

Berechnung von räumlich und zeitlich aufgelösten Durchschnittsgeschwindigkeiten aus FCD zur Effizienzsteigerung im Straßenverkehr

Lukas Spengler*, Marius Madsen, Marc Gennat

*Nachwuchsautor

Hochschule Niederrhein, SWK E² - Institut für Energietechnik und Energiemanagement,
Reinarzstr. 49, 47805 Krefeld, Germany, +49 2151 822 5122, lukas.spengler@hs-niederrhein.de, www.hs-niederrhein.de/swk-e2

Kurzfassung: Um die Energieeffizienz im Straßenverkehr in städtischen und ländlichen Regionen durch die Umstrukturierung vorhandener Mobilitätsformen oder die Etablierung neuer Mobilitätskonzepte zu steigern, fehlen in vielen Kommunen grundlegende Daten, wie zum Beispiel die räumlich und zeitlich aufgelöste Durchschnittsgeschwindigkeit im Straßenverkehr. Dieser Beitrag zeigt eine Methodik, wie diese Daten ermittelt werden können, ohne auf kostenintensive Messungen zurückgreifen zu müssen. Floating Car Data (FCD) werden an vielen Stellen durch Taxiunternehmen oder anderen Firmen mit Fahrzeugflotten aufgezeichnet. Diese Daten werden in der Regel mit einer minimal notwendigen Auflösung gespeichert, die abhängig von der Fahrgeschwindigkeit ist, um Speicherplatz zu sparen. Das führt dazu, dass der Abstand zwischen Datenpunkten und die Genauigkeit des Zeitstempels unzureichend ist für eine einfache arithmetische Berechnung oder ein lineares Gleichungssystem. Daher wird hier eine Methode gezeigt, die ein Parameterschätzverfahren und die Methode der kleinsten Quadrate [4] nutzt, um für jeden Straßenabschnitt zwischen zwei Kreuzungen stündlich die Durchschnittsgeschwindigkeit zu bestimmen.

Keywords: Mobilität, Graphen und Netzwerke, Parameterschätzverfahren, Floating Car Data

1 Einleitung

Verkehrsplanende müssen in der Regel auf der Basis von wenigen Verkehrsdaten Verbesserungen entwickeln und Entscheidungen treffen. Die Erstellung von aussagekräftigen Verkehrsmodellen auf mikroskopischer oder makroskopischer Ebene scheitert häufig an der Datenbasis, die nicht umfangreich genug ist und sehr kostenintensiv erhoben werden müsste.

Dieser Beitrag beschreibt eine Methode zur Berechnung von raum-zeitlichen Durchschnittsgeschwindigkeiten aus der Aufbereitung und Analyse von Floating Car Data (FCD), deren Zeitstempel nur minutenweise gespeichert wurden. Die Daten stammen aus den GPS-Tracks von 38 Taxis in Krefeld (Nordrhein-Westfalen) des Jahres 2019.

Solche Daten sind ebenso in anderen Kommunen verfügbar, sodass die vorgestellte Methode dabei helfen kann, die nutzbare Datengrundlage für Verkehrsplanende zu erhöhen. Die aus den Geschwindigkeiten resultierenden durchschnittlichen Reisezeiten eines jeden Straßenabschnitts können als Kantengewichte in Straßengraphen verwendet werden, um die schnellste Route zwischen zwei Punkten mit einem beliebigen Kürzester-Pfad-Algorithmus aus der Graphentheorie zu berechnen [1]. Darüber hinaus können die berechneten

Geschwindigkeitswerte mit den in OpenStreetMap (OSM) [5] hinterlegten erlaubten Höchstgeschwindigkeiten verglichen werden, um eine Visualisierung des langsam fließenden Verkehrs in Abhängigkeit von der Tageszeit zu ermöglichen.

2 Datenbasis

2.1 Floating Car Data

FCD sind eine Liste mit Geokoordinaten und dazugehörigem Zeitstempel, die in einem Fahrzeug aufgezeichnet werden. Wenn die Daten in einem bestimmten Bereich für längere Zeit in kurzen Intervallen aufgezeichnet werden, können damit Durchschnittsgeschwindigkeiten für die jeweiligen Orte und Zeiträume berechnet werden.

Für das in diesem Beitrag gezeigte Beispiel hat ein Krefelder Taxiunternehmen insgesamt 17 Millionen Datenpunkte von 38 Fahrzeugen bereitgestellt, die neben der GPS-Koordinate und dem Zeitstempel auch die aktuelle Geschwindigkeit und den Besetzungsgrad des Taxis aufgezeichnet haben. Das Intervall, mit dem die Daten aufgenommen wurden, variiert stark in Abhängigkeit von der Geschwindigkeit. Die Datenpunkte wurden entweder im Abstand von 150 oder 300 Meter aufgenommen oder minütlich, wenn das Fahrzeug steht. Die Daten wurden in dem Zeitraum vom 01.01.2019 bis zum 31.12.2019 aufgezeichnet und beschreiben daher das Verkehrsgeschehen vor der Corona-Pandemie.

Aufgrund der un stetigen Datenerhebung können die aufgezeichneten Momentangeschwindigkeiten nicht einfach für jeden Abschnitt gemittelt werden. Außerdem sind die Zeitstempel in dem vorliegenden Beispiel nur minütlich gespeichert worden, sodass eine Durchschnittsberechnung zwischen zwei Zeitpunkten zu ungenau ist.

Ein Parameterschätzverfahren kann das Problem für un stetige und ungenaue FCD lösen, indem mehrere Datenpunkte zu einer Tour zusammengefasst werden.

Nichtbefahrene Straßen können in der Berechnung nicht berücksichtigt werden. Allerdings kann angenommen werden, dass Straßen mit ähnlichen Eigenschaften die gleiche Durchschnittsgeschwindigkeit aufweisen.

In Abbildung 1 sind die Datenpunkte für nur ein Fahrzeug in einem Monat dargestellt. Alle wichtigen Straßen, die zum einen in der Satzung der Straßenreinigung [6] als häufig frequentiert gelistet sind und zum anderen in OpenStreetMap [5] den Tag „highway=secondary“ oder höher aufweisen, wurden befahren.

2.2 Straßennetz

Das Straßennetz wird als gewichteter und gerichteter Graph modelliert, mit Kreuzungen als Knoten und Straßenabschnitten als Kanten auf Basis von OpenStreetMap [5]. Die Erstellung des Graphen erfolgt wie in [7] beschrieben. Die Gewichtung einer Kante entspricht der Länge zwischen den beiden verbundenen Knoten in Metern. Zusätzlich sind Informationen zum Straßentyp, Straßename und erlaubter Höchstgeschwindigkeit hinterlegt. Das ermöglicht die Berechnung der tatsächlichen Fahrdistanz zwischen zwei Punkten unter Berücksichtigung von Einbahnstraßen und Abbiegerestriktionen.

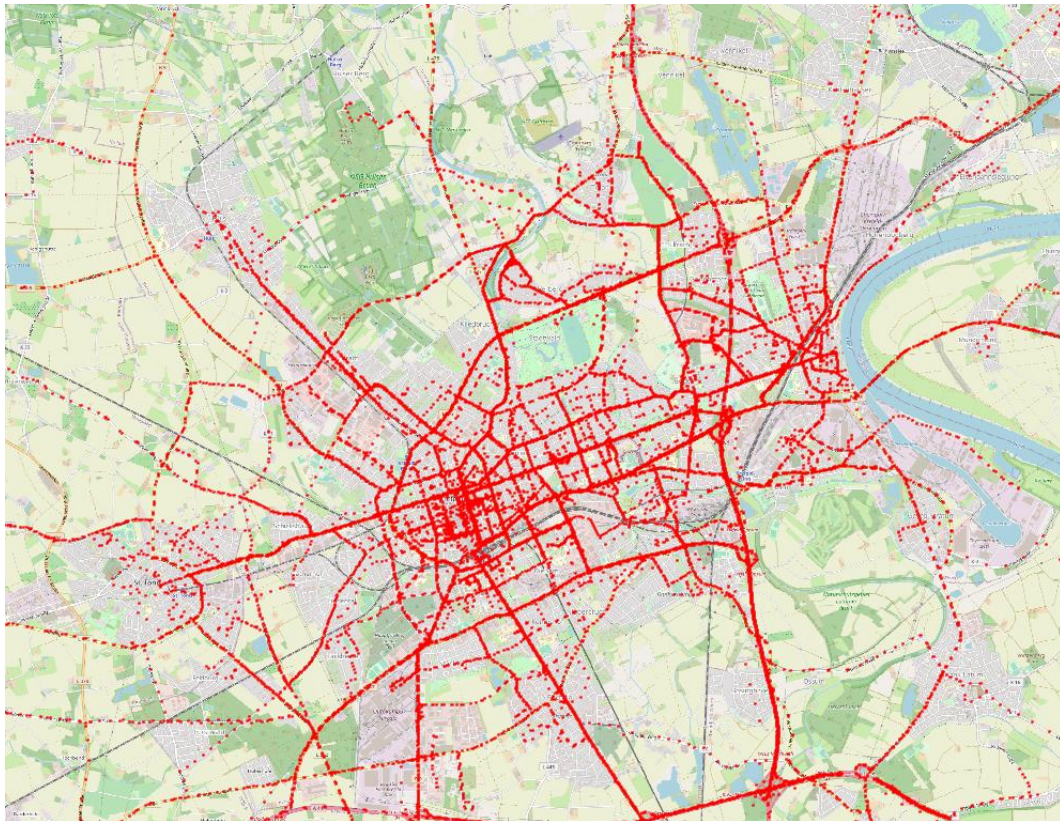


Abbildung 1: FCD-Punkte eines Taxies im Januar 2019 (Karte: OpenStreetMaps [5])

3 Datenaufbereitung

3.1 Formatierung

Die geografischen Koordinaten werden in das Universal Transverse Mercator-System überführt. Dieses kartesische Koordinatensystem mit Metern als Einheit vereinfacht die Berechnung von Distanzen ausreichend genau für stadtgroße Betrachtungsgebiete.

Die FCD-Punkte werden in Touren gruppiert. Eine Tour beginnt und endet mit dem Wechsel des Besetzungsgrades. Relevante Besetzungsgrade sind „besetzt“ und „beauftragt“. Datenpunkte mit irrelevanten Besetzungsgraden „frei“, „Pause“ und „Funk aus“ werden entfernt, da Fahrzeugbewegungen während dieser Zeiten kein normales Fahrverhalten widerspiegeln.

3.2 Sortierung

Dadurch, dass in dem vorliegenden Fall die Sekunden im Zeitstempel fehlen, kommt es vor, dass mehrere Datenpunkte den gleichen Zeitstempel haben. Um hier eine Sortierung innerhalb dieser Minute vornehmen zu können, werden die Distanzen von einem beliebigen Punkt des Zeitschritts $t-1$ zu allen Punkten des Zeitschritts t bestimmt. Es wird angenommen, dass die Touren keine Schleifen oder Umwege innerhalb einer Minute aufweisen, sodass die Reihenfolge der Punkte von Zeitschritt t der Reihenfolge der aufsteigenden Distanzen l zum Punkt des vorigen Zeitschritts entspricht. Diese Sortierung wird für alle Minuten in einem Track, die mehr als einen Datenpunkt beinhalten, durchgeführt. Die

jeweils ersten Zeitschritte einer Tour werden analog, aber mit den absteigenden Distanzen zu einem Punkt im Zeitschritt $t + 1$ sortiert.

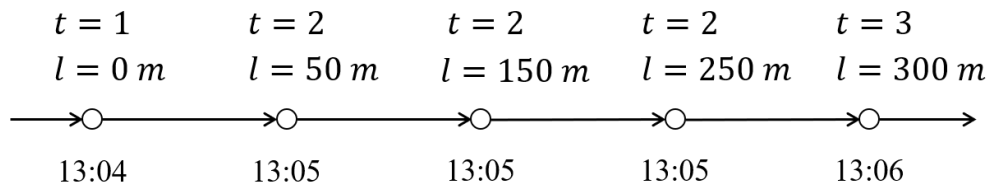


Abbildung 2: Sortierung der Datenpunkte anhand der Distanz l zum Datenpunkt des vorherigen Zeitschritts

3.3 Schätzen der Sekunden im Zeitstempel

Die fehlende Sekunden-Angabe im Zeitstempel wird geschätzt mit der Anzahl der Datenpunkte innerhalb des Zeitschritts. Wenn nur ein Datenpunkt in einem Zeitschritt vorhanden ist, wird dieser in die Mitte gesetzt. D.h. es werden zu der Zeit 30 Sekunden addiert. Wenn zwei Punkte einem Zeitschritt angehören, werden sie gleichmäßig verteilt, das heißt zu dem ersten Punkt werden 15 Sekunden und dem zweiten Punkt 45 Sekunden addiert.

Tabelle 1: Beispiel zur Schätzung der Sekunden im Zeitstempel

Daten ID	Gespeicherter Zeitstempel	Geschätzter Zeitstempel
1	13:04:00	13:04:30
2	13:05:00	13:05:15
3	13:05:00	13:05:45
4	13:06:00	13:05:10
5	13:06:00	13:06:30
6	13:06:00	13:06:50
7	13:07:00	13:07:30
8	13:08:00	13:08:30

Die zu addierenden Sekunden werden für jeden Datenpunkt i mit $i = 1, \dots, n$ eines jeden Zeitschritts und n als Anzahl der Datenpunkte im gleichen Zeitintervall, berechnet mit

$$\Delta t_i = \frac{60}{2n} + (i + 1) \frac{60}{n}. \quad (1)$$

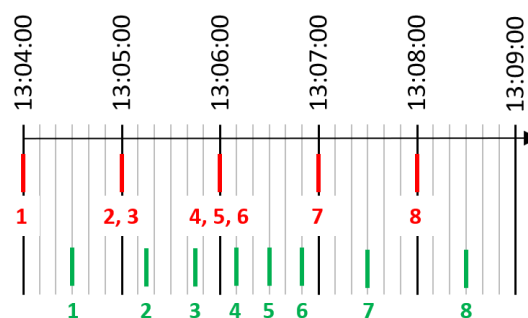


Abbildung 3: Schätzung der Sekunden durch gleichmäßige Verteilung der Punkte innerhalb einer Minute

3.4 Erstellen der Optimierungsvariablen

Wenn die Menge der Datenpunkte ausreichend ist, ist es möglich die Durchschnittsgeschwindigkeiten für einen Tagesgang zu berechnen. Welche Anzahl Daten erforderlich ist, wird in Kapitel 3.6 beschrieben. Die Parameterschätzung muss für jedes Zeitintervall einzeln durchgeführt werden. In diesem Fall werden stündliche Intervalle benutzt, wobei angenommen wird, dass sich die Durchschnittsgeschwindigkeit nachts zwischen 23 Uhr und 6 Uhr, aufgrund des geringen Verkehrsaufkommens, nicht wesentlich unterscheidet. Aus diesem Grund werden die Nachtstunden als ein Zeitintervall zusammengefasst. Die Optimierung wird somit für 18 verschiedene Zeitintervalle durchgeführt. Saisonale Unterschiede oder Urlaubszeiten können separat untersucht werden, was in diesem Beitrag nicht durchgeführt wurde.

Die Aufbereitung der Datensätze kann bis zu diesem Vorbereitungsschritt separat für jedes Fahrzeug auf parallelen Kernen berechnet werden, um die CPU-Auslastung zu maximieren. Nun werden die Punkte der Fahrzeuge für jedes Zeitintervall zusammengeführt. Eine Parallelisierung ist weiterhin möglich, wenn die unterschiedlichen Zeitintervalle parallelisiert berechnet werden.

Für jede Tour werden die FCD-Punkte dem OSM-Graphen zugeordnet, indem jeweils der Lotfußpunkt auf der nächstgelegenen Kante ermittelt wird. Diese Punkte werden mit Hilfe des Dijkstra-Algorithmus [2] verbunden, um eine kontinuierliche Tour zu erzeugen, die sich mit einer Auflistung der Identifikationsnummern der Kanten beschreiben lässt. Anschließend wird die Zuordnung geprüft, da durch die Ungenauigkeiten in den FCD-Koordinaten und den OSM-Daten eine falsche Kante zugeordnet werden kann.

In Abbildung 4 ist ein Beispiel dargestellt, indem der Punkt 2 fälschlicherweise der Abbiegespur zugeordnet wurde und somit ein 10,4 km langer Umweg, blau gestrichelt dargestellt, mit zwei Wendemanövern außerhalb des Darstellungsbereichs entsteht. Die reale Tour ist grün dargestellt.

Für die Überprüfung werden FCD-Punkte nacheinander aus dem Dijkstra-Algorithmus rausgenommen und verglichen, ob dadurch wesentlich kürzere oder längere Routen entstehen. Ist dies der Fall werden die jeweiligen Punkte aus der Tour entfernt. Müssen mehrere Punkte hintereinander entfernt werden, wird die Tour gekürzt oder ganz aus der Analyse entfernt.

Diese Informationen werden zu einer Matrix $A \in \mathbb{R}^{T \times E}$ zusammengefügt mit

$$A_{t,e} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,e} & \cdots & a_{1,E} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,e} & \cdots & a_{2,E} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{t,1} & a_{t,2} & \cdots & a_{t,e} & \cdots & a_{t,E} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{T,1} & a_{T,2} & \cdots & a_{T,e} & \cdots & a_{T,E} \end{pmatrix} \quad (2)$$

wobei der Wert $a_{t,e}$ die Strecke in Metern beschreibt, die während der Tour t auf der Kante e zurückgelegt wurde, mit $t = 1, \dots, T$ und $e = 1, \dots, E$. Eine Zeile repräsentiert somit eine Tour und eine Spalte eine Kante.

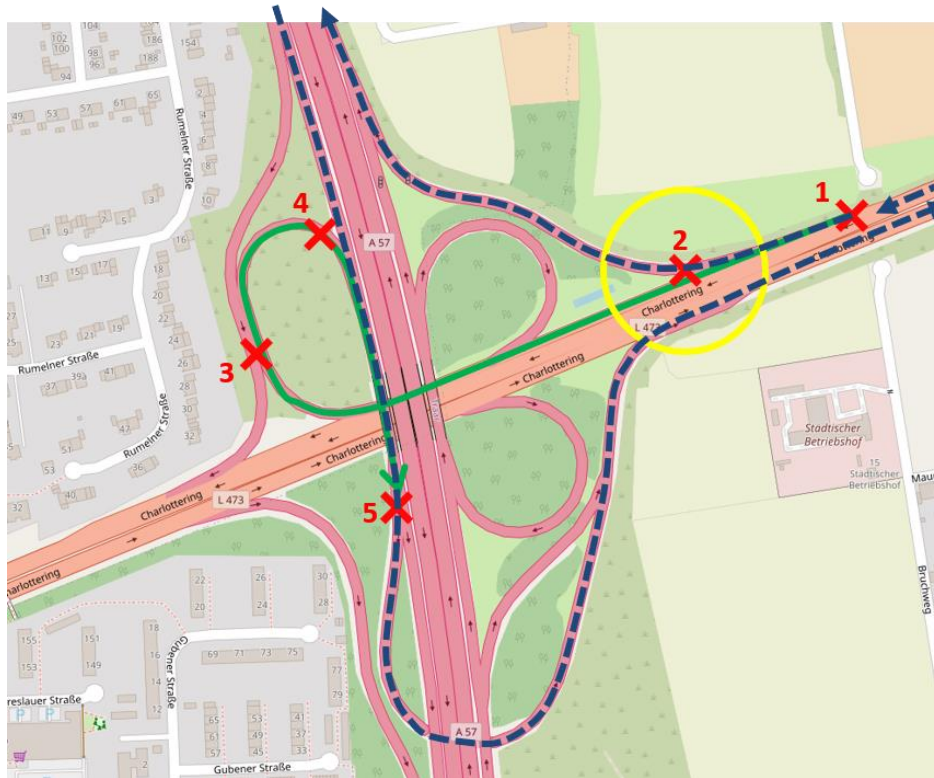


Abbildung 4: Abweichung (blau gestrichelt) zur realen Tour (grün) durch falsch zugeordnete Koordinate des zweiten FCD-Punktes

Für jede Tour t wird die Reisezeit $b \in \mathbb{R}^T$ berechnet, indem Ankunftszeit und Startzeit subtrahiert werden. Die Geschwindigkeit auf der Kante e wird definiert mit $s \in \mathbb{R}^E$. Das führt zu der Gleichung

$$A \cdot \tilde{s} = b \quad (3)$$

mit $\tilde{s}_e = \frac{1}{s_e}$, $\tilde{s} \in \mathbb{R}^E$, $e = 1, \dots, E$.

3.5 Vereinfachung des Gleichungssystems

Um die Anzahl der unbekannt Variablen und Gleichungen für die Optimierung zu minimieren, wird Matrix A vereinfacht. Wenn eine Zeile nur Nullen enthält, weist die jeweilige Tour keine Datenpunkte im Betrachtungsraum auf. Wenn eine Spalte nur Nullen enthält, wurde die entsprechende Kante nicht befahren. Diese Zeilen und Spalten werden aus dem Gleichungssystem entfernt.

Identische Spalten können zusammengeführt werden, indem sie komponentenweise zu einer Spalte summiert werden, denn alle Fahrzeuge sind entweder immer auf beiden Kanten gefahren oder auf keiner der beiden. Eine Unterscheidung in der Durchschnittsgeschwindigkeit ist daher nicht möglich, diese Kanten werden denselben Wert aufweisen.

Wenn die Optimierung zu wenig Gleichungen und zu viele Unbekannte enthält und daher das zugrundeliegende Gleichungssystem unterbestimmt ist, können zwei weitere Schritte angewandt werden, die allerdings den Detailgrad des Ergebnisses reduzieren.

3.5.1 Zusammenführen von unwichtigen Straßenabschnitten

In dem vorliegenden Fall entspricht die Anzahl der befahrenen Kanten in jedem Zeitintervall ungefähr der Anzahl der Fahrten. Daher soll die Analyse auf Hauptverkehrsstraßen konzentriert werden.

Die Statzung der Straßenreinigung [6] kann hierbei herangezogen werden, denn die Krefelder Straßen sind hierin mit Prioritäten versehen worden, womit auch die Hauptverkehrsstraßen festgelegt werden können. Eine andere Möglichkeit der Priorisierung ist die Nutzung von OSM-Tags [5]. Die wichtigsten Straßen sind Autobahnen und hier mit „motorway“ gekennzeichnet. Mit absteigender Relevanz gibt es noch die Tags „trunk“, „primary“, „secondary“, „tertiary“ bis hin zu „unclassified“ und „residential“. Während der erste Ansatz zur Priorisierung auf den Erfahrungswerten der Stadtwerke zur Verschmutzung beruht, basiert der zweite Ansatz auf dem Straßentyp und der Straßengröße.

Die Spalten in A , die die unwichtigen Straßenabschnitte repräsentieren, werden zusammengefügt, wobei die zulässigen Höchstgeschwindigkeiten weiterhin unterschieden werden. Es gibt also zu jeder Stufe der zulässigen Höchstgeschwindigkeit eine Gruppe von Spalten mit unwichtigen Straßen, deren Werte jeweils summiert in die Matrix A einfließen.

3.5.2 Zusammenführen von sehr kurzen Straßenabschnitten

Eine weitere Möglichkeit zur Reduzierung der Anzahl der Straßenabschnitte ist die Summierung der Spalten von Straßenabschnitt, die eine bestimmte Länge nicht überschreiten, mit ihren Nachbarn mit dem gleichen Straßennamen. In diesem Beispiel beträgt die Mindestlänge eines Straßenabschnitts 50 Meter.

Diese Vereinfachungen wurden durchgeführt, um ein Verhältnis von der Anzahl an Touren T zu der Anzahl an Kanten E (T/E -Verhältnis) von ca. 10 für jedes Zeitintervall zu erhalten. Dieser Wert wurde empirisch bestimmt und führt zu einer Konvergenz des Parameterschätzverfahrens mit den gegebenen Daten. Wie dieser Wert des minimalen T/E -Verhältnisses durch eine höhere Genauigkeit im Zeitstempel, eine höhere Anzahl an Daten oder Kanten beeinflusst werden kann, wird in diesem Beitrag nicht weiter untersucht.

4 Berechnung der Durchschnittsgeschwindigkeiten

Die Berechnung der Durchschnittsgeschwindigkeiten im beschriebenen Straßennetz wird mittels eines Parameterschätzverfahren umgesetzt, um das Minimum des Optimierungsproblems, beschrieben mit $\min_{\tilde{s}} f(\tilde{s})$, zu finden, wobei die Kostenfunktion $f(\tilde{s})$ definiert wird durch

$$f(\tilde{s}) = \sum_{t=1}^T (\mathbf{a}_t \cdot \tilde{\mathbf{s}} - b_t)^2 \quad (4)$$

mit $\mathbf{a}_t = a_{t,1}, \dots, a_{t,E}$.

Der Vektor $\tilde{\mathbf{s}}$ ist der Kehrwert des Geschwindigkeitsvektors \mathbf{s} und stellt die Optimierungsparameter dar. Als feste Schranken für Geschwindigkeitswerte gelten 0 bis $35 \frac{m}{s}$.

Damit wird für die Optimierung der Matlab-Constraint-Solver `fmincon` genutzt. Die beschriebenen Vereinfachungen führen ungefähr zu einer Anzahl Touren $T = 16.000$ und einer Anzahl relevanter Kanten $E = 1.500$ für jedes der 18 Zeitintervalle.

Mit dem Ergebnis aus der Optimierung werden schließlich auch Durchschnittsgeschwindigkeitswerte für die Kanten im Betrachtungsraum abgeschätzt, die nicht befahren wurden. Dafür wird angenommen, dass Straßen mit den gleichen Eigenschaften und OSM-Tags ähnliche Durchschnittsgeschwindigkeiten aufweisen. Hierfür werden die relevanten OSM-Keys „highway“, „maxspeed“, „maxspeed:type“, „oneway“ und „lanes“ herangezogen. Zusätzlich wird zwischen innerorts und außerorts unterschieden.

5 Ergebnisse

Der resultierende Vektor s repräsentiert die Durchschnittsgeschwindigkeit für jede Kante des Graphen. Teilt man die Kantenlängen durch die jeweiligen Werte aus dem Geschwindigkeitsvektor erhält man die durchschnittlichen Fahrzeiten für jede Kante. Diese Fahrzeiten können anstatt der Kantenlänge als Gewichtung des Graphen genutzt werden, um mit Hilfe von Kürzester-Pfad-Algorithmen die zu erwartende Reisezeit zwischen zwei Punkten in Krefeld zu bestimmen.

Die durchschnittlichen Geschwindigkeiten auf den Hauptverkehrsachsen in dem Zeitintervall von 8 Uhr bis 9 Uhr sind in Abbildung 5 dargestellt. Hier ist zu erkennen, dass der Verkehr auf manchen Straßen, wie z.B. „Frankenring“, „Kölner Straße“ oder „Gutenbergstraße“ geringe Geschwindigkeiten aufweist.

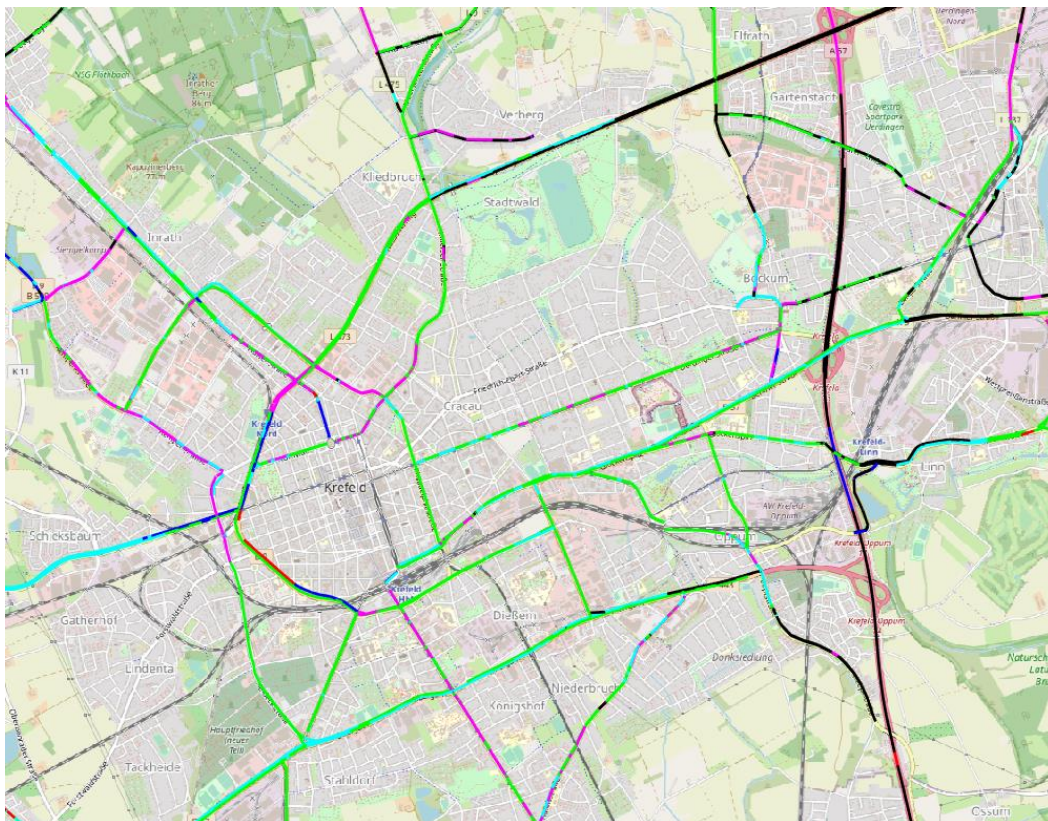


Abbildung 5: Auszug aus den Ergebnissen - Durchschnittsgeschwindigkeiten der Hauptverkehrsstraßen in Krefeld zwischen 8 und 9 Uhr farblich eingestuft

Tabelle 2: Farblegende zu Abbildung 5

Farbe	Geschwindigkeit [km/h]
Rot	≤ 10
Magenta	≤ 20
Grün	≤ 30
Cyan	≤ 50
Blau	≤ 70
Schwarz	> 70

6 Zusammenfassung und Ausblick

Die gezeigte Methode ermöglicht die Berechnung von Durchschnittsgeschwindigkeiten aus vielerorts verfügbaren Floating Car Data, sodass grundlegende Aussagen über ein städtisches oder ländliches Verkehrssystem getroffen werden können. Diese Daten können vielfältig genutzt werden, zum Beispiel um hohes Verkehrsaufkommen zu quantifizieren oder zu erwartende Reisezeiten zu bestimmen. In weiterführenden Analysen können diese Daten mit den Ergebnissen aus anderen Studien wie z.B. [8] verglichen werden.

In einem geplanten Forschungsprojekt sollen die Durchschnittsgeschwindigkeiten und resultierenden Fahrzeiten genutzt werden, um flächendeckende Reisewiderstände nach der Standardisierten Bewertung [3] mit dem Motorisierten Individualverkehr (MIV) und On Demand-Angeboten zu bestimmen, unabhängig von kommerziellen Datenanbietern wie zum Beispiel der Google Distance Matrix API, wie sie in [7] benutzt wurde.

Mit Hilfe dieser MIV-Reisewiderstände und ÖPNV-Reisewiderständen sowie einer multimodalen Quelle-Ziel-Zeit-Matrix soll ein makroskopisches Verkehrsmodell der Krefelder Mobilität erstellt werden, welches die Optimierung von Zukunftsszenarien einer energieeffizienten Verkehrswende ermöglicht.

Referenzen

- [1] Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C.: Introduction to algorithms. Cambridge, Massachusetts: The MIT Press, 2022. Fourth edition. 2022.
- [2] Dijkstra, E.W.: A note on two problems in connexion with graphs. Numerische Mathematik 1, 569-271. Springer 1959.
- [3] Intraplan Consult GmbH: Standardisierte Bewertung von Verkehrswegeinvestitionen im Schienengebundenen öffentlichen Personennahverkehr. München: ITP 2016.
- [4] Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameter. Journal of the society for Industrial and Applied Mathematics. 11(2). 431-441. 1963
- [5] OpenStreetMap contributors: Planet dump [data file from 09.04.2021]. <https://download.geofabrik.de/>

[6] Kommunalbetrieb Krefeld: Satzung über die Reinigung öffentlicher Straßen in der Stadt Krefeld. 7. Änderungssatzung. <https://www.krefeld.de/kbk/inhalt/strassenreinigung/#ym-downloads>

[7] Spengler, L., Gennat, M.: Fahrzeitermittlung im städtischen Raum mittels Google API. in: Proff, H. (eds) Making Connected Mobility Work. Springer Gabler, Wiesbaden 2021.

[8] Wagner, J.M.S., Eschbach, M., Vosseberg, K., Gennat, M.: Travel Time Estimation by means of Google API Data. IFAC-PapersOnLine. Volume 53. 2020.