und sind nicht einfach auf andere Probleme übertragbar. Die Matchingergebnisse wurden in den letzten Jahren kontinuierlich verbessert. Dies nicht zuletzt dank immer besserer Qualität der zu integrierenden Daten.

Um die Integrationsprobleme dieser Arbeit zu lösen werden zwei verschiedene *Matching* Algorithmen verwendet. Es ist nicht Gegenstand dieser Arbeit einen neuen Algorithmus zu entwickeln. Vielmehr sollen bereits bestehende Ansätze auf deren Tauglichkeit geprüft werden. Einerseits soll der von Vivid Solutions Inc. entwickelte Roadmatcher (Vivid Solutions Inc. 2007) und andererseits der Matching Algorithmus nach Marchal *et al.* (2006) benutzt werden. Näheres dazu findet sich im Kapitel 4. Die entsprechende Forschungsfrage zu diesem Kapitel lautet:

Forschungsfrage 3: Sind die verwendeten Matching Algorithmen für die Daten dieser Arbeit geeignet?

2.4 Generierung eines Alternativensatzes

Für die Schätzung eines diskreten Entscheidungsmodells ist ein Satz von Alternativen notwendig. Die wesentlichsten Aspekte und Ansätze für die Generierung von Alternativen werden nachfolgend kurz vorgestellt.

Bovy und Fiorenzo-Catalano (2006) definieren den Alternativensatz (*choice set*) als „... the collection of travel options available to an individual in satisfying his travel demand.“ In der Literatur sind sehr viele Ansätze für die Erzeugung von Alternativen beschrieben. Die meisten dieser Ansätze basieren auf Variationen des kürzesten Pfad Algorithmus. Dabei werden für die Generierung dieser Routen häufig verschiedene Techniken, wie beispielsweise die Minimierung objektiver Funktionen, die Formulierung heuristischer Regeln oder die Implementierung zufälliger Verfahren mit dem k -kürzesten Pfad Algorithmus kombiniert. Bevor einzelne Ansätze etwas näher vorgestellt werden, sollen zuerst einige allgemeine Aspekte über die Generierung von Alternativensätzen erläutert werden.

Bei der Erzeugung eines Satzes von Alternativen müssen gemäss Bovy und Fiorenzo-Catalano (2006) folgende spezifische Charakteristiken berücksichtigt werden:

- Die Anzahl verfügbarer Routen ist in dichten Netzwerken gewöhnlich sehr gross und die meisten Routen sind unbekannt.
- Der Satz plausibler und attraktiver Routen ist häufig sehr umfangreich. Dieser Datensatz ist zudem komplex wegen der Heterogenität der Routenzusammensetzung sowie unterschiedlicher physischer Überlappung zwischen den Routen.
- Wegen den genannten Aspekten ist es für den Analytiker schwierig abzuschätzen wie gross der Satz von Routen ist, welche attraktiv für den Reisenden und relevant in der Entscheidungsmodellierung sind.
- Das Verhalten bei Wahlmöglichkeiten zwischen Routen ist widersprüchlich und besteht aus verschiedenen Formen des Entscheidungsprozesses. Es kann sequentiell (von Entscheidungspunkt zu Entscheidungspunkt), simultan (vom Ausgangs- zum Zielpunkt in einem) oder strategisch (adaptive Wahl basierend auf den aktuellen Netzwerkbedingungen während der Reise) erfolgen.

Der Zweck der Generierung eines Routensatzes sollte ebenfalls betrachtet werden. Diesbezüglich unterscheiden Bovy und Fiorenzo-Catalano (2006) drei Anwendungszwecke:

- Wissenschaftliche Analyse von Reiseoptionen in Netzwerken. Der Planer oder Wissenschaftler ist an verfügbaren Routenalternativen, deren Zahl, Charakteristiken, Komposition etc. interessiert.
- Schätzung von Nachfragemodellen (z.B. Schätzung von Verhaltensparameter der Nutzenfunktion von Entscheidungsmodellen).
- Vorhersage von Wahlwahrscheinlichkeiten in einer Nachfrageanalyse, um die Grösse von Flüssen in Netzwerken zu ermitteln. Hierbei werden Entscheidungsmodelle mit bekannten Parametern aus Schätzungen benutzt.

Insbesondere für Vorhersagezwecke ist es unerlässlich, dass alle attraktiven Alternativrouten im Satz vorhanden sind. Die Vollständigkeit des Alternativensatzes für Schätzungszwecke ist, wie Ben-Akiva und Lerman (1985) gezeigt haben, hingegen etwas weniger wichtig.

Die verschiedenen Verwendungszwecke stellen unterschiedliche Anforderungen in Bezug auf Grösse, Komposition und Variabilität der Alternativen. Die damit verbundene Qualität spielt eine gewichtige Rolle bei der Schätzung eines Entscheidungsmodells. Die Genauigkeit der geschätzten Parameter ist abhängig von der Güte des Alternativensatzes.

Die Ermittlung von Alternativen ist kein geradliniges und einfaches Problem. Vor allem die unterschiedliche Grösse realer Strassennetze bereitet häufig Schwierigkeiten bei der Definition des Alternativensatzes. Der generierte Satz sollte unrealistische Pfade, die ein Reisender niemals wählen würde, ausschliessen. Ebenso sollten Pfade ausgeschlossen werden, die eine zu grosse Ähnlichkeit mit anderen haben, als dass sie von Reisenden unterschieden werden. Ein guter Algorithmus erzeugt demzufolge alle für einen Reisenden relevanten und heterogenen Routen. Insbesondere die tatsächlich gewählte Route müsste im Routensatz vorhanden sein (Bovy und Fiorenzo-Catalano, 2006; Prato und Bekhor, 2006).

2.4.1 Beispiele von Algorithmen

Der kompletteste Ansatz für Pfadgenerierungen beinhaltet unrealistischerweise alle physischen Routen, welche den Start- und Zielpunkt einer Route verbinden. Demgegenüber stehen die selektiven Ansätze, die deterministische und probabilistische Prozeduren für die Generierung eines realistischen Alternativensatzes benutzen. Da in der vorliegenden Arbeit ein deterministischer Ansatz benutzt wird (vgl. Kapitel 5), soll an dieser Stelle lediglich auf diesen Typ von Algorithmen eingegangen werden. Ein guter Überblick über probabilistische Ansätze ist beispielsweise in Cascetta *et al.* (1997) zu finden.

Der geradlinigste deterministische Ansatz sucht nach den k -kürzesten Pfaden. Diese kürzesten Wege Algorithmen gehen von der Kenntnis aller Streckenattribute aus. Eine diesbezügliche Erweiterung bzw. Anpassung ist in Van der Zijpp und Fiorenzo-Catalano (2005) zu finden. Um die plausibelsten k -kürzesten Pfade zu finden, definieren sie verschiedene Klassen von Einschränkungen und werten deren Ergebnisse aus.

Ein anderer Ansatz ist beispielweise derjenige von Ben-Akiva *et al.* (1984). Dabei werden alle Pfade, die für ein bestimmtes Kriterium die beste Route darstellen, markiert. Dieser Algorithmus geht von der Annahme aus, dass die Reisenden verschiedene Zielfunktionen besitzen. Jedes Kriterium gehört zu einer anderen bevorzugten Route und jede Route kann für unterschiedliche Zielfunktionen markiert werden.

Azevedo *et al.* (1993) definieren einen Ansatz, bei dem alle kürzesten Pfade vom Netzwerk eliminiert werden, um danach die nächsten kürzesten Pfade zu finden. Dies wird n Mal ausgeführt (n = Anzahl Ausführungen). Das Hauptproblem besteht darin, dass durch die Eliminierung von Strecken im Netzwerk Verbindungslücken entstehen können. Je nach gelöschter Kante existiert zwischen dem Start- und Zielpunkt dann keine Verbindung mehr. Eine Umgehung dieses Problems eliminiert lediglich einzelne Kanten oder eine Kombination einzelner Kanten vom kürzesten Pfad.

De la Barra *et al.* (1993) beschreiben einen Strecken-Bestrafungs-Ansatz (*link penalty approach*), in welchem die Kostenattribute für den kürzesten Pfad erhöht werden und danach der nächste kürzeste Pfad neu berechnet wird. Der ganze Prozess wird kontinuierlich ausgeführt bis kein neuer Pfad mehr gefunden wird. Park und Rilett (1997) modifizieren diesen Ansatz, indem lediglich für eine bestimmte Anzahl Kanten des kürzesten Weges die Kosten verändert werden.

Nebst den genannten Algorithmen gibt es auch noch komplexere Ansätze. Der von Prato und Bekhor (2006) vorgeschlagene Algorithmus benutzt die sogenannte *branch & bound* Technik, um einen Satz von Alternativen zu generieren. *Branching* Methoden selektieren eine Strecke, welche eine Verzweigung des vorher identifizierten kürzesten Pfades darstellt und konstruieren einen neuen kürzesten Pfad. Dieser führt vom Ausgangspunkt der Route zum Endknoten dieser Verzweigungskante und von dort wieder zum Endpunkt der ursprünglichen Route. Näheres zur *branch & bound* Technik ist in Prato und Bekhor (2006) zu finden. Überdies können weitere Beispiele in Dugge (2006) oder Ramming (2002) nachgelesen werden.

2.4.2 Fazit und Forschungsfrage

Ähnlich wie bei den Matching Algorithmen bieten sich auch bei der Generierung eines Alternativensatzes etliche Methoden an. Einerseits werden bei diesen Ansätzen sehr einfache und pragmatische Methoden verwendet. Andererseits gibt es aber auch weit fortgeschrittene Ansätze wie der *branch & bound* Algorithmus von Prato und Bekhor (2006). Empfehlungen zur Benutzung eines bestimmten Algorithmus sind schwierig abzugeben. Je nach Problemstellung muss einzeln beurteilt werden, welcher Ansatz möglicherweise geeignet ist.

Ebenfalls unterschiedlich sind die Anwendungszwecke und Anforderungen. Ein guter Algorithmus generiert keine unrealistischen Alternativen und liefert somit einen wichtigen Beitrag für eine qualitativ gute Modellschätzung.

Bei der Interpretation der Modellergebnisse dieser Arbeit sollte demzufolge die Entstehungsweise und die Qualität des Alternativensatzes mitberücksichtigt werden. Basierend auf die erwähnten Aspekte stellt sich folgende Forschungsfrage:

Forschungsfrage 4: Generiert ein einfacher Algorithmus einen für diese Arbeit brauchbaren Alternativensatz?

2.5 Einführung in die Entscheidungsmodellierung

Dieses Kapitel soll eine Einführung in die Thematik der diskreten Entscheidungsmodelle und dementsprechend einen Überblick über die grundlegende Terminologie und Theorien vermitteln. Die folgenden Ausführungen wurden, wenn nicht speziell erwähnt, wesentlich auf Basis von Bovy und Stern (1990); Ortúzar und Willumsen (2001) sowie Schüssler und Axhausen (2007) zusammengestellt.

Bis in die frühen 1980er Jahre wurden sogenannte *Aggregate demand (first generation)* Transport Modelle verwendet, die auf beobachtete Beziehungen für eine Gruppe von Reisenden oder auf die durchschnittliche Beziehung auf einer zonalen Stufe basieren. Erst danach begannen sich *disaggregate demand (second generation)* Modelle zu etablieren. Dies obwohl schon früher Warner (1962) oder Oi und Shuldiner (1962) auf die gewichtigen Nachteile der *first generation* Modelle aufmerksam machten. Die *second generation* Modelle fassen auf der beobachteten Wahl eines Individuums. Dies ermöglicht die Entwicklung realistischer Modelle, wie es heutzutage in Form der diskreten Entscheidungsmodelle gemacht wird. Grundsätz-

lich postulieren die Modelle, dass die Wahrscheinlichkeit eines Individuums eine bestimmte Wahl zu treffen, eine Funktion von dessen sozioökonomischen Charakter, der Entscheidungssituation und der relativen Attraktivität der Alternative ist.

Diskrete Entscheidungsmodelle werden in beinahe allen Bereichen der Verkehrsplanung angewendet. Eine diesbezügliche Pionierrolle nehmen Domencich und McFadden (1975) sowie Ben-Akiva und Lerman (1985) ein. Bei einem diskreten Entscheidungsmodell geht es immer um die Simulation der Wahloptionen eines Individuums. Die grundlegende Idee dahinter ist, dass ein Entscheidungsträger mit einem Satz von diskreten Alternativen konfrontiert wird und eine davon auswählen muss. Das Modell selbst schätzt für jede Alternative die Wahrscheinlichkeit gewählt zu werden. Dabei geht man von der Annahme aus, dass der Entscheidungsträger seinen Nutzen stets maximieren will.

Um die Attraktivität einer Alternative repräsentieren zu können, bedient man sich des Konzepts des Nutzens. Der Nutzen jeder Alternative ist charakterisiert durch deren messbare Attribute, die im systematischen Teil V_{in} der Nutzenfunktion U_{in} enthalten sind. Zudem können soziodemographische Attribute des Individuums hier integriert werden. Diejenigen Nutzenkomponenten, die nicht direkt gemessen werden können, werden durch den Zufallsterm ε_{in} erfasst. Diese Zufalls-Komponente entsteht wegen der Heterogenität der Präferenzen zwischen den Individuen, dem unvollständigen Wissen sowie der limitierten Information der Entscheidungsträger. Des Weiteren können auch noch anderen Unsicherheiten mitenthalten sein. Die Nutzenfunktion U_{in} einer Alternative i für eine Person n ist demgemäss wie folgt definiert:

$$U_{in} = V_{in} + \varepsilon_{in} \quad (1)$$

$V_{in} = f(\beta, x_{in})$	systematischer, messbarer Anteil
β	Vektor der Geschmackskoeffizienten
x_{in}	Vektor der Attribute der Alternative i , welche vom Individuum n wahrgenommen werden (soziodemographische Attribute des Individuums können inbegriffen sein)
ε_{in}	Zufallsterm, nicht messbarer Anteil

Die Wahrscheinlichkeit, dass das Individuum n die Alternative i wählt, ist gegeben durch:

$$P(i|C_n) = P(U_{in} \geq U_{jn} \quad \forall j \in C_n) \quad (2)$$

C_n Satz von Routen

U_{in}, U_{jn} Nutzen von Alternative i bzw. j für das Individuum n

Wie weiter oben bereits erwähnt, ist der Nutzen also abhängig von den individuellen Präferenzen des Entscheidungsträgers, der Wahlsituation, der Charakteristik der Alternative und deren Ähnlichkeit mit den anderen verfügbaren Alternativen. Um die Wahl einer Alternative vorhersagen zu können, muss der Wert der Nutzenfunktion in eine Wahrscheinlichkeit zwischen 0 und 1 transformiert werden. Für diesen Zweck existiert eine Vielzahl mathematischer Transformationen, die typischerweise eine S-förmige Kurvenform haben.

2.5.1 Multinomial Logit Model (MNL)

Das einfachste und gebräuchlichste Modell ist das von McFadden (1974) vorgeschlagene Multinomial Logit Model (MNL). Es geht von der Annahme aus, dass die Zufallsterme, oftmals auch Fehlerterme genannt, identisch und unabhängig Gumbel verteilt sind (vgl. Glossar). Die Wahl-Wahrscheinlichkeit jeder Alternativen kann demgemäss wie folgt berechnet werden:

$$P_{in} = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}} \quad (3)$$

$$\mu = \sqrt{\frac{\pi}{6\sigma^2}} \quad \mu = 1; \text{ beim Fehlen einer heterogenen Population}$$

Die Vorteile des MNL Modells sind dessen Flexibilität und die einfache Schätzung der Parameter. Gleichzeitig hat dieses Modell aber auch einen gewichtigen Nachteil, nämlich die IIA-Eigenschaft (vgl. Glossar), wonach die Alternativen unabhängig voneinander betrachtet werden. Dies kann mit dem einfachen „roter Bus - blauer Bus - Paradox“ von Debreu (1960), das zwei Verkehrsmittelwahl-Situationen beschreibt, illustriert werden. In der ersten Situation hat der Entscheidungsträger zwei Optionen: Entweder den roten Bus zu wählen oder das Auto. Dabei wird angenommen, dass die Wahl-Wahrscheinlichkeit jeder Alternative 50% beträgt. Im zweiten Szenario kommt zusätzlich ein blauer Bus, mit denselben relevanten Attributen

wie der rote Bus, hinzu. Da der blaue Bus lediglich eine weitere Alternative zu öffentlichen Verkehrsmitteln ist, würde man als Wahl-Wahrscheinlichkeiten 50% für das Auto und je 25% für die beiden Busse erwarten. Die zusätzlichen Frequenzen im Bus-Netzwerk, aufgrund des blauen Busses, werden dabei vernachlässigt. Wegen der IIA-Eigenschaft modelliert das MNL aber für jede Alternative die gleiche Wahrscheinlichkeit von 33%, dies damit der Quotient zwischen den Wahrscheinlichkeiten für das Auto und den Bussen gleich 1 bleibt. Somit werden die Ähnlichkeiten des roten und blauen Busses komplett ignoriert. Um das beschriebene Problem zu beheben, wurden in jüngster Zeit verschiedene Ansätze entwickelt, die im nächstfolgenden Kapitel näher vorgestellt werden sollen. Dies und auch die Frage, ob Ähnlichkeiten zwischen den Alternativen einen positiven oder negativen Effekt haben, ist Gegenstand aktueller Forschung.

2.5.2 Ähnlichkeit der Alternativen bei diskreten Entscheidungen

In realen Situationen können ganz unterschiedliche Ähnlichkeiten, die abhängig vom Entscheidungskontext sind, auftreten. Im Falle des privaten Verkehrs können Ähnlichkeiten (mathematisch Korrelationen) beispielsweise durch gleiche Wegstücke auftreten. Im öffentlichen Verkehr treten hingegen Ähnlichkeiten durch gleiche Wegstücke, vergleichbare Zeitfenster, gleiche Umsteigemöglichkeiten oder durch den gleichen Anbieter auf. Ein Individuum, das beispielsweise vor der Wahl einer bestimmten Feriendestination steht, sieht sich noch mehr Ähnlichkeiten gegenüber. Diese können die gleiche geographische Region, die Landschaft, die Reiseroute, das Wetter, die angebotenen Produkte/Services etc. betreffen. Selbstverständlich treten Korrelationen auch in ganz anderen Bereichen wie zum Beispiel bei der Wahl einer bestimmten Arbeitsstelle auf.

Modelle, welche die Ähnlichkeit der Alternativen berücksichtigen

Bei der Behandlung der Ähnlichkeiten zwischen Routenalternativen gibt es drei generelle Ansätze. Ersterer teilt die Alternativen in sogenannte Nests (Gruppen) auf. Der zweite Ansatz öffnet die Varianz-Kovarianz Struktur und die dritte Lösung integriert Ähnlichkeitsfaktoren im systematischen Teil der Nutzenfunktion.

Die erste Gruppe beinhaltet grundsätzlich die Generalized Extreme Value (GEV) Modelle. Hier werden die Alternativen in Gruppen (Nests) aufgeteilt. Beim Nested Logit (NL) Modell können beispielsweise Korrelationen in den einzelnen Gruppen, aber nicht zwischen den Gruppen bestehen. Es gibt aber auch Modelle, wie zum Beispiel das Cross Nested Logit (CNL) Modell, welche Korrelationen zwischen den Gruppen abbilden können. Weitere Mo-

delle, die zu dieser Gruppe gehören sind das Generalized Nested Logit (GNL) Modell sowie das Network GEV (NGEV) Modell. Obwohl Modelle wie das CNL beinahe alle erdenklichen Korrelationsarten abbilden können, ist eine realistische Nesting-Struktur sehr komplex und nur mühsam zu schätzen.

Am meisten Forschung wird gegenwärtig für die zweite Gruppe, im Speziellen für die Mixed Multinomial Logit (MMNL) Modelle betrieben. In MMNL Modellen wird der deterministische Teil der Nutzenfunktion neu formuliert, wobei die Gumbel verteilten Fehlerterme des MNL unverändert bleiben. Um trotzdem alle Arten von Korrelationsstrukturen und Geschmacks-Heterogenitäten abbilden zu können, wird zusätzlich ein multivariat zufallsverteilter Fehlerterm eingeführt. Dieser erfasst alle modellierbaren Ähnlichkeiten. Ein weiteres zu dieser Gruppe gehörendes Modell ist das etwas einfachere Multinomial Probit Modell, welches lediglich die Gumbel verteilten Fehlerterme des MNL durch multivariat normalverteilte Fehlerterme ersetzt. Die hier beschriebenen Ansätze und im Besonderen das MMNL erfordern grosse Anstrengungen betreffend der Spezifizierung, Identifizierung und Rechenleistung. Deshalb sind sie für Wahlsituationen mit vielen Alternativen nur schlecht anwendbar.

Die dritte hier vorgestellte Gruppe versucht Korrelationseffekte durch eine Korrektur der systematischen Komponente der Nutzenfunktion zu erfassen. Der Fehlerterm bleibt dabei Gumbel verteilt. Basierend auf das von Cascetta *et al.* (1996) vorgestellte *implicit availability/perception* (IAP) Modell geht man von der Annahme aus, dass der Nutzen einer Alternative mit dem Ähnlichkeitsgrad zu einer anderen Alternative abnimmt. Der ausschlaggebende Aspekt dieser Ansätze ist die adäquate Wahl des Ähnlichkeitsfaktors. Da in der Modellschätzung dieser Arbeit (vgl. Kapitel 9.2) ebenfalls ein Ähnlichkeitsfaktor in die Nutzenfunktion einfließen soll, wird im nächsten Unterkapitel etwas näher auf diese Thematik eingegangen.

Ähnlichkeitsmasse

Sowohl das *Path-Size-Logit* als auch das C-Logit Modell modifizieren die Nutzenfunktion durch einen Ähnlichkeitsfaktor, so dass der Nutzen sich überlappender Routen verkleinert wird. Der Nutzen unabhängiger Routen wird dementsprechend erhöht. Der Unterschied dieser zwei Modelle, welche häufig in der Schätzung der Routenwahl im Strassenverkehr Verwendung finden, liegt in der Berechnung des Ähnlichkeitsfaktors. Im Rahmen dieser Arbeit soll der *Path-Size* Faktor berechnet werden, welcher im Kapitel 6 definiert ist.

Folgende Modelle haben verschiedene Ähnlichkeitsmasse, auf die nicht weiter eingegangen wird, im systematischen Teil der Nutzenfunktion implementiert:

- C-Logit Modell
- *Path-Size* (PS) Logit Modell
- *Competing Destinations* (CP) Modell
- *Prospective Utility* (PU) Modell
- *Independence of a Connection* (IND) *factor* Modell
- *The concept of dominance*

Die Idee, Ähnlichkeitsfaktoren im systematischen Teil der Nutzenfunktion einzubauen, ist sehr attraktiv wegen ihrer Einfachheit und Eleganz. Anstatt die Wahlmöglichkeiten a priori zu strukturieren und riskieren, dass falsche Annahmen über Korrelationen getroffen werden, wird nur der Typ der Ähnlichkeit spezifiziert. Praktische Anwendungen haben gezeigt, dass die IIA-Eigenschaft gut berücksichtigt wird und dass die Modelle auch für eine grosse Anzahl von Alternativen relativ leicht gerechnet werden können.

Nichtsdestotrotz haben die Modelle dieser Gruppe einige Nachteile. So werden zum Beispiel Geschmacksunterschiede nicht berücksichtigt. Das grösste Manko ist die Tatsache, dass diese Modelle für eine spezifische Wahlsituation konstruiert sind und gewöhnlich gewisse Korrelationsaspekte zwischen Alternativen vernachlässigt werden.

2.5.3 Fazit und Forschungsfragen

Wie das Beispiel mit dem roten und blauen Bus gezeigt hat, ist es wichtig bei der Schätzung eines diskreten Entscheidungsmodells die Ähnlichkeit der Alternativen einzubeziehen. Diese kann sich auf verschiedene Aspekte beziehen und für dessen Berücksichtigung gibt es drei generelle Ansätze.

Besonders schwierig ist es einen guten Kompromiss zwischen einem einfachen und komplizierten Modell zu finden. Letztere liefern meistens bessere Ergebnisse als einfache Modelle, doch die Spezifikation derselben erweist sich häufig als grosse Schwierigkeit. Des Weiteren kann auch eine „*black box*“ entstehen, bei der das Zustandekommen der Resultate nicht nachvollziehbar ist.

Aktuelle und zukünftige Forschungsarbeiten beschäftigen sich nebst der Spezifikation von immer besseren Modellen auch intensiv mit dem Einfluss der Ähnlichkeit der Alternativen.

Im Rahmen dieser Arbeit soll der *Path-Size* Faktor berechnet werden und in die Nutzenfunktion einfließen. Die entsprechende Forschungsfrage lautet wie folgt:

Forschungsfrage 5: Hat die Ähnlichkeit zwischen den Alternativen einen positiven oder negativen Einfluss auf die Routenwahl von Fahrradfahrern?

Eine weiter zu untersuchende Frage betrifft die Güte der Modellergebnisse. Wie in den vorherigen Kapiteln erwähnt, hängt die Qualität der Schätzungsergebnisse u.a. von den verwendeten Daten und deren Aufbereitungsart, der Qualität des Alternativensatzes sowie vom Entscheidungsmodell selbst ab. Infolgedessen formuliert sich die nächste Forschungsfrage folgendermaßen:

Forschungsfrage 6: Wie könnte das geschätzte Entscheidungsmodell verbessert werden?

2.6 Routenwahl und Verhalten von Fahrradfahrern

In den letzten Jahren wurden zahlreiche Studien über die Routenwahl und das Verhalten von Fahrradfahrern publiziert. Die Untersuchungen verfügen häufig nur über eine beschränkte Anzahl Beobachtungen und/oder basieren auf hypothetischen Situationen. Folglich sind diese Untersuchungen kritisch zu beurteilen und mit Vorsicht zu genießen. Nichtsdestotrotz sollen in diesem Abschnitt die wichtigsten Ergebnisse zusammengefasst werden.

2.6.1 Untersuchungen zur Routenwahl

Hunt und Abraham (2007) haben eine vollständige Übersicht über die bedeutsamsten Einflüsse auf das Verhalten von Fahrradfahrern, insbesondere bei der Wahl einer bestimmten Route publiziert. In dieser Tabelle werden einerseits die wichtigsten Eigenschaften eines Fahrradweges und andererseits die relevanten Umwelt- und soziodemographischen Einflüsse aufgelistet. Alle Faktoren sind mit entsprechenden Literaturreferenzen belegt. Folgende Tabelle fasst diese Determinanten nochmals zusammen, ohne die Literaturreferenzen einzeln zu nennen. Diese finden sich in Hunt und Abraham (2007).

Tabelle 1 Wichtigste Faktoren bei der Routenwahl von Fahrradfahrern

Faktor	Faktor Beschreibung
Weg- Charakteristiken	Typ des Fahrradweges (gemischt mit anderem Verkehr, Fahrradspur oder separater Fahrradweg)
	Strassencharakteristik (Strassenklasse, Sichtdistanzen, Kurvenradien)
	Existenz von Parkplätzen am Strassenrand
	Strassenbelag und/oder Qualität desselben
	Steigung
	Abstände von Strassenkreuzungen und/oder Konfiguration derselben
	Behandlung von Fahrradfahrern an Signalen, inklusive Wartezeiten und Identifizierung
	Vollständigkeit und Direktheit von Fahrradwegen
	Verfügbarkeit von Duschen am Start- und Zielort
	Verfügbarkeit von Fahrrad-Parkplätzen am Start- und Zielort
Nicht Fahrradspezifische Charakteristiken	Geschwindigkeit des motorisierten Verkehrs und Verhalten der Lenker
	Verkehrsvolumen oder Verhältnis der Verkehrsmittel, inklusive Anteil Lastwagen
	Interaktion mit Fussgängern
Individuelle und Routen- Charakteristiken	Geschlecht
	Alter
	Einkommen
	Grad an Fahrrad-Erfahrung
	Besitz eines Fahrrades
	Sicherheit auf der Strasse
	Persönliche Sicherheit
	Flexibilität bezüglich der Arbeitszeit
	Fahrradtyp (Strassenrad oder Mountainbike)
	Preis des Fahrrads
Routenlänge, Zeit und Distanz	
Umwelt- und sonstige Charakteristiken	Wetter
	Landnutzungsart am Wegrand
	Schönheit der Route
	Grad der politischen Unterstützung des Fahrradfahrers

Grad der öffentlichen Unterstützung des Fahrradfahrers, inklusive Karten, Routenempfehlungen und sanitätischen Nothilfeleistungen
Ausbildung und Förderung von Fahrradfahrern
Verfügbarkeit öffentlicher Verkehrsmittel
Kosten und Umstände durch die Benutzung anderer Verkehrsmittel

Nach: Hunt und Abraham (2007)

Die Aufzählung illustriert sehr schön welche grosse Anzahl an Faktoren die Verkehrsmittelwahl und im speziellen die Routenwahl von Fahrradfahrern beeinflussen können. Hierbei gilt es aber zu beachten, dass nicht jeder Faktor gleich grossen Einfluss hat.

Schon vor einiger Zeit konnten Axhausen und Smith (1986) darauf hinweisen, dass bei Fahrradfahrern qualitative Faktoren eine besondere Rolle bei der Routenwahl einnehmen. Dies weil sie den verschiedenen Umwelteinflüssen mehr exponiert sind als zum Beispiel Autofahrer. Die Ergebnisse ihrer *Stated Preference* Untersuchung zeigen, dass das Verkehrsvolumen, welches stellvertretend für die Sicherheit steht, die wichtigste Determinante für unerfahrene Fahrradfahrer ist. Erfahrene Radfahrer sehen hingegen die Strassenbelagsqualität als wichtigsten Faktor. Der Grund liegt daran, dass höhere Tempi damit verbunden sind. Weiter konnte der Einfluss des Alters belegt werden. Jüngere Fahrradfahrer sind weniger sensibel auf Steigungen als ältere. Die Landnutzung und das Vorhandensein eines separaten Velowegs sind hingegen für alle Altersklassen ungefähr gleich wichtig.

Die *Revealed Preference* Untersuchung von Bovy und Stern (1990) kommt zum Schluss, dass die Distanz und Reisezeit für Fahrradfahrer am wichtigsten ist. Mehr als 20% der gewählten Routen war gleichzeitig auch der kürzeste Weg. Weitere relevante Faktoren für die Wahl einer bestimmten Route waren der Strassenbelag, das Verkehrsvolumen sowie die Sicherheit. Da die Studie in einer mittelgrossen niederländischen Stadt (Delft) erfolgte, weisen Bovy und Stern (1990) darauf hin, dass in anderen Ländern die Topographie (z.B. Steigung) unbedingt mitberücksichtigt werden sollte.

Shafizadeh und Niemeier (1997) zeigen, dass die soziodemographischen Faktoren Alter, Geschlecht und Einkommen den grössten Einfluss auf die Routenwahl von Fahrradfahrern im Pendlerverkehr haben. Zudem ergründen sie, dass separate Velowege wichtig sind und zu längeren Routen animieren. Die Studie basiert auf einer Befragung über den Start- und Zielort der gewählten Route.

Aultman-Hall *et al.* (1997) belegen in ihrer Untersuchung (Kombination von *Revealed Preference* und *Stated Preference* Daten), dass Fahrradfahrer hauptsächlich grössere Steigungen, Eisenbahnkreuzungen und hohe Verkehrsvolumina vermeiden. Brücken spielen im Gegensatz dazu keine signifikante Rolle.

Die Ergebnisse der *Stated Preference* Studie von Tilahun *et al.* (2007) decken auf, dass in der Routenwahl von Radfahrern separate Fahrradspuren den grössten Einfluss haben. Die nächst wichtigen Determinanten sind die Absenz von Parkplätzen entlang der Strasse sowie das Ausweichen auf einen separaten Fahrradweg. Überdies stützen sie die Aussage von Shafizadeh und Niemeier (1997), wonach das Alter, Geschlecht und Einkommen zu den gewichtigsten soziodemographischen Faktoren gehören.

Stinson und Bhat (2003) ermitteln die Reisezeit als wichtigste Determinante bei der Routenwahl von Fahrradfahrern (*Stated Preference* Untersuchung). Einen grossen signifikanten Einfluss haben zudem die Verfügbarkeit einer Fahrradspur oder eines separaten Fahrradwegs, der Verkehrsgrad sowie die Qualität der Strasse bzw. des Strassenbelags.

Die Untersuchung von Dill und Carr (2003) konnte, abgesehen von gewissen Limiten, zeigen, dass neu gebaute Fahrradwege Pendler zur Nutzung eines Fahrrads animieren (Studie basiert auf Zensus-Daten). Ein höheres Angebot an Fahrradwegen hatte in grösseren US Städten einen positiven signifikanten Einfluss. Des Weiteren illustrieren sie die Notwendigkeit der Verbindung von wichtigen Destinationen (z.B. Stadtzentren) durch Fahrradwege. Ebenso wird auf die meisten in der Tabelle 1 aufgelisteten Faktoren hingewiesen.

Interessante Ergebnisse liefert auch die *Stated Preference* Studie von Hochmair (2004). Die 12 bedeutsamsten Kriterien seiner Befragung, sind nach der Wichtigkeit geordnet, die folgenden:

- Fahrradwege
- Sehenswürdigkeiten
- Kurze Routen
- Vermeidung grossen Verkehrs
- Pärke
- Vermeidung steiler Strassen
- Nebenstrassen

- Seen und Flüsse
- Schnelle Routen
- Guter Strassenzustand
- Einfache Routen
- Gute Signalisierung

Wie erwähnt basieren viele Analysen auf hypothetischen Situationen oder sehr kleinen Datensätzen und können deshalb nicht als repräsentativ betrachtet werden. Ein weiteres Problem betrifft die Tatsache, dass Fahrradfahrer und Fussgänger häufig zusammen betrachtet werden. So zum Beispiel in Hyodo *et al.* (2000) sowie Krizek und Johnson (2006). Erstere Studie weist darauf hin, dass beispielsweise in Japan städtische Verkehrsplanungen keinen Unterschied zwischen Fussgängern und Fahrradfahrern machen. Ausserdem ist nicht aus jeder Untersuchung klar, ob mit Fahrradweg eine Fahrradspur auf einer normalen Strasse oder ein separater Weg gemeint ist.

2.6.2 Fazit und Forschungsfragen

Über die Routenwahl von Fahrradfahrern wurden zahlreiche Studien publiziert mit einer beinahe unübersichtlichen Anzahl von Verhaltensdeterminanten. Nachfolgende Aufzählung fasst die am häufigsten genannten Gründe zusammen:

- Separate Fahrradwege oder -spuren
- Kurze Reisezeiten bzw. Distanzen
- Geringer Verkehr
- Gute Strassenbelagsqualität
- Geringe Steigungen

Obwohl bei den Daten dieser Arbeit keine soziodemographischen Informationen vorhanden sind (vgl. Kapitel 3.1), sollen auch diesbezüglich die bedeutsamsten Determinanten nochmals genannt werden. Es sind dies das Alter, das Geschlecht und das Einkommen.

Wegen den erwähnten einschränkenden Voraussetzungen, auf welchen die meisten Studien beruhen, sollten die bisherigen Untersuchungsergebnisse kritisch beurteilt und immer auf die spezifische Situation relativiert werden. Bezüglich dieser Problematik werden in dieser Arbeit

einige Fortschritte erwartet. Der zur Verfügung stehende GPS-Datensatz eröffnet qualitativ vollkommen neue Möglichkeiten, die es in bisherigen Studien nicht gab. Somit besteht die Hoffnung, dass Ergebnisse früherer Untersuchungen kritisch beurteilt und sinnvoll ergänzt werden können. In diesem Zusammenhang lassen sich folgende Forschungsfragen formulieren:

Forschungsfrage 7: Welches sind die wichtigsten Determinanten bei der Routenwahl und was für einen Einfluss haben diese bei der Routenwahl von Fahrradfahrern?

Forschungsfrage 8: Sind die Ergebnisse dieser Arbeit mit Resultaten aus früheren Studien vergleichbar?

3 Datenlage

Die in dieser Arbeit verwendeten Daten stammen aus verschiedenen Erhebungen und Quellen. Einerseits werden GPS-Daten aus einer Zufallsstichprobe der Bevölkerung der Stadt Zürich benutzt sowie andererseits verschiedene Strassendatensätze und GIS-Layer. Letztere stammen von der Stadt Zürich (Tiefbauamt), vom Kanton Zürich (Baudirektion, GIS-Zentrum) sowie vom Bund (Swisstopo). Die GPS-Daten wurden für eine externe Studie erhoben und sind am IVT vorhanden. Eine ziemlich grosse Herausforderung war die Tatsache, dass die verschiedenen Datensätze grosse Qualitätsunterschiede aufweisen.

3.1 GPS-Daten am IVT

Die hier verwendeten GPS-Daten wurden im Jahr 2004 in den Städten Genf, Winterthur und Zürich durch eine freiwillige Zufallsstichprobe der Bevölkerung erhoben. Insgesamt haben 4882 Personen, welche im Durchschnitt rund 6.5 Tage unterwegs waren, an der Studie teilgenommen. Es ist wichtig zu erwähnen, dass die Daten ursprünglich für einen anderen Zweck erhoben wurden. Die Erhebung wurde von den privaten Plakatfirmen der Schweiz in Auftrag gegeben mit dem Ziel einer Qualitätsbewertung ihrer Plakatstellen. Zusammengefasst ging es darum zu messen wie viele Personen an einem bestimmten Plakatstandort vorbei gehen. So wurde die Plakatnutzung mittels eines GPS-Gerätes passiv gemessen. Infolge der eben beschriebenen ursprünglichen Nutzung der Daten, sind pro GPS-Punkt lediglich X,Y,Z-Koordinaten sowie ein Zeitstempel als Information vorhanden. Soziodemographische Daten wurden leider nicht erhoben. Wie sich später bei der Modellschätzung zeigen wird, ist dies ein Nachteil. Solche Informationen wären für diese Arbeit sehr interessant gewesen.

Gemäss Herrn Hagin⁶, welcher am ursprünglichen Projekt beteiligt war, weist das verwendete GPS-Gerät eine Genauigkeit von 2-3 m auf. Er betont zudem, dass besonders in dicht bebauten städtischen Gebieten mit Lageungenauigkeiten von bis zu 10 m zu rechnen sei. Auf die Ursache dieser Fehler wurde bereits in Kapitel 2.1.1 hingewiesen. Die Geräte wurden von der Firma GEOSAT SA in Lausanne angefertigt und sind nicht im Handel erhältlich. Eingebaut

⁶ Mitarbeiter der Firma GEOSAT SA in Lausanne (<http://www.geosat.ch> , Zugriff: 01.12.2007)

wurden GPS-Chips der Firma ublox⁷, welche die Position jede Sekunde speichern, wobei gemäss Hersteller eine Frequenz von bis zu 4 Hz möglich wäre. Des Weiteren sind die Geräte bis zu 10 Stunden ohne nachzuladen verwendbar.

Die Daten wurden aufgrund bestimmter Annahmen, welche detailliert in Schüssler und Axhausen (2008) erklärt sind, in einzelne Wege und Etappen aufgeteilt. Die Wahrscheinlichkeit für die Wahl eines bestimmten Verkehrsmittels erfolgte grundsätzlich auf Basis von Geschwindigkeits- und Beschleunigungsannahmen. Für die hier vorliegende Arbeit standen die daraus hervorgegangenen Fahrradetappen zur Verfügung. Es handelt sich um alle Punkte mit einer Wahrscheinlichkeit für eine Fahrradfahrt grösser als Null. Aufgrund des grossen Umfangs wurden die Daten auf die Stadt Zürich beschränkt. Die entsprechende Struktur ist aus Tabelle 36 im Anhang ersichtlich.

Es sei nochmals darauf hingewiesen, dass der Algorithmus für die Identifikation der Wege und Verkehrsmittel zum Zeitpunkt dieser Arbeit noch nicht vollends ausgearbeitet war.

3.2 Strassennetz

Das für die Untersuchungen dieser Arbeit zugrunde liegende Strassennetz stammt aus vier verschiedenen Quellen. Der Grund liegt darin, dass ein annähernd vollständiges Strassennetz (inkl. Fahrradwege) sowie möglichst viele Sachdaten angestrebt wurden. Um dieses Ziel zu erreichen, wurden verschiedene Datensätze, so zum Beispiel auch Teleatlas- oder Navteq-Daten begutachtet. Mit dem Ergebnis, dass das VECTOR25 Strassennetz des Bundesamtes für Landestopographie Swisstopo, die Strassendaten des Kanton Zürichs sowie die Fahrradwege der Stadt Zürich die besten Informationen bezüglich Geometrie und Sachdaten erbringen. Die Daten über die Fahrradwege der Stadt Zürich sind in zwei verschiedene Datensätze aufgeteilt, welche unterschiedliche räumliche und nicht-räumliche Informationen liefern, dies obschon beide die Fahrradwege als Thematik haben. Der unterschiedliche Verwendungszweck ist die Ursache für diesen Unterschied. Des Weiteren wird in diesen Daten kein Unterschied zwischen einer Fahrradspur auf einer normalen Strasse und einem separaten Fahrradweg gemacht. Deshalb wird künftig der Begriff „markierter Fahrradweg“, welcher für beide Typen steht, gebraucht.

⁷ <http://www.u-blox.com>, Zugriff: 22.11.2007

Nachfolgend werden die einzelnen Datensätze genauer vorgestellt. Ein Überblick über die Attribute derselben ist in den Tabellen 32-35 zu finden.

3.2.1 VECTOR25 Landschaftsmodell der Schweiz

Die VECTOR25 Daten von der Swisstopo decken die gesamte Schweiz sowie das angrenzende Ausland ab und entsprechen dem Perimeter der Landeskarte 1:25000. Für die vorliegende Arbeit wurden die Daten auf die Stadt Zürich beschränkt. Als thematische Ebene ist nur das Strassen- und Wegenetz mit 35 Objektarten interessant. Gemäss Datenherrn hängt die Genauigkeit von der zugrunde liegenden Landeskarte 1:25000 ab. Die Lagegenauigkeit der Punkte und Linien liegt innerhalb von 3 bis 8 m, wobei an wenigen Stellen aus kartographischen Gründen (Generalisierung) grössere Abweichungen auftreten können. Verschiedene Konsistenzbedingungen betreffend Attributwertebereichen und Topologie stellen die problemlose Verwendung der Daten sicher. Insgesamt umfasst die Qualitätssicherung die Aspekte Klassierung (Objektarten), Topologie, Geometrie und Vollständigkeit. Die Datennachführung erfolgt blattweise im 6-Jahres-Zyklus, analog der Landeskarte. Die hier verwendeten Daten sind aus dem Jahr 2000, obschon eine neuere Version bereits vorhanden wäre (Swisstopo 2008).

3.2.2 Strassendaten Kanton Zürich

Der zweite wichtige Datensatz ist derjenige des Kantons Zürich. Sämtliche National- und Staatsstrassen sowie die wichtigsten Gemeindestrassen des Kantons sind darin enthalten. Hier gilt es jedoch anzumerken, dass diese Daten besonders wegen den Verkehrsmodellzahlen von grossem Interesse sind. Die geometrischen Informationen von VECTOR25 werden nur an wenigen Stellen ergänzt. Der Grund liegt darin, dass zum Beispiel nur ein Teil der Gemeindestrassen erfasst wurde. Für diese Arbeit ist wiederum nur der Perimeter der Stadt Zürich relevant. Die visuelle Beurteilung anhand eines Orthophotos (vgl. Kapitel 3.3.2) ergab grössere Lageungenauigkeiten als der VECTOR25 Datensatz. Dies ist erstaunlich, da gemäss Dokumentation des Kantons Zürich alle Strassen auf Basis des Übersichtsplans 1:2500 digitalisiert wurden. Der Stand des Strassennetzes ist vom Jahr 2001, die Verkehrsmodellzahlen stammen hingegen vom Jahr 2003. Analog des VECTOR25 Datensatzes lagen auch diese Daten in Form eines ESRI Shapefiles vor. Für weitere Informationen sei auf die Datendokumentation des Kantons Zürich verwiesen (Kanton Zürich, Amt für Raumordnung und Vermessung (ARV) 2007).

3.2.3 Fahrradwege MAP Zürich

Die Daten aus MAP Zürich, Mobil in Zürich + Stadtpläne beziehen sich auf die gleichnamige, analoge Karte. Herausgeber ist das Tiefbauamt der Stadt Zürich (Stadt Zürich, Tiefbauamt 2007). Die technische Bearbeitung wurde von der Firma Hallwag Kümmerly + Frey AG übernommen. Für die Bearbeitung der Velothematik war hingegen die IG Velo Zürich zuständig. Abgegeben wurden die Daten in einzelnen ESRI Shapefiles, welche je eine thematische Ebene repräsentieren. Der Stand der Karte bzw. Daten bezieht sich auf das Jahr 2004. Digitalisiert wurde auf Basis einer Karte im Massstab 1:10000 vom Vermessungsamt der Stadt Zürich. Die Wichtigkeit dieser Daten liegt darin, dass die beiden ersten Datensätze sowohl räumlich (durch die ausgewiesenen Fahrradwege) als auch über verschiedene Sachdaten ergänzt werden. Auch hier sind die einzelnen Attribute im Anhang aufgelistet. Es gilt wie bereits erwähnt zu beachten, dass nicht unterschieden wird zwischen einer Fahrradspur und einem separaten Fahrradweg parallel zur Strasse. Lediglich bei den Wegen mit Naturbelag kann mit Sicherheit geschlossen werden, dass es sich um einen separaten Weg handelt.

Anders als bei den ersten beiden Datensätzen ist die Qualität dieser Daten unbefriedigend. Fehlende topologische Informationen (z.B. Start- und Endknoten) sowie *Dangling Nodes* (vgl. Abb. 2) erschweren den Gebrauch dieser Daten.

3.2.4 Fahrradwege aus dem kommunalen Richtplan der Stadt Zürich

Die Fahrradwege aus dem kommunalen Richtplan der Stadt Zürich stammen ebenfalls vom Tiefbauamt der Stadt Zürich. Obschon auch diese Datenquelle nur die Fahrradwege innerhalb der Stadt Zürich beinhaltet, gibt es einige Unterschiede zum Vorherigen. Die hier vorgestellte Datenquelle stammt aus dem kommunalen Richtplan und enthält Massnahmen, die umgesetzt wurden oder werden müssen. Somit stellt es ein politisches Instrument dar, im Gegensatz dazu ist die MAP Zürich lediglich als Empfehlung für die Fahrradfahrer und ohne Anspruch auf Vollständigkeit gedacht. Deshalb stehen hier zwei verschiedene Informationen zur Verfügung, die als Kombination die bestmöglichen Angaben gewähren sollen.

Analog zu den anderen Daten lag diese Datenquelle ebenfalls als ESRI Shapefile vor. Topologische Informationen sind keine vorhanden. Digitalisiert wurden die Wege im Jahr 2004 auf Basis der Landeskarte 1:25000. Vom Tiefbauamt wurde eine Darstellung grösser als 1:5000 strengstens abgeraten, da die Lagegenauigkeit nicht sehr gut sei. Diesbezügliche Angaben in Meter können jedoch nicht gemacht werden. Stattdessen beruht diese Aussage auf einer visuellen Überprüfung (Stadt Zürich, Tiefbauamt 2007).

3.3 Weitere GIS-Daten

Zu den ausserdem verwendeten GIS-Daten zählen Daten, welche die Strassennetze um weitere Eigenschaften ergänzen oder für Visualisierungszwecke genutzt werden. Je genauere und je mehr Daten hierbei zur Verfügung stehen, desto besser wird das am Schluss geschätzte Modell, da theoretisch mehr Routeneigenschaften in die Nutzenfunktion einfliessen können.

3.3.1 Lichtsignalanlagen

Die Daten bzw. Standorte der Lichtsignalanlagen wurden von der Stadt Zürich, Dienstabteilung Verkehr (2007) erhoben. Als Grundlage für die Digitalisierung wurde eine 7 Jahre alte Karte (im Rasterformat) der Stadt Zürich von Orell Füssli verwendet. Abgegeben wurden die Daten im CAD-Format dxf ohne räumliches Bezugssystem. Alle ein bis zwei Jahre werden die Daten überprüft und allfällige Ergänzungen vorgenommen. Letztmals erfolgte dies im Februar 2006. Weitere Genauigkeitsangaben sind nicht vorhanden.

3.3.2 Orthophotos

Die verwendeten Orthophotos (SWISSIMAGE) des Bundesamtes für Landestopographie Swisstopo (2008b) weisen eine Auflösung von 0.5 m (Pixel am Boden) auf und stammen von der Flugperiode Ende Mai/Anfangs Juni 2002. Der vorgesehene Nachführungszyklus beträgt sechs Jahre. Die JPG-Bilder haben eine Bildtiefe von 24 Bit (16.7 Mio. Farben) und eine Lagegenauigkeit von 1 bis 3 m (in topographisch schwierigen Gebieten 4 bis 10 m).

3.3.3 Stadtgrenze Zürich

Zum Zwecke der Vereinheitlichung des Perimeters der verschiedenen Datensätze wurden die Gemeindegrenzen der Schweiz (GG25) beigezogen. Enthalten sind sämtliche Landes-, Kantons-, Bezirks- und Gemeindegrenzen der Schweiz in vektorieller Form. Als Format stand ein ESRI Shapefile mit einer Lagegenauigkeit von 3 bis 8 m (analog der Landeskarte 1:25000) zur Verfügung (Swisstopo 2008c). Aufgrund des festgelegten Ziels dieser Arbeit ist nur die Grenze der Stadt Zürich interessant.

3.3.4 DTM-AV

Um die Steigung bzw. das Gefälle der verschiedenen Routen berechnen zu können, wurde das digitale Höhenmodell DTM-AV von der Schweizerischen Landestopographie Swisstopo benutzt. Aufgenommen wurde es mittels Lasermessungen aus Flugzeugen bis zu einer Höhenlage von 2000 m ü. M. Die Höhengenaugigkeit beträgt ± 0.5 m 1σ . Im Gegensatz zum DOM-AV bildet dieses Höhenmodell die Topographie der Erdoberfläche ohne Bewuchs und Bebauung ab. Ausgeliefert wurde das Modell als ASCII X,Y,Z (single space) mit einer Maschenweite von 2 m (Swisstopo 2008e).

3.3.5 Stadtkreise

Die für unterschiedliche Visualisierungen gebrauchten Kreisgrenzen der Stadt Zürich lagen als ESRI Shapefile vor. Gemäss Dokumentation des BFS wurden für dessen Erfassung die Landeskarten 1:25000 sowie verschiedene Karten, Pläne und Unterlagen in grösseren Massstäben verwendet. Angaben über die Genauigkeit dieser digitalen Vektordaten fehlen. Es ist aber anzunehmen, dass diese im Rahmen der übrigen hier benutzten Datensätzen liegt (Bundesamt für Statistik BFS 2008).

4 Datenaufbereitung

Das in Kapitel 2.1 dargelegte Problem unterschiedlicher Qualität von räumlichen Daten trifft auch auf die hier verwendeten Daten zu. Vor dem Gebrauch dieser Grundlagedaten müssen sie also aufbereitet und zusammengespielt werden. Ziel ist es, dass am Ende die bestmöglichen räumlichen und nicht-räumlichen Informationen zur Verfügung stehen.

Die Datenaufbereitung beinhaltet eine Qualitätsüberprüfung mit anschließender Datenverbesserung und/oder -filterung sowie die Integration der verschiedenen Datensätze.

4.1 Strassennetz

Vor der Integration der Strassennetze wurden die Daten auf mögliche Topologiefehler untersucht und wo nötig korrigiert. Erst danach konnten die einzelnen Netze einander zugewiesen werden.

4.1.1 Topologieprüfung und Fehlerkorrektur

Alle Datensätze wurden auf folgende Topologieregeln überprüft (ESRI, 2007):

- Must not self-overlap
- Must not self-intersect
- Must be singlepart

Diese Kontrolle, welche mit Hilfe der Software ArcGIS erfolgte, ergab ziemlich unterschiedliche Resultate. Tabelle 2 fasst die einzelnen Fehler zusammen.

Tabelle 2 Topologieüberprüfung: Anzahl der Fehler

Daten	Must not self-overlap errors	Must not self-intersect errors	Must be singlepart errors
VECTOR 25	0	0	15
Strassennetz Kanton ZH	0	4	3
Fahrradwege (MAP ZH)	0	4	0
Fahrradwege (Kom. R.pl.)	0	2	5

Je nach Fehler können verschiedene Korrekturmöglichkeiten angewendet werden. Teilweise gibt es auch mehrere Lösungen. Sowohl die implementierten Methoden in ArcGIS als auch die Erweiterung durch den ET Geowizards eignen sich hierbei hervorragend. Abbildung 66 im Anhang A2 illustriert den gewählten Ablauf inklusive gewählter Problemlösungen. Weitere Ausführungen zur Lösung dieser Inkonsistenzen werden unterlassen.

Bei den MAP Zürich Daten tritt zusätzlich das Problem sogenannter *Dangling Nodes* auf (vgl. Glossar). Dies bedeutet, dass ein Strassenende nicht mit einer anderen Strasse verbunden ist, obwohl es in Realität so sein müsste. In Abbildung 2 sind die *Dangling Nodes* durch einen Kreis mit einem Kreuz dargestellt. Wie sich erkennen lässt, müssten die einzelnen Strecken bei einem korrekten Netzwerk miteinander verbunden sein.

Abbildung 2 MAP Zürich: Dangling nodes



Quellen: SWISSIMAGE © 2006 Swisstopo (DV033492.2); Tiefbauamt Stadt Zürich

Bei kleinen Datensätzen wäre ein manuelles Editieren die genaueste Lösung. Da es in diesem speziellen Fall sehr lange dauern würde, wurde diese Aufgabe mit der ArcGIS Erweiterung ET Geowizards gelöst. Hierbei muss allerdings ein Toleranzwert bestimmt werden, wonach bis zu dieser Güte ein *Dangling Node* durch das Verbinden mit dem nächsten Knoten oder mit der nächsten Kante aufgehoben wird. Die Festlegung dieses Schwellwertes erfolgte iterativ und visuell. Als Referenz wurde ein Orthophoto der Stadt Zürich beigezogen. Die besten Resultate erzielte ein Schwellwert von 20 m. Abbildung 3 zeigt diesbezügliche Ergebnisse. Rot dargestellt sind die Strassen im ursprünglichen Zustand, hellgrün die verschobenen bzw. verlängerten Strassen mit 20 m Toleranz und blau mit 25 m Toleranz. Sehr gut ersichtlich ist die Veränderung der Geometrie bzw. Lage der Strasse, falls ein zu grosser Toleranzwert (wie bei 25 m) gewählt wird. Die Wahl eines angemessenen Wertes ist somit sehr wichtig.

Abbildung 3 MAP Zürich: Aufhebung der Dangling nodes



Quellen: SWISSIMAGE © 2006 Swisstopo (DV033492.2); Tiefbauamt Stadt Zürich

Diskussion

Diese kurze Diskussion beschränkt sich auf die Problematik der Dangling Nodes, da die Aufhebung der anderen Fehler einfach und problemlos verlief.

Bei der Aufhebung eines *Dangling Nodes* entstehen Fehler, wenn eine Unterbrechung wirklich einen realen Strassenunterbruch repräsentiert. Die Auflösung eines solchen Knotens geschieht selbstverständlich ungewollt. Des Weiteren können aber auch Situationen auftreten, wo der Toleranzwert zu klein ist und der *Dangling Node* bestehen bleibt. Da dieses Strassenetz in einem späteren Schritt mit anderen Strassen zu einem gesamten Netzwerk integriert wird, sollten diese kleinen Fehler aber nicht weiter gravierend sein. Dies weil durch das spätere *Matching* (vgl. Kapitel 4.1.2) ebenfalls Toleranzen angenommen werden und sich dadurch Fehler aufheben oder auch neue entstehen können. Zudem wird nach dem gesamten Matchingprozess das vollständige Netzwerk noch manuell überprüft und allenfalls korrigiert.

Erwähnenswert ist noch der Umstand, dass die verschiedenen Daten unterschiedliche Perimeter abdecken. Aus diesem Grund wurden alle Daten durch einen Clip mit der Grenze der Stadt

Zürich vereinheitlicht. Dies erfolgte vor der eigentlichen Topologieprüfung. Insofern könnte es sein, dass in den Grunddaten noch Fehler ausserhalb der Stadtgrenze auftreten.

4.1.2 Integration der verschiedenen Strassennetze

Nachdem die Grunddaten soweit überarbeitet wurden, dass die Strassennetze nun topologisch konsistent vorlagen, konnten die verschiedenen Netze einander zugewiesen werden. Der gesamte Prozess, welcher über die Wahl einer geeigneten Methode bis zu einer kurzen Diskussion geht, wird im Folgenden beschrieben.

Methodenwahl

Nach Evaluation der Vor- und Nachteile verschiedener Matching-Methoden fiel der Entscheid auf den von Vivid Solutions Inc. entwickelten Roadmatcher (Vivid Solutions Inc., 2007). Über die Gründe soll hier nicht zu sehr ins Detail gegangen werden. Ein Aspekt war allerdings die Tatsache, dass das Ziel dieser Arbeit nicht die Entwicklung eines eigenen *Matching* Algorithmus ist. Deshalb wurde nach einer Methode gesucht, die mit vernünftigem Aufwand erlernt und implementiert werden konnte. Trotzdem wurde versucht eine möglichst genaue und qualitativ hochstehende Lösung zu finden. Wie sich schnell herausstellte, sind die meisten Algorithmen nur für eine spezifische Situation entwickelt und nicht einfach auf andere Probleme übertragbar. Falls dafür mehr Zeit investiert werden könnte, gäbe es möglicherweise einige Algorithmen, die auf das Problem in dieser Arbeit angepasst und erweitert werden könnten.

Als Beispiel für einen erweiterbaren Algorithmus sei der von Lüscher und Burghardt (2006) und in Kapitel 2.3 kurz vorgestellte Algorithmus genannt. Gemäss mündlicher Absprache mit Herrn Lüscher (Gespräch am 18.11.2007) ist die von ihm entwickelte Methode für die Probleme, welche sich in dieser Arbeit stellen, aber in der jetzigen Form ungeeignet.

Ein weiterer Algorithmus, der für diese Aufgabe möglicherweise erweiterbar wäre, ist derjenige von Waldner (2005). Der Algorithmus wurde jedoch für eine frühere Programmversion von ArcGIS erstellt und ist nicht ohne Aufwand auf die Version ArcGIS 9.2 übertragbar. Zusätzlich müsste die Konfigurations-Datenbank und evtl. auch der Algorithmus selbst für vier verschiedene Datensätze erweitert und verbessert werden. Diese Aufgabe erwies sich als zu zeitraubend. Zu guter Letzt fehlt der Nachweis, dass wirklich gute *Matching* Resultate erzielt würden.

Methodik

Zitierbare Referenzen über die genaue Funktionsweise der im Roadmatcher verwendeten Algorithmen wurden leider keine gefunden. Somit muss hier, anders als beim Matching der GPS-Daten, auf eine genaue Erklärung des Algorithmus verzichtet werden. Jedoch soll stellvertretend das Programm aus einer Benutzersicht kurz vorgestellt werden, weil es nicht so weit verbreitet ist.

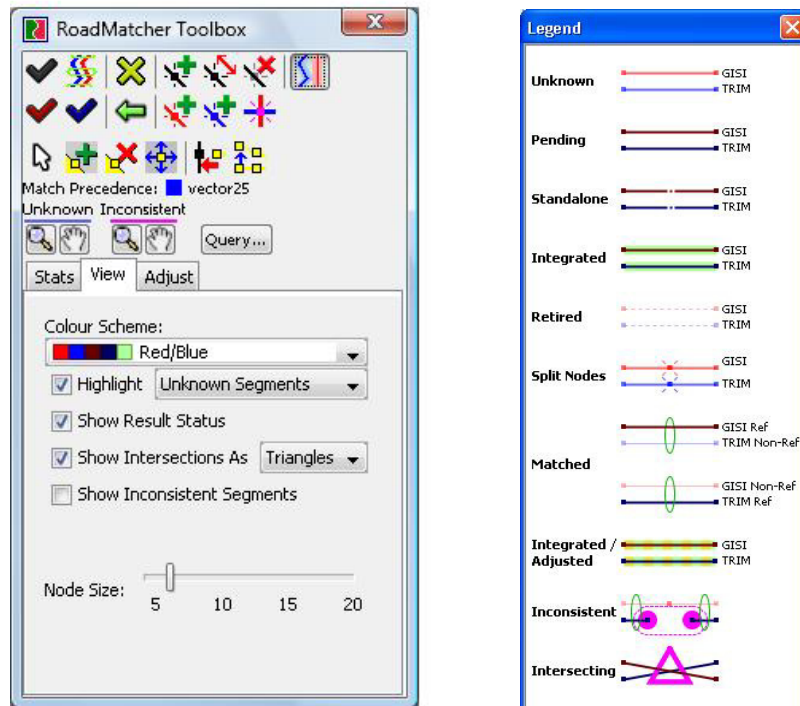
Der Roadmatcher wurde speziell für die Integration von linearen Strassennetzen entwickelt und bietet zahlreiche nützliche Werkzeuge:

- Prioritätenangabe um den Referenzdatensatz festzulegen
- Automatisches und manuelles Matching von Strassensegmenten
- Automatische Detektion topologischer Inkonsistenzen zwischen integrierten Strassensegmenten
- Unterstützung des *many-to-many* Matchings durch Hinzufügen von Knoten
- Ein grosses Set von manuellen Werkzeugen für die Nachbearbeitung
- Ein ausgiebiges visuelles Interface für die Darstellung des *Matching*-Prozesses
- Unterstützung verschiedener In- und Output Formate (Jump GML, GML 2.0, FME GML, WKT, ESRI Shapefile)
- Möglichkeit alle *Matching*-Entscheide oder Veränderungen rückgängig zu machen

Besonders ausgeprägt sind die manuellen Editiertools, welche eine gute Nachbearbeitung ermöglichen (vgl. Abb. 4).

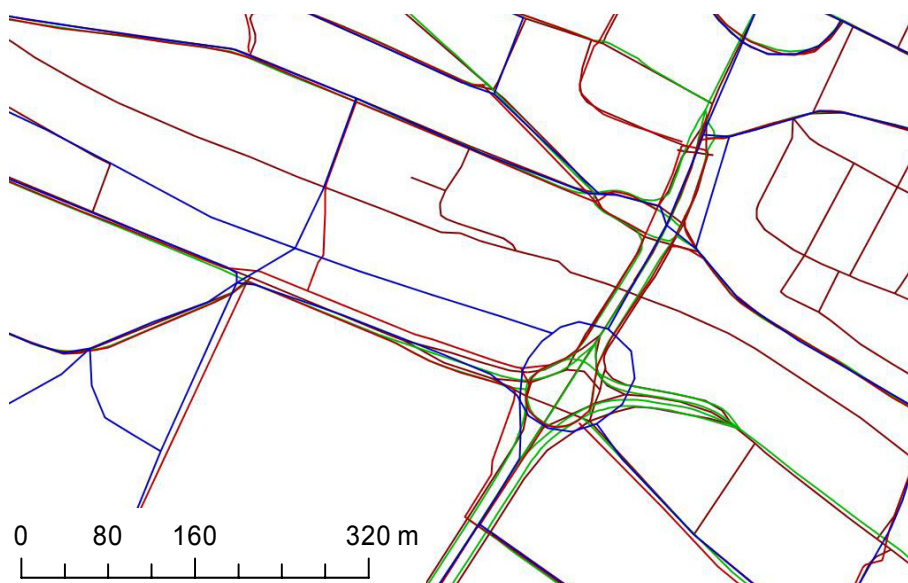
Auch bei dieser Methode treten einige Schwierigkeiten auf, die in einem späteren Abschnitt diskutiert werden. Weitere allgemeine Informationen zur Java Plattform JUMP sowie zum Roadmatcher finden sich im Glossar.

Abbildung 4 Roadmatcher Werkzeuge und Legende



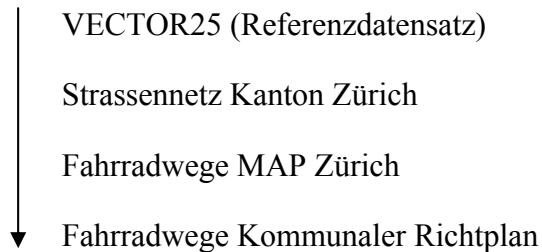
Die Ausgangslage für das Matching präsentierte sich wie in Abbildung 5 dargestellt.

Abbildung 5 Ausgangslage vor dem Matching



Quellen: VECTOR25 © 2006 Swisstopo (DV033492.2); ARV Kanton Zürich; Tiefbauamt Stadt Zürich

Aufgrund der in Kapitel 3 beschriebenen Genauigkeit der Strassennetze wurden die Daten gemäss folgender Reihenfolge gematched:



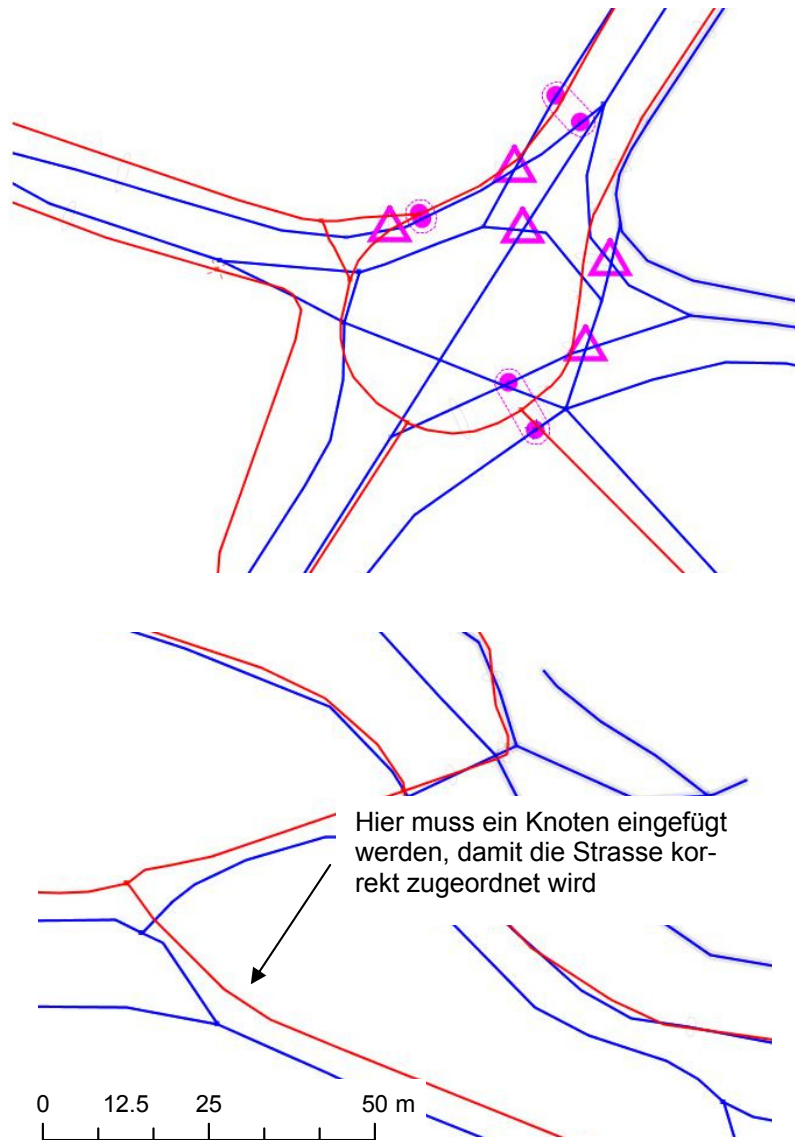
Als zusätzliche Hilfe und zugunsten der besseren Übersicht illustriert das Flussdiagramm im Anhang A2 den gesamten *Matching* Ablauf graphisch (vgl. Abb. 67).

Für das automatische *Matching* sind eine Reihe von Optionen und Toleranzen festzulegen (vgl. Tabelle 3). Nach dem Testen verschiedener Werte wurden eher konservative Toleranzen gewählt. Das heisst, dass die Toleranzen ziemlich klein gehalten wurden, damit Fehler, welche beim automatischen Matchingprozess entstehen können, möglichst vermieden werden. Dementsprechend wird die anschliessende manuelle Überarbeitung etwas aufwändiger. Dafür leiden die Qualität des Netzwerkes einerseits und diejenige des Schlussmodells andererseits weniger. Eine manuelle Nachbearbeitung war besonders an Orten notwendig, wo das Strassennetz dicht und komplex ist (vgl. Abb. 6). Zusätzlich erschwerend war die unterschiedliche Netzkodierung.

Tabelle 3 Automatisches Matching: Optionen und Toleranzen

Optionen und Toleranzen	Toleranzwerte
<i>Minimum Line Segment Length</i>	1 m
<i>Nearness tolerance</i>	10 m
<i>Find standalone Roads: Minimum distance</i>	10 m
<i>Find matched Roads: Minimum distance</i>	10 m
<i>Maximum Adjustment Size</i>	10000 m
<i>Maximum Segment Angle Delta</i>	20°
<i>Minimum Incident Angle</i>	20°
<i>Segment Adjustment Length</i>	20°
<i>Adjustment Method</i>	Warp entire segment

Abbildung 6 Manuelle Nachbearbeitung



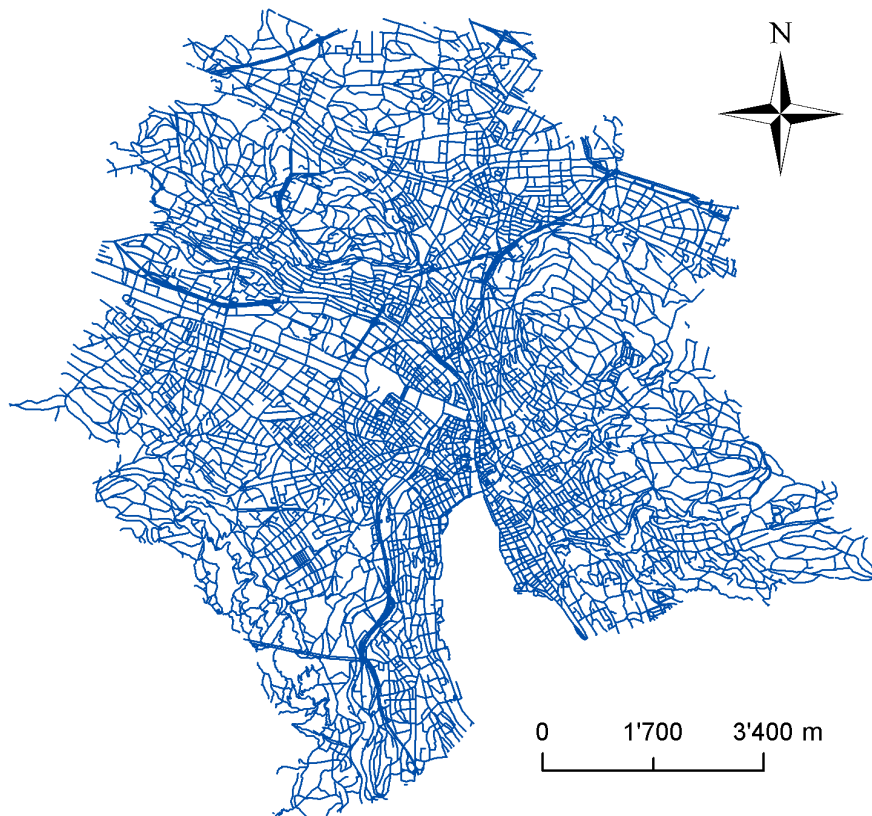
Quellen: VECTOR25 © 2006 Swisstopo (DV033492.2); ARV Kanton Zürich; Tiefbauamt Stadt Zürich

Ergebnisse

Für die drei automatischen Schritte benötigte das Programm mit einem gewöhnlichen PC (Pentium 4, 1.8 GHz, 1 GB RAM) jeweils ca. 4 bis 5 Minuten. Automatisch wurden rund 88% aller Strassensegmente zugeordnet, die übrig gebliebenen Strecken wurden manuell zusammengespielt.

Das am Ende des Integrationsprozesses resultierende Strassennetz (vgl. Abb. 7) enthält alle räumlichen und nicht-räumlichen Informationen der vier Grunddatensätze. Dort wo sich die einzelnen Netze nach dem *Matching* überlagern, sind also mehr Attribute vorhanden, als an Orten wo eine Strasse nur durch eine Quelle repräsentiert wird. Beim Roadmatcher kann angegeben werden welche Attribute für das integrierte und separat erzeugte Strassennetz übernommen werden sollen. Damit keine Information verloren geht, wurden vorerst alle Attribute übertragen. Bei der späteren Verwendung des Netzwerkes können dann alle unnötigen Informationen je nach Bedürfnis eliminiert werden.

Abbildung 7 Integriertes Strassennetzwerk



Quellen: VECTOR25 © 2006 Swisstopo (DV033492.2); ARV Kanton Zürich; Tiefbauamt Stadt Zürich

Diskussion

Nach dem automatischen *Matching* erscheint die Anzahl übrig gebliebener, nicht automatisch integrierter Strassen und Inkonsistenzen ziemlich hoch. Hierbei sollte aber beachtet werden,

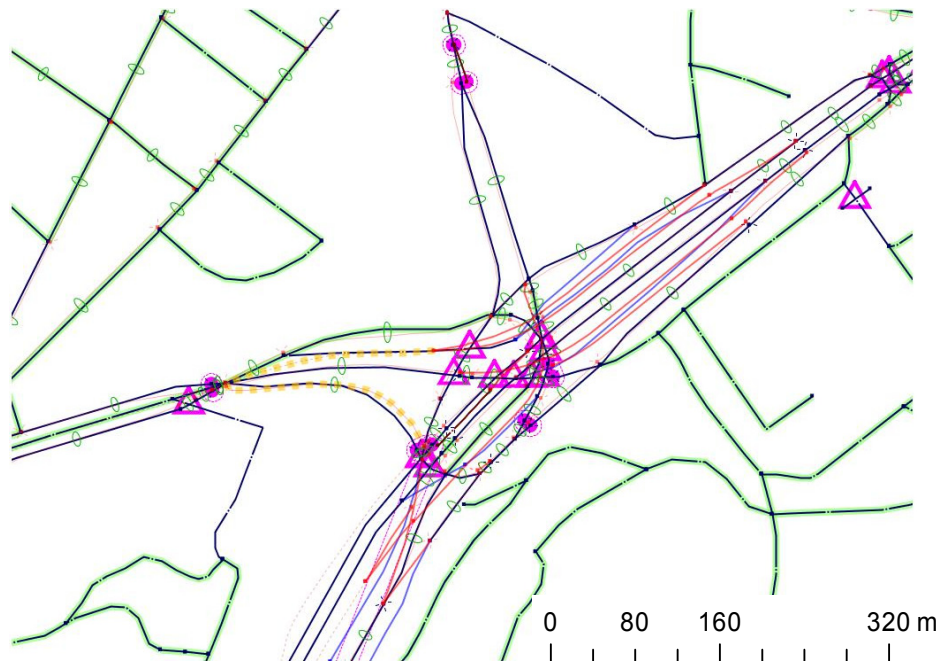
dass die Auflösung einer Inkonsistenz teilweise zwangsläufig andere löst, weshalb sich die Zahl der Überarbeitungsfälle relativ schnell reduziert. Die manuelle Überarbeitung ist ziemlich subjektiv und stark abhängig vom Überarbeiter. Je nach Qualitätsbewusstsein und Ansprüchen dauert dies etwas länger oder kürzer.

Die verschiedenen Qualitätsunterschiede der Daten erschweren das automatische *Matching*. Für die genau gleiche reale Situation kann es sein, dass in einem Datensatz mehr Strassen vorhanden sind als im anderen. Des Weiteren ist die Netzkodierung ziemlich heterogen. Das Fehlen eines (*Vertex* oder) Knotens kann dazu führen, dass zwei Strassen, auch wenn sie sehr nahe beisammen liegen, nicht automatisch zugeordnet werden. Diese Fälle beeinträchtigen die Ausgangslage für das *Matching* und erhöhen den manuellen Überarbeitungsaufwand.

Das nach dem *Matching* resultierende Netz wurde visuell am Computerbildschirm durch Einblenden aller Grunddaten inklusive deren Attribute geprüft. Des Weiteren wurde die Plausibilität aufgrund der Orthophotos von Zürich getestet. Obwohl das Ergebnis für die geplante Anwendung als gut empfunden wurde, sollten mögliche Fehler nicht vergessen gehen. An Orten wo verschiedene Strassen nebeneinander und übereinander verlaufen (z.B. im Bereich der Hardbrücke) gibt es einige Unsicherheiten (vgl. Abb. 8). Solche Unsicherheiten können überall dort auftreten, wo das Netzwerk komplex und dicht ist. Das dadurch bedingte „Rauschen“ sollte aber bei der Modellschätzung am Schluss nicht von Bedeutung sein.

Inwiefern sich dieses Netzwerk für andere Anwendungen verwenden lässt, sei dahingestellt. Es müsste auf jeden Fall genau geprüft werden welche Anforderungen die geplante Verwendung stellt. Überdies sollte nicht vergessen gehen, dass das hier erzeugte Netzwerk für eine Modellerzeugung, die auf einer gewissen Wahrscheinlichkeit beruht, gedacht ist. Somit fallen Fehler in dieser Arbeit möglicherweise weniger stark ins Gewicht als bei einer anderen Anwendung.

Abbildung 8 Matching Probleme und mögliche Fehler



Quellen: VECTOR25 © 2006 Swisstopo (DV033492.2); ARV Kanton Zürich; Tiefbauamt Stadt Zürich

4.2 GPS-Daten

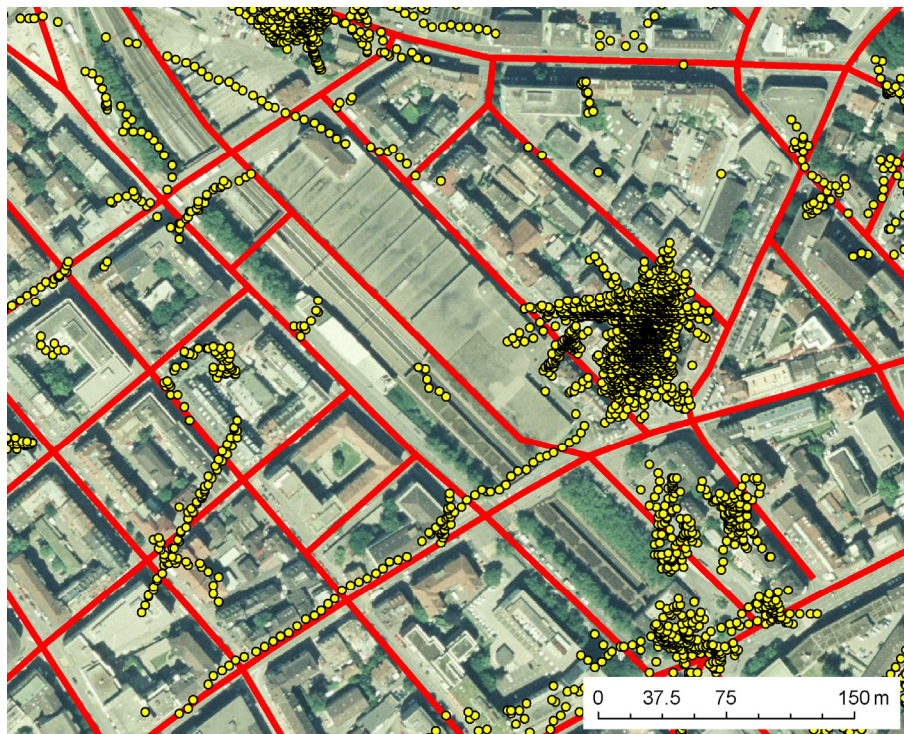
In diesem Abschnitt wird die Vorverarbeitung der GPS-Daten vorgestellt. Analog zur Aufbereitung der Strassendaten wurden auch diese Daten vor dem *Matching* auf das Strassennetz einer Qualitätskontrolle unterzogen. Um die Fehler möglichst vollständig zu eruieren und eliminieren, wurden verschiedene Filterungsschritte auf die Daten angewendet. Dabei wird zwischen einer allgemeinen und einer speziellen Filterung für Fahrradfahrer unterschieden. Letztere soll besonders Fehler, die bei der Identifikation der Verkehrsmittel (hier Fahrradfahrer) entstanden sind, untersuchen und entfernen. Solche Fehler sind dadurch bedingt, dass zum Zeitpunkt der Arbeit der Detektionsalgorithmus von Schüssler und Axhausen (2008) noch nicht vollends ausgearbeitet war.

Die einzelnen Arbeitsschritte sind teilweise sehr aufwändig und werden in der Praxis häufig unterschätzt. Auf die verschiedenen auftretenden Probleme und die gleichzeitige zeitraubende Datenvorverarbeitung wird beispielsweise auch in der Arbeit von Harvey und Krizek (2007) hingewiesen.

4.2.1 Evaluation der GPS-Daten

Die GPS-Daten lagen als Tab-separierte Textdatei vor und wurden in einem ersten Schritt zu einem ESRI Shapefile konvertiert. Nach dieser Umwandlung konnten die Daten ein erstes Mal räumlich visualisiert und evaluiert werden (vgl. Abb. 9). Hierzu wurden die GPS-Punkte sowie die schon oft benutzte Orthophoto mit 50 cm Pixelauflösung, welche als visuelle Referenz im Hintergrund diente, in das Programm ArcGIS eingelesen. Als zusätzliche thematische Ebene wurden die Strassendaten hinzugefügt.

Abbildung 9 Evaluation der GPS-Daten



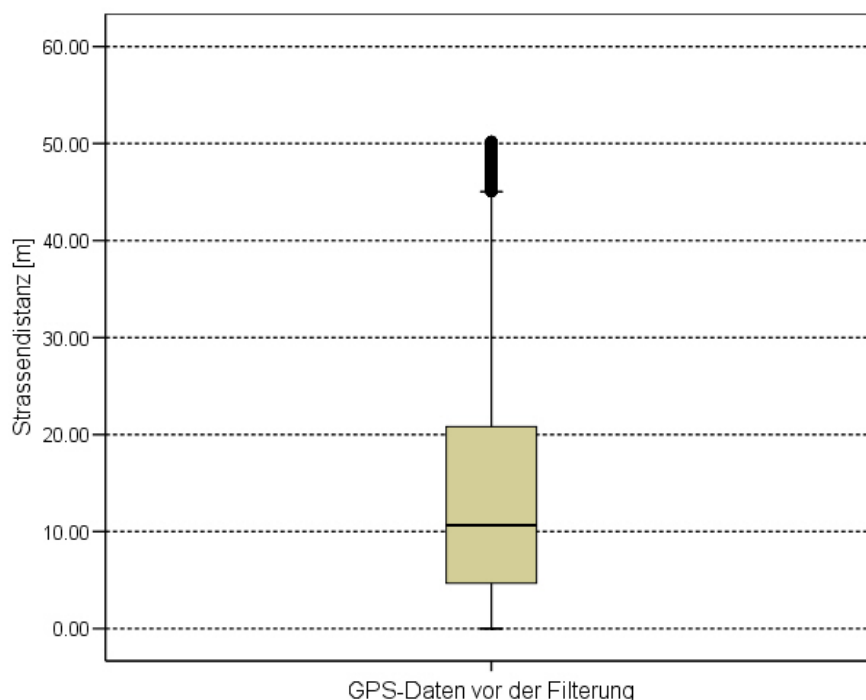
Quellen: SWISSIMAGE, VECTOR25 © 2006 Swisstopo (DV033492.2)

Wie erwartet, ist die Genauigkeit der Daten ziemlich heterogen. An gewissen Orten erscheinen die Messungen plausibel, an anderen Stellen treten hingegen merkwürdige Punktverteilungen oder grosse Abweichungen vom Strassennetz auf. Dies kann selbstverständlich verschiedene Gründe haben, welche bereits in Kapitel 2.1.1 detailliert erwähnt wurden. An dieser Stelle seien die zwei wesentlichsten Fehlerquellen nochmals genannt. Punktwolken können einerseits durch Signalreflexionen in eng bebauten Gebieten oder andererseits durch einen Aufenthalt in einem Gebäude verursacht worden sein. Letzteres kann auftreten, wenn eine Person beispielsweise vergessen hat das Gerät auszuschalten. Die zweite grössere Abwei-

chung, nämlich dass Messungen in grosser Entfernung vom Strassennetz auftreten, hat ebenfalls unterschiedliche Gründe. Es kann durchaus vorkommen, dass eine Strasse im Netzwerk fehlt oder dass ein Fahrradfahrer sich auf offenem Gelände bewegte. Gerade der letzte Punkt ist nicht zu vernachlässigen.

Um sich einen Überblick über das Ausmass der Abweichung vom Strassennetz zu verschaffen, wurde der Abstand der Punkte zu den Strassen berechnet und in einem Boxplot dargestellt (vgl. Abb. 10). Ob ein Punkt links oder rechts von einer Strasse liegt, wurde nicht berücksichtigt.

Abbildung 10 Abstand der GPS-Punkte zum Strassennetz



Der grösste Teil der Punkte, nämlich 50 %, liegt in einem Abstand zwischen 4 und 21 m (unteres und oberes Quartil) von der Strasse entfernt. Weiter ist ersichtlich, dass der Median bei ungefähr 10-11 m liegt. Dieser vermittelt zusätzlich einen Eindruck über die Schiefe der Verteilung. Die Länge der hellbraunen Box lässt auf den Interquartilsabstand schliessen, was ein Mass für die Streuung der Daten ist. Ferner markieren die durchgezogenen horizontalen Linien (*Whisker*) das 1.5-fache des Interquartilsabstands. Diese liegen bei 0 bzw. 45 m. Darüber hinaus folgen die Ausreisser, welche bis ca. 51 m reichen. Auf die Entstehung derselben wurde bereits hingewiesen.

Diskussion

Dias *et al.* (2007) berechneten ebenfalls den Abstand der GPS-Punkte zum Strassennetz um einerseits Fehler der GPS-Daten und andererseits Ungenauigkeiten bei der Digitalisierung des Strassennetzes zu untersuchen. Hierbei haben sie die Position bzw. Strassenseite der Punkte mitberücksichtigt.

Da in der vorliegenden Arbeit Fahrradfahrer untersucht werden, welche sich üblicherweise nicht sehr Verkehrsregel konform verhalten, wurde der Abstand der Punkte zum Strassennetz unabhängig von der Strassenseite berechnet. Des Weiteren ist der von Fahrradfahrern eingenommene Raum weniger eingeschränkt als zum Beispiel bei der Studie von Dias *et al.* (2007), in welcher Fussgänger auf einem schmalen Weg untersucht wurden. Ein Fahrradfahrer in einem städtischen Gebiet kann grössere Bewegungsradien beanspruchen als ein Spaziergänger auf einem engen Parkweg.

Analog zur Untersuchung von Dias *et al.* (2007) gilt es auch hier zu beachten, dass die Strassendaten mit Ungenauigkeiten behaftet sind. Bei der Digitalisierung der einzelnen Strassen wurde nicht immer die Strassenachse als Referenz gewählt und auch die Netzkodierung ist nicht einheitlich. Diese Tatsachen stellen einen weiteren Grund für die Ignorierung der jeweiligen Strassenseiten dar.

4.2.2 Allgemeine Datenfilterung

Die allgemeine Datenfilterung besteht aus drei wesentlichen Schritten. Alle Methoden wurden mit Hilfe der Software ArcGIS ausgeführt. Eine manuelle Durchführung dieser Berechnungsschritte erwies sich als beinahe unmöglich, weshalb ein Python-Skript deren Automatisierung übernimmt. Dieses läuft in Kombination mit ArcGIS und ist im Anhang A7 zu finden. Als weitere Stütze illustriert Abbildung 68 (Anhang A2) den Ablauf graphisch.

Methodik

Der erste Schritt beinhaltet das Eliminieren von GPS-Punkten ausserhalb der Stadt Zürich. Dies erfolgte durch einen Clip mit der Stadtgrenze Zürichs. Notwendig ist dies, da das Strassennetz nur innerhalb der Stadt vollständig vorhanden ist.

Die zweite Filterung wurde mit Hilfe eines Buffers um die Strassen bewerkstelligt (vgl. Abb. 11). Dabei wurde die Grösse des Buffers dem jeweiligen Strassentyp, soweit diese Informati-

on vorhanden ist, und der Genauigkeit des GPS-Gerätes angepasst. Als maximal gültige Abweichung des GPS-Gerätes wurden 10 m auf beiden Strassenseiten angenommen (vgl. Kapitel 3.1). Die Rücksichtnahme auf den Strassentyp dient dazu die Breite der Strasse und die daraus folgenden unterschiedlichen Abweichungen von der Strassenachse zu berücksichtigen. Für Strassen ohne Breitenangaben, was häufig schmale Feld-, Velo- oder Fusswege betrifft, wurde lediglich die GPS-Ungenauigkeit beachtet.

Abbildung 11 Bildung des Strassenbuffers

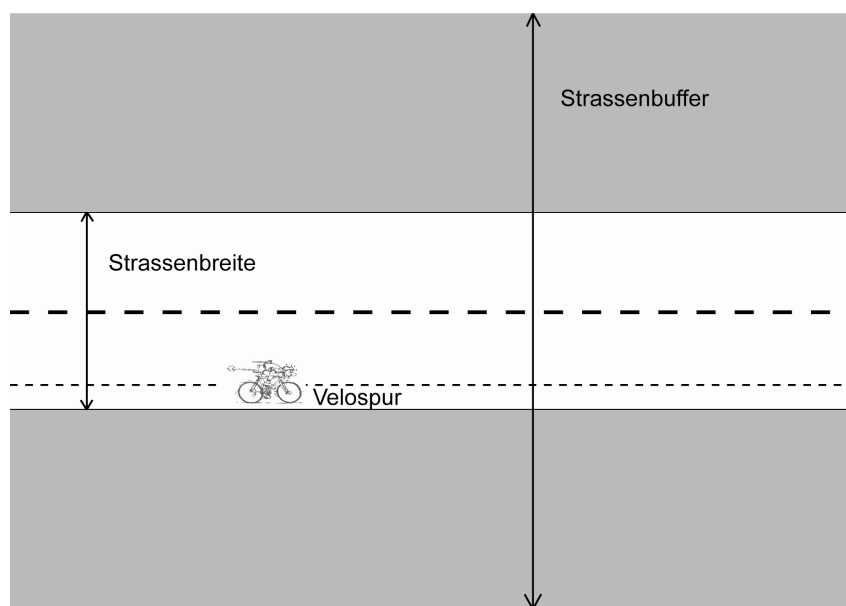


Tabelle 4 vermittelt einen Überblick über die berücksichtigten Strassentypen mit jeweiligen Breitenangaben. Ergänzend dazu ist die gewählte Buffergrösse aufgeführt, welche gemäss obiger Beschreibung wie folgt berechnet wurde:

$$\text{Buffergrösse} = \text{Mindestbreite} + 2 \cdot \text{Maximale Abweichung des GPS - Geräts} \quad (4)$$

Tabelle 4 Strassentypen und Bestimmung der Buffergrösse

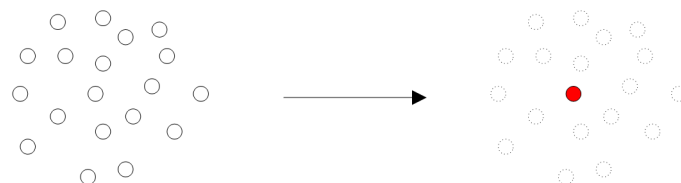
Strassentyp	Mindestbreite ⁸	Buffergrösse
1. Klass Strasse	6 m	26 m
2. Klass Strasse	4 m	24 m
3. Klass Strasse	2.8 m	23 m
4. Klass Strasse	1.8 m	22 m
Quartierstrasse	4 m	24 m
Alle anderen Strassen	Keine Angaben	20 m

Obschon nach den beiden ersten Schritten zahlreiche ungewöhnliche Punktansammlungen entfernt werden, soll der letzte Teil der allgemeinen Datenbereinigung diese Problematik zusätzlich entschärfen. Der dritte Filterungsschritt hat somit zum Ziel sogenannte Punktcluster (vgl. Abb. 9) zu lokalisieren und eliminieren. Zhou and Golledge (2006) bezeichnen die auftretenden Punktcluster auch als systematisches Rauschen. Dies sei die grösste Ursache für spätere Fehler beim Map Matching.

Zu diesem Zweck wird anfangs die Dichte der Punkte (*Kernel Density Estimation*) berechnet. Als Suchradius r wurden 20 resp. 30 m gewählt, da dies einerseits enge Strassenschluchten und andererseits normale Bewegungsradien in Gebäuden berücksichtigt. Um möglichst genaue Resultate zu bekommen, wurde für das Dichteraster 1 m als Zellgrösse gewählt.

Nach der Dichteberechnung werden Punktwolken, die einen bestimmten Schwellwert überschreiten als zu dicht klassiert und durch den jeweiligen Schwerpunkt ersetzt (vgl. Abb. 12).

Abbildung 12 Berechnung des Schwerpunktes

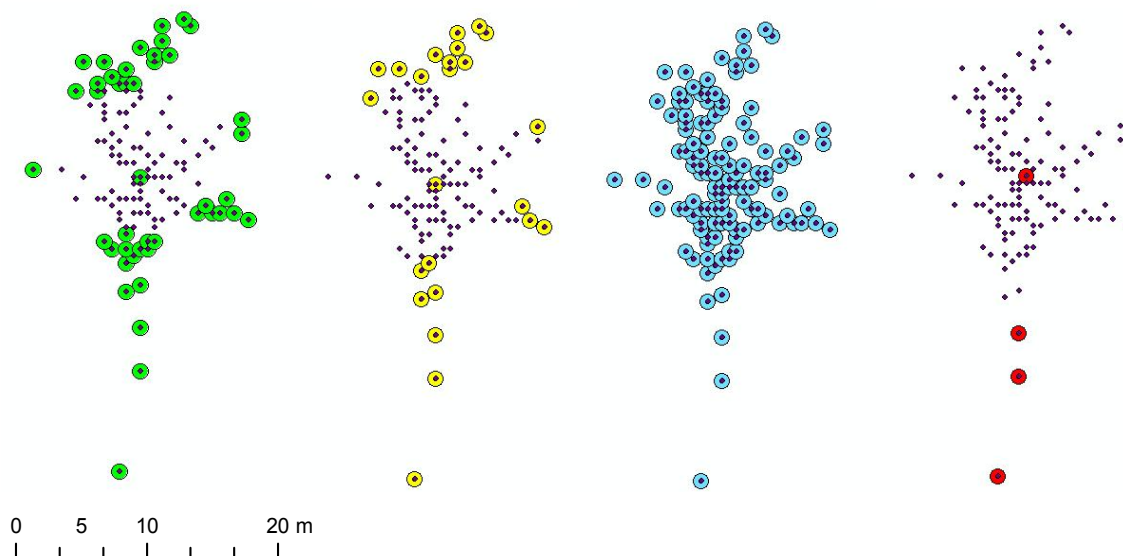


⁸ Mindestbreiten Festlegung: Swisstopo (2008d)

Die Festlegung des Dichte-Schwellwertes ρ ist nicht ganz einfach, weshalb zahlreiche Berechnungen ausgeführt und visuell begutachtet wurden. Schliesslich wurden vier verschiedene Varianten miteinander verglichen. Abbildung 13 zeigt die Ergebnisse für die Filterung mit

- $r = 20$ m und $\rho = 0.09$ (grün)
- $r = 20$ m und $\rho = 0.08$ (gelb)
- $r = 30$ m und $\rho = 0.05$ (blau)
- $r = 30$ m und $\rho = 0.04$ (rot)

Abbildung 13 Dichtefilterung für einen Ausschnitt aus Etappe 1



Die ursprünglichen Punkte sind der Übersicht halber etwas feiner dargestellt. Schnell ersichtlich ist, dass die dritte Variante a priori weggelassen werden kann, da der Schwellwert offensichtlich zu gross ist. Nach der vierten Filterungsversion fallen hingegen zu viele Punkte weg. Daten, welche zum Beispiel ganz oben liegen, sind keine mehr vorhanden. Falls dies der Anfang oder Ende einer Etappe wäre, würde dies zu einer ungewollten Routenverkürzung führen. Die ersten beiden Filterungsvarianten sind sehr ähnlich. Methode 1 scheint die etwas bessere zu sein, da geringfügig weniger Punkte eliminiert werden sowie die Randbereiche bestehen bleiben. Nach Variante 2 wäre beispielsweise der Punkt ganz links nicht mehr berücksichtigt. Da man nicht mit Sicherheit weiss, welche Punkte nun korrekt sind, ist es besser den ganzen Bewegungsradius zu beachten. Dies verhindert, dass Routen möglicherweise unterbrochen werden, da einzelne Punkte zu weit von einer Strassenkante entfernt liegen.

Nicht zuletzt ist es wichtig zu erwähnen, dass eine Vermischung bzw. Überschneidung der Routen der verschiedenen Personen zu grossen Dichten führen kann. Die daraus entstehenden Punktanhäufungen müssen aber keineswegs fehlerhaft sein. Um dieser Tatsache Rechnung zu tragen, wird ein eindeutiger Schlüssel (PTS_ID) gebildet, der sich aus der ID der Person, des Weges und der Etappe zusammensetzt. Die Filterung wurde demzufolge auf Basis dieses Unterscheidungsschlüssels ausgeführt. Dies ist im Python-Skript durch verschiedene Schleifen gelöst.

Ergebnisse

Tabelle 5 vermittelt einen Überblick über die verbleibenden Daten nach den jeweiligen Filterungsschritten. Wie erwartet, dezimiert der erste Clip den ursprünglichen Datensatz am meisten. Ersichtlich ist aber auch, dass nach dem Ausschluss durch den Strassenbuffer nur noch halb so viele Punkte übrig bleiben. Diese Schwellwerte wurden deshalb grundlegend hinterfragt und überprüft.

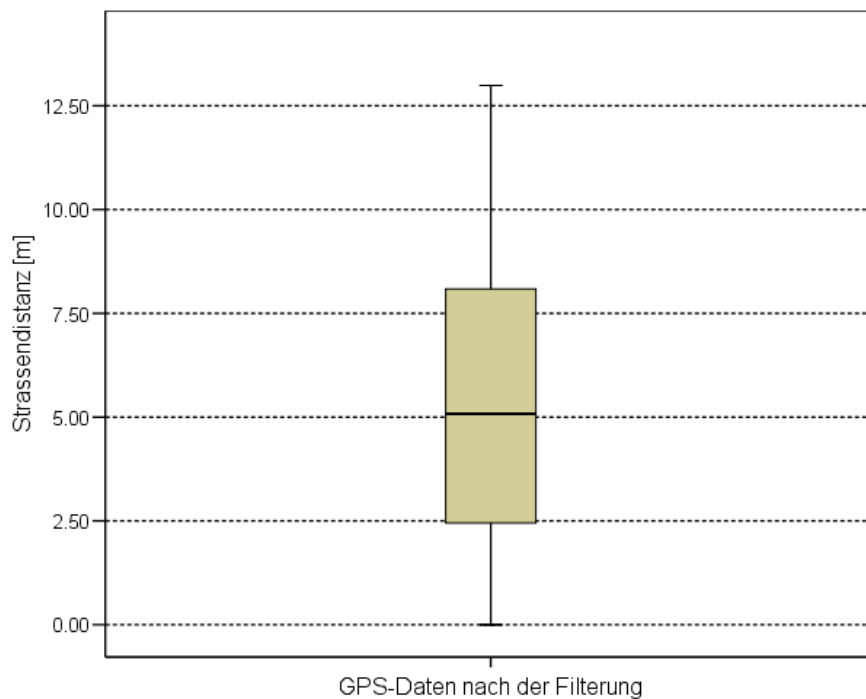
Die Filterung aufgrund der Dichte eliminiert im Gegensatz dazu nur sehr wenige Punkte. Nichtsdestotrotz ist dieser letzte Schritt für das nachfolgende Matching wichtig. Als optimale Lösung wurde die Dichtefilterung mit einem Suchradius von $r = 20$ m und dem Schwellwert $\rho = 0.09$ erachtet und weiterverwendet. Je nach Anwendung und Präferenzen können die Schwellwerte selbstverständlich so angepasst werden, dass mehr oder weniger Daten eliminiert werden.

Nachdem in Kapitel 4.2.1 ein Boxplot der ursprünglichen GPS-Daten vorgestellt wurde, soll dasselbe mit den gefilterten Daten erfolgen (vgl. Abb. 14). Sehr gut sichtbar ist, dass der Median nun bei rund 5 m liegt. 50 % der Punkte weisen einen Abstand von 2.5 bis 8 m von der Strasse auf. Des Weiteren ist die Verteilung symmetrischer geworden. Grössere Abstände als 13 m kommen aufgrund der hier festgesetzten Buffergrösse logischerweise nicht mehr vor.

Tabelle 5 Ergebnisse nach der Filterung des GPS-Datensatzes

Filterungsschritt	Anzahl Punkte	Anzahl Etappen	Übrig gebliebene Punkte/Etappen in %
Vor der Filterung	1169546	52423	100/100
Clip mit der Stadtgrenze ZH	265241	12688	22.68/24.20
Clip mit dem Strassenbuffer	136407	9365	11.67/17.86
Dichtefilterung ($r = 20$ m; $\rho = 0.09$)	134235	9336	11.48/17.81
Dichtefilterung ($r = 20$ m; $\rho = 0.08$)	133539	9336	11.42/17.81
Dichtefilterung ($r = 30$ m; $\rho = 0.05$)	135278	9336	11.57/17.81
Dichtefilterung ($r = 30$ m; $\rho = 0.04$)	133330	9336	11.40/17.81

Abbildung 14 Abstand der gefilterten GPS-Punkte zum Strassennetz



Diskussion

Bei einigen Strassen, namentlich bei schmalen Feld-, Velo- oder Fusswegen wurde lediglich die GPS-Abweichung für die Buffergrösse berücksichtigt. Dies erscheint nicht weiter gravierend, da diese Strassen nicht sonderlich breit sind und die maximale GPS-Abweichung ohnehin ziemlich grosszügig gewählt wurde. Zusätzlich kommt hinzu, dass die gesamte Strassenbreite von einem Zweiradfahrer niemals ausgenützt wird. Letzterer Einwand war auch ein Grund weshalb lediglich die Mindestbreite der Strassen und nicht beispielsweise Durchschnittsbreiten bei der Bufferbildung berücksichtigt wurden. Der Clip der Daten durch einen Strassenbuffer hat auch noch einen anderen Vorzug. Es wurden Punkte eliminiert, welche verstreut im Raum liegen ohne eine Beziehung zum Strassennetz zu haben. Somit wird die Problematik vermindert, wonach sich Fahrradfahrer häufig nicht an Verkehrsregeln halten oder auf offenem Gelände bewegen.

In der Literatur finden sich einige Studien, die sich mit der Problematik der GPS-Datenfilterung auseinandergesetzt haben. Dias *et al.* (2007) zum Beispiel versuchten ebenfalls GPS-Punkte, welche einem Weg zugeordnet werden sollten, zu filtern. Sie liessen dabei Daten weg, welche eine Distanz von zweimal der theoretischen Genauigkeit des GPS-Gerätes aufweisen. Dieser Wert wurde durch eine visuelle Validierung der GPS-Tracks festgelegt und entsprach analog dieser Anwendung 10 m. Da die GPS-Datenerhebung entlang eines schmalen Weges erfolgte, wurde die Breite der Strasse nicht berücksichtigt. In der hier vorgestellten Anwendung macht dies aber Sinn, da die verschiedenen Strassen einerseits ganz unterschiedliche Breiten aufweisen und andererseits ein sehr komplexes und dichtes Netzwerk mit vielen Kreuzungen bilden.

Einen ähnlichen Ansatz für die Reduktion von Punktclustern in GPS-Daten haben Zhou and Golledge (2006) angewandt. Wie in dem hier vorgestellten Verfahren wurden zu dicht beisammen liegende Punkte durch den jeweiligen Schwerpunkt ersetzt. Die Identifizierung der Punktcluster erfolgte hingegen mit einer anderen Methode. Für diesen ersten Schritt wurde der DBSCAN (Ester *et al.*, 1997) Clustering Algorithmus verwendet. Die hier vorgestellte Variante erscheint ebenso angemessen, da der Schwellwert und der Suchradius sehr schnell und flexibel veränderbar sind. Des Weiteren ist die Rechengeschwindigkeit für den verwendeten Datensatz auch akzeptabel. Selbstverständlich ist die angesprochene Rechenzeit aber von verschiedenen Faktoren, wie der Anzahl Punkte ausserhalb der Stadtgrenze abhängig. Diese Daten werden als Erstes eliminiert und machen die nachfolgenden Berechnungen dementsprechend schneller. Auch sonstige auf dem Computer laufende Anwendungen wirken sich

auf die Rechenzeit aus. Weitere Ratschläge um die Leistung von ArcGIS Desktop zu verbessern finden sich in ESRI (2008).

Alternative Möglichkeiten zur allgemeinen Filterung

Bevor zur speziellen Datenbereinigung für Fahrradfahrer übergegangen wird, sollen in diesem Abschnitt kurz zwei alternative Filterungsvarianten angedeutet werden.

Eine Alternative zur eben beschriebenen Filterungsmethode besteht darin, die GPS-Punkte der einzelnen Etappen zu einer Polylinie zu verbinden. Analog zum vorherigen Filterungsschritt werden auch hier die aufgezeichneten Etappen anhand des eindeutigen Unterscheidungskriteriums PTS_ID getrennt. Punkte mit gleicher PTS_ID werden abhängig von der Aufzeichnungszeit zu einer Polylinie verbunden. Die entstandenen Polylinien lassen sich dann auf Selbstverschneidungen und -überlappungen untersuchen. Hierzu ist die Topologie-Überprüfung von ArcGIS sehr hilfreich. Fehler können einfach detektiert und je nach Situation entfernt werden. Es gilt jedoch zu beachten, dass nur mehrfache Überschneidungen eliminiert werden sollten. Diese deuten auf einen Aufenthalt in einem Gebäude oder ungewollte Signalreflexionen in engen Strassenschluchten hin.

Diese Methode wurde deshalb nicht angewendet, weil Selbstverschneidungen nicht in jedem Fall fehlerhaft sein müssen. Besonders bei Fahrradfahrern kann dies häufig auftreten und bei einer Automatisierung möglicherweise zu einer ungewollten Punkteliminierung führen. Die hier empfehlenswerte Beurteilung jedes Einzelfalls wäre viel zu aufwändig.

Die zweite mögliche Alternative bestünde in der Generalisierung der aus den GPS-Etappen erzeugten Polylinien (z.B. mit der Generalisierungsfunktion von ArcGIS). Dies wurde beispielsweise in Harvey und Krizek (2007) gemacht. Ähnlich wie bei der obigen Variante besteht aber auch bei dieser Möglichkeit die Gefahr einer unkontrollierten Eliminierung von Punkten.

4.2.3 Spezielle Filterung für Fahrradfahrer

Nach der eben beschriebenen allgemeinen Bereinigung wurden die Daten auf mögliche Fehler, die bei der Identifikation der Fahrradfahrer entstanden sind, untersucht und weiter gefiltert. Auch diese Schritte, welche wiederum mittels verschiedener Funktionen der Software ArcGIS ausgeführt wurden, sind im Anhang A2 graphisch dargestellt (Abb. 69).

Methodik

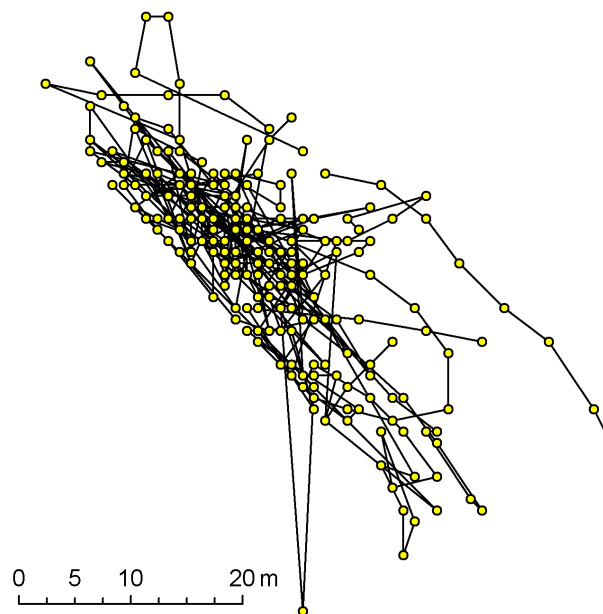
Da genug GPS-Punkte zur Verfügung standen und wie schon mehrfach erwähnt der Identifikationsalgorithmus zum Zeitpunkt der Arbeit noch fehlerhaft war, wurden als erste Massnahme alle Punkte entfernt, welche eine Wahrscheinlichkeit für Fahrradfahrer $p_{\text{Bike}} < 0.5$ aufweisen.

Als nächster Schritt wurden die Höhe, die Steigung und das Gefälle der GPS-Punkte bzw. Etappen untersucht. Aufgrund ziemlich unrealistischer Höhen war es offensichtlich, dass einige Z-Werte fehlerhaft sein müssen. Etwa 100 Punkte wiesen Höhen um die 200 m ü. M. auf, was deutlich unter dem tiefsten Punkt von Zürich liegt (392 m ü. M. bei Oberengstringen an der Limmat). Auch Steigungen und Gefälle zwischen einzelnen Punkten von knapp 90° wurden gemessen. Dieses Problem wurde durch das Ersetzen der Z-Werte mit den Höhen aus dem DTM-AV von der Swisstopo gelöst.

Nachdem die Höhen nun plausible Werte aufwiesen, wurde die Geschwindigkeit bei Steigungen und Gefällen kontrolliert. Dabei sollten unrealistische Werte für Fahrradfahrer (z.B. eine Geschwindigkeit von 10 m/s für eine Steigung von 20°) ausgemacht werden.

Schliesslich wurde die Topologie der einzelnen Etappen (Selbstverschneidungen) untersucht. Die auftretenden Fehler wurden einzeln und manuell begutachtet sowie nötigenfalls entfernt. Abbildung 15 zeigt einen typischen Fall, bei dem die Daten manuell beseitigt wurden. Notwendig war dieser Schritt, da solche Cluster durch die Dichtefilterung nicht behoben wurden (die Dichte lag öffentlich unter dem festgelegten Schwellwert)

Abbildung 15 Mehrere Selbstverschneidungen einer Etappe



Ergebnisse

Die Auswahl der Daten mit einer Wahrscheinlichkeit $p_{\text{Bike}} \geq 0.5$ eliminiert mehr als die Hälfte der Punkte, weshalb noch 33651 übrig blieben.

Die Untersuchung der maximalen Steigung (vgl. Abb. 16) und des maximalen Gefälles erzielte keine überraschenden Ergebnisse. Die maximal auftretende Neigung (Steigungen und Gefälle zusammen betrachtet) beträgt rund 20° . Diese Zahl ist zwar ziemlich hoch, doch betrifft sie lediglich Einzelfälle und rechtfertigt keine Fehler. Ausserdem weisen rund 50 % der Daten Neigungen zwischen 0.5 und 3° auf. Somit wurden keine weiteren Punkte eliminiert.

Der Plausibilitätsvergleich der Geschwindigkeit mit der Steigung bzw. dem Gefälle zeigte ebenfalls keine unerwarteten Werte. Etwas eigenartig ist einzig die Tatsache, dass die Geschwindigkeit, wie in Abbildung 18 dargestellt, mit zunehmendem Gefälle tendenziell nicht zunimmt. Anscheinend waren die Probanden vorsichtige Fahrradfahrer. Bei der Beurteilung der Steigung lässt sich hingegen erkennen, dass mit zunehmender Neigung die Geschwindigkeit abnimmt.

Abbildung 16 Maximale Neigung der GPS-Etappen

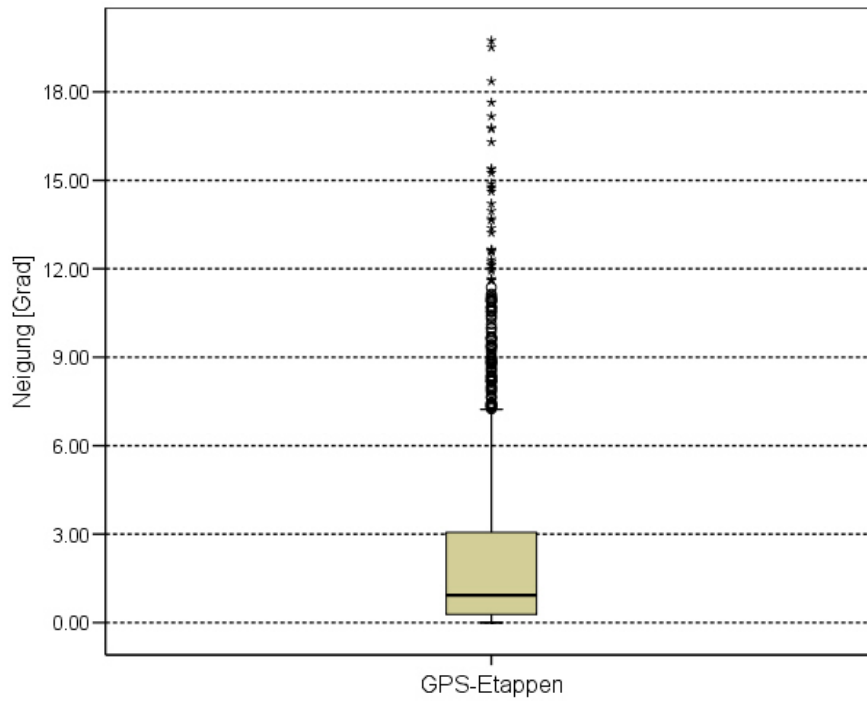


Abbildung 17 Geschwindigkeit verglichen mit der Steigung

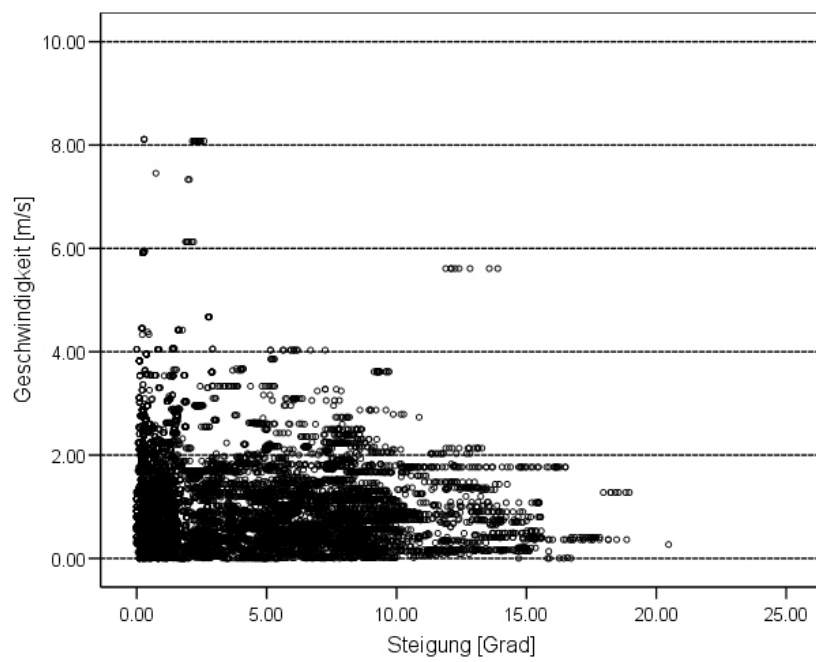


Abbildung 18 Geschwindigkeit verglichen mit dem Gefälle

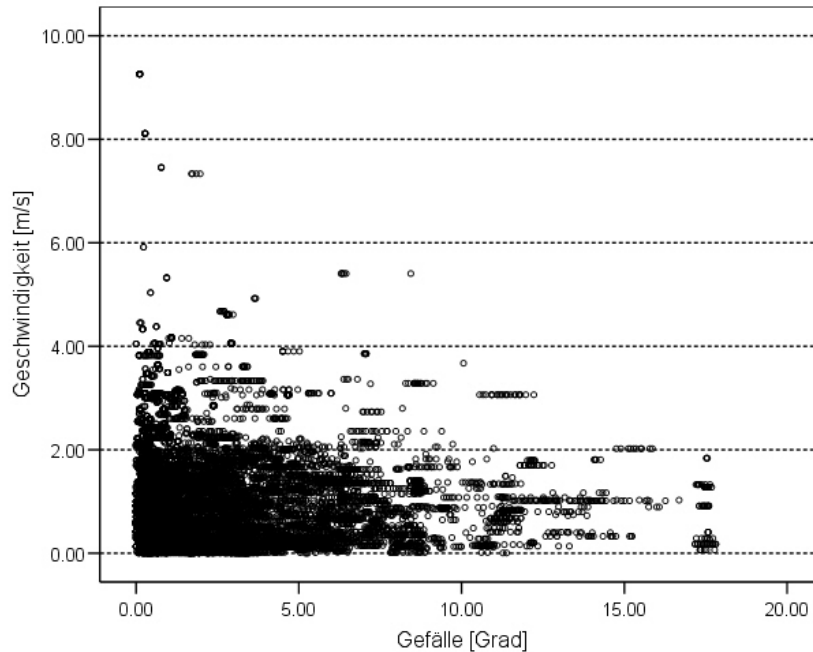


Tabelle 6 fasst die Filterungsergebnisse zusammen. Gut ersichtlich ist, dass ein grosser Teil der Daten eliminiert wurde. Diese Tatsache ist nicht zu beschönigen, doch ab diesem Zeitpunkt ist die Wahrscheinlichkeit, dass es sich bei den Daten wirklich um Fahrradfahrer handelt, sicherlich beträchtlich erhöht. Überdies sind immer noch genug Daten für die Schätzung eines diskreten Entscheidungsmodells vorhanden. Auf eine weitere separate Diskussion wird in diesem kurzen Unterkapitel verzichtet.

Tabelle 6 Ergebnisse nach der speziellen Filterung des GPS-Datensatzes

Filterungsschritt	Anzahl Punkte	Anzahl Etappen	Übrig gebliebene Punkte/Etappen in %
Ursprüngliche Daten (nach allg. Filterung)	134057	9336	11.46/17.81
$p_{\text{Bike}} > 0.5$	33651	2672	2.88/5.10
Manuelle Filterung aufgrund fehlerhafter Topologie	22952	1624	1.96/3.10

4.2.4 Matching der GPS-Daten

Nachdem die GPS-Grunddaten bereinigt wurden, sollten diese Daten nun auf das Strassennetz zugewiesen werden. Ziel ist es die aufgezeichneten GPS-Koordinaten in bestmöglicher Genauigkeit auf das bestehende Strassennetzwerk zu matchen.

Methodenwahl

Analog zur Wahl eines *Matching* Algorithmus bei den Strassendaten, wurden auch bei diesem Schritt mögliche Alternativen sorgfältig abgewägt. Die Entscheidungsgründe sind dabei mehrheitlich dieselben wie beim ersten *Matching* Problem. Zwei Hauptgründe gaben den Ausschlag für die Wahl des Algorithmus nach Marchal *et al.* (2006). Erstens ist der gesamte Code am IVT vorhanden und zweitens wurde in früheren Studien gezeigt, dass gute Resultate innert kurzer Zeit erreicht werden können. Nach Marchal *et al.* (2006) ist der vorgestellte Ansatz fähig ganze Datenbanken mit Tausenden von Wegen in akzeptabler Zeit zu prozessieren. Es können rund 2000 Punkte pro Sekunde verarbeitet werden, was rund 1000 Mal schneller ist als die GPS-Datenaufzeichnung selbst. Dies entspricht ungefähr der doppelten Geschwindigkeit gegenüber anderen Algorithmen wie diejenigen von Nielsen und Jørgensen (2004) oder Wolf *et al.* (2004).

Methodik

Die genaue Funktionsweise und der Aufbau des Algorithmus von Marchal *et al.* (2006) sind im Anhang A4 zu finden. Analog zum *Matching* Prozess bei den Strassendaten wird hier lediglich auf die technische Ausführung eingegangen.

Vor der Ausführung der *Matching Software* musste einige Vorarbeit geleistet werden, was beispielsweise die Installation einer PostGIS Datenbank (vgl. Glossar) bedingte. Zudem musste die Opensource Software GDAL⁹ auf dem Rechner installiert werden, damit auch die Export Funktion unterstützt wird.

Nach der Installation dieser Zusatzprogramme konnten das Strassennetz und die GPS-Daten in die PostGIS Datenbank importiert werden.

⁹ <http://www.gdal.org>, Zugriff: 10.01.2008

Für das Strassennetz werden die Formate von Navteq und Tele Atlas unterstützt. Da die hier verwendeten Daten weder das eine noch das andere Format besaßen, mussten die Daten zunächst auf eine dieser Formen gebracht werden. Hierbei wurde das Navteq Format als Referenz gewählt und das Netzwerk mit den entsprechenden Attributen ergänzt. Zumal für das *Matching* nicht alle Informationen relevant sind, wurden gewisse Attributspalten leer gelassen. Sehr wichtig ist aber zum Beispiel der Anfangs- und Endknoten jeder Strassenkante, welche nach dem Strassen-*Matching* fehlten. Diese Berechnung erfolgte in ArcGIS.

Gleich wie die Strassendaten wurden auch die GPS-Punkte in die PostGIS Datenbank importiert. Für diesen Zweck unterstützt die Software zwei etwas unterschiedliche Textformate sowie den GPS-spezifischen Typ NMEA (vgl. Glossar). Die genaue Datenstruktur mit den notwendigen Attributen ist in der Tabelle 36 im Anhang A1 aufgelistet.

Nachdem die Grunddaten erfolgreich in die PostGIS Datenbank importiert wurden, erfolgte das eigentliche *Matching*. Hierbei lassen sich einige relevante Parameter bestimmen. So können beispielsweise das räumliche Bezugssystem, die minimale Distanz oder der minimale zeitliche Abstand zwischen zwei Punkten festgelegt werden. In der Tabelle 7 sind die verschiedenen Parameter mit den in dieser Arbeit gewählten Werten beschrieben.

Tabelle 7 *Matching*-Eigenschaften

Eigenschaft	Beschreibung	Wert
srIdOutTable	SRID (Spatial Reference ID) der Outputtabelle	21781(= CH1903)
orientedLink	Berücksichtigung der Orientierung der Strecken	false
keepOnes	Pfade mit einer Strecke beibehalten?	false
whereCondition	Bedingung für GPS-Daten (z.B. „speed >5“)	-
minCruisedRatio	Minimal berücksichtigtes Verhältnis für die gefahrene Strecke (0.5 bedeutet z.B., dass mindestens die Hälfte der Kante befahren wurde)	0.5
minPoints	Minimale Anzahl Punkte	8
maxDist	Maximale Distanz zwischen 2 Punkten [m]	200
maxTimeGap	Maximale Zeitdifferenz zwischen 2 Punkten [ms]	120000
minCandidates	Minimale Anzahl Kandidaten für das Matching	8 (default)
maxCandidates	Maximale Anzahl Kandidaten für das Matching	100 (default)
adaptativeFactor	Faktor um Anzahl Kandidaten festzusetzen	0.001 (default)
coeffDistance	Distanzkoeffizient (Differenz zwischen 2 Punkten und 2 dem Strassennetz zugewiesenen Punkten)	0.25 (default)

Ergebnisse

Aufgrund der kurzen Prozessierungszeit (vgl. Tabelle 8), konnte mit den Parameterwerten ein wenig experimentiert werden. Dabei wurde mittels der Open Source Software Quantum GIS¹⁰ eine Verbindung zur PostGIS Datenbank hergestellt, damit die Resultate fortlaufend visualisiert werden konnten. Dies ermöglichte die Auswahl der besten Lösung, ohne einzeln alle Ergebnisse aus der PostGIS Datenbank exportieren zu müssen.

¹⁰ <http://www.qgis.org> , Zugriff: 10.01.2007

Als optimalste Lösung erwies sich diejenige mit den Werten gemäss Tabelle 7. Auf die Rechenleistung haben die Parameter keinen erwähnenswerten Einfluss. Diese wird hauptsächlich von der Datenmenge bestimmt und war wie weiter oben angedeutet sehr gut (vgl. Tabelle 8).

Tabelle 8 Rechenzeit für das Matching und Import der Daten

Arbeitsschritt	Benötigte Zeit
Import der GPS-Daten (22952 Punkte)	18.89s
Import des Strassennetzes (24866 Strecken, 8686 Knoten)	17.29s
<i>Map Matching</i>	27.93s

Abbildung 19 zeigt die *Matching*-Ergebnisse für Punkte, welche einen guten Datenfluss bilden. Mit roter Farbe sind die Strecken hervorgehoben auf denen Punkte zugewiesen wurden.

Abbildung 19 Matching der GPS-Punkte



Quelle: VECTOR25 © 2006 Swisstopo (DV033492.2)

Da in späteren Arbeitsschritten für die ermittelten Routen mögliche Alternativen berechnet werden sollen, wurden einige kleine Routen entfernt. Es handelt sich um diejenigen, welche

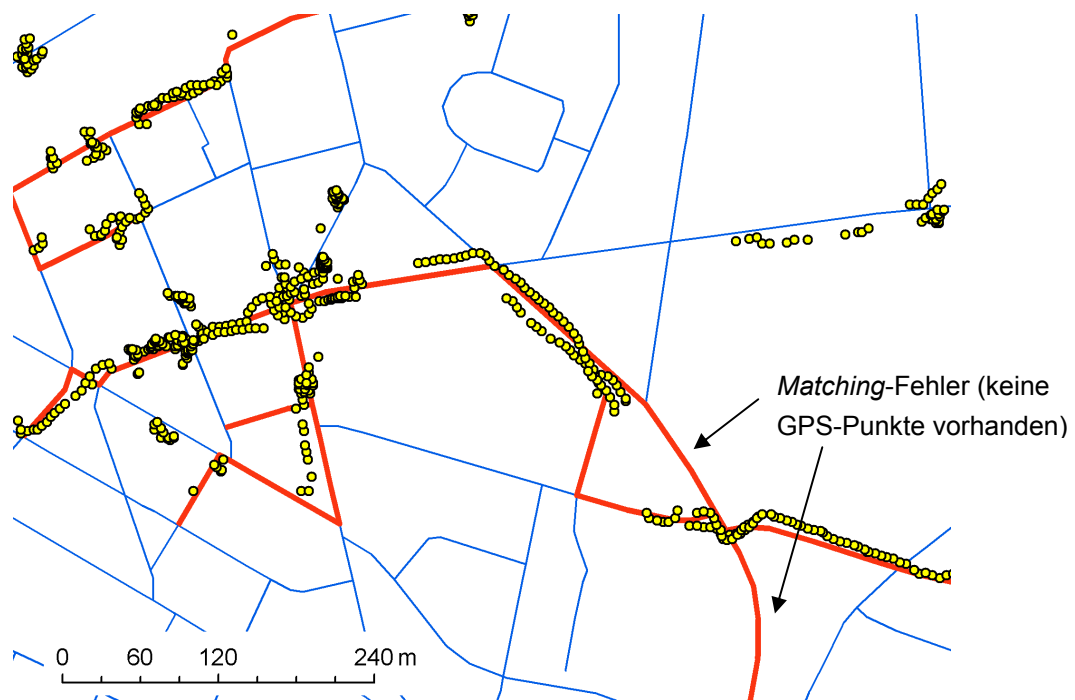
weniger als drei Strassenkanten beinhalten. Ferner wurde untersucht, ob jede Route mindestens eine Kreuzung (anhand des Knotentyps) beinhaltet. Wäre dies nicht der Fall, könnten später keine Routenalternativen berechnet werden.

Aufgrund dieses Kriteriums blieben für die späteren Berechnungen 258 Routen übrig. Die Länge derselben ist in Abbildung 36 des Abschnitts 8.3 dargestellt.

Diskussion

Etwas problematisch erwies sich die Situation, in der die GPS-Punkte keinen durchgehenden Datenfluss bilden. Der Grund dieser Zerstückerung liegt mit grösster Wahrscheinlichkeit am Algorithmus für die Detektion der Verkehrsmittel. Obwohl die Parameter so gesetzt wurden, dass beispielsweise mindestens die Hälfte einer Strassenkante gefahren sein sollte, wurde diese Limite durch den Algorithmus teilweise missachtet (vgl. Abb. 20). An gewissen Orten wurden einige wenige Punkte einer deutlich längeren Strecke zugeordnet. Dieser Fehler kann aber auch ein Vorteil des *Matchings* sein, da Zerstückerungen bzw. Löcher in den Daten durch die Zuweisung einer Strassenkante teilweise aufgehoben werden. Hierbei ist die Wahl der *Matching*-Parameter aber besonders wichtig und ständig zu überprüfen.

Abbildung 20 Matching der GPS-Punkte: Probleme und Fehler

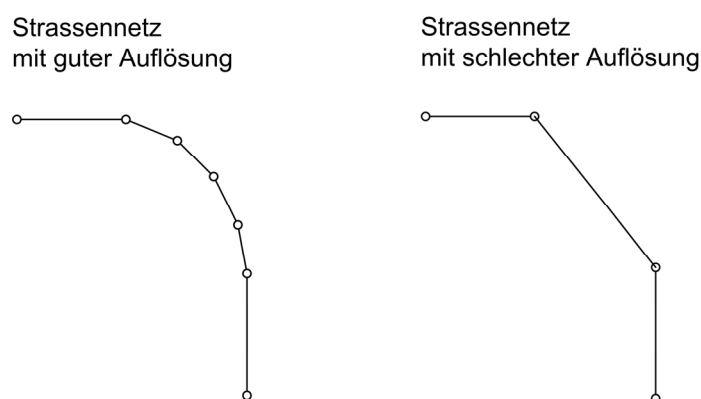


Quelle: VECTOR25 © 2006 Swisstopo (DV033492.2)

Weitere Fehler können bei komplexen Strassensituationen mit mehreren Kreuzungen oder Fahrspuren entstehen. Sehr gute Resultate ergeben sich hingegen in Situationen, in denen die Punkte gut entlang der Strasse verlaufen und wenig zerstückelt (durchgehender Datenfluss ohne Lücken) sind.

Nebst den Fehlern durch den Algorithmus sollten beim *Map Matching* stets zwei mögliche Quellfehler beachtet werden. Erstens haben die GPS-Geräte ihre eigenen Limiten und zweitens ist das Strassennetz eine Approximation der physischen Umwelt und daher lückenhaft und unpräzise. Strassen können fehlen oder deren Beschreibung kann in schlechter Auflösung sein (vgl. Abb 21).

Abbildung 21 Strassennetz Auflösung



Die Länge der generierten Routen erscheint für Fahrradfahrer auf den ersten Blick eher kurz (vgl. Abb. 36). Die mittlere Länge der gewählten Routen beträgt knapp 500 m. Auch die maximale Länge ist mit rund 1 km etwas kurz. Ein Vergleich mit den Mikrozensus Daten des Jahres 2000, wo durchschnittliche Fahrradrouten 1.34 km lang sind, akzentuiert diese Tatsache zusätzlich. Diese Problematik kann etwas abgeschwächt werden, wenn die genaue Entstehung der Routen berücksichtigt wird. Als maximale Zeitlücke (TimeGap) zwischen einzelnen Punkten wurden 120 Sekunden gewählt. Somit werden Wege, bei denen ein Fahrradfahrer beispielsweise längere Zeit an einer Lichtsignalanlage hielt, aufgeteilt. Auch sonstige Zwischenhalte führen zu mehreren Routen. Diese Tatsachen relativieren deshalb die etwas kurzen Routen. Nichtsdestotrotz sollten die Resultate kritisch beurteilt werden. Zum wiederholten Male wird auf die Mängel des Algorithmus für die Detektion des Verkehrsmittels hingewiesen.

4.3 Ergänzung des Strassennetzes um weitere Eigenschaften

Zu den Grunddaten für die spätere Modellschätzung zählen nicht nur die GPS-Punkte und das Strassennetz, sondern auch andere Daten, welche das Strassennetz um wichtige Eigenschaften ergänzen. In diesem Kapitel wird als Erstes das *Matching* der Lichtsignalanlagen auf das Strassennetz beschrieben. Ferner folgen Ausführungen über die Ergänzung des Strassennetzes mit genauen Höhendaten sowie Erklärungen über die Zuordnung von Kategorien des durchschnittlichen Tagesverkehrs.

4.3.1 Matching der Lichtsignalanlagen

Methodenwahl

Aufgrund der etwas speziellen Art dieser Daten wurde hier auf einen komplizierten *Matching* Algorithmus verzichtet. Der Grund liegt daran, dass die als Punktdaten vorhandenen Lichtsignalanlagen ohne grössere Probleme dem Strassennetz zugeordnet werden können. Die nachfolgend beschriebenen Aufbereitungsschritte wurden alle mittels der Software ArcGIS ausgeführt. Abbildung 70 im Anhang A2 illustriert, analog zu den anderen Aufbereitungsschritten, die weiter unten folgende Methodik graphisch.

Methodik

Vor der Erläuterung der eigentlichen Vorgehensweise soll einleitend eine in diesem Zusammenhang wichtige Annahme dargelegt werden.

Die erwähnte Überlegung betrifft die Problematik an Strassenkreuzungen, wo ein Ampel-Punkt möglicherweise mehrere Lichtsignalanlagen repräsentiert (vgl. Abb. 22). Hierbei wird angenommen, dass jede Strasse, die in eine Kreuzung mündet, eine Lichtsignalanlage besitzt. Somit kann ein einzelner Punkt mehrere Anlagen repräsentieren. Aufgrund dieser Gegebenheit sollten solche Lichtsignalanlagen exakt auf den Knoten der Kreuzung zugewiesen werden. So ist gewährleistet, dass alle in eine Kreuzung mündenden Strassen einen Ampel-Punkt besitzen. Alle übrigen Ampel-Punkte, welche nicht an einer Kreuzung liegen, können dem am nächsten liegenden Knoten oder auf der nächst liegenden Strassenkante zugeordnet werden.

Abbildung 22 Lichtsignalanlagen an Strassenkreuzungen



Quellen: SWISSIMAGE, VECTOR25 © 2006 Swisstopo (DV033492.2); Dienstabteilung Verkehr Stadt Zürich

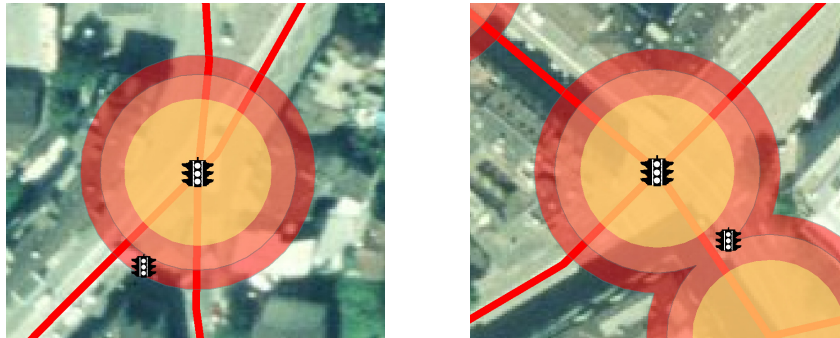
Nach einer Formatumwandlung (Transformation des CAD-Formates dxf in ein Shapefile) der Daten und anschliessender Georeferenzierung wurde das eben beschriebene Problem folgendermassen gelöst:

Als Erstes wurden alle Knoten des Strassennetzes extrahiert. Über die Annahme, dass in eine Strassenkreuzung mindestens drei Strassen münden, konnten anschliessend die Kreuzungen ausgemacht werden. Um diese Kreuzungen wurde nun ein Buffer von 24 m erzeugt. Alle Lichtsignalanlagen, die innerhalb dieses Buffers lagen, wurden dann dem Knoten der Strassenkreuzung zugewiesen.

Die Festlegung der Buffergrösse geschah visuell und iterativ durch die Erprobung mehrerer Grössen. Abbildung 23 zeigt das Problem eines zu klein gewählten Buffers. Der gelbliche Kreis repräsentiert eine Buffergrösse von 16 m, Orange 20 m und Rot schliesslich 24 m. Die ursprünglichen Lichtsignalanlagen sind etwas kleiner dargestellt. Wie auf der linken Abbildung sichtbar, wird die Lichtsignalanlage nur beim grössten Buffer korrekt zugeordnet. Hingegen zeigt das rechte Bild das auftretende Problem, falls die Knoten sehr nahe beieinander

liegen. In solchen Fällen wurde die Lichtsignalanlage ganz einfach auf den am nächsten liegenden Knoten verschoben.

Abbildung 23 Matching der Lichtsignalanlagen



Quellen: SWISSIMAGE, VECTOR25 © 2006 Swisstopo (DV033492.2); Dienstabteilung Verkehr Stadt Zürich

Ergebnisse

Von total 394 Punkten waren 190 im 24 m Buffer enthalten, 157 waren es bei 20 m und 109 bei 16 m Buffergrösse. Diese grosse Anzahl ist nicht weiter erstaunlich, da Lichtsignalanlagen gewöhnlich an Strassenkreuzungen liegen. Insofern sollten sie auch im Buffer enthalten sein.

Insgesamt konnten 29 von 394 Punkten nicht zugeordnet werden. Dies entspricht rund 7 %. Diejenigen Punkte, welche nicht automatisch integriert werden konnten, wurden schliesslich manuell der wahrscheinlichsten Strassenkante zugeordnet.

Diskussion

Anhand einer visuellen Überprüfung wurde der ganze Ansatz evaluiert. Dass ein paar Fehler entstehen können, ist offensichtlich. Schwierig ist es aber diese ausfindig zu machen, da man nicht genau weiss, wie viele Lichtsignalanlagen eine Strassenkreuzung in der realen Welt tatsächlich besitzt. Nur eine Begehung vor Ort könnte eine absolute Genauigkeit für die Überprüfung garantieren.

Des Weiteren kann es sein, dass das Strassennetz nicht vollständig ist oder dass die Lichtsignalanlagen ausserhalb der visuell festgesetzten Toleranz von 24 m auf beiden Strassenseiten liegen. Zu guter Letzt wurden auch Ungenauigkeiten bei der Georeferenzierung eingeführt. Hierbei gilt es den dabei entstanden RMS-Fehler von 0.029 m zu erwähnen (vgl. Glossar). Als Transformationsmethode wurde eine Affine Transformation gewählt.

Trotz der genannten Unsicherheiten erscheint der hier gewählte Ansatz zufriedenstellend, denn ein möglicherweise genauerer Matching Algorithmus würde nicht unbedingt bessere Resultate versprechen. Dies weil bereits die Grunddaten einige Fehler und Lücken aufweisen.

4.3.2 Ergänzung des Strassennetzes mit topographischen Daten

Zum Zwecke der Berechnung verschiedener Routeneigenschaften wurde das Strassennetz mit Höhen aus dem DTM-AV ergänzt. Darauf basierend lassen sich für den später generierten Schätzdatensatz (gewählte Routen und Alternativen) nun folgende Eigenschaften berechnen:

- Maximale und minimale Steigung
- Maximales und minimales Gefälle
- Durchschnittliche Steigung
- Durchschnittliches Gefälle
- Kumulierte Höhenmeter aufwärts
- Kumulierte Höhenmeter abwärts

4.3.3 Bestimmung des durchschnittlichen Tagesverkehrs für alle Strassenklassen

Die für diese Arbeit wichtigen Angaben über den durchschnittlichen Tagesverkehr (DTV) liegen lediglich für die Strassen des Netzes vom Kanton Zürich vor. Somit enthalten die Strassen aus den anderen Netzen (ca. 2/3 aller Strassen) keine solchen Informationen. Dies ist bedauernd, weshalb sich eine Abschätzung für alle anderen Strassen aufdrängt.

Durch das Matching der unterschiedlichen Netze konnten die vorhandenen DTV-Werte mit den Strassenklassen aus VECTOR25 verglichen werden und entsprechende Grössen abgeleitet werden. Eine Sonderbeachtung gebührte dabei denjenigen Strassen, die nicht einer VECTOR25-Klasse zugeordnet werden konnten. Bei den Strassen ohne Klassierung handelt es sich um Fahrradwege von MAP Zürich oder vom kommunalen Richtplan. Diese ergänzen die übrigen Strassen durch neue geometrische Informationen und überlappen nicht mit den übrigen Daten. Es sind also separate Fahrradwege, weshalb kein Verkehr angenommen wurde.

Die folgende Tabelle 9 illustriert einige statistische Grössen zum eben genannten Vergleich. Auf Basis dieser Gegenüberstellung erfolgte die Klassierung aller Strassen (vgl. Tabelle 10).

Hierbei wurde versucht ähnliche Strassenklassen (in Bezug auf die Verkehrswerte) zu einer einzigen DTV-Kategorie zusammenzufassen.

Tabelle 9 Statistische Grössen für die Festlegung der DTV-Kategorien

Strassenklasse	DTV: Durchschnitt	DTV: Standardabw.	DTV: Minimalwert	DTV: Maximalwert
1. Klass Strasse	15455	11618	48	65726
2. Klass Strasse	6188	5509	2	26146
3. Klass Strasse	51	358	2	2995
4. Klass Strasse	0.7	23.4	0	808
5. Klass Strasse	0	0	0	0
6. Klass Strasse	0	0	0	0
Quartier Strasse	668	2008	2	20729
Parkweg	0	0	0	0
Steg	0	0	0	0
Autobahn	30882	14826	13487	95264
Autostrasse	27941	26600	281	95264
Ein- und Ausfahrt	7783	10637	43	81478
Autobahn Zufahrt	4833	4789	2220	15059
k.A.	0	0	0	0

Tabelle 10 DTV-Kategorien

Kategorie	Werte (Autos/Tag)	Strassenklasse	Beschreibung
1	0	5. Klass-, 6.Klass- Strassen, Velowege	Kein Verkehr
2	1-19	4. Klass-Strassen, Parkweg	Geringer Verkehr
3	20-999	3. Klass-Strasse, Quartier-Strasse	Mässiger Verkehr
4	1000-19999	1. Klass-Strasse, 2. Klass-Strasse	Stark befahrene Strasse
5	20000-100000	Autobahn, Autostrasse	Sehr stark befahrene Strasse

Diskussion

Die Klassierung erwies sich aufgrund der grossen Standardabweichung der DTV-Werte pro Strassenklasse als sehr schwierig. Aus diesem Grund gilt es einige Punkte zu beachten. Zum einen liegt einer solchen Klassierung immer eine gewisse Subjektivität zugrunde. Des Weiteren führt die grosse Standardabweichung zu einer Überlappung der verschiedenen Klassen, welche aber nicht berücksichtigt werden kann. Eine besondere Stellung nehmen die Strassen der Kategorie 5 (Autobahn, Autostrasse) ein. Diese wird beim später verwendeten Schätzdatensatz nicht vorkommen, weil Autobahnen und Autostrassen von Fahrradfahrern nicht befahren werden dürfen.

Eine alternative Klassierungsmöglichkeit könnte beispielsweise über das schweizerische Hektarraster erfolgen. Hierzu könnte das Verkehrsvolumen auf Basis der Einwohnerzahlen entlang der Strassen abgeschätzt werden. Diese Methode wäre aber wegen der grobmaschigen Rasterauflösung (100 m x 100 m) mit grosser Wahrscheinlichkeit ungenauer.

5 Ermittlung der Alternativen

Nach der Aufbereitung aller Grunddaten wurden für die beobachteten Routen die jeweils möglichen Alternativen ermittelt. Zu diesem Zweck wurden für die gewählten Routen alle Start- und Zielpunkte, sogenannte Origin-Destination-Paare (OD-Paare), identifiziert. Diese Information, welche aus ArcGIS als Textdatei exportiert wurde, dient neben dem vollständigen Strassennetz als Input für die nachfolgenden Berechnungen. Das Strassennetz wurde ebenfalls als Textdatei, mit einer ID pro Kante sowie deren Start- und Endknoten, ausgegeben. Da als Kostenattribute lediglich die Länge der Strecken berücksichtigt wird, ist dies die einzig notwendige Zusatzinformation für das Netz.

Wie in Kapitel 5 erwähnt, existieren für diese Berechnung etliche Methoden. In dieser Arbeit wird ein Algorithmus verwendet, der in seiner Grundform beispielweise schon in Azevedo *et al.* (1993) vorgestellt wurde. Der verwendete Ansatz, welcher nun detailliert vorgestellt werden soll, wurde als Modul in MATSim implementiert. Näheres zu MATSim findet man im Glossar.

5.1 Methodik

Das generelle Vorgehen soll anhand eines einfachen Beispiels, welches in Abbildung 24 dargestellt ist, erklärt werden.

In einem ersten Schritt (Baumlevel B_0) wird für das erste OD-Paar, basierend auf dem ursprünglichen Netz N_0 , die kürzeste Route berechnet. Wie in Abbildung 24 ersichtlich, besteht diese Route aus den Strecken l_i (hier l_1 , l_2 und l_3). Es gilt dabei zu beachten, dass die kürzeste Route denjenigen Pfad darstellt, welcher die Kostenattribute (hier wird lediglich die Länge berücksichtigt) minimiert.

In der nächsten Hierarchieebene (B_1) werden nun drei Subnetze N_1 mit je einer fehlenden Kante (l_i) der vorher berechneten kürzesten Route gebildet (selbstverständlich ist die Anzahl Subnetze von der Anzahl Kanten der kürzesten Route abhängig). Für jedes Subnetz N_1 wird dann die neue kürzeste Route (l_j) gerechnet. Ist eine Route in der Liste bereits vorhanden, so wird diese nicht nochmals gespeichert und bei diesem Ast abgebrochen (vgl. durchgestrichenes Rechteck in der Abbildung 24).

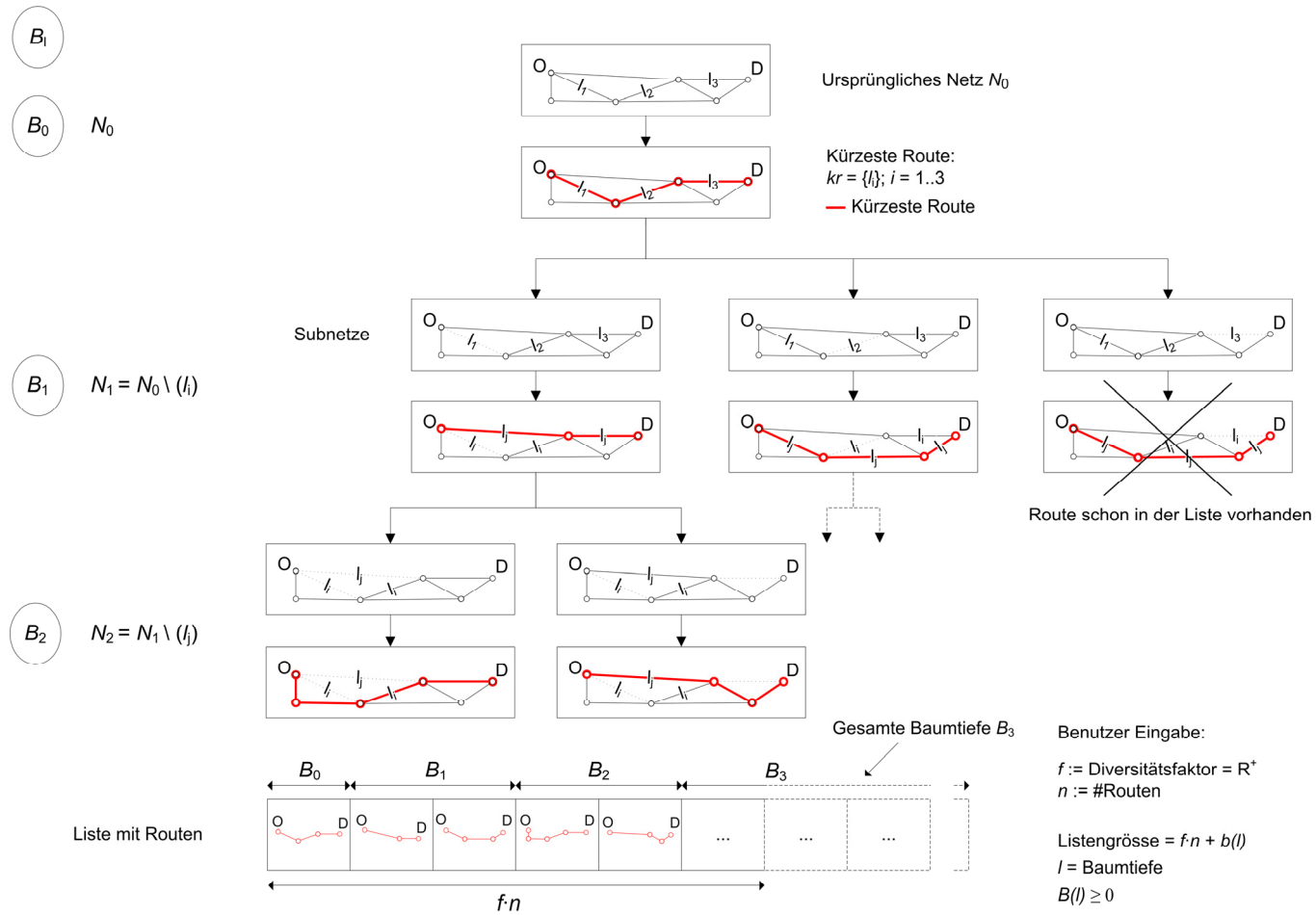
Die nächste Bauebene (B_2) berechnet wiederum neue kürzeste Routen. Diesmal aufgrund der Subnetze N_2 mit zwei fehlenden Kanten (l_i und l_j). Der Baum wird solange expandiert, bis mindestens $f \cdot n$ Alternativen gefunden wurden. Dabei ist n das Abbruchkriterium (Anzahl Alternativen) und f ein Diversitätsfaktor, welcher mit n multipliziert wird. Weitere Erklärungen zum Diversitätsfaktor f folgen weiter unten.

Eine wichtige Eigenschaft des Algorithmus ist, dass eine Breitensuche (*breadth-search*) angewendet wird. Dies um möglichst unterschiedliche Routen zu erhalten. Da also von der Wurzel des Baumes aus immer eine ganze Bauebene expandiert wird, kann es sein, dass mehr als die wirklich erforderlichen Routen berechnet werden. Eine Tiefensuche (*deep-search*) wäre nicht geeignet, da z.B. die letzte Kante des OD-Paares eine viel kleinere Wahrscheinlichkeit hat in den Subnetzen eliminiert zu werden, als die erste Kante. Somit wären die daraus generierten Alternativen für diese Art von Aufgabe unbrauchbar.

Weiter oben wurde als Abbruchkriterium $n = 20$ angegeben. Aufgrund der Breitensuche würden so aber sehr viele kurze Routen mit kleiner Variabilität generiert werden. Um die Unterschiede zwischen den einzelnen Alternativen zu verstärken, wurde deshalb ein Diversitätsfaktor f eingeführt, der mit n multipliziert wird. Folglich wird der ursprünglich erforderliche Alternativensatz f -mal vergrößert. Die Expansion um $f = 3$ erwies sich für diese Arbeit als ideal. Der gesamte Alternativensatz hat nun folgende Grösse: $f \cdot n + b(l)$, wobei l die Baumtiefe ist und $B(l) \geq 0$. Der Term $b(l)$ kommt wie bereits erwähnt dadurch zustande, dass immer eine ganze Bauebene expandiert wird. Im Minimum stehen also für jedes OD-Paar 60 Alternativen zur Verfügung.

Schliesslich beinhaltet der letzte Schritt die Auswahl von 20 Routen aus dem gesamten Alternativensatz von $60 + b(l)$. Dabei wird die kürzeste Route immer ausgewählt, die anderen 19 Routen werden hingegen zufällig ausgesucht (vgl. nächste Seite).

Abbildung 24 Algorithmus für die Generierung des Alternativensatzes

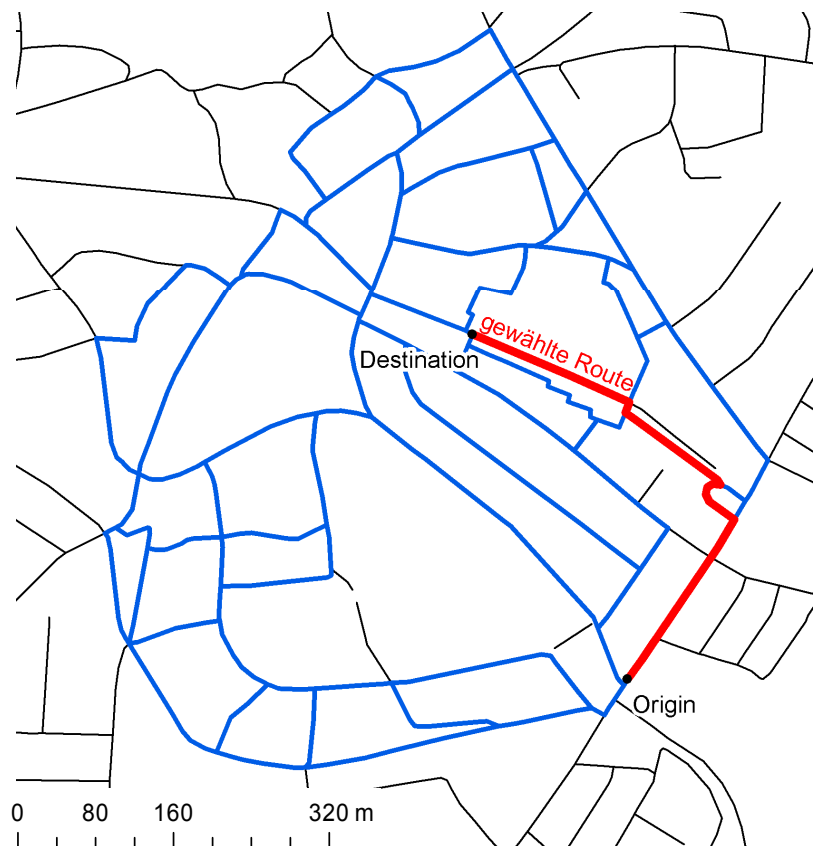


5.2 Ergebnisse

Insgesamt wurden für 165 OD-Paare je 20 Alternativen ermittelt. Für den definitiven Schätzdatensatz kommt zusätzlich die gewählte Route hinzu. Ist diese bereits vorhanden, so wird eine weitere zufällige Alternative hinzugefügt, so dass pro OD-Paar immer 21 Routen zur Verfügung stehen. In Zukunft sind mit dem Begriff Alternativen die nicht gewählten Routen gemeint.

Abbildung 25 illustriert das Ergebnis für ein solches Paar. In blauer Farbe sind die von den 20 Alternativen befahrenen Strecken dargestellt, rot hervorgehoben ist die jeweils gewählte Route.

Abbildung 25 Alternativen und gewählte Route fuer ein OD-Paar



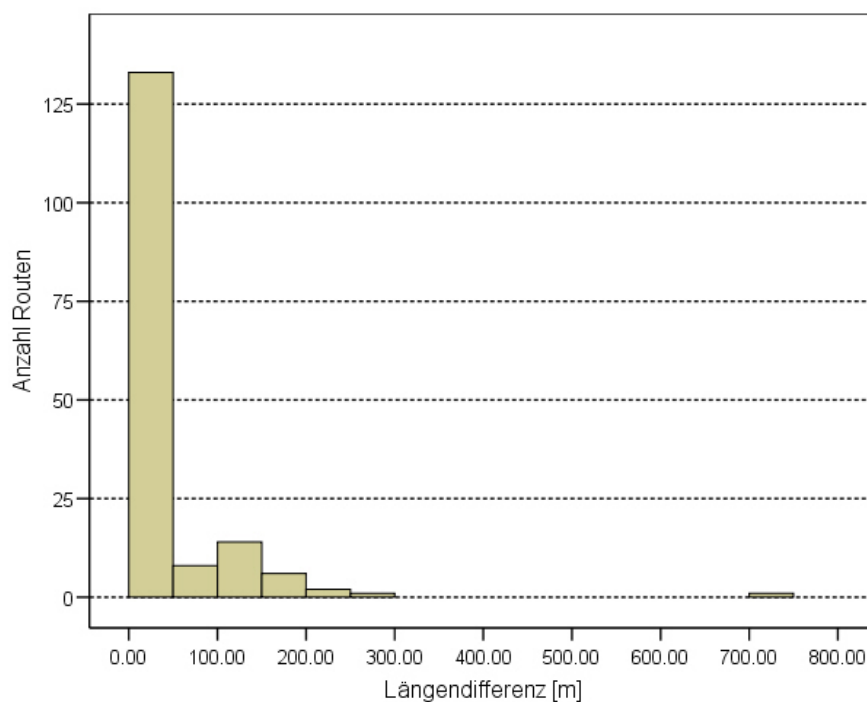
Quelle: VECTOR25 © 2006 Swisstopo (DV033492.2)

Die Länge der Alternativen variiert gut und ist in Abbildung 36 dargestellt. Verglichen mit den gewählten Routen sind die Alternativen um einiges länger. Hierbei darf jedoch nicht vergessen werden, dass jeweils mindestens 60 Alternativen pro OD-Paar generiert wurden und daraus 20 zufällig ausgewählt wurden. Folglich ist es klar, dass die Routen dementsprechend länger sein können.

Wirklich aufschlussreich ist lediglich der Vergleich der gewählten Route mit der kürzesten Route, welche im Schätzdatensatz speziell gekennzeichnet wurde.

Im Durchschnitt sind die gewählten Routen rund 60 % länger als die kürzesten Verbindungen. Jedoch sollte dieser Wert mit Vorsicht interpretiert werden, da Ausreisser dieses Ergebnis verfälschen können. Zur besseren Interpretation illustriert Abbildung 26 diesen Längenunterschied anhand eines Balkendiagramms. Sehr gut ersichtlich ist, dass der grösste Teil der gewählten Routen eine Differenz von kleiner gleich 50 m zur kürzesten Verbindung aufweisen. Grössere Unterschiede treten hingegen seltener auf. Dies lässt vorwegnehmen, dass kurze Routen für Fahrradfahrer anscheinend sehr wichtig sind.

Abbildung 26 Längendifferenz zwischen der gewählten und der kürzesten Route

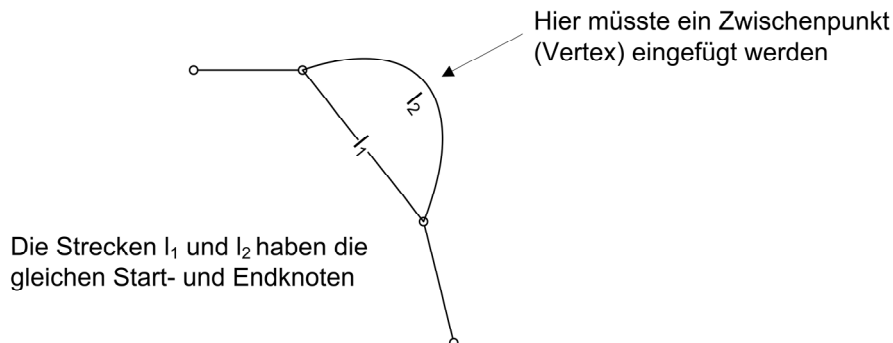


5.3 Diskussion

Durch die Integration von vier verschiedenen Strassennetzen unterschiedlicher Qualität sind bei der Berechnung der Alternativen einige wenige Probleme aufgetaucht.

Offenbar konnten bei den Strassen von MAP Zürich nicht alle *Dangling Nodes* (vgl. Kapitel 4.1) eliminiert werden. Die Folge ist, dass nicht jeder Knoten jeden anderen erreicht. Der Grund solcher Fehler liegt daran, dass vor allem schmale Feldwege, aber auch Fahrradwege auf offenem Gelände, eigenständig im Raum liegen können und so keine Verbindung zu anderen Strassen aufweisen. Ferner kann die Situation auftreten, dass zwei Kanten die gleichen Start- und Endpunkte aufweisen (vgl. Abb. 27). Korrekterweise müsste an einer Strassenkrümmung (Kurve) ein *Vertex* eingefügt werden. Die Ursache dieser Probleme liegt an der unterschiedlichen Netzkodierung der verschiedenen Strassendatensätze. Eine visuelle Überprüfung zeigte auf, dass dies vor allem eine Folge von Qualitätsmängeln im Netz von MAP Zürich ist.

Abbildung 27 Zwei Strecken mit gleichen Start- und Endknoten



Die genannten Probleme wurden von MATSim detektiert und gelöst, indem einige Strassenkanten für die späteren Berechnungen unberücksichtigt blieben. Überdies wurde jede Kante verdoppelt, damit die Fahrt in beide Richtungen möglich ist. Der Grund ist, dass sich viele Fahrradfahrer nicht sonderlich an Verkehrsregeln halten und häufig auch Einbahnstrassen in verkehrter Richtung benutzen (mittlerweile sind jedoch viele Einbahnstrassen für den Radverkehr beidseitig befahrbar). Dieses Verhalten soll auch durch den Alternativensatz abgedeckt werden.

Infolge der beschriebenen Probleme wurde das ursprüngliche Strassennetz mit 24866 Kanten (alle Kanten verdoppelt, damit beide Richtungen befahrbar sind) auf 24680 Strecken reduziert. Dieses resultierende Netz war die Basis für die Berechnung des Alternativensatzes.

Weiter gilt es zu erwähnen, dass der angewendete Algorithmus ein ziemlich pragmatischer und einfacher Ansatz ist. Es ist klar, dass im generierten Alternativensatz auch Routen enthalten sind, die von einem Radfahrer möglicherweise nicht als Alternative berücksichtigt werden. Dies weil sie beispielsweise viel zu lang im Vergleich zur kürzesten Route sind. Interessant zu untersuchen wären die Unterschiede dieses Ansatzes im Vergleich zu einem komplizierteren und möglicherweise besseren Algorithmus (z.B. derjenige von Prato und Bekhor, 2006).

Schliesslich werden hier je 20 Alternativen pro gewählte Route generiert. In der Praxis kann es selbstverständlich sein, dass ein Fahrradfahrer bedeutend weniger Alternativen in Erwägung zieht.

5.4 Ermittlung von Alternativen mit ArcGIS

Nebst der Lösung mit MATSim wurden auch die Möglichkeiten von ArcGIS (Network Analyst) für die Erzeugung eines Alternativensatzes getestet. Automatisiert wurde die Berechnung mit Hilfe eines Pythonskriptes, welches im Anhang A8 zu finden ist.

Der angewendete Ansatz unterscheidet sich von der MATSim-Lösung und beruht auf einer iterativen, zufälligen Verrauschung des ursprünglichen Netzes und anschliessender Berechnung der jeweils kürzesten Route. Die Verrauschung bestand darin, die ursprüngliche Länge jeder Kante zufällig zwischen 0 und 50% zu verlängern. Als Kostenattribut wurde wiederum nur die Länge der Strecken berücksichtigt.

ArcGIS bietet mit dem Network Analyst die Möglichkeit auf Basis eines Netzwerkes mit bestimmten Kostenattributen die kürzeste Route für ein OD-Paar zu berechnen. Als Eingabedaten sind wiederum die OD-Paare sowie das Strassennetz (als ArcGIS Network) erforderlich. Die Lösung mit ArcGIS wurde wegen sehr grossen Rechenzeiten abgelehnt. Folgende Gründe können zu grossen Rechenzeiten führen:

- Es kann immer nur für ein OD-Paar die kürzeste Route berechnet werden. Würden beispielsweise die Paare O_1D_1 sowie O_2D_2 gleichzeitig hinzugefügt werden, würde die kürzeste Route von O_1 nach D_1 , von D_1 nach O_2 und von O_2 nach D_2 ermittelt werden. Der Pfad von D_1 nach O_2 wäre dabei nutzlos und überflüssig.
- Die Verrauschung jeder einzelnen Kante ist ziemlich rechenintensiv.

- Nach jeder Verrauschung muss das Netzwerk neu erzeugt werden (*build network*), was wiederum viel Rechenzeit in Anspruch nimmt.

Für die genannten Gründe gibt es nicht immer eine erdenkliche Lösung. So können das erste und letzte Problem keineswegs gelöst werden, da diese abhängig von der verwendeten Software sind. Für den zweiten Punkt bestünde die Möglichkeit, jeweils nur die Kanten der gerade berechneten Route zu verrauschen oder einen ähnlichen Ansatz wie beim weiter oben erklärten Algorithmus zu verwenden.

Aufgrund der speziellen Ausrichtung von MATSim auf Routenberechnungen, wurde eine Erweiterung und/oder Verbesserung der Methode mit ArcGIS ausser Acht gelassen. Des Weiteren wurden keine brauchbaren Literaturreferenzen für die Lösung ähnlicher Probleme mit ArcGIS gefunden. Das im Anhang zu findende Pythonskript soll dennoch einen Anhaltspunkt darstellen, falls Routenberechnungen in Kombination mit ArcGIS angestrebt werden. Einfachere Routenkalkulationen lassen sich nämlich ohne Probleme mit ArcGIS lösen.

6 Berechnung der Ähnlichkeit sich überlappender Routen

Wie in Abschnitt 2.5 dargelegt, ist für ein realistisches Entscheidungsmodell zwingend notwendig die Ähnlichkeit der Routen zu berücksichtigen. Die grundsätzliche Idee dahinter fusst auf der Annahme, dass der Nutzen einer Alternative mit dem Ähnlichkeitsgrad zu einer anderen Alternative abnimmt. Je nach Situation können ganz unterschiedliche Ähnlichkeiten auftreten. Im privaten Verkehr wird häufig die Ähnlichkeit durch Benutzung gleicher Wegstücke betrachtet. Dieser Korrelationswert wurde auch für die hier vorliegenden Routen ermittelt. Dabei wurde der *Path-Size*-Faktor, welcher später in den systematischen Teil der Nutzenfunktion einfließt, berechnet.

6.1 Methodik

Um die besagten Ähnlichkeiten zu berechnen, wurde das Programm BioRoute (vgl. Glossar) verwendet. Die Software hilft zudem die Eingabedaten für das Programm BIOGEME, welche für die Schätzung des Entscheidungsmodells benutzt werden, vorzubereiten.

Als Eingabe für BioRoute wird eine Textdatei, die in verschiedenen Abschnitte (Sektionen) gegliedert ist, benötigt. Ein Beispiel einer solchen Datei ist im Anhang A5 dargestellt (vgl. Tabelle 47). Im Teil [OverlapMeasure] kann angegeben werden, wie die Überlappung berechnet werden soll. Hierbei können theoretisch alle kanten-additiven Attribute des Strassennetzes verwendet werden. Typischerweise wird jedoch eine Kombination aus Reisezeit und Länge benutzt (Bierlaire, 2005). Da die Reisezeit in dieser Arbeit nicht als Routeneigenschaft vorhanden ist, wurde die Ähnlichkeit auf Basis der Streckenlänge ermittelt.

BioRoute verwendet als Ähnlichkeitsmass die *Path-Size* (PS) Grösse nach Ben-Akiva und Bierlaire (1999), welche wie folgt definiert ist:

$$PS_{in} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj} \frac{L_{C_n}^*}{L_j}} \quad (5)$$

PS_{in}	Path-Size (Mass für die Ähnlichkeit)
l_a	Länge der Strecke a von Route i
L_i	Länge der Route i
Γ_i	Gesamtheit der Strecken von Route i
δ_{aj}	Indikator mit Wert 1, falls die Strecke Teil von Route j ist, sonst 0
$L_{C_n}^*$	Länge der kürzesten Route für das OD-Paar n
L_j	Länge der Route j
C_n	Gesamtheit der Routen für das OD-Paar n

Beim Path-Size Faktor wird jede Route mit allen anderen Alternativen verglichen, welche das gleiche OD-Paar verbinden und mindestens eine gemeinsame Kante haben. Erwartet werden Werte zwischen 0 und 1, wobei 0 vollständige Überlappung und 1 keine Überlappung bzw. Eigenständigkeit bedeutet.

Wie in Abschnitt 2.5.2 erwähnt, wird der Path-Size Faktor benutzt, um den systematischen Teil der Nutzenfunktion zu ergänzen. Die Wahrscheinlichkeit Route i zu wählen, ist dann wie folgt gegeben:

$$P_i = \frac{e^{V_i + \ln PS_{in}}}{\sum_j^n e^{V_j + \ln PS_{in}}} \quad (6)$$

BioRoute berechnet den Ähnlichkeitswert für alle gewählten Routen und deren Alternativen. Zusätzlich werden die vorhandenen Attribute des Strassennetzes als Routeneigenschaft hinzugefügt. Beim Ähnlichkeitsmass sollte beachtet werden, dass der Logarithmus von PS ausgegeben wird. Der Grund ist, dass üblicherweise beim *Path-Size-Logit-Modell* die Nutzenfunktion des MNL-Modells mit dem Logarithmus von PS ergänzt wird (vgl. (6)).

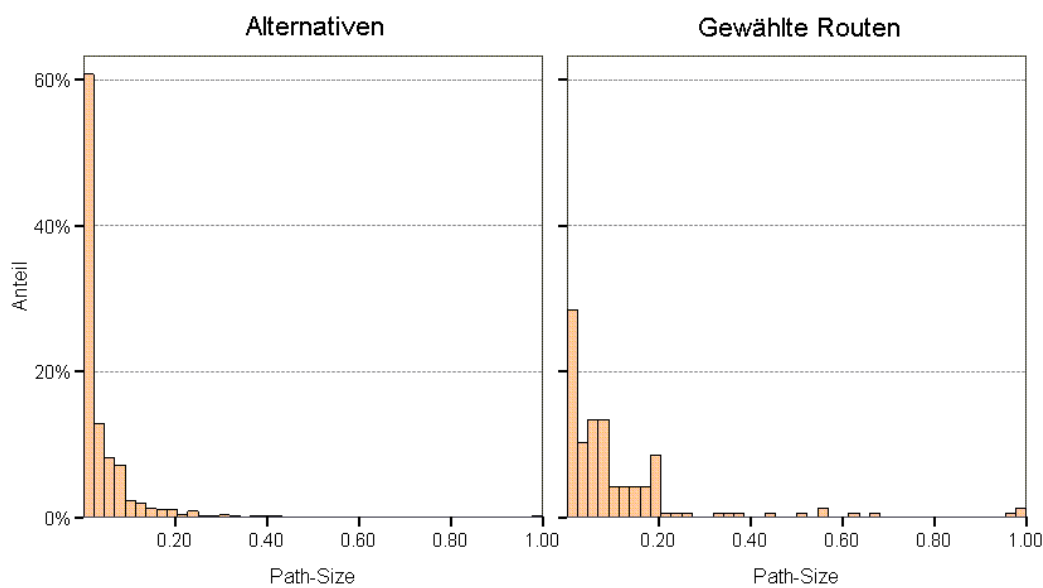
Weiter werden von BioRoute die Modellspezifikationsdatei für BIOGEME, welche noch manuell bearbeitet werden muss, sowie eine Log-Datei für eine mögliche Fehlersuche ausgegeben.

6.2 Ergebnisse

Die Verteilung der berechneten Ähnlichkeiten ist in den Abbildungen 28 und 29 dargestellt. Aufgrund der unterschiedlichen Anzahl Alternativen ($n = 165 \cdot 20 = 3300$) verglichen mit den gewählten Routen ($n = 165$), wurden prozentuale Vergleiche gebildet.

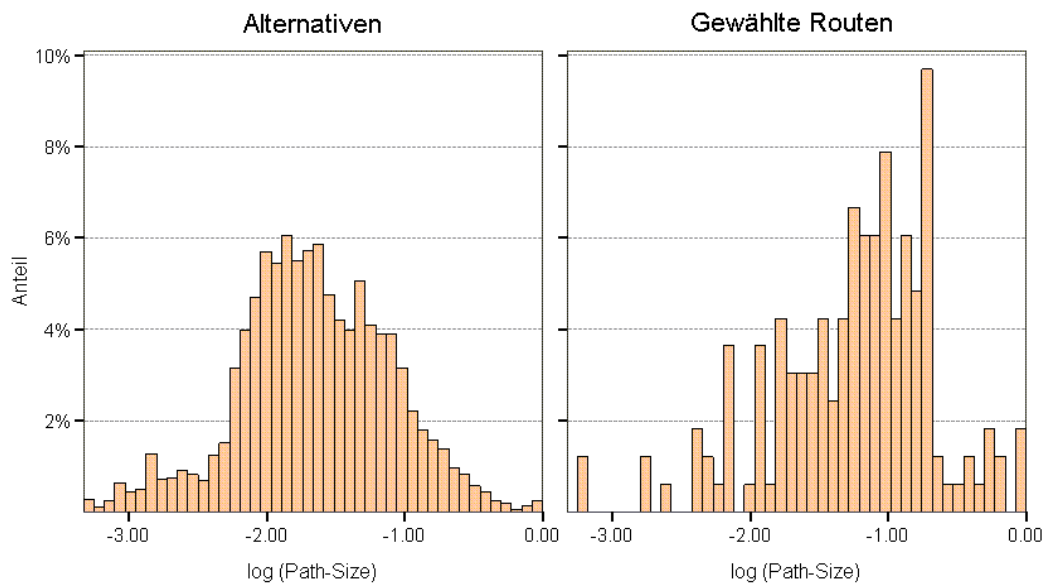
Auf den ersten Blick ist ersichtlich, dass die Alternativen grössere Ähnlichkeiten untereinander aufweisen als die gewählten Routen. Lediglich ein sehr geringer Prozentanteil ist hier eigenständig ($Path-Size = 1$). Die gewählten Routen weisen, prozentual gesehen, mehr eigenständige Routen auf.

Abbildung 28 Path-Size Faktor



Die erwähnten Feststellungen sind etwas besser in Abbildung 29 mit dem natürlichen Logarithmus des Path-Size Wertes ersichtlich. Je negativer ein Wert, desto grösser ist die Ähnlichkeit. Werte von 0 bedeuten hier vollständige Eigenständigkeit.

Abbildung 29 Logarithmus des Path-Size Faktors



Aufgrund dieser ersten Erkenntnisse ist es interessant zu ermitteln, ob die Ähnlichkeit bzw. Unabhängigkeit einer Route von der Routenlänge abhängig ist. Zu diesem Zweck wurden für den gesamten Schätzdatensatz (gewählte Routen und Alternativen) Boxplots für unterschiedliche Routenlängen gebildet (vgl. Abb. 30 und 31). Es zeigt sich die Tendenz, dass kürzere Routen geringere Ähnlichkeiten aufweisen. Dies widerlegt die in Weis (2006) gefundene Eigenschaft, wonach für kürzere Routen eher ähnliche Alternativen existieren. Die Untersuchung in Weis (2006) basiert jedoch auf Routen des öffentlichen Verkehrs.

Abbildung 30 Path-Size für unterschiedliche Routenlängen

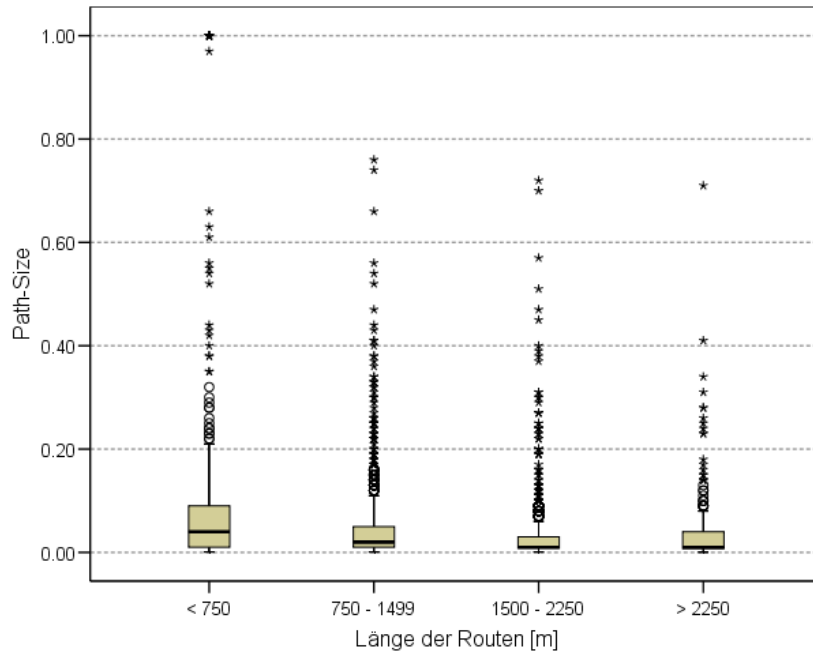
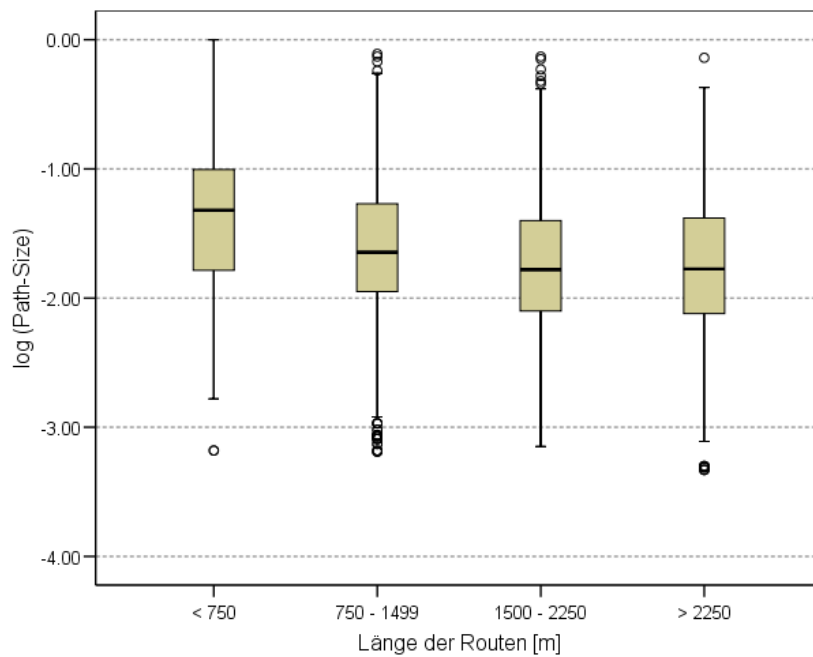


Abbildung 31 Log(Path-Size) für unterschiedliche Routenlängen



6.3 Diskussion

Die gefundene Eigenschaft, dass kurze Routen geringere Ähnlichkeitswerte aufweisen, kann mit der geringen Routenlänge bzw. der geringen Anzahl Kanten teilweise erklärt werden. Kurze Routen weisen weniger Kanten auf, weshalb die Überlappungsmöglichkeiten reduziert werden. Folglich ergeben sich für solche Routen kleinere Ähnlichkeitswerte. Beim hier vorliegenden Schätzdatensatz ist diese Problematik besonders akzentuiert, da die Routen zum Teil sehr wenige Kanten besitzen.

Die Qualität des Routensatzes widerspiegelt sich also nicht nur bei späteren Modellschätzungen, sondern auch bei der Berechnung gewisser Routeneigenschaften. Dies will aber nicht heißen, dass bei qualitativ besseren Routen (hier längere Routen) andere Ergebnisse erzielt würden. Es soll lediglich das Zustandekommen bestimmter Eigenschaften kritisch beleuchten.

Mit Hilfe des hier berechneten PS-Faktors kann ein gewichtiger Nachteil, wonach die Alternativen als unabhängig voneinander gelten, von MNL Modellen teilweise behoben werden. Hierzu wird das MNL Modell mit dem berechneten PS-Faktor ergänzt. Das resultierende PS-Logit Modell gehört, wie in Abschnitt 2.5.2 erläutert, zur dritten Gruppe von Modellen, welche die Ähnlichkeit berücksichtigen. Ein interessanter Vergleich wäre die Berechnung der C-Logit Korrektur, welche ebenfalls die Nutzenfunktion durch einen Ähnlichkeitsfaktor ergänzt. Aus Zeitgründen musste aber darauf verzichtet werden.

7 Berechnung weiterer Routeneigenschaften

Damit bei der Modellschätzung möglichst viele Routeneigenschaften berücksichtigt werden können, wurden die im Schätzdatensatz vorhandenen Routen um verschiedene Eigenschaften ergänzt. Die einzelnen Berechnungen erfolgten mittels ArcGIS, MATSim und SPSS.

Einige Eigenschaften, wie beispielsweise die Ähnlichkeit oder Länge der Routen, sind bereits vorhanden und wurden dementsprechend so belassen. Andere Attribute, welche von BioRoute mittels des Strassennetzes kalkuliert wurden, mussten hingegen neu berechnet oder angepasst werden. BioRoute lässt die Länge der einzelnen Strassenkanten bei der Berechnung der Routeneigenschaften ausser Acht. Dies ist aber ziemlich wichtig und wurde nachträglich getan. Als Beispiel sei die Berechnung der durchschnittlichen Steigung einer Route illustriert. Diese wurde wie folgt berechnet:

$$S_i = \sum_{j \in \Gamma_i} \frac{L_j \cdot S_j}{L_i} \quad (7)$$

- S_i Durchschnittliche Steigung der Route i
- S_j Durchschnittliche Steigung der Strecke j
- L_i Länge der Route i
- L_j Länge der Strecke j
- Γ_i Gesamtheit der Strecken von Route i

Analog zu dieser einfachen Gewichtungsfunktion wurden die anderen Routeneigenschaften berechnet (vgl. Tabelle 13).

Eine etwas spezielle Stellung nehmen die Kategorien des durchschnittlichen Tagesverkehrs (DTV) ein. Hier wurden zunächst für jede Route die jeweiligen Kategorienanteile kalkuliert. Basierend auf diesen Anteilen erfolgte dann eine Dummy-Kodierung. Der höchste Kategorienanteil erhält dabei den Wert 1, alle anderen den Wert 0. Nachfolgende Tabellen illustrieren diese Sachlage anhand von Beispielwerten etwas genauer.

Tabelle 11 Anteil DTV-Kategorien für Route i

Routen	Anteil DtvKat1	Anteil DtvKat2	Anteil DtvKat3	Anteil DtvKat4
Route _{i}	0.2	0.1	0.6	0.1

Tabelle 12 Dummy-Kodierung der DTV-Kategorien von Route i

Routen	Dummy DtvKat1	Dummy DtvKat2	Dummy DtvKat3	Dummy DtvKat4
Route _{i}	0	0	1	0

Bei der Schätzung der Modelle gilt es zu berücksichtigen, dass von n Kategorien lediglich $n-1$ verwendet werden. Die nicht verwendete Kategorie dient dabei als Referenzkategorie. Mit dieser werden dann die Parameter der anderen Kategorien verglichen.

Nach Beendigung aller Berechnungsschritte weist der gesamte Schätzdatensatz folgende Eigenschaften resp. Variablen auf (vgl. Tabelle 13):

Tabelle 13 Variablen des Schätzdatensatzes

Variable	Beschreibung
Length	Routenlänge [m]
RiseAv	Durchschnittliche Steigung der Route [m/100m]
RiseMax	Maximale Steigung der Route [m/100m]
FallAv	Durchschnittliches Gefälle der Route [m/100m]
FallMax	Maximales Gefälle der Route [m/100m]
Uphill	Kumulierte Höhenmeter aufwärts [m]
Downhill	Kumulierte Höhenmeter abwärts [m]
DtvMax	Maximale DTV-Kategorie der Route [1 - 4]
DtvKat1	Dummy für DTV-Kategorie 1 [1/0]
DtvKat2	Dummy für DTV-Kategorie 2 [1/0]
DtvKat3	Dummy für DTV-Kategorie 3 [1/0]
DtvKat4	Dummy für DTV-Kategorie 4 [1/0]
CovAv	Anteil Naturbelag an der Route [0 - 1]
BikeAv	Anteil markierter Fahrradwege an der Route [0 - 1]
BridgeAv	Anteil Brücken an der Route [0 - 1]
TunnelAv	Anteil Tunnels an der Route [0 - 1]
TLights	Anzahl Lichtsignalanlagen [\mathbb{R}^+]
lnPS	Logarithmus der Path-Size Grösse [\mathbb{R}^-]

8 Deskriptive und räumliche Analyse

Bevor zur Schätzung der Entscheidungsmodelle übergegangen wird, werden in diesem Kapitel Ergebnisse der beschreibenden Statistik des Schätzdatensatzes vorgestellt. Hierbei wurde zwischen den gewählten Routen und jeweiligen Alternativen unterschieden. Wie bei der Darstellung der Ähnlichkeit wurden prozentuale Vergleiche wegen unterschiedlicher Datengrösse (Alternativen: $n = 3300$; gewählte Routen: $n = 165$) gemacht.

Um die Ergebnisse der einzelnen Analysen besser einordnen zu können, werden in einem separaten ersten Kapitel einige in diesem Zusammenhang sinnvolle Eigenschaften des gesamten Strassennetzes aufgelistet. Entsprechende Zahlen können für die späteren Analysen als ungefähren Erwartungswert betrachtet werden. Des Weiteren wird die räumliche Verteilung der gewählten Routen in der Stadt Zürich kurz vorgestellt.

Die in diesem Kapitel vorgestellte Untersuchung ist insofern wichtig, da für die spätere Modellschätzung erste Erkenntnisse gezogen werden können. Diese können bei der Definition der Nutzenfunktion einfließen.

8.1 Eigenschaften des gesamten Strassennetzes

Bei dieser Berechnung gilt es zu beachten, dass jede Kante doppelt vorhanden ist, damit beide Fahrrichtungen ermöglicht werden. Somit können die Steigung und das Gefälle nicht unterschieden werden. Stellvertretend für diese Werte steht die Neigung. Einige Grössen, wie zum Beispiel die Länge des gesamten Netzes, die totale Anzahl Lichtsignalanlagen etc. wurden bewusst weggelassen, da sie für diese Untersuchung keinen Sinn machen. Tabelle 14 illustriert die einzelnen Ergebnisse.

Tabelle 14 Eigenschaften des gesamten Strassennetzes

Eigenschaft	Wert
Durchschnittliche Neigung	0.02
Maximale Neigung	0.74
Anteil markierter Fahrradwege	0.46
Anteil Brücken	0.02
Anteil Tunnel	0.01
Anteil Naturbelag	0.04
Anteil DTV-Kategorie 1	0.22
Anteil DTV-Kategorie 2	0.18
Anteil DTV-Kategorie 3	0.38
Anteil DTV-Kategorie 4	0.20
Anteil DTV-Kategorie 5	0.02

8.2 Räumliche Verteilung der gewählten Routen

Die räumliche Verteilung der gewählten Routen (in schwarzer Farbe) in der Stadt Zürich ist in den Abbildungen 32 und 33 dargestellt. Dieser Verteilung wurde die Netzdichte pro Kreis (Zürich hat 12 Stadtkreise) resp. die Neigung hinterlegt.

8.2.1 Gewählte Routen verglichen mit der Netzdichte

Die für diese Analyse zugrunde liegende Netzdichte wurde folgendermassen berechnet:

$$\rho_i = \frac{LK_i}{AK_i} \quad i = 1..12 \quad (8)$$

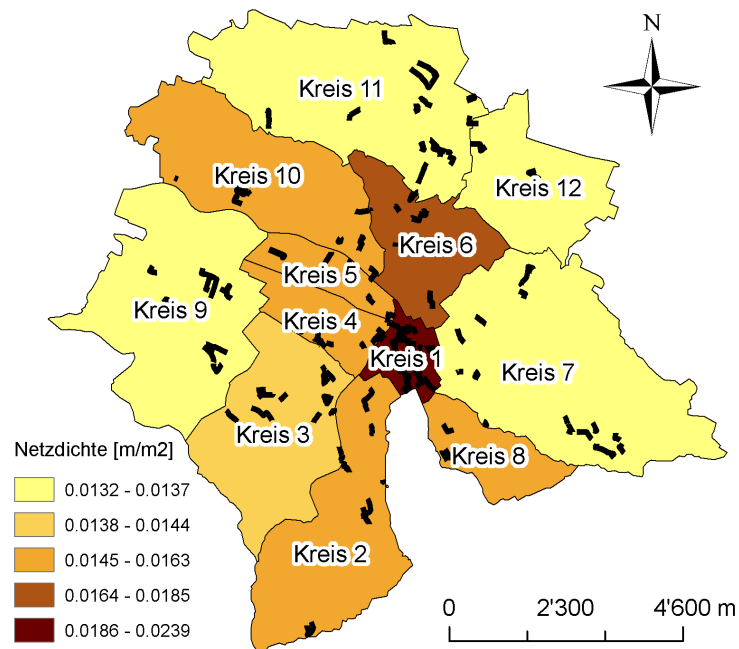
LK_i Länge des Strassennetzes für Kreis i

AK_i Fläche des Kreises i

Die Betrachtung der räumlichen Verteilung der gewählten Routen lässt die Vermutung zu, wonach die Netzdichte eine untergeordnete Rolle zu spielen scheint. Einzig im Kreis 1 scheint

die grössere Netzdichte einen positiven Einfluss zu haben. In diesem Gebiet liegen eine beträchtliche Anzahl der gewählten Routen. In den übrigen Kreisen ist hingegen kein Einfluss betreffend der Netzdichte erkennbar.

Abbildung 32 Gewählte Routen verglichen mit der Netzdichte

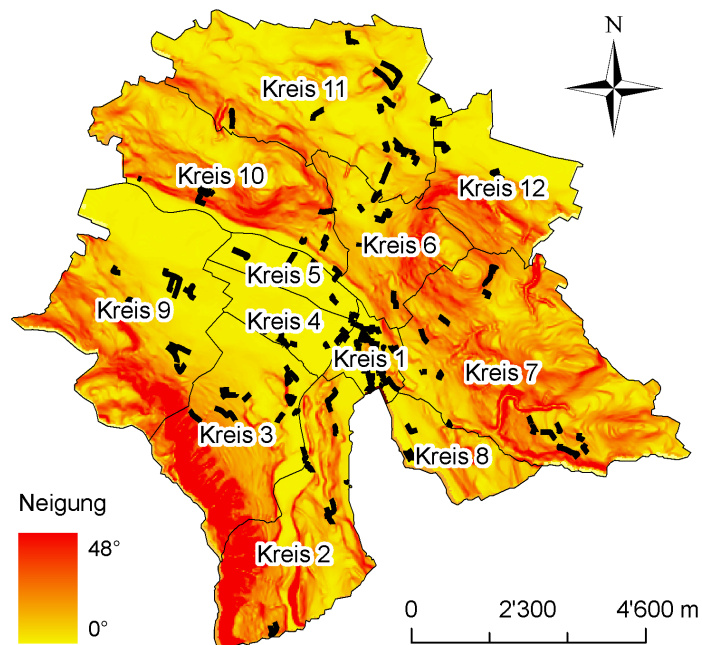


Quelle: Quartiergrenzen 1994/2004/2006, BFS Geostat

8.2.2 Gewählte Routen verglichen mit der Neigung

Die Betrachtung der Neigung führt zum Schluss, dass steile Neigungen offensichtlich gemieden werden. Ob dies auch wirklich zutrifft, wird die Verteilung der Abbildungen 37 und 38 im Abschnitt 8.3 mit der deskriptiven Analyse detailliert zeigen.

Abbildung 33 Gewählte Routen verglichen mit der Neigung

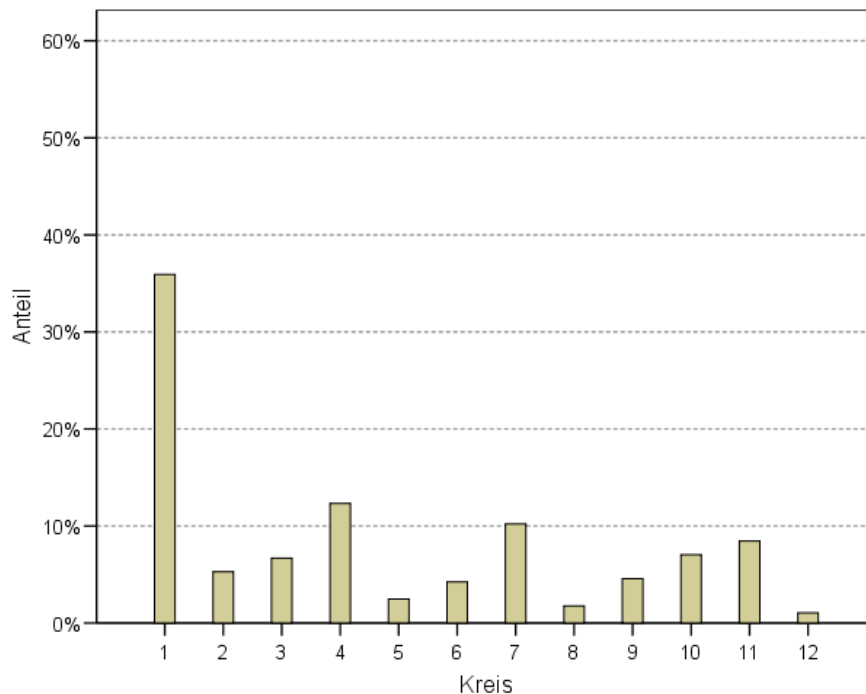


Quellen: DTM AV © 2006 Swisstopo (DV033492.2); Quartiergrenzen 1994/2004/2006, BFS Geostat

8.2.3 Anteil Routen pro Kreis

Der grösste Anteil der gewählten Routen (über 35%) geht durch den Kreis 1 (vgl. Abb. 34). Aufgrund der zentralen Lage dieses Gebietes ist dies keineswegs erstaunlich. Möglicherweise erklärt diese Tatsache, dass das Fahrrad häufig für Fahrten zum Arbeitsplatz oder für Einkaufsmöglichkeiten genutzt wird. Am wenigsten Routen führen durch die Kreise 5, 8 und 12. Die Arbeits- und Einkaufsmöglichkeiten in diesen Gegenden sind deutlich geringer als in Zentrumsnähe.

Abbildung 34 Anteil Routen pro Kreis

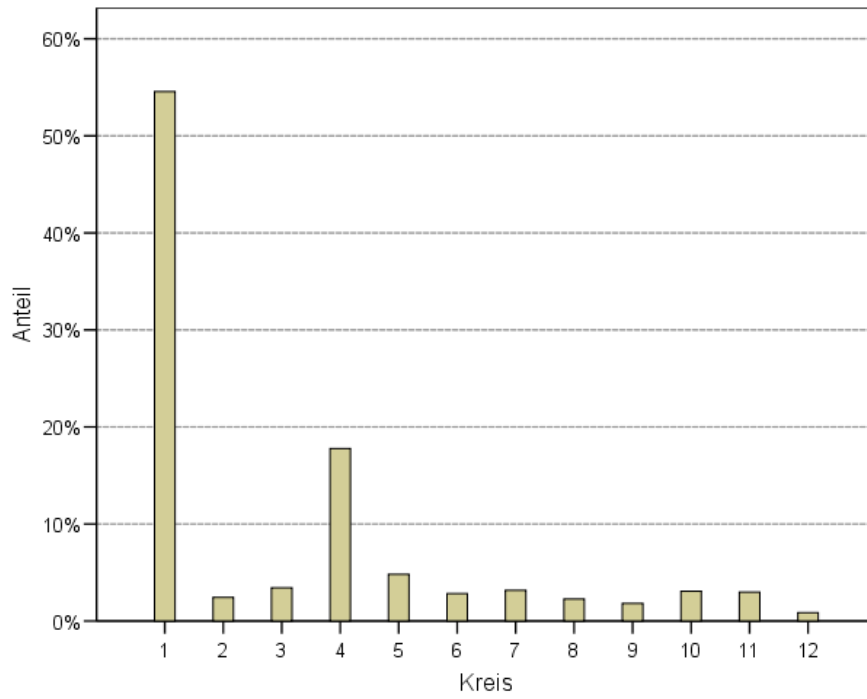


Da die Länge des zur Verfügung stehenden Strassennetzes bezüglich der Routenanteile eine Rolle spielt, wurden die Anteile mit der Netzlänge pro Kreis normiert (vgl. Abb. 35). Somit erhalten Routen, welche beispielsweise durch einen Stadtkreis mit geringerer Netzlänge führen ein grösseres Gewicht. Dies weil das Angebot an Routenmöglichkeiten in solchen Gebieten eingeschränkt ist.

Grosse Änderungen gegenüber der vorherigen Abbildung können nicht festgestellt werden. Die Anteile am Kreis 1 wurden sogar noch verstärkt. Ebenso an Bedeutung gewonnen haben die Routen durch den Kreis 4. Geringere Anteile sind hingegen beim Kreis 7 festzustellen.

Abschliessend kann gesagt werden, dass der grösste Teil der gewählten Routen durch zentrale Gebiete mit vielen Arbeitsplätzen und Einkaufsstätten führt. Zudem sind zentrale Stadtteile auch relativ flach.

Abbildung 35 Normierter Anteil Routen pro Kreis

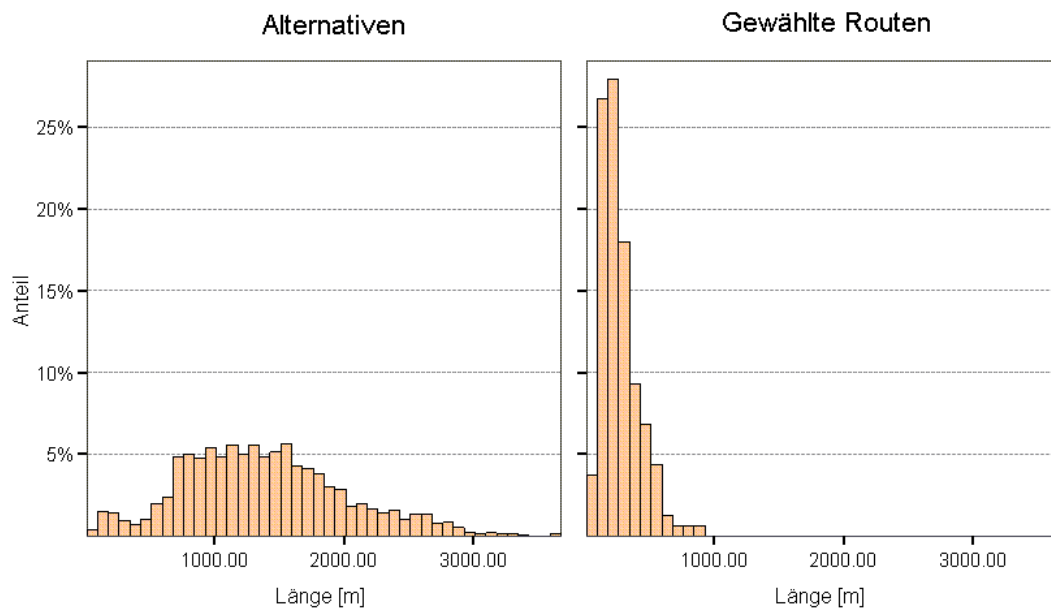


8.3 Deskriptive Analyse des Schätzdatensatzes

8.3.1 Routenlänge

Wie in Kapitel 5.3 bereits erwähnt, sind die gewählten Routen deutlich kürzer als die entsprechenden Alternativen. Die Verteilung der Alternativen zeigt, dass deren Längen gut variieren und somit eine grosse Standardabweichung haben. Folglich sind im Alternativensatz sowohl kurze als auch lange Routen vorhanden.

Abbildung 36 Länge der Alternativen und gewählten Routen

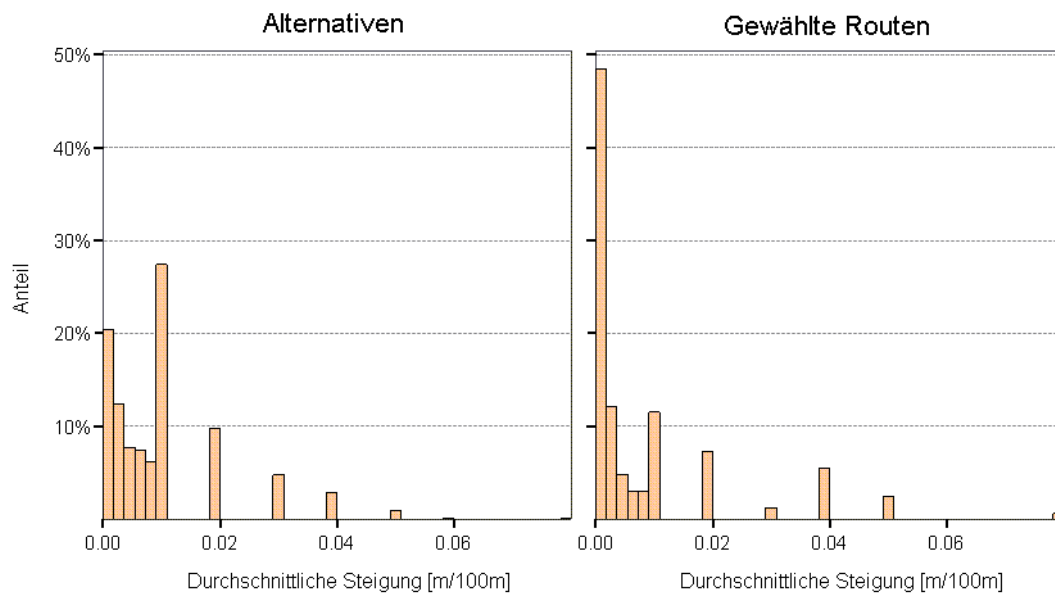


8.3.2 Steigung der Routen

Durchschnittliche Steigung

Sehr gut ersichtlich ist, dass rund 50 % der gewählten Routen durchschnittliche Steigungen von gegen 0 aufweisen. Demgegenüber stehen lediglich rund 20 % der Alternativen. Verglichen mit der durchschnittlichen Steigung des Gesamtnetzes (0.02) kann gedeutet werden, dass die gewählten Routen vor allem über flaches Gelände führen.

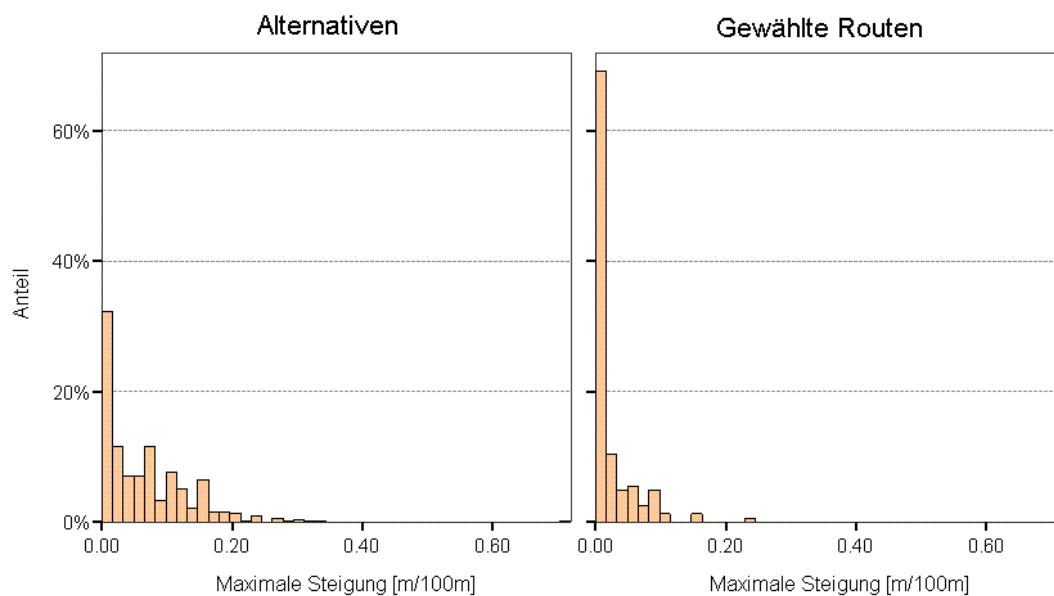
Abbildung 37 Durchschnittliche Steigung



Maximale Steigung

Beinahe 70 % der gewählten Routen weisen maximale Steigungswerte von gegen 0 auf. Der Unterschied zu den Alternativen ist hier wesentlich grösser als beim vorherigen Durchschnittswert. Somit kann geschlossen werden, dass die im vorherigen Abschnitt dargestellten Durchschnittswerte (vgl. Abb. 37) von möglichen Ausreißern in den Daten beeinflusst wurden. Die Betrachtung der maximalen Steigung scheint für die spätere Modellschätzung also geeigneter.

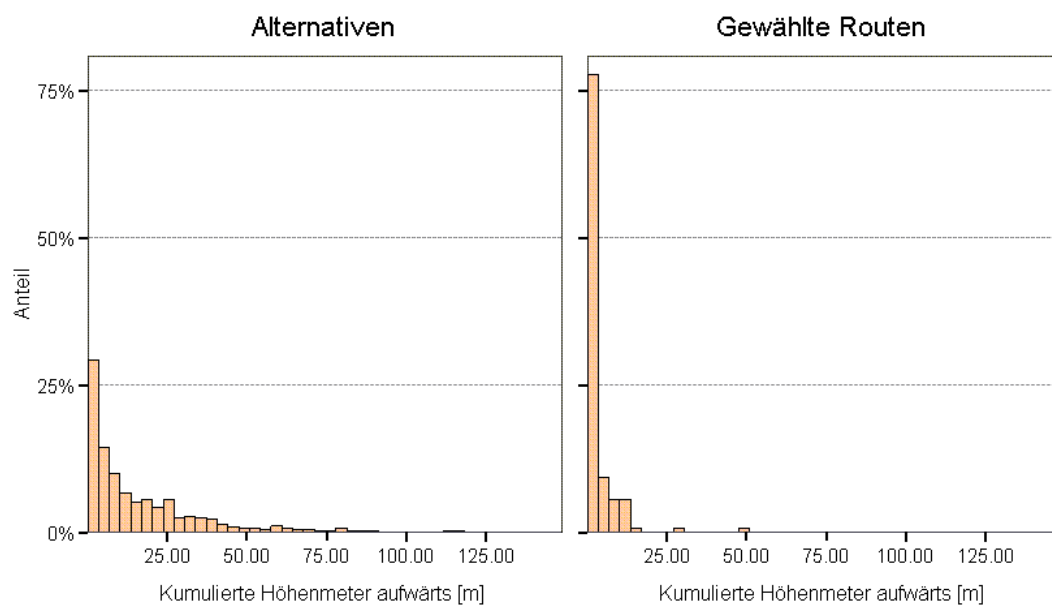
Abbildung 38 Maximale Steigung



Kumulierte Höhenmeter aufwärts

Die an dieser Stelle vorgestellte Analyse zeigt die kumulierten Höhenmeter, die aufwärts zurückgelegt wurden. Deutlich erkennbar ist, dass die gewählten Routen weniger Höhenmeter überwinden als die Alternativen. Selbstverständlich korreliert diese Grösse mit der Steigung und Länge der Routen.

Abbildung 39 Kumulierte Höhenmeter aufwärts

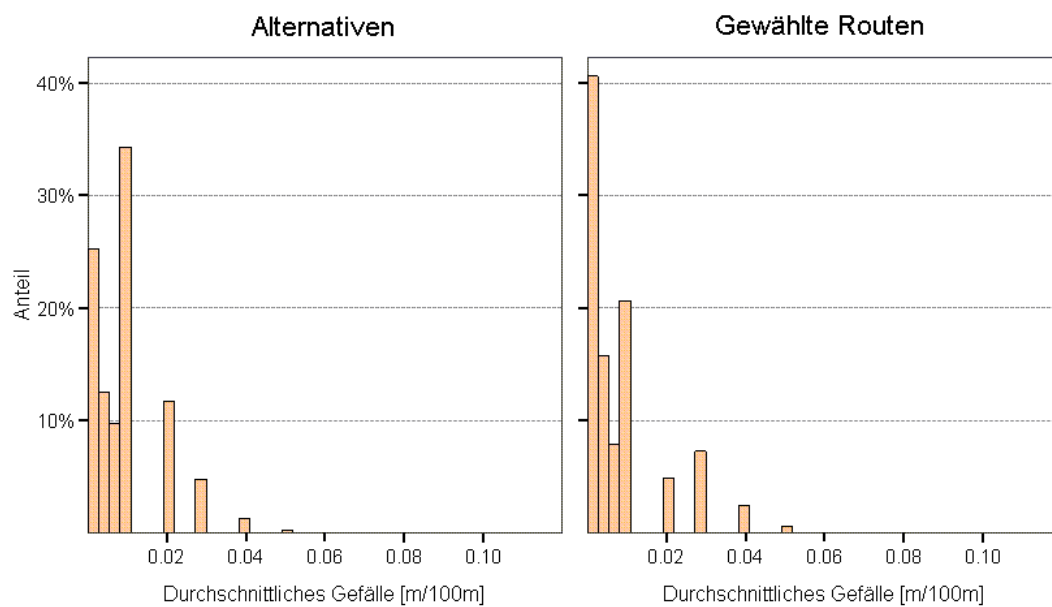


8.3.3 Gefälle der Routen

Durchschnittliches Gefälle

Die Verteilung des Gefälles lässt im Gegensatz zur Steigung keine solch deutlichen Schlüsse zu. Tendenziell weisen die gewählten Routen kleinere Gefälle auf als die Alternativen. Prozentual betrachtet gibt es bei den gewählten Routen aber auch mehr Routen mit grösseren Gefällen (vgl. Balken bei 0.04).

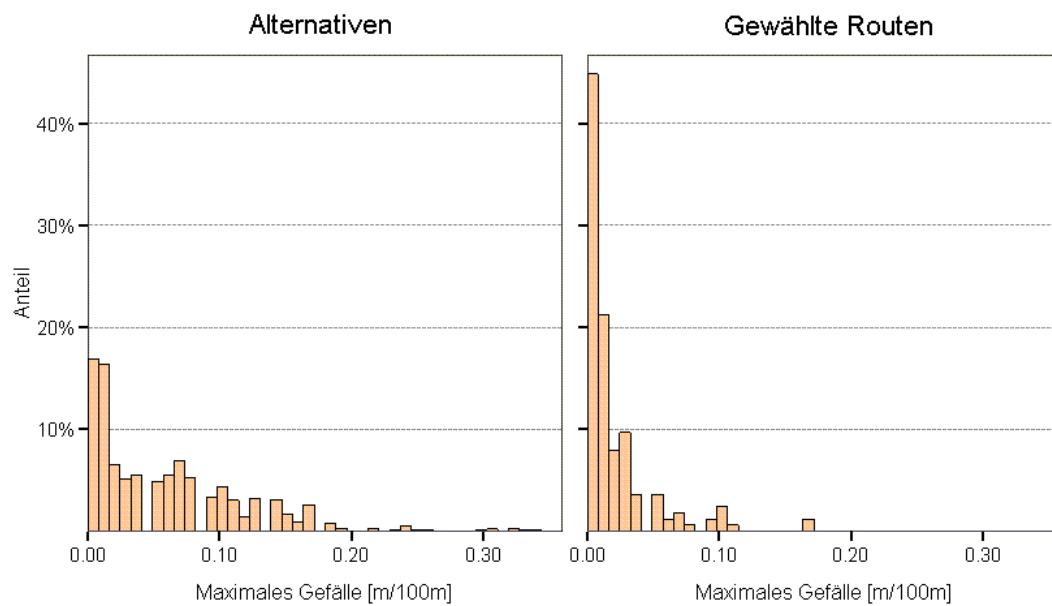
Abbildung 40 Durchschnittliches Gefälle



Maximales Gefälle

Unterschiede in der Verteilung sind beim maximalen Gefälle, analog zur maximalen Steigung, deutlich sichtbar. Die gewählten Routen führen klar über kleinere maximale Gefälle als die Alternativen.

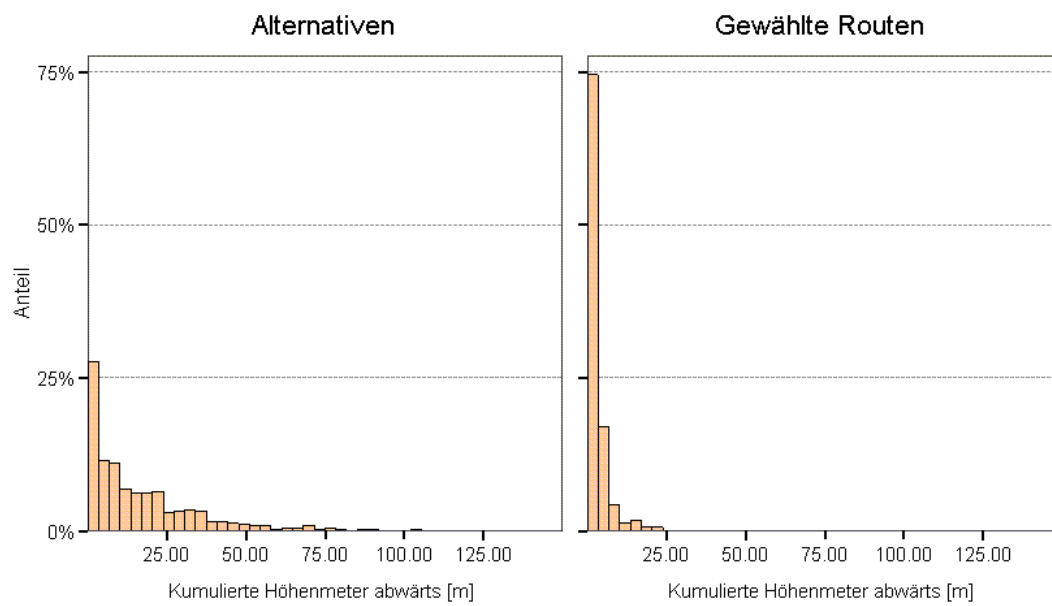
Abbildung 41 Maximales Gefälle



Kumulierte Höhenmeter abwärts

Aufgrund der Korrelation mit der Routenlänge und dem Gefälle weisen die untenstehenden Verteilungen keine Überraschungen auf. Die Alternativen bewältigen wie erwartet mehr Höhenmeter abwärts.

Abbildung 42 Kumulierte Höhenmeter abwärts

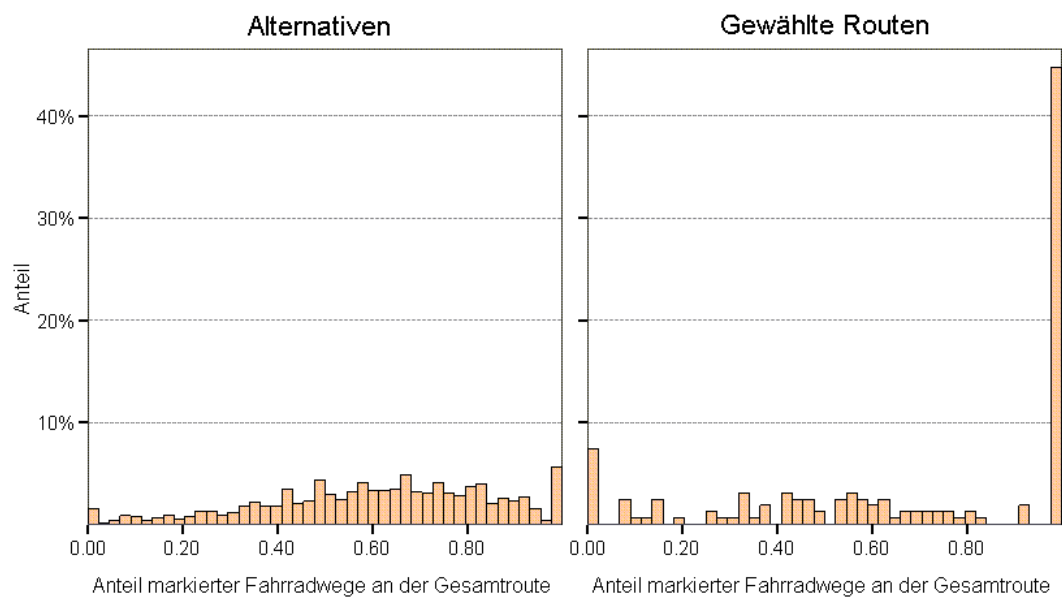


8.3.4 Anteil markierter Fahrradwege

Auffallend bei dieser Abbildung ist der hohe Anteil bei 1 für die gewählten Routen. Beinahe 50 % der gewählten Routen führen vollständig oder annähernd vollständig über einen markierten Fahrradweg. Hierbei soll die bereits erwähnte Tatsache wiederholt werden, wonach mit markiertem Fahrradweg sowohl Fahrradspuren auf einer normalen Strasse als auch separate Fahrradwege gemeint sind.

Verglichen mit dem Erwartungswert (Durchschnitt Gesamtnetz) von 0.46 weist der grösste Teil der gewählten Routen grössere Anteile markierter Fahrradwege auf. Bei den Alternativen liegt hingegen etwa die Hälfte der Routen unter diesem Wert. Nichtsdestotrotz sollte beachtet werden, dass auch ungefähr 8 % der gewählten Routen keinen markierten Fahrradweg benutzen.

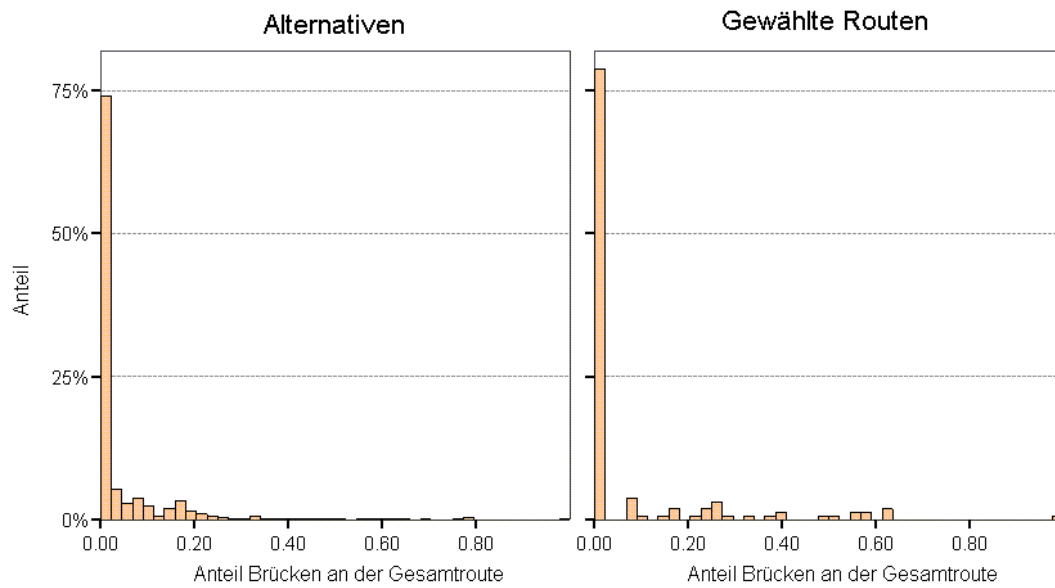
Abbildung 43 Anteil markierter Fahrradwege an der Gesamtroute



8.3.5 Anteil Brücken

Der Anteil Brücken an den jeweiligen Routen lässt, wie in Abbildung 44 ersichtlich, keinen bedeutsamen Schluss zu.

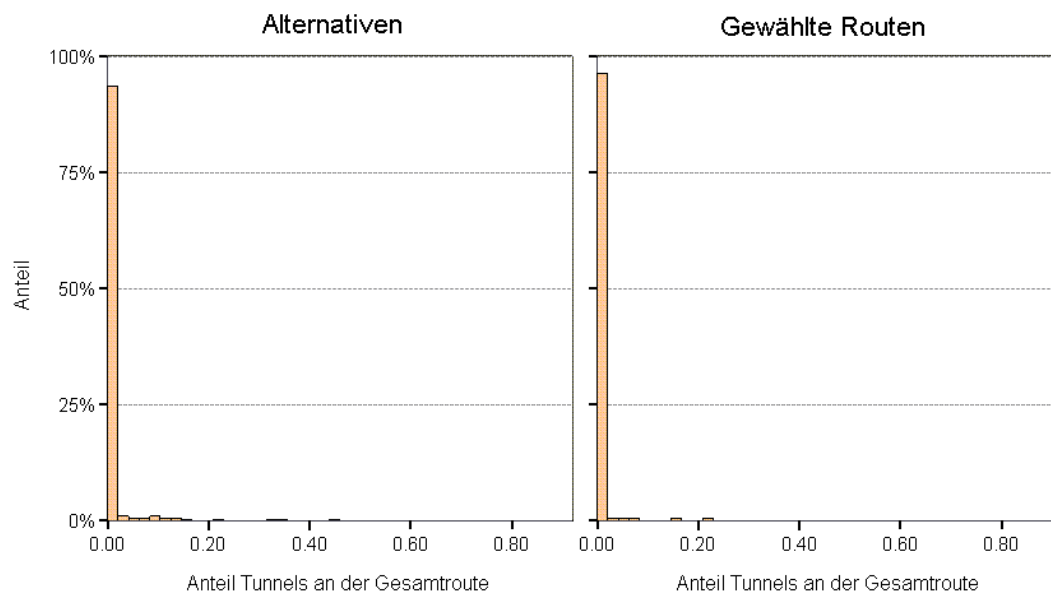
Abbildung 44 Anteil Brücken an der Gesamtroute



8.3.6 Anteil Tunnels

Der Anteil an Tunnels illustriert, dass auch die Tunnels beim hier verwendeten Schätzdatensatz eine belanglose Rolle zu spielen scheinen.

Abbildung 45 Anteil Tunnels an der Gesamtroute

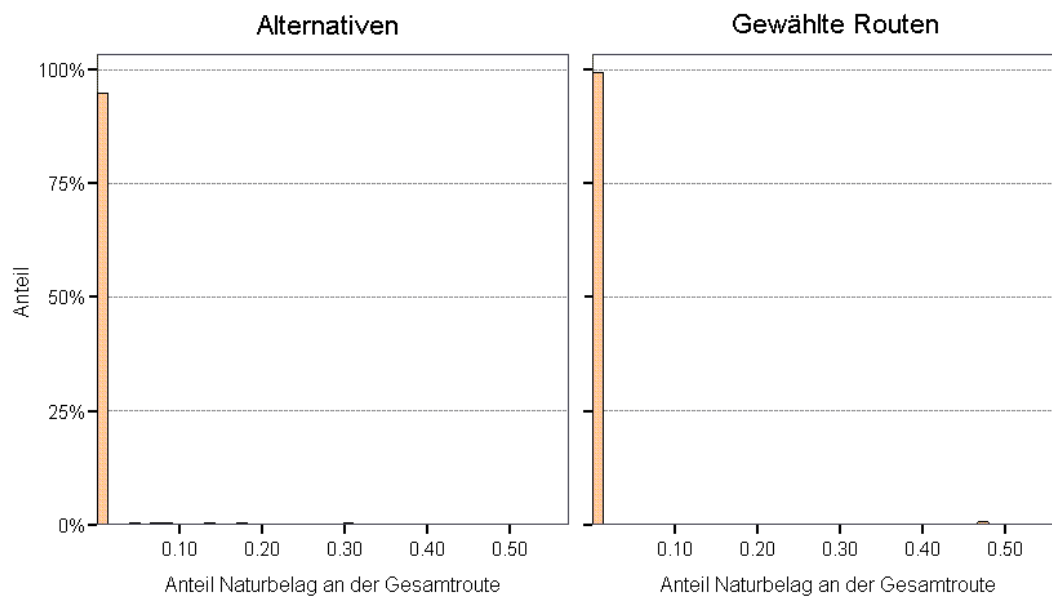


Sowohl bei den Tunnels als auch bei den Brücken könnte anstatt des Anteils auch die entsprechende Anzahl pro Route ermittelt werden. Diese wurde aber bewusst nicht dargestellt, da die Korrelation mit der Routenlänge zu grosse Ausmasse annimmt.

8.3.7 Anteil Naturbelag

Analog zu den beiden vorherigen Verteilungen lässt sich auch über den Belag der Strasse keine bedeutsame Aussage machen. Die gewählten Routen führen zu praktisch 100 % über asphaltierte Strassen. Der Durchschnittswert für das Gesamtnetz liegt bei 0.04. Somit kann lediglich eine geringe Tendenz festgestellt werden, wonach die gewählten Routen Naturbelag meiden.

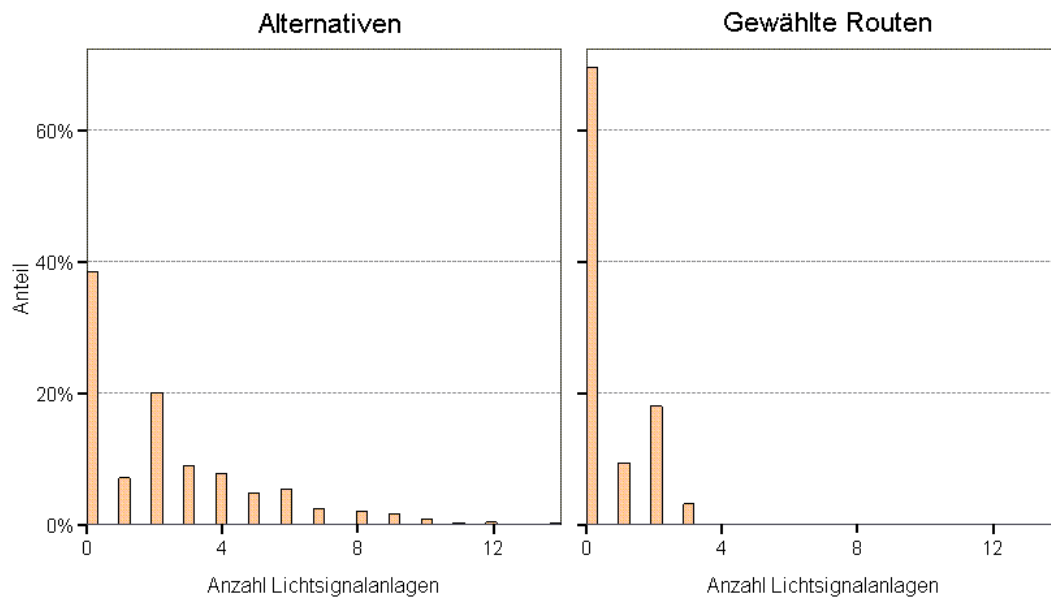
Abbildung 46 Anteil Naturbelag an der Gesamtroute



8.3.8 Anzahl Lichtsignalanlagen

Die Verteilung der Anzahl Lichtsignalanlagen pro Route illustriert sehr schön, dass die gewählten Routen über deutlich weniger Lichtsignalanlagen führen als die Alternativen. Über 60 % der gewählten Routen kreuzen gar keine Lichtsignalanlagen. Mehr als drei Lichtsignalanlagen kommen bei den gewählten Routen nicht vor. Diese Eigenschaft korreliert wahrscheinlich mit der Routenlänge.

Abbildung 47 Anzahl Lichtsignalanlagen



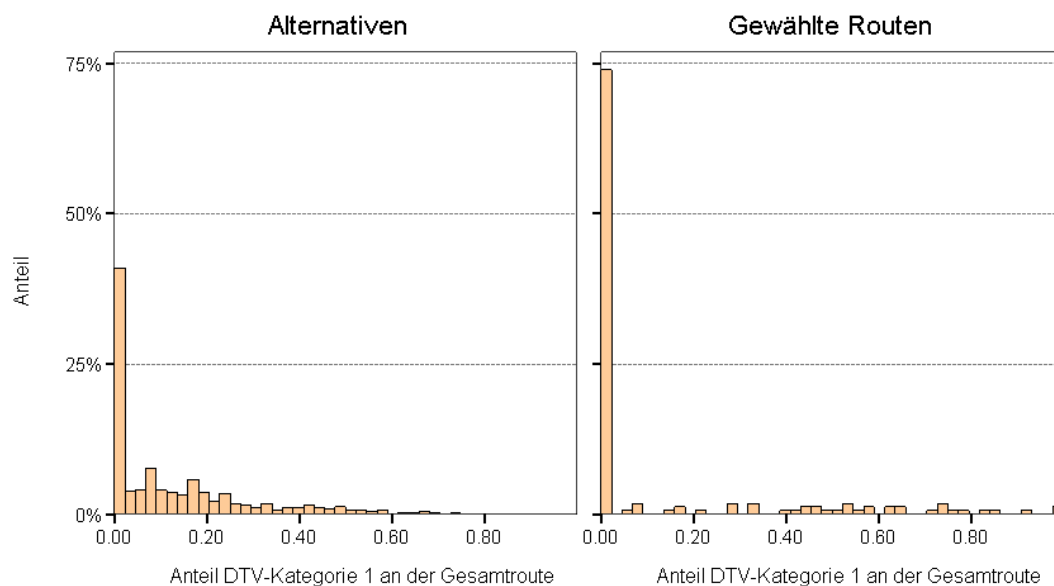
8.3.9 Anteil DTV-Kategorien

Zur Wiederholung sei an dieser Stelle darauf hingewiesen, dass der Verkehr von Kategorie 1 bis 4 zunimmt. Am meisten Verkehr weist also DTV-Kategorie 4 auf.

Anteil DTV-Kategorie 1

Beinahe 75 % der gewählten Routen weisen einen Anteil von 0 für die DTV-Kategorie 1 auf. Verglichen mit dem Durchschnittswert des gesamten Netzes von 0.22 ist dies erstaunlich. Umso bemerkenswerter ist dieses Ergebnis, weil es die Kategorie mit dem geringsten Verkehrsaufkommen darstellt. Dessen ungeachtet kommen bei den gewählten Routen aber auch einige Routen mit Anteilen über 0.75 vor, wo hingegen bei den Alternativen keine solchen Routen vorliegen.

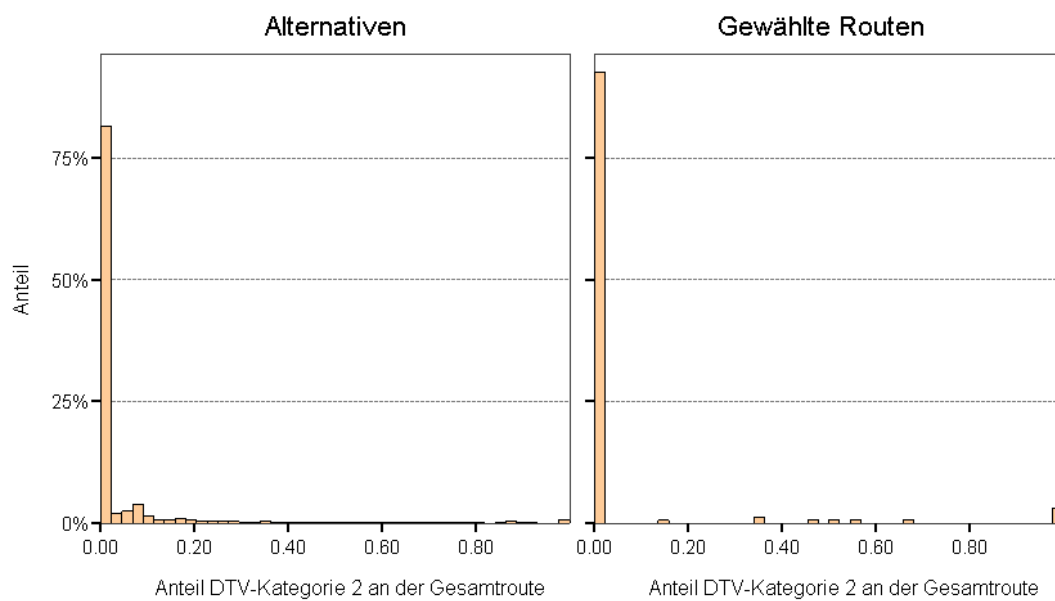
Abbildung 48 Anteil DTV-Kategorie 1



Anteil DTV-Kategorie 2

Der Anteil an DTV-Kategorie 2 zeigt, dass über 80 % der gewählten Routen nicht über Straßen dieser Kategorie führen. Ein geringer Anteil gewählter Routen weist aber auch einen Anteil von 1 auf. Verglichen mit dem Erwartungswert von 0.18 (Gesamtnetz) liegen sowohl die Alternativen als auch die gewählten Routen tendenziell darunter.

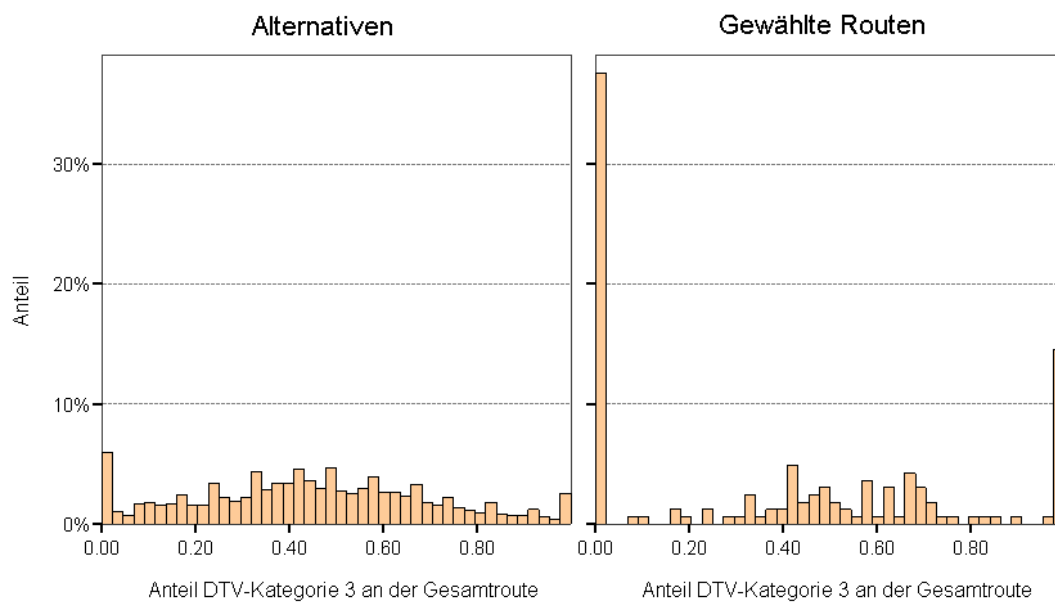
Abbildung 49 Anteil DTV-Kategorie 2



Anteil DTV-Kategorie 3

Auffallend bei diesen Verteilungen sind die grossen Anteile bei den Werten 0 und 1 der gewählten Routen. Die Verteilung der Alternativen ist hingegen etwas symmetrischer.

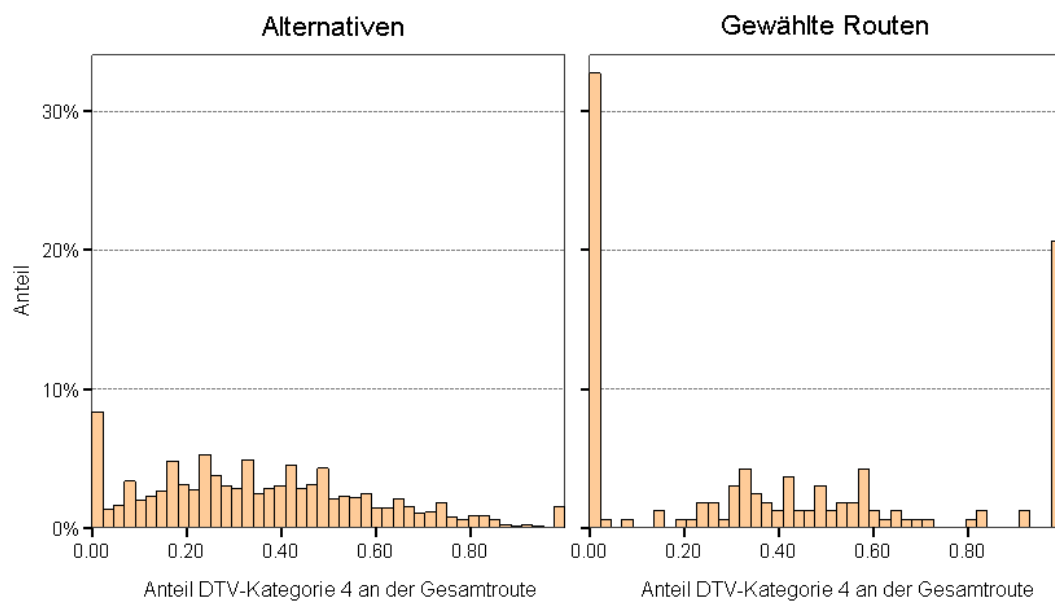
Abbildung 50 Anteil DTV-Kategorie 3



Anteil DTV-Kategorie 4

Eine der DTV-Kategorie 3 sehr ähnliche Verteilung weisen die hier vorgestellten Balkendiagramme auf.

Abbildung 51 Anteil DTV-Kategorie 4



Abschliessend kann der vielleicht etwas gewagte Schluss gezogen werden, wonach die DTV-Kategorien keinen starken Einfluss auf die Routenwahl ausüben. Die Verteilung der einzelnen Kategorien scheint eher zufällig entstanden zu sein. Ob diese Behauptung wirklich standfest ist, wird die Modellschätzung im nächsten Kapitel zeigen.

9 Schätzung der Entscheidungsmodelle

In diesem Kapitel sollen die Modelle, welche im Rahmen dieser Arbeit geschätzt wurden, vorgestellt werden.

Aufgrund der grossen Variablenzahl des Schätzdatensatzes (vgl. Tabelle 13) wird als Erstes eine Korrelationsanalyse durchgeführt. Dabei soll die Korrelation der einzelnen Variablen untersucht werden, um danach eine Auswahl der besten Variablen zu treffen. Für stark korrelierende Grössen soll nur eine Variable bei den späteren Schätzungen berücksichtigt werden.

Nachdem die Variablenzahl mittels der Korrelationsanalyse reduziert ist, werden für jede Variable separate Modelle geschätzt. Die Ergebnisse der einzelnen Modellschätzungen sollen eine weitere Hilfe für die Auswahl der relevantesten Routeneigenschaften bzw. Variablen darstellen.

Erst nach dieser Variablenselektion werden erste Modelle geschätzt, welche die Routenwahl von Fahrradfahrern erklären. Basierend auf einem einfachen linearen Modell mit nur wenigen Parametern in der Nutzenfunktion werden iterativ neue Parameter hinzugefügt und die Nutzenfunktion kontinuierlich erweitert. Weiter sollen aufgrund gewisser Annahmen unterschiedliche Variablentransformationen und Wechselwirkungen getestet werden. Die Auswahl der besten Modelle erfolgt aufgrund der Güte der entsprechenden Schätzung sowie auf Basis der Signifikanz der verschiedenen Parameter.

Für die Berechnung der Modelle wurde das Programm BIOGEME (vgl. Glossar) verwendet. Die Nutzenfunktion und deren Parameter werden in der Modellspezifikationsdatei definiert (vgl. Modellbeispiel der Tabelle 48, Anhang A6). Die verwendeten Variablennamen sind in Tabelle 13 erklärt.

BIOGEME berechnet die Werte der geschätzten Parameter und deren Einfluss auf das Modell. Weiter werden für die Bestimmung der Signifikanz der Parameter die jeweiligen Standardabweichungen und die Resultate des t-Tests ausgegeben. Berechnet wird die Testgrösse t wie folgt:

$$t = \frac{\beta}{\sigma} \tag{9}$$

σ Standardfehler von β

Standardmässig wird ein Parameter als signifikant erachtet, wenn der Betrag von t grösser als 1.96 ist (95%-Fraktilwert der t -Verteilung).

Schliesslich sei noch der von BIOGEME ausgegebene Log-Likelihood Wert erwähnt. Um die Modellgüte zu bestimmen wird die Null-Log-Likelihood und die Schluss-Log-Likelihood Grösse berechnet. Der erste Log-Likelihood Wert nimmt dabei an, dass die Grössen der verschiedenen Parameter keinen Einfluss haben, was also eine zufällige Routenwahl bedeuten würde. Letzterer Wert gibt hingegen die Log-Likelihood unter der Annahme an, dass die Grössen einen Einfluss haben. Je kleiner die Schluss-Log-Likelihood, desto besser ist das geschätzte Modell (grössere Auswahlwahrscheinlichkeiten verringern diesen Wert). Die genaue Funktion der Likelihood Schätzung sowie Erklärungen dazu sind im Glossar zu finden.

9.1 Korrelationsanalyse

Wie erwähnt, wurde als Erstes kreuzweise ermittelt wie die Variablen korrelieren. Dabei wurde mittels des Programms SPSS der Rangkorrelationskoeffizient nach Spearman für jeweilige Variablenpaare berechnet. Dieser nichtparametrische Test setzt, im Gegensatz zu Pearsons Korrelationskoeffizienten, keine Normalverteilung voraus. Des Weiteren können die Variablen auf einer Ordinalskala gemessen worden sein und die Beziehung derselben braucht nicht linear zu sein. Für diese Analyse ebenso geeignet wäre Kendalls Tau Rangkorrelationskoeffizient.

Auf Basis dieses Vergleichs werden die Variablen *Uphill*, *Downhill* und *DtvMax* für die weiteren Berechnungen ausgeschlossen. Die Korrelation mit der Routenlänge und anderen Variablen (*Uphill* korreliert z.B. stark mit *RiseAv* und *Downhill*) nimmt zu grosse Ausmasse an. Andere Routeneigenschaften wie *RiseAv*, *RiseMax*, *FallAv* sowie *FallMax* werden vorläufig beibehalten, obwohl die Korrelation untereinander teilweise sehr gross ist. Der Grund liegt daran, dass stellvertretend für diese vier Variablen diejenige mit der besten Modellgüte ausgewählt werden soll. Dies wird im nächsten Abschnitt beschrieben. Aus der Tabelle 41 im Anhang A3 sind die Korrelationsergebnisse der ausgeschlossenen Variablen ersichtlich.

9.2 Modellschätzungen

9.2.1 Modellschätzung für einzelne Variablen

Die bisherige Variablenselektion soll anhand der Schätzung von Modellen mit jeweils nur einer berücksichtigten Routeneigenschaft verfeinert werden. Ziel dieser Analyse ist es diejenigen Eigenschaften auszuwählen, deren Modelle die besten Ergebnisse erzielen. Die entsprechenden Nutzenfunktionen nehmen alle in folgender Form an:

$$U_i = \beta_{Routeneigenschaft_i} \cdot Routeneigenschaft_i \quad (10)$$

Für das Ähnlichkeitsmass Path-Size werden zwei Transformationen untersucht. Die Variable wird sowohl als logarithmische als auch als Box-Cox-Transformation (vgl. Glossar) getestet. Wie in Kapitel 6.1 erwähnt, ist die Ergänzung der Nutzenfunktion mit dem logarithmisch transformierten Ähnlichkeitsmass die in der Literatur übliche Modellform. Tabelle 15 illustriert die verschiedenen Modellergebnisse. In den Nutzenfunktionen dieser Arbeit steht $T(\dots)$ für eine Box-Cox-Transformation.

Auf die Vorzeichen der verschiedenen Parameter wird bei dieser Analyse keinen Wert gelegt. Vorerst interessiert allein die Modellgüte. Hierzu wird die Null Log-Likelihood mit der Final Log-Likelihood verglichen und der entsprechende ρ^2 -Wert betrachtet. Der Wert des t-Tests gibt schliesslich Auskunft über die Signifikanz des Parameters.

Tabelle 15 Ergebnisse der Modellschätzung mit einzelnen Variablen

Nutzenfunktion U =	ln(L)	ρ^2	t-Test
$\beta_{Length} \cdot Length$	-150.158	0.701	-9.93
$\beta_{PS} \cdot PS$	-475.839	0.053	7.72
$\beta_{\ln(PS)} \cdot \ln(PS)$	-442.560	0.119	10.91
$\beta_{T(PS)} \cdot T(PS)$	-442.535	0.119	3.91
$\beta_{CovAv} \cdot CovAv$	-497.236	0.010	-2.21
$\beta_{BikeAv} \cdot BikeAv$	-489.708	0.025	4.98
$\beta_{BridgeAv} \cdot BridgeAv$	-492.039	0.021	4.72
$\beta_{TunnelAv} \cdot TunnelAv$	-501.649	0.001	-1.11
$\beta_{TLights} \cdot TLights$	-388.093	0.227	-9.51
$\beta_{RiseMax} \cdot RiseMax$	-326.593	0.350	-9.17
$\beta_{RiseAv} \cdot RiseAv$	-492.045	0.021	-4.33
$\beta_{FallMax} \cdot FallMax$	-362.425	0.279	-9.44
$\beta_{FallAv} \cdot FallAv$	-494.999	0.015	-3.82
$\beta_{DtvKat1} \cdot DtvKat1 +$			1.88
$\beta_{DtvKat2} \cdot DtvKat2 +$	-492.903	0.019	1.45
$\beta_{DtvKat3} \cdot DtvKat3$			-2.64

Wie erwartet erreicht der Ansatz mit der Länge die beste Modellgüte. Erstaunlich ist vor allem der hohe Wert von $\rho^2 = 0.7$. Zurückzuführen ist dies auf die kurze Distanz der gewählten Routen. In Kapitel 5.2 wurde bereits angedeutet, dass die Routenlänge eine wichtige Rolle einnimmt.

Weitere gute Ergebnisse erzielen der Reihe nach die maximale Steigung, das maximale Gefälle, die Anzahl Lichtsignalanlagen sowie die transformierten Path-Size Werte. Die Schätzung des Box-Cox-transformierten Ähnlichkeitsmasses erreicht praktisch identische Resultate wie die logarithmische Umwandlung (vgl. Tabelle 15). Das λ bei der Box-Cox-Transformation beträgt -0.0179, weshalb die Funktion annähernd der logarithmischen Funktion gleichkommt (vgl. Abb. 64 im Glossar). Folglich lassen sich die ähnlichen Modellergebnisse gut erklären.

Etwas schlechter schneiden die anderen Modelle ab. Aufgrund der deskriptiven Analyse in Abschnitt 8.3 wären vor allem für den Anteil der markierten Fahrradwege bessere Resultate erwartet worden.

Für die Dummy-kodierten DTV-Kategorien wurde die Kategorie 4 als Referenz gewählt, weshalb sie in der Tabelle 15 nicht aufgelistet ist. Die Parameter der anderen Kategorien werden also mit dieser Referenzkategorie verglichen. Die entsprechende Nutzenfunktion enthält alle drei ersten Kategorien zusammen, da die Wirkung der gesamten Variablen untersucht werden soll.

Wie im vorherigen Kapitel erwähnt, korrelieren die Steigungs- und Gefällegrößen untereinander ziemlich stark. Da die maximale Steigung die besten Schätzergebnisse liefert, soll die besagte Variable stellvertretend für alle anderen topographischen Größen ausgewählt werden.

Ziel der Arbeit ist es, möglichst viele Determinanten für die Routenwahl von Radfahrern auszumachen. Hierbei muss aber eine adäquate Auswahl an Variablen getroffen werden, damit das Modell auch möglichst verallgemeinbar ist. Eine hohe Anzahl an berücksichtigten Variablen erreicht möglicherweise eine höhere Modellgüte, doch nimmt die spezifische Anpassung auf den vorhandenen Schätzdatensatz immer mehr zu. Folglich ist es schwierig die Ergebnisse auf andere Daten zu übertragen. Weiter gilt es, schwierig zu erhebende Eigenschaften und somit „teure“ Variablen von den leicht zu erhebenden Größen zu unterscheiden. Aufgrund dieser Argumentation und der weiter oben beschriebenen Analyse werden für die weiteren Modelle folgende Variablen berücksichtigt: *Length*, *RiseMax*, *TLights*, *ln(PS)*, *BikeAv*.

9.2.2 Lineare Ansätze

Als Erstes werden einfache lineare Modelle geschätzt, deren Nutzenfunktion schrittweise um einen weiteren Parameter erweitert wird. Die Reihenfolge beim Hinzufügen der Parameter stützt sich auf die Ergebnisse des vorherigen Abschnitts. Diesbezügliche Nutzenfunktionen sind nachfolgend aufgelistet:

Tabelle 16 Nutzenfunktionen der linearen Modelle

U =

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax \quad (11)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (12)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights + \beta_{\ln(PS)} \cdot \ln(PS) \quad (13)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (14)$$

Die besten Schätzergebnisse erzielt das letzte Modell (14). Dessen Resultate sind in der Tabelle 17 ersichtlich.

Um den Nutzeneinfluss der einzelnen Variablen ausmachen zu können, werden die Parameterwerte mit dem Mittelwert der entsprechenden Variablen (vgl. Tabelle 38 im Anhang A3) multipliziert. Dies löst die Problematik unterschiedlicher Skalierung und erlaubt den Vergleich einzelner Routeneigenschaften. Die hier verwendeten Mittelwerte basieren auf dem gesamten Schätzdatensatz. Es wird also für eine durchschnittliche Route die Zusammensetzung des Nutzens kalkuliert. In Abschnitt 9.3 ist eine detaillierte Analyse über verschiedene Einflüsse abgehandelt.

Tabelle 17 Schätzergebnisse für das Modell (14)

Parameter	Wert	Wert · Mittelwert ¹¹	t-Test
β_{Length}	-0.006	-8.121	-7.26
$\beta_{RiseMax}$	-17.700	-1.115	-2.03
$\beta_{TLights}$	-0.385	-0.823	-1.55
$\beta_{\ln(PS)}$	0.754	-1.227	2.23
β_{BikeAv}	0.864	0.533	1.36
Log-Likelihood: $\ln(L) = -142.984$, $\rho^2 = 0.715$			

¹¹ Mittelwert gesamter Schätzdatensatz

Wie erwartet fließen die Länge der Routen, die maximale Steigung, die Anzahl Lichtsignalanlagen sowie der Ähnlichkeitsfaktor negativ in die Nutzenfunktion ein. Auch erwartet wurde, dass der Anteil markierter Fahrradwege einen positiven Einfluss hat.

Es gilt zu beachten, dass die beiden Parameter β_{BikeAv} und $\beta_{TLights}$ eigentlich nicht signifikant sind. Da sie aber für die Fragestellung dieser Arbeit relevant sind und zusätzlich die Modellgüte verbessern, werden sie trotzdem beibehalten. Der jeweilige t-Test Wert weicht zudem nicht erheblich vom 95%-Fraktilwert (1.96) ab.

Dieses Modell (14) wird demzufolge als Grundlage für die weiteren Schätzungen dienen.

Einbau weiterer Transformationen

Als nächstes wird versucht die Modellschätzung durch den Einbau von Variablentransformationen zu verbessern. In einem ersten Ansatz wird davon ausgegangen, dass der Einfluss der Steigung nicht linear ist. Für dieses Modell wird postuliert, dass mit zunehmender Steigung deren Einfluss abnimmt. Folglich wird die besagte Eigenschaft als Box-Cox-transformierte Variable in die Nutzenfunktion einfließen. Dies weil die Funktion bzw. Kurve dieser Transformation mit zunehmenden Werten flacher wird (vgl. Abb. 64). Es wurden zwei verschiedene Nutzenfunktionen (15), (16) aufgestellt, die in der Tabelle 18 ersichtlich sind.

Die Güte dieser beiden Modelle wurde gegenüber Modell (14) nicht verbessert. Im Gegenteil, sie wurde beinahe bei allen Schätzungen leicht schlechter. Dabei erzielt die Schätzung des Modells (16) die besten Resultate (vgl. Tabelle 19). Mit einer Log-Likelihood von -142.728 erzielt dieses Modell nahezu identische Resultate wie das ursprüngliche Modell (14). Allerdings ist bei diesem Ansatz die Signifikanz der geschätzten Parameter deutlich schlechter.

Analoge Überlegungen wurden für die Länge der Routen gemacht und entsprechende Variable transformiert. Die Nutzenfunktionen (17) und (18) dieser Transformation sind ebenfalls in Tabelle 18 zu finden. Auch diese Transformation bringt, wie die Ergebnisse zeigen, keine Verbesserung der Schätzung.

Eine Transformation der beiden Variablen $TLights$ und $BikeAv$ macht keinen Sinn, weil diese Eigenschaften auch Werte von 0 aufweisen. Die Transformationsfunktion ist für Werte von 0 nicht definiert, da die Kurve ins Unendliche führt (vgl. Abb. 64 im Glossar).

Tabelle 18 Nutzenfunktionen der linearen Modelle mit Transformationen

U =

$$\beta_{Length} \cdot Length + T(RiseMax) + \beta_{Tlights} \cdot Tlights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (15)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot T(RiseMax) + \beta_{Tlights} \cdot Tlights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (16)$$

$$T(Length) + \beta_{RiseMax} \cdot RiseMax + \beta_{Tlights} \cdot Tlights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (17)$$

$$\beta_{Length} \cdot T(Length) + \beta_{RiseMax} \cdot RiseMax + \beta_{Tlights} \cdot Tlights + \beta_{\log PS} \cdot \log PS + \beta_{BikeAv} \cdot BikeAv \quad (18)$$

Tabelle 19 Schätzergebnisse für die Modelle (15), (16), (17) und (18)

Modell	Final L	ρ^2
(15)	-145.706	0.710
(16)	-142.728	0.716
(17)	-338.691	0.322
(18)	-281.673	0.439

9.2.3 Nichtlineare Ansätze

Zwecks weiterer Verbesserung der Modellgüte werden die Routeneigenschaften nun als nicht-lineare Terme in die Nutzenfunktion eingefügt. Es wird wiederum das Modell (14) als Basis verwendet und durch den Einbau von Exponenten erweitert. Dabei werden einerseits Nutzenfunktionen aufgestellt, bei denen jeweils nur eine Variable einen Exponenten erhält und andererseits wird ein Modell getestet, welches alle Variablen exponentiell erweitert. Die logarithmisch transformierte Ähnlichkeitsvariable $\ln(PS)$ wird nicht mit einem Exponenten ergänzt. Stattdessen wird die ursprüngliche Grösse PS mit einem Exponenten erweitert. Die Nutzenfunktionen sehen wie folgt aus:

Tabelle 20 Nutzenfunktionen der nichtlinearen Modelle

U =

$$\beta_{Length} \cdot Length^{e_{Length}} + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (19)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax^{e_{RiseMax}} + \beta_{TLights} \cdot TLights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (20)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights^{e_{TLights}} + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (21)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights + \beta_{PS} \cdot PS^{e_{PS}} + \beta_{BikeAv} \cdot BikeAv \quad (22)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv^{e_{BikeAv}} \quad (23)$$

$$\beta_{Length} \cdot Length^{e_{Length}} + \beta_{RiseMax} \cdot RiseMax^{e_{RiseMax}} + \beta_{TLights} \cdot TLights^{e_{TLights}} + \beta_{PS} \cdot PS^{e_{PS}} + \beta_{BikeAv} \cdot BikeAv^{e_{BikeAv}} \quad (24)$$

$$\beta_{Length} \cdot Length^{e_{Length}} + \beta_{RiseMax} \cdot RiseMax^{e_{RiseMax}} + \beta_{TLights} \cdot TLights^{e_{TLights}} + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv^{e_{BikeAv}} \quad (25)$$

Die beste Log-Likelihood liefert wiederum das letzte Modell (25). Hier besteht aber das Problem, dass die geschätzten Parameter keineswegs signifikant sind. Die Standardabweichungen der Parameter sind teilweise ausserordentlich gross (z.B. $1.8 \cdot 10^{308}$), weshalb die Werte des t-Tests sehr klein werden. Folglich ist auch das Modell nicht wirklich brauchbar.

Bezüglich Signifikanz der Parameter ist einzig das Modell (20) akzeptabel (vgl. Tabelle 21). Jedoch ist die Güte dieses Ansatzes leicht schlechter als die lineare Schätzung des Modells (14). Somit stellt das Modell (14) immer noch die beste Schätzung dar.

Tabelle 21 Schätzergebnisse für das Modell (20)

Parameter	Wert	Wert · Mittelwert	t-Test
β_{Length}	-0.007	-9.053	-7.14
$\beta_{RiseMax}$	-3.530	-1.035	-1.13
$\beta_{TLights}$	-0.382	-0.817	-1.53
$\beta_{\ln(PS)}$	0.733	-1.193	2.17
β_{BikeAv}	0.960	0.592	1.52
$e_{RiseMax}$	0.443		1.98
Log-Likelihood: $\ln(L) = -143.448$, $\rho^2 = 0.714$			

9.2.4 Modelle mit Wechselwirkungen zwischen den Routeneigenschaften

Bei der Modellierung von Wechselwirkungen soll untersucht werden, ob Interaktionen zwischen einzelnen Routeneigenschaften bestehen. Die Möglichkeiten die einzelnen Eigenschaften zu kombinieren sind sehr umfangreich, weshalb eine sinnvolle Auswahl davon getroffen werden muss. Des Weiteren sollten die einzelnen Terme auch eine Interpretation zulassen. Folgende Wechselwirkungsterme wurden als zweckmässig erachtet:

- $Length \cdot RiseMax$ „maximale“ Höhenmeter [m]
- $BikeAv \cdot Length$ Absolute Länge des markierten Fahrradweges [m]
- $BikeAv \cdot Length \cdot RiseMax$ „maximale“ Höhenmeter auf einem markierten Fahrradweg [m]
- $\frac{TLights}{Length}$ Anzahl Lichtsignalanlagen pro Länge [m]

Lineare Ansätze

Basierend auf dem bisher besten Modell (14) werden die Wechselwirkungsterme zunächst als linearer Term hinzugefügt. Die entsprechenden Nutzenfunktionen sind aus Tabelle 22 ersichtlich.

Tabelle 22 Nutzenfunktionen mit Wechselwirkungstermen

U =

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (26)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WRiseLength} \cdot RiseMax \cdot Length$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (27)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WBikeLength} \cdot BikeAv \cdot Length$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (28)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WBikeLengthRise} \cdot BikeAv \cdot Length \cdot RiseMax$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (29)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WTLightsLength} \cdot \frac{TLights}{Length}$$

Lediglich das Modell (26) führt zu einem verbesserten Log-Likelihood. Entsprechende Ergebnisse sind in der nachfolgenden Tabelle ersichtlich.

Tabelle 23 Schätzergebnisse für das Modell (26)

Parameter	Wert	Wert · Mittelwert	t-Test
β_{Length}	-0.005	-6.909	-5.29
β_{RiseMax}	0.309	0.019	1.56
β_{TLights}	-0.440	-0.941	-1.79
$\beta_{\text{ln(PS)}}$	0.725	-1.179	2.18
β_{BikeAv}	0.934	0.576	1.48
$\beta_{\text{WRiseLength}}$	-0.060	¹² -5.041	-2.29

Log-Likelihood: $\ln(L) = -140.161$, $\rho^2 = 0.721$

Es besteht also eine Interaktion zwischen der maximalen Steigung und der Länge der Route. Deren Nutzeneinfluss ist negativ und lediglich 1.4 Mal kleiner als die Länge. Verringert wurde nicht nur die Einflussnahme der Länge sondern auch diejenige der maximalen Steigung. Die anderen Variablen fließen hingegen mit ungefähr der gleichen Gewichtung wie Modell (14) in die Nutzenfunktion ein.

Bei der Interpretation der Wechselwirkungsterme ist Vorsicht geboten, da die involvierten Variablen nicht Mittelwert-befreit sind. Die Mittelwerte der Variablen sind konstant und kommen sowohl in der Variablen selbst als auch im Interaktionsterm der Variablen vor. Folglich ist eine Interpretation schwierig. Eine mögliche Lösung wäre die Daten Mittelwert-befreit zu machen und dann nur die Abweichung der Variablen zu untersuchen. So könnte der Einfluss der einzelnen Terme besser interpretiert werden.

Transformationen

Analog der Vorgehensweise bei den bisherigen Ansätzen werden auch die Interaktionen als transformierte Terme in die Nutzenfunktion eingefügt (vgl. Tabelle 24). Als Umwandlungsfunktion wurde eine Box-Cox-Transformation gewählt, da diese auch eine logarithmische Transformation miteinschliesst (falls $\lambda = 0$).

¹² Einfluss des gesamten Interaktionsterms $\beta_{\text{WRiseLength}} \cdot \text{RiseMax} \cdot \text{Length}$

Tabelle 24 Nutzenfunktionen mit transformierten Wechselwirkungstermen

U =

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (30)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WRiseLength} \cdot T(RiseMax \cdot Length)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (31)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WBikeLength} \cdot T(BikeAv \cdot Length)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (32)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WBikeLengthRise} \cdot T(BikeAv \cdot Length \cdot RiseMax)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (33)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WTLightsLength} \cdot T\left(\frac{TLights}{Length}\right)$$

Die Transformation der Wechselwirkungsterme brachte keine Verbesserung der Log-Likelihood. Somit werden diese Ansätze nicht weiter verfolgt.

Nichtlineare Ansätze

Wiederum mit dem Ziel die bisherigen Modellergebnisse zu verbessern, werden als nächstes die Wechselwirkungsterme mit einem Exponenten erweitert (vgl. Tabelle 25).

Tabelle 25 Nutzenfunktionen mit nichtlinearen Wechselwirkungstermen

U =

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (34)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WRiseLength} \cdot (RiseMax \cdot Length)^{e_{WRiseLength}}$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (35)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WBikeLength} \cdot (BikeAv \cdot Length)^{e_{WBikeLength}}$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (36)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WBikeRise} \cdot (BikeAv \cdot Length \cdot RiseMax)^{e_{BikeLengthRise}}$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights \quad (37)$$

$$+ \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WTLightsLength} \cdot \left(\frac{TLights}{Length}\right)^{e_{WTLightsLength}}$$

Auch das Hinzufügen von Exponenten brachte keine Zunahme der Modellgüte. Bei den Ansätzen mit Interaktionstermen stellt also das Modell (26) die beste Schätzung dar.

9.2.5 Modellschätzung ohne die Routenlänge

Da die Routenlänge bei den vorangehenden Modellen die mit Abstand wichtigste Routeneigenschaft ist und somit das Verhalten der Fahrradfahrer fast im Alleingang erklären kann, soll diese Variable bei den folgenden Modellen weggelassen werden. Der Grund liegt daran, dass die Bedeutung der übrigen Routeneigenschaften etwas näher untersucht werden soll.

Die Vorgehensweise ist dieselbe wie bei den bisherigen Ansätzen. Im Unterschied zum vorherigen Kapitel werden hier lediglich die Ansätze mit den besten Ergebnissen vorgestellt.

Lineare Ansätze

Die Nutzenfunktion des besten linearen Modells sieht wie folgt aus:

$$U = \beta_{RiseMax} \cdot T(RiseMax) + \beta_{TLights} \cdot TLights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (38)$$

Entsprechende Ergebnisse sind aus Tabelle 26 ersichtlich.

Tabelle 26 Schätzergebnisse für das Modell (38)

Parameter	Wert	Wert · Mittelwert	t-Test
$\beta_{RiseMax}$	-13.100	¹³ -19.618	-3.48
$\beta_{TLights}$	-0.956	-2.044	-6.14
$\beta_{\ln(PS)}$	0.612	-0.996	2.42
β_{BikeAv}	0.551	0.339	1.96
λ	0.501		8.00
Log-Likelihood: $\ln(L) = -235.981$, $\rho^2 = 0.53$			

¹³ Einfluss des Box-Cox-transformierten Mittelwertes mit $\lambda = 0.501$

Im Unterschied zu den Ansätzen, welche die Länge mitberücksichtigen, führt hier eine Box-Cox-Transformation der Variable *RiseMax* zu einer Verbesserung der Log-Likelihood-Schätzung des Modells. Bemerkenswert ist die hohe Einflussnahme der maximalen Steigung. Deren Nutzeneinwirkung ist rund 6 Mal grösser als diejenige der übrigen Routeneigenschaften zusammen.

Nichtlineare Ansätze

Das beste nichtlineare Modell erweitert die maximale Steigung und die Anzahl Lichtsignalanlagen um einen Exponenten. Die Nutzenfunktion sieht dann wie folgt aus:

$$U = \beta_{RiseMax} \cdot RiseMax^{e_{RiseMax}} + \beta_{TLights} \cdot TLights^{e_{TLights}} + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (39)$$

Diesbezügliche Schätzergebnisse sind in Tabelle 27 illustriert.

Tabelle 27 Schätzergebnisse für das Modell (39)

Parameter	Wert	Wert · Mittelwert	t-Test
$\beta_{RiseMax}$	-26.400	¹⁴ -6.610	-5.74
$\beta_{TLights}$	-0.270	-1.426	-2.28
$\beta_{\ln(PS)}$	0.633	-1.029	2.52
β_{BikeAv}	0.432	¹⁵ 0.267	1.97
$e_{RiseMax}$	0.501		8.30
$e_{TLights}$	2.190		3.05
Log-Likelihood: $\ln(L) = -233.052$, $\rho^2 = 0.536$			

Wiederum hat die maximale Steigung den mit Abstand grössten Einfluss auf die Nutzenzusammensetzung. Gegenüber dem Modell (38) wurde aber deren Nutzenanteil deutlich verkleinert.

¹⁴ Einfluss des gesamten Terms $\beta_{RiseMax} \cdot RiseMax^{e_{RiseMax}}$

¹⁵ Einfluss des gesamten Terms $\beta_{BikeAv} \cdot BikeAv^{e_{BikeAv}}$

9.3 Analyse der verschiedenen Modelle

In diesem Kapitel werden die Modelle (14), (26), (38) sowie (39), welche die besten Schätzergebnisse erbrachten, etwas näher untersucht. Die ersten beiden Ansätze berücksichtigen die Routenlänge, während die letzten zwei Modelle die Länge ausser Acht lassen. Tabelle 28 fasst deren Nutzenfunktionen nochmals zusammen.

Tabelle 28 Nutzenfunktionen mit nichtlinearen Wechselwirkungstermen

U =

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (14)$$

$$\beta_{Length} \cdot Length + \beta_{RiseMax} \cdot RiseMax + \beta_{TLights} \cdot TLights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv + \beta_{WRiseLength} \cdot RiseMax \cdot Length \quad (26)$$

$$\beta_{RiseMax} \cdot T(RiseMax) + \beta_{TLights} \cdot TLights + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (38)$$

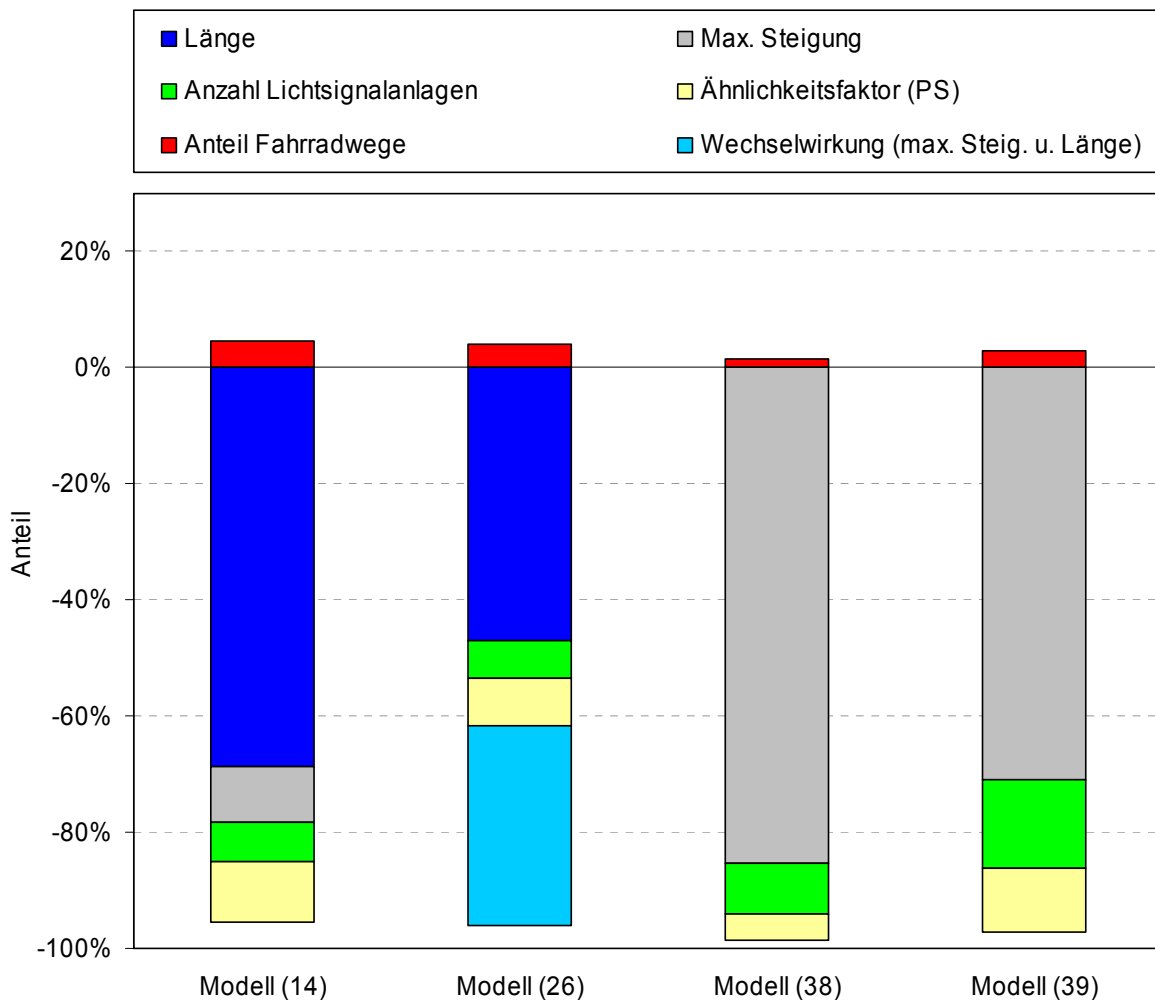
$$\beta_{RiseMax} \cdot RiseMax^{e_{RiseMax}} + \beta_{TLights} \cdot TLights^{e_{TLights}} + \beta_{\ln(PS)} \cdot \ln(PS) + \beta_{BikeAv} \cdot BikeAv \quad (39)$$

9.3.1 Zusammensetzung des Nutzens

Wie in den vorhergehenden Ausführungen angedeutet, ist es interessant zu ermitteln wie der Nutzen einer Route zustande kommt. Hierzu können die Variablenmittelwerte mit den Parameterwerten der verschiedenen Modelle multipliziert werden und danach deren prozentualer Einfluss auf den gesamten Nutzen berechnet werden. Die folgenden drei Abbildungen illustrieren die mittlere Nutzenzusammensetzung für den gesamten Schätzdatensatz, für die gewählten Routen sowie für die Alternativen. Positive Anteilswerte deuten darauf hin, dass die entsprechende Eigenschaft zu einer Nutzenerhöhung beiträgt, während negative Anteile den Nutzen verringern.

Zusammensetzung des mittleren Nutzens aller Routen

Abbildung 52 Mittlere Nutzenzusammensetzung aller Routen



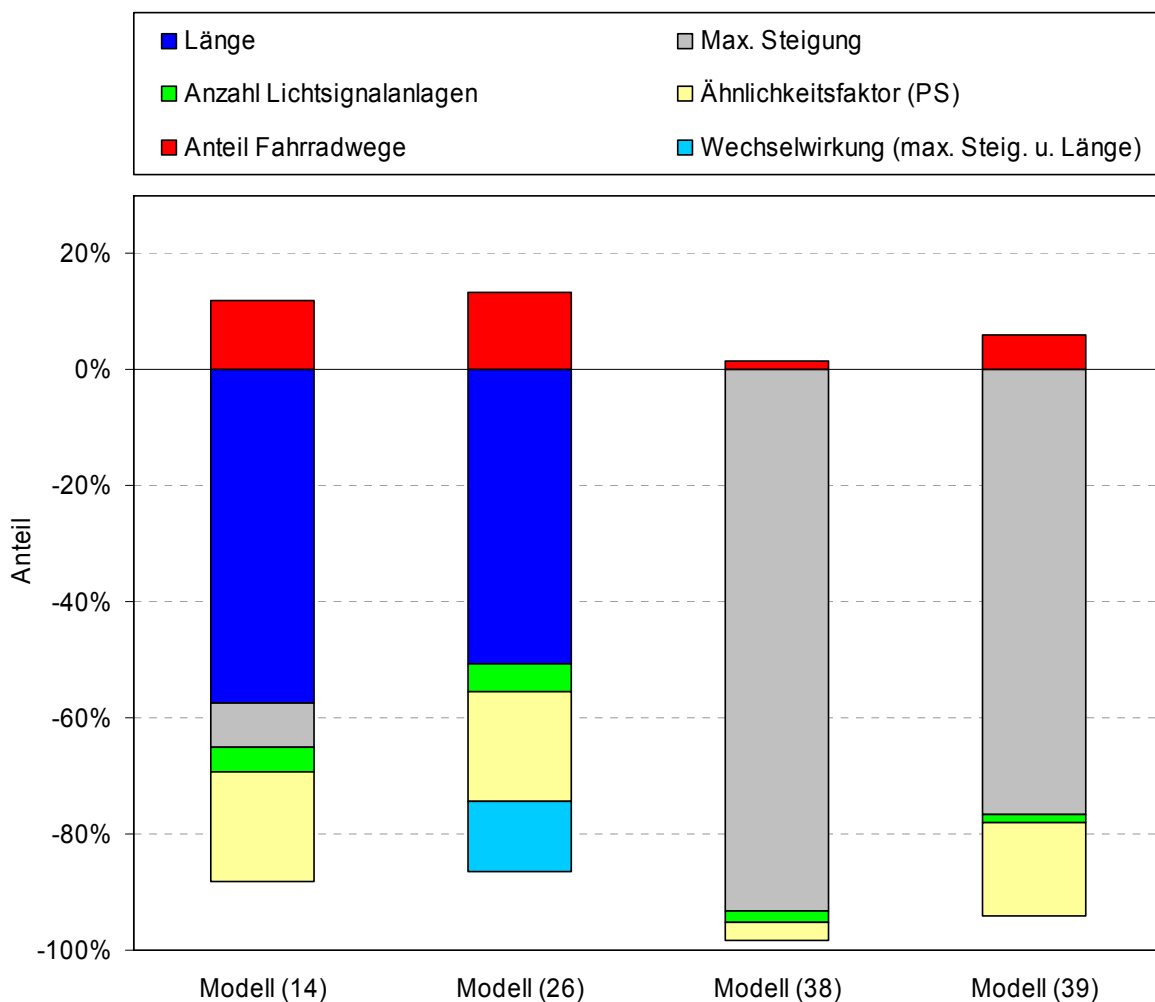
Zum grössten negativen Teil der Nutzenzusammensetzung trägt im Modell (14) die Länge bei. Die maximale Steigung, die Anzahl Lichtsignalanlagen sowie der Ähnlichkeitsfaktor tragen ebenfalls zu einer Nutzenverringerung bei, jedoch in einem viel kleineren Ausmass. Zu einer Nutzenerhöhung führt der Anteil Fahrradwege, dies mit einem Anteil von ca. 5%.

Im Modell (26) ist der Einfluss der Interaktion zwischen der Routenlänge und der maximalen Steigung gut ersichtlich. Die Einflussnahme der maximalen Steigung alleine wird so klein, dass sie in der Abbildung nicht sichtbar ist. Der positive Anteil durch die Fahrradwege ist ungefähr gleich gross wie beim Modell (14).

Der Nutzen der Modelle (38) und (39) wird zu über 80% resp. 70% von der maximalen Steigung bestimmt. Ebenfalls negativ tragen der Ähnlichkeitsfaktor und die Anzahl Lichtsignalanlagen bei. Der Anteil der Lichtsignalanlagen beim Ansatz (39) ist ungefähr doppelt so gross verglichen mit den Modellen (14) und (26). Einen sehr geringen Nutzeneinfluss übt der Anteil markierter Fahrradwege aus.

Zusammensetzung des mittleren Nutzens der gewählten Routen

Abbildung 53 Mittlere Nutzenzusammensetzung der gewählten Routen



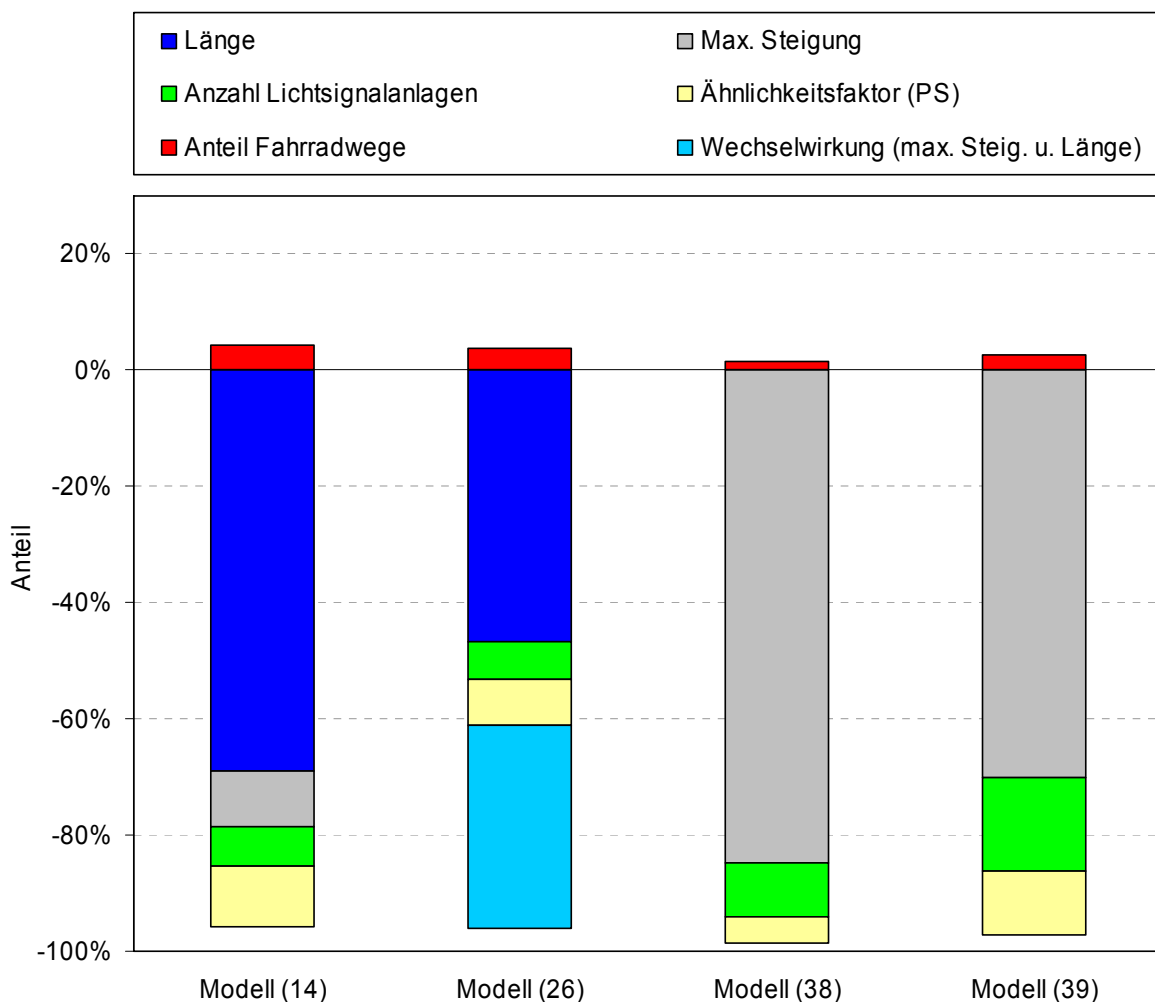
Die mittlere Nutzenzusammensetzung der gewählten Routen unterscheidet sich vom gesamten Schätzdatensatz. Sehr gut sichtbar ist, dass der positive Anteil durch die markierten Fahrradwege bei den Modellen (14) und (26) mehrheitlich auf Kosten der Routenlänge zugenommen

hat. Bei diesen Ansätzen wurde auch der negative Einfluss der Ähnlichkeit erhöht. Etwas geringer ist der Nutzen verringernde Anteil der Lichtsignalanlagen.

In den Modellen (38) und (39) hat der negative Beitrag der maximalen Steigung zugenommen. Der Nutzen von Modell (38) wird zu über 90% von der maximalen Steigung zusammengesetzt. Erhöht hat sich aber auch bei diesen Ansätzen der positive Anteil markierter Fahrradwege.

Zusammensetzung des mittleren Nutzens der Alternativen

Abbildung 54 Mittlere Nutzenzusammensetzung der Alternativen



Die Abbildung über die mittlere Nutzenzusammensetzung der Alternativen unterscheidet sich nicht gross von derjenigen des gesamten Schätzdatensatzes. Der Grund liegt daran, dass die

Anzahl Alternativrouten verglichen mit den gewählten Routen viel grösser ist und somit einen höheren Einfluss auf den gesamten Schätzdatensatz ausübt. Folglich weisen die Mittelwerte der Alternativen resp. des gesamten Schätzdatensatzes geringe Unterschiede auf. Einzig der Nutzenanteil durch die markierten Fahrradwege scheint etwas geringer als beim gesamten Schätzdatensatz zu sein. Dies ist aber nichts Aussergewöhnliches, da bei dieser Analyse die gewählten Routen, welche hohe Anteile an Fahrradwegen besitzen, fehlen.

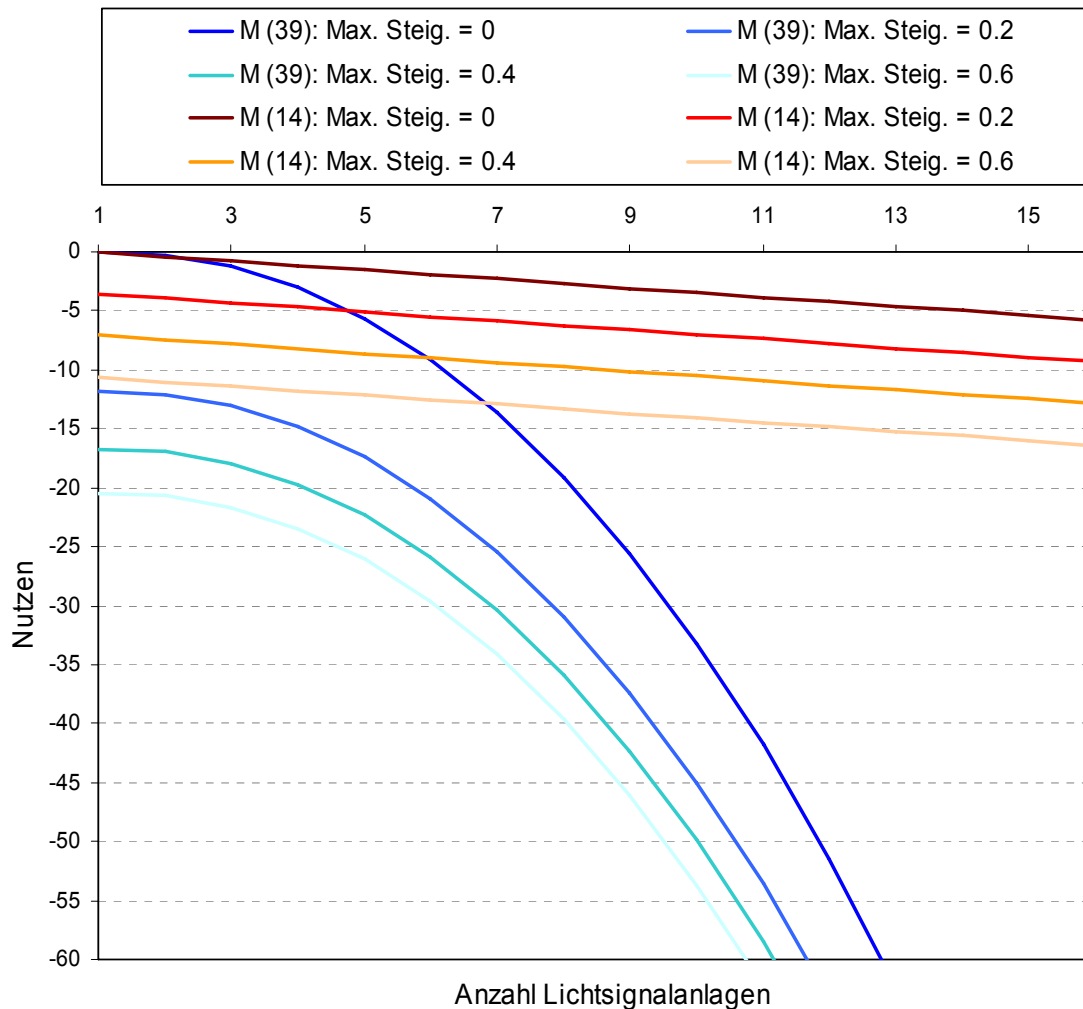
9.3.2 Einfluss verschiedener Routeneigenschaften

In diesem Kapitel folgen weitere Analysen über die Art des Einflusses verschiedener Variablen. Es werden immer die Nutzenfunktionen zweier Modelle miteinander konfrontiert.

Anzahl Lichtsignalanlagen verglichen mit der maximalen Steigung

Abbildung 55 zeigt sehr gut wie sich der Nutzen mit der Anzahl Lichtsignalanlagen für verschiedene Werte der maximalen Steigung verändert. Verglichen werden das lineare Modell (14) sowie der exponentielle Ansatz (39). Mit zunehmender maximaler Steigung und Anzahl Lichtsignalanlagen verringert sich wie erwartet der Nutzen. Der Unterschied der beiden Modelle manifestiert sich besonders in einer hohen Anzahl Lichtsignalanlagen. Aufgrund der eingeführten Exponenten für die maximale Steigung und die Anzahl Lichtsignalanlagen im Modell (39) ist dies nicht weiter erstaunlich. Ebenfalls darauf zurückzuführen ist die Tatsache, dass der Nutzen des Modells (39) exponentiell mit kleineren Werten der maximalen Steigung zunimmt.

Abbildung 55 Einfluss der Anzahl Lichtsignalanlagen und max. Steigung

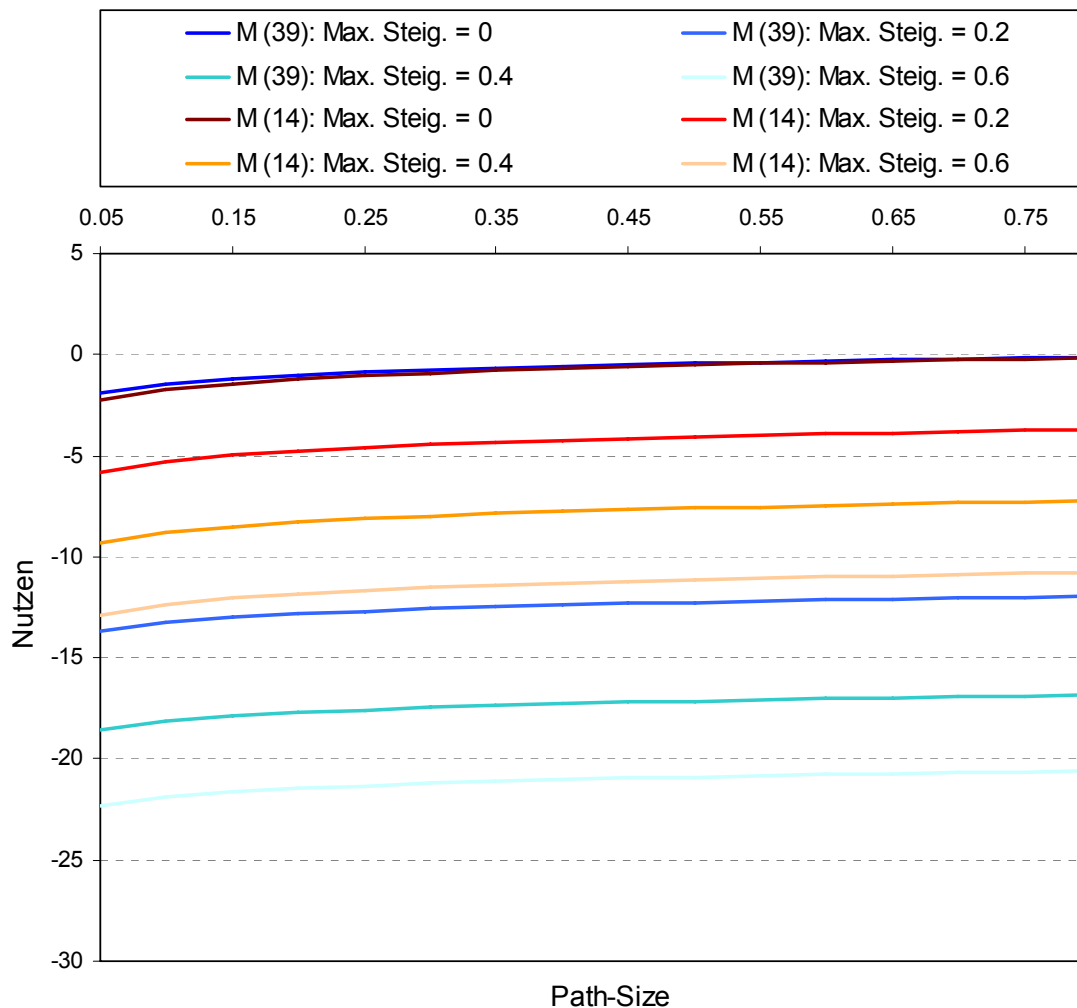


Ähnlichkeit verglichen mit der maximalen Steigung

Bei dieser Analyse soll untersucht werden, wie sich der Nutzen bei unterschiedlichen Ähnlichkeits- und maximalen Steigungswerten verhält. In Abbildung 56 ist ersichtlich, dass die Unterschiede der beiden Ansätze (14) und (39) bei unterschiedlichen maximalen Steigungswerten liegen. Auffallend ist dabei der hohe Nutzen des Modells (39) bei einer maximalen Steigung von Null. Diese Zunahme lässt sich, analog zur vorherigen Untersuchung, mit dem Exponenten, der dieser Variablen angefügt wurde, erklären ($e_{RiseMax} = 0.501$; entspricht einer Wurzelfunktion). Dessen Einfluss wird erst bei sehr kleinen maximalen Steigungen relevant.

Der Einfluss des Ähnlichkeitsfaktors ist bei beiden Modellen gleich, da diese Grösse bei beiden Ansätzen logarithmisch in die Nutzenfunktion einfliesst.

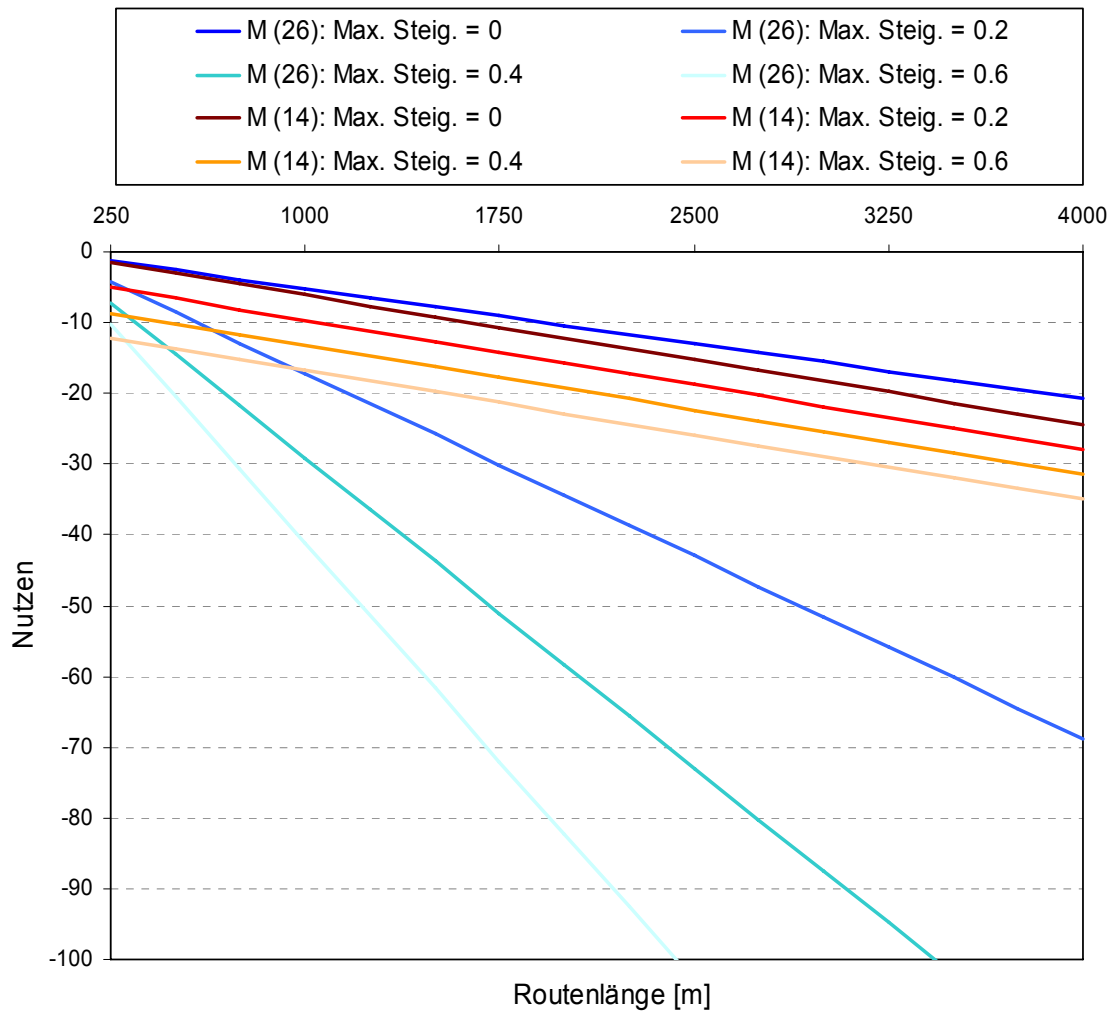
Abbildung 56 Einfluss der Ähnlichkeit und der max. Steigung



Routenlänge verglichen mit der maximalen Steigung

Bei dieser Analyse werden die Ansätze (14) und (26), welche beide die Routenlänge integriert haben, näher betrachtet. Wie in Abbildung 57 ersichtlich, führt die Einführung des Interaktionsterms zwischen der Länge und der maximalen Steigung dazu, dass der Nutzen mit zunehmender Routenlänge stärker abnimmt. Beim Modell (26) offenbart sich auch die Tatsache, dass bei grösseren Routenlängen die Unterschiede zwischen den verschiedenen maximalen Steigungswerten grösser werden.

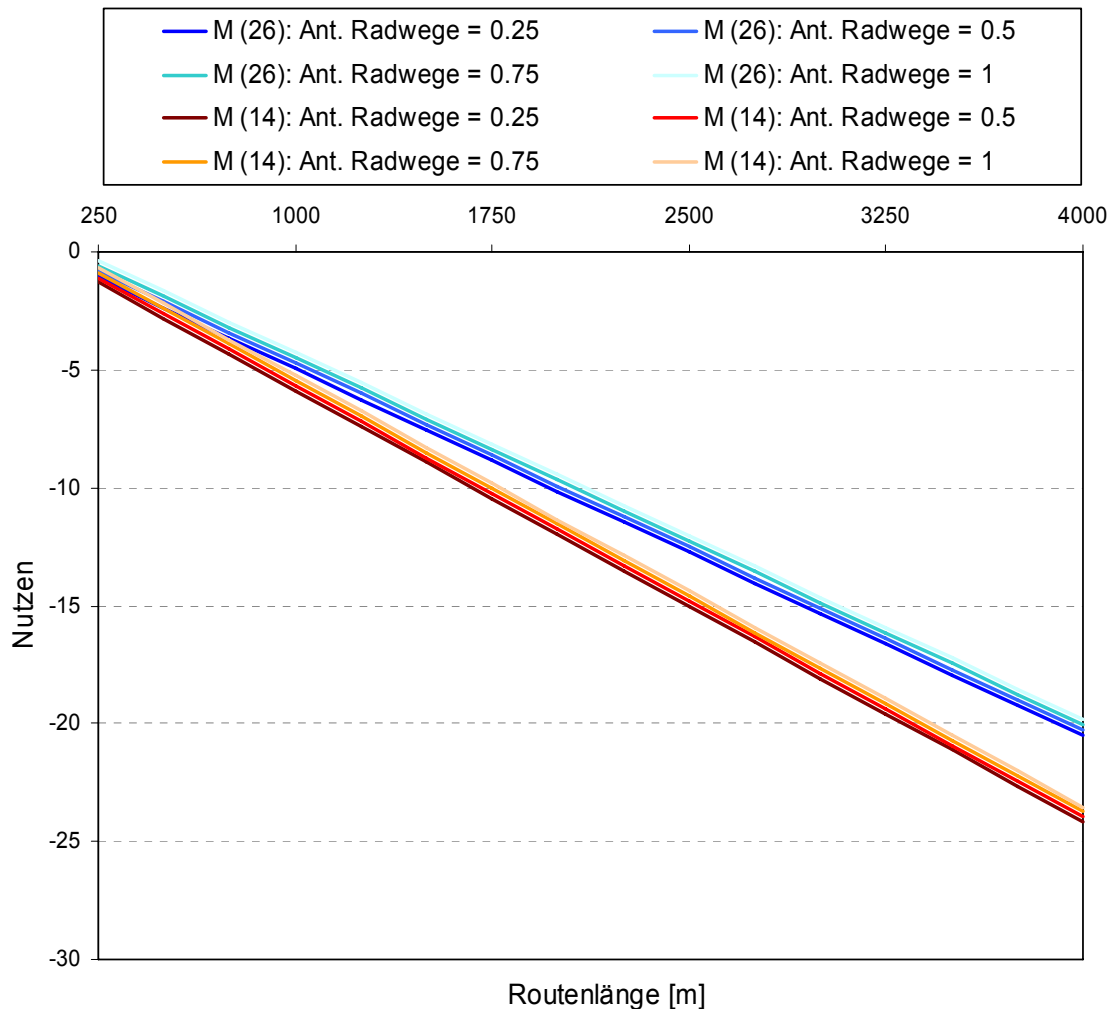
Abbildung 57 Einfluss der Länge und max. Steigung



Routenlänge verglichen mit dem Anteil Fahrradwege

Abbildung 58 lässt erkennen, dass die Nutzengeraden beim Modell (26) analog zur vorherigen Grafik steiler verlaufen. Der Nutzen nimmt also mit zunehmender Routenlänge schneller ab. Sehr gut ersichtlich ist auch, dass verschiedene Anteile an Fahrradwegen einen äusserst geringen Einfluss ausüben. Die Nutzenunterschiede sind wirklich minimal.

Abbildung 58 Einfluss der Länge und des Anteils markierter Fahrradwege



9.3.3 Nutzendifferenz zwischen den gewählten Routen und den Alternativen

In den letzten Abschnitten wurden die Zusammensetzung des Nutzens sowie der Einfluss der verschiedenen Routeneigenschaften dargelegt. Hierbei sollte beachtet werden, dass der absolute Nutzenwert bei einer Entscheidung keine Rolle spielt. Die Entscheidung, ob eine Route gewählt wird oder nicht, erfolgt aufgrund der Nutzendifferenz und nicht auf Basis des absoluten Nutzenwertes.

Folgendes sehr einfaches Beispiel soll diesen Sachverhalt näher erläutern. Gegeben seien 2 Modelle A und B, welche je zwei Alternativen zur Auswahl stellen (vgl. Tabelle 29). Der Nutzenunterschied zwischen den Routen ist für beide Modelle gleich, der absolute Wert je-

doch verschieden. Nun wird gezeigt, dass die Wahrscheinlichkeit die erste Route zu wählen bei beiden Ansätzen gleich ist.

Tabelle 29 Beispiel Nutzendifferenz

	Nutzen Route 1	Nutzen Route 2
Modell A	1	2
Modell B	100	101

Nach Formel (3) ist die Wahrscheinlichkeit Route 1 zu wählen gegeben durch:

$$P_{U_1} = \frac{e^1}{e^1 + e^2} \text{ resp. } P_{U_1} = \frac{e^{100}}{e^{100} + e^{101}} \quad (40)$$

Schreibt man die rechte Gleichung in die Form $P_{U_1} = \frac{e^{99} \cdot e^1}{e^{99} \cdot (e^1 + e^2)}$ um, so kann der Term e^{99} weggekürzt werden und es wird ersichtlich, dass die Wahrscheinlichkeit bei beiden Modellen gleich ist. Folglich sollte klar sein, dass für die Routenwahl lediglich die Nutzendifferenz entscheidend ist.

Abbildung 59 illustriert die Zusammensetzung der mittleren Nutzendifferenz zwischen den gewählten Routen und den Alternativen.

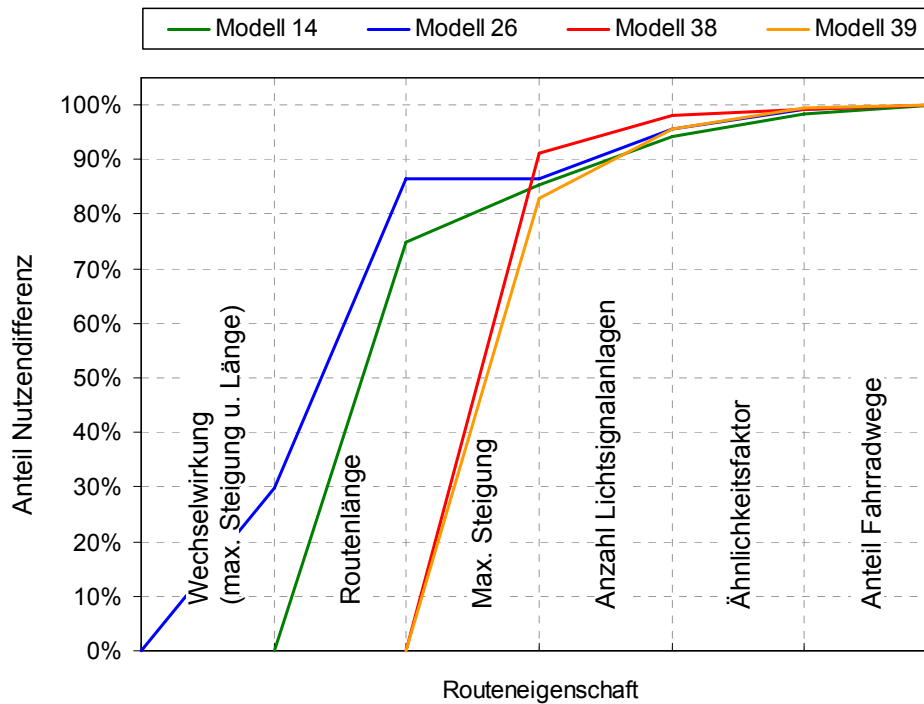
Die einzelnen Werte wurden wie folgt berechnet:

$$U_{Variable_i} = Parameter_i \cdot (Mittelwert_{Variable_i, Alternativen} - Mittelwert_{Variable_i, Gewählte Routen}) \quad (41)$$

$i = \text{Länge, maximale Steigung, Anzahl Lichtsignalanlagen etc.}$

Zwecks besseren Verständnisses wurden die Werte in Anteile zwischen 0 und 100% umgerechnet. Ein Vergleich der Steigung der Geraden erlaubt nun eine Beurteilung der Variablen Einflüsse.

Abbildung 59 Mittlere Nutzendifferenz zwischen den gewählten Routen und Alternativen



Sehr gut ersichtlich ist der grosse Einfluss der Routenlänge bei den Modellen (14) resp. (26). Bei diesen Ansätzen kommt die Nutzendifferenz also grössten Teils durch den mittleren Längenunterschied zustande. Beim Modell (26) ist zusätzlich der Einfluss des Interaktionsterms ziemlich gross.

Bei den anderen beiden Modellen liefert die maximale Steigung den grössten Beitrag zur Nutzendifferenz, wie der steile Kurvenverlauf illustriert.

Die Anzahl Lichtsignalanlagen sowie der Ähnlichkeitsfaktor tragen bei allen vier Modellen ungefähr gleich viel zur Nutzendifferenz bei.

Schliesslich wird noch illustriert, dass der Anteil Fahrradwege nur einen geringen Anteil beisteuert.

9.3.4 Nutzen der gewählten Routen

Die Schätzung eines Modells ist eine Optimierungsaufgabe, welche so gelöst wird, dass die Gesamtwahrscheinlichkeit für die Wahl einer beobachteten Route am grössten ist. Da also die Gesamtwahrscheinlichkeit betrachtet wird, können Einzelfälle auftreten, bei denen die gewählte Route nicht wirklich den grössten Nutzen aufweist.

Eine interessante Grösse ist deshalb die Anzahl bzw. der Anteil gewählter Routen, welche den grössten Nutzenwert besitzen. Tabelle 30 illustriert entsprechende Ergebnisse.

Tabelle 30 Gewählte Routen mit grösstem Nutzen

Modell	Anzahl OD-Paare, bei welchen die gewählte Route den grössten Nutzen aufweist (Total: 165 OD-Paare)
14	120 (72.7%)
26	120 (72.7%)
38	91 (55.2%)
39	94 (57.0%)

Bei den Modellen (14) und (26) besitzt in rund 73 % der Fälle die gewählte Route auch wirklich den grössten Nutzen. Bei den Ansätzen, welche die Routenlänge ausser Acht lassen, sind es lediglich 55.2 % resp. 57 % der Routen.

9.3.5 Elastizitäten

Bei der Modellierung diskreter Entscheidungen können die direkten Elastizitäten eine weitere gute Interpretationshilfe darstellen. Deshalb soll auch für die hier vorliegenden Routen diese Grösse berechnet werden.

Die direkte Elastizität gibt die prozentuale Änderung der Auswahlwahrscheinlichkeit einer Route bei der Änderung eines Attributs um 1 % an. Die anderen Routencharakteristiken bleiben dabei identisch. Nach Ortúzar *et al.* (2001) ist die allgemeine Elastizitätsformel wie folgt gegeben:

$$E_{P_i, x_i} = \frac{\delta P_i}{\delta x_i} \cdot \frac{x_i}{P_i} \cdot (1 - P_i) \quad (42)$$

Für ein lineares Modell ergibt sich die Form:

$$E_{P_i, x_i} = \beta \cdot x_i \cdot (1 - P_i) \quad (43)$$

β geschätzter Parameter für das Attribut x

P_i Auswahlwahrscheinlichkeit für die Alternative i

Bei einem nichtlinearen Ansatz müssen zusätzlich die einzelnen Exponenten berücksichtigt werden, weshalb sich folgende Elastizitätsformel ergibt:

$$E_{P_i, x_i} = e_i \cdot \beta_i \cdot x_i^{e_i} \cdot (1 - P_i) \quad (44)$$

Basierend auf diesen Grundlagen wurden für die Modelle (14), (26), (38) und (39) die Elastizitäten der gewählten Routen berechnet. Die Ergebnisse wurden anhand von Boxplots visualisiert und sind in den Abbildungen 60 bis 63 dargestellt. Die genauen statistischen Werte sind zudem aus den Tabellen 43 bis 46 im Anhang A3 ersichtlich.

Beim Ansatz (14) weist die Länge die grösste Elastizität auf. Der Median liegt bei -0.75, was bedeutet, dass die Wahrscheinlichkeit für die Wahl einer beobachteten Route im Mittel um 0.75 % abnimmt, wenn die Länge um 1 % zunimmt. Ebenfalls ziemlich elastisch ist der Ähnlichkeitsfaktor. Die geringsten Werte sind bei der Anzahl Lichtsignalanlagen zu finden. Für den Anteil markierter Fahrradwege ergibt sich eine positive Elastizität, da diese Variable auch positiv in die Nutzenfunktion einfließt.

Auch beim Modell (26) weist die Länge die grösste Elastizität auf. Gut erkennbar ist, dass die Elastizität der maximalen Steigung sehr stark abnimmt. Erklärbar ist dies mit der Einführung des Wechselwirkungsterms, welcher ziemlich elastisch reagiert. Wie in vorangehenden Analysen gezeigt, ist der Einfluss der maximalen Steigung alleine beim Modell (26) sehr gering. Folglich könnte diese weggelassen und nur als Interaktionsterm eingefügt werden.

Der Vergleich der Modelle (38) und (39) offenbart vor allem Unterschiede bei der maximalen Steigung. Die Box-Cox-Transformation der maximalen Steigung bewirkt anscheinend, dass die Elastizität der entsprechenden Variablen, verglichen mit Modell (39), abnimmt. Ein kleiner Unterschied ist auch bei der Anzahl Lichtsignalanlagen zu finden. Praktisch identische Elastizitätswerte ergeben sich hingegen für die Ähnlichkeit und den Anteil markierter Fahrradwege.

Abbildung 60 Elastizitäten Modell 14

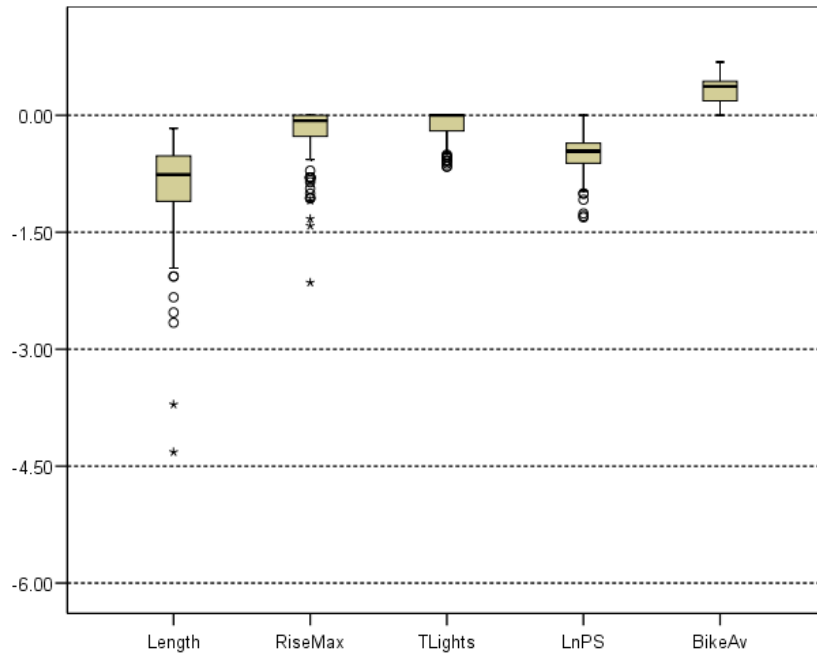


Abbildung 61 Elastizitäten Modell 26

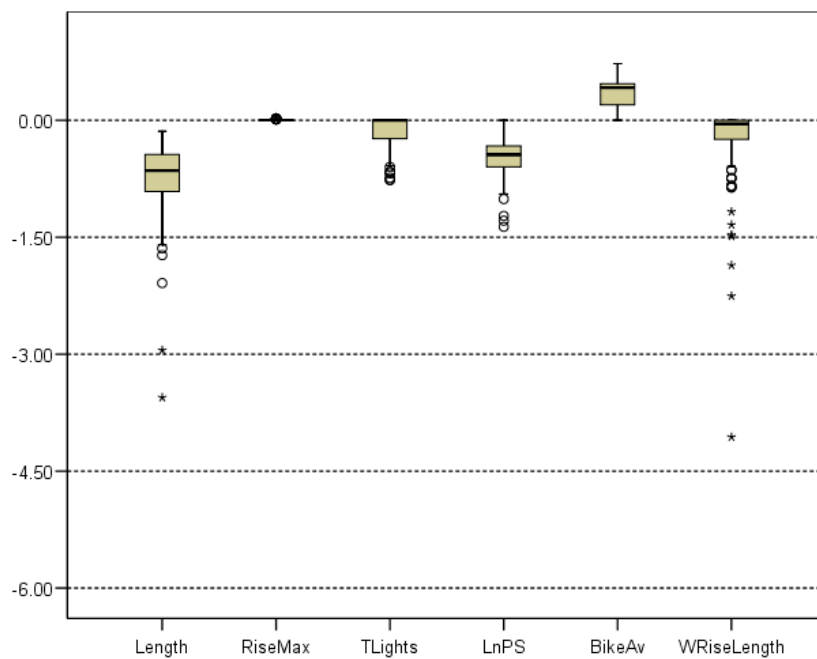


Abbildung 62 Elastizitäten Modell 38

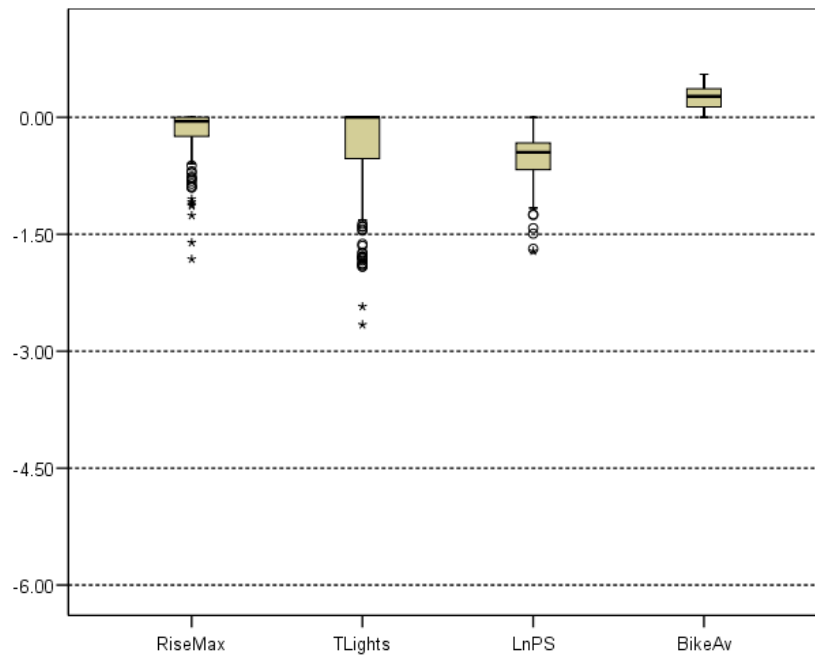
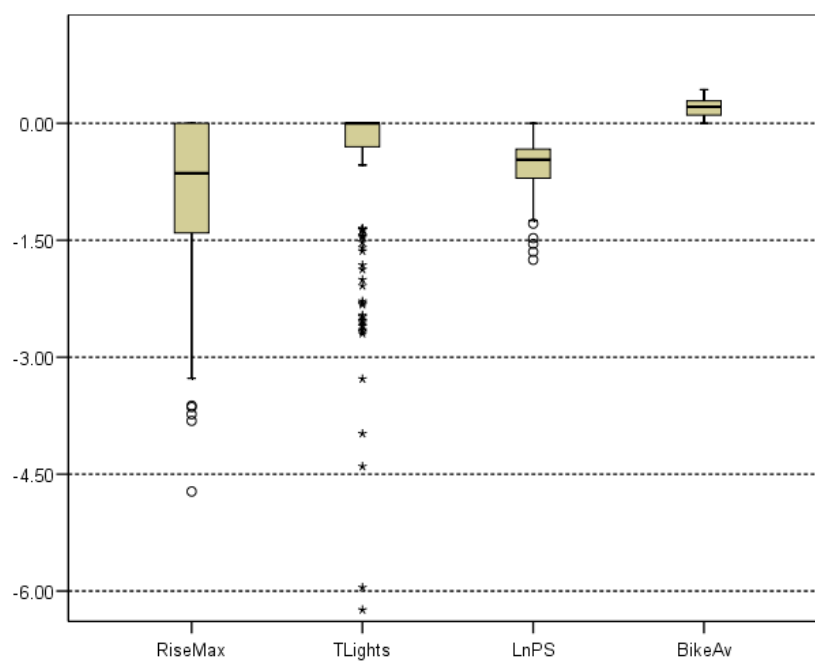


Abbildung 63 Elastizitäten Modell 39



9.3.6 Eigenschaftsäquivalente

Für einen Planer ist es relevant zu wissen, mit welcher Längenverkürzung zum Beispiel eine zusätzliche Lichtsignalanlage verbunden ist (bei gleich bleibendem Nutzen). Deshalb sollen in diesem Abschnitt verschiedene „Äquivalente“ gebildet werden, die für Planer von Bedeutung sein können.

Betrachtet wird lediglich das Modell (14), da in diesem Ansatz die meisten Routeneigenschaften berücksichtigt werden. Zudem ist es leichter zu interpretieren als beispielsweise das Modell (26) mit dem Interaktionsterm. Tabelle 31 illustriert die verschiedenen Ergebnisse.

Tabelle 31 Eigenschaftsäquivalente für das Modell (14)

Quotienten	Werte
$\frac{\beta_{TLights}}{\beta_{Length}}$	64.17
$\frac{\beta_{BikeAv}}{\beta_{Length}}$	-144.00
$\frac{\beta_{RiseMax}}{\beta_{Length}}$	2950.00
$\frac{\beta_{TLights}}{\beta_{BikeAv}}$	-0.44
$\frac{\beta_{TLights}}{\beta_{RiseMax}}$	0.02

Es ist ersichtlich, dass eine zusätzliche Lichtsignalanlage einer Routenverkürzung von rund 64 m entspricht. Wird der Anteil Fahrradwege betrachtet, so heisst dies, dass derselbe um 0.44 erhöht werden muss, falls die Route eine Lichtsignalanlage mehr passiert.

Weiter erkennbar ist, dass wenn eine Route um 14.4 m länger wird, der Anteil Fahrradwege um 0.1 zunehmen sollte. Oder anders ausgedrückt, eine Erniedrigung des Anteils Fahrradwege um 0.1 bedingt eine um 14.4m kürzere Route, wenn der Nutzen dieser Route gleich bleiben soll.

Ferner entspricht eine Zunahme der maximalen Steigung von 0.01 einer um 29.5 m kürzeren Route. Führt eine Route hingegen über eine zusätzliche Lichtsignalanlage, so bedingt dies eine um 0.02 niedrigere maximale Steigung.

9.4 Zusammenfassung der Modellergebnisse

Die verschiedenen Modellschätzungen und Analysen der vorherigen Kapitel haben zu einer Vielzahl von Erkenntnissen geführt, welche hier zusammengefasst werden.

Bei Modellen, welche die 5 Eigenschaften Routenlänge, maximale Steigung, Anzahl Lichtsignalanlagen, Ähnlichkeit und Anteil Fahrradwege berücksichtigen, lassen sich folgende Erkenntnisse formulieren:

- Das lineare Modell (14) mit dem logarithmisch transformierten Ähnlichkeitsfaktor führt zu den einträglichsten und verständlichsten Resultaten.
- Gute Ergebnisse erzielt auch der lineare Ansatz (26) mit einem Wechselwirkungsterm zwischen der maximalen Steigung und Länge der Routen.
- Die Interpretation des Interaktionstermes ist ziemlich schwierig, weshalb das lineare Modell (14) vorgezogen werden sollte.
- Ausser des Anteils Fahrradwege fliessen alle Routeneigenschaften negativ in die Nutzenfunktion ein.
- Der grösste Einfluss auf die Routenwahl hat die Länge des Weges.
- Die kleinste Einflussnahme übt der Anteil markierter Fahrradwege aus.
- Etwa gleich gross ist die negative Wirkung der Anzahl Lichtsignalanlagen und der Ähnlichkeit.
- Die Routenlänge ist auch bezüglich der Elastizität die wichtigste Grösse. Kleine Änderungen wirken sich relativ stark auf die Auswahlwahrscheinlichkeit einer Route aus.
- Der Ähnlichkeitsfaktor reagiert ebenfalls ziemlich elastisch auf Änderungen.
- Die ideale Fahrradroute ist gemäss diesen Erkenntnissen möglichst kurz, flach, eigenständig und kreuzt eine geringe Anzahl an Lichtsignalanlagen. Des Weiteren verfügt sie über einen hohen Anteil an markierten Fahrradwegen.

Wenn die Routenlänge ausser Acht gelassen wird, ändern sich die Variableneinflüsse. Die wichtigsten Erkenntnisse aus den Schätzungen der Modelle ohne die Routenlänge können folgendermassen zusammengefasst werden:

- Bei linearen Ansätzen führt eine Box-Cox-Transformation der maximalen Steigung zu einer verbesserten Modellgüte.

- Die Erweiterung der maximalen Steigung und der Anzahl Lichtsignalanlagen mit einem Exponenten verbessert die Schätzung für nichtlineare Modelle.
- Mit Ausnahme des Anteils Fahrradwege fließen wiederum wie erwartet alle Eigenschaften negativ in die Nutzenfunktion ein.
- Den mit Abstand grössten Einfluss übt die maximale Steigung aus.
- Geringfügig ist die Einflussnahme des Anteils Fahrradwege.
- Die grössten Elastizitätswerte besitzen die maximale Steigung und der Ähnlichkeitsfaktor.
- Wenn die Routenlänge ausgeblendet wird, sollte die ideale Fahrradroute als wichtigste Eigenschaft zumindest flach sein. Weiter sollte sie wenn möglich eigenständig sein, über wenige Lichtsignalanlagen führen und hohe Anteile an markierten Fahrradwegen besitzen.

Bei der Planung eines neuen Fahrradweges sollte demnach die Direktheit einer Route oberste Priorität geniessen. Überdies sollten steile Steigungen und eine hohe Anzahl an Lichtsignalanlagen vermieden werden. Mit gleich grosser Priorität sollte die Ähnlichkeit (vgl. Kapitel 6) zwischen den Routen berücksichtigt werden. Viele ähnliche Wege werden negativ bewertet. Folglich sollte besser auf eine ähnliche Route verzichtet werden. Dafür müsste aber die bestehende Route mit den genannten Kriterien optimiert werden. Obwohl der positive Einfluss des Anteils markierter Fahrradwege geringen Einfluss hat, sollte dieser Aspekt nicht vernachlässigt werden. Ein hoher Anteil an markierten Fahrradwegen erhöht die Sicherheit bezgl. möglicher Unfälle, weshalb er zusätzliche Beachtung bekommen sollte. Schliesslich können die in Abschnitt 9.3.6 berechneten Äquivalente bei der Planung hilfreich sein. Im nächsten Abschnitt werden Punkte aufgegriffen, die bei der Schätzung von Modellen zu beachten sind.

9.5 Diskussion

Zunächst sollte beachtet werden, dass die Ergebnisse der Modellschätzungen von der Qualität des Schätzdatensatzes in starkem Masse abhängig sind. Dieser wiederum basiert auf Daten, welche durch verschiedene Ansätze und Algorithmen aufbereitet oder erzeugt wurden. Das wohl grösste Problem dieser Arbeit ist die kurze Distanz der gewählten Routen. Auf dessen Gründe wurde schon mehrfach hingewiesen und diese sollen hier nicht mehr erwähnt werden. Ebenfalls kritisch zu betrachten ist der Alternativensatz. Dessen Qualität ist einerseits durch die Funktionsweise des Algorithmus selbst bedingt und andererseits von der Distanz zwischen den OD-Paaren abhängig.

Bei der Schätzung von Modellen gibt es meistens ein bestes Modell und ein am besten geeignetes Modell. Ersteres liefert die besten Ergebnisse bezüglich Log-Likelihood und Signifikanz der einzelnen Parameter, doch ist es häufig schwierig zu interpretieren. Das am besten geeignete Modell generiert ebenfalls gute Schätzergebnisse, doch im Unterschied zum komplizierten Ansatz ist dessen Aufbau simpler und damit einfacher zu verstehen. Das in dieser Arbeit am besten geeignete Modell ist der lineare Ansatz (14), dessen Modellgüte nur geringfügig schlechter ist als diejenige des besten Modells (26). Mit einem ρ^2 von 0.715 erreicht das Modell (14) gute Schätzergebnisse. Über 70% der Varianz der Routeneigenschaften kann damit erklärt werden. Zu dieser Thematik soll noch erwähnt werden, dass die Nutzenfunktion des Modells (14) auch „verallgemeinbar“ ist und somit möglicherweise für andere Situationen brauchbar. Würden hoch komplexe Nutzenfunktionen, welche auf jede kleine Besonderheit der Daten eingehen, aufgestellt werden, so wären deren Ergebnisse kaum für andere Arbeiten brauchbar.

Offensichtlich ist, dass einige Variablen untereinander korrelieren. Dies kann einerseits eine Schwäche der Daten sein oder andererseits in der Natur der Daten liegen. So ist wohl klar, dass die maximale Steigung mit der Routenlänge korreliert und kaum vermeidbar ist. Die Korrelation des Ähnlichkeitsfaktors mit der Länge ist hingegen schwierig zu beurteilen. Es kann nicht deutlich erkannt werden, ob dies eine spezielle Datenschwäche ist. Etwas erstaunlich ist, dass die Anzahl Lichtsignalanlagen nur sehr schwach mit der Routenlänge korreliert (vgl. Tabelle 42). Dafür korreliert die besagte Anzahl mit der maximalen Steigung, was auf den ersten Blick merkwürdig vorkommt. Es kann jedoch damit erklärt werden, dass steilere Wege in Zürich eher an der Stadtperipherie liegen, wo es nicht viele Kreuzungen mit Lichtsignalanlagen gibt. Auch die Korrelation des Anteils markierter Fahrradwege mit der maximalen Steigung kann ähnlich begründet werden. Höchstwahrscheinlich gibt es weniger markierte Fahrradwege in zentralen Stadtgebieten mit vielen Lichtsignalanlagen.

In zahlreichen früheren Studien (Aultman-Hall *et al.*, 1997; Axhausen und Smith, 1986; Bovy und Stern, 1990; Hochmair, 2004; Stinson und Bhat, 2003) wurde das Verkehrsvolumen, welches auch für die Sicherheit steht, als ein wichtiges Entscheidungskriterium genannt. In dieser Arbeit erwiesen sich die Kategorien des durchschnittlichen Tagesverkehrs (DTV) als unbedeutend. Hierbei sollte beachtet werden, dass die Klassierung derselben auf einer Schätzung beruht (vgl. Kapitel 4.3.3). Eine Reklassierung verbunden mit anderen Klassengrenzen und/oder -größen könnte möglicherweise zu anderen Ergebnissen führen. Eine weitere Überlegung wäre die Verwendung eines anderen Ansatzes für die Schätzung des Verkehrs.

10 Schlussfolgerungen und Ausblick

Der Weg von rohen GPS-Daten zu brauchbaren Routenwahl Modellen ist ziemlich lang und beinhaltet eine Reihe von Berechnungsschritten. Insbesondere die verschiedenen Datenaufbereitungsschritte sind aufwändig und werden häufig unterschätzt. Bei dieser Arbeit gehörten, neben diversen kleineren Aufbereitungsarbeiten, die Integration von vier Strassennetzen und das *Matching* von GPS-Punkten dazu. Nach der Datenaufbereitung wurde ein Alternativensatz generiert sowie die Ähnlichkeit zwischen den Routen berechnet. Auch diese Schritte erwiesen sich als entscheidend für die Schlussergebnisse. Zuletzt wurden verschiedene Modelle geschätzt und deren Resultate sorgfältig analysiert.

Im nächsten Unterkapitel werden die am Anfang gestellten Forschungsfragen aufgegriffen und beantwortet. Die Antworten zu diesen Fragen können gleichzeitig als Fazit betrachtet werden. Abgerundet wird die Arbeit mit einem Blick in die Zukunft und möglichen künftigen Forschungsschwerpunkten.

10.1 Beantwortung der Forschungsfragen

Forschungsfrage 1: Wie wirken sich die qualitativ unterschiedlichen Daten auf die Integration sowie auf nachfolgende Berechnungen aus?

Die unterschiedliche Netzkodierung der verschiedenen Strassennetze erwies sich für den verwendeten Matchingansatz als grösste Schwierigkeit. Entsprechende Qualitätseinbussen sind deren Folge. Konkret ausgewirkt hat sich dies bei der Erzeugung des Alternativensatzes. Hierbei mussten einige Strassenkanten ausgeschlossen werden.

Bei den GPS-Daten erwies sich die Zerstückelung der Daten und die daraus resultierenden kurzen Routen als arbeitserschwerend. Zurückzuführen ist dies auf den Verkehrsmittel- bzw. Wege-Identifikationsalgorithmus. Bei weiterer Verbesserung des Algorithmus sollten aber solche Fehler nicht mehr vorkommen. Die Lagegenauigkeit der GPS-Koordinaten verursachte keine besonderen Schwierigkeiten.

Forschungsfrage 2: Sind GPS-Daten für Untersuchungen über die Routenwahl von Fahrradfahrern geeignet?

Diese Frage kann klar bejaht werden. Die räumliche und temporale Auflösung solcher Daten ist besonders gut, weshalb sich vielfältige Modellierungsmöglichkeiten anbieten. Dessen Ergebnisse sind den *Stated Preference* Untersuchungen, welche auf hypothetischen Situationen beruhen, deutlich überlegen. Als einschränkend erwiesen sich die fehlenden soziodemographischen Informationen. Dies ist aber nicht ein Nachteil von GPS-Daten allgemein, sondern eine Besonderheit der hier verwendeten Daten. Die Limiten von GPS-Daten sollten jedoch nicht vergessen werden (vgl. Abschnitt 2.1.1). Besonders in dicht bebauten städtischen Gebieten treten ungewollte Signalreflexionen auf, welche herausgefiltert werden müssen.

Forschungsfrage 3: Sind die verwendeten Matching Algorithmen für die Daten dieser Arbeit geeignet?

Der verwendete Ansatz für das *Matching* der Strassennetze kann nicht als ungeeignet abgestempelt werden. Jedoch ist dieser stark abhängig von der Qualität der zu integrierenden Daten. Als erschwerend erwies sich die unterschiedliche Auflösung der verschiedenen Netze (vgl. Abb. 21). Schlecht aufgelöste Netzbereiche können auch zu Fehlern bei der späteren Zuordnung der GPS-Punkte führen. Dennoch erscheint das nach dem *Matching* erzeugte Netz für das Ziel dieser Arbeit brauchbar, da bei der Modellschätzung kleine Fehler untergehen. Interessant zu untersuchen wären die Integrationsergebnisse von anderen Matchingansätzen.

Für das Matching der GPS-Daten wurde der Algorithmus nach Marchal *et al.* (2006) verwendet. Dieser produziert allgemein gute Resultate, wobei die GPS-Punkte einen fließenden Datenfluss bilden müssen. Zerstückelte Daten (Datenlücken), wie sie hier teilweise vorkommen, können zu Matchingfehlern führen.

Forschungsfrage 4: Generiert ein einfacher Algorithmus einen für diese Arbeit brauchbaren Alternativensatz?

Diese Frage kann weder klar bejaht noch verneint werden. Als brauchbar gelten Alternativen, deren Länge gut variiert und trotzdem nicht erheblich von den gewählten Routen abweicht. Die hier erzeugten Routen erfüllen dieses Kriterium, wobei sie nebst dem Strassennetz von der Distanz zwischen den OD-Paaren abhängig sind. Je näher diese beisammen liegen, desto weniger Kanten liegen dazwischen und desto schwieriger ist es realistische Alternativen zu

generieren. Ein Vergleich mit einem komplizierteren Ansatz wäre sehr aufschlussreich und würde die Beantwortung der Frage erleichtern. Obwohl in der Praxis höchstwahrscheinlich nicht alle hier erzeugten Routen als realistische Alternativen angesehen werden, erlaubte der hier verwendete Alternativensatz erste vielversprechende Routenwahlmodelle aufzustellen.

Forschungsfrage 5: Hat die Ähnlichkeit zwischen den Alternativen einen positiven oder negativen Einfluss auf die Routenwahl von Fahrradfahrern?

Die Ähnlichkeit zwischen den Routen erwies sich als Nutzen verringernd. Ähnliche Routen werden negativ bewertet, weshalb Planer auf viele ähnliche Wege (in Bezug auf die physische Überlappung, vgl. Abschnitt 6) verzichten sollten. Stattdessen sollten bestehende Wege nach den weiter unten genannten Kriterien (vgl. Forschungsfrage 7) optimiert werden.

Trotz der klaren Resultate wäre ein Vergleich mit längeren Routen (grössere Distanz zwischen den OD-Paaren) sehr wichtig. Somit könnte die Relevanz dieser gefundenen Eigenschaft besser beurteilt werden.

Forschungsfrage 6: Wie könnte das geschätzte Modell verbessert werden?

Für die Beantwortung dieser Frage gibt es viele mögliche Antworten. Zum einen könnte die Qualität des Schätzdatensatzes verbessert werden, indem beispielsweise ein anderer Algorithmus für die Generierung des Alternativensatzes verwendet würde. Auch eine bessere Qualität der GPS-Daten und des Strassennetzes würden möglicherweise zu einem verbesserten Schätzdatensatz beitragen. Eine weitere Überlegung in diese Richtung wäre den Schätzdatensatz aufzuteilen und so für jeden Teil ein anderes Modell zu schätzen. So könnten unter Umständen verschiedene Modelle für verschiedene Routentypen erzeugt werden. Die Aufteilung nach Routentypen könnte beispielsweise nach den gefahrenen Höhenmetern erfolgen. So entstünden vielleicht Modelle für „Bergetappen“ und „Flachetappen“.

Die Qualität des Schätzdatensatzes könnte aber auch durch mehr berücksichtigte Eigenschaften erhöht werden. Das Vorhandensein von soziodemographischen Daten wäre ein diesbezügliches Beispiel.

Eine weitere Möglichkeit die Modellergebnisse zu optimieren, wäre der Gebrauch anderer Ansätze (z.B. NL, CNL, MMNL etc.). Zwingend für die Verwendung anderer Modelle ist aber die Berücksichtigung der Ähnlichkeit zwischen den Routen.

Ferner könnte eine Faktorenanalyse für die zu berücksichtigenden Variablen helfen. Die daraus bestimmten neuen synthetischen Variablen würden möglicherweise zu besseren Ergebnissen führen. Erschwerend wäre aber die Interpretation der Resultate, da die Variablen eine Kombination verschiedener Eigenschaften darstellen.

Forschungsfrage 7: Welches sind die wichtigsten Determinanten bei der Routenwahl und was für einen Einfluss haben diese bei der Routenwahl von Fahrradfahrern?

Die mit Abstand wichtigste Determinante für die Routenwahl von Fahrradfahrern ist die Routenlänge. Kurze und damit direkte Routen sind demgemäss besonders wichtig. Weiter folgen nach der Wichtigkeit geordnet die maximale Steigung, die Ähnlichkeit zwischen den Routen, die Anzahl Lichtsignalanlagen und der Anteil markierter Fahrradwege. Mit Ausnahme des Anteils Fahrradwege haben alle Routeneigenschaften einen negativen Einfluss. Eine für Fahrradfahrer ideale Route ist den Ergebnissen dieser Arbeit zufolge möglichst kurz, flach, eigenständig und kreuzt eine geringe Anzahl an Lichtsignalanlagen. Des Weiteren verfügt sie über einen hohen Anteil an Fahrradwegen.

Forschungsfrage 8: Sind die Ergebnisse dieser Arbeit mit Resultaten aus früheren Studien vergleichbar?

Auf ähnliche Ergebnisse bezüglich der Routenlänge kamen die Untersuchungen von Bovy und Stern (1990), Hochmair (2004) sowie Stinson und Bhat (2003). Diese Studien haben ebenfalls die Distanz als wichtige Determinante ausgemacht.

Ein häufig genanntes Kriterium ist auch der Anteil Fahrradwege (vgl. Dill und Carr, 2003; Hochmair, 2004; Shafizadeh und Niemeier, 1997; Stinson und Bhat, 2003; Tilahun *et al.*, 2007). Dieses ist auch in der hier vorliegenden Arbeit relevant, doch nicht so entscheidend wie in diesen früheren Studien.

Auf die Wichtigkeit von möglichst flachen Routen und damit die Vermeidung von Steigungen hingewiesen haben bereits Aultman-Hall *et al.* (1997) sowie Hochmair (2004). Diese Ergebnisse decken sich also mit den hier erhaltenen Resultaten.

In einigen früheren Studien (Aultman-Hall *et al.*, 1997; Axhausen und Smith, 1986; Bovy und Stern, 1990; Hochmair, 2004; Stinson und Bhat, 2003) wurden nebst anderen Eigenschaften der Strassenbelag und das Verkehrsvolumen als wichtig für die Routenwahl von Fahrradfahrer angegeben. Diese Eigenschaften erwiesen sich bei den Modellschätzungen dieser Arbeit

als nicht signifikant. Folglich können die Ergebnisse dieser früheren Untersuchungen nicht gestützt werden.

Referenzen, welche die Ähnlichkeit zwischen den Routen sowie die Anzahl Lichtsignalanlagen als Determinanten angeben, wurden keine gefunden. In dieser Hinsicht hat die hier vorliegende Arbeit eine neue Erkenntnis zu Tage gefördert.

Nicht vergleichbar sind soziodemographische Kriterien sowie andere Eigenschaften (z.B. die Landnutzungsart am Wegrand etc.), da solche Angaben in dieser Studie fehlten.

Zusammenfassend lässt sich folgern, dass die Ergebnisse dieser Arbeit sich durchaus im wissenschaftlichen Kontext einbetten.

10.2 Ausblick

Im vorangehenden Abschnitt wurden bereits einige Punkte für zukünftige Arbeiten genannt. Hier soll nochmals erwähnt werden, dass die hier gefundenen Erkenntnisse mit anderen Studien schwierig zu vergleichen sind. Dies weil die früheren Untersuchungen auf ganz anderen Voraussetzungen beruhen. Insofern müssten die künftigen Schwerpunkte auf weiteren Studien mit GPS-Daten liegen. Erkenntnisse aus *Stated Preference* Analysen gibt es bereits zur Genüge.

10.2.1 Empfehlungen für neue GPS-Studien

Ein Vorteil von neuen Studien mit GPS-Daten wäre, dass gewisse Einschränkungen der hier vorliegenden Daten behoben werden könnten. So könnten im Rahmen einer neuen Erhebung auch soziodemographische Informationen (Alter, Geschlecht etc.) einbezogen werden sowie Fehler bei der Identifikation der Verkehrsmittel vermieden werden. Der letzte Aspekt liesse sich beispielsweise gut durch eine Erhebung im Rahmen der Aktion „bike to work“ berücksichtigen. Als Probanden wären aber lediglich Personen geeignet, die eine Route vollständig mit dem Fahrrad bewältigen (nicht in Kombination mit öffentlichen Verkehrsmitteln). Besonders vielversprechend wären also die grosse Wahrscheinlichkeit, dass alle Beobachteten wirklich Fahrradfahrer sind und die mögliche Stichprobengrösse. Gemäss Pro Velo Schweiz (2008) hat die Aktion im letzten Jahr 607 Betriebe (über 33000 Teilnehmende) zur Teilnahme bewogen. Für dieses Jahr haben sich bereits 867 Betriebe angemeldet (Stand 08.05.2008). Nebst der Grösse der Stichprobe wäre auch die räumliche Ausdehnung sehr aussichtsreich. So

liessen sich mögliche Heterogenitäten zwischen städtischen und ländlichen Gebieten bei der Routenwahl untersuchen. Die Finanzierung einer solchen Datenerhebung müsste sicherlich auch beachtet werden und sollte nicht unterschätzt werden. Ein Blick auf die Sponsoren und Wettbewerbspreise der diesjährigen „bike to work“ Aktion zeigt jedoch, dass von Seite verschiedener Geldgeber Interesse besteht. Zudem könnte finanzielle Hilfe von Städten und Kantonen erwartet werden, da Erkenntnisse über die Routenwahl die Planungsarbeiten erleichtern würden. Weitere Informationen finden sich in Pro Velo Schweiz (2008).

Um Datenschutzverstösse zu vermeiden, sollte die Identität der Probanden stets unbekannt bleiben. Die Kombination von soziodemographischen Informationen wie das Alter, Geschlecht, Einkommensstärke etc. mit hausgenauen Adressdaten (vgl. nächster Abschnitt) ist hierbei besonders heikel. Über die genaue Adresse sowie das Geschlecht und Alter einer Person könnte möglicherweise auf den Namen dieser Person geschlossen werden. Bei einer Erhebung wie sie hier angesprochen wurde, würden die aufgezeichneten Fahrten vom Wohn- zum Arbeitsort erfolgen. Folglich müsste genaustens abgeklärt werden, dass die Anonymität der teilnehmenden Personen bestehen bleibt. Unproblematisch würden diese Daten, wenn die entsprechenden Personen die Verwendung solcher Daten bewilligen.

10.2.2 Reverse Geocoding

An dieser Stelle soll der Fokus wieder auf die hier vorliegenden GPS-Daten gelegt und die Frage aufgegriffen werden was aus diesen Daten noch herauszuholen wäre. Da soziodemographische Informationen fehlen und diese unmöglich rekonstruiert werden können, richtet sich diese Empfehlung auf die Anreicherung der Daten durch genaue Adressangaben.

Potenzial für die Zukunft hat das sogenannte *reverse geocoding*. Aus den daraus gewonnenen Adressdaten könnten möglicherweise Angaben über den Wohn- und Arbeitsort der Personen abgeleitet werden und somit der Fahrzweck ausgemacht werden. Datenschutzaspekte müssen nur beachtet werden, wenn auch soziodemographische Daten vorliegen, da in solchen Fällen die Anonymität nicht zwingend gewährleistet wäre (siehe Abschnitt weiter oben). Im nächsten Abschnitt soll das *reverse geocoding* etwas näher betrachtet werden.

Mit *reverse geocoding* wird die Umwandlung respektive Beziehung von geographischen Koordinaten zu Adressen bezeichnet. Im Gegensatz dazu steht das *geocoding*, wo von Adressen auf Koordinaten geschlossen wird.

Das Google Map API offeriert einen Geocodierungsdienst über die GClientGeocoder Klasse (Google 2007). Für die inverse Beziehung, das *reverse geocoding*, bietet das API jedoch keinen entsprechenden Dienst. Hierfür gibt es aber ein API, welches auf die beiden Klassen GClientGeocoder und GDirections beruht (Geominne 2007). Mit Hilfe dieser Klassen lassen sich von Koordinaten, welche im WGS84¹⁶-Koordinatensystem vorliegen, Adressen ableiten. Standardmässig wird für jeden Punkt die nächst liegende Strasse gesucht und dessen Adresse übernommen. Es ist aber möglich, dass für gewisse Punkte keine Adresse gefunden wird, dies weil der Punkt zu weit von einer Strasse entfernt ist oder weil das Strassennetz von Google unvollständig ist. Über die genaue Berechnungsgrundlagen und die maximale Distanz, das ein Punkt von einer Strasse entfernt sein kann, bestehen leider keine Informationen. Ferner sollte beachtet werden, dass das API für das *reverse geocoding* relativ neu ist, weshalb kleine Fehler nicht ausgeschlossen werden dürfen. Da Anfragen über die GDirections Klasse auf 10000 pro 24 Stunden und IP limitiert sind, ist dies auch für das *reverse geocoding* der Fall (Die Limite für die „normale“ Geokodierung beträgt 15000 pro 24 Stunden und IP). In der neuesten API Version besteht seit neustem die Möglichkeit die Koordinaten hausgenau zuzuordnen. Dies limitiert aber die Anfragehäufigkeit auf 2000 pro Tag und IP. Nähere Informationen finden sich in Google (2007b).

Die so erhaltenen Adressen könnten schliesslich mit dem Gebäude- und Wohnungsregister der Schweiz verglichen werden. Über die Häufigkeit einer vorkommenden Adresse könnte so möglicherweise der Zweck einer Fahrt abgeleitet werden. Damit die genannte Möglichkeit nicht bloss ein theoretisches Konstrukt bleibt, wurde der beschriebene Ansatz mit einem kleinen Testdatensatz erprobt. Nachfolgend ist kurz die technische Vorgehensweise erklärt.

Zuerst müssen die GPS-Punkte ins WGS84-Koordinatensystem transformiert und als csv-Datei gespeichert werden. Danach erfolgt das *reverse geocoding* über das Aufrufen einer HTML Datei, welche im Anhang zu finden ist. Diese benutzt wiederum eine Javascript-Datei (greversegeocoderv102.js), die für das eigentliche *reverse geocoding* zuständig ist. Die Javascript-Datei sowie die Testdaten finden sich in der Beilage-CD dieser Arbeit.

Die gefundenen Adressen werden in einem neuen Browserfenster angezeigt und können nachfolgend als Textdatei gespeichert werden (Achtung: Browser sollte für diese Anwendung Po-

¹⁶ Für die Umrechnung von WGS84-Koordinaten ins Schweizer Koordinatensystem gibt es eine Näherungslösung von der Swisstopo (http://www.swisstopo.ch/pub/down/basics/geo/system/ch1903_wgs84_de.pdf), Zugriff: 10.12.2007.

pup-Fenster zulassen). Es gilt zu beachten, dass die Punkte (mit Breiten- und Längenangaben), die HTML- sowie die Javascript-Datei am selben Ort gespeichert werden müssen.

Die daraus generierten Resultate erwiesen sich als vielversprechend. Aus Zeitgründen musste aber auf eine weitere Verfolgung dieses Ansatzes verzichtet werden. Dies ist der Grund, weshalb im Kapitel Ausblick auf diese Thematik eingegangen wurde. Für die Zukunft wäre es sehr interessant zu sehen, welches Potenzial im *reverse geocoding* steckt.

10.2.3 Verbesserung bestehender Software

In eine komplett andere Richtung zielen die Empfehlungen, die als Letztes angesprochen werden.

Bezüglich der verwendeten Software für die Ähnlichkeitsberechnung wäre es hilfreich, wenn durch BioRoute auch andere Ähnlichkeitsmasse als der PS-Faktor berechnet werden könnten. Dieser berücksichtigt nämlich lediglich die physische Überlappung der verschiedenen Routen. Wie in Abschnitt 2.5.2 erwähnt, können aber auch noch andere Ähnlichkeiten (z.B. gleiche Reisezeiten) auftreten. Weiter müsste die Gewichtung mit der Länge der Strassenkanten bei der Berechnung der verschiedenen Routeneigenschaften möglich werden.

Auch für die Schätzsoftware BIOGEME gäbe es hilfreiche Entwicklungsmöglichkeiten. Es wäre ideal, wenn alle Variablenkombinationen (linear, nichtlinear etc.) automatisch getestet und lediglich die besten Nutzenfunktionen ausgegeben würden. Somit könnte eine Menge Handarbeit gespart werden.

11 Literatur

- Anders, K.-H. und J. Bobrich (2004) MRDB approach for automatic incremental update, Vortrag, *ICA Workshop on Generalisation and Multiple Representation*, Leicester, August 2004.
- Aultman-Hall, L., F.L. Hall und B.B. Baetz (1997) Analysis of bicycle commuter routes using geographic information systems: Implications for bicycle planning, *Transportation Research Record*, **1578**, 102-110.
- Axhausen, K.W. und P. Keller (2003) Hinweise für die Erstellung von studentischen Berichten, *Arbeitsberichte Verkehrs- und Raumplanung*, **76**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.
- Axhausen, K.W. und R.L. Smith (1986) Bicyclist link evaluation: A stated preference approach, *Transportation Research Record*, **1085**, 7-25.
- Azevedo, J.A., M.E.O. Santos Costa, J.J.E.R. Silvestre Madera und E.Q. Vieira Martins (1993) An algorithm for the ranking of shortest paths, *European Journal of Operational Research*, **69** (1) 97-106.
- Badard, T. (1999) On the automatic retrieval of updates in geographic databases based on geographic data matching tools, *Bulletin du Comité français de cartographie*, **162** (17), 34-40.
- Bentley, J.L. und H.A. Mauer (1980) Efficient worst-case data structures for range searching, *Acta Informatica*, **13** (2) 155-168.
- Ben-Akiva, M.E. und M. Bierlaire (1999) Discrete choice methods and their applications to short term travel decisions, in W. Hall (Hrsg.) *Handbook of Transportation Science*, 5-35, Kluwer, Dordrecht.
- Ben-Akiva, M.E. und S.R. Lerman (1985) *Discrete Choice Analysis*, MIT Press, Cambridge.
- Ben-Akiva, M.E., M.J. Bergman, A.J. Daly und R. Ramaswamy (1984) Modeling inter-urban route choice behaviour, in J. Volmuller und R. Hamerslag (Hrsg.) *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, 299-330, VNU Science Press, Utrecht.
- Bernstein D. und A. Kornhauser A. (1996) An introduction to map matching for personal navigation assistants, <http://www.njtide.org/reports/mapmatchintro.pdf> , Zugriff: 15.12.2007.
- Bierlaire, M. (2005) An introduction to BIOGEME (Version 1.5), <http://transpor2.epfl.ch/biogeme/doc/tutorial.pdf> , Zugriff: 03.01.2008.

- Bierlaire, M. (2003) BIOGEME: a free package for the estimation of discrete choice models, Vortrag, 3rd Swiss Transport Research Conference, Ascona, März 2003.
- Bovy, P.H.L. und S. Fiorenzo-Catalano (2006) Stochastic route choice set generation: behavioral and probabilistic foundations, Vortrag, 11th International Conference on Travel Behaviour Research, Kyoto, August 2006.
- Bovy, P.H.L. und E. Stern (1990) *Route Choice and Way Finding in Transport Networks*, Kluwer, Dordrecht.
- Bundesamt für Statistik BFS (2008) Quartiergrenzen von Schweizer Städten, GEOSTAT-Datenbeschreibung, <http://www.bfs.admin.ch/bfs/portal/de/index/regionen/22/lexi.Document.98673.pdf>, Zugriff: 14.04.2008.
- Burrough, P.A. und R.A. McDonnell (1998) *Principles of Geographical Information Systems*, Oxford University Press, Oxford.
- Cascetta, E., F. Russo und A. Vitetta (1997) Stochastic user equilibrium assignment with explicit path enumeration : Comparison of models and algorithms, Vortrag, *Proceedings of the 8th IFAC Symposium on Transportation Systems*, 1078-1084, Chania, Juni 1997.
- Cascetta, E., A. Nuzzola, F. Russo und A. Vitetta (1996) A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks, in J. B. Lesort (Hrsg.) *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, 697-711, Pergamon, Oxford.
- Chen, W., M. Yu, Z.-L. Li und Y.-Q. Chen (2003) Integrated vehicle navigation system for urban applications, Vortrag, *Proceedings of the 7th International Conference on Global Navigation Satellite Systems (GNSS)*, 15-22, European Space Agency, Graz, April 2003.
- Chung, E. und A. Shalaby (2004) Development of a trip reconstruction tool to identify traveled links and used modes for GPS-based personal travel surveys, Vortrag, 83rd Annual Meeting of the Transportation Research Board, Washington D.C, Januar 2004.
- Czerniak, R.J. (2002) NCHRP Synthesis of Highway Practice 301: Collecting, Processing, and Integrating GPS Data into GIS, *NCHRP Report*, **301**, National Research Council, Washington D.C.
- De La Barra, T., B. Perez und J.Anez (1993) Multidimensional path search and assignment, Vortrag, 21st PTRC Summer Annual Meeting, Manchester, September 1993.
- Debreu, G. (1960) Review of R.D. Luce individual choice behavior, *American Economic Review*, **50** (1) 186-188.
- Devogele, T, C. Parent und S. Spaccapietra (1998) On spatial database integration, *International Journal of Geographical Information Science*, **12** (4) 335-352.

- Dias, E.S., A.J. Edwardes und R.S. Purves (2007) Analysing and aggregating visitor tracks in a protected area, Vortrag, 5th *International Symposium*, Enschede, Juni 2007.
- Dill, J. und T. Carr (2003) Bicycle commuting and facilities in major U.S. cities: If you build them, commuters will use them, *Transportation Research Record*, **1828**, 116-123.
- Domencich, T.A. und D. McFadden (1975) *Urban Travel Demand: A Behavioural Analysis*, North Holland, Amsterdam.
- Dugge, B. (2006) Ein simultanes Erzeugungs-, Verteilungs-, Aufteilungs- und Routenwahlmodell, Dissertation, Technische Universität Dresden, Institut für Verkehrsplanung und Straßenverkehr, Dresden.
- ESRI (2008) ArcAktuell, <http://esri-germany.de/news/arcaktuell/index.html> , Zugriff: 10.01.2008.
- ESRI (2007) ArcGIS Desktop Help 9.2, <http://webhelp.esri.com/arcgisdesktop/9.2> , Zugriff: 30.12.2007.
- Ester, M., H. Kriegel, J. Sander und X. Xu (1997) A density-based algorithm for discovering clusters in large spatial databases with noise, in M.O. Scholl und A. Voisard (Hrsg.) *Advances in Spatial Databases: 5th International Symposium, SSD '97 Berlin*, 226-232, Springer, Berlin.
- Geominne, N. (2007) Reverse Geocoder for Google Maps Documentation, <http://nicogoeminne.googlepages.com/documentation.html> , Zugriff: 10.12.2007.
- Google (2007) Google Maps API, <http://code.google.com/apis/maps> , Zugriff: 10.12.2007.
- Google (2007b) Google Maps API Terms of Service, <http://code.google.com/apis/maps/terms.html> , Zugriff: 10.12.2007.
- Gösseln, G.V. und M. Sester (2003) Semantic and geometric integration of geoscientific data sets with ATKIS - applied to geo-objects from geology and soil science, Vortrag, *ISPRS Commission IV Joint Workshop "Challenges in Geospatial Analysis, Integration and Visualization II"*, Stuttgart, September 2003.
- Gower, J.C. (1985) Measures of similarity, dissimilarity, and distance, in S. Kotz, N. L. Johnson und C. B. Read (Hrsg.) *Encyclopedia of Statistical Sciences*, **5**, 397-405, John Wiley & Sons, New York.
- Greenfeld, J.S. (2002) Matching GPS observations to locations on a digital map, Vortrag, 81st *Annual Meeting of the Transportation Research Board*, Washington D.C., Januar 2002.
- Hochmair, H. (2004) Decision support for bicycle route planning in urban environments, Vortrag, 7th *AGILE Conference on Geographic Information Science*, Heraklion, April 2004.
- Harvey, F. und K. Krizek (2007) *Commuter Bicyclist Behavior and Facility Disruption*, Research Services Section, Minnesota Department of Transportation, Minneapolis.

- Honey, S.K., W.B. Zavoli, K.A. Milnes, A.C. Phillips, M.S. White und G.E. Loughmiller (1989) Vehicle navigational system and method, United States Patent No., 4796191.
- Hunt, J.D. und J.E. Abraham (2007) Influences on bicycle use, *Transportation*, **34** (4) 453-470.
- Hyodo, T., N. Suzuki und K. Takahashi (2000) Modeling of bicycle route and destination choice behavior for bicycle road network plan, *Transportation Research Record*, **1705**, 70-76.
- Jo, T., M. Haseyamai und H. Kitajima (1996) A map matching method with the innovation of the Kalman filtering, *IEICE - Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **E79-A** (11) 1853-1855.
- JUMP (2007) The JUMP Projekt, <http://www.jump-project.org> , Zugriff: 29.11.2007.
- Kanton Zürich, Amt für Raumordnung und Vermessung (ARV) (2007) Geodaten des GIS-ZH, Strassennetz, http://www.giszh.zh.ch/internet/bd/arv/gis/de/gisdaten/nach_thema/strassennetz.html , Zugriff: 22.11.2007.
- Kim, W., G. Jee und J. Lee (2000) Efficient use of digital road map in various positioning for ITS, Vortrag, *IEEE Symposium on Position Location and Navigation*, San Diego, März 2000.
- Kim, S. und J. Kim (2001) Adaptive fuzzy-network based C-measure map matching algorithm for car navigation system, *IEEE Transactions on industrial electronics*, **48** (2) 432-440.
- Kim, S., J.-H. Kim und I.-H. Hyun (1998) Development of a map matching algorithm for car navigation system using fuzzy Q-factor algorithm, Vortrag, *World Congress Intelligent Transport System*, Seoul, Korea, Oktober 1998.
- Kowoma (2007) GPS-Infos, Fehlerquellen bei GPS, <http://www.kowoma.de/gps/Fehlerquellen.htm> , Zugriff: 30.11.2007.
- Krizek, K.J. und P.J. Johnson (2006) Proximity to trails and retail: Effects on urban cycling and walking, *Journal of the American Planning Association*, **72** (1) 33-42.
- Lüscher, P. und D. Burghardt (2006) Matching von Strassendaten stark unterschiedlicher Massstäbe und Aufbau einer MRDB, *Mitteilungen des Bundesamtes für Kartographie und Geodäsie*, **36**, 79-88.
- Mantel, D. und U. Lipeck (2004) Matching cartographic objects in spatial databases, Vortrag, *XXth Congress of the ISPRS*, Istanbul, Juli 2004.
- Marchal, F., J.K. Hackney und K.W. Axhausen (2006) Efficient map-matching of large GPS data sets - Tests on a speed monitoring experiment in Zurich, *Transportation Research Record*, **1935**, 93-100.
- MATSim (2008) Multi-Agent Transport Simulation Toolkit, <http://www.MATSim.org> , Zugriff: 01.02.2008.

- McFadden, D. (1974) Conditional logit analysis of qualitative choice-behaviour, in P. Zarembka (Hrsg.) *Frontiers in Econometrics*, 105-142, Academic Press, New York.
- Meng, Y. (2006) Improved positioning of land vehicle in ITS using digital map and other accessory information, *Dissertation*, Department of Land Surveying and Geoinformatics, Hong Kong Polytechnic University, Hong Kong.
- Najjar, M.E. und P. Bonnifait (2003) A roadmap matching method for precise vehicle Localization using belief theory and Kalman filtering, Vortrag, *11th International Conference in Advanced Robotics*, Coimbra, Juni 2003.
- Nielsen, O.A. und R.M Jørgensen (2004) Map matching algorithms for GPS-data – methodology and test on data from the AKTA roadpricing experiment in Copenhagen, Vortrag, *5th TRI annual Symposium on Transportation Analysis (TRISTAN V)*, Le Gosier, Juni 2004.
- Ochieng, W.Y., M.A. Quddus und R.B. Noland (2004) Map matching in complex urban road networks, *Brazilian Journal of Cartography* (Revista Brasileira de Cartografia), **55** (2) 1-18.
- Oi, K.I.Y und P.W Shuldiner (1962) *An Analysis of Urban Travel Demands*, Northwestern University Press, Evanston.
- Ortúzar, J. de D. und L.G. Willumsen (2001) *Modelling Transport*, 3. Auflage, John Wiley, Chichester.
- Park, D. und L.R. Rilett (1997) Identifying multiple and reasonable paths in transportation networks: a heuristic approach, *Transportation Research Record*, **1607**, 31-27.
- Pro Velo Schweiz (2008) Interessengemeinschaft Velo Schweiz, <http://www.pro-velo.ch> , Zugriff: 08.04.2008.
- Phuyal, B. (2002) Method and use of aggregated dead reckoning sensor and GPS data for map matching, Vortrag, *Proceedings of the Institute of Navigation (ION)*, Portland, Oregon, September 2002.
- PostGIS (2007) What is PostGIS, <http://postgis.refrations.net> , Zugriff: 30.11.2007.
- Prato, C. G. und S. Bekhor (2006) Applying branch & bound technique to route choice set generation, *Transportation Research Record*, **1985**, 19-28.
- Pyo, J.-S., D.-H. Shin und T.-K. Sung (2001) Development of a map matching method using the multiple hypothesis technique, Vortrag, *Proceedings of the 4th International IEEE Conference on Intelligent Transport Systems*, Oakland, August 2001.
- Quddus, M.A., W.Y. Ochieng und R.B. Noland (2007) Current map matching algorithms for transport applications: State-of-the art and future research directions, *Transportation Research Part C*, **15** (5) 312-328.

- Quddus, M. A., W.Y. Ochieng und R.B. Noland (2006) Map matching algorithms for intelligent transport systems applications, Vortrag, *13th World Congress on Intelligent Transport Systems and Services*, London, Oktober 2006.
- Quddus, M.A., R.B. Noland und W.Y. Ochieng (2006b) A high accuracy fuzzy logic-based map matching algorithm for road transport, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, **10** (3) 103-115.
- Quddus, M.A., W.Y. Ochieng, L. Zhao und R.B. Noland (2003) A general map matching algorithm for transport telematics applications. *GPS Solutions*, **7** (3) 157-167.
- Ramming, M.S. (2002) Network Knowledge and Route Choice, Dissertation, Massachusetts Institute of Technology, Cambridge.
- Schüssler, N. und K.W. Axhausen (2008) Identifying trips and activities and their characteristics from GPS raw data without further information, Vortrag, *8th International Conference on Survey Methods in Transport*, Annecy, May 2008.
- Schüssler, N. und K. W. Axhausen (2007) Recent developments regarding similarities in transport modeling, Vortrag, *6th Swiss Transport Research Conference*, Ascona, September 2007.
- Shafizadeh, K. und D. Niemeier (1997) Bicycle journey-to-work: Travel behavior characteristics and spatial attributes, *Transportation Research Record*, **1578**, 84-90.
- Shepherd, I. (1991) Information Integration in GIS, in P. Longley, M. Goodchild, D. Maguire und D. Rhind (Hrsg.) *Geographical Information Systems*, 337-360, First Edition, John Wiley, New York.
- Stadt Zürich, Tiefbauamt (2007) Velorouten - das städtische Veloroutennetz, http://www.stadt-zuerich.ch/internet/taz/home/mobilitaet/fuss_veloverkehr/velorouten.html , Zugriff: 22.11.2007.
- Stadt Zürich, Dienstabteilung Verkehr (2007) Knotenplan der Lichtsignalanlagen, <http://www.stadt-zuerich.ch/internet/pd/dav/home/strukturell/re.html> , Zugriff: 23.11.2007.
- Stigmar, H. (2005) Matching route data and topographic data in a real-time environment, Vortrag, *10th Scandinavian Research Conference on Geographical Information Science*, Stockholm, Juni 2005.
- Stinson, M.A. und C.R. Bhat (2003) Commuter bicyclist route choice: Analysis using a stated preference survey, *Transportation Research Record*, **1828**, 107-115.
- Swisstopo (2008) VECTOR 25 Landschaftsmodell, <http://www.Swisstopo.admin.ch/internet/Swisstopo/de/home/products/landscape/vector25.html> , Zugriff: 20.02.2008.

- Swisstopo (2008b) SWISSIMAGE,
<http://www.Swisstopo.admin.ch/internet/Swisstopo/de/home/products/images/ortho/swissimage.html> , Zugriff: 20.02.2008.
- Swisstopo (2008c) GG25 - Gemeindegrenzen 25,
<http://www.Swisstopo.admin.ch/internet/Swisstopo/de/home/products/landscape/gg25.html> , Zugriff: 21.02.2008.
- Swisstopo (2008d) Landeskarte 1:25000, Zeichenerklärung,
<http://www.Swisstopo.admin.ch/internet/Swisstopo/de/home/products/maps/national/25.html> , Zugriff: 21.02.2008.
- Swisstopo (2008e) DOM/DTM-AV,
http://www.Swisstopo.admin.ch/internet/Swisstopo/de/home/products/height/dom_dtm-av.html , Zugriff: 05.02.2008.
- Syed, S. und M.E. Cannon (2004) Fuzzy logic-based map matching algorithm for vehicle navigation system in urban canyons, Vortrag, *Proceedings of the Institute of Navigation (ION) National technical meeting*, San Diego, Januar 2004.
- Taylor, G., G. Blewitt, D. Steup, S. Corbett und A. Car (2001) Road reduction filtering for GPS-GIS navigation, *Transactions in GIS*, **5** (3) 193-207.
- Tilahun, N.Y., D.M. Levinson und K.J. Krizek (2007) Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey, *Transportation Research A*, **41** (4) 287-301.
- Tsui, S.Y.A. und A.S. Shalaby (2006) Enhanced system for link and mode identification for personal travel surveys based on global positioning systems, *Transportation Research Record*, **1972**, 38-45.
- Universität Rostock (2007) Geoinformatik-Service, <http://www.geoinformatik.uni-rostock.de/lexikon.asp> , Zugriff: 01.12.2007.
- Van der Zijpp, N.J. und S. Fiorenzo-Catalano (2005) Path enumeration by finding the constrained k-Shortest paths, *Transportation Research B*, **39** (6) 545-563.
- Veregin, H. (1999) Quality of spatial data, in P. Longley, M. Goodchild, D. Maguire und D. Rhind (Hrsg.) *Geographical Information Systems*, 177-189, Second Edition, John Wiley, New York.
- Vivid Solutions Inc. (2007) Roadmatcher,
<http://www.vividsolutions.com/products.asp?catg=spaapp&code=roadmatcher> , Zugriff: 21.11.2007.
- Volz, S. (2006) An iterative approach for matching multiple representations of street data, Vortrag, *ISPRS Workshop on Multiple Representation and Interoperability of Spatial Data*, Hannover, Februar 2006.

- Waldner, U. (2005) Automatic matching of large detailed street networks – An algorithm with curve and attribute affinity and its evaluation, Vortrag, *5th Swiss Transport Research Conference*, Ascona, März 2005.
- Walter, V. (1997) Zuordnung von raumbezogenen Daten - am Beispiel der Datenmodelle ATKIS und GDF, *Dissertation*, Fakultät Bauingenieur- und Vermessungswesen, Universität Stuttgart, Stuttgart.
- Warner, S.L. (1962) *Strategic Choice of Mode in Urban Travel: A Study of Binary Choice*, Northwestern University Press, Evanston.
- Weis, C. (2006) Routenwahl im ÖV, *Diplomarbeit*, IVT, ETH Zürich, Zürich.
- White, C.E., D. Bernstein und A.L. Kornhauser (2000) Some map matching algorithms for personal navigation assistants, *Transportation Research C*, **8** (1) 91-108.
- Wolf, J., S. Schönfelder, U. Samaga, M. Oliviera und K.W. Axhausen (2004) Eighty weeks of GPS traces: Approaches to enriching trip information, *Transportation Research Record*, **1870**, 46-54.
- Xiong, D. und J. Sperling (2004) Semiautomated matching for network database integration, *ISPRS Journal of Photogrammetry and Remote Sensing*, **59** (1-2) 35-46.
- Yang, D., B. Cai und Y. Yuan (2003) An improved map-matching algorithm used in vehicle navigation system, Vortrag, *Proceedings of the International IEEE Conference on Intelligent Transport Systems*, Shanghai, Oktober 2003.
- Zhao, Y. (1997) *Vehicle Location and Navigation System*, Artech House Publishers, Norwood.
- Zhang, M., L. Meng und H. Qian (2007) A structure-oriented matching approach for the integration of different road networks, Vortrag, *XXIII International Cartographic Conference (ICC)*, Moskau, August 2007.
- Zhang, M. und L. Meng (2006) Implementation of a generic road-matching approach for the integration of postal data, in *Proceedings of the 1st ICA Workshop on Geospatial Analysis and Modeling*, 141-154, Wien, Österreich, Juli 2006.
- Zhang, M., W. Shi und L. Meng (2006) A matching approach focused on parallel roads and looping crosses in digital maps, Vortrag, *International Symposium of Theoretical Cartography and Geoinformation Science*, Wuhan, Oktober 2006.
- Zhou, J. und R. Golledge (2006) A three-step general map matching method in the GIS environment: Travel/transportation study perspective, *International Journal of Geographical Information System*, **8** (3) 243-260.

12 Glossar

12.1 BIOGEME

BIOGEME (Bierlaire's Optimization Toolbox for GEV Model Estimation) ist ein objektorientiertes Softwarepaket für die Schätzung von diskreten Entscheidungsmodellen und im Speziellen für Generalized Extreme Value (GEV) Modellen. Obschon es auf Linux entwickelt wurde, existiert auch eine Version für Windows. Folgende Funktionen werden gegenwärtig vom Programm unterstützt:

- Schätzung von GEV Modellen
 - Multinomial Logit Modelle
 - Nested Logit Modelle
 - Cross-Nested Modelle
 - Netzwerk-GEV Modelle
- Schätzung von Binary Probit Modellen
- Kontinuierliche und diskrete Mischungen von GEV Modellen
- Panel Daten: Es ist möglich, spezifisch individuelle Zufallsterme zu schätzen
- Keine Skalierungsprobleme: Die Spezifikation der Nutzenfunktion ist vollständig unabhängig von der Verteilung der Fehlerterme. Der Nutzer braucht sich keine Sorgen über Skalenprobleme, welche beispielsweise in einschränkenden Parametern von Nested Logit Modellen vorkommen können, zu machen.
- Unterstützung nicht-linearer Nutzenfunktionen

Weitere Informationen sowie Literaturreferenzen sind zu finden in Bierlaire (2005).

12.2 BioRoute

BioRoute ist ein Dienstprogramm von BIOGEME und hilft die Eingabedateien für die Modellschätzung vorzubereiten. Es benötigt eine vollständige Beschreibung des Netzwerkes, Start- und Endpunkte von Routen (inklusive deren Alternativen) sowie eine Funktion für die

Ähnlichkeitsberechnung. Die Ausführung von BioRoute erzeugt die Modellspezifikations- und die Sample-Datei (Bierlaire 2005).

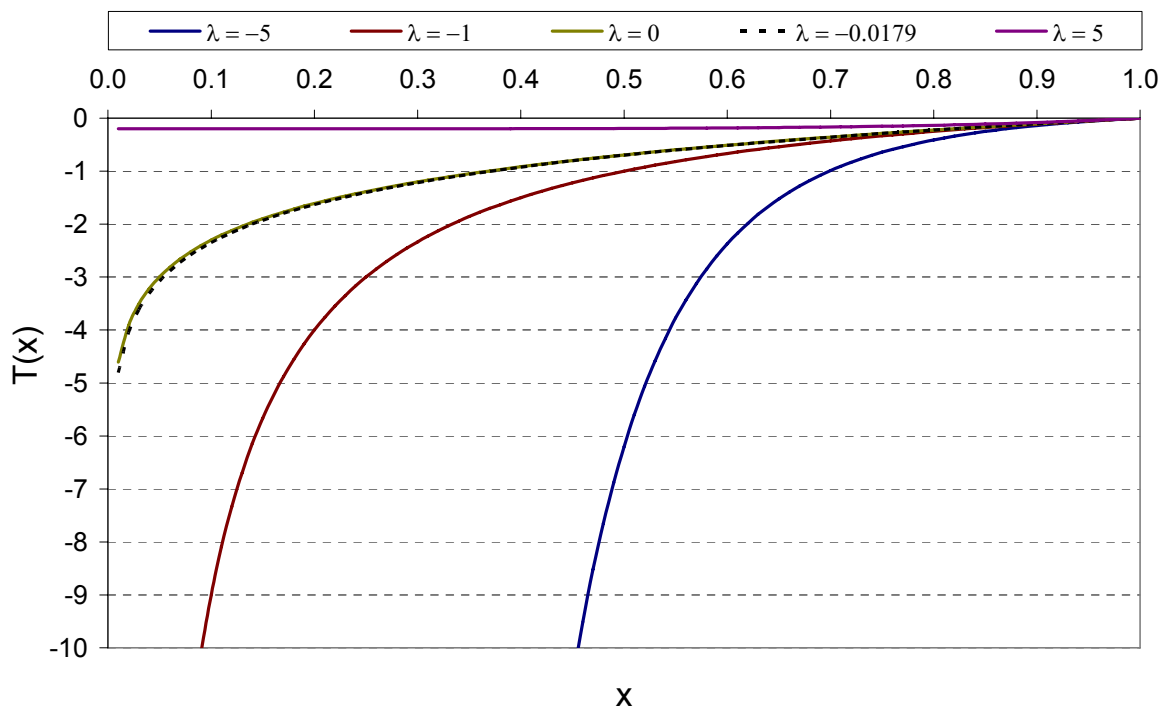
12.3 Box-Cox-Transformation

Die Box-Cox-Transformation ist wie folgt definiert:

$$T_{\lambda}(x) = \begin{cases} \frac{x^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \ln x & \lambda = 0 \end{cases} \quad (45)$$

Es ist ein Verfahren für die Linearisierung einer Variablen, welches das Fitting der Modellparameter verbessert. Die Korrelation einer Einflussgröße auf eine davon abhängige wird so verringert. Abbildung 64 illustriert den Verlauf für verschiedene Werte von λ . Ein λ -Wert von 0 erzeugt die logarithmische Funktion (Weis 2006).

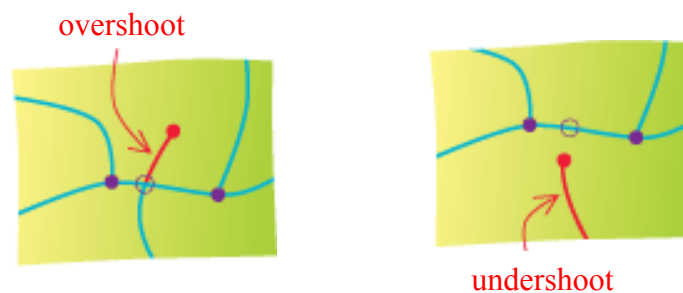
Abbildung 64 Box-Cox-Transformation für verschiedene λ



12.4 Dangling Node

Unter einem Dangling Node versteht man einen Endpunkt einer Linie, der keine Verbindung zu einer anderen Linie besitzt. Es wird als Fehler betrachtet, wenn die in Abbildung 65 illustrierten Situationen auftreten.

Abbildung 65 Dangling node



Quelle: ESRI (2007)

12.5 Gumbel-Verteilung

Die Gumbel-Verteilung ist eine stetige Wahrscheinlichkeitsverteilung mit folgender Verteilungsfunktion:

$$F(x) = e^{-x} \quad (46)$$

Die Ableitung der Verteilungsfunktion führt zur Dichtefunktion, welche wie folgt definiert ist:

$$f(x) = e^{-x} \cdot e^{-x} \quad (47)$$

Weitere wichtige Größen, welche nachfolgend erläutert werden, sind der Erwartungswert und die Varianz:

$$E(X) = \gamma \quad (48)$$

$$\gamma = 0.5772 \quad \text{Euler-Mascheroni-Konstante}$$

$$V(X) = \frac{\pi^2}{6} \quad (49)$$

Quelle: Weis (2006)

12.6 IIA-Eigenschaft

Die IIA-Eigenschaft (Independence of irrelevant alternatives) postuliert, dass die Alternativen unabhängig voneinander sind. Somit hängt die Auswahlwahrscheinlichkeit einer Alternative lediglich vom ihrem Nutzen ab. Einflüsse von anderen Alternativen werden ausser Acht gelassen. Zur Beiseitigung der IIA-Eigenschaft gibt es nach Schuessler und Axhausen (2007) drei verschiedene Ansätze:

- Gruppierung der Alternativen in Nests und Kombination einer der folgenden zwei Methoden
- Freigabe/Öffnung der Varianz-Kovarianz-Matrix
- Integration von Ähnlichkeitsfaktoren im systematischen Teil der Nutzenfunktion

12.7 Integration

In der vorliegenden Arbeit versteht man unter Integration das Kombinieren von Daten unterschiedlicher Herkunft und verschiedenartiger Struktur. Dabei beschreibt die Integration den Prozess der Abbildung verschiedener Daten in ein gemeinsames Datenmodell. Ziel ist es dem Anwender einen gleichartigen und gleichzeitigen Zugriff auf die Daten zu ermöglichen (Universität Rostock, 2007).

12.8 JUMP

Die Basis des JUMP-Projektes bilden eine Reihe von freien open source Anwendungen, welche eine graphische Benutzeroberfläche (GUI) für das Visualisieren und Manipulieren räumlicher Daten bereitstellen. Aktuell besteht das JUMP-Projekt aus folgenden drei Hauptmodulen:

- JUMP (Unified Mapping Plattform): Stellt einerseits die Hauptbenutzeroberfläche für das Visualisieren und Prozessieren räumlicher Daten sowie andererseits eine er-

weiterbare Programmierschnittstelle für die Entwicklung eigener Applikationen bereit.

- JTS (Java Topology Suite): Die Java Topology Suite liefert die Algorithmen für die Verarbeitung von Geodaten (z.B. Verschneidungen, Buffer-Bildung etc.). Das dabei verwendete räumliche Objektmodell ist OGC-konform.
- JCS (Java Conflation Suite): Die Java Conflation Suite besteht aus einem Set interaktiver Werkzeuge für die Verschmelzung (Matching) räumlicher Daten.

Verschiedene ebenfalls frei verfügbare Plug-Ins (z.B. der in dieser Arbeit benutzte Roadmatcher) optimieren die bestehenden Werkzeuge und Funktionen für bestimmte Anwendungsgebiete (JUMP, 2007).

12.9 MATSim

Das Multi-Agent Transport Simulation Toolkit MATSim bietet eine Toolbox, um grossskalige agenten-basierte Transportsimulationen zu implementieren. Die Toolbox besteht aus mehreren Modulen, die alleine oder kombiniert benutzt werden können. Des Weiteren können die Module durch eigene Implementierungen ersetzt werden. Zur Zeit bietet MATSim eine Toolbox für Nachfragemodellierungen, Mobilitätssimulationen, Neuplanungen, einen Controller um Simulationen iterativ ablaufen zu lassen sowie Methoden für die Analyse der Resultate (MATSim 2008).

12.10 Maximum Likelihood Schätzung durch BIOGEME

Die Schätzung von unbekanntem Parametern erfolgt vom Programm BIOGEME durch die Maximum Likelihood Methode. Folgende Annahmen werden dabei getroffen:

Eine Alternative k besteht aus einem Satz von Werten für die Attribute x_{in} , bezeichnet als x_{in}^k , und einer beobachteten Wahl. Die Attribute sind sowohl dem Individuum n als auch der Alternative i assoziiert. Die Wahrscheinlichkeit des Modells die gewählte Route zu reproduzieren ist wie folgt gegeben:

$$P_{in}^k(\alpha) = P_{in}(\alpha, x_{in}^k) \quad (50)$$

wobei P_{in} die Wahrscheinlichkeitsfunktion des betrachteten Modells und α der Vektor der zu schätzenden Modellparameter darstellt.

Steht nun einer ganzer Satz von Alternativen K zur Verfügung, so wird die Wahrscheinlichkeit den ganzen Alternativensatz zu reproduzieren Likelihood genannt und ist gegeben durch:

$$L(\alpha) = \prod_{k=1}^K P_{in}^k(\alpha) \quad (51)$$

Die genannte Funktion (51) wird optimiert bzw. maximiert. Hierbei wird deren natürlicher Logarithmus, die Log-Likelihood, als zu maximierende Grösse benutzt:

$$\ln L(\alpha) = \sum_{k=1}^K \ln P_{in}^k(\alpha) \quad (52)$$

Sind die Alternativen gewichtet, so wird in der Summenfunktion (52) zusätzlich ein Gewichtungsfaktor w_k eingefügt.

Zur Beurteilung der Güte des geschätzten Modells berechnet BIOGEME verschiedene Grössen:

- $L(\beta)$ = Log-Likelihood der geschätzten Modellparameter β
- $\rho^2 = 1 - \frac{L(\beta)}{L(0)}$

$L(0)$ ist die sogenannte Null Log-Likelihood. Diese berechnet die Log-Likelihood mit allen Parametern = 0. Je näher ρ^2 an 1 ist, desto besser ist die Güte des geschätzten Modells. Ein Wert von $\rho^2 = 0$ würde bedeuten, dass die Einführung der Variablen verglichen mit dem Nullmodell keine Verbesserung darstellt.

Weiter zu erwähnen gilt, dass bei bestimmten Modellen (z.B. NL oder CNL) gewisse Bedingungen bezüglich der Funktion (52) zu beachten sind. Diesbezüglich erlaubt BIOGEME die Festlegung von Einschränkungen in den Parametern (Bierlaire 2003; Weis 2006).

12.11 NMEA

NMEA steht für die Institution "National Marine Electronics Association" und ist ein Quasi-standardformat im Zusammenhang mit GPS-Daten. Die Daten enthalten Positions-, Richtungs-, Geschwindigkeitsangaben sowie Angaben zum GPS-Status und -Qualität (HDOP, PDOP). Das NMEA-Format wird von beinahe allen bekannten geodätischen und Navigationsempfängern unterstützt (Universität Rostock, 2007).

12.12 PostgreSQL und PostGIS

PostgreSQL ist ein relationales Datenbankverwaltungssystem, welches aus Ingres hervorgegangen ist und als open source verfügbar ist. Das Zusatzmodul PostGIS ist eine Erweiterung um räumliche Objekte (analog Oracle Spatial für Oracle) und wird kontinuierlich weiter entwickelt. Es unterstützt Geometriedaten gemäss der Simple Feature Spezifikation des OGC und ermöglicht räumliche Abfragen und Analysen (PostGIS, 2007).

12.13 Revealed und Stated Preferences

Revealed Preferences Daten beschreiben tatsächliches, empirisches Verhalten. Dazu werden Mobilitätsprotokolle oder -tagebücher von Testpersonen ausgefüllt. In der Verkehrsplanung werden die Mobilität im Allgemeinen und die Verkehrsmittelwahl im Besonderen üblicherweise in der Form der Revealed Preferences erhoben und analysiert.

Stated Preferences Untersuchungen machen hingegen Aussagen über das Verhalten unter Bedingungen, die in der Realität (noch) nicht existieren. Durch die Betrachtung von hypothetischen Entscheidungen können bei dieser Art von Daten grössere Attributbereiche und Bedingungen abgedeckt werden als in einem realen System.

Häufig werden beide Methoden in komplementärem Sinne benutzt, damit man die jeweiligen Stärken kombinieren kann (Ortúzar, 2001).

12.14 RMS - Fehler

Der RMS-Fehler (root mean square Fehler) misst die Differenz zwischen bekannten Positionen (Punktkoordinaten) und Positionen, die interpoliert oder digitalisiert wurden. Die Berechnung erfolgt gemäss folgender Formel:

$$RMS - Fehler = \sqrt{(x_r - x_i)^2 + (y_r - y_i)^2} \quad (53)$$

wobei x_r, y_r die Koordinaten der bekannten Punkte und x_i, y_i diejenigen der interpolierten Punkte darstellen (ESRI, 2007).

12.15 Roadmatcher

Der auf eine Java Bibliothek basierende Roadmatcher benutzt die JUMP Plattform, um eine interaktive GUI Umgebung mit reichhaltigen Visualisierungs- und räumlichen Manipulationsmöglichkeiten anbieten zu können. Das in dieser Arbeit relevante Matching von unterschiedlichen Objekten wird durch die Java Conflation Suite unterstützt. Hierfür gibt es auf Basis der JCS ein spezielles Modul für die Integration von linearen Objekten (Strassen), den hier benutzten Roadmatcher. Weitere Informationen sind in Vivid Solutions Inc. (2007) zu finden.

Anhänge

A 1 Datenbeschreibungen

Tabelle 32 Attributkatalog von VECTOR25

Feldname	Beschreibung	Feldtyp
FID	Feature ID (ArcGIS)	Integer
Shape	Geometrie	Geometrie
FNODE	Startknoten	Integer
TNODE	Endknoten	Integer
LPOLY	Linkes Polygon	Integer
RPOLY	Rechtes Polygon	Integer
LENGTH	Länge (ArcGIS)	Double
STR	Strada-Nummer	Integer
STR_ID	Strada-ID	Integer
BRDIDGETYPE	Brückentyp	Text
OBJECTID	Objekt ID	Integer
OBJECTORIG	Objektherkunft	Text
OBJECTVAL	Objektart	Text
STRADA_ID	Strada-DB Attribut	Text
TUNNELTYPE	Tunneltyp	Text
YEAROFCHAN	Nachführungsjahr	Integer

Tabelle 33 Attributkatalog des Strassendatensatzes vom Kanton Zürich

Feldname	Beschreibung	Feldtyp
FID	Feature ID (ArcGIS)	Integer
Shape	Geometrie	Geometrie
FNODE	Startknoten	Integer
TNODE	Endknoten	Integer
LPOLY	Linkes Polygon	Integer
RPOLY	Rechtes Polygon	Integer

LENGTH	Länge in m	Double
STRASS_2_5_TE	Internes Attribut (k.A.)	Integer
STRASS_2_6_TE-ID	Internes Attribut (k.A.)	Integer
ID	Identifikationsschlüssel (ID)	Integer
LINIENSCHL_2003	Linienschlüssel Verkehrsmodelldaten 2003	Integer
FVMODE_20003	Verkehrsmodellschlüssel für Startknoten	Integer
TVMODE_2003	Verkehrsmodellschlüssel für Endknoten	Integer
TYP_03	Streckentyp Verkehrsmodell	Integer
LAENGE_03	Streckenlänge in m aus Verkehrsmodell	Double
KAP_03	Max. Durchflussmenge in Fahrzeugen	Integer
DTV_03	Summe des durchschnittlichen Tagesverkehrs in Mfz aus beiden Richtungen 2003	Integer
DWV_03	Summe des durchschnittlichen Werktagsverkehrs in Mfz aus beiden Richtungen 2003	Integer
T_VO_03	Basisgeschwindigkeit in km/h (Reisegeschwindigkeit im unbelasteten Netz) 2003	Integer
F_VO_03	Basisgeschwindigkeit in km/h (Reisegeschwindigkeit im unbelasteten Netz; 2003)	Integer
F_DWV_03	Durchschnittlicher Werktagsverkehr in Mfz 2003	Integer
T_DWV_03	Durchschnittlicher Tagesverkehr in Mfz 2003	Integer
F_BEL_ASP_03	Belastung in Mfz in der Abendspitzenstunde 2003	Integer
T_BEL_ASP_03	Belastung in Mfz in der Abendspitzenstunde 2003	Integer
F_BEL_MSP_03	Belastung in Mfz in der Morgenspitzenstunde 2003	Integer
T_BEL_MSP_03	Belastung in Mfz in der Morgenspitzenstunde 2003	Integer
F_LW_ANT_DWV_03	Durchschnittlicher Lastwagenanteil am F_DWV in % 2003	Integer
T_LW_ANT_DWV_03	Durchschnittlicher Lastwagenanteil am T_DWV in % 2003	Integer
GEAENDERT	k.A.	Integer
RICHTPLANTYP	Aus Strassentyp und Eigentum berechnet, für Symbolisierung/Darstellung auf dem GIS- Browser	Integer
STRASSENNAME	Eindeutige Strassenamenbez. für National- und Staatsstrassen	Text
STRASSENTYP	Strassentyp	Integer
UNTERHALT	Für den Strassenunterhalt verantwortliche Stelle	Text
EIGENTUM	Eigentümer der Strasse	Text
ANSCHLUSS_NR_BUN	Anschlussnummer vom Bund	Text

PRIO	Internes Attribut für Druckpriorität	Integer
BFS	Gemeindenummer gemäss Bundesamt für Statistik	Integer
KT_RTE_KEY	Routenschlüssel für Routenthema «kantstreet»	Double
UNTERBRUCH	k.A.	Integer
VORTRITT_BRB_N	Strassennummern des Bundes	Text
DWV_98	Summe des durchschnittlichen Werktagsverkehrs in Mfz aus beiden Richtungen 1998	Integer
DWV_96	Summe des durchschnittlichen Werktagsverkehrs in Mfz aus beiden Richtungen 1996	Integer
AUSN_TRANSP_ROUT	Attribut für Ausnahmetransportrouten	Integer
TALSTRASSE1	Unterscheidung der Talstrassen für Abt. Störfallvorsorge	Double
TALSTRASSE2	Unterscheidung der Talstrassen für Abt. Störfallvorsorge	Double
UBEZ_NOK	k.A.	Integer
VM_PROB	Attribut für die Anpassung des VM-Netzes an das GIS-Netz	Text
RICHTPL2005	Attribut zum Darstellen des neuen Verkehrsplan für die Richtplanrevision 2005	Integer
ROUTE_SCHLUESSEL	Schlüssel für National- und Staatsstrassen Routen	Integer
RICHTPL	k.A.	Integer
TUNNEL	Bezeichnung für Tunnel	Integer
KT_RTE_K_1	k.A.	Text
KT_RTE_K_2	k.A.	Float
STRADAT_TY	k.A.	Text
STRADAT_NR	k.A.	Text
LENGTH_KM	Länge in km	Float
MEAS_TEXT	k.A.	Text

Tabelle 34 Attributkatalog von MAP Zürich

Feldname	Beschreibung	Feldtyp
FID	Feature ID (ArcGIS)	Integer
Shape	Geometrie	Geometrie
ANGLE	k.A.	Float
Map ZH Typ	Strassentyp gemäss Map ZH	Text

Tabelle 35 Attributkatalog der Fahrradwege vom kommunalen Richtplan

Feldname	Beschreibung	Feldtyp
FID	Feature ID (ArcGIS)	Integer
Shape	Geometrie	Geometrie
OBJECTID	Objekt ID	Integer
Shape_Leng	Länge (ArcGIS)	Double
id_geom_li	k.A.	Integer
OBJECTID_1	k.A.	Integer
FESTLEGUNG	k.A.	Integer
KATEGORIE	Strassenkategorie für die Darstellung	Integer
TEILPLAN	k.A.	Integer
BEZEICHNUN	k.A.	Integer
ORTSBEZUG	Ortsbezug/ Strassenname	Text
fk_geom_pu	k.A.	Integer
fk_geom_li	k.A.	Integer
fk_geom_fl	k.A.	Integer
fk_massnah	k.A.	Integer
mut_da	Datum für Veränderungen	Date
mut_user	Personenname für Veränderungen	Text
id_richtpl	Richtplan ID	Integer
OBJECTID_2	k.A.	Integer
id_bezeich	Bezeichnungs ID	Integer
bezeichn_1	Unterscheidung der Elemente des regionalen und kantonalen Richtplans	Text

Tabelle 36 Attributkatalog der GPS-Daten (für Import in PostGIS)

Feldname	Beschreibung	Feldtyp
PersonID	ID der Person	Integer
TripID	ID des Trips	Integer
SegmentID	ID des Segmentes	Integer
X	X-Koordinate	Double
Y	Y-Koordinate	Double
Z	Z-Koordinate	Double
Speed	Geschwindigkeit	Double
Time	Datum und Zeit	Double

Tabelle 37 Attributkatalog der nach dem Matching erhaltenen Routen

Feldname	Beschreibung	Feldtyp
GID	ID einer Strecke	Integer
NUMLINK	Position der Strecke in der Route	Integer
IDSEGMENT	ID der Route	Integer
IDLINK	ID der Strecke des Strassennetzes	Integer
FIRSTPOINT	ID des ersten GPS-Punktes, der auf eine Strecke gematched wurde	Integer
ENDPOINT	ID des letzten GPS-Punktes, der auf eine Strecke gematched wurde	Integer
ERROR	Fehler: Summe der Distanz (in m) aller Punkte zur Strassenkante	Double
POURCENT_Extremity	Prozentanteil der Strassenkante, die gefahren wurde (z.B. 0.5, falls U-turn in der Mitte der Strecke)	Double

A 2 Flussdiagramme für die Datenaufbereitung

Abbildung 66 Strassendaten: Topologieprüfung und Fehlerkorrektur

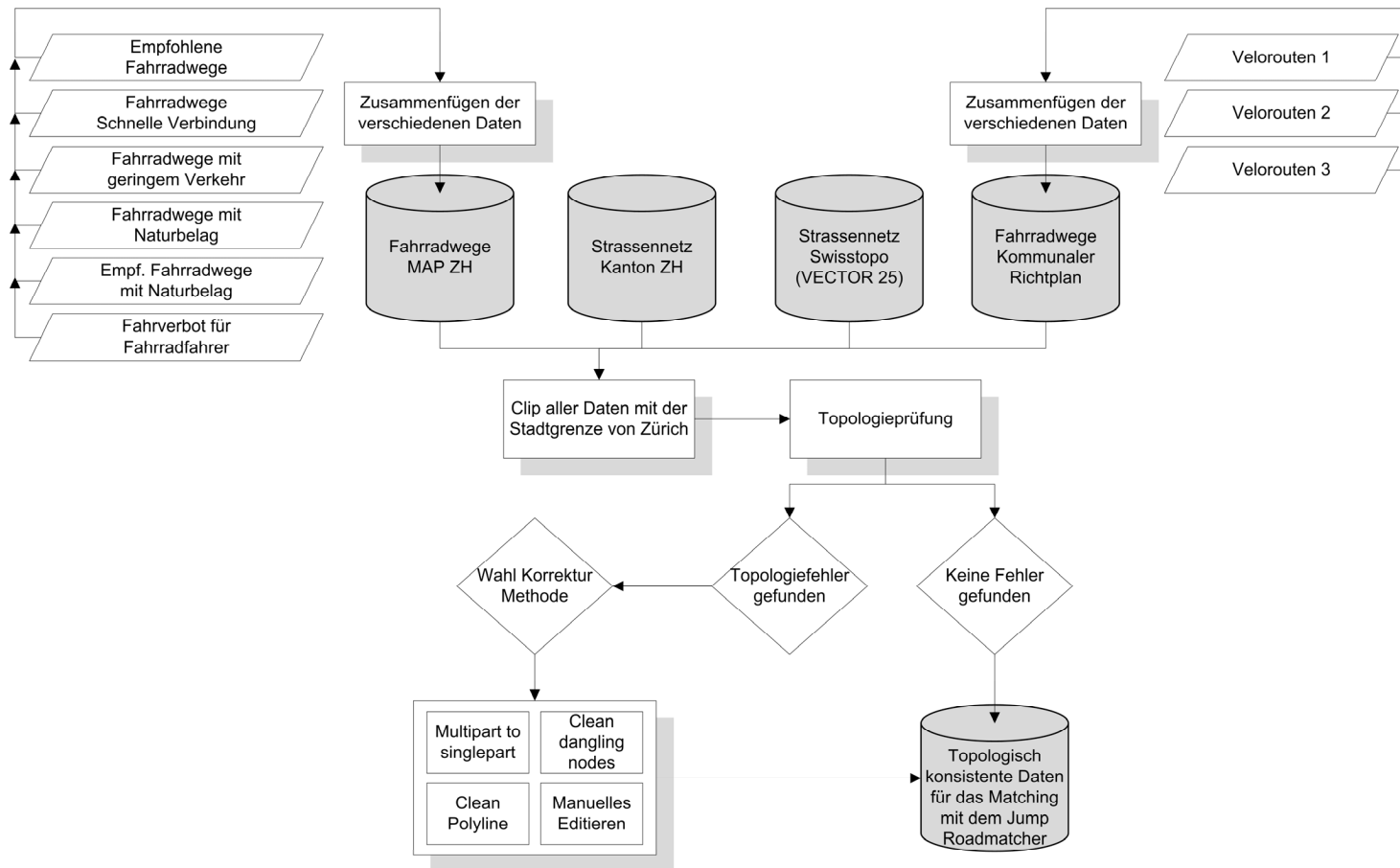


Abbildung 67 Matching der Strassennetze

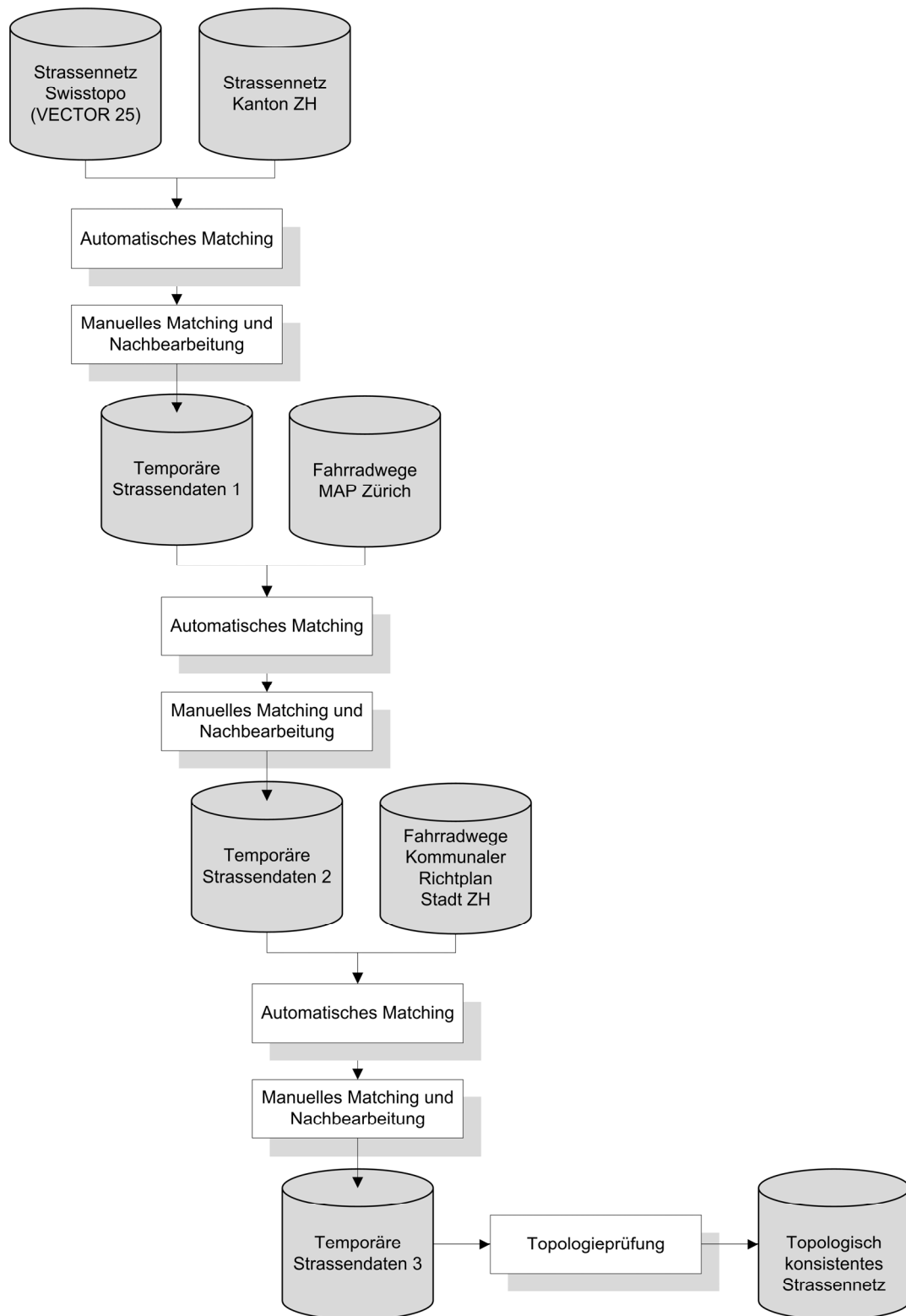


Abbildung 68 Allgemeine GPS-Daten Filterung

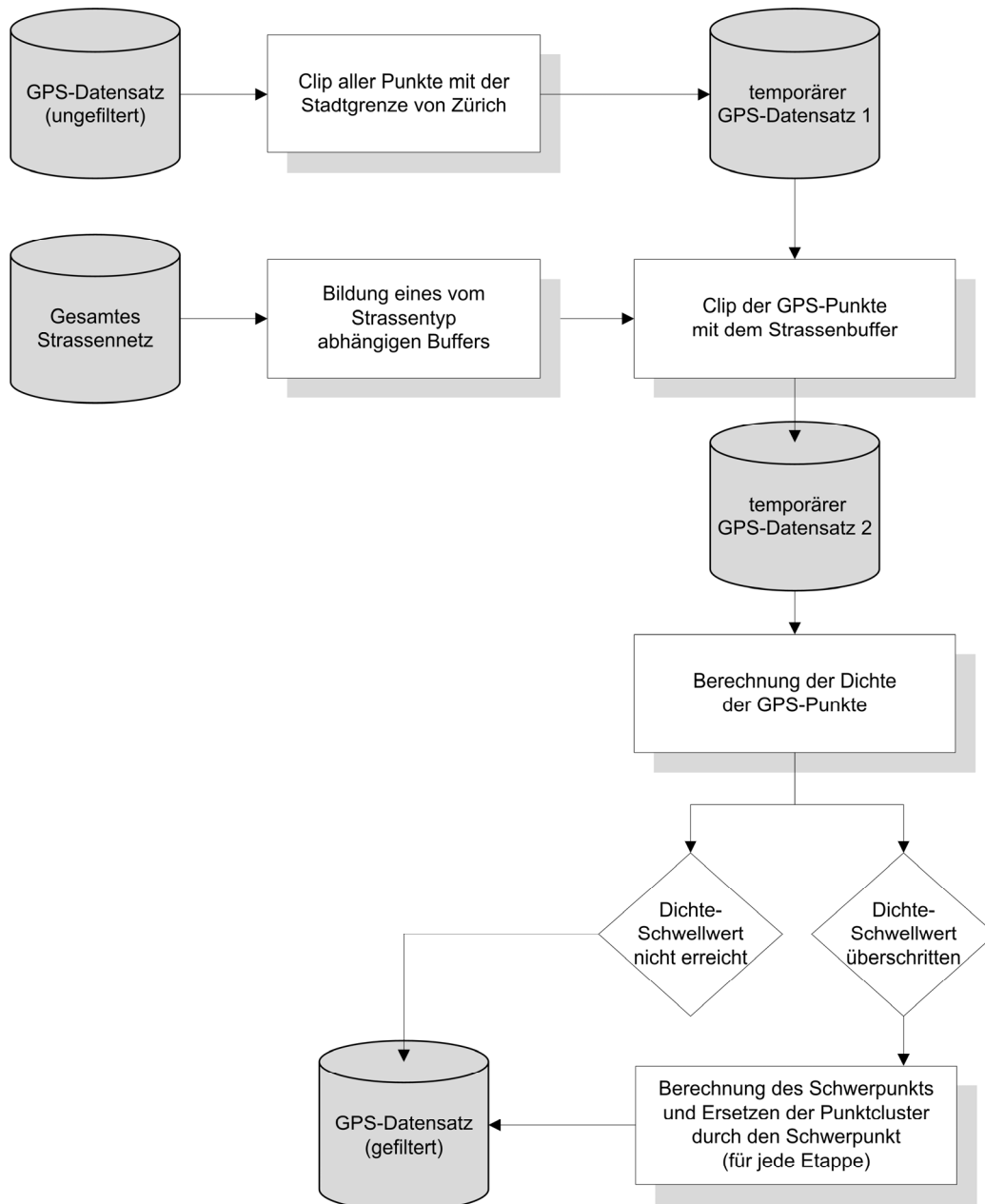


Abbildung 69 Spezielle GPS-Daten Filterung für Fahrradfahrer

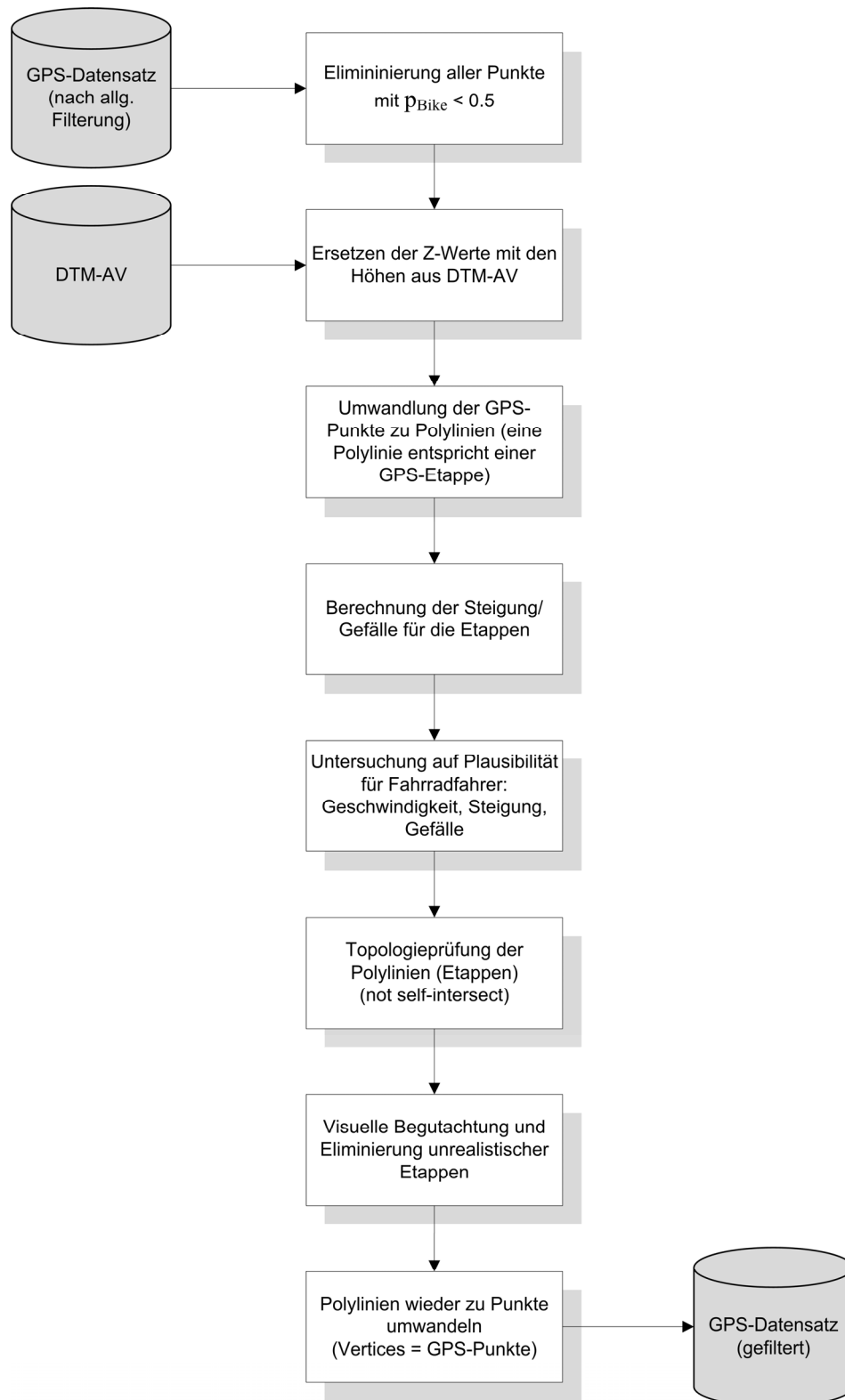
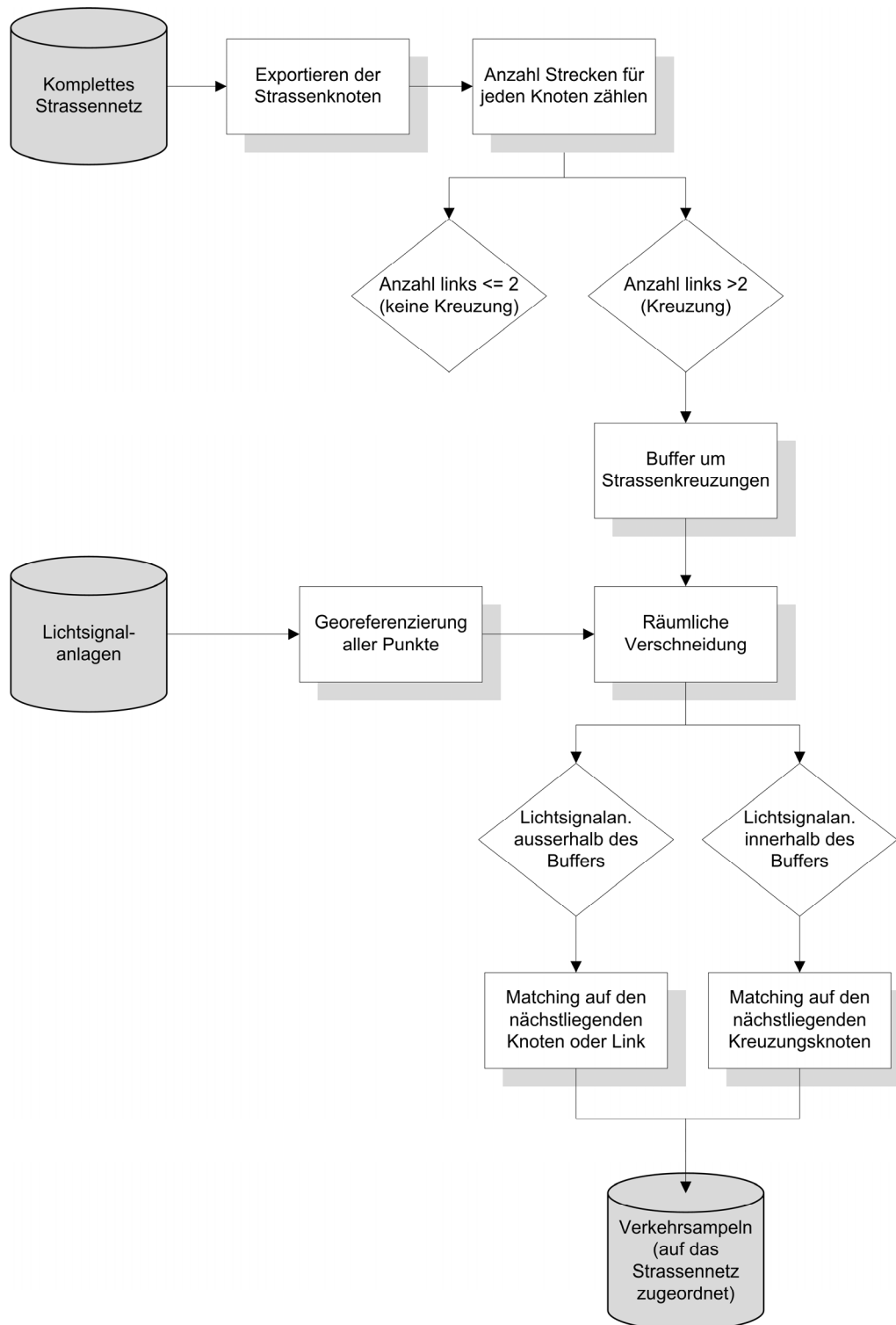


Abbildung 70 Matching der Lichtsignalanlagen



A 3 Statistische Resultate

Tabelle 38 Statistik der Modellvariablen für den gesamten Schätzdatensatz

	Min	Max	Mittelwert	Standardabw.
Length	10.810	3679.360	1330.910	656.519
RiseMax	0.000	0.720	0.063	0.064
TLights	0.000	14.000	2.138	2.515
Ln(PS)	-3.330	0.000	-1.634	0.561
BikeAv	0.000	1.000	0.617	0.243

Tabelle 39 Statistik der Modellvariablen für die gewählten Routen

	Min	Max	Mittelwert	Standardabw.
Length	68.160	989.840	467.037	143.942
RiseMax	0.000	0.240	0.021	0.035
TLights	0.000	3.000	0.533	0.887
Ln(PS)	-3.180	0.000	-1.249	0.594
BikeAv	0.000	1.000	0.681	0.349

Tabelle 40 Statistik der Modellvariablen für die Alternativen

	Min	Max	Mittelwert	Standardabw.
Length	10.810	3679.360	1384.106	626.183
RiseMax	0.000	0.720	0.065	0.065
TLights	0.000	14.000	2.218	2.543
Ln(PS)	-3.330	0.000	-1.653	0.553
BikeAv	0.000	1.000	0.614	0.236

Tabelle 41 Spearman Rangkorrelationskoeffizient für ausgeschlossene Variablen

	Length	RiseAv	RiseMax	FallAv	FallMax	Uphill	Downhill	DtvMax
Length	1.000	0.243	0.305	0.213	0.320	0.579	0.597	0.454
RiseAv	0.243	1.000	0.893	0.808	0.814	0.907	0.803	0.180
RiseMax	0.305	0.893	1.000	0.772	0.836	0.871	0.806	0.208
FallAv	0.213	0.808	0.772	1.000	0.896	0.787	0.879	0.144
FallMax	0.320	0.814	0.836	0.896	1.000	0.833	0.880	0.216
Uphill	0.579	0.907	0.871	0.787	0.833	1.000	0.933	0.325
Downhill	0.597	0.803	0.806	0.879	0.880	0.933	1.000	0.313
DtvMax	0.454	0.180	0.208	0.144	0.216	0.325	0.313	1.000

Tabelle 42 Spearman Rangkorrelationskoeffizient für die Modellvariablen

	Length	RiseMax	TLights	Ln(PS)	BikeAv
Length	1.000	0.305	0.012	-0.238	0.041
RiseMax	0.305	1.000	-0.306	-0.078	-0.280
TLights	0.012	-0.306	1.000	-0.082	0.215
Ln(PS)	-0.238	-0.078	-0.082	1.000	0.046
BikeAv	0.041	-0.280	0.215	0.046	1.000

Tabelle 43 Elastizitäten des Modells (14)

Variable	Min	Max	Mittelwert	Standardabw.
Length	-4.32	-0.17	-0.909	0.593
RiseMax	-2.15	0.00	-0.209	0.335
TLights	-0.66	0.00	-0.110	0.187
LnPS	-1.31	0.00	-0.501	0.239
BikeAv	0.00	0.68	0.322	0.176

Tabelle 44 Elastizitäten des Modells (26)

Variable	Min	Max	Mittelwert	Standardabw.
Length	-3.56	-0.14	-0.745	0.469
RiseMax	0.00	0.04	-0.004	0.006
TLights	-0.77	0.00	-0.125	0.211
LnPS	-1.37	0.00	-0.479	0.231
BikeAv	0.00	0.73	0.346	0.187
WRiseLength	-4.06	0.00	-0.221	0.464

Tabelle 45 Elastizitäten des Modells (38)

Variable	Min	Max	Mittelwert	Standardabw.
RiseMax	-1.82	0.00	-0.197	0.324
TLights	-2.66	0.00	-0.367	0.651
LnPS	-1.72	0.00	-0.514	0.316
BikeAv	0.00	0.55	0.249	0.153

Tabelle 46 Elastizitäten des Modells (39)

Variable	Min	Max	Mittelwert	Standardabw.
RiseMax	-4.72	0.00	-0.953	1.004
TLights	-6.24	0.00	-0.522	1.104
LnPS	-1.75	0.00	-0.529	0.323
BikeAv	0.00	0.43	0.195	0.120

A 4 Map matching Algorithmus nach Marchal *et al.* (2006)

In diesem Kapitel soll der in Marchal *et al.* (2006) vorgestellte Algorithmus detailliert beschrieben werden.

Als Erstes wird angenommen, dass $G(V, E)$ der gerichtete Graph des Strassennetzwerkes darstellt. Dabei ist V ein Satz von Knoten und E ein Satz von Kanten. Der GPS-Datenstrom, welcher aus einem Satz von Punkten besteht, wird $Q_i (i = 1..T)$ genannt. Jeder Punkt besteht aus einem Koordinatenpaar (x_i, y_i) und einem Zeitstempel t_i . Die Distanz von einem Punkt Q zur gerichteten Strecke AB wird gemäss White *et al.* (2000) gemessen und ist wie folgt definiert:

$$d(Q, AB) = \begin{cases} d_e(Q, Q') & \text{wenn } Q' \in [AB] \\ \min \{d_e(Q, A), d_e(Q, B)\} & \text{sonst} \end{cases} \quad (54)$$

Dabei ist d_e die euklidische Distanz und Q' die Projektion von Q auf AB .

Mit dieser Definition wird postuliert, dass $d_{Q,AB} = d_{Q,BA}$ ist und ein Punkt gleich weit entfernt von den zwei gegenläufig gerichteten Strecken eines Strassensegmentes ist. Deshalb definieren Marchal *et al.* (2006) eine minime parallele Verschiebung der gerichteten Pfade, welche die Fahrriichtung berücksichtigt. Dadurch wird bei der Berechnung der Distanz eine gerichtete Strecke AB durch $A'B'$ ersetzt, wo $A' = A + \lambda 1_{\perp}$, $B' = B + \lambda 1_{\perp}$ und $1_{\perp} \perp AB$ ist. λ wird verwendet, um die physische Distanz zwischen der Mitte der Strasse und der Mitte der Fahrspur zu repräsentieren (vgl. Abb. 71). Marchal *et al.* (2006) schlagen hier eine Distanz von 3m vor, wobei diese den eigenen Bedürfnissen anpassbar ist. Abbildung 72 illustriert die hier beschriebene Berechnung, wobei Q , Q_1 und Q_2 gleich weit entfernt sind von AB .

Abbildung 71 Distanz zwischen der Strassenmitte und Mitte der Fahrspur

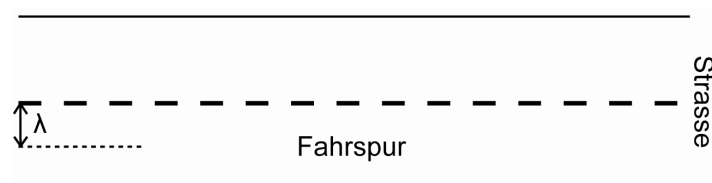
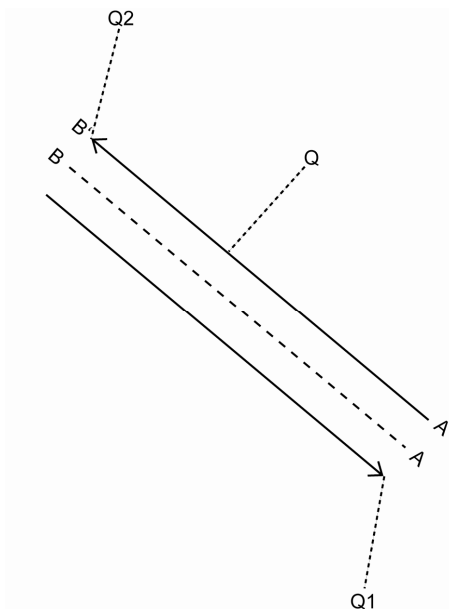


Abbildung 72 Distanz zwischen einem Punkt und einem Strassensegment



Quelle: Marchal *et al.* (2006)

Gewichtung eines Pfades

Der nächste Schritt beinhaltet die Gewichtung des Pfades P . $P\{E_1, E_2, \dots, E_p\}$ bezeichnet den Pfad, der aus p aufeinander folgenden Kanten E_1, E_2, \dots, E_p besteht. Der absolute Gewichtungswert F beim Matching der Sequenz Q_i ist definiert als

$$F = \sum_{j=1}^p \sum_{i=1}^T d(Q_i, E_j) \delta_{ij}, \quad (55)$$

wobei $\delta_{ij} = 1$ wenn Q_i der Strecke E_j zugeordnet wurde, ansonsten gilt $\delta_{ij} = 0$.

Das Problem besteht nun darin, den Pfad P^* zu finden, der F minimiert. Der relative Wert F_r wird definiert als $F_r = \frac{F}{T}$. Dieser wird benutzt für die Schätzung des Fehlers des Algorithmus. Es gilt dabei zu beachten, dass F_r stark abhängig ist von der Auflösung des Strassennetzwerkes. Des Weiteren wird jeder Punkt einer einzigen Kante des Pfades zugeordnet, so dass gilt

$$\sum_{j=1}^p \sum_{i=1}^T \delta_{ij} = T. \quad (56)$$

Initialisierung des Algorithmus

Der eigentliche Algorithmus beginnt mit dem Finden des Satzes S der N Strecken, die am nächsten zum ersten Datenpunkt Q_1 liegen. Somit wird als nächstes für jede Strecke in S ein sogenannter single-link Pfad erstellt. Die Pfade (P_1, P_2, \dots, P_N) werden dann entsprechend ihrem Gewichtungswert rangiert: $F_1 \leq F_2 \leq \dots \leq F_n$. Rein technisch gesehen, werden sie in einer sortierten Liste gespeichert. Weisen zwei Pfade denselben Gewichtungswert auf, so erfolgt eine alphabetische Rangierung.

Aufteilung von Routen in Pfade

Da im GPS-Datenfluss Unterbrechungen vorkommen können, wird in den meisten Fällen nicht eine ganze Route auf einen einzelnen kontinuierlichen Pfad zugewiesen. Wie in Abschnitt 2.1.1 bereits erwähnt, können Unterbrechungen durch Tunnels, dichte Baumkronen, schwache Signale etc. verursacht werden. Aus diesem Grund wird eine Route oftmals durch eine Sequenz von einzelnen Pfaden gematched.

Dabei sorgt die äussere Schleife des Algorithmus dafür, dass die Punkte sequentiell prozessiert werden. Wenn nun $d_e(Q_{i-1}, Q_i)$ grösser als ein vorgegebener Schwellwert (z.B. 300 Meter) oder $t_{i-1} - t_i$ länger als eine bestimmte Periode (z.B. 120 Sekunden) ist, dann wird der Matching-Prozess für den aktuellen Pfad unterbrochen und neu gestartet mit dem Punkt Q_{i+1} . Die beiden Schwellwerte können anfangs festgelegt und je nach Matchingresultat beliebig verändert werden.

Folgen der Netzwerk-Topologie

Für jeden neuen Punkt Q_i , wird der nachfolgend beschriebene Prozess angewendet:

1. Es wird angenommen, dass Q_i der letzten Kante E_p aus P zugeordnet wurde ($\delta_{ij} = 1$).
2. Der Wert des Pfades P wird gemäss (55) berechnet.
3. P wird in einer neu sortierten Liste V eingefügt.

4. Der letzte Prozessierungsschritt prüft, ob die Route den nächsten Knoten, welcher das Ziel E_p darstellt, erreicht hat.
 - Ist dies der Fall, werden die neuen Pfade P_k^{FS} für die nächste Kante E_p erstellt.
 - Der Wert der Pfade P_k^{FS} wird wieder gemäss (55) berechnet.
 - Die Pfade P_k^{FS} werden in der neu sortierten Liste V eingefügt.

Ist die Bedingung im 4. Schritt erfüllt, so werden neue Pfade erstellt, welche abhängig von der Netzwerk-Topologie sind.

Die Bestimmung, ob eine Route wirklich das Ende der letzten Strecke erreicht hat, ist aber immer noch nicht eindeutig. Marchal *et al.* (2006) benutzen deshalb die sogenannte crow-fly Distanz, welche die Punkte seit dem Beginn von E_p abdeckt. Dies um heraus zu finden, ob ein Fahrzeug (in diesem Falle ein Fahrradfahrer) möglicherweise die ganze Strecke gefahren ist.

Es wird nun angenommen, dass Q_{p_0} der erste Punkt ist, welcher E_p zugeordnet wurde und $L(E_p)$ die physische Länge von E_p ist. Wenn die folgende Bedingung erfüllt ist,

$$\sum_{k=p_0}^{i-1} d_e(Q_k, Q_{k+1}) > \alpha \cdot L(E_p) \quad (57)$$

dann kann angenommen werden, dass das Fahrzeug das Ziel von E_p erreicht hat. Ein Problem ergibt sich allerdings, wenn ein Gefährt einen u-turn in der Mitte einer langen Strecke macht. Um dies abzudecken, wird der Parameter $\alpha = 0.5$ eingeführt.

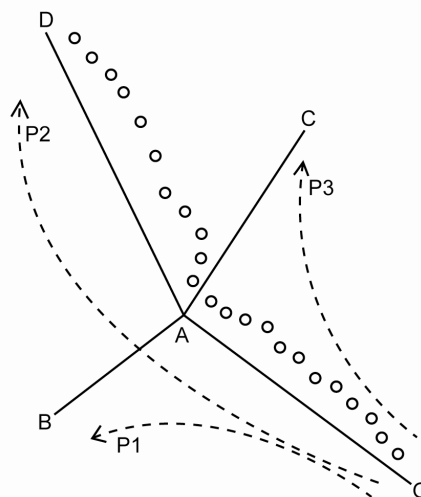
Als nächstes wird angenommen, dass $FS(E_p)$ der Satz von nachfolgenden Kanten von E_p (nächste Kante nachdem das Ende erreicht wurde) darstellt. Ist die Bedingung von (57) erfüllt, so wird ein neuer Pfad P_k^{FS} für jede Strecke in $FS(E_p)$ erzeugt. Diese Kind-Pfade unterscheiden sich von den Vorfahren dadurch, dass sie eine Strecke mehr besitzen:

$$E_{p+1}^k \in FS(E_p), \quad (58)$$

wo E_{p+1}^k die letzte Strecke des Pfades P_k^{FS} ist.

Die Kind-Pfade beginnen mit dem letzten Punkt, so dass $\delta_{ip+1} = 1$ gilt. Abbildung 73 illustriert die Erzeugung eines Pfades beim Fahren von O zu D. Bei der Schnittstelle A werden drei Kind-Pfade (P_1, P_2, P_3) erzeugt. Es ist natürlich klar, dass die ersten Punkte nahe bei A noch nicht Auskunft darüber geben, welche Richtung wirklich gewählt wurde, da das Netzwerk je nach Auflösung feine Krümmungen nicht repräsentieren kann. Welcher Pfad schlussendlich gewählt wurde, ergibt sich somit aus den berechneten Gewichtungswerten. In diesem Fall dominiert der Wert von P_2 diejenigen von P_1 und P_3 .

Abbildung 73 Pfaderzeugung bei Schnittstellen



Quelle: Marchal *et al.* (2006)

Beibehaltung eines Satzes von möglichen Pfaden

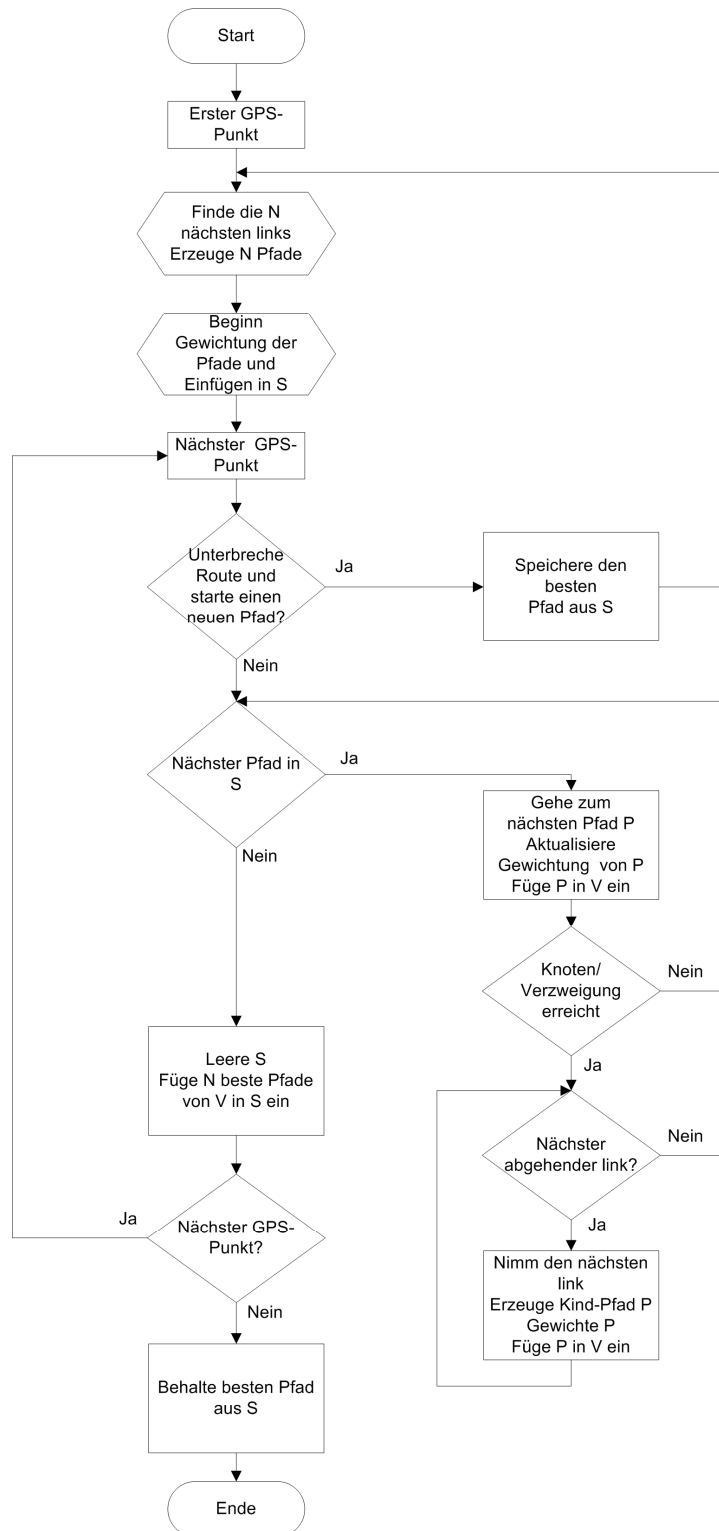
Kind-Pfade werden in der sortierten Liste V eingefügt und entsprechend ihrem Gewichtungswert rangiert. Infolgedessen hat V mehr Elemente als S . Wird Q_i prozessiert, so wird S geleert und nur die besten N Elemente von V werden in S eingefügt. Dies verhindert ein unbegrenztes Wachstum der sortierten Liste S . In der Praxis kann die Verzweigung und Pfaderzeugung für eine Reihe verschiedener Q_i geschehen, solange Bedingung (57) erfüllt ist. Die Grösse dieser Reihe ist abhängig vom Parameter α . Aus diesem Grund sind generell mehrere Instanzen von Pfaden P_1, P_2, P_3 in S , welche abhängig vom Punkt Q_i bei der Verzweigung sind.

Am Schluss werden die ineffizientesten Pfade verworfen. Die sortierte Liste S muss gross genug sein, damit auch diejenigen Pfade nachverfolgt werden, die anfangs noch nicht als richtig angenommen werden. Abbildung 73 zeigt sehr schön, dass die ersten Punkte AC als Richtung suggerieren würden. Erst nach der Betrachtung der anderen Punkte ist ersichtlich und auch durch den Gewichtungswert repräsentiert, dass die korrekte Richtung OD ist.

Bei der Prozessierung des letzten Punktes Q_i wird nun der beste Pfad aus S als Resultat des Matching Prozesses gespeichert. Analog erfolgt der Verlauf, wenn die Route unterbrochen ist und ein neuer Pfad gestartet wird. Auch hier wird nur der beste Pfad aus S als bestes Matching Ergebnis behalten.

Die nächstfolgende Abbildung 74 illustriert den gesamten Algorithmus graphisch. An dieser Stelle wird nochmals darauf hingewiesen, dass der Algorithmus von den beiden Parametern α (Parameter für die Berücksichtigung von u-turns) und N (Grösse des Satzes möglicher Kandidaten) abhängig ist.

Abbildung 74 Flussdiagramm des Map Matching Algorithmus



Quelle: Marchal *et al.* (2006)

A 5 Beispiel für die Syntax einer BioRoute Eingabedatei

Tabelle 47 Beispiel für die Syntax einer BioRoute Eingabedatei

[Nodes]

// Hier werden die Knoten des Strassennetzes definiert.

// Jede Zeile entspricht einem Knoten.

// ID Name

1 "Node 1"

2 "Node 2"

3 "Node 3"

4 "Node 4"

5 "Node 5"

6 "Node 6"

... ...

[ODpairs]

// An dieser Stelle sind die Start- und Endknoten der Routen angegeben.

// OrigID DestID

1 4

4 12

7 16

... ...

[Links]

// Hier werden die Strassenkanten mit ihren Attributen definiert.

// Die erste Spalte entspricht einer eindeutigen ID, es folgen die Start- und Endknoten der

// jeweiligen Kanten sowie deren Attribute.

//ID	FromNode	ToNode	Laenge	Steigung	VWeg	...
12	1	2	2.34	0.25	0	
13	2	4	1.55	0.02	0	
14	4	5	4.58	0.32	0	
15	5	6	8.39	0.03	0	
16	6	7	10.74	0.09	1	
17	7	12	6.36	0.12	1	
18	12	13	9.65	0.06	0	
19	13	16	26.03	0.00	1	

[OverlapMeasure]

// Hier wird angegeben wie die Überlappung (Ähnlichkeit) berechnet werden soll.

Laenge

[LinkNames]

// An dieser Stelle können Namen für die Strassenkanten vergeben werden.

// ID	Name
14	Strasse A
16	Strasse B

[Subpaths]

// Hier werden sogenannte Subpfade definiert.

// Subpfade können die Definition der eigentlichen Pfade erleichtern.

// ID	Subpfad_1	Subpfad_2	Subpfad_3	...
140	14	15	16	
155	17	18		

[ExplicitPaths]

// Hier werden alle Alternativen, inklusive der gewählten Routen definiert.

// Nach den Kanten bzw. Subpfaden können Attribute, wie beispielweise eine

// Dummy-Variable für die kürzeste Route, angegeben werden.

// ID	Strecke_1	Strecke_2	...	Shortest	...
10001	140	17		1	
10002	12	13		0	
10003	155	19		0	
...	

[Observations]

// Hier werden alle Beobachtungen (gewählte Routen) mit entsprechenden Alternativen definiert.

// Die erste Zahl entspricht der ID der gewählten Route, die nachfolgenden ID's

// entsprechen der ID der Alternativen. Ferner können am Ende weitere Attribute (meist soziodemographische

// Variablen wie z.B. das Alter der Person) definiert werden.

// ChoiceID	ID_Alt1	ID_Alt2	...	Alter	...
10001	10005	10008		25	
10002	10009	10012		36	
...	

[PathSize]

// Wert des Gammmaparameters, empfehlenswert nicht zu verändern.

// Für weitere Informationen vgl. Bierlaire (2005).

A 6 Beispiel einer BIOGEME Modelldatei

Tabelle 48 Beispiel einer BIOGEME Modelldatei

```
[ModelDescription]
// Hier kann das Modell mit Kommentaren beschrieben werden.
-----
[Choice]
// Hier wird definiert wie die Spalte des Schätzdatensatzes heisst, wo die ID der gewählten Route steht
choice
-----
[Beta]
// Hier werden die Namen, Startwerte sowie untere und obere Grenzwerte der
// Modellparameter definiert. Zudem kann über den Status angegeben werden, ob der
// Parameter vorgegeben (Status = 1) ist oder geschätzt werden soll (Status = 0).
//Name                Value                LowerBound            UpperBound            Status
CONST                 0.0                 -10000                10000                 0
BETA_length           0.0                 -10000                10000                 0
BETA_rise_max         0.0                 -10000                10000                 0
BETA_tlights          0.0                 -10000                10000                 0
...                   ...                 ...                   ...                   ...
-----
[Utilities]
// An dieser Stelle wird die lineare Komponente der Nutzenfunktion definiert.
// ID                Name                avail                linear expression
// Nr. der           Name der            Dummyvariable        Lineare Funktion
// Alternative       Alternative         „Alter. vorhanden“
1                   alt1               avail1              CONST * one + BETA_Length * x1_Length +
                    BETA_RiseMax * x1_RiseMax + BETA_TLights *
                    x1_TLights + BETA_BikeAv * x1_BikeAv
2                   alt2               avail2              CONST * one + BETA_Length * x2_Length +
                    BETA_RiseMax * x2_RiseMax + BETA_TLights *
                    x2_TLights + BETA_BikeAv * x1_BikeAv
...                   ...                 ...                 ...
-----
[GeneralizedUtilities]
// An dieser Stelle wird der nichtlineare Teil der Nutzenfunktionen angegeben.
//ID                non-linear expression
1                   BETA_pathSize1 * LOG (pathSize1)
2                   BETA_pathSize2 * LOG (pathSize2)
...                   ...
-----
```

[Expressions]

// Hier werden diejenigen Ausdrücke, die nicht im Datensatz definiert sind, angegeben

one = 1

[Model]

// Hier wird die zu berechnende Modellform definiert.

\$MNL // Multinomial Logit Modell

A 7 Pythonskript für die allgemeine GPS-Daten Filterung

```
# Tool Name: Filterung der GPS-Daten
# Beschreibung: Filterung der GPS-Daten aufgrund eines Strassenbuffers und der Dichte der GPS-Punkte
# Version: 1
# Autor: Gianluca Menghini
# Erforderliche Input-Daten: Stadtgrenze Zuerich, Strassennetzwerk der Stadt Zuerich, GPS-Datensatz mit Feld "PTS_ID"
# (PTS_ID = eindeutiger Schlüssel, welcher sich aus PersonID, TripID und SegmentID zusammensetzt)
# Erforderliche Programmversion: ArcGIS 9.2
# Erstellungsdatum: 1. November 2007
# Modifikationen: 5. November 2007

# Import der System Module
import sys, string, os, arcgisscripting, math
import traceback

# Erstellen des Geoprocessor Objektes
gp = arcgisscripting.create()

# Notwendiges Produkt festlegen
gp.SetProduct("ArcInfo")

# Notwendige Lizenzen bereitstellen
gp.CheckOutExtension("spatial")
gp.CheckOutExtension("3D")

# Laden der notwendigen Toolboxes
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Spatial Analyst Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Spatial Statistics Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Conversion Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Data Management Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/3D Analyst Tools.tbx")

#Festlegung der Input Parameter
gps_pt = sys.argv[1]
strassennetz = sys.argv[2]
grenze_zh = sys.argv[3]

# Festlegung lokaler Variablen
```

```
path1 = "C:\\Documents and Settings\\menghing\\My Documents\\Masterarbeit\\"
path = "C:\\Documents and Settings\\menghing\\My Documents\\Masterarbeit\\GPS_Points\\"
strassennetz_buffer = path1 + "strassennetz_buffer.shp"
strassennetz_select_1 = path1 + "strassennetz_select_1.shp"
strassennetz_buffer_1 = path1 + "strassennetz_buffer_1.shp"
strassennetz_select_2 = path1 + "strassennetz_select_2.shp"
strassennetz_buffer_2 = path1 + "strassennetz_buffer_2.shp"
strassennetz_select_3 = path1 + "strassennetz_select_3.shp"
strassennetz_buffer_3 = path1 + "strassennetz_buffer_3.shp"
strassennetz_select_4 = path1 + "strassennetz_select_4.shp"
strassennetz_buffer_4 = path1 + "strassennetz_buffer_4.shp"
strassennetz_select_5 = path1 + "strassennetz_select_5.shp"
strassennetz_buffer_5 = path1 + "strassennetz_buffer_5.shp"
strassennetz_buffer_dissolved = path1 + "strassennetz_buffer_dissolved.shp"
gps_pt_clip_zh = path + "gps_points_clip_zh.shp"
gps_pt_gefiltert = path + "gps_pt_gefiltert.shp"
gps_pt_dens = path + "gps_pt_dens"
gps_pt_den_re = path + "gps_pt_den_re"
#gps_pt_den_re_shp = path + "gps_pt_den_re_shp.shp"
density_layer = path + "density_layer"
gps_layer = path + "gps_layer"
gps_pts = path + "gps_pts.shp"
```

try:

Process: Clip der GPS-Punkte mit der Stadtgrenze von Zuerich

```
gp.AddMessage ("Clip der GPS-Punkte mit der Stadtgrenze von Zuerich")
gp.Clip_analysis(gps_pt, grenze_zh, gps_pt_clip_zh, "")
gp.AddMessage ("Ende des Clips mit der Stadtgrenze von Zuerich")
```

Beginn der ersten Filterung der GSP-Punkte aufgrund eines Strassenbuffer

```
gp.AddMessage ("Beginn der ersten Filterung der GSP-Punkte aufgrund eines Strassenbuffer")
```

Festlegung der anfaenglichen Buffergrosse auf 10 bzw. 20 m (10 m auf beiden Strassenseiten)

```
buffergrosse = 10
```

Process: Selektiere alle Strassen und mache einen Buffer von 20 m (dieser wird durch die spaeteren Schritte erweitert)

```
gp.AddMessage ("Buffer von 20 m um alle Strassen (dieser wird dann durch die anderen Schritte erweitert)")
gp.Buffer_analysis(strassennetz, strassennetz_buffer, buffergrosse, "FULL", "ROUND", "ALL", "")
```

Process: Selektiere alle 1.Klass Strassen und mache einen Buffer von 26 m

```
gp.AddMessage ("Selektiere alle 1.Klass Strassen und mache einen Buffer von 26 m")
sqlExpression = "\"OBJECTVAL\" = '1_Klass'"
if gp.Select_analysis(strassennetz, strassennetz_select_1, sqlExpression):
    gp.Buffer_analysis(strassennetz_select_1, strassennetz_buffer_1, buffergrosse + 3, "FULL", "ROUND", "ALL", "")
# Process: Append zum vorherigen Buffer
gp.Append_management(strassennetz_buffer_1, strassennetz_buffer, "NO_TEST")
```

Process: Selektiere alle 2.Klass Strassen und mache einen Buffer von 24 m

```
gp.AddMessage ("Selektiere alle 2.Klass Strassen und mache einen Buffer von 24 m")
sqlExpression = "\"OBJECTVAL\" = '2_Klass'"
if gp.Select_analysis(strassennetz, strassennetz_select_2, sqlExpression):
```



```
gp.Buffer_analysis(strassennetz_select_2, strassennetz_buffer_2, buffergroesse + 2, "FULL", "ROUND", "ALL", "")
# Process: Append zum vorherigen Buffer
gp.Append_management(strassennetz_buffer_2, strassennetz_buffer, "NO_TEST")

# Process: Selektiere alle 3.Klass Strassen und mache einen Buffer von 23 m
gp.AddMessage ("Selektiere alle 3.Klass Strassen und mache einen Buffer von 23 m")
sqlExpression = "\"OBJECTVAL\" = '3_Klass'"
if gp.Select_analysis(strassennetz, strassennetz_select_3, sqlExpression):
    gp.Buffer_analysis(strassennetz_select_3, strassennetz_buffer_3, buffergroesse + 1.5, "FULL", "ROUND", "ALL", "")
    # Process: Append zum vorherigen Buffer
    gp.Append_management(strassennetz_buffer_3, strassennetz_buffer, "NO_TEST")

# Process: Selektiere alle 4.Klass Strassen und mache einen Buffer von 22 m
gp.AddMessage ("Selektiere alle 4.Klass Strassen und mache einen Buffer von 22 m")
sqlExpression = "\"OBJECTVAL\" = '4_Klass'"
if gp.Select_analysis(strassennetz, strassennetz_select_4, sqlExpression):
    gp.Buffer_analysis(strassennetz_select_4, strassennetz_buffer_4, buffergroesse + 1, "FULL", "ROUND", "ALL", "")
    # Process: Append zum vorherigen Buffer
    gp.Append_management(strassennetz_buffer_4, strassennetz_buffer, "NO_TEST")

# Process: Selektiere alle Quartierstrassen und mache einen Buffer von 24 m
gp.AddMessage ("Selektiere alle Quartierstrassen und mache einen Buffer von 24 m")
sqlExpression = "\"OBJECTVAL\" = 'Q_Klass'"
if gp.Select_analysis(strassennetz, strassennetz_select_5, sqlExpression):
    gp.Buffer_analysis(strassennetz_select_5, strassennetz_buffer_5, buffergroesse + 2, "FULL", "ROUND", "ALL", "")
    # Process: Append zum vorherigen Buffer
    gp.Append_management(strassennetz_buffer_5, strassennetz_buffer, "NO_TEST")

# Process: Dissolve des gesamten Buffers
gp.AddMessage ("Dissolve des gesamten Buffers")
gp.Dissolve_management(strassennetz_buffer, strassennetz_buffer_dissolved, "", "", "MULTI_PART")

# Process: Clip der GPS-Punkte
gp.AddMessage ("Clip der GPS-Punkte")
gp.Clip_analysis(gps_pt_clip_zh, strassennetz_buffer_dissolved, gps_pt_gefiltert, "")

# Process: Loeschen temporaerer Daten
gp.AddMessage ("Loeschen temporaerer Daten")
gp.Delete_management(strassennetz_select_1, "FeatureClass")
gp.Delete_management(strassennetz_select_2, "FeatureClass")
gp.Delete_management(strassennetz_select_3, "FeatureClass")
gp.Delete_management(strassennetz_select_4, "FeatureClass")
gp.Delete_management(strassennetz_select_5, "FeatureClass")
gp.Delete_management(strassennetz_buffer, "FeatureClass")
gp.Delete_management(strassennetz_buffer_1, "FeatureClass")
gp.Delete_management(strassennetz_buffer_2, "FeatureClass")
gp.Delete_management(strassennetz_buffer_3, "FeatureClass")
gp.Delete_management(strassennetz_buffer_4, "FeatureClass")
gp.Delete_management(strassennetz_buffer_5, "FeatureClass")
#gp.Delete_management(strassennetz_buffer_dissolved, "FeatureClass")
#gp.Delete_management(gps_pt_clip_zh, "FeatureClass")
```

```
gp.AddMessage ("Ende der Filterung der GPS-Punkte anhand des Strassenbuffers")
```

```
# Beginn der zweiten Filterung der GPS-Punkte aufgrund der Dichte
```

```
gp.AddMessage ("Beginn der Filterung der GPS-Punkte aufgrund einer bestimmten Dichte")
```

```
# Process: Erzeugen einer neuen Feature Class, um die gefilterten Punkte dort zu speichern
```

```
gp.AddMessage ("Erzeugen einer neuen Feature Class")
```

```
gp.CreateFeatureclass_management(path, "gps_pts", "POINT", gps_pt_gefiltert, "DISABLED", "DISABLED",  
"PROJCS[CH1903_LV03',GEOGCS[GCS_CH1903',DATUM[D_CH1903',SPHEROID[Bessel_1841',6377397.155,299.  
1528128]],PRIMEM[Greenwich',0.0],UNIT[Degree',0.0174532925199433]],PROJECTION[Hotine_Oblique_Mercator_  
Azimuth_Center'],PARAMETER[False_Easting',600000.0],PARAMETER[False_Northing',200000.0],PARAME-  
TER[Scale_Factor',1.0],PARAMETER[Azimuth',90.0],PARAMETER[Longitude_Of_Center',7.439583333333333],PAR  
AMETER[Latitude_Of_Center',46.95240555555556],UNIT[Meter',1.0]],IsHighPrecision")
```

```
# Festlegung des Field Namens um eine Liste von Unique Values zu erstellen
```

```
fieldname = "PTS_ID"
```

```
# Search Cursor um alle Unique Values zu identifizieren
```

```
rows = gp.SearchCursor(gps_pt_gefiltert)
```

```
row = rows.Next()
```

```
# List-Variable, welche alle Unique Values beinhaltet
```

```
L = [ ]
```

```
# While loop, geht durch alle Records und fuegt Unique Values der List-Variable hinzu
```

```
while row:
```

```
    value = row.GetValue(fieldname)
```

```
    if value not in L:
```

```
        L.append(value)
```

```
    row = rows.Next()
```

```
# Sortieren der List Variable
```

```
L.sort()
```

```
# Wenn ein Wert in der Liste fehlt ist, dann wird er von der Liste geloescht
```

```
if '' in L:
```

```
    L.remove('')
```

```
# Loop durch die List-Variable
```

```
i = 0
```

```
for item in L:
```

```
# Process: Selektion der GPS-Punkte fuer eine bestimmte Route
```

```
gp.AddMessage ("Selektion der GPS-Punkte von Route " + str(L[i]))
```

```
sqlExpression = "\"PTS_ID\" = " + str(L[i])
```

```
gp.MakeFeatureLayer(path + "gps_pt_gefiltert" + ".shp", path + "gps_layer" + str(L[i]))
```

```
gp.SelectLayerByAttribute(path + "gps_layer" + str(L[i]), "NEW_SELECTION", sqlExpression)
```

```
# Process: Berechnung der Dichte der GPS-Punkte mit einem Suchradius von 20 m und einer Zellgroesse von 1 m
```

```
gp.AddMessage ("Berechnung der Dichte der GPS-Punkte von Route " + str(L[i]))
```

```
gp.PointDensity_sa(path + "gps_layer" + str(L[i]), "NONE", path + "gps_dens" + str(L[i]), "1", "Circle 20 MAP",
```



```
gp.Delete_management(path + "gps_pt_clip" + str(L[i]) + ".shp", "FeatureClass")
gp.Delete_management(path + "gps_pt_select" + str(L[i]) + ".shp", "FeatureClass")
gp.Delete_management(path + "gps_pt_erased" + str(L[i]) + ".shp", "FeatureClass")
gp.Delete_management(path + "centroid" + str(L[i]) + ".shp", "FeatureClass")
gp.Delete_management(path + "gps_dens" + str(L[i]), "RasterDataset")
gp.Delete_management(path + "gps_re" + str(L[i]), "RasterDataset")
```

Falls die obige Bedingung nicht erfuehlt ist, dann mache Folgendes:

else:

Process: Auswahl der ungefilterten Punkte

```
gp.AddMessage ("Auswahl der Punkte von Route " + str(L[i]) + " (hier wurden keine zu dicht beisammen liegende
Punkte gefunden)")
gp.Select_analysis(gps_pt_gefiltert, path + "gps_pt_select" + str(L[i]) + ".shp", sqlExpression)
```

Process: Append der ungefilterten Punkte

```
gp.AddMessage ("Append der ungefilterten Punkte von Route " + str(L[i]))
gp.Append_management(path + "gps_pt_select" + str(L[i]) + ".shp", gps_pts, "NO_TEST")
```

Process: Loeschen temporaerer Daten

```
gp.AddMessage ("Loeschen temporaerer Daten der Route " + str(L[i]))
gp.Delete_management(path + "gps_dens" + str(L[i]), "RasterDataset")
gp.Delete_management(path + "gps_pt_select" + str(L[i]) + ".shp", "FeatureClass")
```

i = i + 1

Abfangen moeglicher Fehler

except:

Fehlerhinweise

```
gp.addmessage(gp.GetMessages())
gp.addmessage("Error in line: " + str(traceback.tb_lineno(sys.exc_info()[2])))
gp.AddError(str(sys.exc_info()[1]))
gp.addmessage(gp.GetMessages(0))
print gp.GetMessages(0)
```

A 8 Pythonskript für die Generierung von Alternativen

```
# Tool Name: Berechnung von zufaelligen Routenalternativen
# Beschreibung: Berechnung zufaelliger Routen (Alternativen) fuer mehrere Start- und Endpunktbeziehungen
# Version: 1
# Autor: Gianluca Menghini
# Erforderliche Input-Daten: Strassennetz inkl. Netzwerk (als ArcGIS network), Start- und Endpunkte
# Erforderliche Programmversion: ArcGIS 9.2
# Erstellungsdatum: 30. Dezember 2007
# Modifikationen: 5. Januar 2008

# Import der Sytem Module
import sys, string, os, arcgisscripting, math
import traceback, random

# Erstellen des Geoprocessor Objektes
gp = arcgisscripting.create()

# Notwendiges Produkt festlegen
gp.SetProduct("ArcInfo")

# Notwendige Lizenzen bereitstellen
gp.CheckOutExtension("spatial")
gp.CheckOutExtension("3D")
gp.CheckOutExtension("Network")

# Laden der notwendigen Toolboxes
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Spatial Analyst Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Spatial Statistics Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Conversion Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Data Management Tools.tbx")
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/3D Analyst Tools.tbx")

# Festlegung der Input Parameter
strassennetz = sys.argv[1]
routen_start_ende = sys.argv[2]
strassennetz_network = sys.argv[3]

# Festlegung lokaler Variablen
path = "C:\\Documents and Settings\\menghing\\My Documents\\Masterarbeit\\routen_test\\"
alle_routen = path + "alle_routen.shp"
alle_stops = path + "alle_stops.shp"
strassennetz_Layer = path + "strassennetz_layer.lyr"
routen_start_ende_layer = path + "routen_start_ende_lyr.lyr"
routen_berechnet = path + "routen_berechnet.shp"
stops = path + "stops_berechnet.shp"

try:

    # Schleife fuer die Anzahl Berechnungen moeglicher Alternativen (hier 5 Durchlaeufer)
    i=0
```

```
for i in range(0,5):

    # Pfad fuer die Zwischenspeicherung temporaerer Routenlayer
    routen = path + "routen" + str(i) + ".lyr"

    # Process: Mache einen neuen Feature Layer fuer das Strassennetz sowie fuer die Anfangs- und Endpunkte
    gp.addmessage("Make Feature Layer")
    gp.MakeFeatureLayer_management(strassennetz, strassennetz_Layer)
    gp.MakeFeatureLayer_management(routen_start_ende, routen_start_ende_layer)

    # Anzahl Records fuer das Strassennetz zaehlen
    Nblinks = gp.GetCount_management(strassennetz)

    # Neue Schleife fuer die zufaellige "Verrauschung" des Strassennetzes
    j=0
    for j in range(1,Nblinks):

        # Selektiere die Strecke j
        gp.addmessage("Select layer by attribute")
        sqlExpression = "[OBJECTID_12] = " + str(j)
        gp.SelectLayerByAttribute_management(strassennetz_Layer, "NEW_SELECTION", sqlExpression)

        # Generierung der Zufallszahl (zwischen 0.0 und 1)
        randomNbr = random.random()
        gp.addmessage("Random Number = " + str(randomNbr))

        # Berechnung des Feldes "traveltime" (=Attribut fuer die Berechnung der kuerzesten Route) mit der Zufallszahl
        gp.addmessage("Calculate Field with random Number")
        gp.CalculateField_management(strassennetz_Layer, "traveltime", "[traveltime] + [traveltime]*" + str(randomNbr),
            "VB", "")

        j=j+1

    # Bildung des Netzwerkes mit dem "verrauschten" Strassennetz
    gp.addmessage("Build Network")
    gp.BuildNetwork_na(strassennetz_network)

    # Erzeuge einen neuen Routenlayer
    gp.addmessage("Make Route Layer")
    gp.MakeRouteLayer_na(strassennetz_network, routen, "traveltime", "USE_INPUT_ORDER", "PRESERVE_BOTH",
        "NO_TIMEWINDOWS", "traveltime", "ALLOW_UTURNS", "", "NO_HIERARCHY", "",
        "TRUE_LINES_WITH_MEASURES", "")

    # Anzahl Records des Start- und Endpunktedatensatzes zaehlen
    NbLocations = gp.GetCount_management(routen_start_ende)

    # Neue Schleife fuer das sequentielle Hinzufuegen der Anfangs- und Endpunkte
    k=0
    for k in range(0,NbLocations-1):

        # Selektiere die Punkte k (=Start) und k+1 (=Ende)
        gp.addmessage("Select layer by attribute")
```

```
sqlExpression2 = "\"FID\" = " + str(k) + " OR " + "\"FID\" = " + str(k+1)
gp.SelectLayerByAttribute_management(routen_start_ende_layer, "NEW_SELECTION", sqlExpression2)

# Fuege die Start- und Endpunkte den "Locations" hinzu
gp.addmessage("Add Locations")
gp.AddLocations_na(routen, "Stops", routen_start_ende_layer, "CurbApproach # 0;Attr_Length # 0;Name N #",
"5000 Meters", "", "strassennetz_def SHAPE;strassennetz_Junctions NONE", "MATCH_TO_CLOSEST",
"CLEAR", "NO_SNAP", "5 Meters")

# Berechne die kuerzeste Route (aufgrund des Attributs "traveltime")
gp.addmessage("Solve Routes")
gp.Solve_na(routen, "SKIP")

# Kopiere die erzeugten Routen und Stops
gp.addmessage("Copy Routes and Stops")
gp.CopyFeatures(routen + "\\Routes", routen_berechnet)
gp.CopyFeatures(routen + "\\Stops", stops)

# Fuege die Routen und Stops den Feature Classes "alle_routen" und "alle_stops" hinzu
gp.addmessage("Append Routes and stops")
gp.Append_management(routen_berechnet, alle_routen, "NO_TEST")
gp.Append_management(stops, alle_stops, "NO_TEST")

# Loesche temporaere Daten
gp.addmessage("Loesche temporaere Daten")
gp.Delete_management(routen_berechnet, "FeatureClass")
gp.Delete_management(stops, "FeatureClass")

k=k+2

# Loesche temporaere Daten
gp.addmessage("Loesche temporaere Daten")
gp.Delete_management(strassennetz_Layer)
gp.Delete_management(routen_start_ende_layer)

# Setze die Werte des Attributs "traveltime" auf die urspruenglichen Werte zurueck
gp.CalculateField_management(strassennetz, "traveltime", "[traveltime] = [Shape_Length]", "VB", "")

# Compact der Datenbank mit den Strassendaten
gp.compact(path + "strassennetz.mdb")

i=i+1

# Abfangen der Fehler
except:
gp.addmessage(gp.GetMessages())
gp.addmessage("Error in line: " + str(traceback.tb_lineno(sys.exc_info()[2])))
gp.AddError(str(sys.exc_info()[1]))
gp.addmessage(gp.GetMessages(0))
print gp.GetMessages(0)
```

A 9 Reverse geocoding

Das dargestellte Beispiel basiert weitgehend auf die Dokumentation von Geominne (2007).

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>GReverseGeocoder v1.0.3 Example</title>
    <script src="http://maps.google.com/maps?file=api&v=2&key=..." type="text/javascript"></script>
    <script src="greversegeocoderv103.js" type="text/javascript"></script>
    <script type="text/javascript">
```

```
var xmlhttp;
var eingabe;
var zeit;
var ausgabedat;
var ergebnis = "Lat;Lng;Address\n";
```

// Funktion fuer das Laden der CSV-Datei

```
function loadCSV(url){
    xmlhttp=null;
    if (window.XMLHttpRequest) {
        xmlhttp=new XMLHttpRequest();
    }
    else if (window.ActiveXObject) {
        xmlhttp=new ActiveXObject("Microsoft.XMLHTTP");
    }
    if (xmlhttp!=null){
        xmlhttp.onreadystatechange=state_Change;
        xmlhttp.open("GET",url,true);
        xmlhttp.send(null);
    }
    else {
        alert("Your browser does not support XMLHTTP.");
    }
}
```

// Funktion fuer die Statusaenderung

```
function state_Change(){
    if (xmlhttp.readyState==4){
        eingabe = xmlhttp.responseText;
        zeit = setInterval('processInput()', 2000);
    }
}
```

// Funktion fuer die Prozessierung der Inputdaten

// Auf jeder neuen Zeile sind lat und lng Werte eines Punktes gespeichert

```
function processInput() {
    var pos = eingabe.indexOf("\n");
    var latlng = eingabe;

    if (pos != -1) latlng = eingabe.substring(0,pos);
```



```
var lat = latlng.substring(0,latlng.indexOf(";"));
var lng = latlng.substring(latlng.indexOf(";")+1);
var punkte = new GLatLng(lat,lng);
rg.reverseGeocode(punkte);

if (pos != -1) eingabe = eingabe.substring(pos+1);
else {
    clearInterval(zeit);
    setTimeout('output()', 5000);
}
}

// Funktion fuer die Ausgabe der Resultate
function output(){
    ausgabedat=window.open("','Reverse Geocoder');
    ausgabedat.document.write(ergebnis);
    ausgabedat.document.close();
    ausgabedat.focus();
}

// GMap2 Objekt (Karte)
var map;
// GReverseGeocoder Objekt
var rg;
// Text Eingabe-Felder
var lat;
var lng;
// Resultat div
var info;

// Funktion fuer das Laden der Karte und reverse geocoding
function load() {
    if (GBrowserIsCompatible()) {
        lat = document.getElementById("lat");
        lng = document.getElementById("lng");
        info = document.getElementById("info");

        map = new GMap2(document.getElementById("map"));
        map.setCenter(new GLatLng(47.36905939398391, 8.542041778564453), 13);
        map.addControl(new GLargeMapControl());
        rg = new GReverseGeocoder(map);

        // Fuer Hausnummer genaue Geokodierung
        // rg.setExperimentalHouseNumber(true);

        // Listener um die Resultate hinzuzufuegen
        GEvent.addListener(rg, "load", goodergebnis);
        GEvent.addListener(rg, "error", badergebnis);

        loadCSV('gps.csv');
```

```
}
}

// wird mit einem Placemark aufgerufen, wenn die reverse geocode Anfrage erfolgreich war
// setzt das Resultat div
function goodergebnis(placemark) {
  var html = placemark.address + '<br />' + '<b>Country code:</b>' + placemark.AddressDetails.Country.CountryName
  Code;
  info.innerHTML = html;
  var temp = placemark.address;
  var lt = Number(placemark.RequestPoint.coordinates[1]);
  var ln = Number(placemark.RequestPoint.coordinates[0]);
  ergebnis += lt + ";" + ln + ";" + temp + '<br>';
}

// wird aufgerufen, wenn die reverse geocode Anfrage erfolglos war
function badergebnis() {
  info.innerHTML = "Unable to reverse geocode";
  var temp = "Unable to reverse geocode";
  var lt = Number(rg.getLastPoint().lat());
  var ln = Number(rg.getLastPoint().lng());
  ergebnis += lt + ";" + ln + ";" + temp + "\n";
}

// Erlangt die Inputs der lat lng Felder und macht das reverse geocoding
// Anfrage
function reverse(){
  var punkte = new GLatLng(lat.value,lng.value);
  rg.reverseGeocode(punkte);
}

</script>

</head>

<body onload="load()" onunload="GUNload()">
  <p><b><H2>Reverse Geocoding Beispiel durch Benutzung des GReverseGeocder v1.0.3</H2></b><br />
  Es sucht nach einem csv-File und macht ein Reverse Geocoding fuer jeden einzelnen GPS-Punkt.<br />
  Das csv-File muss gps.csv benannt werden und am gleichen Ort wie das HTML- und Javascript-File gespeichert sein.<br
  /><br />
  Nachdem alle Punkte prozessiert wurden, wird ein Pop-up Fenster mit den Resultaten aufgerufen.<br />
  Dies sollte nicht durch den Browser blockiert werden.<br />
<br />

<div id="info">
</div><div id="map" style="width: 500px; height: 500px;"></div>

</body></html>
```