Institute of Transport Economics
Norwegian Centre for Transport Research

Efficient Map-Matching for Urban Freight Applications

Øyvind Lothe Brunstad, Bo Dong

2097/2025

Title: Efficient Map-Matching for Urban Freight Applications
Tittel: Effektiv Map-Matching for anvendelsesområder i urban godstransport
Author: Øyvind Lothe Brunstad, Bo Dong
Date: 05.2025
TØI Report: 2097/2025
Pages: 34
ISSN Electronic: 2535-5104
ISBN Electronic: 978-82-480-1784-4
Project: 4310
Project Manager: Øyvind Lothe Brunstad
Quality Manager: Inger Beate Hovi
Commissioning: Trude Kvalsvik
Research Area: Industry and Freight
Keywords: Map-matching, distribution, Vehicle Kilometers, Cruising for Parking

Summary

We present a framework to simplify the process of offline map-matching, applied to real GPS data from vans in Oslo and Akershus in Norway. Filters are established to exclude GPS points associated with stops and minor movements. We also highlight challenges related to sparse and noisy GPS data and discuss advantages and disadvantages using different open-source map-matching tools. Map-matching resulted in a driving distance 11% longer, compared to straight-line GPS measurements. Furthermore, we review existing research on Vehicle Kilometres Travelled (VKT) and 'cruising for parking' for commercial vehicles and provide practical examples of how such analysis may be implemented using map-matching. Regarding 'cruising for parking' analyses for commercial vehicles, we find little or no research applying map-matching techniques. Therefore, we see a potential to supplement 'cruising for parking' research on commercial vehicles, by utilizing advanced map-matching techniques.

Kort sammendrag

Rapporten presenterer et rammeverk for offline map-matching, anvendt på GPS-data fra varebiler i Oslo og Akershus. Filtre brukes for å ekskludere GPS-punkter knyttet til stopp og små bevegelser. Vi framhever også utfordringer relatert til sparsomme og støyende data og diskuterer fordeler og ulemper ved ulike map-matching-verktøy med åpen kildekode.

Map-matching resulterte i gjennomsnitt med 11 % lengre kjøreavstand sammenliknet med rettlinjede GPS-målinger. Vi gjennomgår eksisterende forskning på «Vehicle Kilometres Traveled» (VKT) og «cruising for parking» for kommersielle kjøretøy og gir praktiske eksempler på hvordan slike analyser kan gjennomføres ved hjelp av map-matching. Når det gjelder analyser av cruising for parking for kommersielle kjøretøy, finner vi lite eller ingen forskning som bruker map-matching-teknikker. Vi ser derfor et potensial i å supplere cruising for parking-forskning for kommersielle kjøretøy med avanserte analyseteknikker.

Preface

This project was undertaken as an internal project at the Institute of Transport Economics (TØI) to develop a knowledge base on map-matching, as well as associated challenges and application areas. The main purpose of this project was to establish a framework to simplify and guide the process of map-matching, and to investigate how matched paths may form a basis for traffic work and 'cruising for parking' analyses in urban areas. GPS data collected in LIMCO (Hovi et al., 2021) formed the basis for the project, enabling us to perform and evaluate map-matching on real-world GPS data from vans in the Norwegian counties of Oslo and Akershus.

The project was carried out by Øyvind Lothe Brunstad and Bo Dong. Brunstad carried out map-matching analyses, evaluation of results, a literature review on 'cruising for parking' and most of the methodology part. He also contributed to the design of the map-matching framework and the introduction part. Dong carried out most of the introduction part, a literature review on VKT (Vehicle Kilometres Travelled), and contributed to the methodology part, the design of the map-matching framework, and evaluation of results, by providing extensive knowledge on the field and quality assurance. Inger Beate Hovi and Daniel Ruben Pinchasik contributed with proofreading and advice to ensure the relevance of this report.

Oslo, May 2025
Institute of Transport Economics

Bjørne Grimsrud	Kjell W. Johansen
Managing Director	Head of Research

Contents

Institute of Transport Economics
Norwegian Centre for Transport Research

ENGLISH **Summary**

Efficient Map-Matching for Urban Freight Applications

TØI Report 2097/2025 • Authors: Øyvind Lothe Brunstad, Bo Dong • Oslo 2025 • 34 pages

- Map-matching resulted in a driving distance, 11% higher compared to straight-line GPS measurements. However, both GPS quality and Hidden Markov Model (HMM)-assumptions greatly affect this measurement.
- Based on literature, we applied filters on GPS points associated with stops and minor movements at stop locations, and emphasized the importance of removing such points - as some of them may suffer from jiggle effect (when the vehicle is idling) or pickup/deliveries located at parking lots/infrastructure that may not be present as road segments in OpenStreetMap.
- PyTrack and similar Python-based tools offer easy solutions for map-matching but may not meet the speed requirements for large-scale applications. High-performance platforms like Open-Source Routing Machine (OSRM), Fast Map-matching (FMM) and Valhalla provide faster processing but require more complex setups.
- We revealed little or no existing research focusing on 'cruising for parking' analysis for commercial vehicles, using map-matching techniques. Therefore, we see a potential to supplement 'cruising for parking' research on commercial vehicles, by utilizing advanced map-matching techniques.

Introduction

Map-matching is a critical data preprocessing technique or cleaning method. It is the process of aligning GPS data points from a vehicle with the most likely road segments that could have produced those coordinates. This technique is crucial for several reasons, particularly in handling errors and improving the accuracy of GPS-based vehicle tracking. Map-matching is essential because GPS data is prone to two primary sources of error: Measurement errors, and sampling errors.

Map-matching algorithms can be broadly categorized into two types: online and offline map-matching. Online map-matching algorithms process GPS points in real-time, focusing on the vehicle's movement as it happens. Online map-matching is crucial for applications requiring immediate location updates and real-time decision-making, such as navigation systems and real-time traffic monitoring (Zhang et al., 2021). In contrast, offline map-matching algorithms analyse stored GPS data. This approach allows for more sophisticated processing techniques and can utilize the entire dataset to improve accuracy. Offline map-matching is ideal for applications that do not require real-time updates but need high accuracy, such as historical data analysis and transportation planning (Zhang et al., 2021).

In this report, we employed offline map-matching techniques, analysing GPS data from vans, considering freight transport in urban areas. We had access to GPS data from vans in the Norwegian counties of Oslo and Akershus, collected in LIMCO (Hovi et al.,2021). Based on this, we first implement and discuss a framework to simplify and guide the process of map-matching, referring to pre-processing steps, matching-techniques as well as examples of how matching-results can be used to model driving patterns. Secondly, we investigate how the resulting matched paths can be utilized in the application of traffic work estimations. On a general level, we also discuss the potential of map-matching in 'cruising for parking' analysis.

In a practical manner, we seek to answer the following research questions:

1) How well does the proposed map-matching framework perform on sparse and noisy GPS data retrieved from commercial vans in urban logistics?
2) What are the associated challenges in applying the map-matching framework to the given data?
3) How do matched paths form a basis for investigating traffic work and 'cruising for parking' analyses in urban areas?

Framework

We performed map-matching using PyTrack[1]. PyTrack and HMM-based map-matching formulated by Newson and Krumm (2009), formed the basis for our framework, with some modifications and additions, mainly concerning pre-preprocessing steps and the application of map-matching results. Our suggested framework starts by removing stop observations associated with pickup/ delivery activities. Next, for the algorithm to be able to search for candidates, we include a road network graph from OpenStreetMap. Next, we decide upon a given search radius for candidates. The identified candidates are then evaluated using the Viterbi Algorithm and emission and transition probabilities, where the goal is to find the most likely path, through the candidates at t times. This results in a matched path, that can serve various application areas.

Depending on the type of map-matching tool used, the process may differ a bit from our suggested framework. However, we believe this framework provide insight on the map-matching process on a general and practical level.

Additionally, it is important to note that Open-Source map-matching platforms like OSRM, FMM or Valhalla, offers faster matching performance compared to PyTrack. These platforms are optimized for speed, making them suitable for applications requiring quick processing times. On the other hand, the setup time and required skills are likely to increase. Also, many existing platforms are non-Python-based, required to run on private server and mainly designed for the industry (Tortora et al., 2022).

[1] https://github.com/cosbidev/PyTrack

Pre-processing steps

In general, our GPS trajectories were quite noisy and sparse. Based on literature, we applied filters on GPS points associated with stops and minor movements at stop locations, and emphasized the importance of removing such points - as some of them may suffer from jiggle effect (when the vehicle is idling) or pickup/deliveries located at parking lots/infrastructure that may not be present as road segments in OpenStreetMap. This pre-preprocessing step led to an exclusion of about 50% of all GPS points. Interestingly, the excluded GPS points only accounted for about 10% of the total Great Circle Distance and are likely even lower, due to jiggle effect.

Furthermore, removing GPS points resulting from pickup/delivery activities and stops over 60 minutes, resulted in even more sparsity in our data. This made it difficult to identify the vehicles trajectory, when we were manually interpreting the GPS points. In fact, based on the data, we do not know the actual routes.

Map-matching results

It is important to note that the HMM map-matching algorithm assumes that vehicles always take the shortest path between matched candidates (Newson & Krumm, 2009). This assumption can lead to inaccuracies, particularly when the distance between GPS points is substantial (i.e., the Great Circle Distance). In such cases, significant information about the true path is lost, and the algorithm's assumption of the shortest path may not reflect the actual route taken by the vehicle.

Following our presented framework, we calculated an average map-matching distance, about 11% higher compared to the distance between the consecutive GPS points. Exploring each individual vehicle route, this difference ranged from a minimum of 2.5 percent to a maximum of 30.6 percent. However, both GPS quality and HMM-assumptions greatly affect these measurements.

By modelling a view of the routes, we also illustrated that main roads, such as highways, are used more frequently compared to smaller inner city roads. This is because main roads serve as transit routes between delivery areas and terminal locations, while smaller roads are used more randomly for specific delivery missions.

Vehicle kilometres travelled (VKT)

The Vehicle Kilometres Travelled or VKT[2] is a crucial metric for quantifying road infrastructure usage and is considered one of the most important statistical measures in traffic and transport analysis (Chen et al., 2017). Traditionally, a wide range of methods have been applied with the focus of measuring VKT. This includes Traffic Volume Counts, using fuel consumption to estimate VKT, Odometer Readings or Household and Workplace Travel Surveys. However, recent advancements in positioning technology and data processing have significantly improved the estimation of VKT. One of the most promising innovations in this field is the integration of GPS data with digital road networks to enhance accuracy and reliability in VKT measurement (Fan et al., 2019). Unlike traditional survey-based methods, GPS data collection

[2] Vehicle kilometers travelled (VKT) is also referred to as traffic work.

eliminates the need for periodic manual reporting, allowing for more dynamic and scalable monitoring of road usage patterns.

We attempted to illustrate how our matched paths from commercial vans in Oslo and Akershus form a basis for estimating VKT. Generally, we consider VKT to be one of several application areas of map-matching. Compared with publicly available statistics collected using traditional methods, our approach highly deviates in terms of VKT. Our VKT estimation is therefore best suited as an example. High deviations is likely due to uncertain assumptions we make, such as the number of departures/arrivals at terminal locations daily, comparable industry segments, the number of days in a year that a particular van is used or the fact that our data only covers one particular day. These uncertainties are described in greater detail in chapter 5.1. However, by using larger samples of GPS data also covering seasonal variations, we believe these uncertainties can be reduced and potentially reflect more realistic aggregated VKT measures compared to publicly available statistics collected using traditional methods. This in turn would require faster matching-performance, where Open-Source map-matching platforms like OSRM, Valhalla or FMM, could be possible candidates.

Cruising for parking

Another application area of map-matching might be to detect the well-known ‘cruising for parking’, which in urban areas, can cause significant congestions and pollutions (Dalla Chiara and Goodchild, 2020). We reviewed existing literature focusing on ‘cruising for parking’. Our aim was to identify previous research primarily focusing on commercial vehicles that applied GPS data and preferably map-matching techniques as a basis to perform such analysis. However, most literature have been focused on passenger vehicles, while for commercial vehicles, the literature is scarce (Dalla Chiara and Goodchild, 2020). Many have assumed that commercial vehicles park in travel lanes or loading zones close to the destination, and therefore do not ‘cruise for parking’ (Amer and Chow, 2017; Iwan et al., 2018; as cited in Dalla Chiara and Goodchild, 2020). For this reason, we also consider other research focusing on non-commercial vehicles as relevant.

Considering non-commercial vehicles, we identified seven articles focusing on ‘cruising for parking’, using map-matching techniques. In contrast, we identified little or no similar research applying the same techniques when focusing on commercial vehicles. However, several researchers have addressed commercial vehicle’s parking behaviour, ‘cruising for parking’ or closely related topics, by utilizing raw GPS data or digital tachograph data. Out of these, we identified 4 articles somehow focusing on ‘cruising for parking’ analysis, utilizing raw GPS data.

We see a potential to supplement ‘cruising for parking’ research on commercial vehicles, by utilizing advanced map-matching techniques. Future research may also consider map-matching techniques applied in previous research focusing on non-commercial vehicles. For instance, Weinberger et al. (2020) focused on ‘cruising for parking’ for non-commercial vehicles and noticed that current map-matching software sometimes results in errors, when applied on low-quality or circling GPS data. For this reason, they developed a new map-matching algorithm named ‘pgMapMatch’, to better cope with U-turns and circling, by also allowing for analysis on low-resolution GPS data.

Transportøkonomisk institutt
Stiftelsen Norsk senter for samferdselsforskning

NORSK Sammendrag

Effektiv Map-Matching for anvendelsesområder i urban godstransport

TØI rapport 2097/2025 • Forfattere: Øyvind Lothe Brunstad, Bo Dong • Oslo, 2025 • 34 sider

- Anvendelse av Map-matching på GPS-data fra varebiler i bydistribusjon resulterte i en estimert kjørelengde som er 11 % høyere sammenlignet med rettlinjede GPS målinger. Resultatene påvirkes imidlertid i stor grad av både GPS-kvalitet og forutsetninger brukt i map-matching med en Hidden Markov Model (HMM).
- Vi understreker viktigheten av å filtrere GPS-observasjoner knyttet til stopp og mindre bevegelser på laste/losseplasser, da disse kan være påvirket av «jiggle»-effekten (når kjøretøyet står på tomgang), eller laste-/losse-operasjoner på parkeringsplasser som kanskje ikke er representert som vegsegmenter i OpenStreetMap.
- PyTrack og lignende Python-baserte verktøy tilbyr enkle løsninger for map-matching, men oppfyller sannsynligvis ikke ytelseskravene for store applikasjoner. Høyytelsesplattformer som Open-Source Routing Machine (OSRM), Fast Map-matching (FMM) og Valhalla gir raskere beregning, men krever mer komplekse oppsett.
- Vi fant lite eller ingen eksisterende forskning med fokus på analyser av 'cruising for parking' for kommersielle kjøretøy ved bruk av map-matching-teknikker. Vi ser derfor et fremtidig forskningspotensial ved å supplere 'cruising for parking'-forskning for kommersielle kjøretøy med avanserte map-matching-teknikker.

Innledning

Map-matching er et viktig prosesseringsverktøy eller rensemetode for GPS data. Gitt et sett av GPS koordinater fra et kjøretøy, beregner metoden veisegmenter med størst sannsynligheten for å ha produsert GPS koordinatene. Dette er spesielt viktig for å håndtere feilaktige GPS data. GPS data er i hovedsak utsatt for to typer støy/feil: målefeil og utvalgsfeil.

Det skilles i hovedsak mellom to typer map-matching: online og offline map-matching. Online map-matching prosesserer GPS data i sanntid, og fokuserer dermed på kjøretøyets bevegelse mens det skjer. Online map-matching er viktig for applikasjoner med krav til umiddelbar oppdatering av posisjon og beslutningstaking i sanntid, slik som navigasjonssystemer og trafikkovervåkning i sanntid (Zhang et al., 2021). I motsetning, analyserer offline map-matching lagrede GPS data. Dette muliggjør mer avanserte prosesseringsteknikker og hele GPS datasettet kan utnyttes til forbedring av punktlighet. Offline map-matching er nyttig for applikasjoner uten krav til GPS i sanntid, men som krever høy punktlighet, f.eks. historiske analyser og transportplanlegging (Zhang et al., 2021).

Vi har benyttet offline map-matching-teknikker til å analysere GPS data fra varebiler, med fokus på godstransport i urbane områder. GPS-dataene stammet fra varebiler i Oslo og Akershus, innsamlet i forbindelse med LIMCO-prosjektet (Hovi et al.,2021). Basert på dette har vi først implementert og diskutert et rammeverk som forenkler og veileder map-matchings-prosessen, med fokus på forbehandling av data, matching-teknikker, samt eksempler på hvordan matchingsresultater kan brukes til å modellere kjøremønstre i kart.

I neste steg undersøkte vi hvordan de matchede kjøremønstrene kan anvendes i analyser av trafikkarbeid. På et overordnet nivå vurderer vi også map-matchingspotensialet i analyser av 'cruising for parking'.

På en praktisk måte søker vi å svare på følgende forskningsspørsmål:

1) Hvor godt presterer det foreslåtte rammeverket for map-matching på sparsomme og støyende GPS-data hentet fra kommersielle varebiler i urban logistikk?
2) Hvilke utfordringer er forbundet med å benytte dette rammeverket på de gitte dataene?
3) Hvordan kan matchede kjøremønstre danne grunnlag for å undersøke trafikkarbeid og «cruising for parking» i byområder?

Rammeverk

Map-matching er utført ved hjelp av programvaren PyTrack[3]. PyTrack, og HMM-basert map-matching formulert av Newson og Krumm (2009), dannet grunnlaget for rammeverket vårt, med enkelte modifikasjoner og tillegg, hovedsakelig knyttet til forbehandling av data og anvendelsesområder for map-matching. Rammeverket begynner med å fjerne stoppobservasjoner knyttet til laste- og losseaktiviteter. Deretter, for å muliggjøre søk etter kandidater, inkluderer vi en veinettgraf fra OpenStreetMap. En søkeradius defineres, og innenfor denne identifiseres aktuelle kandidater som blir evaluert ved hjelp av Viterbi-algoritmen, basert på emisjons- og overgangssannsynligheter, med mål om å finne den mest sannsynlige ruten gjennom kandidatene over tid. Dette resulterer i en matchet rute som anvendes i ulike analyser og bruksområder.

Avhengig av hvilken programvare som benyttes til map-matching, kan prosessen avvike noe fra vårt foreslåtte rammeverk. Vi mener imidlertid at dette rammeverket gir innsikt i map-matchingsprosessen på et generelt og praktisk nivå.

Det er også viktig å merke seg at Open-Source map-matchingsplattformer som OSRM, FMM eller Valhalla, tilbyr raskere ytelser sammenlignet med PyTrack. Disse plattformene er optimalisert for hastighet, noe som gjør dem egnet for applikasjoner som krever raske behandlingstider. På den annen side vil tiden det tar å sette disse opp og nødvendige ferdigheter sannsynligvis øke. Dessuten er mange eksisterende plattformer ikke-Python-baserte, krever å kjøre på privat server og hovedsakelig designet for industrien (Tortora et al., 2022).

[3] https://github.com/cosbidev/PyTrack

Forbehandlingstrinn

GPS dataene vi hadde tilgang til, var generelt sparsomme og preget av betydelig støy. Samtidig understreker vi viktigheten av å filtrere GPS observasjoner knyttet til stopp og mindre bevegelser på laste/losseplasser, da disse kan være påvirket av «jiggle»-effekten (når kjøretøyet står på tomgang), eller av laste-/losseoperasjoner på parkeringsplasser som kanskje ikke er representert som vegsegmenter i OpenStreetMap. Dette forbehandlingstrinnet førte til at rundt 50 % av alle GPS-punkter ble ekskludert. Interessant nok utgjorde de ekskluderte GPS-punktene kun rundt 10 % av den totale 'Great Circle Distance'-avstanden og det er sannsynlig at denne er enda lavere, nettopp på grunn av «jiggle»-effekten.

Det er også verdt å merke seg at ekskluderingen av GPS-punkter knyttet til laste- og losse-aktiviteter, samt stopp over 60 minutter, førte til ytterligere sparsomme data. Dette gjorde det utfordrende å identifisere kjøretøyets rute, spesielt ved manuell tolkning av GPS-punktene. Basert på det tilgjengelige datagrunnlaget kjenner vi faktisk ikke de eksakte rutene kjøretøyene har fulgt.

Map-matching resultater

Det er viktig å merke seg at HMM map-matching algoritmen antar at kjøretøy alltid tar den korteste ruten mellom matchede kandidater (Newson & Krumm, 2009). Denne antagelsen kan føre til unøyaktigheter, spesielt når avstanden mellom GPS-punktene er betydelig (dvs. 'Great Circle Distance'). I slike tilfeller går betydelig informasjon om det sanne rutevalget tapt, og algoritmens antakelse om den korteste veien gjenspeiler kanskje ikke den faktiske ruten kjøretøyet tok.

Basert på det presenterte rammeverket, beregnet vi en gjennomsnittlig map-matchings-distanse, omtrent 11 % høyere enn den rettlinjede distansen mellom de påfølgende GPS-punktene. Ved utforsking av hver enkelt kjøretøyrute, varierte denne forskjellen fra minimum 2,5 prosent til maksimalt 30 prosent. Målingene påvirkes imidlertid sterkt av både GPS-kvalitet og HMM-map-matching antakelser.

Ved å visualisere de matchede rutene i et kart, illustrerte vi også at hovedveier, som f.eks. motorveier, brukes hyppigere sammenlignet med mindre byveier. Dette skyldes at hovedveier fungerer som transittruter mellom leveringsområder og terminallokasjoner, mens mindre veier brukes mer tilfeldig for spesifikke leveringer.

Vehicle Kilometres Travelled (VKT)

"Vehicle Kilometres Travelled" (VKT), også kalt trafikkarbeid, er avgjørende for å kvantifisere bruken av veiinfrastruktur og regnes som et av de viktigste statistiske målene i trafikk- og transportanalyse (Chen et al., 2017). Tradisjonelle metoder med søkelys på å kvantifisere VKT, inkluderer gjerne vegtrafikktellinger, bruk av drivstofforbruk for å estimere VKT, avlesning av kilometertellere, eller reiseundersøkelser blant husholdninger og arbeidsplasser. Nylige fremskritt innen posisjoneringsteknologi og databehandling har imidlertid forbedret estimeringen av VKT betydelig. En av de mest lovende innovasjonene på dette feltet er integreringen av GPS-data med digitale veinett for å øke nøyaktigheten og påliteligheten i VKT-målinger (Fan et al., 2019). I motsetning til tradisjonelle undersøkelsesbaserte metoder, eliminerer GPS-datainnsamling behovet for periodisk manuell rapportering, noe som muliggjør mer dynamisk og skalerbar overvåking av veibruksmønstre.

Vi har forsøkt å illustrere hvordan våre matchede ruter fra varebiler i Oslo og Akershus kan danne grunnlag for å estimere VKT. Generelt anser vi VKT for å være en av flere mulige bruksområder for map-matching. Sammenlignet med offentlig tilgjengelig statistikk, innsamlet med tradisjonelle metoder, finner vi store avvik med hensyn på VKT. Vår VKT-estimering er derfor best egnet som eksempel. Høye avvik skyldes sannsynligvis usikre forutsetninger vi la til grunn for beregningene, som for eksempel antall avganger/ankomster på terminallokasjoner daglig, sammenlignbare næringstyper bedrifter tilhører, antall dager i et år som en bestemt varebil brukes eller det faktum at våre benyttede GPS data kun inkluderer en dag. Disse usikkerhetene er nærmere beskrevet i kapittel 5.1. Vi tror imidlertid at større utvalg av GPS data som også dekker sesongvariasjoner, kan redusere disse utfordringene og potensielt reflektere mer realistiske aggregerte tall på VKT, sammenlignet med offentlig tilgjengelig statistikk innsamlet med tradisjonelle metoder. Dette vil igjen kreve raskere matching-ytelse, der Open-Source kart-matching-plattformer som OSRM, Valhalla eller FMM kan være mulige verktøy.

Cruising for parking

Et annet bruksområde for map-matching kan være identifisering av det velkjente begrepet 'cruising for parkering', som i urbane områder kan forårsake betydelige køer og forurensninger (Dalla Chiara og Goodchild, 2020). Vi har gjennomgått eksisterende litteratur med fokus på 'cruising for parkering'. Målet var å identifisere tidligere forskning, primært med fokus på kommersielle kjøretøy som benyttet GPS data og fortrinnsvis map-matching-teknikker som grunnlag for å utføre slike analyser. Det meste av litteraturen har imidlertid vært fokusert på ikke-kommersielle kjøretøy, mens for kommersielle kjøretøy er litteraturen knapp (Dalla Chiara og Goodchild, 2020). Mange har antatt at kommersielle kjøretøy parkerer i kjørefelt eller lastesoner nær destinasjonen, og at de derfor ikke foretar 'cruise for parking' (Amer og Chow, 2017; Iwan et al., 2018; som sitert i Dalla Chiara og Goodchild, 2020). I lys av dette, vurderer vi også annen forskning med fokus på ikke-kommersielle kjøretøy som relevant.

For ikke-kommersielle kjøretøy, identifiserte vi syv artikler med fokus på 'cruising for parkering', og som benyttet map-matching-teknikker. Derimot identifiserte vi lite eller ingen lignende forskning som implementerte de samme teknikkene for kommersielle kjøretøy. Flere forskere har imidlertid analysert kommersielle kjøretøys parkeringsatferd, 'cruising for parking' eller nært beslektede emner, ved å bruke rå GPS data eller digitale fartsskriverdata. Av disse identifiserte vi fire artikler som på en eller annen måte fokuserer på 'cruising for parkering'-analyser, ved å benytte rå GPS data.

Vi ser et fremtidig forskningspotensial ved å supplere 'cruising for parking'-forskning for kommersielle kjøretøy, med avanserte map-matchingsteknikker. Fremtidig forskning kan også vurdere map-matching-teknikker benyttet i tidligere forskning med fokus på ikke-kommersielle kjøretøy. For eksempel, Weinberger et al. (2020) fokuserte på 'cruising for parkering' for ikke-kommersielle kjøretøy og la merke til at daværende map-matching-programvarer noen ganger resulterer i feil når de anvendes på GPS data av lav kvalitet eller GPS sirkler. Derfor utviklet de en ny map-matching algoritme kalt 'pgMapMatch', for bedre å takle U-svinger og GPS sirkler, som også muliggjør analyser av lavoppløselige GPS data.

1 Introduction

The modernization of the transport sector offers potential for efficiency and innovation, driven by the implementation of state-of-the-art Artificial Intelligence (AI) and Machine Learning (ML) algorithms. These technologies have the capability to revolutionize transportation systems, enhance decision-making processes, and improve operational efficiency. However, the quality and effectiveness of AI-augmented transport models are heavily dependent on the quality of the data they are fed with. Therefore, acquiring high-quality data is a critical first step for any intelligent transport application. In the era of big data, travel data from traditional transport surveys remain essential. Among the diverse types of big data, GPS data have emerged as one of the most significant and widely used data sources. GPS data provide critical insights into vehicle locations, travel patterns, and distribution routes.

While big data offers tremendous potential, it also presents significant challenges in understanding how to effectively utilize and interpret these vast datasets (Bonnel and Munizaga, 2018). First and foremost, the quality of GPS data can vary significantly. It is often noisy, incomplete, or inaccurate, which poses challenges for directly utilizing it in Data-Driven Transport Systems (DDTS). Low-quality GPS data can introduce significant errors into DDTS. Uncleaned GPS data may generate incorrect conclusions, such as inefficient vehicle routing or poor resource allocation. In worst-case scenarios, the reliance on erroneous GPS data can lead to real-world consequences, such as accidents or system failures, particularly in critical applications like autonomous driving or emergency response systems. Addressing the challenges of noisy or low-quality data is essential to unlock the full potential of future technologies. By prioritizing data cleaning and preprocessing, the transport sector can confidently adopt AI and ML solutions to enhance innovation, efficiency, and safety.

A workflow diagram (Figure 1.) illustrates how GPS data may form the basis for different application areas. The process starts by collecting GPS data for a specific vehicle type. Next, GPS traces are matched to a road network using map-matching techniques. This results in matched paths, which may serve as basis for various analysis, such as calibration of Vehicle Kilometres Travelled (VKT) or analysis of parking behaviour.

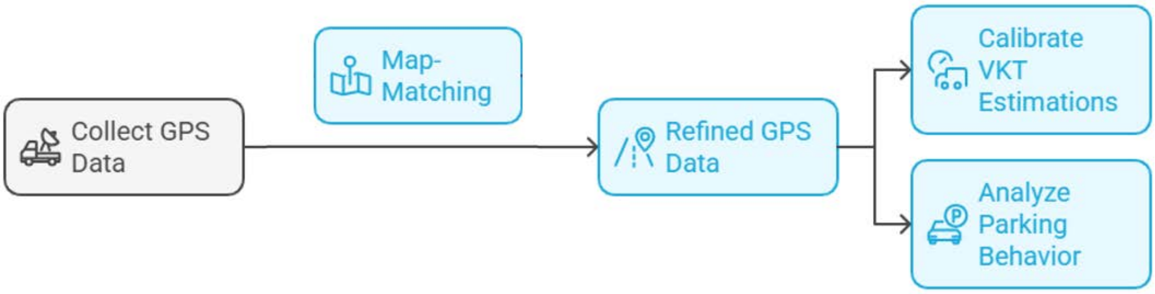

Figure 1.1: GPS data processing workflow including Map-Matching. The refined GPS data can be applied to calibrate VKT estimations and/or analyse parking behaviour among other things.

Map-matching is a critical data preprocessing technique or cleaning method. It is the process of aligning GPS data points from a vehicle with the most likely road segments that could have produced those coordinates. This technique is crucial for several reasons, particularly in handling errors and improving the accuracy of GPS-based vehicle tracking. Map-matching is essential because GPS data is prone to two primary sources of error: Measurement errors, and sampling errors.

Measurement errors occur when recorded GPS points do not accurately reflect the vehicle's actual location. Factors contributing to measurement errors include satellite signal multipath effects, atmospheric conditions, and receiver quality.

Sampling errors arise because GPS points are recorded at discrete intervals, typically every few seconds. This interval-based recording results in the loss of information about the vehicle's trajectory between recorded points (Plaudis et al., 2021; Saki and Hagen, 2022).

As an example, one can consider a scenario where a vehicle turns or deviates from its path between two recorded points. Since these manoeuvres might not be captured, the data can be incomplete or potentially misleading. By employing map-matching techniques, we can correct measurement and sampling errors, thereby improving the accuracy and reliability of GPS data. This can be of particular value in areas such as urban transport, where planners can use map-matched data to analyse transportation networks, identify bottlenecks, and design more efficient road systems.

Map-matching algorithms can be broadly categorized into two types: online and offline map-matching. Online map-matching algorithms process GPS points in real-time, focusing on the vehicle's movement as it happens. Online map-matching is crucial for applications requiring immediate location updates and real-time decision-making, such as navigation systems and real-time traffic monitoring (Zhang et al., 2021). In contrast, offline map-matching algorithms analyse stored GPS data. This approach allows for more sophisticated processing techniques and can utilize the entire dataset to improve accuracy. Offline map-matching is ideal for applications that do not require real-time updates but need high accuracy, such as historical data analysis and transportation planning (Zhang et al., 2021).

In our research, we employ offline map-matching techniques. The choice of offline map-matching is driven by the nature of our report, which focuses on analysing stored GPS data from vans, in the context of freight transport in urban areas, i.e. data stemming from the LIMCO project (Hovi et al., 2021). Working with offline map-matching techniques allows us to leverage the full LIMCO-dataset, applying complex algorithms to correct for measurement and sampling errors, and thereby ensuring the highest possible accuracy.

Map-matching processes necessitate the use of a digital map, which provides the framework for aligning GPS data points with road segments. There are several sources for digital maps, each offering different levels of detail and coverage. OpenStreetMap (OSM) is a widely used source, and structures road segments into nodes at intersections and links between them. OSM is an open-source platform, and its data are freely available, making it a popular choice for many map-matching applications. For a more detailed representation of roads in Norway, the National Road Data Bank offers comprehensive road data. These data can be downloaded from the Norwegian Public Roads Administration[4]. While OSM is commonly used, the National Road Data Bank provides more specific and detailed information, which can be advantageous for certain applications. In our research, we utilize OSM for the map-matching process due to its accessibility, and because of its extensive support for various map-matching applications that build further upon OSM-data.

Table 1.1 provides a comparative overview of selected studies on map-matching applications. It highlights various application areas such as taxi services, logistics companies, and private vehicle fleets, with GPS frequency ranging from 1 second to 2 minutes. The test regions for these studies span multiple locations, including Nagoya (Japan); Central Stockholm (Sweden); San Francisco (USA); Wuhan (China), and Nottingham and Birmingham (UK). The table also includes a comparison of our contribution in the context of logistics companies in Oslo, Norway, with a GPS frequency of 2

[4] https://vegkart.atlas.vegvesen.no/

minutes. This comprehensive comparison underscores the diversity in GPS sampling intervals and geographic regions covered by map-matching research.

Table 1.1: Selected studies on map-matching applications identified in scientific literature.

Article	Application area	GPS frequency	Test Region
Miwa et al. (2012)	Taxi	from 5 s to 90 s	Nagoya (Japan)
Rahmani and Koutsopoulos (2013)	Taxi	up to 2 minutes	Central Stockholm (Sweden)
Hunter et al. (2014)	Taxi	between 1 s and 2 min	San Francisco (USA)
Chen et al. (2014)	A logistics company	from 10 s to 120 s	Wuhan (China)
Quddus and Washington (2015)	A fleet of private cars, buses or light duty vehicles or smartphones	1 s, 5 s, 30 s, 60 s sampling intervals	Urban, suburban and rural areas in Nottingham and Birmingham (UK)
Bierlaire et al. (2013)	Smartphone by the same person	10 s	Not specified
Our contribution	**Logistics companies**	**2 minutes**	**Oslo (Norway)**

In this report, we first implement and discuss a framework to simplify and guide the process of map-matching applied on a real case. Here, we refer to pre-processing steps and matching-techniques, as well as examples of how matching-results can be used to model a view of driving patterns. Secondly, we investigate how the resulting matched paths can be utilized in the application of traffic work estimations (described in chapter 2.1). On a general level, we also discuss the potential of map-matching in ‘cruising for parking’ analysis (described in chapter 2.2). In a practical manner, we seek to answer the following research questions:

1) How well does the proposed map-matching framework perform on sparse and noisy GPS data retrieved from commercial vans in urban logistics?
2) What are the associated challenges in applying the map-matching framework to the given data?
3) How do matched paths form a basis for investigating traffic work and ‘cruising for parking’ analyses in urban areas?

The remainder of this report is organized as follows. Chapter 2 provides a literature review on the existing literature regarding VKT and ‘cruising for parking’. Chapter 3 describes the map-matching algorithm based on the Hidden Markov Model (HMM), data preprocessing steps and proposes our framework. Chapter 4 reports results of our computational experiments. Chapter 5 discusses the application of traffic work estimations and ‘cruising for parking’ analyses, based on map-matching. Lastly, the conclusion is given in Chapter 6.

List of abbreviations

Abbreviations	
AI:	Artificial Intelligence
DDTS:	Data-Driven Transport Systems
FCD:	Floating Car Data
FMS:	Fleet Management System
GPS:	Global Positioning System
HGV:	Heavy goods vehicle
HMM:	Hidden Markov Model
ML:	Machine Learning
OSM:	OpenStreetMap
SSB:	Statistics Norway
VKT:	Vehicle Kilometers Travelled, also referred to as traffic work.

2 Literature review

2.1 Vehicle kilometres travelled

The Vehicle Kilometres Travelled or VKT[5] is a crucial metric for quantifying road infrastructure usage and is considered as one of the most important statistical measures in traffic and transport analysis (Chen et al., 2017). This is because VKT measures the total distance travelled by vehicles within a specific geographic area over a given period and directly reflects the extent of spatial interaction within society and the economy, serving as a key indicator of how road networks are utilized (Bäumer et al., 2018). Accurate measurement of VKT is essential for effective transportation planning, policymaking, and infrastructure development. It provides insights into travel demand, helps in assessing the efficiency of transportation systems, and aids in the allocation of resources for maintenance and expansion projects (Mun and Jung, 2024). Moreover, VKT data is instrumental in evaluating environmental impacts, such as vehicle emissions, and in formulating strategies to mitigate congestion and improve air quality (Zhang et al., 2025).

2.1.1 Traditional Methods of Estimating VKT

Traditional methods of collecting VKT data have relied on a variety of techniques, each offering different levels of accuracy, feasibility, and cost-effectiveness. Figure 2.1 illustrates the four primary methods commonly used for VKT estimation: (1) Traffic Volume Counts; (2) Fuel Consumption Method; (3) Odometer Readings; and (4) Travel Surveys.

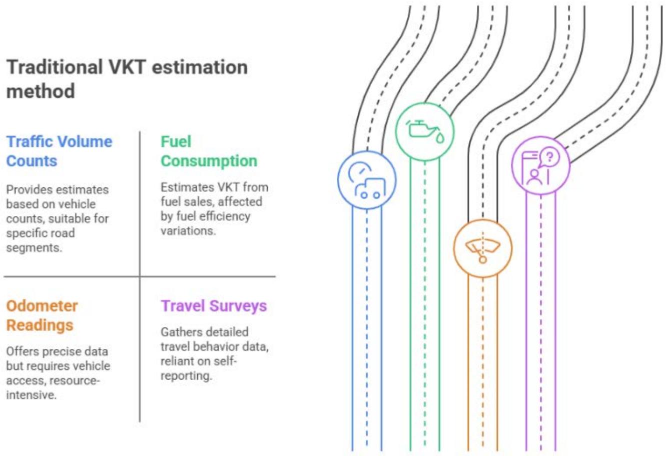

Figure 2.1: Traditional Methods for Collecting Vehicle Kilometres Travelled Data.

1. **Traffic Volume Counts** have been widely used to estimate VKT by counting the number of vehicles passing a specific point on a roadway over a set period (Fu et al., 2017). This data is then multiplied by the length of the road segment to obtain an estimate of total VKT. These counts can be conducted manually by human observers or automatically using technologies such as pneumatic tubes, loop detectors, and infrared sensors. The Norwegian Public Roads Administration (NPRA) maintains a large number of traffic counting stations along the main

[5] Vehicle kilometers travelled (VKT) is also referred to as traffic work.

road network, and the data from these stations are made available through the National Road Data Bank[6] (Nasjonal vegdatabank, NVDB). Studies have demonstrated the effectiveness of spatial interpolation and remote sensing in refining traffic volume estimates for VKT calculation (see Jung et al., 2017; Fu et al., 2017). For example, Fu et al., 2017 utilizes traffic volume counts to estimate street level annual average daily traffic values (AADTs) and convert AADTs to VKMs for all vehicle types for each road segment by multiplying the AADTs for vehicle types by the length of that road segment.

2. **Odometer Readings** Collecting odometer readings from vehicles during periodic inspections or surveys provides a direct measurement of VKT (Alberini et al., 2021). This method is used by Statistics Norway (SSB) for producing their road traffic volume statistics[7] (kjørelengde-statistikk). This method, while highly accurate, requires large-scale data collection efforts and regulatory frameworks to ensure comprehensive coverage. Studies analyzing odometer data from the National Household Travel Survey highlight its role in estimating vehicle miles traveled and overall travel demand (Alberini et al., 2021). A practical application related to VKT estimation is found in Fleet Management Systems, which typically use odometer readings to report the distances traveled by vehicles.
3. **Fuel Consumption Method** Another traditional approach estimates VKT based on fuel consumption statistics and average vehicle fuel economy (Hakimelahi et al., 2020). By dividing the total fuel consumed by the estimated average fuel efficiency, VKT can be derived. Although this method provides a macro-level estimate of travel demand, variations in fuel efficiency and vehicle use introduce potential inaccuracies. For instance, Hakimelahi et al., 2020 has explored how fuel consumption data (i.e. 7.5 million records of refueling data at the 64 fuel stations) can be integrated into travel demand models to enhance VKT estimation accuracy.
4. **Household and Workplace Travel Surveys** Household and workplace travel surveys involve self-reported data collection from vehicle users regarding their travel behavior. These surveys provide valuable insights into vehicle use patterns and trip purposes but are subject to potential biases due to self-reporting inaccuracies (Eisenmann and Kuhnimhof, 2018). Recent studies have demonstrated how such surveys can be integrated with vehicle ownership cost models to refine transport behavior analyses (Eisenmann and Kuhnimhof, 2018).

Each of these traditional methods has its advantages and limitations, with ongoing research focusing on improving their accuracy and cost-effectiveness by integrating emerging data sources.

2.1.2 GPS and Map-Matching for Enhanced VKT Estimation

Recent advancements in positioning technology and data processing have significantly improved the estimation of VKT. One of the most promising innovations in this field is the integration of GPS data with digital road networks to enhance accuracy and reliability in VKT measurement (Fan et al., 2019). GPS technology provides precise and continuous location data, enabling researchers to track vehicle movements over time and space. Unlike traditional survey-based methods, GPS data collection eliminates the need for periodic manual reporting, allowing for more dynamic and scalable monitoring of road usage patterns. Several studies (see Hovi et al. 2020; Fan et al., 2019; Jagadeesh

---

[6] Accessible at https://nvdb.atlas.vegvesen.no/

[7] https://www.ssb.no/transport-og-reiseliv/landtransport/statistikk/kjorelengder

et al., 2004) have explored the application of GPS-based models in transport research, demonstrating their potential to improve VKT estimation accuracy.

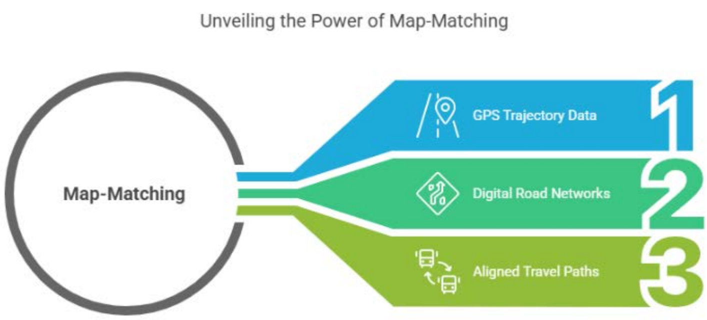

Figure 2.2: The key components of the map-matching process, which enhances the accuracy of GPS-based vehicle tracking. The three main elements include: (1) GPS Trajectory Data, which captures vehicle movement patterns; (2) Digital Road Networks, which provide the reference framework for aligning travel paths; and (3) Aligned Travel Paths, which result from processing raw GPS data through map-matching algorithms to correct positional inaccuracies and accurately represent vehicle movement on the road network.

Figure 2.2 shows map-matching algorithms are essential for aligning raw GPS data with known road networks, correcting inaccuracies caused by signal drift and other errors. In practice, regulatory authorities and transport planners increasingly rely on GPS-derived VKT data to develop congestion pricing strategies, assess road maintenance needs, and evaluate environmental impacts. For example, Fan et al. (2019) introduced a geo-computing framework using big GPS trajectory data analytics for vehicle miles traveled estimation, showcasing the application of GPS data in transport modeling and traffic management.

Despite the advantages, challenges such as data privacy concerns, the need for extensive infrastructure to process large datasets, and discrepancies in GPS accuracy across different environments persist. Advancements in map-matching algorithms, as discussed by Millard-Ball et al. (2019), aim to address these issues by improving the accuracy of GPS data in urban environments. Additionally, the development of open-source map-matching approaches, such as the one presented by Saki and Hagen (2022), offers scalable solutions for processing extensive GPS datasets.

The integration of GPS and map-matching technologies represents a significant advancement in VKT estimation, offering more accurate, efficient, and scalable alternatives to traditional methods. It makes it possible to calculate traffic work for defined geographical areas. These emerging methodologies provide unprecedented accuracy, enabling policymakers and researchers to analyze travel behavior, optimize infrastructure usage, and develop sustainable mobility solutions in real time.

2.2 Parking behaviour

A possible way to utilize GPS data might be to detect the well-known 'cruising for parking', which in urban areas can cause significant congestion and pollution (Dalla Chiara and Goodchild, 2020). We reviewed existing research on 'cruising for parking', primarily focusing on research that utilized GPS data, and preferably map-matching techniques. Based on this review, we find that map-matching offers a potential for meaningful applications in 'cruising for parking' analyses. Whilst our goal was to review existing literature with focus on commercial vehicles in general, and freight vehicles in particular (i.e. vans, trucks, HGVs, etc), we find that most literature has focused on passenger vehicles. For commercial vehicles, the literature is scarce (Dalla Chiara and Goodchild, 2020). Many have assumed that commercial vehicles park in travel lanes or loading zones close to their destinations and therefore do not ‘cruise for parking’ (Amer and Chow, 2017; Iwan et al., 2018; both as cited in Dalla Chiara and Goodchild, 2020). Due to literature focusing on commercial vehicles being scarce, we also consider reasearch focusing on non-commercial vehicles as relevant. Some of the methodologies focusing on non-commercial vehicles may also be relevant to consider for commercial vehicles.

Generally, existing research seems fragmented and includes many different terms and concepts. Although our review might therefore not provide a complete set of all the relevant literature, we believe it forms a reasonable basis to possibly discover new research areas.

To identify literature subsets applying map-matching techniques, we first screened for the search words ‘map-matching’ with and without hyphen. Next, articles including at least one of these search words were explored in greater detail for confirmation. Although this approached ensures the coverage of most relevant literature, it might miss articles in which map-matching is referred to under different terms. Often, map-matching is also referred to as a pre-processing step, and therefore not described in detail.

A notable challenge observed during our review, is that several studies with high map-matching relevance did not provide detailed or sufficient information about vehicle types. In cases where research did not provide detailed vehicle information or only terms such as ‘cars’, we therefore assumed that literature focused on non-commercial or private vehicles.

2.2.1 Private vehicles

Considering research focusing on non-commercial vehicles and ‘cruising for parking’ (or closely related topics), we identified ten studies utilizing GPS data (or similar data), of which seven utilized map-matching techniques. Table 2.1 illustrates our identified studies and whether they applied map-matching techniques.

Table 2.1: Research on 'cruising for parking' focusing on non-commercial vehicles utilizing GPS data (or similar data). Italicized text indicates direct quotation from the articles.

Authors/aim of study	Data/vehicle type/limitations?	Relevant results on 'cruising'	Map-matching
Milia et al. (2023) examines the detection of cruising, using vehicle GPS traces.	GPS data/ cars	After the cruising detection, about 800,000 GPS trips were used to estimate and validate an offline machine learning algorithm to forecast the cruising time in three different urban areas in the City of Copenhagen, Denmark, with clear distinct parking conditions. Neighborhoods were divided into spatial cells for which hourly cruising times were estimated. Feed-Forward Neural Network (FFNN) and eXtreme Gradient Boosting (XGBoost) architectures were tested as machine learning algorithms and outperformed a simple moving average (RMSE gains from 62.01 to 52.57 s). The present study paves the ground for the exploration of large datasets with GPS trajectories in urban areas for tackling the lack of information on parking search. Despite the improved overall prediction power, the potential errors from the cruising detection method, lack of data needed to capture patterns when cruising time is high or the existence of many missing values due to aggregation of data could be the reason for the observed algorithm's inability to predict the larger values of cruising time.	Yes
Paidi et al. (2022) aims to estimate CO_2 emissions as well as cruising distances at an open parking lot.	Videos collected by a thermal camera at a parking lot/unknown vehicle type	The majority of drivers tend to choose parking spaces near a shopping center, and they prefer to cruise non-optimal distances to find an empty parking space near the shopping center. The observed mean non-optimal cruising distance is 2.7 times higher than the mean optimal cruising distance. Excess CO_2 emissions and non-optimal cruising were mainly observed during visitor peak hours when there were limited or no empty parking spaces.	Yes
Millard-ball et al. (2020) argues that cruising for parking may not be as onerous as previously thought and that few vehicle travels are caused by "cruising". Using a dynamic programming model, a large-scale GPS-dataset retrieved from San Francisco and a Household Travel Survey in California, the authors reconcile these perspectives.	GPS data and video/cars	The model predicts that when parking is perceived to be scarce, drivers are more willing to take a convenient available space, even if it is some distance from their destination. We conclude that cruising for parking is self-regulating, and that in certain circumstances parking scarcity can even reduce vehicle travel.	Yes
Weinberger et al. (2020) aims to develop a method using GPS data to provide policy-relevant insights on cruising	GPS data and video/cars	Our results suggest that cruising in San Francisco, CA and Ann Arbor, Michigan is acute in some locations but overall experienced in less than 5–6% of vehicle trips, and that it accounts for less than 1% of vehicle travel in these cities–considerably less than in previous estimates.	Yes
Jones et al. (2017) aims to reduce parking search times using an automated real-time parking system developed by the authors and which they refer to as ParkUs 2.0.	GPS data/cars	In the city of Bristol alone, we have shown, using our collected trip and publicly available census data that over 790 metric tons of CO2 is generated every year due to cruising. At a total cost of £368, 000 (US$467, 000) in terms of fuel wasted.	Yes
Using FCD data **Mannini et al. (2017)** proposes a model for on-street parking search time estimation, before vehicles reach their destination.	Floating Car Data/unknown/ anonymized vehicle type	The proposed model is suitable to be used either in off-line applications for planning or in on-line tools to assess user's information and dynamic routing. Future developments mainly regard the use of more efficient algorithm of mapmatching, the specification and calibration of different gap functions as well as the application to more detailed zoning systems. Moreover, if actual privacy constraints on FCD data will be solved, it will be also interesting to specify and calibrate models for different vehicle types.	Yes
Hampshire et al. (2016) used videos and GPS for analysis of Parking Search Behaviour.	GPS data and video/cars	It was found that 30% of the drivers generated more than 70% of the meters cruised. This finding suggests that the search strategies of a few drivers disproportionately affect the many.	Yes

Authors/aim of study	Data/vehicle type/limitations?	Relevant results on 'cruising'	Map-matching
Bisante et al. (2023) provides a literature review on cruising detection and a new modelling approach for cruising behaviour.	GPS data/cars/only 9 users in the dataset and no information on the timestamp when "cruising" starts.	Our final cruising detection system achieved high accuracy and can be used to provide drivers with parking availability information and reduce traffic congestion. The final cruising for parking detection system can recognize, with a high level of accuracy (93%) and a nominal delay rate (177.43s), when a driver is looking for parking, addressing RQ2.	No
Mantouka et al. (2021) aimed at identifying factors affecting parking search time duration and compares the performance of different modelling approaches.	GPS data collected by smartphone/ unknown vehicle type	Findings reveal that the time of day in which the trip was performed, as well as trip duration and length, significantly affect parking searching duration.	No
van der Waerden et al. (2014) investigates parking search behaviour using GPS data.	GPS data/cars	It appears that GPS tracking can be used to investigate both the temporal and spatial aspects of parking search behavior. The average parking search time found in this study is 1 min and 18 s (approximately 14 % of the total travel time). The use of street segments for parking is influenced by distance to the city center, distance to nearest parking facility, presence of shops, and parking tariffs.	No

Paidi et al. (2022) used thermal cameras on an open parking lot to identify 'cruising trajectories' for drivers searching for parking. Their aim was to investigate both cruising distances and CO_2 emissions from the vehicles at the parking lot. Based on the videos, they applied 'Mod Yolo and Kalman filters' to calculate trajectories of 316 drivers. Map-matching was further used to calculate cruising distances based on the calculated trajectories. Results of the study indicated that non-optimal cruising and excess CO_2 emissions mostly occurred when parking space was limited or empty, during peak hours.

Weinberger et al. (2020) utilized GPS data in Michigan, San Francisco, CA and Ann Arbor to provide relevant policy insight from cruising. They applied map-matching to calculate the length of each path. However, they noticed that current map-matching software sometimes results in errors, when applied on low-quality or circling GPS data. For this reason, they developed a new map-matching algorithm named 'pgMapMatch', to better cope with U-turns and circling, by also allowing for analysis on low-resolution GPS data. A separate paper documents their new algorithm (Millard-ball et al. (2019). The study's results indicate that cruising occurred in less than 5-6% of all vehicle trips (Weinberger et al., 2020).

Millard-ball et al. (2020) proposed a 'dynamic programming model of parking search', applying the methodologies and data described in Weinberger et al. (2020). Their results suggested that drivers may be more willing to take a parking spot some distance from their destination, when parking possibilities are scarce. Milia et al. (2023) also applied the 'pgMapMatch' algorithm (Millard-ball et al., 2019) on GPS data. Their aim was to identify 'cruising', by utilizing GPS traces from large GPS datasets. First, they identified 'cruising', using GPS data. Secondly, a machine learning algorithm was estimated and validated on GPS data and used to predict urban cruising times in Copenhagen.

Jones et al. (2017) developed a real-time parking system (ParkUs 2.0) with the aim of reducing parking search times. Based on location data collected via smartphones carried by drivers, a machine learning model was used to classify cruising behaviour. GPS data was pre-processed by map-matching. Referring to cruising detection, their system obtained an accuracy of 81%.

Manninni et al. (2017) utilized Floating Car Data to model parking search time. Map-matching was used to connect the Floating Car Data to a transport network. Hampshire et al. (2016) utilized both video and GPS data to analyse parking search behaviour. Analysing GPS traces involved the use of map-matching. The results indicated that above 70% of the cruising distance was caused by 30% of the drivers.

Other examples for which we were not able to identify the use of map-matching techniques, include Bisante et al. (2023), Mantouka et al. (2021) and Van der Waerden et al. (2014). Bisante et al. (2023) used a 'Boosted Tree classifier' to develop a smartphone application, able to detect 'cruising for parking' automatically. Mantouka et al. (2021) and van der Waerden et al. (2014) utilized GPS data to investigate parking search times, amongst other things.

In summary, we identified seven articles focusing on 'cruising for parking' for non-commercial vehicles, using map-matching techniques. For three additional articles utilizing GPS data and focusing on 'cruising for parking', we were not able to detect the use of map-matching techniques.

2.2.2 Commercial vehicles

In contrast to our identified research focusing on non-commercial vehicles, we identified little or no similar research applying the same techniques when focusing on commercial vehicles. However, several researchers have addressed commercial vehicle's parking behaviour, 'cruising for parking' or closely related topics, by utilizing raw GPS data or digital tachograph data.

Ghizzawi et al. (2024) performed a literature review on research considering parking behaviour for commercial vehicles. They identified six articles that utilized either GPS data or digital tachograph data. We reviewed and summarized these articles below:

- Sharman et al. (2012) focused on predicting stop durations for commercial vehicles using GPS data, both with and without establishment data.
- Mahmud et al. (2020) used GPS data to study the impact of amenity availability on parking usage.
- Mjøsund and Hovi (2022) utilized GPS data in urban areas to derive pickup/delivery locations and delivery times from stops.
- Seya et al. (2020) used digital tachograph data to investigate parking durations and parking decisions.
- Dalla Chiara and Goodchild (2020) provided the first extensive empirical analysis of commercial vehicles searching for parking, with the aim of empirical evidencing searching for parking and how this phenomenon is influenced by curb-space allocation. Their methodology was to detect parking behaviour between several departure and arrival locations for commercial vehicles, by calculating time deviation between the actual recorded time (i.e between the departure and arrival location), and the fastest time retrieved (also between the departure and arrival location) from a routing engine. They performed regression analysis to evaluate how parking infrastructure, private loading bays and off-street parking influence time deviations. Results indicated, among other things, a cruising time of 2.3 minutes per trip on average.
- Parts of the methodology in Dalla Chiara and Goodchild (2020) were applied in Dalla Chiara et al. (2022) to investigate how cruising for parking was affected by providing drivers with information on real-time curb availability, by also accounting for distance.

Other relevant research we identified includes Dalla Chiara et al. (2021). For this study, the authors travelled along with commercial vehicle drivers in Seattle and retrieved GPS data, to collect information on decisions made by drivers. They also estimated ‘cruising for parking’ and found an average cruising time of about 3.8 minutes. Results showed that about 80% of the operating time, the vehicles stayed parked.

Saki and Hagen (2024), in turn, identified parking search routes by using a deep learning neural network on GPS data. Their dataset included, amongst others, GPS data collected for delivery trucks, mostly in Frankfurt. Additionally, some arguments suggests that they applied map-matching techniques on their GPS data by mapping ‘Mean Parking Search Duration’ onto road segments in OpenStreetMap. However, we have not been able to investigate this further.

Ranjbari et al. (2023) applied a ‘heavy vehicle driving simulator’ to investigate parking and delivery-behaviours among commercial vehicle drivers. Nevland et al. (2020) used GPS data collected by ‘heavy commercial vehicles’. Based on the GPS data, they aimed to identify parking locations. Another study proposed analysing parking searches using a non-spatial model, as well as a geo-simulation model (Martens et al., 2010). For this study, we were unable to establish whether it considered commercial or non-commercial vehicles.

To the best of our knowledge and except for the study by Saki and Hagen (2024) for which we were unable to establish whether map-matching techniques had been used, none of the aforementioned studies focusing on commercial vehicles and GPS data, applied map-matching techniques. They were rather based on raw GPS data, or only departure/arrival locations retrieved from GPS data. Table 2.2 summarizes four research articles focusing on commercial vehicles, for which we were able to identify the use of GPS data in ‘cruising for parking’ analysis or closely related topics.

Table 2.2: 'Cruising for parking' research focusing on commercial vehicles, utilizing GPS data. Italicized text indicates a direct quotation from the articles.

Authors/aim of study	Data/vehicle type	Relevant results on 'crusing'	Map-matching
Saki and Hagen (2024) gathered ground truth data on parking search durations using a self-developed app and used a model to identify parking search routes along with search duration.	GPS data/Delivery trucks	(...) we developed a deep learning neural network model that accurately identifies parking search routes in GPS data and predicts search duration. Our model outperforms existing parking search identification models proposed in previous studies. This generates the first reliable statistics on parking search durations and reveals key insights about parking search patterns in this city. Notably, the predicted mean parking search duration from this extensive dataset, comprising over 860,000 journeys, is approximately 1.5 min.	Unclear
Dalla Chiara et al. (2022) investigated how providing delivery drivers with information on curb availability, affects cruising for parking	GPS data/ Commercial vehicles	The data collected showed that when curb availability information was provided to drivers, their cruising for parking time significantly decreased by 27.9 percent, and their cruising distance decreased by 12.4 percent. These results demonstrate the potential for implementing intelligent parking systems to improve the efficiency of urban logistics systems.	No
Dalla Chiara et al. (2021) travelled along with commercial vehicle drivers in Seattle and retrieved GPS data, to collect information on decisions made by drivers. They also estimated 'cruising for parking'.	GPS data/ Commercial vehicles	Cruising: Drivers searched for available parking; given the observed data, the estimated average cruising for parking time was 3.8 minutes.	No
Dalla Chiara and Goodchild (2020) provided the first extensive empirical analysis of commercial vehicles searching for parking, with the aim of empirical evidencing searching for parking and how might search for parking are influenced by curb-space allocation	GPS data/ Commercial vehicles	We obtain an average estimated parking cruising time of 2.3 min per trip, contributing on average for 28 percent of total trip time. We also found that cruising for parking decreased as more curb-space was allocated to commercial vehicles load zones and paid parking and as more off-street parking areas were available at trip destinations, whereas it increased as more curb space was allocated to bus zone.	No

3 Hidden Markov Model Map-Matching

Figure 3.1 shows an example of map-matching in an OpenStreetMap graph. The Appendix provide an extensive description of the HMM-algorithm. In the figure, GPS points are marked as black dots along the road segments. The orange points display candidates within 30 meters of each GPS point. Each GPS point corresponds to at least one candidate, given that road segments exist within a certain radius (in this case 30 meters) of the GPS point. Consequently, candidates and GPS data are used to evaluate the probability of a given route choice. The optimal sequence for candidates (i.e. the sequence of candidates resulting in the highest joint probability) is illustrated with their respective IDs. The green line indicates transitions between the candidates, i.e. the matched route.

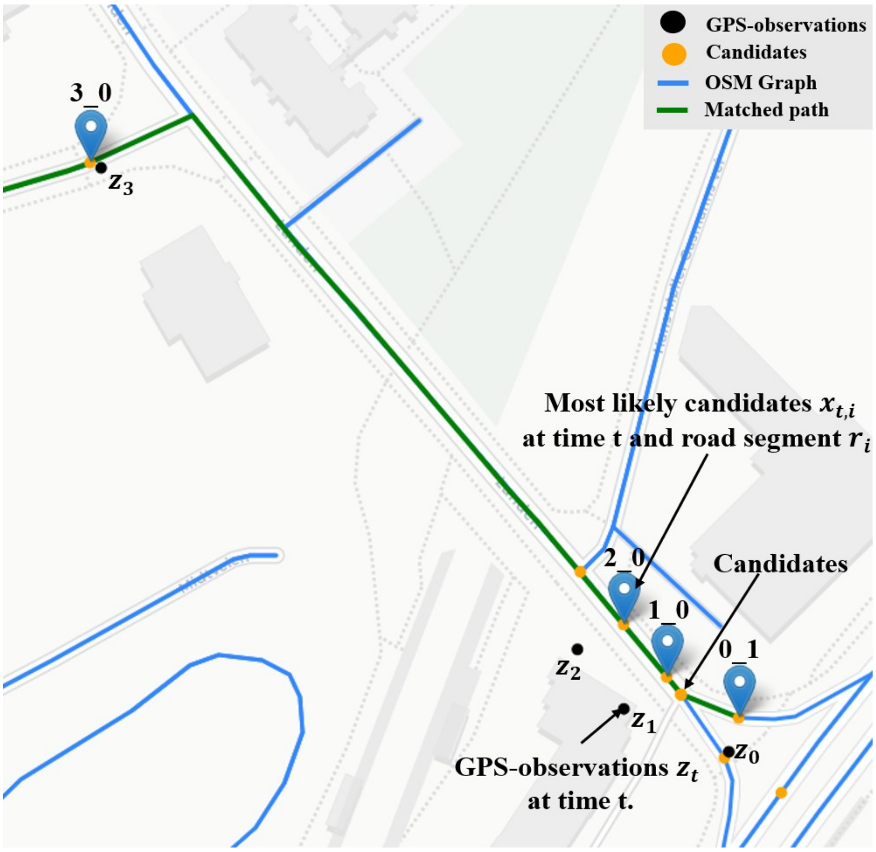

Figure 3.1: Example of a matched path retrieved from PyTrack (Tortora et al., 2022).

3.1 Data and pickup/delivery activities

In this report, we utilise GPS data collected in LIMCO (Hovi et al., 2021), originally collected from multiple freight forwarders in Norway during the period 2019-2021. The data consist of GPS data retrieved from approximately 1650 trucks/HGVs and 200 vans. Mjøsund and Hovi (2022) utilized the data retrieved in 2019. Although they had access to data from a large number of vehicles, they pointed out that the user pattern of these vehicles is not necessarily representative given that GPS devices were more commonly used in HGVs, compared to vans.

By analysing the GPS datapoints, combined with timestamps, Mjøsund and Hovi (2022) aimed at separating actual stops (i.e. resulting from pickup/delivery activities), from what is considered as driving activities for vans and HGVs. As such, several criteria were applied (e.g. cut-offs when vehicles were moving at under or over 8 km/h between GPS points). To prevent GPS errors and minor vehicle movement from affecting which events are considered as stops, the authors built upon methodology from Gingerich et al. (2016). By doing so, GPS points registered within a radius of 250 meters from the first GPS point were also considered as stops. For stops taking place within 'inner city areas', this radius was reduced to 50 meters. Along with the GPS points considered as stops, the time difference between timestamps was accumulated for each point. New trips were generated each time the stop-time exceeded 60 minutes. Mjøsund and Hovi (2022) found that short stops are typically not associated with pickup and deliveries, but rather traffic stops along the road (e.g. at traffic lights). For that reason, they defined primary stops as stops over 2 minutes for vans and over 5 minutes for HGVs, with all stops below these thresholds being considered stops during driving activities. In the remainder, we refer to primary stops as pickup/delivery stops.

3.2 Data and preprocessing steps

Based on the data obtained in the LIMCO project (Hovi et al., 2021), we selected a representative day of GPS observations in 2020, only considering van activities in the Norwegian counties of Oslo and Akershus. Furthermore, we selected GPS data coming from 31 distinct vehicles that day, with approximately 20 000 GPS observations. This resulted in a subset of data that we could study using map-matching algorithms.

To eliminate data-points resulting from pickup/delivery activities, we decided to exclude all GPS points belonging to stops over 2 minutes. This includes all the primary stops as defined in Mjøsund and Hovi (2022). As mentioned, these authors also defined new routes whenever stop times exceeded 60 minutes for a particular vehicle. Instead of defining new routes, we decided to exclude the points belonging to stops over 60 minutes and continuing the route of the particular vehicle. In this way, we obtained a single route for each vehicle on the selected day.

Considering 'cruising for parking' analysis, excluding GPS points belonging to stops over 2 minutes, may not be advisable as some of them may involve 'cruising'. However, as mentioned in chapter 2, Weinberger et al. (2020) noticed that current map-matching software sometimes results in errors when applied on low-quality or circling GPS data, and consequently developed a new map-matching algorithm to better cope with these challenges (documented in Millard-ball et al., 2019). With this in mind, it is likely that our algorithm would result in errors if applied on GPS points involving pickup/delivery activities. On the other hand, excluding GPS points belonging to stops over 2 minutes may provide a more accurate basis for estimating traffic work. This is strongly suggested by the fact that our GPS data belonging to stops over 2 minutes only accounts for a small distance (i.e. the distance between consecutive GPS points), but a considerable high fraction of the number of GPS points in total. The exact figures of excluded GPS stops over 2 minutes, are described in greater detail at the end of this subsection, referring to Table 3.1.

Referring to our pre-processing steps based on Mjøsund and Hovi (2022), similar techniques have been used by other researchers as well. Huang et al. (2021), reviewed different approaches:

1. The points are removed if the speed is below 3 m/s and the consecutive distance between the points does not exceed the threshold
2. Another approach was to detect parking behaviour if the vehicle was travelling less than 50 meters within 30 minutes.

Others, like Newson and Krumm (2009) and Koller et al. (2015), included only points that were within 2 standard deviations from the previous point. Another example includes Yang and Gidòfalvi (2018), who removed points where the trajectory consisted of only 1 point, the average speed exceeded 150 km/h, or the boundaries of their road network were not sufficient to cover their geometries.

As such, our approach may differ in that we removed points resulting from pickup/delivery stops and long stops exceeding 60 minutes, while stops below these thresholds were considered part of the driving activity. In this way, we sought to model only the driving paths resulting from pickup/delivery stops in urban areas.

In our data, most vehicles both started and ended their journeys at Alfaset in Oslo, where multiple Norwegian transport service providers are located. Additionally, most vehicles periodically departed from and arrived at Alfaset multiple times daily, with an average of 1.7 times on the particular day. This was calculated by counting the number of cases with GPS point departures and arrivals to/from a 500-meters radius from Alfaset. On average, the vehicles used about 4.2 hours between each departure and arrival at Alfaset. However, due to noise and sparsity in the data, the number of departures and arrivals at Alfaset is likely higher. In general, the actual number of departures and arrivals at a terminal location might also be higher due to the fact that not all vehicles stop at Alfaset, and there may be multiple other terminal locations as well. Even though most of the vehicle routes consisted of multiple departures and arrivals daily, we considered only the vehicles' overall routes for the particular day.

Table 3.1: Selected GPS observations for unique vehicles, retrieved from Hovi et al. (2021).

Vehicle ID	Number of GPS points			Length of Route (km)		
	Cleaned	Original	Δ (%)	Cleaned	Original	Δ (%)
1	615	1093	43.7	74.4	80.8	7.9
2	137	441	68.9	41.0	49.2	16.7
3	227	794	71.4	83.6	103.7	19.4
4	496	1013	51.0	98.0	111.0	11.7
5	631	1193	47.1	173.3	188.5	8.1
6	463	878	47.3	145.9	156.8	7.0
7	154	416	63.0	115.8	129.5	10.6
8	166	277	40.1	196.0	202.9	3.4
9	45	102	55.9	31.5	34.0	7.4
	2934	6207	54.3	959.5	1056.4	10.2

Table 3.1 illustrates 9 of the 31 individual vehicles on the selected day, along with the number of selected GPS points, the original number of GPS points, and distance between the selected GPS points. These distances were calculated by using Geopandas in Python, along with a UTM coordinate reference system. This gave us a good approximation of the Great Circle Distance between the GPS points[8].

On average, around 50 percent of the GPS points (considering all the 31 individual vehicles) was removed due to assumed stop activity resulting from pickup/deliveries and stops over 60 minutes. Even though this seems a lot, we found that the average difference between the two distances (i.e. before and after cleaning), was only about 10 percent (considering all the 31 individual vehicles). This means that about 50 percent of the excluded GPS points only account for about 10 percent of the distance, while the remaining GPS points account for ca. 90 percent of the distance. Our identified GPS stops may also be affected by signal inaccuracy, causing GPS positions to vary over time, even though the vehicle is idling. This phenomenon, also known as jiggle (McCormack, 2014), may lead us to overestimate the GPS distance at stops over 2 minutes if they were included in the data. This might suggest that the excluded GPS stops accounts for even less than 10 percent of the distance.

[8] This methodology was used for all cases for which we refer the "Great Circle Distance", except for distances used in the PyTrack map-matching process.

Furthermore, excluding GPS stops over 2 minutes may also avoid cases with GPS points located at parking lots/infrastructure, that may not be present as road segments in OpenStreetMap[9].

3.3 Data-quality

In general, the GPS trajectories are quite noisy and sparse. Removing points resulting from pickup/delivery activities and stops over 60 minutes resulted in even more sparsity in our data. This made it difficult to identify the vehicles' trajectory when manually interpreting the GPS points. In fact, based on the data, we do not know the actual routes.

Figure 3.2 illustrates a histogram of the length (measured in UTM) and time (in seconds) between consecutive GPS points, considering all GPS data from the 31 vehicles. Nearly 90 percent of the GPS observations are within 700 meters or 2 minutes and 20 seconds. On average, the distance between GPS points is 311 meters, with an average time difference of 2 minutes and 3 seconds. However, median values provide a more accurate representation as they are less influenced by outliers, which could result from excluding points associated with pickup/delivery activities and extended stops. The median values are approximately 102 meters for the distance and 18 seconds for the time interval between consecutive GPS points. Using these median values, we calculated an average speed of about 20 km/h. In contrast, Mjøsund and Hovi (2022) calculated average speeds of about 18 km/h in inner cities and 20 km/h in outer cities, which is in line with our estimate. The average speed of 20 km/h aligns with expected speeds for urban delivery trucks. Urban freight vehicles often encounter numerous factors that influence their speed, such as high traffic volumes, especially during peak hours, which can significantly reduce average speeds.

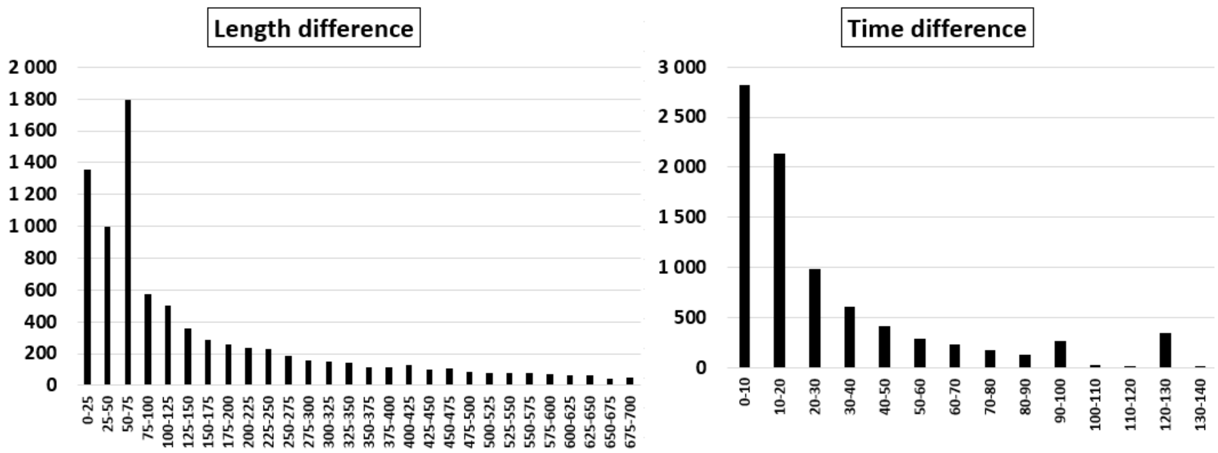

Figure 3.2: Histograms by length (in meter intervals) and time (in second intervals) between consecutive GPS points, considering all the GPS data from the 31 vehicles.

Based on the matched vehicle routes described later, we attempted to estimate the magnitude of GPS noise by measuring the Great Circle Distance between GPS points and their respective matched candidates. Our findings indicate that, on average, GPS points deviated by 8.36 meters from the matched candidates. The standard deviation of these differences was 6.34 meters, and the median absolute deviation was 2.62 meters. However, due to the lack of exact ground truth paths in our data, there is a significant degree of uncertainty in these measurements.

[9] We noticed that parking lots were often not present as road segments in our road network graph from OpenStreetMap.

Despite the noise and sparsity in our GPS data, previous research supports the robustness of map-matching algorithms under such conditions. For instance, Newson and Krumm (2009) found that their HMM algorithm performed well even with intervals of up to 30 seconds between consecutive GPS points or GPS noise with a standard deviation of up to 50 meters. This finding suggests that sparse and noisy GPS data can still yield reliable results when processed with sophisticated algorithms like HMM. However, Quddus and Washington (2015) point out that map-matching algorithms are generally most suitable for high-frequency GPS data, and that the accuracy level may drop to about 70 percent if these algorithms are applied to low-frequency GPS data. Additionally, as mentioned in chapter 2, Weinberger et al. (2020) aimed at providing relevant policy insight from 'cruising' and noticed that current map-matching software sometimes results in errors, when applied on low-quality or circling GPS data. They developed a new map-matching algorithm to better cope with these challenges (documented in Millard-ball et al., 2019). Furthermore, it is important to note that the HMM algorithm assumes that vehicles always take the shortest path between matched candidates (Newson and Krumm, 2009). This assumption can lead to inaccuracies, particularly when the distance between GPS points is substantial (i.e., the Great Circle Distance). In such cases, significant information about the true path is lost, and the algorithm's assumption of the shortest path may not reflect the actual route taken by the vehicle. Quddus and Washington (2015) also noted that especially buses tend to follow pre-defined routes that may not correspond to the shortest path between GPS points.

A closer inspection of the GPS data also revealed instances where vehicles appeared to drive back and forth on the same road links multiple times. This might be due to 'cruising for parking'. However, our data includes no information on the actual pickup/delivery stops, i.e. we assumed that stops over 2 minutes are pickup/deliveries. Therefore, it has not been possible to investigate the studied vehicles' 'cruising for parking' activity in greater detail. Chapter 5.2 nevertheless illustrates how such analysis might be performed on a general level for commercial vehicles, using map-matching.

The back-and-forth movements of the vehicles might also be due some other reasons. This includes operational needs such as multiple deliveries on the same street but can also stem from errors in the GPS data collection process.

3.4 Proposed framework

Based on the literature regarding map-matching and GPS data, we suggest a framework to simplify and guide the process for map-matching. Our suggested framework is illustrated in Figure 3.1 and outlines the different required steps in performing map-matching. Depending on the type of tool used, the process may differ a bit from our suggested framework. We performed map-matching using PyTrack[10]. PyTrack, and HMM-based map-matching formulated by Newson and Krumm (2009), formed the basis for our framework, with some modifications and additions, primarily concerning preprocessing steps, and, on a general level, the application of map-matching results. We believe that this framework can give insights into the map-matching process at both a general and practical level.

[10] https://github.com/cosbidev/PyTrack

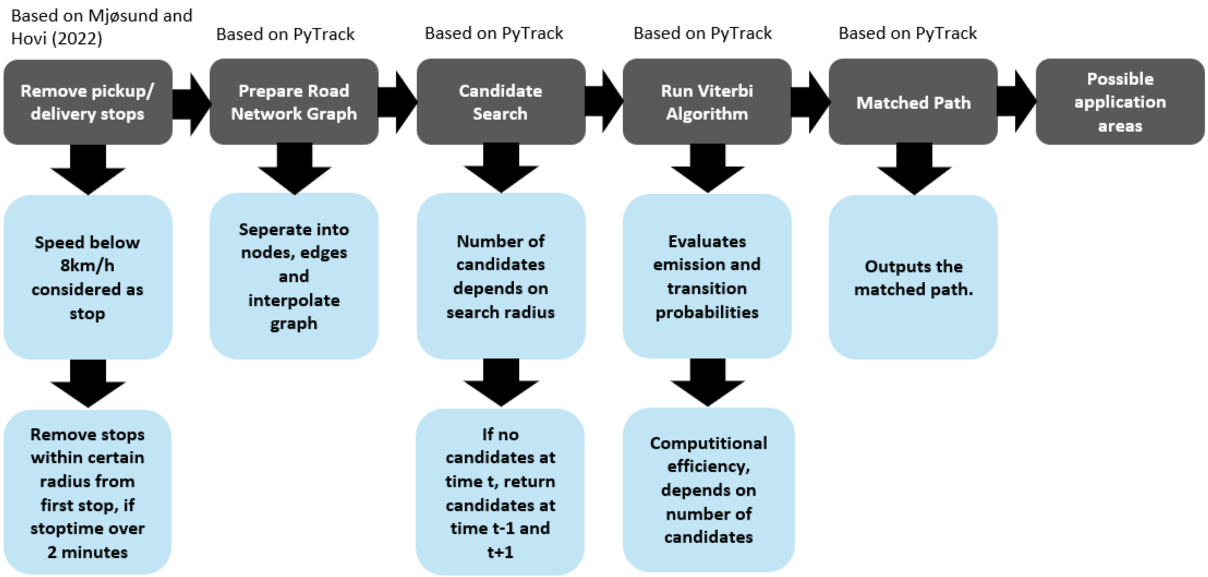

Figure 3.1: A Map-Matching Framework on Sparse GPS Data.

The framework starts by removing stop-observations associated with pickup/delivery activities, as described in chapter 3.2. Next, to enable the algorithm to search for candidates, we include a road network graph from OpenStreetMap. Thereafter, we define a search radius for candidates, described in greater detail in chapter 4.1. If no candidates are found for a given observation at time t, the algorithm returns only candidates at time t-1 and t+1. Next, the identified candidates are evaluated using the Viterbi Algorithm and emission and transition probabilities, with the goal of finding the most likely path, through the candidates at t times. This results in a matched path that can provide a basis for various application areas. Chapters 4.2 and 4.3 provide extensive discussions on what we can learn from map-matching in urban areas as well as aggregated driving paths. On a general level, resulting matched paths may also form the basis for traffic work estimations and ‘cruising for parking’ analysis, as we describe in chapter 5. However, considering ‘cruising for parking’ analysis, it may not be advisable to implement our suggested pre-processing steps. Due to our exclusion of stops over 2 minutes (and GPS-points within a certain range of these stops), it is likely that the excluded points involve ‘cruising’ activity.

4 Results

4.1 Performance of the algorithm

The calculation time represented a significant challenge when running the Hidden Markov Model (HMM) algorithm. This is a known issue in the field, and several researchers have proposed solutions to address it. Koller et al. (2015) suggested replacing the Viterbi algorithm with a Bidirectional Dijkstra algorithm to speed up the matching process. This modification enabled them to reduce the processing time by up to 45% without negatively affecting matching results. Yang and Gidòfalvi (2018) introduced the Fast Map-matching (FMM) method, which uses precomputation combined with an HMM. During the precomputation stage, a table is created that stores origin-destination pairs of shortest paths within a certain length in their road network. This precomputed table then allows for a hash table search instead of routing queries, which are known bottlenecks in map-matching. Experiments in the same study demonstrated that the FMM algorithm could significantly improve the performance of map-matching. Liao (2023), in turn, developed a three-dimensional HMM model that integrates GPS altitude data with a neural network. This approach improved both accuracy and processing time, outperforming other algorithms.

To speed up our algorithm, we implemented a simplistic approach by setting a 30-meter search radius around each GPS point. PyTrack also enabled us to interpolate the OpenStreetMap graph to speed up the process (see details in chapter 4.2). Like the method used by Newson and Krumm (2009), a reduced search radius resulted in fewer candidate points that our algorithm had to consider, thereby reducing the running time. Given a search radius of 30 meters, our algorithm was able to match about 3.3 GPS points per second on average, when using a 6 core Intel i9 processor, along with 32 GB of memory. The calculation time includes candidate search and running the Viterbi algorithm in PyTrack. We also did some initial testing using Open-Source Routing Machine[11] (OSRM) for map-matching, via Docker Desktop[12] that we set up on our computer. OSRM is written in C++ and enables high performance matching. In our testing, OSRM outperformed PyTrack in terms of speed. However, we ultimately chose PyTrack due to its convenient approach to exploring the matching process, as the main focus of this report is on simplifying and guiding the map-matching process.

Setting a smaller search radius resulted in some GPS points having no candidates. In such cases, we calculated the vehicle route between the candidates (based on GPS) at time t-1 and t+1. As this scenario was rare, it did not significantly alter the overall matched path. Some vehicle routes failed to complete due to the limited search radius. By increasing the search radius up to 100 meters, we managed to match all routes except two, namely for *Vehicle 14* and *15*. However, further increasing of the search radius proved to be extremely time-consuming, leading us to terminate the map-matching for these vehicle routes. We identified two potential reasons for the failed matching:

- High levels of GPS noise and sparse data points can lead to difficulties in finding suitable candidates within the search radius.
- Missing road segments in OpenStreetMap or road segments not connected to the rest of the network can result in failed candidate searches.

The noise and sparsity of the data introduced several difficulties. In some cases, this made it challenging to manually identify the most likely paths of the vehicles, particularly when there were

[11] https://github.com/Project-OSRM/osrm-backend

[12] https://www.docker.com/products/docker-desktop/

considerable distances between consecutive GPS points. This issue posed a significant challenge in evaluating the performance of our map-matching algorithm.

To evaluate our algorithm's accuracy, we graphically displayed the GPS points alongside the matched paths. This visual inspection allowed us to identify cases where the matching appeared unlikely compared to the actual GPS points. Through this methodology, we were able to practically assess the algorithm's performance for several vehicle routes. In most instances, the matched paths seemed plausible and consistent with the sequence of GPS points. This suggests that the algorithm was generally successful in reconstructing the likely routes of the vehicles. The algorithm appeared to manage GPS noise effectively, producing matched paths that aligned well with the GPS data despite the inherent noise. However, a significant challenge in our evaluation was the absence of the true path of the vehicles. This lack of ground truth introduced uncertainty into our assessment, as we could not definitively confirm the accuracy of the matched paths. A similar issue has been discussed by Saki and Hagen (2022), who faced difficulties in measuring the accuracy of their map-matching due to unknown actual GPS locations. They proposed the use of artificial GPS points combined with a deviation metric as a solution.

As mentioned in chapter 3.3 on Data Quality, some vehicles appeared to drive back and forth on the same road links multiple times. These occurrences, likely due to 'cruising for parking' activity, made our practical evaluation even more challenging. It is also possible to implement filters to remove redundant GPS points that indicate back-and-forth movements without significant progress.

Additionally, some GPS points led to unlikely turns due to missing road segments in OSM, as illustrated in Figure 4.1, which shows the matching of an example route from Vehicle 31. The blue lines represent the road network graph from OSM, while the green lines show the matched vehicle route. The GPS points are depicted as black dots. By examining the candidates, we discovered that the GPS points at the right corner were assigned to road segments that were not connected to other parts of the road network. However, because we extended the search radius to 100 meters for this particular vehicle route, the algorithm also considered other candidates located on connected road segments. This extension resulted in an unlikely and convoluted path. In this case, the GPS points likely constituted a different and shorter path than what was mapped, primarily due to missing road segments in OSM.

Figure 4.1: Example of a matched path that was prone to missing road segments (Vehicle 31).

4.2 What can we learn from map-matching in urban logistics?

Table 4.1, Vehicle IDs, along with detailed information regarding the search radius for candidates, the computed map-matching distance, the distance measured between consecutive GPS points (i.e., the cleaned sample), and the calculation time including the time that the algorithm uses for candidate search and the Viterbi-search. The map-matching distances were calculated using Geopandas in Python and a UTM coordinate reference system. More specific, the distance was calculated between each matched node-ID, connected to an interpolated OpenStreetMap graph. This approach provided an indication of the true routing distance, although the interpolation lowered the accuracy of the measured distance. PyTrack enabled us to interpolate the graph to speed up the process by reducing precision. Practically, this means that the distance is calculated as a straight-line distance between each starting and ending node of a link. While this method increased processing speed, it also introduced some inaccuracies in the distance measurements. The trade-off between accuracy and speed is a critical consideration in map-matching. By interpolating the road graph and calculating straight-line distances, we achieved faster computation times at the expense of precision in the routing distances.

Table 4.1: Selected map-matching results.

Vehicle ID	Search radius (m)	Length of Route (km)	Map Matched Cleaned	Δ(%)	Calculation time (min)
1	60	84.3	74.4	13.3	1.39
2	30	46.7	41	13.8	0.70
3	60	93.5	83.6	11.9	6.39
4	30	110.6	98	12.9	1.43
5	60	190.1	173.3	9.7	8.19
6	100	162.1	145.9	11.1	32.77
7	30	134.7	115.8	16.3	1.63
8	30	232.3	196	18.5	3.01
9	100	41.2	31.5	30.6	2.27
Total		1095.5	959.5	15.3	6.4

As seen in the table, the average calculation time for selected instances is 6.4 minutes. A smaller search radius representing a smaller search region and usually fewer candidate points results in a shorter calculation time. For 93.5% of the instances, a match can be found with a search radius of up to 100 meters. Initially we set a 30-meter radius to find a matched route. If the matching failed, we increased the radius to 60 meters and further extended to 100 meters if necessary. However, as mentioned previously, the matching failed for two of the vehicle routes even though we used a search radius of 100 meters.

The weighted average distance, combining all the 29 matched paths, was about 11% higher compared to the distance between the consecutive GPS points. This difference ranged from a minimum of 2.5 percent and a maximum of 30.6 percent. This outcome highlights the strength of map-matching algorithms in accurately estimating driving distances, as opposed to the substantially underestimated distances calculated from raw GPS data. In contrast, Saki and Hagen (2022) reported only a 1.4 percent longer map-matching distance compared to their GPS distance. This smaller difference is likely due to their dataset having denser GPS trajectories. With denser GPS data, more road turns and details are captured, reducing the underestimation of driving distances. The sparsity of GPS data significantly impacts the accuracy of distance measurement. In our report, the greater the distance between consecutive GPS points, the more likely it is that important road details, such as turns and curves, are missed. This leads to a higher underestimation when calculating straight-line distances between GPS points. Map-matching algorithms help mitigate this underestimation by aligning GPS points with the actual road network, thus providing a more accurate representation of the true driving distance. However, as mentioned, the HMM always takes the shortest path between matched candidates, which is not necessarily the true path taken by the vehicle, especially when the data is considerably sparse.

Even though we aimed to indicate a true routing distance, including GPS points resulting from pickup/delivery activities and long stops would have led to a higher matched distance. However, if included this may lead us to overestimate the matched path distance, due to the jiggle effect (when the vehicle is idling) described in chapter 3.2. Additionally, GPS positions may be located at parking lots/infrastructure (i.e. usually pickup/delivery locations), having no road segments in OpenStreetMap (as described in chapter 3.2).

4.3 Aggregated driving paths

Figure 4.2 provides a comprehensive view of the combined 29 matched paths within Oslo's inner city, revealing patterns in road usage and delivery activities. Although a few of the vehicle routes extended far out in the metropolitan area, this figure provides a detailed view of the urban core. Each link in the combined matched path was used 2.56 times on average, with a standard deviation of 3.13. Darker lines on the map indicate links that are used more frequently, representing more standard deviations from the mean. This graphical differentiation helps identify heavily used roads. By categorizing into standard deviations from the mean, we were able to graphically differentiate between areas. We illustrate that main roads, such as highways, are used more frequently compared to smaller inner city roads. This is because main roads serve as transit routes between delivery areas and terminal locations, while smaller roads are used more randomly for specific delivery missions. Frequent use of road links is particularly noticeable around terminal locations at Alfaset (north-eastern part of Oslo), where multiple professional transport providers are located.

By utilizing GPS data retrieved from vans in Oslo and Akershus, we have demonstrated the feasibility of accurately estimating driving distances in urban areas. Although PyTrack was shown to be slower compared to other solutions, it is important to note that many available open-source solutions for map-matching are based on the same HMM introduced and described by Newson and Krumm (2009), albeit with some modifications. Open-Source map-matching platforms like OSRM, FMM or Valhalla[13], offer faster matching performance compared to PyTrack. These platforms are optimized for speed, making them suitable for applications requiring quick processing times. On the other hand, the setup time and required skills are likely to increase. Also, many existing platforms are non-Python-based, required to run on private servers and mainly designed for the industry (Tortora et al., 2022). In contrast, platforms like PyTrack and LeuvenMapMatching offer Python-based map-matching and might be easier to use.

[13] https://valhalla.github.io/valhalla/api/map-matching/api-reference/

Figure 4.2: A comprehensive view of the combined 29 matched paths within Oslo's inner city, revealing patterns in road usage and delivery activities.

5 Application areas for map-matching

5.1 Potential traffic work estimation

We practically illustrate traffic work (i.e. VKT) estimations as one of several application areas of map-matching, using our matched paths. Then, we compare our estimates with publicly available statistics on VKT, collected using traditional methods. Due to our exclusion of GPS points belonging to stops over 2 minutes, our estimations are solely based on the driving paths resulting from the assumed pickup/delivery activities. By extracting the test instances, we calculated an average daily vehicle distance of about 107 km using our 29 matched paths. As mentioned earlier, most of the vehicles also periodically departed and arrived at Alfaset multiple times daily, with an average of 1.7 times on a particular day. However, this factor was associated with a large degree of uncertainty and is likely to be higher in practice. For simplicity, we therefore assumed this value to be 2 times daily. Dividing by this factor, we estimated an average distribution trip length of about 53.5 km. A survey undertaken by Statistics Norway in 2018[14], focused on transportation by vans, showed that distribution trips with vans, accounted for about 491.3 million km annually. This mileage was distributed over 8.5 million trips. Given these SSB figures, we estimated an average trip length of58 km, about 8 % higher than our GPS-based estimate. The statistics also allowed for braking estimates down between Norwegian counties[15], although for this level of detail, further subdivisions between distribution and non-distribution trips were not available. Breaking the statistic down to Oslo and Akershus, we found that vans were driven about 583.8 million km annually (including load), distributed over 19.2 million trips (including load). These estimates leave us with an estimated trip length of about 30.4 km, which is considerably lower than our GPS-based estimate, however they don't account for empty driving. Other challenges might also arise when comparing our matched paths with publicly available statistics, for example comparable industry segments, private ownership etc.

Another statistic undertaken by Statistics Norway, is the driving length statistics[16], which allowed us to measure the average annual driving distances of small freight vehicles (i.e. mostly vans), by age of the vehicle. Considering base year 2020, and vehicles that were between 0 and 4 years old, the statistics provided us with an annual driving distance around 17 000 km. Assuming that our vehicles were used about 230 days a year and with an average daily distance of about 107 km, we estimated an annual driving distance of about 24 584 km, almost 50 percent higher than the statistic. However, our small sample of vehicles and the fact that we only considered vans used for distribution and one day of a year may affect representativeness. Seasonal variation may also play a crucial role. Assuming that our vehicles were used 230 days a year might also be overly optimistic, considering that not all vehicles are necessarily in use during off-peak seasons.

Even though our estimations of traffic work suffered from a lot of uncertainty, we believe that our considerations could be utilized and implemented using larger GPS samples, that to a larger degree cover seasonal variations. This, in turn, would require faster matching-performance, where Open-Source map-matching platforms like OSRM, Valhalla or FMM, could be possible candidates.

[14] https://www.ssb.no/statbank/table/07293/
[15] https://www.ssb.no/statbank/table/07298/
[16] https://www.ssb.no/statbank/table/12575/

5.2 Detecting 'cruising for parking'

Another way to utilize our map-matching results might be to detect 'cruising for parking' for our commercial vehicles in the dataset. As mentioned previously, we revealed instances of GPS data where vehicles appeared to drive back and forth on the same road links multiple times, likely due to 'cruising for parking' activity. However, a significant lack in our ground truth data is that we lack information about the actual pickup/delivery locations, and only have an assumption that all stops above 2 minutes are pickup/deliveries.

Furthermore, it is likely that our excluded GPS points belonging to stops over 2 minutes, involve 'cruising' activity. For more accurate 'cruising' detection, we therefore advise to adjust our pre-processing steps to not exclude GPS-points that may involve 'cruising' activity. For our case, observations by Weinberger et al. (2020) on possible errors when applying map-matching software on low-quality or circling GPS data, means that inclusion of GPS points belonging to stops over 2 minutes would potentially result in more errors. Furthermore, the authors referred to in Weinberger at al. (2020), developed a map-matching algorithm to better cope with these challenges (documented in Millard-ball et al., 2019). This algorithm has been used in later research as well (Millard-ball et al., 2020; Milia et al., 2023). Therefore, a significant effort should also be put into a consideration of the most appropriate map-matching engine when we consider 'cruising for parking'. Other map-matching techniques applied in previous research focusing on 'cruising for parking' for non-commercial vehicles, might also be relevant to consider when performing similar analyses for commercial vehicles.

Figure 5.1 shows an example of a 'cruising for parking' analysis, by partly adopting of the methodlogy used in Dalla Chiara and Goodchild (2020), Dalla Chiara et al. (2021) and Dalla Chiara et al. (2022). The grey grid represents the road segments. The green line shows the matched path between the two pickup/delivery stops, marked as dark points. The orange line represents the fastest path between the two pickup/delivery stops. In this example, the matched path obviously doesn't constitute the shortest path. The excess distance might be caused by 'cruising for parking' activity.

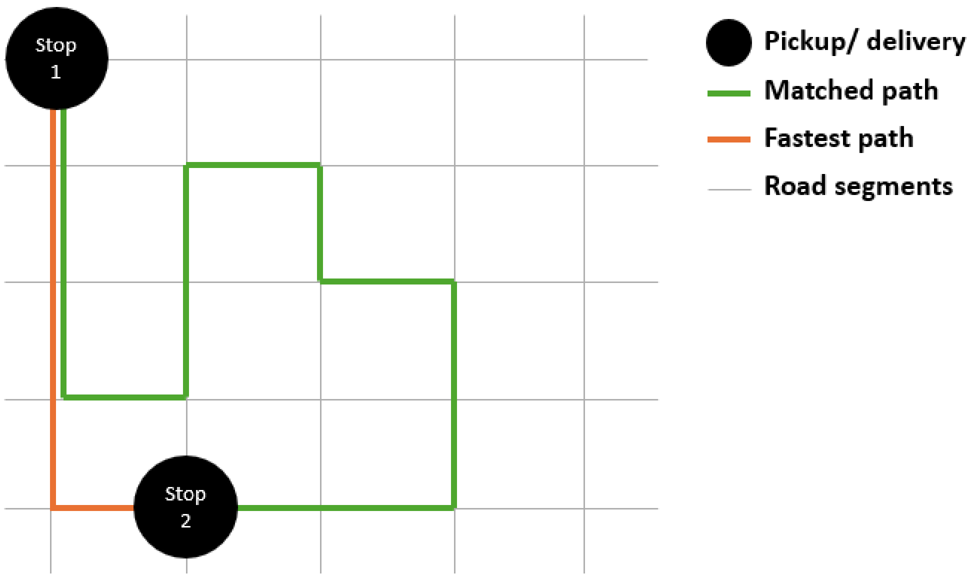

Figure 5.1: Fictive matched path, fastest path and pickup/delivery stops.

Calculating the matched distance and the fastest path distance between the two pickup/delivery stops enables us to calculate a distance deviation. Retrieving timestamps for both the matched path and the shortest path allows us to calculate the time deviation. According to Dalla Chiara and

Goodchild (2020), Dalla Chiara et al. (2021), and Dalla Chiara et al. (2022), this allows us to evaluate parking behaviour. Both time deviation and distance deviation can be thought of as estimates of ‘cruising for parking’.

Referring to our literature review and to the best of our knowledge, little or no research has applied map-matching techniques on ‘cruising for parking’ analysis for commercial vehicles. Additionally, several studies, like Ghizzawi et al. (2024) and Dalla Chiara & Goodchild (2020), have mentioned a gap in the research on parking behaviour/searching for parking for commercial vehicles. Therefore, we see a potential to supplement ‘cruising for parking’ research on commercial vehicles by utilizing advanced map-matching techniques. In this context, map-matching techniques applied in the research focusing on non-commercial vehicles, might also be relevant to consider.

6 Conclusions

This report demonstrated the use of offline map-matching techniques, applied on GPS data from vans in the Oslo and Akershus counties, using PyTrack (an Open-Source Python-based map-matching tool). We present a framework for simplifying and guiding the process of map-matching and pre-processing steps, as well as challenges in applying the presented framework. We also provide a view of matched driving patterns and discuss advantages and disadvantages in applying different open-source map-matching tools. Furthermore, we explore how map-matching may form the basis for two different application areas: calculation of traffic work, and analyses of 'cruising for parking'.

Our framework illustrated how challenges such as GPS noise and data sparsity can be partially mitigated using an adaptive search radius, as well as the main steps in performing map-matching. Based on literature, we also applied filters on GPS points associated with stops and minor movements at stop locations, and emphasized the importance of removing such points - as some of them may suffer from jiggle effect (when the vehicle is idling) or pickup/deliveries located at parking lots/infrastructure that may not be present as road segments in OpenStreetMap. This pre-preprocessing step led to an exclusion of about 50% of all GPS points in our own analyses. Interestingly, the excluded GPS points only accounted for about 10% of the total Great Circle Distance and are likely even lower, due to jiggle effect.

Furthermore, we illustrated how matched paths may be plotted on a map and thereby provide insights into use frequencies of different roads. Map-matching resulted in a driving distance that was about 11% higher compared to straight-line GPS measurements. The size of this increase, however, may be greatly affected by the quality of the GPS points and important road turns that may not be captured by low-density GPS data. Additionally, it is important to note that the HMM algorithm assumes that vehicles always take the shortest path between matched candidates (Newson and Krumm, 2009). This assumption can lead to inaccuracies, particularly when the distance between GPS points is substantial (i.e., the Great Circle Distance). In such cases, significant information about the true path is lost, and the algorithm's assumption of the shortest path may not reflect the actual route taken by the vehicle.

Regarding different Open-Source map-matching tools, PyTrack and similar Python-based tools offer easy solutions for map-matching but may not meet the speed requirements for large-scale applications. High-performance platforms like OSRM, FMM and Valhalla provide faster processing, but require more complex setups.

Referring to our first suggested application area of map-matching, we reviewed different approaches modelling Vehicle kilometres travelled (VKT). Traditionally, a wide range of methods have been applied with the focus of measuring VKT. This includes Traffic Volume Counts, using fuel consumption to estimate VKT, Odometer Readings, or Household and Workplace Travel Surveys. However, one of the most promising innovations in this field is the integration of GPS data with digital road networks to enhance accuracy and reliability in VKT measurement (Fan et al., 2019). We attempted to illustrate how our matched paths from commercial vans in Oslo and Akershus form a basis for estimating VKT. VKT forms one of several application areas of map-matching. Compared with publicly available statistics collected using traditional methods, our approach highly deviates in terms of VKT. Our VKT estimation is therefore best suited as an example. High deviations are likely due to uncertain assumptions we make.

Referring to our second suggested application area of map-matching, we reviewed literature focusing on 'cruising for parking'. Our literature review revealed little or no existing research focusing on 'cruising for parking' analysis for commercial vehicles, using map-matching techniques. Generally, a lack of studies focusing on parking behaviour or searching for parking for commercial vehicles has been noted also by other researchers. On the other hand, focusing on non-commercial vehicles, we identified 7 articles using map-matching techniques. With this in mind, we see a potential to supplement 'cruising for parking' research on commercial vehicles by utilizing advanced map-matching techniques. In this context, techniques applied in the research on non-commercial vehicles might also be relevant to consider.

References

Alberini, A., Teja Burra, L., Cirillo, C., Shen, C., 2021. Counting vehicle miles traveled: What can we learn from the NHTS? Transportation Research Part D: Transport and Environment 98, 102984. https://doi.org/10.1016/j.trd.2021.102984

Amer, A., & Chow, J. Y. J. (2017). A downtown on-street parking model with urban truck delivery behavior. *Transportation Research Part A: Policy and Practice*, *102*, 51–67. https://doi.org/10.1016/j.tra.2016.08.013

Bierlaire, M., Chen, J. & Newman, J. (2013). A probabilistic map-matching method for smartphone GPS data. Transportation Research Part C: Emerging Technologies 26, 78–98.

Bisante, A., Panizzi, E., & Zeppieri, S. (2023). Cruising-for-Parking Detection on the Smartphone Based on Implicit Interaction and Machine Learning. *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 93–102.

Bonnel, P. & Munizaga, M.A. (2018). Transport survey methods - in the era of big data facing new and old challenges. Transportation Research Procedia 32, 1–15.

Bäumer, M., Hautzinger, H., Kuhnimhof, T., Pfeiffer, M., 2018. The German Vehicle Mileage Survey 2014: Striking the balance between methodological innovation and continuity. Transportation Research Procedia, Transport Survey Methods in the era of big data:facing the challenges 32, 329–338.

Chen, B.Y., Yuan, H., Li, Q., Lam, W.H., Shaw, S.L. & Yan, K. (2014). Map-matching algorithm for largescale low-frequency floating car data. International Journal of Geographical Information Science 28, 22–38. Publisher: Taylor & Francis.

Chen, F., Wu, J., Chen, X., Wang, J., 2017. Vehicle kilometers traveled reduction impacts of Transit-Oriented Development: Evidence from Shanghai City. Transportation Research Part D: Transport and Environment 55, 227–245. https://doi.org/10.1016/j.trd.2017.07.006

Dalla Chiara, G. & Goodchild, A. (2020). Do commercial vehicles cruise for parking? Empirical evidence from Seattle. Transport Policy, 97, 26–36. https://doi.org/10.1016/j.tranpol.2020.06.013

Dalla Chiara, G., Krutein, K. F., Ranjbari, A. & Goodchild, A. (2021). Understanding Urban Commercial Vehicle Driver Behaviors and Decision Making. Transportation Research Record, 2675(9), 608–619. https://doi.org/10.1177/03611981211003575

Dalla Chiara, G., Krutein, K. F., Ranjbari, A. & Goodchild, A. (2022). Providing curb availability information to delivery drivers reduces cruising for parking. Scientific Reports, 12(1), 19355. https://doi.org/10.1038/s41598-022-23987-z

Edward, M. (2014). Developing Roadway Performance Measures Using Commercial GPS Data from Trucks. *Institute of Transport Engineers. ITE Journal;* Jun 2014; 84; 6; Agricultural & Environmental Science Collection. pg 36.

Eisenmann, C., Kuhnimhof, T., 2018. Some pay much but many don't: Vehicle TCO imputation in travel surveys. Transportation Research Procedia, Transport Survey Methods in the era of big data:facing the challenges 32, 421–435. https://doi.org/10.1016/j.trpro.2018.10.056

Fan, J., Fu, C., Stewart, K., Zhang, L., 2019. Using big GPS trajectory data analytics for vehicle miles traveled estimation. Transportation Research Part C: Emerging Technologies 103, 298–307. https://doi.org/10.1016/j.trc.2019.04.019

Fu, M., Kelly, J.A., Clinch, J.P., 2017. Estimating annual average daily traffic and transport emissions for a national road network: A bottom-up methodology for both nationally-aggregated and spatially-disaggregated results. Journal of Transport Geography 58, 186–195. https://doi.org/10.1016/j.jtrangeo.2016.12.002

Ghizzawi, F., Galal, A. & Roorda, M. J. (2024). Modelling parking behaviour of commercial vehicles: A scoping review. Transport Reviews, 44(4), 743–765. https://doi.org/10.1080/01441647.2024.2305202

Gingerich, K., Maoh, H. & Anderson, W. (2016). Classifying the purpose of stopped truck events: An application of entropy to GPS data. Transportation Research Part C: Emerging Technologies 64, 17–27.Hakimelahi, A., Rao, K.V.K., Dhingra, S.L., Peiravian, F., 2020. Updating of a Travel Demand Model Based on Fuel Consumption Data. Transportation Research Procedia, Recent Advances and Emerging Issues in Transport Research – An Editorial Note for the Selected Proceedings of WCTR 2019 Mumbai 48, 1401–1417. https://doi.org/10.1016/j.trpro.2020.08.171

Hampshire, R. C., Jordon, D., Akinbola, O., Richardson, K., Weinberger, R., Millard-Ball, A. & Karlin-Resnik, J. (2016). Analysis of Parking Search Behavior with Video from Naturalistic Driving. *Transportation Research Record*, *2543*(1), 152–158. https://doi.org/10.3141/2543-18

Hovi, I.B., Mjosund, C.S., Pinchasik, D.R., Grønland, S.E., 2020. Smart data capture to reduce reporting burden, increase data quality in national truck surveys, and increase analysis capability. Presented at the 2020 Forum on Integrated and Sustainable Transportation Systems (FISTS), IEEE, Delft, South Holland Province, Netherlands, pp. 107–113. https://doi.org/10.1109/FISTS46898.2020.9264888

Hovi, I.B., Mjøsund, C.S., Bø, E., Pinchasik, D.R. & Grønland, S.E. (2021). Logistikk, miljø og kostnader: Kjøretøydata som grunnlag for forskning, transportplanlegging og forbedringsarbeid.

Huang, Z., Qiao, S., Han, N., Yuan, C.a., Song, X. & Xiao, Y. (2021). Survey on vehicle map-matching techniques. CAAI Transactions on Intelligence Technology 6, 55–71.

Hunter, P., Abbeel, A. & Bayen, A. (2014). The path inference filter: Model-based low-latency map-matching of probe vehicle data. IEEE Transactions on Intelligent Transportation Systems 15, 507–529.

Iwan, S., Kijewska, K., Johansen, B. G., Eidhammer, O., Małecki, K., Konicki, W. & Thompson, R. G. (2018). Analysis of the environmental impacts of unloading bays based on cellular automata simulation. *Transportation Research Part D: Transport and Environment*, *61*, 104–117. https://doi.org/10.1016/j.trd.2017.03.020

Jagadeesh, G.R., Srikanthan, T., Zhang, X.D., 2004. A Map Matching Method for GPS Based Real-Time Vehicle Location. The Journal of Navigation 57, 429–440. https://doi.org/10.1017/S0373463304002905

Jones, M., Khan, A., Kulkarni, P., Carnelli, P. & Sooriyabandara, M. (2017). ParkUs 2.0: Automated Cruise Detection for Parking Availability Inference. *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 242–251. https://doi.org/10.1145/3144457.3144495

Jung, S., Kim, Jounghwa, Kim, Jeongsoo, Hong, D., Park, D., 2017. An estimation of vehicle kilometer traveled and on-road emissions using the traffic volume and travel speed on road links in Incheon City. Journal of Environmental Sciences 54, 90–100. https://doi.org/10.1016/j.jes.2015.12.040

Koller, H., Widhalm, P., Dragaschnig, M. & Graser, A. (2015). Fast Hidden Markov Model Map-Matching for Sparse and Noisy Trajectories, pp. 2557–2561.

Liao, J. (2023). Optimization of Map-matching Algorithm in Various Road Conditions. Highlights in Science, Engineering and Technology 78, 59–66.

Millard-Ball, A., Hampshire, R. C. & Weinberger, R. R. (2019). Map-matching poor-quality GPS data in urban environments: The pgMapMatch package. *Transportation Planning and Technology*, *42*(6), 539–553. https://doi.org/10.1080/03081060.2019.1622249

Millard-Ball, A., Hampshire, R. C. & Weinberger, R. (2020). Parking behaviour: The curious lack of cruising for parking in San Francisco. *Land Use Policy*, *91*, 103918. https://doi.org/10.1016/j.landusepol.2019.03.031

Mun, H., Jung, J., 2024. Relationships between Vehicle Kilometers Traveled and Carbon Emissions based on Urban Size. KSCE Journal of Civil Engineering 100112. https://doi.org/10.1016/j.kscej.2024.100112

Mjøsund, C.S. & Hovi, I.B. (2022). GPS data as a basis for mapping freight vehicle activities in urban areas – A case study for seven Norwegian cities. Research in Transportation Business & Management 45, 100908.

Mahmud, S., Akter, T. & Hernandez, S. (2020). Truck Parking Usage Patterns by Facility Amenity Availability. Transportation Research Record, 2674(10), 749–763. https://doi.org/10.1177/0361198120937305

Mannini, L., Cipriani, E., Crisalli, U., Gemma, A. & Vaccaro, G. (2017). On-Street Parking Search Time Estimation Using FCD Data. *Transportation Research Procedia*, *27*, 929–936. https://doi.org/10.1016/j.trpro.2017.12.149

Mantouka, E. G., Fafoutellis, P. & Vlahogianni, E. I. (2021). Deep survival analysis of searching for on-street parking in urban areas. *Transportation Research Part C: Emerging Technologies*, *128*, 103173. https://doi.org/10.1016/j.trc.2021.103173

Martens, K., Benenson, I. & Levy, N. (2010). The Dilemma of On-Street Parking Policy: Exploring Cruising for Parking Using an Agent-Based Model. I B. Jiang & X. Yao (Red.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics* (s. 121–138). Springer Netherlands. https://doi.org/10.1007/978-90-481-8572-6_7

Milia, K., Hedengran, M. D., Jansson, T., Rodrigues, F. & Azevedo, C. L. (2023). Forecasting Parking Search Times Using Big Data. I C. Antoniou, F. Busch, A. Rau, & M. Hariharan (Red.), *Proceedings of the 12th International Scientific Conference on Mobility and Transport* (s. 175–197). Springer Nature. https://doi.org/10.1007/978-981-19-8361-0_12

Miwa, T., Kiuchi, D., Yamamoto, T. & Morikawa, T. (2012). Development of map-matching algorithm for low frequency probe data. Transportation Research Part C: Emerging Technologies 22, 132–145.

Nevland, E. A., Gingerich, K. & Park, P. Y. (2020). A data-driven systematic approach for identifying and classifying long-haul truck parking locations. *Transport Policy*, *96*, 48–59. https://doi.org/10.1016/j.tranpol.2020.04.003

Newson, P. & Krumm, J. (2009). Hidden Markov map-matching through noise and sparseness, in: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Association for Computing Machinery, New York, NY, USA. pp. 336–343.

Paidi, V., Håkansson, J., Fleyeh, H. & Nyberg, R. G. (2022). CO2 Emissions Induced by Vehicles Cruising for Empty Parking Spaces in an Open Parking Lot. *Sustainability*, *14*(7), Art. 7. https://doi.org/10.3390/su14073742

Plaudis, M., Azam, M., Jacoby, D., Drouin, M.A. & Coady, Y. (2021). An Algorithmic Approach to Quantifying GPS Trajectory Error, in: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), IEEE, Montreal, BC, Canada. pp. 3902–3909.

Quddus, M. & Washington, S. (2015). Shortest path and vehicle trajectory aided map-matching for low frequency GPS data. Transportation Research Part C: Emerging Technologies 55, 328–339.

Rahmani, M. & Koutsopoulos, H.N. (2013). Path inference from sparse floating car data for urban networks. Transportation Research Part C: Emerging Technologies 30, 41–54.

Ranjbari, A., Goodchild, A., McCormack, E., Verma, R., Hurwitz, D., Liu, Y. & Jashami, H. (2023). *INSIGHTS FROM COMMERCIAL DRIVER PARKING DECISIONS IN A TRUCK SIMULATOR TO INFORM CURB MANAGEMENT DECISIONS.* University of Washington & Oregon State University. https://urbanfreightlab.com/wp-content/uploads/2023/04/Truck-Simulator-for-Curb-Mngt-Decisions.pdf

Saki, S. & Hagen, T. (2022). A Practical Guide to an Open-Source Map-Matching Approach for Big GPS Data. SN Computer Science 3, 415.

Saki, S. & Hagen, T. (2024). Parking search identification in vehicle GPS traces. Journal of Urban Mobility, 6, 100083. https://doi.org/10.1016/j.urbmob.2024.100083

Seya, H., Zhang, J., Chikaraishi, M. & Jiang, Y. (2020). Decisions on truck parking place and time on expressways: An analysis using digital tachograph data. Transportation, 47(2), 555–583. https://doi.org/10.1007/s11116-018-9899-y

Sharman, B. W., Roorda, M. J. & Habib, K. M. N. (2012). Comparison of Parametric and Nonparametric Hazard Models for Stop Durations on Urban Tours with Commercial Vehicles. Transportation Research Record, 2269(1), 117–126. https://doi.org/10.3141/2269-14

Tortora, M., Cordelli, E. & Soda, P. (2022). PyTrack: A Map-Matching-Based Python Toolbox for Vehicle Trajectory Reconstruction. IEEE Access 10, 112713–112720.

van der Waerden, P., Timmermans, H. & Van Hove, L. (2015). GPS Data and Car Drivers' Parking Search Behavior in the City of Turnhout, Belgium. I I. Ivan, I. Benenson, B. Jiang, J. Horák, J. Haworth, & T. Inspektor (Red.), *Geoinformatics for Intelligent Transportation* (s. 247–256). Springer International Publishing. https://doi.org/10.1007/978-3-319-11463-7_18

Weinberger, R. R., Millard-Ball, A. & Hampshire, R. C. (2020). Parking search caused congestion: Where's all the fuss? *Transportation Research Part C: Emerging Technologies*, *120*, 102781. https://doi.org/10.1016/j.trc.2020.102781

Yang, C. & Gidòfalvi, G. (2018). Fast map-matching, an algorithm integrating hidden Markov model with precomputation. International Journal of Geographical Information Science 32, 547–570.

Zhang, D., Guo, Z., Guo, F. & Dong, Y. (2021). An offline map-matching algorithm based on shortest paths. International Journal of Geographical Information Science 35, 2238–2261.

Zhang, L., van Lierop, D., Ettema, D., 2025. The effect of electric vehicle use on trip frequency and vehicle kilometers traveled (VKT) in the Netherlands. Transportation Research Part A: Policy and Practice 192, 104325. https://doi.org/10.1016/j.tra.2024.104325

Appendix

Hidden Markov Model map-matching

Newson and Krumm (2009) introduced and described a novel map-matching methodology by using a HMM and the Viterbi algorithm. Their approach involves identifying the most likely road segments that could have produced the GPS points. More specifically, given a road segment r_i and a GPS point z_t, their algorithm locates the closest point on the nearest road segments by measuring the Great Circle Distance between z_t and r_i. This results in a candidate point for that GPS point, which the authors refer to as $x_{t,i}$. The candidates can also be thought of as hidden states.

For each road segment corresponding to a GPS point, Newson and Krumm (2009) calculate an emission probability, $p(z_t|r_i)$, that tells us the likelihood of observing a GPS point z_t, given that the vehicle was actually traversing on road segment r_i. In practice, the road segments are determined by candidates located at that road segment. For each GPS point z_t, there can be multiple candidates $x_{t,i}$, each corresponding to their own road segment r_i. However, Newson and Krumm (2009) emphasize that we in practice only consider road segments that are within a reasonable reach of the GPS points. By setting the emission probability to zero for any road segments longer away from z_t than a given threshold, they reduced the number of candidate matches. This in turn sped up their algorithm by having to consider fewer candidates.

Newson and Krumm (2009) describe transition probabilities as the probability of a vehicle moving between two candidates (candidate locations), at two consecutive t times. The transition probability takes into account the difference in Great Circle Distance between two consecutive GPS points and the respective shortest path distance between two consecutive candidates. Since each GPS points may involve multiple candidates, the shortest path distance is calculated for each combination of consecutive candidates (i.e. at time t and t+1) and compared to the Great Circle Distance. Then, the smaller the differences between the Great Circle Distance and the shortest path distances, the higher the transition probability. The final goal of HMM is to find the path of candidates (hidden states) that maximizes the product of emission and transition probabilities.

In our report, we have used PyTrack[17], which is an Open-Source Python-package for HMM map-matching. PyTrack used the Viterbi-algorithm for finding the path that maximizes emission and transition probabilities (Tortora et al., 2022).

[17] https://github.com/cosbidev/PyTrack

TØI is an applied research institute that carries out research and study assignments for businesses and public agencies. TØI was established in 1964 and is organized as an independent foundation. TØI develops and disseminates knowledge about transport with scientific quality and practical application. The department has an interdisciplinary environment with 90+ highly specialized researchers.

The department conducts research dissemination through TØI reports, articles in scientific journals, books, seminars, as well as posts and interviews in the media. The TØI reports are available free of charge on the department's website www.toi.no.

TØI covers all means of transport and thematic areas within transport, including traffic safety, public transport, climate and environment, tourism, travel habits and travel demand, urban planning, ITS, public decision-making processes, business transport and general transport economics. The institute participates actively in international research collaboration, with particular emphasis on the EU framework programs.

The Department of Transport Economics requires copyright for its own work and emphasizes acting independently of the clients in all professional analyses and assessments.

Postal Address:
Institute of Transport Economics
Postboks 8600 Majorstua
N-0349 Oslo
Norway

Email: toi@toi.no

Business Address:
Forskningsparken
Gaustadalléen 21

Web address: www.toi.no